Situating Data

Inquiries in Algorithmic Culture

Edited by
Karin van Es and
Nanna Verhoeff

Amsterdam University Press

MediaMatters

MediaMatters is an international book series published by Amsterdam University Press on current debates about emerging and transforming cultural practices that engage with (new) media technologies. Contributions to the series critically analyse and theorise the materiality, spatiality, mobility and performativity of these practices in book projects that engage with today's dynamic digital media culture.

MediaMatters focuses on objects and practices such as: installation art; (digital) performance; site-specific theater; time-based art; experimental film and video; digital and new media art; motion capture; telematics; looping media and digital GIFs; glitch media; cybernetics, robots and AI; virtual reality, augmented and mixed reality; screen media; interactive media, haptic/tactile media; mobile media; tactical media; ecological art and media; media architecture; new museum and exhibition practices.

Cover illustration: Getty Images, 2020

Cover design: Coördesign, Leiden
Lay-out: Crius Group, Hulshout

ISBN 978 94 6372 297 1
e-ISBN 978 90 4855 544 4
DOI 10.5117/9789463722971
NUR 670

Table of Contents

List of Figures

Acknowledgements

In 2021 we had a clear task at hand: to develop two new critical data studies tracks for the teaching programs offered by the Department of Media and Culture Studies within the Faculty of Humanities at Utrecht University. Many of our colleagues are working on questions regarding the impact of datafication and algorithmization on our culture from various disciplinary fields within media and culture studies, but this had not yet included a shared conversation about the common ground of these approaches as specific to, as well as for, cultural inquiry. Indeed, what is the unique contribution of media and culture studies to the emerging field of critical data studies? But also, the other way around: what role do these cultural transformations of datafication and algorithmization play in the theoretical and methodological development of cultural inquiry itself? This sparked the ambition for a book that positions this double perspective and that showcases the richness, diversity, and productivity of situating data *in* cultural inquiry.

First and foremost, we want to thank our inspiring colleagues who enthusiastically participated in brainstorming and discussions that helped shape the collection before you. We are very thankful for the support of the Centre for Digital Humanities, the Research Institute for Cultural Inquiry (ICON), and the Utrecht University Library's Open Access Fund. The financial support of the latter enabled us to publish this book in open access.

We want to thank Maryse Elliott and the editorial board of Amsterdam University Press's *MediaMatters* book series for their support. A special thanks to the anonymous reviewers of our manuscript, and to Eggo Müller, Eef Masson, and Iris van der Tuin for providing feedback on the collection as a whole and on the introduction in particular. The book would not have been possible without the support and efforts of our editorial assistant Chris van der Vegt, who provided invaluable feedback and helped lighten the administrative burden.

We dedicate this book to past, present, and future students. Their inquisitive and critical minds keep the field of cultural inquiry looking sideways and forward.

Karin van Es & Nanna Verhoeff
Amsterdam, Utrecht, Rotterdam, July 2022

Introduction: Situating Data as Cultural Inquiry

Nanna Verhoeff and Karin van Es

Abstract
This collection examines the impact of data, datafication, and algorithmization on contemporary culture, and thereby also on the agenda of the broader field of cultural inquiry. Our perspective is double. We ask: Where and how do data and algorithms shape and transform culture? But also, where and how do they impact and transform scholarly practices in the study of culture? Situating data as cultural inquiry, thus, is not only an act of localizing data both *in* and *as* culture, but also an act of situating our perspective on, and knowledge about, this culture. With the selection of chapters in this collection, contributing to the diversification of cultural inquiry, we propose conceptual and methodological directions for exploring where, when, and how data and algorithms (re)shape cultural practices, create (in)justice, and produce knowledge.

Keywords: cultural inquiry, situatedness, performativity, materiality, criticality, creativity

The datafication and algorithmization of culture and society transforms and expands the field of cultural inquiry. Both the objects of study as well as our scholarly practices—questions, approaches, and methods—change. Alongside new objects and questions, our traditional "objects" of cultural inquiry are increasingly digitized, mediatized, and given lives as data (Posner and Klein 2017). Indeed, data and algorithms are fundamentally cultural. On the one hand, they (re)produce ideologies, values, and beliefs (Beer 2019), and on the other hand, in processes of selection, abstraction, and translation, data are shaped by material, historical, and discursive conditions and restraints. While the processes of datafication are not always visible, their consequences

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_INTRO

have far-reaching material and discursive effects. These unfold on various levels and scales—on the level of individual, everyday life; on community and institutional levels; and on the level of our planetary ecosystem.

A proliferation of new conceptual terms articulates perspectives on contemporary technological, social, and cultural transformations. The "computational turn" (Berry 2012) or "algorithmic turn" (Uricchio 2011) unleashed a "data revolution" (Kitchin 2022) and contribute to the advent of what we can discern as the "algorithmic condition" (Colman et al. 2018) of contemporary culture that shapes our "Datafied Society" (Schäfer and van Es 2017). These terms are accompanied by critical evaluations of, and creative interventions in, the fundamental consequences of datafication and algorithmization of culture and society. Critical scholarship has called out the logic of "surveillance capitalism" in the networked sphere (Zuboff 2015) and, perhaps also more fundamentally, forms of "data colonialism" and the "new social order" resulting from the appropriation of human life through data (Couldry and Mejias 2019). Creative responses from new fields such as the creative humanities, artistic research, critical making, and research by design have experimented not only with making visible and debatable the presence of data-driven processes and their direct impact on our environments and ways of living, but also with proposing or designing alternative interfaces to data for, and in, public spaces (Verhoeff et al. 2019).

These critical statements and creative interventions that address the cultural impact of data and algorithms have a shared agenda. They signal radical change, articulate questions and concerns, and call for new perspectives by urgently asking to take seriously the socio-material, epistemological, and ecological implications of the ongoing fundamental changes caused by the "deep mediatization" (Hepp 2020) of almost all societal realms. How can we understand the quality and significance of current socio-technical transformations that result from datafication and algorithmization? How can we explore the changing conditions and contours for living within such new and changing frameworks? How can, or should we, think and act within, but also in response to, these conditions? What are the ethics they call for? Or, to put it in Harawayan's (2016) terms, what is our response-ability?

Cultural Inquiry as Perspective

Taking up the challenges of the datafication of culture, as well as of the scholarship of cultural inquiry itself, this collection contributes to the critical debate about data and algorithms by engaging with these bigger and

more fundamental questions from a specifically situated perspective. This perspective locates data not in abstract terms as "out there," unreachable, invisible, and immaterial, yet ubiquitous and all-pervasive, but instead as fundamentally situated: present, emergent, and relational. Herein it pushes back against claims of data as "big" and "universal" and conceptual metaphors such as "cloud," "liquid," and "fuel" that encode how we make sense of data (Puschmann and Burgess 2014). Such claims and metaphors obfuscate the socio-political, ecological, and epistemological realities that situated data practices entail. As such it risks reproducing structural biases, blind spots, and inequalities.

However, as feminist science and technology studies scholar Donna Haraway (1988) has pointed out, a situated perspective prompts us to look both ways: at the object as well as subject of study. For our inquiry here, this entails locating data as embedded and embodied in cultural practices, but it also critically and self-reflexively situates us as researchers of those practices as also embedded and embodied within the same contemporary, datafied, and algorithmicized culture. Situating data, thus, is not only an act of localizing data both *in* and *as* culture, but it is also an act of situating our perspective on, and knowledge about, data. Such a perspective, we hold, is much needed for understanding how we think, practice, and conceptualize data in our contemporary culture and society, but it is also needed to conceive of possible interventions or ways to reshape these practices. Data are, thus, produced by culture and culture-producing. Taking data seriously as a question of culture—an approach also inherently embedded within this culture—then requires a situated, reflexive, critical, and a productive (or, creative) perspective. This collection responds to this demand by collectively formulating and demonstrating the value of such a perspective for, and approach to, data as *cultural inquiry*. Exposing the realities of situated data practices can be a first step toward meaningful intervention and change with the aim of a more responsible and just datafied society.

As digital humanities scholar Ted Underwood (2018) points out, the current age of datafication, algorithms, and machine learning needs the humanities, specifically for a traditional skepticism about numbers, but also for their ability to contextualize and deal with complexity (related to their cultural and critical analyses). Digital humanities scholars Miriam Posner and Lauren Klein (2017, 3) similarly contend that the field of media studies has particular sensitivities and concerns that enrich the broader field of data studies. Media theory brings in a perspective on medium specificity that contributes to new methods of critiquing data sets as produced both by technologies and cultural frameworks (cf. Poirer 2020) and the close

reading of code (cf. Marino 2006, 2020). Conceptualizations of performativity (Barad 2007; Drucker 2013) offer directions for thinking about and with the generative processes of data. These build (on) perspectives on the materiality, mediality, and performativity of data, and bring attention to the particular contexts—and thus politics—from which data and algorithms emerge. Moreover, the study of pressing societal issues of these times (e.g., fake news, polarization, predictive policing, surveillance, market domination) should not be left to the sciences only but requires a fundamental and deep collaboration with the humanities (Parks 2020). In a special issue on the study of Artificial Intelligence and machine cultures for the journal *Media, Culture & Society*, the editors stress the relevance of media and cultural studies specifically. As they claim, these fields within the humanities offer three important correctives to current debates as they help to question the notion of the "human," contextualize machine cultures, and support ethical and responsible AI practices.

As an introduction to this collection, in the following we sketch the contribution of cultural inquiry in situated reflexive, critical, and productive approaches to data and algorithms. Fundamentally non-exhaustive and selective, we zoom in on inquiries specifically informed by the sensitivities of media studies. These sensitivities guide the book's inquiries into how processes of datafication and algorithmization shape practices, create (in)justice, produce knowledge, and call for new research agendas and methodological directions for cultural inquiry. As outlined below, these are centered on questions, concepts, and debates around the materiality, power, and affect of data.

Approaching Data Critically

The impact of data and algorithms has contributed to the re-emergence of empiricism in science, policymaking, and public opinion and debate. It is grounded in a set of false assumptions—perpetuated by the metaphors mentioned earlier, e.g., big, universal, cloud, liquid, fuel—that data can be exhaustive, transparent, and unbiased and that it they capture a whole domain, arise from "nowhere" without agenda, are generated independently, and can speak for themselves (Kitchin 2022, 115–17). However, their etymological root in Latin notwithstanding, *data* are never really "given" without some trade-off. Or, to emphasize the act *before* data: data are always already framed when sought, and never simply "raw" (Bowker 2005), to be found "out there." Moreover, data are also not just a result *of* but are also

resulting *in*: they have a past, a present, and a future. Media theorists Lisa Gitelman and Virginia Jackson formulated this eloquently: "Data need to be imagined *as* data to exist and function as such, *and that imagination of data entails an interpretive base*" (2013, 3; emphasis added). While such critiques immediately bubbled to the surface, they remained scattered at first.

Calls for a more systematic approach to data criticism followed. Within media studies, danah boyd and Kate Crawford were early to forward an agenda for the critical study of big data in their seminal article "Critical Questions for Big Data" (2012). Within critical geography, Craig Dalton and Jim Thatcher (2014) provided an initial agenda for what they termed Critical Data Studies. Since then, Critical Data Studies has developed as an interdisciplinary field drawing input from anthropology, law, information studies, political sciences, (new) media studies, and gender studies. It has played an important role in demystifying myths of accuracy and objectivity, exposing data as always-also partial, selective, and biased. It has tasked itself with exploring the ethical, cultural, and critical challenges posed by so-called "Big Data" (Iliadis and Russo 2016) and raising questions about the generation, circulation, and deployment of data by charting and unpacking more complex data assemblages (Kitchin and Lauriault 2018).

While Critical Data Studies is a blossoming and rich field, digital media scholar Yanni Alexander Loukissas sees limits to the critical reflection at the core of most branches, which he finds "can be detached rather than responsible, analytic rather than affective, or conceptual rather than hands on" (2019, 9). His work on the *locality* of data is a direct response to this issue. Inspired by feminist ethics of care, he embraces material engagement and affect and calls attention to neglected things. Apart from these aspects related to criticality, geographer and specialist in the critical study of the datafication of (smart) cities Rob Kitchin (2022, 302–6) urges for "decentering data" in critical data studies. It entails understanding data and its assemblages as constructions and expressions of society and culture. This, he finds, can be recognized in the work of feminist and critical race scholarship, but it is not common in most data studies. As such, the approach to data as culture and the collection of cultural inquiries in this collection are a productive contribution to and expansion of the field of, and discussions within, the critical study of data. For an approach to data-as-culture, the aim is not so much to understand what data *are*, but rather how they came to be and what they *do* as they are entangled with algorithms. Indeed, data and algorithms are operational in that they influence our lifeworld and are woven into everyday practices (Loukissas 2020; Rettberg 2020). This brings to the fore the *performativity* of data – specifically in a focus on data

practices and the socio-political, ecological, and epistemological conditions for, and implications of, such practices. Analyzing data and algorithms as performative entails a critical unpacking of how they shape, define, and maintain the world.

In contemporary critical approaches to data, we identify three recurring and interrelated foci of questions, concepts, and debates around the materiality, power, and affect of data that are pertinent to data as a cultural inquiry. Regarding materiality in relation to power, important work has, for instance, been produced in critical algorithm studies (Bucher 2012; Gillespie, 2014; Noble 2018; Pasquale 2015; Rieder 2015) that examines the social and political dimensions of algorithms. It has resulted in insightful contributions like that of Safiya Noble in *Algorithms of Oppression* (2018), which explore how search engines reinforce racism, and in *Automating Inequality* (2017), in which Virgina Eubanks exposes how digital tracking and automated decision-making profile, police, and punish the poor. Similar concern for the materiality of media can be traced to work in the field of software studies (Fuller 2003, 2008; Manovich 2001, 2013), with conceptual origins in the work of figures such as Harold Innis (1951) and Marshall McLuhan (1964).

Software studies has become a field concerned with analyzing the social and cultural impact of software systems and now also includes critical work done in interface studies (Chamberlain 2011; Galloway 2012; Stanfill 2015). It is closely related to platform studies (Monfort and Bogost 2009), which is concerned with interrogating the relationships between hardware and software. These fields share an interest in the *materiality* of media technologies and push back against metaphors and imaginaries of data as immaterial. This scholarly work is important because examining data as material "draws specific attention to the historical particularities, cultural specificities, and political consequences" (Dourish and Mazmanian 2013, 4) of data. Indeed, data are caught up in complex socio-technical assemblages. It is necessary to attend both to materiality and to representational and rhetorical aspects of data.

With its analyses of such assemblages, platform studies have moved away from a game studies perspective to include content sharing websites and social media applications (Plantin et al. 2018). This latter branch of platform studies (Gillespie 2010; Langlois and Elmer 2013; van Dijck 2013) has generated a spinoff in what could be referred to as platformization studies (Helmond 2015; Nieborg and Poell 2018; Poell, Nieborg, and Duffy 2022), which examines the extension of digital platforms into different spheres of life and cultural production.

Departing from a similar concern with the material, media studies saw a surge in infrastructure studies in part through the influential work of Brian Larkin (2008) and Lisa Parks and Nicole Starosielski (2015). They are concerned with the intersection of everyday life with the material artifacts of media infrastructures. The study of infrastructure has also found its bearing within digital media studies as digital platforms increasingly operate as infrastructures (Plantin et al. 2018; Plantin and Punathambekar 2019). A particular strand of infrastructure studies is concerned with the harm of data and AI to our environment, further entrenching inequality and acting as a driving force for undemocratic governance. The book *Atlas of AI* (2021) by Kate Crawford is exemplary hereof. While these studies depart from the materiality of data, they ultimately bring this to bear on questions of governance, ownership, and business models. This research results in interest for critical questions pertaining to issues of power such as labor, discrimination, exploitation, and surveillance. We now also increasingly encounter studies exploring the adverse environmental (material) effects of streaming media (Lobato 2019; Marks et al. 2020) and prompts for intensifying such concerns in environmental media studies or green(ing) media studies (Keilbach and Pabiś-Orzeszyna 2022).

Questions of power are also central in scholarship that calls for decentering data universalism and a de-Westernization of critical data studies (Milan and Treré 2019). This connects to concerns addressed in the research agenda of data activism (Milan and van der Velden 2018), data justice (Dencik et al. 2019), and data feminism (D'Ignazio and Klein 2020). While data activism is concerned with analyzing and intervening in socio-technical practices that critique big data collection, data justice focuses on social justice in data-driven systems and big data, and data feminism provides principles for a data science informed by feminist theory and activism.

Beside an initial focus on the material and discursive power of data and its consequences (Kennedy 2018) on a larger societal scale, new research agendas have been put forward that are looking at the "street level" (McCullough 2013) of digitization, datafication, and algorithmization. This encompasses, for instance, an interest in data's "mundane" everyday experiences (Pink et al. 2017), its embeddedness in everyday practices of "vernacular data cultures" (Burgess et al. 2022), and an understanding of living with data "from the bottom up" (Couldry and Powell 2014). As a result of this research, there have been calls for attention to the affective dimensions of data (Kennedy 2018; Lupton 2017). This interest in data practices and affects builds on a longer tradition in media studies and the wider field of cultural inquiry to examine culture in and through everyday practices and experiences.

Cultural Inquiries, Plural

The digitization and datafication of culture and society has created new opportunities for research in the humanities and given rise to various computational approaches (cf. Dobson 2019; D'Ignazio and Klein 2020; Lindgren 2020; Rogers 2013). However productive these distant approaches can be for mapping phenomena and isolating patterns, this collection examines the close encounter between researcher, concept, and object. It entails the fundamental question of how we can do research on the situatedness of data while simultaneously considering our own situated position within a datafied society. Such a doubly situated and situating approach to data requires methods that build on the ethical commitment of critical humanities scholarship (Barad 2007; Geerts and Carstens 2019; Verhoeff and van der Tuin 2020) as well other disciplines that acknowledge data as also small, partial, local, material, and embodied. In line with this approach, this collection explores some areas within media studies, cultural theory, gender and postcolonial studies, and philosophy of science that study and work with data as part of the larger project of cultural inquiry. These entails various qualitative methods, situated readings, and creative methods that inquire into the material, power-full, and affective dimensions of data and data practices. It pushes forward the project of situated and "local readings" of data (Loukissas 2019). Such local readings immediately demonstrate how power, however historically determined, plays out differently in different contexts.

The contributions to this collection all respond to what we may consider the challenging "newness" of data for cultural inquiry and, specifically, its subfields of media studies. The datafication and algorithmization of culture not only demands a fundamental (re)assessment of our research objects and research practices, but also of how these processes have a fundamental impact on the research agenda, conceptual vocabulary, and methodological scope of cultural inquiry itself. As the chapters in this collection demonstrate, together this emergence of "data" as its own object of cultural inquiry entails not only a broad scoping of new "objects"—e.g., interstitial data and algorithmic serendipity—but also an articulation and mobilization of variously shaded new or revisited concepts and methods that help to take seriously the ecological, ethical, and epistemological implications of datafication and algorithmization as a pluriform and tentacular cultural process. As such, datafication and algorithmization prompts scholars to position and redefine themselves, in the process blurring and redrawing disciplinary boundaries.

Above we have sketched the way datafication involves cultural transformations that prompt questions about interrelated aspects and implications of materiality, power, and affect. These themes are central and recurring throughout this book. This collection brings together various perspectives on the datafication and algorithmization of culture from debates and disciplines within the field of cultural inquiry, specifically (new) media studies, game studies, urban studies, screen studies, gender studies, and postcolonial studies. The contributions are clustered around the aforementioned foci of *practices*, *justice*, and *knowledge* and an additional section on *agendas*, which explores future lines of research and new methods. Each section shows theoretical and conceptual tools for examining and understanding these aspects of data in contemporary culture as manifold and divergent but also connected, while also demonstrating links and partial overlap with the other sections. Therefore, we should speak of cultural inquiries, plural.

The first three chapters of section one investigate various contexts in which data are imagined and practiced: these include coffee roasting, precision farming, and urban design. The following two chapters—one on video streaming platforms and the other on electronic waste—explore the implications that the reliance of their articulation and circulation on material platforms and infrastructures has on the environment.

From these environmental concerns about waste and the impact of streaming, we shift attention to questions of (in)justice embedded in, and resulting from, working with data and algorithms. Section two is concerned with the "fairness in the way people are made visible, represented and treated as a result of their production of digital data" (Taylor 2017). Here, the focus is on different forms of inclusion and exclusion that are maintained or challenged by datafication, including datafication as boundary work, and the logics of race in governmental data systems. The last two articles address the role of art in uncovering the politics of facial recognition and Afrofuturist activism. These inquiries explore possible approaches and routes to "break" with problematic, oppressive, and exclusionary norms and assumptions that often infuse our data technologies and practices.

In the third section, four chapters examine the impact that data and algorithms have as part of the media technologies we work with for, and in, practices of knowledge production. This is illustrated through an analysis of knowledge production with metadata as active agents, a tracing of metadata in archival search systems, a proposal for diachronic affordance analysis, and an analysis of adaptive learning platforms. These contributions underscore the symbiotic relationship between data, algorithms, and knowledge. A

recurring question in these chapters is how principles of mediation and translation occur in the complex assemblages of human and non-human actors and how this has fundamental epistemic consequences.

The last section of this collection contains several shorter statements that each, from various perspectives, provide an outlook on new research directions, approaches, and methods—including new collaborative and interdisciplinary ways of working. These entail a call for entrepreneurial research, creative methods, collaboration and dialogue between the Global North and Global South, and an ecosystemic approach to data. With these chapters and these mission statements, we hope to have signaled and gestured toward future directions for cultural inquiries into data.

References

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.

Beer, David. 2019. *The Data Gaze: Capitalism, Power and Perception*. London: SAGE.

Berry, David M. 2012. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, edited by David M. Berry, 1–20. New York: Palgrave Macmillan.

Bowker, Geoffrey C. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, And Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5: 662–79.

Bucher, Taina. 2012. "Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook." *New Media & Society* 14, no. 7: 1164–80. https://doi.org/10.1177/1461444812440159.

Burgess, Jean, Kath Albury, Anthony McCosker, and Rowan Wilken. 2022. *Everyday Data Cultures*. Cambridge, UK: Polity Press.

Chamberlain, Daniel. 2011. "Scripted Spaces: Television Interfaces and the Non-Places of Asynchronous Entertainment." In *Television as Digital Media*, edited by James Bennett and Niki Strange, 230–54. Durham, NC: Duke University Press.

Colman, Felicity, Vera Bühlmann, Aislinn O'Donnell, and Iris van der Tuin. 2018. *Ethics of Coding: A Report on the Algorithmic Condition* [*EoC*]. Brussels: European Commission. https://cordis.europa.eu/project/rcn/207025_en.html.

Couldry, Nick, and Alison Powell. 2014. "Big Data from the Bottom Up." *Big Data & Society* 1, no. 1: 1–5. https://doi.org/10.1177/2053951714539277.

Couldry, Nick, and Ulises A. Mejias. 2019. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television and New Media* 20, no. 4: 336–49. https://doi.org/10.1177/1527476418796632.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Dalton, Craig, and Jim Thatcher. 2014. "What Does a Critical Data Studies Look Like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data.'" *Society and Space*. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care.

Dencik, Lina, Arne Hintz, Joanna Redden, and Emiliano Treré. 2019. "Exploring Data Justice: Conceptions, Applications and Directions." *Information, Communication & Society* 22, no. 7: 873–81. https://doi.org/10.1080/1369118X.2019.1606268.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.

Dobson, James. 2019. *Critical Digital Humanities: The Search for a Methodology*. Urbana-Champaign, IL: University of Illinois Press.

Dourish, Paul, and Melissa Mazmanian. 2013. "Media as Material: Information Representations as Material Foundations for Organizational Practice." In *How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies*, edited by Paul R. Carlile, Davide Nicolini, Ann Langley, and Haridimos Tsoukas, 92–118. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199671533.003.0005.

Drucker, Johanna. 2013. "Performative Materiality and Theoretical Approaches to Interface." *DHQ: Digital Humanities Quarterly* 7, no. 1. http://digitalhumanities.org:8081/dhq/vol/7/1/000143/000143.html.

Eubanks, Virginia. 2017. *Automating Inequality*: *How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press.

Fuller, Matthew. 2003. *Behind the Blip: Essays on the Culture of Software*. Brooklyn, NY: Autonomedia.

Fuller, Matthew. 2008. *Software Studies: A Lexicon*. Cambridge, MA: MIT Press.

Galloway, Alexander. 2012. *The Interface Effect*. Cambridge, UK: Polity Press.

Geerts, Evelien, and Delphi Carstens. 2019. "Ethico-onto-epistemology." *Philosophy Today* 63, no. 4: 915–25. https://doi.org/10.5840/philtoday202019301.

Gitelman, Lisa, and Virginia Jackson. 2013. "Introduction." In *"Raw Data" Is an Oxymoron,* edited by Lisa Gitelman, 1–14. Cambridge, MA: MIT Press.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms.'" *New Media & Society* 12, no. 3: 347–64. https://doi.org/10.1177/1461444809342738.

Gillespie, Tarleton. 2014. "Algorithm [draft] [#digitalkeywords]." *Culture Digitally* (blog), June 25, 2014. http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword.

Haraway, Donna J. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3: 575–99. https://doi.org/10.2307/3178066.

Haraway, Donna J. 2016. *Staying with the Trouble: Making Kin in the Chthulucene.* Durham, NC: Duke University Press. https://doi.org/10.1215/9780822373780.

Helmond, Anne. 2015. "The Web as Platform: Data Flows in Social Media." PhD diss., University of Amsterdam.

Hepp, Andreas. 2020. *Deep Mediatization*. New York: Routledge.

Iliadis, Andrew, and Federica Russo. 2016. "Critical Data Studies: An Introduction." *Big Data and Society* 3, no. 2 (December). https://doi.org/10.1177/2053951716674238.

Keilbach, Judith, and Michał Pabiś-Orzeszyna. 2021. "Green(ing) Media (Studies)." In *NECSUS_European Journal of Media Studies* 10, no. 2: 105–12. https://doi.org/10.25969/mediarep/17278.

Kennedy, Helen. 2018. "Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication." *Krisis: Journal for Contemporary Philosophy* 1: 18–30.

Kitchin, Rob. 2022. *The Data Revolution. A Critical Analysis of Big Data, Open Data and Data Infrastructures*. London: SAGE.

Kitchin, Rob, and Tracey Lauriault. 2018. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." In *Thinking Big Data in Geography*, edited by Jim Thatcher, Josef Eckert, and Andrew Shears, 3–20. Lincoln, NE: University of Nebraska Press.

Langlois, Ganaele, and Greg Elmer. 2013. "The Research Politics of Social Media Platforms." *Culture Machine* 14: 1–17.

Larkin, Brian. 2008. *Signal and Noise: Media, Infrastructure, and Urban Culture in Nigeria*. Chapel Hill, NC: Duke University Press.

Lindgren, Simon. 2020. *Data Theory.* Cambridge, UK: Polity Press.

Liu, Alan. 2017. "Toward Critical Infrastructure Studies." Paper presented at the University of Connecticut, Storrs, CT, February 23, 2017.

Lobato, Ramon. 2019. *Netflix Nations: The Geography of Digital Distribution.* New York: NYU Press

Loukissas, Yanni Alexander. 2020. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

Lupton, Deborah. 2017. "Feeling Your Data: Touch and Making Sense of Personal Digital Data." *New Media & Society* 19, no. 10: 1599–614.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: MIT Press.

Manovich, Lev. 2013. *Software Takes Command*. New York: Bloomsbury.

Marino, Mark C. 2006. "Critical Code Studies." *Electronic Book Review.* https://electronicbookreview. com/essay/critical-code-studies.

Marino, Mark C. 2020. *Critical Code Studies*. Cambridge, MA: MIT Press.

Marks, Laura U., Joseph Clark, Jason Livingston, Denise Oleksijczuk, and Lucas Hilderbrand. 2020. "Streaming Media's Environmental Impact." *Media+Environment* 2, no. 1. https://doi.org/10.1525/001c.17242.

McCullough, Malcolm. 2013. *Ambient Commons: Attention in the Age of Embodied Information*. Cambridge, MA: MIT Press.

Milan, Stefania, and Emiliano Treré. 2019. "Big Data from the South(s): Beyond Data Universalism." *Television and New Media* 20, no. 4: 319–35. https://doi.org/10.1177/1527476419837739.

Milan, Stefania, and Lonneke van der Velden, eds. 2018. "Data Activism." *Krisis: Journal for Contemporary Philosophy* 1. https://archive.krisis.eu/issue-1-2018-data-activism.

Montfort, Nick, and Ian Bogost. 2009. *Racing the Beam: The Atari Video Computer System*. Cambridge, MA: MIT Press.

Natale, Simone, and Andrea L. Guzman. 2022. "Reclaiming the Human in Machine Cultures: Introduction." *Media, Culture & Society* 44, no. 4: 627–37. https://doi.org/10.1177/01634437221099614.

Nieborg, David B., and Thomas Poell. 2018. "The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity." *New Media & Society* 20, no. 11: 4275–92. https://doi.org/10.1177/1461444818769694.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Parks, Lisa, and Nicole Starosielski, eds. 2015. *Signal Traffic: Critical Studies of Media Infrastructures.* Urbana-Champaign, IL: University of Illinois Press.

Parks, Lisa. 2020. "Field Mapping: What Is the 'Media' of Media Studies?" *Television and New Media* 21, no. 6: 642–49. https://doi.org/10.1177/1527476420919701.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge, MA: Harvard University Press.

Pink, Sarah, Shanti Sumartojo, Deborah Lupton, and Christine Heyes La Bond. 2017. "Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living." *Big Data and Society* 4, no. 1: 1–12. https://doi.org/10.1177/2053951717700924.

Plantin, Jean-Christophe, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* 20, no. 1: 293–310. https://doi.org/10.1177/1461444816661553.

Plantin, Jean-Christophe, and Aswin Punathambekar. 2019. "Digital Media Infrastructures: Pipes, Platforms, and Politics." *Media, Culture & Society* 41, no. 2: 163–74. https://doi.org/10.1177/0163443718818376.

Poell, Thomas, David Nieborg, and Brooke Erin Duffy. 2021. *Platforms and Cultural Production*. Cambridge, UK: Polity Press.

Poirier, Lindsay. 2021. "Reading Datasets: Strategies for Interpreting the Politics of Data Signification." *Big Data and Society* 8, no. 2: 1–19. https://doi.org/10.1177/20539517211029322.

Posner, Miriam, and Lauren F. Klein. 2017. "Data as Media." *Feminist Media Histories* 3, no. 3: 1–8.

Powell, Alison. 2018. "The Data Walkshop and Radical Bottom-up Data Knowledge." In *Ethnography for a Data-Saturated World*, edited by Hannah Knox and Dawn Nagus, 212–32. Manchester: University of Manchester Press.

Puschmann, Cornelius, and Jean Burgess. 2014. "Metaphors of Big Data." *International Journal of Communication* 8: 1690–709.

Rieder, Bernhard. 2015. "What Is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index." *Computational Culture* 2. http://computationalculture.net/article/what_is_in_pagerank.

Rogers, Richard. 2013. *Digital Methods*. Cambridge, MA: MIT Press.

Schäfer, Mirko Tobias, and Karin van Es, eds. 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam: Amsterdam University Press.

Stanfill, Mel. 2015. "The Interface as Discourse: The Production of Norms Through Web Design." *New Media & Society* 17, no. 1: 1059–74. https://doi.org/10.1177/1461444814520873.

Taylor, Linnet. 2017. "What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally." *Big Data & Society* 4, no. 2: 1–14. https://doi.org/10.1177/2053951717736335.

Underwood, Ted. 2018. "Why an Age of Machine Learning Needs the Humanities." Public Books, December 5, 2018. https://www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities.

Uricchio, William. 2011. "The Algorithmic Turn: Photosynth, Augmented Reality and the Changing Implications of the Image." *Visual Studies* 26, no. 1: 25–35.

van Dijck, José. 2013. *Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.

Verhoeff, Nanna, and Iris van der Tuin. 2020. "Interfaces (for) Diffracting Technobodies: A Science-Humanities-Design Perspective for an Algorithmic Somatechnics." *Somatechnics: Journal of Bodies – Technologies – Power* 10, no. 3: 374–96.

Verhoeff, Nanna, Sigrid Merx, and Michiel de Lange, eds. 2019. "Urban Interfaces: Media, Art, and Performance in Public Spaces." Special issue, *Leonardo Electronic Almanac* 22, no. 4.

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30, no. 1: 75–89.

About the Authors

Nanna Verhoeff is Professor of Screen Cultures and Society in the Department of Media and Culture Studies at Utrecht University. She initiated the research group [urban interfaces] and is co-lead of the Open Cities platform at Utrecht University. Her research on urban media combining perspectives from (digital) media and performance studies contributes to the interdisciplinary methodological development of the creative humanities.
>n.verhoeff@uu.nl

Karin van Es is Associate Professor of Media and Culture Studies at Utrecht University and is project lead Humanities at Utrecht Data School. Her main research concerns the datafication and platformization of culture and society.
> k.f.vanes@uu.nl

Part 1

Practices

1. Coffee Roasters' Data Vernacular

On the Entanglement of Digital Data and Craft

Markus Stauff, Pauline van Romondt Vis, and Karin van Es

Abstract
The symbolic opposition between data/datafication and human perception or reasoning is a key feature of contemporary data discourse. This article suggests analyzing how such dominant ideas about data get articulated in specific contexts. We take the use of computerized data in small-scale coffee roasting as an example of a "data vernacular" that reproduces, uses, but also modifies elements of the dominant data discourse. While data's promise of efficiency and consistency is taken up in coffee roasting, the data are embedded in the context of a craft whose insistence on the superiority of human senses actively constrains the impact of data. This ultimately adds vernacular voice and variation to the human versus data semantic.

Keywords: discourse, vernacular, coffee roasting, digital data, human senses, craft

Contemporary debates about digital data are structured by a set of assumptions and key concepts— what we will call "data discourse"—that enable the voicing of different opinions within a shared framework. One conspicuous example for such widely shared assumptions is the symbolic opposition between data/datafication and human perception or reasoning. According to Thylstrup et al. (2020), for instance, the amount of data in society has become too numerous to be accessible to the human mind, and datafication's abstractions and correlations offer insights that are radically different from what is accessible to human senses. Interestingly, this "problematization" grounds opposing takes on data. On the one hand, there are those who praise big data as a chance to gain new insights avoiding

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH01

human biases, and on the other hand, there are others who are concerned with the amounts of behavioral data collected on digital platforms and harnessed by machine learning to nudge constant attention and with how that threatens authentic social life and human values. The former plead for more datafication to guarantee social progress; the latter call for strategies of "digital detox" or "disconnectivity" to protect what they consider social and human values (Syvertsen and Enli 2020; Hesselberth 2018). And thus, while the two positions disagree on the evaluation of data, they share the basic assumption that humans and ("big") data have opposing characteristics and epistemic potentials.

However, to get a fuller picture of contemporary data culture, it is worth looking at how such basic ideas about data get articulated in specific contexts that combine human practices and data-based procedures in different ways. In this chapter, we therefore zoom in on the specific context of specialty coffee roasting. While coffee roasting machines, with their steely looks and levers, feel like remnants from the early industrial age, the monitoring and manipulation of the roasting process is supported by digital technology. Sensors and computer screens allow for precise gauging of temperature, time, and color of the roast. Such datafication promises to facilitate consistency and the deliberate development of specific roasting profiles for different beans. However, its overall impact is kept in check by the persistent relevance of manual practices and sensual experiences—like smell, vision, and sound—that are central to coffee roasting's self-characterization as a craft. Crafts are thus an especially interesting example here, since they insist on the inalienability of human senses and manual labor while also increasingly using computerized data to fine-tune production processes. To organize and legitimize this combination, elements of the dominant data discourse, with its opposition of human versus data, are used, reproduced, but also modified. These specific practices thus add to contemporary culture's ideas about, and modes of, using data.

Coffee connoisseurship, as a particular subculture, creates its own "data vernacular" that is related to but also distinct from other small-scale use of operational data by individuals and organizations. Like "vernacular photographies" (Batchen 2000) or "vernacular creativity" (Burgess 2007), the more idiosyncratic and rather local manners of using data, at least in public perception, often get overshadowed by strongly formalized and institutionalized applications. A data vernacular builds on but also contributes to the wider data discourse. Instead of taking any characteristics of digital data for granted, one needs to situate their features and potentials in the changing relationship with other practices and technologies. After all, the

impression that data are universal and the enticing power of "big data" only result from the "cleaning" of data that eradicates their local embeddedness and heterogeneity. In contrast, "[l]earning to look for the local in data can help us see data infrastructures as composites" (Loukissas 2019, 90).

In this chapter, we analyze the datafication of coffee roasting and how it ambivalently entangles data with the display and cultivation of human skills. First, we unpack what is meant by data discourse. Subsequently, we explore coffee roasting as a data-driven craft. In the final three sections, we reflect on the data practices of different roasters based on our interviews and contextual corpus. Here, we seek to understand how the specific combination of machines, knowledge exchange, apprenticeship, the senses, and collectively shared ideas gives meaning and function to datafication. More specifically, we aim to analyze how these coffee roasters negotiate their own use of (and reflection on) data with the generally shared ideas about characteristics and (dis-)advantages of data. The vernacular data practices of these coffee roasters can be considered a contribution to the wider data discourse that shapes societies' engagement with new forms of knowledge.

Data Discourse

Many scholars have underlined that the impact of data on contemporary culture cannot be separated from their "mythologies" (boyd and Crawford 2012), "ideologies" (van Dijck 2014), or "imaginaries" (Beer 2016). Like all technologies, processes of datafication are embedded in and achieve cultural significance through representations and ideas that delineate their alleged potentials and connect them to wider maps of social meaning and "epistemological hierarchies" (Millington and Millington 2019). *Moneyball*, for example, a popular non-fiction book (Lewis 2008) that was made into a successful movie (Miller 2011), introduced a wider audience to the idea that fine-grained statistics can help a sports team to identify players whose qualities have been overlooked by traditional scouts who base their decisions on mere observation. This superiority of data is similarly articulated when tracking apps offer new insights into aspects of everyday life from sleep patterns to learning progress. This always implies a performance of data: they need to be displayed and staged to convince a specific audience not only of the validity of the facts but also of the superiority of datafied knowledge (Ruppert and Scheel 2019). A broad variety of data visualizations translates abstract statistics into immediately comprehensive and often affective forms. Curve charts in particular display the temporal developments of

everything from COVID-19 infections to stock prices to individual fitness practices in dramatic ups and downs that make progress and regression immediately evident (Link 2004).

More generally, the design of all technologies is guided by "collectively imagined forms of social life and social order" (Jasanoff and Kim 2009, 120). On the one hand, technologies react to and aim to temper concerns about inhuman rationalization; on the other hand, they are made to embody the promise of social well-being and progress. Often, such cultural embedding of technologies is part of their institutional, coordinated implementation. Yet, it can just as much arise from vernacular applications and popular narratives. Bruno Latour underlined that such symbolic or rhetorical layers that connect technologies with imagined modes of use and promises of progress are no less substantial than the "actual" technical features of some machinery; rather, it depends on the conjunctural circumstances if (and to what extent) a "technical" element or a "rhetorical" element becomes essential for the durability and impact of a technology (1991, 114–16).

The subtle but important analytical and theoretical differences of scholarly approaches to such rhetorical and symbolic layers of technologies are beyond the scope of this chapter. However, applying their main insights to data, we use the concept of "data discourse" to highlight two aspects. First, in line with discourse analysis more broadly, we understand individual "opinions" and "ideas" to be connected to a patterned and structured way of speaking about data. Instead of purely idiosyncratic "sense making," the vernacular uses and meanings of data in each individual instance are of interest for how they emerge from and position themselves within the broader "problematizations" (Foucault 1997) characterizing contemporary data culture. Second, we understand data (their technologies, uses, and meanings) to be linked to questions of power and knowledge. The capability of shaping things (or behaviors) is both a condition for and a consequence of the knowledge that is enabled (and promised) through "big" or "small" data.

To understand coffee roastings' specific contribution to, and inflection of a wider data discourse, we base our analysis on semi-structured in-depth interviews (Kvale and Brinkmann 2014) with nine coffee roasters at five different sites in Amsterdam and observations of their coffee roasting process. We also looked at the broader context within which their situated data practices took shape by examining homepages of roasters and the coffee roasting handbooks by Scott Rao (2014) and Rob Hoos (2015) that are recurring reference points in the field. The interviews took place before and during the roasting of coffee and were recorded and transcribed. They were structured around our topic list (Galletta 2013) but moved freely between

topics. We asked our interviewees how they had learned to roast coffee, to describe the roasting process, and about the role of software and the human senses in that process. We took notes, photographs, and short video clips which, together with the transcribed interviews and the homepages of the different roasters, formed our corpus.

Coffee Roasting as Data-Driven Craft

Coffee first arrived in Europe during the seventeenth century when the colonial trade companies of Great Britain and the Netherlands compensated for the decreasing profit from spices and cotton with the import of coffee and tea (Reinhard 2016). Fueled by the advent of the more resilient Robusta coffee plant, different national coffee cultures emerged during industrialization (Morris 2017). Jonathan Morris (2017) links present-day roast type preferences to historical developments. The post–World War II era popularized the "cup of Joe" in the USA, which was served in large volume to complement food consumption. The light roasts predominant in Scandinavia, he speculates, are linked to the persistence of home roasting, which created a desire to reduce wastage. The French roast involved a dark roast to counteract the bitterness of the Robusta bean imported from their colonies. This is similarly true for countries such as the Netherlands and Belgium, which roasted medium-dark. The Italian coffee culture was driven by technological innovations and is distinct through the emergence of espresso, an elite beverage at the time, and the crema layer.

Rather than a collective national coffee culture, postmodern consumers "use coffee as a prop for the expression of individual personality" (Morris 2017, 487). Herein, however, Morris identifies a quest for authenticity which spurs a narrative that "present[s] a shift away from drinking commodity coffee as not so much a break with, as a reconnection to, the routines, rituals, and meanings that were manufactured around coffee consumption in the past" (2017, 488). Local coffee roasting is now part of a wider set of practices that harness a certain discontent with standardized mass consumption and transform it into both a subculture and an element of the creative industries. Coffee culture was part of a broader trend in consumer products that demanded more transparency regarding the resources, production lines, and sometimes the work conditions used. Like microbreweries, artisanal markets, and urban gardening, it injects local flavor, authenticity, and individualizing taste differences into consumer culture (Reckwitz 2010). Instead of merely returning to a pre-industrial form of coffee roasting,

this process opened many previously inaccessible or black-boxed aspects of coffee making to create craft-based approaches: the roasting, preparing, and even the drinking of coffee became connected to skill, knowledgeability, and distinction.

Until the end of the eighteenth century, craft was largely understood in terms of replication and "variation from norm was seen as a mark of poor quality" (Adamson 2014, 144). But today, craft is associated with the creative application of skill, experience, and attention to detail. This includes the public performance of "quality." In many branches, the use of (and knowledge about) either old-fashioned, "original" techniques or of state-of-the-art tools and technology highlights the artisanal character of the work. After all, there is nothing that can display the quality and originality of work like a set of exclusive tools. Manual skills and tacit knowledge become visible when they are organized around a set of specialized devices whose selection and pertinent application demand and thus embody depth and breadth of knowledge.

The use of data as a key tool in the craft of coffee roasting—and thus the contribution of coffee roasting to the wider data discourse—gained relevance through the emergence of "specialty coffee," a term used to signal and to technically standardize high quality coffee. Deviating from coffee sold in supermarkets and traditional Italian coffee culture, specialty coffee tends towards a "light" roasting of the beans to create a more complex flavor. Internationally well-known coffee expert and book author Scott Rao states: "The lighter one roasts, the more challenging it is to fully develop the bean centers" (2014, 178–79). Specialty coffee roasting presents itself as a craft combining manual skill, sensual awareness, and the appropriate application of tools and knowledge. As such, it highlights so-called "human" characteristics which are in rhetorical opposition to the "cold rationality" of data: datafication aims at abstraction and its power results from its distant and aggregating objectivity that is markedly different from local, embodied, and qualitatively rich human subjectivity (Peters 2001).

While technical knowledge—following a distinction by Oakeshott—"can be learned from a book," the practical knowledge characterizing craft and skill is "only imparted and acquired" (Oakeshott in Adamson 2014, 63) through repetitive practice. This allows (and requires) one to pay attention to the differences in material qualities and utilize them as a starting point to learn not only how to do things, but also to develop a sense of self and reflect on what we consider to be "good" (Sennett 2008, 8). But as we will see, the skills are practiced, performed, and disciplined through the use of data.

The craft of roasting is impossible without a diligent training of the senses. Controlling the roasting process involves smell (during roasting, the smell of the beans transitions from grass to hay to bread), sight (the color of the beans goes from green to yellow and then cinnamon—but preferably not to the dark brown or even black like industrial roasts), taste (identifying the degree of sweetness and acidity in the roasted beans), and even hearing (the beans make a cracking noise twice during the roasting process). The training of the senses is supported by note-taking (on paper or in spreadsheet software) that connects them with basic measurements: time and temperature of the roasting process. This enforces the quantification and thus commensuration of endless varieties of tastes and smells.

Far from being an individual process, this goes hand in hand with certain standardization of observations and evaluations. Not unlike the scientific communities analyzed by Lorraine Daston (2008), the coffee roaster community, as divulged by our informants, is a "well-trained collective" with shared manners of creating distinctions and identifying entities where the novice would only experience sensual chaos. During coffee tasting ("cupping"), the roasters all use a score sheet from the Specialty Coffee Association to determine the quality by quantifying fourteen different dimensions of their coffee (e.g., fragrance, aftertaste, acidity); international workshops and competitions contribute to the adjustment of individual observations and the acquisition of a collectively shared vocabulary.

Next to developing and partly standardizing the individual senses, a craft also implies careful attention to and knowledge about the raw material one works with—first, the beans, but eventually this extends to the beans' environment and the devices used. For specialty coffee roasting, the aim is to extract the best possible taste out of the particular bean. This triggers interest in using technology that allows roasters to perform their craft with insights into and the ability to manipulate ever more minute details of the process. Such focus on quality, taste differentiation, and technological knowledgeability creates a fertile ground for the application of data that safeguards and communicates high standards but also offers additional and fine-tuned ways of manipulating the object, fostering innovation and creativity.

Expensive devices allow the roasters to measure the moisture and density of the green beans before roasting or to determine the color and thus the "roast degree" of the roasted coffee. Most conspicuously, the entire roasting process is supported by computer software that logs and potentially steers the temperature and duration of coffee roasting. On the one hand, this is just a more efficient and precise way to handle data that before were collected

and noted manually. On the other hand, though, the use of computerized data has a different status symbolically and practically, and it changes the balance between data and sensorial observations. Both the pre-established analogue data practices and the relevance of sensual expertise characterize the "data vernacular" of coffee roasting and trigger an explicit reflection on data and its relation to craft. The next sections discuss 1) how data's promise of efficiency and consistency is taken up in coffee roasting, 2) how the data are embedded in the context of a craft, and 3) how the craft's insistence on the superiority of human senses actively constraints the impact of data.

The Promise of Efficiency and Consistency

Coffee roasting software helps to collect and log data while roasting coffee. What are the specific promises of (digital) data's application in coffee roasting and how does datafication increase the self-identification of roasting as a craft? Scott Rao (2014, 7–8) praises data-logging software as key to "a systematic, objective, evidence based" approach to roasting coffee that sits uneasily with intuitive roasters who celebrate the "feel" for roasting. This fits within larger debates about the relation between craft and technology, between intuition and data-based insights that characterize data discourse far beyond coffee roasting: In many fields, ranging from teaching and journalism to sports or medicine, the tension between (or ideal combination of) holistic human judgment and dissecting quantified analysis remains an ongoing issue. In 1996, Malcom McCullough foresaw a future in which digital craft was possible. Back then, the computer was regarded as a "tool for the mind not the hands" and technology as "order imposed on skill" (1996, 17–21). Adamson (2014, 166) remarks that the promise of completely computer-based craftsmanship foreseen by McCullough has not been realized. While he points to the public perception of craft as an intuitive practice centered on *making* rather than *thinking,* our example of coffee roasting might help to see how digital data get integrated into (partly) analogue craft.

Roasting software such as *Cropster* (market leader) and *Artisan* (open source) offer roasters a curve displaying the temporal development of the temperature. Of relevance for the taste of the coffee are not only the total duration and temperature of the roasting process, but also the time between the different phases of the roasting process. As outlined above, this can be classified through smell, color, or a cracking noise, which all are related to chemical processes like the Maillard reaction or "caramelization." On their

computers, roasters annotate the curve to record the moments when they increase the temperature or airflow. This creates a roasting *profile* that can easily be reproduced.

Additionally, the curve supports the development of pertinent roasting profiles for different beans. When roasters get a new bean, they might first apply the profile of a similar bean. Often, they roast smaller batches with three slightly different profiles to determine which one delivers the best taste and then fine-tune this profile. Thus, it combines the craft-typical attention to each bean (different from industrial coffee roasting) with a data-based form of reproducibility. In line with the interest to increase the aspects one can "craft," the curve enables roasters, as one of our informants put it, to determine not only the destination of your journey (e.g., the darkness of the roasted bean) but to understand and fine-tune the path there.

Peter tells us how the software increases control over the process.[1] Roasting coffee sometimes felt like "steering a large boat": the effect of changing direction is only noticeable later. Coffee roasting is similar in that the effect of adding heat or airflow only becomes apparent with a delay. The Cropster software offers a metric (the "rate of rise") that continuously indicates how quickly the temperature is rising and thus allows for much more subtle and rapid intervention. Mark confirms this; he now mostly looks at the curve rather than the temperature gauge of the roasting machine. Consequently, roasters mostly abstain from taking out sample beans during the process to check color and taste—a process that has the disadvantage of impacting the temperature in the roasting drum.

The rationale for using data in coffee roasting is largely in line with the contemporary data discourse: Digital data promises a certain non-invasive efficiency and consistency compared to the analogue notation of data. At the same time, and of special importance to a self-understanding as craft, the data open new, refined incentives and possibilities for constant experiments with more details becoming accessible through the digital data. It is mostly economic reasons that deter experimentation. The people at *Kaffee* mentioned that it becomes infeasible to find the optimal roasting profile if they get a small batch of very expensive coffee. For both objectives, consistency or experimentation, the digital data are firmly embedded in analogue protocols and human taste and decision making.

Ultimately, coffee roasters are in search of consistency to satisfy customers with good coffee (Schenker and Rothgeb 2017, 265). Consistency is considered the main advantage of the roast curve. It ensures that batches

1 All names of interviewed roasters and their roasteries are pseudonyms.

of the same bean are roasted the same way. As Mark explained, this is also an economic necessity because it guarantees reliable quality without constant experimentation and a lot of waste. It furthermore allows him to monitor his staff, who need to deliver a roast within a certain margin of the benchmark profile.

Digital Data are Embedded in Specific Contexts

The digital data and their visualization in roasting curves change what roasters can manipulate, what they pay attention to, and even how they define the quality of coffee. As long as this figures as a thoughtful, intentional use of a pertinent tool, it is in accordance with the notion of craft. As is well known from other fields, however, measurements with their seemingly objective authority tend to replace other (more sensual) ways of decision-making and shape the values that can be imagined (Beer 2016, 9). Our roasters restrain the looming authority of data through a self-understanding of roasting as a craft. The use of (digital) data is carefully embedded in a context that highlights the persistent relevance of ("analogue") skills for the achievement of quality.

The roasters underlined that, to guarantee consistency, the digital curve needs to be combined with analogue practices before, during, and after roasting. Paul called it "his protocol" several times, which referred to following a systematic and strict procedure. This protocol included how long to heat up the roaster, what he did in between batches, and when to mark the first crack with the software (e.g., when you hear some beans crack or wait for full-fledged cracking). Barry provided a similar reflection, discussing coffee roasting as choreography.

Additionally, the coffee roasters all highlighted the necessity to interpret and constantly adapt the data depending on the circumstances. They discussed the limitations of their tools and especially the sensors. Peter explained how the data-logging software only reads "relative measuring points," whereas the color of the beans is an actual result of energy transmission, an actual value. Two other roasters stated the data are mere representations of "something" happening in the drum. That "something" is specific for each machine, because the data are output from specific thermocouples. The type of sensor, the sensitivity, the positioning, and the number of sensors are different for *every* machine, even of the same type. This means that the data these sensors produce are *entangled* with a specific machine and the particular beans being roasted. Roasters highlight the role of data as a tool

that asks for careful, well-trained application, undermining the equation of data with automation and the objectivity that is dominant in the wider data discourse.

Moreover, the coffee roasters need to get a feel for the machine and how these data are specific to it. They know, for instance, that the drum is colder for the first batch than for consecutive batches. Exemplifying the craft approach to data, Paul explains how he has created different profiles for the first, second, and third roast. Similarly, Barry says that the first batch is always the worst. He compares it to the first pancake. Knowing it is the first batch changes their interpretation of the data output. We witnessed how this interpretation also takes place during roasting, as Paul remarked that he is a bit under the curve line: "Last week it was really a lot colder, which can make a difference in what you see by a degree or two." The readings are thus also influenced by the environment in which they operate (e.g., a hot or cold day). Data can help to cope with such volatility. Measuring the density and moisture of the bean before roasting can, for example, inform the choice of a roasting profile. However, a recurring motif in our interviews was context. This limits the explanatory power and the transferability of metrics. In accordance with the common characteristics of a craft, the insistence on the organic complexity of coffee beans—whose quality changes during the seasons because of temperature and humidity—undermines the authority of data and precludes data from replacing skill.

In the end, the curves and values are not seen as the secret sauce of coffee roasting, because they can be interpreted differently by someone in a different context. Additionally, the specifics of each machine foreclose a "blind" transfer of data from one machine or from one roaster to another. The roasters use both digital and sensory data. The interplay between these showed that digital data are not the holy grail, but that they are firmly embedded in specific contexts and personal tastes and are used in relation to the knowledge and experience roasters accumulate over the years.

Constraining the Role of Digital Data

An emphasis on the inconclusiveness of data was a shared sentiment among all roasters, creating symbolic space within the data discourse that can and needs to be filled with human skills and senses. This can be compared to streaming services like Netflix, where data analytics plays an increasing role in decision making. Here, despite the surrounding rhetoric and hype, the acquisition of a series like *House of Cards* was "a very human decision" (Frey

2021, 108). In their daily practices, roasters develop different strategies to combine the promises of datafication (consistency, multiplication of access points) with the articulation of specific human qualities, thus offering subtle variations on the human versus data semantic that structures contemporary data discourse. Underlining the status of the human–data relationship as a "problematization" rather than a consistent and unanimously shared ideology, roasters position themselves differently toward the relationship between humans and data. Most radically, Amsterdam coffee roaster *Rovers* (whom we did not interview) state on their homepage: "We are craft coffee roasters, there is [*sic*] no computers involved in our coffee roasting profiles, we see, smell and taste our coffees like no other." *Brandmeesters* similarly emphasizes the senses: "Real coffee roasting is not guided by a computer, but by the senses; our eyes, our ears and our nose." Both clearly put craft and computerization in opposition, whereas *Stooker* expresses their entanglement: "But in addition to craftsmen we also see ourselves as researchers. Using new innovations in the coffee world we keep on testing our knowledge and scrutinize our product. So, craftsmanship with a healthy dose of high-tech innovation." *Rum Baba* does the same: "Roasting coffee at *Rum Baba* is a precise and man-operated job, manual work with [*sic*] help of digital technology." Notably, we have not interviewed roasters who promote full automation, but we know of roasters who mostly base the roasting process on the execution of a profile they get with each batch of beans.

Returning to our interlocutors, Peter for example finds there is a danger in working too much with the screen, because it creates distance from the *actual process* of roasting coffee. Discussing fully automatic roasting, Paul states that he does not want the simplified reality captured by the computer. He explains how the senses are important because certain events in the process cannot be recorded automatically. Even though the software can predict the first crack, it is still registered based on a host of indicators: sound, smell, and looking at smoke development. Moreover, the senses are more accurate: "If it doesn't crack, it doesn't crack even if it says 180 degrees." We witnessed how the roasters continuously oscillated between their computer screen and looking at, listening to, and smelling the beans.

At *Roast*, they struggled to explain what they use the data for. Feeling and intuition clearly dominated their work. Repeatedly they answered "intuition" when we asked how they decided to adjust their roasting. They use the curve to get an idea of what aspects they could tweak and less to identify a particular problem.

At *Kaffee*, the unavoidable subjective dimension of the data logging was highlighted, as it is the human who makes the notation of the transition

from grass to hay. An aversion toward too much quantitative standardization was explicitly connected to the social and human characteristics of the specialty coffee circle. Mark wants to keep it "more poetic and less scientific." Significantly, several of the people we spoke to told us that they learned the craft by traveling and meeting people. If data discourse more generally tends to place human and (digital) data in symbolic opposition, the craft of coffee roasting describes itself as so based on social, personal, and sensual human qualities that digital data have a limited role to play.

The complex entanglement of human senses and data in the coffee roasting craft is especially visible in the relation between the taste and the adaptations to the roasting process. Peter remarks, "There is no direct relation between a nice curve and a nice coffee, even though on average the ones with the nice curve are better coffees." Barry, for instance, had an underdeveloped roast with a seemingly perfect curve. A desired change in taste cannot be achieved by a specific intervention in the roasting process.

The personal and sometimes volatile nature of human taste is another reason why some of the roasters doubt the increasing "scientificity" of coffee roasting through datafication. Thomas highlighted that the first cup of coffee of the day always tastes better than the second cup. At *Kaffee*, we heard that a different roast might be more appropriate for coffee with milk. This focus on taste also questions the quantified standardization and the respective protocols. At *Roast*, they highlighted their own taste as the definite guiding principle. They worry that collective taste will develop along predefined criteria through the scoring sheets (which they use as default) and too many rules. The exchange of profiles for them is mostly to get inspiration about radical or surprising alternative approaches or perspectives.

Although one of the roasters mentioned that the curve makes it easier to train new roasters, everyone with whom we spoke emphasized that learning how to roast coffee is best done through an apprenticeship. There are some basic principles, but most skills are learned by doing and experimenting. Barry went through a great deal of trial and error, working his way through hundreds of roasts, reading about roasting, embodying the knowledge. Kevin states that it is about "Doing time on the machine." Paul compared this process to building a reference library through experience: "You cannot write it down, read and then have that. It's different from doing it." Mark similarly spoke of creating a library of tastes.

Roasters use their experience when roasting new coffees. However, as discussed, it is possible to use existing roasting profiles from similar beans, and some of our interviewees mentioned liking roasting new beans by staying within the parameters of their practical knowledge. Relying

on experience alone when roasting coffee could be too expensive, as it is more difficult to reproduce the same taste in all batches. This is where the curve comes in handy. So, while data's role in coffee roasting is constrained through the focus on craft and personal taste, roasters still enact the promise of reproducibility. We see again how data is negotiated: earlier, it was by embedding data in the specific context of the craft, and now through the senses and the experience of the roasters.

Conclusion

In contrast to scientific disciplines and industrial production, specialty coffee roasting is a proudly and publicly displayed craft; the informal conversations at the counter and the homepages of coffeehouses and roasteries all contribute to broader concerns with achieving the appropriate balance of the authenticity of the human senses with the latest, advanced technology. At points, the use of advanced technology, including data, becomes a conspicuous element of craft: it extends the skill set for honing a product, thus performing individuality and creativity. Building on, and combined with, analogue forms of notation, standardization, and commensuration (e.g., taste vocabulary), datafication is here constrained as a tool for increasing human agency. In line with the symbolic opposition between data and humans in the wider data discourse, this impression of human agency, however, can only be sustained through an emphasis on the limitations of data. The organic unpredictability of the bean, the complexity of the roasting process (combining analogue and digital elements with natural ones like the weather), and, most importantly, the individuality of human senses and taste support a radical questioning of data's authority. This leaves ample room for experience, intuition, and other allegedly "human" qualities.

The specific requirements and traditions of embedded practices, like coffee roasting, add a vernacular voice and variation to the seemingly binary data discourse structured by a symbolic opposition between data/datafication and human ways of reasoning and thinking. Contemporary data discourse is far from homogeneous with respect to its conceptualization and evaluation of data. Most characteristically, this data vernacular, the use (and reflection) of data in specific contexts and practices, navigates a certain, intentional use of computerized data with a clear skepticism towards the overreach of data. Between the utopian and dystopian voices of the current data discourse, this might sound like a more "realist" middle ground resulting from vernacular data practices. Instead, however, we suggest understanding

coffee roasting as one of many actualizations of a wider data discourse that comprises different ways of balancing the human versus data distinction emerging from and made plausible by specific practices and contexts.

Our interviews and observations enabled us to consider how coffee roasting uses, enrolls, and appropriates digital tools. We are beyond an essential distinction between analogue and digital; the two are constantly embedded in hybrid situations. In this respect, coffee roasting is part of a broader emergence of "postdigital" practices. Especially in the arts, the label postdigital has been coined to describe aesthetic strategies that intentionally combine analogue and digital techniques to question the assumption of a "digital revolution" and the alleged singularity of all things digital (e.g., Berry and Dieter 2015). Without denying the epistemic and political challenges that result from large scale data traces and their algorithmic ordering, we would argue that the vernacular of coffee roasting—and quite probably other craft-based uses of digital data—offer important insights into the entanglement of human and data, too.

References

Adamson, Glenn. 2014. *The Invention of Craft*. London: Bloomsbury Academic.

Batchen, Geoffrey. 2000. "Vernacular Photographies." *History of Photography* 24, no. 3: 262–71. https://doi.org/10.1080/03087298.2000.10443418.

Beer, David. 2016. *Metric Power*. London: Palgrave Macmillan.

Berry, David M., and Michael Dieter, eds. 2015. *Postdigital Aesthetics*. London: Palgrave Macmillan.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5: 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Burgess, Jean. 2007. *Vernacular Creativity and New Media*. PhD diss., Queensland University of Technology. https://eprints.qut.edu.au/16378.

Daston, Lorraine. 2008. "On Scientific Observation." *Isis* 99: 97–110.

Foucault, Michel. 1997. "Polemics, Politics and Problematizations. An Interview with Michel Foucault by Paul Rabinow in May 1984." In *Ethics: Subjectivity and Truth. The Essential Works of Michel Foucault, 1954–1984*, edited by Paul Rabinow, 111–19. New York: The New Press.

Frey, Mattias. 2021. *Netflix Recommends: Algorithms, Film Choice, and the History of Taste.* Berkeley, CA: University of California Press.

Galletta, Anne. 2013. *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. New York: NYU Press.

Hesselberth, Pepita. 2018. "Discourses on Disconnectivity and the Right to Disconnect." *New Media & Society* 20, no. 5: 1994–2010. https://doi.org/10.1177/1461444817711449.

Hoos, Rob. 2015. *Modulating the Flavor Profile of Coffee*. Morrisville, NC: Lulu Press.

Jasanoff, Sheila, and Sang-Hyun Kim, eds. 2009. *Dreamscapes of Modernity*. Chicago: University of Chicago Press.

Kvale, Steinar, and Svend Brinkmann. 2009. *InterViews: Learning the Craft of Qualitative Interviewing*. London: SAGE.

Latour, Bruno. 1991. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.

Lewis, Michael. 2004. *Moneyball: The Art of Winning an Unfair Game*. New York: W. W. Norton and Company.

Link, Jürgen. 2004. "The Normalistic Subject and Its Curves: On the Symbolic Visualization of Orienteering Data." *Cultural Critique* 57: 47–67.

Loukissas, Yanni A. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

McCullough, Malcolm. 1996. *Abstract Craft: The Practiced Digital Hand*. Cambridge, MA: MIT Press.

Miller, Bennett, director. 2011. *Moneyball*. Columbia Pictures.

Millington, Brad, and Rob Millington. 2015. "'The Datafication of Everything': Toward a Sociology of Sport and Big Data." *Sociology of Sport Journal* 32, no. 2: 140–60.

Morris, Jonathan. 2017. "We Consumers – Tastes, Rituals and Waves." In *The Craft and Science of Coffee*, edited by Britta Folmer, 457–91. London: Academic Press/Elsevier.

Peters, John Durham. 2001. "'The Only Proper Scale of Representation': The Politics of Statistics and Stories." *Political Communication* 18, no. 4: 433–49.

Rao, Scott. 2014. *The Coffee Roaster's Companion*. Scott Rao.

Reckwitz, Andreas. 2010. "The Self-Culturalisation of the City: On the Transformation of Modern Urbanity in the 'Creative City.'" In *ESA Research Network Sociology of Culture Midterm Conference*, October 30, 2010. https://ssrn.com/abstract=1699808.

Reinhard, Wolfgang. 2016. *Die Unterwerfung der Welt: Globalgeschichte der europäischen Expansion 1415–2015*. Munich: C.H. Beck.

Ruppert, Evelyn, and Stephan Scheel. 2019. "The Politics of Method: Taming the New, Making Data Official." *International Political Sociology* 13, no. 3: 233–52.

Schenker, Stefan, and Trish Rothgeb. 2017. "The Roast: Creating the Beans' Signature." In *The Craft and Science of Coffee*, edited by Britta Folmer, 245–72. London: Academic Press/Elsevier.

Sennett, Richard. 2008. *The Craftsman*. New Haven, CT: Yale University Press.

Syvertsen, Trine, and Gunn Enli. 2020. "Digital Detox: Media Resistance and the Promise of Authenticity." *Convergence* 26, no. 5–6: 1269–83. https://doi.org/10.1177/1354856519847325.

Thylstrup, Nanna Bonde, Daniela Agostinho, Annie Ring, Catherine D'Ignazio and Kristin Veel. 2020. "Big Data as Uncertain Archives." In *Uncertain Archives: Critical Keywords for Big Data*, edited by Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel, 1–7. Cambridge, MA: MIT Press.

van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2. https://doi.org/10.24908/ss.v12i2.4776.

About the Authors

Markus Stauff works at the University of Amsterdam's Media Studies Department. His main research interests are television and digital media, governmentality, and visual culture of media sports.
> m.stauff@uva.nl

Pauline van Romondt Vis works at the University of Amsterdam's Media Studies Department. Her (research) interests include media representation and reception, construction of knowledge, power relations, and coffee.
> p.j.c.romondtvis@uva.nl

Karin van Es is Associate Professor of Media and Culture Studies at Utrecht University and is project lead Humanities at Utrecht Data School. Her main research concerns the datafication and platformization of culture and society.
> k.f.vanes@uu.nl

2. The Agricultural Data Imaginary

Precision Farming's Reinforcement of the Productivist Approach to Agriculture

Eggo Müller

Abstract

Big Data come with the promise of a better future. In the agricultural discourse on smart technologies and data-based applications in farming, so-called "precision farming" is envisioned as a "revolution" of traditional agricultural mass production of crops and livestock. Big Data are imagined as making the agrifood industry more efficient, more profitable, and more sustainable. Drawing on David Beer's concept of the "data imaginary" (2019), this chapter examines discourses on precision farming in corporate advertisements, lobbyist agricultural journals, and review articles in academic journals in the field of agriculture and computing. It argues that data-based agrifood production is seen as the next technological fix of the broken system of traditional industrial farming, while it in fact reinforces the devastating environmental and social damages that traditional industrial farming has caused.

Keywords: big data, smart farming, data imaginary, productivist agriculture, technological solutionism

While the famous metaphor of Big Data as "the new oil" of digital economies (e.g., van't Spijker 2014) has conclusively been criticized by a variety of academic commentators (c.f. Bucher 2018, 88), it is far from losing its social and discursive power in the business world and related sciences. On the contrary, the more data available and the more that data collecting technologies and practices proliferate, the more players in the looming data business invest in the development of new business opportunities based on the power of massive data-driven and algorithmically processed solutions (Beer 2019).

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH02

As Stefania Milan and Lonneke van der Velden note, "Big data evokes a broad set of socio-technical phenomena enveloped in quasi-mythological narratives that univocally emphasize possibility and magnitude" (2016, 60). In the realm of Big Data, such quasi-mythological narratives create what David Beer (2018, 2019) has called a "data imaginary"—a presentation of "a series of problems and inadequacies to which data analytics are offered as the solution" (Beer 2018).

One of these sectors in which data-based technologies are presented as the solution is the agrifood industry. Agriculture accounts for 4% of the global domestic product (GDP) worldwide (World Bank 2020) and is responsible for more than 25% of the greenhouse gas emissions, mainly caused by livestock production (Willet et al. 2019). However, as the driving companies of data-based technologies in this sector claim, data-based solutions will help fix the most urgent food- and sustainability-related issues of our planet—once the sector embraces its data-driven future. An article published in *Forbes* in 2019 summarizes this imagination of a future datafied agrifood sector as follows:

> In just 30 years' time, it is forecasted that the human population of our planet will be close to 10 billion. Producing enough food to feed these hungry mouths will be a challenge, and demographic trends such as urbanization, particularly in developing countries, will only add to that. To meet that challenge, agricultural businesses are pinning their hopes on technology, and that idea that increasingly sophisticated data and analytics tools will help to drive efficiencies and cut waste in agriculture and food production. (Marr 2019)

Indeed, the "datafication" (Mayer-Schönberger and Cukier 2013, 78) of agriculture is imagined as making food production and distribution more effective and, consequently, more sustainable. This so-called "precision agriculture" (Carolan 2016, 138) is the "revolution" for which the agrifood industry is striving, supported by international governmental institutions (Zaruo-Tejada et al. 2014) and the applied sciences in this sector (c.f. Himesh et al. 2018; Sponchioni et al. 2014). In the Netherlands, one of the world's leading countries in this sector, precision farming already covered about 65% of the arable farmland in 2015 (Michalopulos 2015).

Addressing the digital revolution in industrial farming, Kelly Bronson and Irena Knezevic (2016) have advocated for critical data scholarship in food and agriculture. This critical scholarship would include research into how "the images circulating in the promotion of Big Data tools normalize

hegemonic farming systems" (3), as they argue with respect to manufacturing company John Deere's visionary *Farm Forward* marketing video from 2012. In this chapter, I discuss such images as an agricultural data imaginary that started forming a dominant discourse in public relations, journalism, and science in the agrifood sector since the 2010s. First, I discuss the theoretical background of my approach, inspired by the work of David Beer (2016, 2019), in addressing how we should understand and analyze the work of affirmative discourses on the revolutionizing power of data and corresponding industrial practices and institutions. I revisit two marketing videos from John Deere's *Farm Forward* campaign that promote smart farming in an imaginative way and can be seen as the popularizing representation of the data imaginary of precision farming. After this, I review two other types of sources in more depth to reconstruct the agricultural data imaginary: articles on smart farming in the lobbyist online magazine *Future Farming* and scientific literature review articles on precision farming published in academic journals during the past ten years. Like the *Forbes* article quoted above, the diverse types of sources unanimously represent data-based precision farming as a profitable solution for major environmental problems. At the same time, they legitimize and reinforce what is known as the "productivist" approach to agriculture (Kneen 1995). This shows that the data imaginary in agriculture has formed a powerful discourse that infiltrates all three areas thoroughly: public relations, specialized journalism, and academic research. In the final section of this chapter, I then critically discuss the role of the agricultural data imaginary in reinforcing the disastrous productivist approach to food production.

The aim of this chapter is twofold: it presents an approach to the Big Data discourse in agriculture and analyzes the politics of the Big Data imaginary in that sector. In other words, the data themselves are not the object of my analysis but rather the discourse on the data-based agricultural technologies and applications. In doing so, this chapter develops a media studies perspective on Big Data in agriculture that critically discusses a blind spot in agricultural science that neglects the discursive work of Big Data. This focus on discourse and the data imaginary implies that I will not discuss the current developments of data-driven precision farming in depth (c.f. a short overview in Carolan 2015, 137ff. and in more detail García et al. 2020; Miles 2019; Sponchioni et al. 2014; Wolfert et al. 2017). Herein I follow David Beer's suggestion that it is the data imaginary that legitimizes and shapes data-led practices. However, as my findings show, the data imaginary as described by Beer is not universal but develops situated sets of ideologies legitimizing datafication in different sectors.

The Data Imaginary as Productive Discourse

The meaning of Big Data technologies is created in narratives and practices that situate these technologies in concrete everyday contexts. I draw on David Beer's (2016, 2018, 2019) work on discourses of Big Data, conceptualizing the "data imaginary" not just as the communicative "mirror" or "overflow" of actual Big Data practices in social, political, or economic reality, but as a productive power in shaping data-driven practices. Taking Foucault's *Birth of the Clinic* as a model, Beer explores in *The Data Gaze* "how data-led processes spread, how data-informed knowledge is legitimated and how this industry approaches and frames data" (2019, 1). Particularly the latter is consequential, since, as Beer emphasizes, mythological discourses on Big Data fed by the ideology of technological solutionism (Morozov 2013) are critical for the introduction and adaptation of data-driven technologies in the business world. Following Beer's discourse analytical approach, it is these mythological discourses and their rationales— defined as the "data imaginary"—that shape the realities and practices of Big Data:

> The data imaginary can be understood to be part of how people imagine data and its existence, as well as how it is imagined to fit within norms, expectations, social processes, transformations and ordering. (2019, 18)

With his concept of the "data imaginary," Beer draws on Charles Taylor's elaboration of "social imaginaries" as discussed in the book *Modern Social Imaginaries* (2004). Taylor defines "social imaginaries" as the ways people imagine their social world, including how they interact, communicate, and expect their environment to act based on shared norms and values. Social imaginaries thus have the power of ordering the social world and people's interactions, and they lend legitimacy to shared social practices. Beer's conceptualization of the data imaginary is designed to:

> reveal the embedded rationalizing discourses that are deeply woven into data analytics. This rationalizing discourse—which reflects wider norms, modes of calculative thinking, forms of governance and political ideas—is doing a significant amount of work to shape the integration and realization of data analytics in different settings. (2019, 7)

As I show in the following sections, Beer's concept of the data imaginary is instructive in identifying the features of the agricultural data imaginary as they are promoted in the productivist discourse on smart farming. However,

instead of departing from the six characteristics that Beer (2019) distilled from Big Data industries' self-promotion—namely "speedy," "accessible," "revealing," "panoramic," "prophetic," and "smart"—my analysis follows a bottom-up approach to unravel the specificity of the data imaginary in the agrifood sector. In the following section, I take a closer look at how two promotional videos by the machine and smart technologies manufacturer John Deere envision future farming.

Imagining Future Farming with John Deere

In agricultural production, data do not occur as side effects of everyday activities and interactions, unlike in computing, internet, and social media, domains that are the focus of most scholarship on Big Data within Media, Data, and Communication Studies (c.f. van Dijck 2014). One of the major developers and advocates of data-based precision farming is the John Deere company, the largest agriculture machinery producer worldwide founded in 1837 in Grand Detour, Illinois. In 2012, the company started equipping their agricultural machines with sensors to collect data about soil quality and crop condition and connect these data with other sets of information about weather, agricultural markets, and price developments (van Rijmenam 2013; Carolan 2017). To promote their data-driven systems of precision farming, John Deere launched a marketing campaign under the slogan "farm forward" that same year (Bronson and Knezevic 2016).

Central to this campaign was a video entitled *Farm Forward* illustrating the company's vision of data-based precision farming of the future. This six-minute video (John Deere 2012) establishes what the end title articulates below the company's yellow-green brand logo: "The future of farming is in sight." It describes the start of a day on a future farm. In this futuristic vision of farming, smart technologies and linked data processed by John Deere's platform have completely replaced heavy physical labor on the farm. A farmer's job is to make decisions based on the suggestions from the proprietary system that processes huge amounts of diverse data. With this algorithmically generated information, farming, it is suggested, becomes more effective, productive, and secure. Farmers can adapt to local circumstances such as weather, soil quality, and the growth of crops in real time, but the system also calculates external information about developments of markets and prices. In this vision of the future, the labor of farming is depicted first and foremost as managing information in a somewhat sterile environment. Except for a short virtual exchange with

his son out on the fields, the only human trace in this technology-loaded vision are the automatic female voices of John Deere's platform *Farm Site*.

While the first *Farm Forward* video from 2012 depicts a mix of already available and envisioned technologies linked to the virtual John Deere platform, the level of sci-fi in the 2019 video is far smaller. Most of the depicted technologies and services were up and running at this point and only some of them were still under development. What is called "The John Deere Farm Site" in the 2012 video was launched in that same year as the *MyJohnDeere.com* platform, designed to collect huge amounts of data from the buyers of John Deere's equipment and services. Combining these with weather and market data, the platform allows one to optimize production based on algorithmic calculations. In this respect, John Deere's strategy can be seen as a perfect example of what boyd and Crawford have identified as the "deep government and industrial drive toward gathering and extracting maximum value from data" (2012, 675).

The 2012 video was not well-received by its target group. Farmers felt that they were reduced to white collar workers that manage information instead of being in touch with nature, animals, and machines. As John Stone, SVP of John Deere's Intelligent Solutions Group (ISG), stated in an interview with Bernard Maar (2019), "the farmer has been the primary "sensor" on a farm for years – and so much of farming is visual." However, in John Deere's vision of the future, smart technologies take over and do a better job than any farmer before.

Not surprisingly, the 2019 revision of the video with the title *Farm Forward 2.0* (John Deere 2019) created a more lively and communicative representation of future farming that included women and family life on the farm. This time, the video starts with a scene where the farmer and his wife are out in the fresh air observing the rain falling on their fields and discussing how to approach the new day's tasks. The futuristic displays from the first video have shrunk to a real-size portable tablet that now provides the necessary data-based and algorithmically processed information. Life on the farm is represented in a more traditional, pastoral way, while the technology and data-based innovations are implied in emphatic interpersonal communication between human actors. The farmer interacts with John Deere's smart farming platform on a virtual screen in his pickup truck, suggesting that he is still out in the fields and in contact with nature. The scenes now include automated processes such as tractors performing a software update during the night or smart self-riding "see and spray systems" that apply pesticides effectively at night while the farmer's family is enjoying rest.

With few newly developed smart technologies and machines added, the 2019 video articulates the same discourse that Bronson and Knezevic have identified in the first video of the campaign as a traditional "productivist" approach to agriculture. More precisely, John Deere's vision of future farming implies claims of enhanced efficiency, security, resilience, and—new and rather explicitly in the second video from 2019—sustainability, while at the same time advocating a traditional productivist approach. This traditional productivist approach has been criticized for creating a treadmill of production and profit maximization (Ward 1993) and is, as Geoff A. Wilson argues, "strongly rooted in memories of wartime hardships" (2001, 79). It has resulted in an unsustainable system of industrial overproduction of food in the global West, causing massive health issues and irreparable environmental damage (Willet et al. 2019) while leaving significant parts of the world population with draughts, malnutrition, and starvation (Bronson and Knezevic 2016, 3). However, as the *Farm Forward* campaign imagines, future data-based farming technologies will help to fix at least the environmental problems. In the next section, I discuss lobbyist discourses in the agricultural magazine *Future Farming* before I analyze exemplary academic review articles on precision farming.

Agricultural Data Imaginary in the Expert Magazine *Future Farming*

Future Framing is an online platform and magazine that, according to its own marketing, forms the "gateway to the world of smart farming" (www.futurefarming.com). Together with several "content partners" in the precision agrifood industry, it covers and promotes smart technological and data-based innovations in the agricultural production chain. Along with the website *Future Farming*, it runs four other websites with expert and industry information about innovations in diverse sectors of livestock production: *Pig Progress*, *Dairy Global*, *Poultry World*, and *All About Feed*. The platform's close connection to the industry is not seen as problematic but is instead featured as an asset: well informed experts from the smart agrifood industries regularly publish on these platforms, including *Future Farming*. For this chapter, I have reviewed articles published in the section "Smart farming" that address data-related innovations and Big Data.

Ofir Schlam's commentary on the "4 ways big data analytics are transforming agriculture" (2019) can be seen as exemplary of *Future Farming*'s discourse. It states from the start that "[d]ata-driven farming is on course to reshape the entire agricultural economy." The author, president, and co-founder of Taranis,

a company offering AI and machine learning systems for precision monitoring crop growth (including stand count, insect damage, weed detection, nutrient deficiencies disease pressure; https://taranis.ag), identifies the following four aspects of the data-driven transformation of the agricultural economy:

1. Boosting productivity and innovation
2. Managing environmental challenges
3. Cost savings and business opportunities
4. Better supply chain management (Schlam 2019)

These four aspects cover all dimensions of a productivist discourse: Big Data guarantee that production becomes more efficient, fertile, and sustainable and will be more profitable for those who embrace the new data-led technologies. In his explanation of the four aspects, Schlam reproduces typical tropes that regularly surface in *Future Farming*'s coverage of data-driven solutions:

– The growing world population and "global food demand"
– Better management of "key resources including seed, fertilizer, and pesticides" implying that fewer resources will be wasted during the production process
– The claim that data from soil and plant sensors "gaining unprecedented visibility" outperform the farmer's eyesight
– The possibility to adapt to "climate change and other environmental challenges"
– More income and thus the opportunity to save money and manage risks of volatile markets
– A supply chain that "will be better equipped to tailor their product offerings and services according to the needs of the agricultural market"

The final statement of the article summarizes all the central tropes of this discourse as follows:

> That's the benefit of precision agriculture and data-driven farming: It doesn't just make farmers smarter, more productive, and more efficient. It's on course to reshape the entire agricultural economy—and to help feed billions of people in the process. (Schlam 2019)

It is not surprising that Schlam, as a representative of a start-up in precision farming, reproduces a discourse that promotes data-driven technologies as the solution to challenges that extensive industrial mass production

of crops and livestock in the Western world have caused. These ideas are typical for the professional discourse in this field, as a study based on forty interviews with US farmers by Christopher Miles (2019) has demonstrated: "Big data, and automation will create more accurate, efficient, transparent and environmentally friendly food production," as Miles (2019, 1) summarizes the farmers' beliefs. However, this discourse implies that diets will not change, that populations and up-and-coming economies will follow Western patterns of food consumption, and that industrial agriculture will remain the standard form of the production of food. Before addressing the problematic dimensions of this productivist discourse, I will have a short look at the academic discourse reviewing studies on the development of data-driven precision agriculture.

Big Data Imaginary in Scientific Literature Reviews on Precision Farming

It is not surprising that lobbyist publications promote big farming companies' vision of and approach to data-based precision farming technologies and solutions. However, one might expect a different discourse in scientific publications on Big Data in agriculture. And indeed, papers and statements linked to alternative, sustainable agriculture and the *Right to Repair* movement (Bloomberg 2017; Carolan 2016; Wanstreet 2018) indicate that there is a critical scholarship regarding the social and economic consequences of data-based precision agriculture. However, my analysis of articles published in leading academic journals in agronomy reveals a dominantly affirmative discourse embracing and reproducing the industrial Big Data imaginary of precision farming. My sample is taken from the extensive bibliometric literature review of the "Digital Agricultural Revolution" by Bertoglio et al. (2021). I will examine one article that I find exemplary of this as the main source for my analysis.

In their review of academic literature on "the use of machine learning in precision livestock farming" of the past 10 years, Rodrigo García et al. (2020) introduce precision livestock farming as the "fourth industrial revolution, also known as Industry 4.0" (1) and summarize its main advantages as follows:

> (i) to identify the most appropriate livestock feeding, (ii) reduce environmental impact through efficient management, (iii) manage crop processes to make a perfect synergy with livestock feeding, (iv) ensure food safety through traceability [...] of products, and (v) improve animal health and crop efficiency. (García et al. 2020, 1)

It is immediately obvious that this condensed overview, based on publications in scientific journals over the past ten years, reproduces the typical tropes of the productivist discourse: enhanced efficiency including synergy between different sectors, enhanced food security, enhanced health of livestock and crops, and enhanced sustainability. Sustainability relates in this context specifically to "improved productivity," which is enabled by adequate data management. Obviously, the data imaginary also here does its discursive work:

> To improve efficiency, it is essential to, correctly, manage data generated every day in livestock farms [...]. A correct data management can result in improved productivity, in terms of grazing lot management, livestock nutrition, and animal health. (García et al. 2020, 1)

Again, the argument is that connected data sensors can deliver Big Data information in real-time that generates better insights than a farmer could access ever before, since in

> traditional livestock farming, decisions are often based—only—on the experience of the producer. In PLF [precision livestock farming; E.M.], such decisions are based on quantitative data, such as liters of milk per milking. In addition, quantitative data can be obtained in real-time. To obtain and study such data, PLF systems use data analysis, *machine learning* (*ML*), control systems, and ICT. (García et al. 2020, 1)

And the central legitimizing trope of data-based precision farming is not missing in the introduction to this literature review:

> At present, PLF seeks, through technological solutions in agricultural livestock production systems, to supply adequate food for the expected world population of more than nine billion inhabitants by 2050 [...]. (García et al. 2020, 2)

This claim then is supported by repeating the argument that precision livestock farming will also enhance sustainability by improving animal health, and it is then added that

> PLF allows producers to maintain an optimum number of animals per farm, find prompt solutions to animal diseases, and define a more efficient production model. (García et al. 2020, 2)

Again, the trope of sustainability is linked to the tropes of productivity and efficiency, which means optimizing the livestock per farm. Although, technically speaking, that could include reducing the number of animals per square meter, it seems that this sentence does not suggest this rhetorically. More radical steps towards enhanced sustainability, like the reduction or the abolition of livestock production, are certainly not what this productivist discourse proposes. On the contrary, the trope of the growing world population again functions as the rational and moral legitimation of an intensified productivist approach to farming.

My review of a broader sample of literature largely showed the same patterns: productivist discourses based on the Big Data imaginary prevail and exclude critical voices in the discussion about the sustainability of the industrial mass-production of food (e.g., Himesh et al. 2020; Wolfert 2017). This is the case despite mass production of food being responsible for major damage of the environment and of the health of human and nonhuman animals (c.f. Willet 2019). Such exclusion seems to be the most powerful discursive effect of the amalgamation of the agrifood industry's data imaginary with the discourse of productivist industrial production. There seems to be no alternative, since data-based precision farming is imagined as more effective and at the same time more sustainable than traditional, analogue farmer-based agriculture.

However, as I show in the final section of this chapter, there are also critical and alternative voices addressing problematic ownership-related and environmental implications of this productivist data imaginary. Theses voices are rooted in different scholarly traditions, such as the sociology of food and agriculture (c.f. Carolan 2022), political economy and environmentalism (Dauvergne 2020), and critical data studies (Bronson 2022; Bronson and Knezevic 2016; Wanstreet 2018).

Productivist Data Imaginary Reinforcing Unsustainable Food Systems

As discussed above, Ofir Schlam's commentary on the "4 ways big data analytics are transforming agriculture" (2019) describes John Deere as an exemplary company helping farmers with their innovative technologies and access to Big Data to increase their production by 30%. John Deere's services work not only with data generated by the individual farmer covering his own soil, seeds, and plants, but, as Schlam emphasizes, "the portal also includes data from outside sources, including other farmers, offering insight into

productivity under a wide range of conditions" (2019). John Deere owns these data, and though the company has signed the *Privacy and Security Principles for Farm Data* formulated by the American Farm Bureau, commentators expect that the services will in fact turn out to be a closed system on which farmers will be dependent once they have subscribed to the services. As Rian Wanstreet comments in an article discussing John Deere's policy:

> Equipment manufacturers know their customers will find it almost impossible to leave their precision agriculture data platforms once they've joined, and almost as hard to stay away. […] The general belief is that those who buy-in to a precision data platform will have no choice but to stay in, and as more come onboard, the more it will seem that everyone *has* to join. Think about it like Facebook, but for agricultural equipment. (2018)

This view is supported by a sociological study by Michael Carolan, who interviewed fourteen professionals involved in the large-scale precision agrifood industry and nineteen regional food entrepreneurs engaged in making precision farming accessible for sustainable small-scale farms. As Carolan shows, representatives of the large-scale agro industry believe in Big Data as "the next 'big thing'" (137), while those involved in regional or local initiatives trying to adapt smart technologies for sustainable farming and food products are critical of the proprietary systems that will reinforce and probably intensify farmers' existing dependencies on the dominant, globally operating companies. As one representative of a precision tech company stated in one of Carolan's interviews,

> Farmers needn't to worry about losing control of the data. […] What we provide, and what we want farmers coming back for year in and year out, are our tools, our platforms, algorithms, and our expertise. (2016, 147)

However, when Carolan then asked whether farmers would become dependent on the data-based services and thus be forced to come back, the industry representant's self-confident, if not threatening, answer was: "It's always their choice. If they want to remain profitable, they'll keep coming back" (2016, 148). Not surprisingly, entrepreneurs in regional or local food initiatives, aimed at challenging the dominant productivist food system using AI to support sustainable developments, expressed their concerns about farmers' intensified dependencies once they have subscribed to corporate systems of data-based precision technologies, even if this is accomplished in vague terms of community building.

Like Carolan, Peter Dauvergne (2020) emphasizes in his book *AI in the Wild* (2020) that farmers who subscribe to the technologies and services of John Deere or other leading precision farming companies would be locked into the company's system and would become totally dependent on that company's policy. This was already John Deere's business model—heavily criticized by the *Right to Repair* movement—in the analogue era with their "hardware," the agricultural machines. No external service, and not even farmers themselves, were allowed to fix a broken machine from the John Deere company. Not surprisingly, the *Right to Repair* movement also fights John Deere's data policy that copies the company's infamous "hardware" policy (Bloomberg 2017). The movement has since proliferated widely to other sectors including ICT and was recently picked up as a European Union directive (Hernandez, Miranda, and Goñi 2020).

However, buying a John Deere means to subscribe and pay for the company's services. As the *Farm Forward* videos by John Deere illustrate, too, farmers using the "John Deere Farm Site," or in the real world the MyJohnDeere.com platform, are attached to the company's services. These include not only selling agricultural equipment, seeds, fertilizers, pesticides, and fuel, but also providing loans for buying these resources or leasing new, expensive data-ready equipment (Wanstreet 2018). What the *Farm Forward* videos and Schlam's article imply are the new dependencies for farmers: dependencies on the "needs of the agricultural market"—as Schlam (2019) addresses these—are shaped by the mayor players on that market, with John Deere being one of these mayor players ready to exploit the new dependency of the individual farmer.

This, too, is an effect of the power of the pervasive Big Data discourse on productivist agriculture that Dauvergne discusses from a critical political economy perspective in *AI in the Wild* (2020). While acknowledging the potential role of AI for the future of global sustainability (2020, 112), Dauvergne also points to "a risk of smart city and farming technologies reinforcing global forces of unsustainable consumption and production." (102) With the enhanced dependency of farmers on technological innovations and on data-based governance and the big transnational conglomerates in the agrifood industry, Dauvergne sees precision farming as a powerful discourse that will foster the traditional productivist approach to agriculture. As he states,

> [...] more often than not, the environmental gains from the commercial applications of artificial intelligence are rebounding into greater extraction, production, and consumption, doing more to prop up failed models of technocratic management than truly advance global sustainability. (2020, 10)

And while smart agriculture cannot "fix the broken system" (2020, 113), i.e., the unsustainability of the global extractive agrifood industry, the data imaginary of precision farming has generated a powerful discourse that supports this very broken system by promising that AI and Big Data in the sector will solve one of mankind's most urgent problems: feeding a growing world population while respecting the limits of our planet. This powerful discourse is today evident in the institutionalization of systems that companies such as John Deere or Monsanto, now acquired by the German multinational pharmaceutical and life sciences company Bayer, have developed. It is evident in agricultural equipment such as drones that scan the soil, and in self-steering tractors that do the sowing and the harvesting on the fields. And it is manifest in governmental policies (e.g., Zarco-Tejada et al. 2014) and managerial and financial infrastructures that push data-based precision farming as the only futureproof approach to agriculture. Rian Wanstreet therefore speaks of the momentum of a "'treadmill-like' discourse that prevails in industry" (2018).

Conclusion

As I have shown, the particular data imaginary of precision farming comprises the tropes of enhanced effectivity, accuracy, safety, and thus profitability and links these to the topics of sustainability, climate change, and of the moral obligation to care about a vast, growing world population. These three topics, perhaps the most urgent global issues, are thus presented as the core problems for which data-based precision farming is presented as the solution. Or as David Beer has put it when referring to the promotional discourse of the data analytics industry:

> A life without data is left unimaginable, and a life with data is glossy, shiny, and full of hope. That is the image that is conjured. The result is that data analytics become much harder to turn away from. (Beer 2018)

The same principle applies for agriculture where production without data-based precision farming is left unimaginable, although the legitimizing ideologies are situated in the particular sector: not only will farming be more effective and a farmer's life "shinier" and without all the uncertainty of traditional farming, data-based precision farming will save mankind and our planet. This agricultural data imaginary is widely shared amongst industry representatives, lobbyists, policy makers, and academic experts

in agricultural science. It is an imaginary that must be deconstructed vis-à-vis the devastating effects of productivist agrifood industries in the Anthropocene: deforestation, overfertilization, pollution of water and soil, greenhouse gas emission, climate change. These problems result in ever more draughts, wildfires, floods, and famines particularly, but not only, in less developed regions of the world.

Bronson and Knezevic (2016) are right: we do need a critical data scholarship in food and agriculture that includes a media studies–inspired critical perspective on the discourses of Big Data in agrifood production (cf. Miles 2019). We need to ask critical questions about which players embrace and develop data-based technologies; how existing infrastructures transform by being datafied and what new infrastructures emerge; what discourses promote and structure these transformations; and what are the societal and, in the specific case of this chapter, environmental effects. Those are questions that a critical data scholarship poses when discussing the Big Data discourse. For a critical analysis of the politics of the Big Data imaginary in the agrifood sector, Beer's approach is instructive, as I have shown in this chapter with regard to the productivist discourse in agriculture. The concept can be applied as a critical analytical tool to other fields of society where data-led transformations, based on the ideology of technological solutionism, are embraced as the "next big thing." The danger is that such transformations do reinforce existing unsustainable, undemocratic, and discriminatory systems.

References

Beer, David. 2016. "How Should We Do the History of Big Data?" *Big Data & Society* 6, no. 1: 1–10.

Beer, David. 2018. "The Data Imaginary: Six Reasons Why Data Analytics Have Become So Powerful." *LSE British Politics and Policy at LSE* (blog), *London School of Economics and Political Science*. July 4, 2018. https://blogs.lse.ac.uk/politicsandpolicy/data-imaginary.

Beer, David. 2019. *The Data Gaze: Capitalism, Power and Perception*. London: SAGE.

Bertoglio, Riccardo, Chiara Corbo, Filippo M. Renga, and Matteo Matteucci. 2021. "The Digital Agricultural Revolution: A Bibliometric Analyses Literature Review." *IEEE Access* 9. https://doi.org/10.1109/ACCESS.2021.3115258.

Bloomberg, Jason. 2017. "John Deere's Digital Transformation Runs Afoul of Right-To-Repair Movement." *Forbes,* April 30, 2017. https://www.forbes.com/sites/jasonbloomberg/2017/04/30/john-deeres-digital-transformation-runs-afoul-of-right-to-repair-movement/#1a5256ab5ab9.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocation for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5: 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Bronson, Kelly. 2022. *The Immaculate Conception of Data: Agribusiness, Activists, and Their Shared Politics of the Future*. Montreal: McGill-Queen's University Press.

Bronson, Kelly and Irena Knezevic. 2016. "Big Data in Food and Agriculture." *Big Data & Society* 3, no. 1: 1–5.

Bucher, Tania. 2020. "(Big) Data and Algorithms: Looking for Meaningful Patterns." In *Handbook of Digital Media and Communication*, edited by Leah A. Lievrouw and Brian D. Loader, 87–98. Los Angeles, CA: SAGE.

Carolan, Michael. 2017. "Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition." *Sociologia Ruralis* 75, no. 2: 135–54.

Dauvergne, Peter. 2020. *AI in the Wild: Sustainability in the Age of Artificial Intelligence*. Cambridge, MA: MIT Press.

García, Rodrigo, Jose Aguilar, Mauricio Toro, Angel Pinto, and Paul Rodríguez. 2020. "A Systematic Literature Review on the Use of Machine Learning in Precision Livestock Farming." *Computers and Electronics in Agriculture* 179.

Hernandez, Ricardo J., Constanza Miranda, and Julian Goñi. 2020. "Empowering Sustainable Consumption by Giving Back to Consumers the 'Right to Repair.'" *Sustainability* 12, no. 3: 850–65. https://doi.org/10.3390/su12030850.

Himesh, Shivappa, E.V.S. Prakasa, K.C. Gouda, K.V. Ramesh, V. Rakesh, G.N. Mohapatra, B. Kantha et al. 2020. "Digital Revolution and Big Data: A New Revolution in Agriculture." *CAB Reviews: Perspectives in Agriculture, Veterinary Sciences, Nutrition and Natural Resources* 13, no. 21: 1–7.

John Deere. 2012. "Farm Forward." YouTube. Video, 2:44. August 23, 2012. https://www.youtube.com/watch?v=to8nOEkrX-I

John Deere. 2019. "Farm Forward 2.0." YouTube. Video, 3:54. John Deere. February 20, 2019. https://www.youtube.com/watch?v=nKAz-g7MAxs.

Kneen, Brewster. 1995. *From Land to Mouth: Understanding the Food System*, 2nd ed. Toronto, ON: NC Press Ltd.

Marr, Bernard. 2019. "The Amazing Ways John Deere Uses AI and Machine Vision to Help Feed 10 Billion People." *Forbes*, March 15, 2019.
https://www.forbes.com/sites/bernardmarr/2019/03/15/the-amazing-ways-john-deere-uses-ai-and-machine-vision-to-help-feed-10-billion-people/#6e20f0a52ae9.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Boston: Mariner Books.

Michalopoulos, Sarantis. 2015. "Europe Entering the Era of 'Precision Agriculture.'" *EurActiv.com*, October 23, 2015.
https://www.euractiv.com/section/science-policymaking/news/europe-entering-the-era-of-precision-agriculture.

Milan, Stefania, and Lonneke van der Velden. 2016. "The Alternative Epistemologies of Data Activism." *Digital Culture & Society* 2, no. 2: 57–74. https://doi.org/10.14361/dcs-2016-0205.

Miles, Christopher. 2019. "The Combine Will Tell the Truth: On Precision Agriculture and Algorithmic Rationality." *Big Data & Society* 1, no. 12. https://doi.org/10.1177/205395171989444.

Morozov, Evgeny. 2013. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: Public Affairs.

Schlam, Ofir. 2019. "4 Ways Big Data Analytics Are Transforming Agriculture." *Future Farming*, July 15, 2019. https://www.futurefarming.com/Tools-data/Articles/2019/7/4-ways-big-data-analytics-are-transforming-agriculture-450440E.

Sponchioni, Gianluca, Matteo Vezzoni, Andrea Bacchetti, Maria Pavesi, and Filippo M. Renga. 2014. "The 4.0 Revolution in Agriculture: A Multi-Perspective Definition." *Summer School F. Turco-Industrial Systems Engineering*: 143–49.

Taylor, Charles. 2004. *Modern Social Imaginaries*. Durham, NC: Duke University Press.

van Dijck, José. 2014. "Datafication, Dataism, and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2: 197–208.

van Rijmenam, Mark. 2013. "John Deere Is Revolutionizing Farming with Big Data." *Datafloq*. February 21, 2013. https://datafloq.com/read/john-deere-revolutionizing-farming-big-data.

van 't Spijker, Arent. 2014. *The New Oil: Using Innovative Business Models to Turn Data into Profit*. New York: Technics Publications.

Wanstreet, Rian. 2018. "America's Farmers Are Becoming Prisoners to Agriculture's Technological Revolution." *Vice*, March 8, 2018. https://www.vice.com/en_us/article/a34pp4/john-deere-tractor-hacking-big-data-surveillance.

Ward, Neil. 1993. "The Agricultural Treadmill and the Rural Environment in the Post-Productive Era." *Sociologica Ruralis* 33, no. 3–4: 348–64.

Willet, Walter et al. 2019. "Food in the Anthropocene: The EAT-Lancet Commission on Healthy Diets from Sustainable Food Systems." *The Lancet* 393, 10170. https://doi.org/10.1016/S0140-6736(18)31788-4

Wilson, Geoff A. 2001. "From Productivism to Post-Productivism… and Back Again? Exploring the (Un)changed Natural and Mental Landscapes of European Agriculture." *Transactions of the Institute of British Geographers* 26, no. 1: 77–102.

Wolfert, Sjaak, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. 2017. "Big Data in Smart Farming: A Review." *Agricultural Systems* 153: 69–80.

World Bank. 2020. "Agriculture and Food: Overview." https://www.worldbank.org/en/topic/agriculture/overview.

Zarco-Tejada, Pablo J., Neil Hubbard, and Philippe Loudjani. 2014. *Precision Agriculture: An Opportunity for EU Farmers.* Brussels: Joint Research Centre of the European Commission.

About the Author

Eggo Müller is Professor of Media and Communication at Utrecht University. His research covers television entertainment, participatory cultures, digital audio-visual archives, and food communication. He is member of Utrecht University's Centre for Environmental Humanities and of Future Food Utrecht.
> e.mueller@uu.nl

3. Controversing Datafication through Media Architectures

Corelia Baibarac Duignan, Julieta Matos Castaño, Anouk Geenen, and Michiel de Lange

Abstract

In this chapter, we discuss a speculative and participatory "media architecture" installation that engages people with the potential impacts of data through speculative future images of the datafied city. The installation was originally conceived as a physical combination of digital media technologies and architectural form—a "media architecture"—that was to be situated in a particular urban setting. Due to the COVID-19 pandemic, however, it was produced and tested for an online workshop. It is centered on "design frictions" (Forlano and Mathew, 2014) and processes of controversing (Baibarac-Duignan and de Lange, 2021). Instead of smoothing out tensions through "neutral" data visualizations, controversing centers on opening avenues for meaningful participation around frictions and controversies that arise from the datafication of urban life. The installation represents an instance of how processes of controversing may unfold through digital interfaces. Here, we explore its performative potential to "interface" abstract dimensions of datafication, "translate" them into collective issues of concern, and spark imagination around (un)desirable datafied urban futures.

Keywords: datafication, controversing, public engagement, urban futures, smart city, media architecture

Imagine yourself walking on the street in a city sometime in the near future. In fact, make it today. Surveillance cameras are likely installed on every corner. How would you feel about this? Do these cameras make you feel safe in a public space, or do they make you feel spied on? Maybe you are

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH03

wondering why no one seems to have asked you anything about installing surveillance cameras, what happens with the data that is captured, or who has access to the data and for how long. Or perhaps you simply try to ignore their presence altogether. The omnipresence of "smart technologies" in cities today, where media technologies are part and parcel of urban architecture, is controversial. The ongoing datafication of cities leads to a variety of contestations, for instance about how new forms of knowledge production coincide with new kinds of in- and exclusion, about the agency of citizens in such developments, and about societal friction regarding public values. We believe that one of the things hampering the discussion of these questions is that the datafication of cities happens largely under the radar. In other words, there are little to no opportunities for people to engage in issues and debates about datafied smart cities. Another hindrance is that the datafication of urban life is often presented as a neutral and efficient technological solution to complex urban problems. This effectively sweeps any potential normative discussions about what kind of urban futures we find acceptable or desirable under the rug, as well as the frictions and contestations that come along with it.

By contrast, we feel that it is imperative to develop ways for teasing out those discussions and engaging a multiplicity of voices in the debates about the futures of our datafied cities. As our cities today have become hybrids of architectural form and media interfaces, we must find ways to debate this through critical "media architectures." In this chapter, we discuss such a media architecture, a research-by-design installation called *Future Frictions* that is meant to do exactly that. The question raised is how a media architecture installation can contribute to fostering civic engagement in datafied smart city futures through a deliberate strategy of "controversing" (a strategy for making controversies publicly debatable). We analyze how this controversing strategy, explained in more detail below, can help to generate public discussions about datafied urban futures and public values.

Datafied Smart City Futures, Value Frictions, and Controversies

Smart technologies and big data have taken a central role in efforts to curtail the impacts of cities on wider contemporary societal challenges like climate change, resource depletion, and increased green-house-gas emissions. Datafication is presented as delivering much-needed seamless solutions by addressing alleged inefficiencies in the urban system in frictionless ways (Powell 2021). Aims of streamlining and optimizing urban infrastructures

and services underpin the "smart city" as a predominant urban imaginary, centered on homogenizing visions of quantified and techno-oriented urban futures (Sadowski and Bendor 2019). Moreover, most smart city visions portray media technologies in general—and data in particular—as means to solve the problems of the city *as is*, rather than creating opportunities for radically re-imagining and transforming our urban futures by the people who inhabit these cities (Miller 2020). After all, using data for optimization always departs from what is already available. Ironically, institutional approaches to engage citizens in smart city developments are often intended to smoothen out and overcome tensions that may arise from the implementation of technologies themselves.

The increased power of technology companies together with corporate and policy visions advocating for the use of "smart" technologies to address urban problems has led to significant concerns in the academic arena with how "big data" may affect public values and create social inequalities (Kitchin 2014). Such concerns have raised critical debates around "smart cities" (Townsend 2013), the "datafied society" (van Es and Schäfer 2017), "platform society" (van Dijck, Poell, and de Waal 2018), and "surveillance capitalism" (Zuboff 2019). In particular, the imaginary of the "smart city" as a generic technology-optimized vision for future cities is often used to justify political choices and trigger new economic paradigms benefiting corporate actors to the detriment of citizens (Sadowski and Bendor 2019; Vanolo 2014). This critique is supported by studies that show how images and expectations of the future structure actual decision-making and social organization (Jasanoff and Sang-Hyun 2015). This effectively *performs* the future in the present and often becomes a self-fulfilling prophecy (Meyer 2019). In other words, the ways in which we imagine the future—our "social imaginaries" (Taylor 2004)—shape how we act in the present. The performative dimension of smart city imaginaries is a theme that our installation attempts to engage with, as will be explained below.

This leads to several questions. First, we inquire who gets to define and articulate controversial issues. All too often, frictions and contestations associated with datafication tend to be framed externally and not by the people interacting with the data (Rettberg 2020). Instead, we ask: what might be the conditions that allow citizens to identify and debate their own issues of concern? Second, we investigate how the datafied city as a predominant urban imaginary instigates a continual renegotiation and redefinition of public values by multiple and diverse "publics" (Latour 2005). Different groups have conflicting viewpoints on the issues involved in datafication and the values they attach to them, as the security camera example shows.

We depart from the idea that "socio-technical controversies" (Jasanoff and Sang-Hyun 2009) can bolster civic engagement. Specifically, we focus on an approach to civic engagement that places friction and ongoing contestations around public values at the center (cf. van Dijck et al. 2019). We propose the notion of *controversing* as a deliberately frictional strategy for civic engagement that addresses the interlinked needs for recontextualizing, meaning making, and agency in debates around datafication (for a detailed discussion, see Baibarac-Duignan and de Lange 2021). This moves away from a singular data-optimized smart urban vision and helps to tease out a plurality of possible futures imagined by very diverse inhabitants. This is in line with other recent pleas to "stay with the trouble" (Haraway 2016) by "undoing optimization" in smart cities (Powell 2021) and valorizing the inherent messiness in interfacing with "smart" urban data (Mattern 2021). Controversing, we argue, has the potential to generate relational and dynamic forms of collective agency in reconfiguring urban futures.

This chapter analyzes how this research-by-design strategy for increasing civic engagement with the datafied smart city works by looking at the immersive installation *Future Frictions*.[1] The installation comprises an interactive digital interface combined with an immersive scenario-based web experience that engages the participants with the potential impacts of datafication through speculative future images of the datafied city. The intervention was developed as part of the NWO-funded project "Designing for Controversies in Responsible Smart Cities," developed by the University of Twente, Utrecht University, and a consortium of public and commercial partners such as the Amersfoort Municipality.[2] It was tested at the Media Architecture Biennale on June 28, 2021 during an online workshop with about 15 participants. Often, future-oriented design methods, such as "techniques of futuring" (Hajer and Pelzer 2018) aim to bring together actors around one or more imagined futures to support certain orientations for action. Our installation aims instead to create and support spaces for participants to imagine and debate desirable smart city futures, formulate potential controversies, and reflect on value clashes.

1 The development of the installation unfolded through an iterative co-creation process in which the research team worked together with a design agency (Design Innovation Group) and a collective of creative coders and programmers to develop the installation (Creative Coding Utrecht/Katpatat). We tested the installation as part of a workshop during the Media Architecture Biennale 2020 (MAB20), held online in June/July 2021 due to the COVID-19 pandemic.

2 The project is developed as a collaboration between University of Twente and Utrecht University, together with a consortium of public and private partners. It is aimed at developing a collaboration platform for envisioning and developing responsible smart cities, including ethical reflection on issues connected to urban datafication. See http://www.responsiblecities.nl.

In what follows, we reflect on the capacity of the installation to make typically abstract socio-technical controversies tangible, to challenge tacit assumptions, and to generate alternative images of desired futures that bring together different perspectives. The installation builds on critical and speculative design and supports diverse participants in visibilizing desirable smart city futures. The notion of "visibilizing" derives from STS (Prasad 2005) and is frequently used in design research for interventions that make tangible something that was abstract and intangible before, like technologies shaping the city (Matos-Castaño, Geenen, and van der Voort 2020). Visibilizing, in Latourian terms, entails "making things public by revealing and stimulating multiple perspectives to be expressed" (Latour 2005). With our approach, we move beyond the logic of solutionism, pervasive in smart city discourses, toward a space for material engagement with datafication, which we see as a precursor to collective imagination and action.

Methodological Inspirations: Speculative Design and Experiential Futures

The development of the *Future Frictions* installation was informed and inspired by previous work on speculative design (Dunne and Raby 2013; Auger 2013) and experiential futures (Candy 2010; Candy and Dunagan 2017). Speculative design revolves around creating artifacts based on future scenarios to materialize future social implications and aims to establish debate about (un)desirable futures and the potential for a plurality of actions. Instead of focusing on developing products or services on the basis of their functionality, speculative design fosters ethical reflection and responsibility. Making plausible futures tangible enables discussion about relevant ethical issues. In the context of smart cities, recent projects have explored the potential of speculative design to address, for instance, the lack of awareness regarding data nudging and its social implications (Park 2020). Speculative design focuses on opening spaces to discuss alternative futured by provoking social, ethical, and emotional questions that are often neglected in top-down smart city debates (Forlano and Matthew 2014).

Recently, speculative design has laid the foundation for experiential futures. Instead of designing objects or artifacts, experiential futures engage people with experiences or immersive situations. Experiential futures revolve around creating experiences that bring the worlds of tomorrow into the present to make futures "richer, more accessible, and immediate" (Candy

2010, 86). By engaging people with an experience, analogue or digital, it is possible to explore concrete manifestations of potential futures to instigate debate and gain insights about current actions that could be taken to avoid or achieve these futures. As opposed to written reports and presentations, experiential futures can make potential futures tangible and concrete (Pelzer and Versteeg 2019). In the words of Kuzmanovic and Gaffney, "[e]xperiential futures are a tool to crack open the door to multiple possibilities for change in the present" (2017, 107).

Although experiential futures rely on physical materiality to engage people in conversations about futures (Hajer and Pelzer 2018), videogames or web experiences can also offer possibilities for civic engagement. For instance, the use of interactive media may shape social imaginaries by providing inspiring alternatives (Bendor 2018). The practical potential of these forms of engagement derives from their scale and accessibility through online applications, as well as multisensory engagement with potential futures (Vervoort 2019). More conceptually, their performative potential lies in the "what if" question by opening up imaginative spaces for thinking about alternative futures. This matters because it helps to steer away from the suggested inevitability of techno-optimized futures that is performed by smart city visions, as discussed above.

Future Frictions: First Prototype

Based on these methodological inspirations, we developed an immersive first prototype. While the initial plan was to create a physical and spatially situated installation, due to COVID-19, we had to resort to a purely online 3D experience. It addresses potential smart city futures participants can easily relate to, in line with Auger (2013), but with the addition of a pinch of uncanniness for the sake of controversing smart and datafied urban futures.

In its first prototype, participants enter a virtual world and go on a quest to explore a neighborhood where a new smart city technology will be implemented. The task is to decide how the technology should be implemented, accomplished by making a choice among three potential outcomes, observing changes in the environment, and listening to what some of the residents have to say about the effects of the technology.

The technology we used for deliberately stirring up the debate following our strategy of controversing is a drone. Upon finding it, participants encounter three options of what should happen with the images that the drone takes and where they should be uploaded. Each of these three options exposes

Fig. 1. Screenshot of the *Future Frictions* interface (created by the authors).

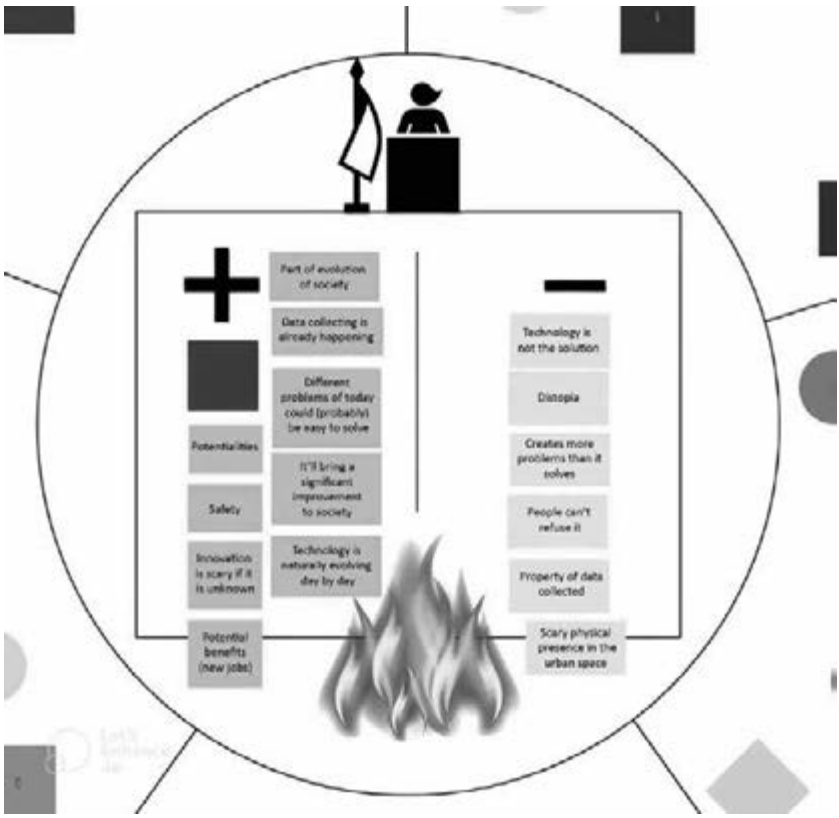
potential tensions: around corporate, community, and government control of the data. While these may initially appear straightforward, participants are sent back into the neighborhood to see how their choice has changed it, as well as the neighbors' experiences. This introduces a level of ambiguity and friction and provokes participants to formulate their own controversy. Toward the end, participants are prompted to reflect on the experience and the controversy by writing a postcard from their future neighborhood to a loved one. The postcard compels users to reflect on how they feel about the technology and the controversy identified and stimulates the imagination of different possible futures.

Identifying Smart City Controversies with *Future Frictions*

The authors tested *Future Frictions* in a participatory workshop to explore its potential for civic engagement during the Media Architecture Biennale 2020. The participants started by individually experiencing *Future Frictions*. This was followed by a collective discussion about smart city controversies, as raised through the individual experiences. To support interaction and debate, we used a digital canvas, Mural.co, and explicitly asked participants to reflect on:

a) *The controversies that surfaced in the web experience.* Examples noted by the participants included: the tension between anonymity and

Fig. 2. Screenshot of the digital canvas (Step 3) (created by workshop participants).



surveillance, sociability and isolation, and the boundary between collective and personal interests.

b) *How technology, as portrayed in Future Frictions, changed the experience of urban life*. Participants highlighted how a current experience of "not being known outside" could change into an experience of pervasive monitoring and less room for anonymity in the city.

c) *The responses and feelings about the impacts of technology on urban life that Future Frictions evoked*. Participants mentioned a wide array of feelings: "powerlessness, playfulness (experiment), fear, uncanniness, and endless possibilities."

d) *Aspects of urban life that were affected in the web experience and should be taken into consideration.* Participants discussed potential impacts on public street life, such as sociability and unexpectedness, and the fact that technology could make citizens more dependent on public authorities and government.

Fig. 3. The collage co-created by participants (Step 4) (created by workshop participants).

Following the discussion on controversies, the participants decided on the implementation of drone technology, in the context of the neighborhood displayed in the installation. We divided them into two groups, one supporting the implementation of the drone and one group against it. Moreover, one participant acted as a mayor to document the debate and make a final decision. The debate had two rounds during which participants switched roles so that those supporting the technology would be against it in the first round, and vice versa (Figure 2). Following the debate, the mayor decided to implement the drone technology, provided that certain conditions were met. In particular, she emphasized the importance of accepting the unavoidable evolution of technology, which demands a constant need for revisiting and reshaping powers to avoid losing control over technology and its impacts. Some of the participants suggested that establishing clear boundaries around the implementation of technology and seeing the city as a space for experimentation could help achieve this outcome.

The last step consisted in co-creating a collage to visibilize how the mayor's decision might shape the future city. Each participant added an element in the collage from an extensive collection of images we provided, and briefly discussed their choice in relation to the mayor's decision and the other participants' items. Moreover, the participants collectively had to suggest a title for their collage (Figure 3).

Collectively creating the collage consisted of navigating through different meanings and understandings of what the outcome of the visibilization meant. The various titles given to the collage highlighted this. They included: "The human factor," reflecting on the potential dehumanizing effects of technology in the city; "A 'perfect' place," suggesting how technology usually "repairs" seemingly inefficient aspects of the city; "Sweet community," imagining a utopian future for urban communities in control of, and serviced by, technology; and "Sweet troubles," proposing a controversy-fueled future by accepting the inevitability of technological innovations and the frictions these would likely cause.

Assessing the Installation's Potential Based on the Controversing Framework

Let us now analyze how the installation, which was built on the concept of controversing, allows us to analyze civic engagement with smart city issues and allows the debate to center around public values in the responsible "smart city." We do this by addressing the three elements of the controversing framework we developed: recontextualizing, meaning making, and agency (Baibarac-Duignan and de Lange 2021). *Recontextualization* involves the re-urbanizing of delocalized big urban data by situating contestations around datafication in specific spatiotemporal settings. *Meaning making* acknowledges the epistemological necessity to meaningfully "interface" with abstract datafication and to "translate" data into collective issues of concern, which are almost never univocal but instead rife with tensions. *Agency* considers the active role controversies can play in serving as a "glue" for engagement and collective action, where the onus in participatory processes is in on the conditions that enable participation in the shaping of smart city futures.

Future Frictions recontextualizes smart tech in tangible ways by making its implications material and concrete through changes in the actual environment, the characters' experiences, and social interactions between them. As one of the participants in the workshop noticed, the installation "gives body to the technology." The impacts of technology become tangible by visibly modifying the surrounding environment and therefore the user's experience. Moreover, the user is not presented with an un-relatable reality, for instance that of a sleek techno-futuristic environment, which could potentially alienate rather than engage them in the experience. Instead, the installation brings controversies at a "human scale," using images of existing urban environments and inhabitants to shape an imagined future.

Experiential methods such as data walking have highlighted the benefits of walking in raising awareness around datafication (Houston, Gabrys, and Pritchard 2019; Powell 2018; van Zoonen et al. 2017). Walking in the physical environment can have an important role in grounding data as material, situated and embedded in everyday life practices (van Es and de Lange 2020). We propose that our immersive installation presents a similar potential, helping to recontextualize the abstract notion of datafication by bringing it closer to people's everyday urban realities.

Future Frictions fosters meaning making by rendering visible the effects of the drone on the environment and on the characters' experiences; in this way, the installation offers tangible evidence of what tend to be ephemeral datafication processes. The installation does not provide clear-cut answers or pre-defined controversies. Each of the three options presents the user with both positives and negatives. For instance, the fact that public authorities have access to the images recorded by the drone offers a feeling of safety to the teenage girl, but it results in undesired help for the elderly woman and thus a feeling of powerlessness. This ambiguity enables the user to reflect on their values and identify their own controversies as points where the control of technology impacts the imaginary boundaries preserving these values. Moreover, the workshop reinforced ambiguity through an agonistic element when participants changed roles and metaphorically stepped into the shoes of participants with opposing views via the Mural canvas. This process allowed the creation of a shared situation and a common baseline of knowledge (i.e., based on the shared experience of the installation) for participants coming from different backgrounds. In this first iteration, the digital canvas acted as a "meaningful interface" (de Lange 2019) helping to generate group discussions about emerging controversies around datafication and translating them into a shared matter of concern.

As for agency, *Future Frictions* allows participants to formulate their own concerns through collective interactions. While we developed the installation and an overall simple narrative, participants have the freedom to follow their own path, make sense of the technology proposed for debate, and articulate controversies arising from its use. Through this, stories emerge that become the basis for the postcards from the future. From this perspective, participation is not equated with how we as researchers tell *our* story or involve the user in an experiential journey through the speculative neighborhood. Participation emerges from the opportunities that *Future Frictions* affords for interactions and shared reflections through the postcards and the workshop canvas. This strategy is specifically developed to counterbalance power relations between the creator and the user and the

presumed increased participation in initiatives presenting data visualizations that frame the story in particular ways (Rettberg 2020; Söderström, Paasche, and Klauser 2014). The MAB workshop added a further dimension to the individual reflection by providing a space in the digital canvas for collective debate (an agonistic element) and acting on the outcome of the debate through making a collage using a collection of images. The collage represents the outcome of their group debate on the controversies raised by the installation, visibilizing an image of a future city that reflects the participants' diverse values, worries, and hopes. By creating conditions for collective reflection on the mediating roles of technology in the city (Verbeek 2015), the canvas materializes the potential for collective action.

Discussion: Widening Engagement in Shaping Urban Futures

This chapter has shown how *Future Frictions* as a frictional media architectural interface challenges singular visions of techno-oriented futures and serves to increase awareness, debate, and reflection. We analyzed how *Future Frictions* engaged people around otherwise abstract and intangible issues of datafication in today's cities and allowed participants to imagine alternative urban futures through *controversing* as the purposeful use of friction and contestation.

As a first reflection, the *controversing* framework centers on value plurality and controversies. This allows us to move beyond normative and prescriptive futuring techniques aimed at providing pre-defined images of urban futures (Oomen, Hoffman, and Hajer 2021). The goal of *Future Frictions* is to empower people to imagine, shape, and reflect on alternative futures by engaging with controversies. From a methodological perspective, the installation, as addressed in the context of the MAB workshop, offers an element of agonism through role-playing and aspects of critical making through the collage, which allows participants to act on the emerging controversies. Thus, participants do not gather around an externally formulated desirable future (e.g., Hajer and Pelzer 2018) but rather engage around multiple futures as an "issue" (Marres 2007). The web experience supports the participants in identifying their own values and controversial issues and becomes a means to collectively "make" an image of a desirable urban future. This future reflects their diverse values, which materialize in aspects of the city and urban life that they find important.

A second reflection relates to the role of critical and speculative design in making the future accessible. There is a need to widen the debate on our

socio-technical futures, and approaches like speculative design and critical making, together with media architectures as interfaces, offer meaningful and accessible entry points to achieve this. Although speculative design aims to spark debate and reflection by opening alternative futures, it is often perceived as elitist and distant from the realities of a wider audience (Forlano and Matthew 2014; Beattie et al. 2020; Kozubaev et al. 2020). By using accessible language and recognizable urban elements, *Future Frictions* speculates about provocative yet relatable urban futures that trigger reflections on the impacts of technology on everyday life, now and in the long term. Through speculation, *Future Frictions* brings to the surface social interactions and potential power relations that stakeholders may have in potential futures. This way, the installation raises questions not only about a specific technology but also about the socio-technical context in which it exists. These insights are in line with Wong et al. (2020), who acknowledge the potential of infrastructural speculations for civic engagement.

A third reflection relates to the relevance of making abstract phenomena like "controversies" and "datafication" tangible and accessible. Our experience with *Future Frictions* shows the value of material and embodied engagement in involving a wider audience in socio-technical controversies, even if it takes place in the digital realm. The installation supports communicating complex socio-technical theories in a way that allows for a more even relationship between researchers, citizens, and other stakeholders. By *controversing* through speculation, *Future Frictions* highlights the politics existing in smart cities, moving away from homogenous perspectives around technological impacts that focus on utopian or dystopian consequences. This fosters constructive ambiguity to enable participants to reflect on the values they consider important in the city, as well as challenging tacit assumptions to generate images of desired futures that bring together different perspectives. Combined, these three points make up the performative dimension of the *Future Frictions* installation in producing possible alternatives.

We acknowledge the limitations of the purely digital interactions and methods discussed in this paper. Yet we feel there is a need for critical and creative design methods that stimulate the imagination beyond externally formulated urban visions and toward a plurality of potential futures. Visualizing data in meaningful ways or even providing immersive experiences of desirable futures is not sufficient to challenge current practices. *Future Frictions* as presented here is no silver bullet. It brings together participants already willing to debate controversies and values together and to envision alternatives. Actual tenacious controversies rarely spawn from controlled environments. In a next iteration, we hope to bring *Future Frictions* into

public space to explore how tensions emerge as people reflect on technological impacts and express their values on the ground.

Despite its limitations, as a frictional interface, *Future Frictions* offers a glimpse into how media architectures could become mediators in processes of widening participation in imagining futures we desire for our cities. We hope our approach inspires other researchers to become attuned to frictions arising from the datafication of cities, study controversies through experimental and co-creative settings, and create the conditions for people to formulate their own issues, tensions, and values around new technologies and to use their imagination for speculative criticality. Ultimately, it is about challenging privileged positions in our collective imagination and "staying with the trouble" of having other actors at the table. Freeing collective imagination then becomes an act of social emancipation, which might be just the key to building more inclusive urban futures together.

Acknowledgements

This work is partly financed by the Dutch Research Council (NWO; project number: CISC.CC.012). We thank our partners (Aerovision, Design Innovation Group, Future City Foundation, Marxman, and Municipality of Amersfoort) for their valuable contributions to the research. We thank Creative Coding Utrecht and Katpatat for their collaboration on creating the installation and Beatrijs Hinloopen for the images rendered.

References

Baibarac-Duignan, Corelia, and Michiel de Lange. 2021. "Controversing the Datafied Smart City: Conceptualising a 'Making-Controversial' Approach to Civic Engagement." *Big Data & Society* 8, no. 2: https://doi.org/10.1177/20539517211025557.

Beattie, Hamish, Daniel Brown, and Sara Kindon. 2020. "Solidarity Through Difference: Speculative Participatory Serious Urban Gaming (SPS-UG)." *International Journal of Architectural Computing* 18: 141–54.

Bendor, Roy. 2018. *Interactive Media for Sustainability*. London: Palgrave Macmillan.

Candy, Stuart. 2010. "The Futures of Everyday Life: Politics and the Design of Experiential Scenarios." PhD diss., University of Hawai'i at Mānoa. https://doi.org/10.13140/RG.2.1.1840.0248.

Candy, Stuart, and Jake Dunagan. 2017. "Designing an Experiential Scenario: The People Who Vanished." *Futures* 86: 136–53.

de Lange, Michiel. 2019. "The Right to the Datafied City: Interfacing the Urban Data Commons." In *The Right to the Smart City*, edited by Paolo Cardullo, Cesare Di Feliciantonio, and Rob Kitchin, 71–83. Bingley, UK: Emerald.

Forlano, Laura, and Anijo Mathew. 2014. "From Design Fiction to Design Friction: Speculative and Participatory Design of Values-Embedded Urban Technology." *Journal of Urban Technology* 21, no. 4: 7–24.

Forlano, Laura. 2019. "Cars and Contemporary Communications | Stabilizing/Destabilizing the Driverless City: Speculative Futures and Autonomous Vehicles." *International Journal of Communication* 13: 2811–38.

Hajer, Maarten A., and Peter Pelzer. 2018. "2050 – An Energetic Odyssey: Understanding 'Techniques of Futuring' in the Transition towards Renewable Energy." *Energy Research and Social Science* 44 (October): 222–31. https://doi.org/10.1016/j.erss.2018.01.013.

Haraway, Donna J. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham, NC: Duke University Press.

Houston, Lara, Jennifer Gabrys, and Helen Pritchard. 2019. "Breakdown in the Smart City: Exploring Workarounds with Urban-Sensing Practices and Technologies." *Science, Technology, & Human Values* 44, no. 5: 843–70. https://doi.org/10.1177/0162243919852677.

Jasanoff, Sheila, and Sang-Hyun Kim. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications. https://doi.org/10.4135/9781473909472.

Kozubaev, Sandjar, C. Elsden, N. Howell, M.L.J. Søndergaard, N. Merrill, B. Schult and R.Y. Wong. 2020. "Expanding Modes of Reflection in Design Futuring." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Kuzmanovic, Maya, and Nick Gaffney. 2017. "Enacting Futures in Postnormal Times." *Futures* 86: 107-17.

Latour, Bruno. 2005. "From Realpolitik to Dingpolitik or How to Make Things Public." In *Making Things Public. Atmospheres of Democracy*, edited by Bruno Latour and Peter Weibel, 14–43. Cambridge, MA: MIT Press.

Marres, Noortje. 2007. "The Issues Deserve More Credit: Pragmatist Contributions to the Study of Public Involvement in Controversy." *Social Studies of Science* 37, no. 5: 759–80. https://doi.org/10.1177/0306312706077367.

Matos-Castaño, Julieta, Anouk Geenen, and Mascha van der Voort. 2020. "The Role of Participatory Design Activities in Supporting Sense-Making in the Smart City." In *Synergy – DRS International Conference 2020*, edited by Stella Boess, Ming Cheung, and Rebecca Cain. Held online.

Mattern, Shannon. 2021. *A City is Not a Computer: Other Urban Intelligences*, 1st ed. Princeton, NJ: Princeton University Press.

Oomen, Jeroen, Jesse Hoffman, and Maarten A. Hajer. 2021. "Techniques of Futuring: On How Imagined Futures Become Socially Performative." *European Journal of Social Theory* (January 27). https://doi.org/10.1177/1368431020988826.

Park, Joo Young. 2020. "Smart Bench: A Speculative Design to Create Critical Awareness of Data-driven Nudging in the Smart City." MA thesis, Delft University of Technology.

Pelzer, Peter, and Wytske Versteeg. 2019. "Imagination for Change: The Post-Fossil City Contest." *Futures* 108: 12–26.

Powell, Alison. 2018. "The Data Walkshop and Radical Bottom-up Data Knowledge." In *Ethnography for a Data-Saturated World*, edited by Hannah Knox and Dawn Nafus, 212–32. Manchester: Manchester University Press. https://doi.org/10.7765/9781526127600.00018.

Powell, Alison. 2021. *Undoing Optimization: Civic Action in Smart Cities*. New Haven, CT: Yale University Press.

Prasad, Amit. 2005. "Making Images/Making Bodies: Visibilizing and Disciplining through Magnetic Resonance Imaging (MRI)." *Science, Technology, & Human Values* 30, no. 2: 291–316. https://doi.org/10.1177/0162243904271758.

Rettberg, Jill Walker. 2020. "Situated Data Analysis: A New Method for Analysing Encoded Power Relationships in Social Media Platforms and Apps." *Humanities and Social Sciences Communications* 7, no. 5. https://doi.org/10.1057/s41599-020-0495-3.

Sadowski, Jathan, and Roy Bendor. 2019. "Selling Smartness: Corporate Narratives and the Smart City as a Sociotechnical Imaginary." *Science, Technology, & Human Values* 44, no. 3: 540–63. https://doi.org/10.1177/0162243918806061.

Schäfer, Mirko Tobias, and Karin van Es. 2017. *The Datafied Society: Studying Culture Through Data*. Amsterdam: Amsterdam University Press.

Söderström, Ola, Till Paasche, and Francisco Klauser. 2014. "Smart Cities as Corporate Storytelling." *City* 18, no. 3: 307–20. https://doi.org/10.1080/13604813.2014.906716.

Taylor, Charles. 2004. *Modern Social Imaginaries*. Durham, NC: Duke University Press.

Townsend, Anthony. 2013. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: W.W. Norton and Company.

van Dijck, José, Thomas Poell, and Martijn de Waal, eds. 2018. *The Platform Society: Public Values in a Connected World*. Oxford: Oxford University Press.

van Es, Karin, and Michiel de Lange. 2020. "Data with Its Boots on the Ground: Datawalking as Research Method." *European Journal of Communication* 35, no. 3: 278–89. https://doi.org/10.1177/0267323120922087.

Vanolo, Alberto. 2014. "Smartmentality: The Smart City as Disciplinary Strategy." *Urban Studies* 51, no. 5: 883–98. https://doi.org/10.1177/0042098013494427.

van Zoonen, Liesbet, Fadi Hirzalla, Jiska Engelbert, Linda Zuijderwijk, and Luuk Schokker. 2017. "'Seeing More than You Think': A 'Data Walk' in the Smart City." *Bang the Table* (blog). https://www.bangthetable.com/blog/data-walk-in-smart-city/.

Verbeek, Peter-Paul. 2015. "Beyond Interaction: A Short Introduction to Mediation Theory." *Interactions* 22, no. 3: 26–31. https://doi.org/10.1145/2751314.

Vervoort, Joost M. 2019. "New Frontiers in Futures Games: Leveraging Game Sector Developments." *Futures* 105 (January): 174–86. https://doi.org/10.1016/j.futures.2018.10.005.

Wong, Richmond. Y., Vera Khovanskaya, Sarah E. Fox, Nick Merrill, and Phoebe Sengers. 2020. "Infrastructural Speculations: Tactics for Designing and Interrogating Lifeworlds." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

About the Authors

Corelia Bairbarac-Duignan is Assistant Professor in the Department of Technology, Policy and Society, University of Twente. Previously, she researched creative forms of civic engagement in datafied smart cities at Utrecht University. Originally trained as an architect, she has research experience in the fields of urban sustainability and technology-mediated civic participation.
> c.e.baibarac@utwente.nl

Julieta Matos Castaño a postdoctoral researcher at DesignLab, Human Centred Design group [u twente]. Her research interests lie in the development of approaches to support the making of collaborative futures. In this project, she explores ways to develop responsible smart city futures, drawing on participatory and speculative design practices, inclusive imaginaries, and sensemaking processes.
> j.matoscastano@utwente.nl

Anouk Geenen is a PhD candidate at DesignLab, Human Centred Design group [u twente]. Although she is trained as a theoretical physicist, she is keen on working interdisciplinarily and finding the spot where science

and society meet. In this project, Anouk explores designerly ways to enable diverse stakeholders to reflect ethically upon smart city technologies.
> a.j.p.geenen@utwente.nl

Michiel de Lange is Assistant Professor at Utrecht University in the Department of Media and Culture Studies and co-PI in the project "Designing for Controversies in Responsible Smart Cities." He is co-founder of The Mobile City, a platform for the study of new media and urbanism; a core member of the [urban interfaces] research group at Utrecht University; and a researcher in the field of (mobile) media and urban culture.
> m.l.delange@uu.nl

4. Streaming against the Environment

Digital Infrastructures, Video Compression, and the Environmental Footprint of Video Streaming

Marek Jancovic and Judith Keilbach

Abstract

Building on an infrastructural approach, this chapter investigates the environmental impact of video streaming. It clarifies some of the less obvious relationships between media infrastructures, video compression standards, and electronics supply chains and demonstrates how their interactions unfold material effects on the environment. Complicating recent critical research on data centers, we posit that existing models for calculating the ecological footprint of video streaming cannot capture its full extent and advocate for an interdisciplinary approach to data, computation, and infrastructure. This approach informs our argument that the development of new compression standards redistributes environmental responsibility in a way that benefits streaming providers and data centers at the expense of end users and hardware manufacturers.

Keywords: video streaming, video compression, infrastructural inversion, environmental footprint, data infrastructure

In 2020, while many were staying at home due to the COVID-19 pandemic, internet traffic rose to an unprecedented high. Studies reported momentary surges of web conferencing, gaming, and video streaming of up to 300% and a general increase of internet traffic by about 40% above the expected annual growth (Feldmann et al. 2021). Internet service providers had to take short-term measures and increase their capacity. The European Commission even asked streaming services and internet users to relieve the pressure on internet infrastructure (European Commission 2020), and Netflix and YouTube lowered their bandwidth demands to help prevent data overload.

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH04

This bandwidth crisis foregrounds the physical realm of data traffic upon which digital culture rests. While diaphanous metaphors like "the cloud" or "streaming" evoke the impression of immateriality (Carruth 2014; Blanchette 2011), media theorists have long been pointing out that all data is bound to hardware (e.g., Kittler 1995). In this chapter, we take up on these insights and investigate the materiality of video streaming. We emphasize computational processes and their impact on the environment, thereby echoing Paul Edwards (2021), who views algorithms as a core element of digital infrastructures.

As part of our argument, we posit that existing models for calculating the greenhouse gas emission of video streaming cannot capture the full extent of its ecological footprint. To better understand the environmental effects of our digital media culture, it is important to develop approaches that bring together perspectives from various disciplines including media studies, critical data studies, science and technology studies, environmental studies, and information science. Such an interdisciplinary approach informs our argument that new compression standards benefit streaming providers and data centers at the expense of increased energy use on the users' side. This allows streaming services to gradually divert environmental responsibility to consumers and hardware manufacturers, even as they continue developing and advocating for increasingly energy-hungry video standards.

Following Lisa Parks and Nicole Starosielski (2015) and aligning methodologically with what science and technology studies call "infrastructural inversion" (Bowker et al. 2009, 98), we are taking on an infrastructural perspective to emphasize the materiality of video streaming and understand how the physical and sociotechnical arrangements of electronic devices, data centers, and other network facilities interconnect with other technical systems, computational processes, and technological standards.[1,2] Such an approach is not only suitable to render transparent technical systems visible, but it also explores the environmental impact of digital media culture. Previous research into the ecological footprint of digital media has focused, among other issues, on the extraction of raw materials and the enormous amount of waste that the rhetoric of immateriality usually conceals (e.g., Gabrys 2011; Maxwell and Miller 2012; Cubitt 2017), but our goal is to clarify

1 Such an inversion can be understood as a "figure-ground gestalt shift" (Star and Ruhleder 1996, 113) that aims to make infrastructures visible. In science and technology studies, this is one of the methods to study them. See Bowker et al. 2009, 98.

2 The infrastructure of video streaming is what Paul Edwards calls a second-order system (2021, 317), because it relies on other telecommunication and electricity infrastructures.

some of the less obvious relationships between video compression standards, media infrastructures, and electronics supply chains and demonstrate how their interactions unfold material effects on the environment. More specifically, we posit that video compression algorithms play a crucial role in the ongoing shift of the environmental costs of streaming from data centers to end users. Conceptually, this claim ties in with Jonathan Sterne's (2012, 250) suggestion that research on data compression techniques can serve as a point of entry toward richer theories and histories of media. The consideration of tangible material elements (such as raw materials or electronic waste) is important, but we should also not lose sight of computational processes, whose environmental effects, as we argue, cannot be fully captured by tools like carbon footprint calculators. The situation in the Netherlands offers a particularly compelling example, because it demonstrates the complex position of data centers in the debate.

After briefly delineating the infrastructure of video streaming and addressing some recent controversies surrounding Dutch data centers, we discuss the complexity of calculating the CO_2 footprint of streaming and then focus on the environmental ramifications of compression standards, addressing in particular the notion of compression efficiency that drives the development of new video standards.[3] We show how streaming services' standards-making activities result in increased energy consumption by end devices. In effect, each new compression standard gradually shifts the responsibility for sustainable action away from data centers and streaming services and onto viewers and end users. Our argument touches upon the limitations of existing ways of calculating and conceptualizing environmental impact, and we hope to increase awareness of the role our media habits and media devices play in contributing to energy consumption.

The Infrastructure of Video Streaming

Infrastructures are socio-technical systems that provide critical services to our society (Edwards 2021, 314), with power grids, water supply, railroads, and telecommunication networks as archetypal examples. Infrastructures emerge from an interplay of technology and socio-political factors (such as social transformations, consumer demands, regulations, and policies) and comprise a variety of elements, including technologies, institutions,

3 For more information, see https://www.washingtonpost.com/climate-environment/2022/05/28/meta-data-center-zeewolde-netherlands.

financial schemes, built environments, work processes, etc.[4] Once they are built, infrastructures sink "into an invisible background": they are "just there, ready-to-hand, completely transparent" (Star and Ruhleder 1996, 112), unless a breakdown, like a power outage or the collapse of a bridge, makes them visible.[5]

The infrastructure that facilitates video streaming is similarly invisible and taken for granted. However, an increasing interest in the physical, social, and political materiality of media distribution is rendering it more and more visible (e.g., Holt and Sanson 2013; Lobato 2019). Data centers have become a preferred object of study when investigating digital infrastructures, not least because they are the site of intersection for a range of pressing issues such as data mining, large-scale surveillance, geopolitics, and data sovereignty. Data centers have provoked discussions about the corporate use of public services (Hogan 2015; Brodie 2020) and stimulated reflections on land use and physical space (Johnson 2019; Vonderau 2019; Mayer 2020), on power sources and cooling systems (Hogan 2015; Velkova 2016) and on the energy demand of machine learning (Rohde et al. 2021; Tarnoff 2020).

Data centers' environmental impact is ambiguous, particularly in the context of video streaming, as we will discuss in more detail below. On the one hand, data center operators in some regions, such as the Netherlands, are increasingly committing to carbon neutrality and energy efficiency (DDA 2020; Kamiya 2020). On the other hand, their corporate environmentalism has been criticized as a strategy to "reduce, refuse, and redistribute the relations between carbon and data," which preempts ecological critique (Pasek 2019, 2). Research on, for example, failed plans for the reuse of data center–generated waste heat in the Netherlands (van Kessel 2021a) or Anna Pasek's probe into Microsoft's renewable energy purchases and system of carbon offsetting (Pasek 2019) have demonstrated this point convincingly. Focusing on video compression, we argue that video streaming services are similarly reshaping environmental relations by passing on responsibility to hardware manufacturers and end users. Laura Marks et al. (2021) advocate that in addition to data centers, end user devices must be included in calculations of the energy consumption and carbon footprint of video streaming. Somewhat counterintuitively, this is not always the case, as consumer

4 Brian Larkin points to the conceptual unruliness of infrastructures that are things and, at the same time, the relation between things (2013, 329). Scholars of science and technology studies emphasize this relationality as well, since the "work of one person is the infrastructure of another" (Bowker et al. 2009, 98); see also Star and Ruhleder (1996, 122–23).

5 Repair and maintenance are similar moments that make infrastructures visible; see Henke and Sims (2020).

devices are often seen as outside the internet system boundary. We concur with this view, and in the following analysis of streaming infrastructure and compression algorithms, we provide further reasons as to why end devices are critical in these considerations.

Public and academic interest in streaming infrastructures has been growing in recent years, and the global chip shortage caused by the COVID-19 pandemic has prominently drawn attention to some of its lesser-known elements, such as semiconductor manufacturers. But the topography of streaming also comprises other elements that continue operating in relative obscurity and have yet to be addressed by critical humanities scholarship. To sketch the close entanglements between hardware, data, infrastructure, and standards, let us briefly recapitulate a part of the life cycle and supply chain of a chip, like the graphics processing unit that a smartphone or television might use to decompress video.

A chip manufacturer—a company such as Mediatek or NVIDIA—designs the chip hardware. Chip manufacturers closely follow the development of video compression standards (in fact, they might actively take part in their creation, as NVIDIA does) and will design their chips to allow the processing of new and emerging video formats. Neither Mediatek nor NVIDIA physically produce any chips but outsource their fabrication to semiconductor foundries like TSMC in Taiwan. The chip might be bought and further handled by a hardware and consumer electronics manufacturer like MSI or Sony, who assembles it into larger components such as graphics cards, smartphones, or televisions, or by an integrator, who installs firmware on devices such as set-top boxes. Telecommunications and pay television companies offer such devices with their services, which often include partnerships with streaming providers such as Netflix. A company like Netflix, in turn, purchases computing power from data center operators like Amazon Web Services. Data centers then deliver video to end users through various forms of wired and wireless infrastructures, which are maintained by network operators. The Netflix app fetches video signals from the data center, which are processed by the chip and displayed on your screen.

All these actors maintain complex logistical relationships with each other, and many of them enter partnerships to develop new industry standards. These partnerships take the form of consortia such as the Alliance for Open Media, whose influence on the energy use of our electronics we address further below. The standards development process results in new video formats through which economic cooperations and rivalries are negotiated (Volmar 2020). At this point, what is important to us is that video compression standards, despite seeming like abstract documents that only deal with

the computation of data as disconnected from issues of materiality, actually exert significant material effects on the environment.

As we show in the following section, grasping and calculating the full extent of these material effects is troublesome. The complexity of the streaming infrastructure and supply chains outlined above, as well as their relationship with, and dependency on, other technologies, complicates the assessment of video streaming's environmental impact. By focusing on video compression, we want to emphasize that streaming services like Netflix or YouTube are not only built on top of (landline and undersea) cable communications systems, cellular networks, and power grids. Video streaming is also enabled by standards, protocols, and compression algorithms and software—which all need to be factored in when taking on an environmental disposition.

Calculating the Environmental Impact of Video Streaming

Modeling the environmental footprint of video streaming is notoriously complex. It is difficult to quantify the pressure on landscape, water quality, and biodiversity that data centers, cable installations, cooling systems, and energy supply place on the environment. Even when it comes to carbon emissions and energy consumption, estimates of CO_2 emitted and kWh consumed per hour of video streaming differ by up to three orders of magnitude, depending on whom and when you ask (Aslan et al. 2018; Marks et al. 2021).[6]

A comparison of two recent sources illustrates the scope of the uncertainty. According to a white paper by the London-based Carbon Trust (2021), streaming one hour of video in Europe produces 55g of CO_2-equivalent emissions. Obringer et al. (2021) found a value of 441g for the same activity, more than eight times as much.[7] The large discrepancy between these studies, both of which claim to be using up-to-date data, can be partially explained by differences in the underlying assumptions, such as disparities in the proportion of sustainable electricity in a particular region's energy mix or emission factors. But these differences matter not only in a numerical sense. It is important to recognize that they are also used to underscore

6 Existing models have been criticized for a range of reasons: severely over- or underestimating variables like bitrates and wattage, overlooking parameters like device type and screen size, building upon obsolete data, mistakenly correlating data traffic with energy use, erroneously extrapolating energy use growth from storage capacity growth, or failing to account for energy efficiency gains (Kamiya 2020; Masanet et al. 2020; Carbon Trust 2021).

7 The difference of 386g corresponds to the CO_2 emitted by driving 3.1 kilometers in a recently manufactured passenger car (European Environment Agency 2021).

specific attitudes toward climate action: Obringer et al. use their numbers to emphasize the need for regulatory intervention, whereas the Carbon Trust report concludes that "the carbon footprint of viewing one hour of video streaming is very small compared to other everyday activities" (2021, 8), implying that the video streaming industry does not merit immediate attention. It comes as no surprise that reactions in high-profile media outlets to scientific research have been just as contradictory, covering the entire spectrum from alarmist to appeasing (e.g., Daigle 2020; cf. Kaufman 2020).

Reliable data about the environmental impacts of video streaming are not only difficult to obtain; they also become obsolete very quickly. Changing consumption patterns, fluctuations in power production and demand, and the rapid succession of new technologies confound existing models and necessitate continuous adjustments in the calculations. This can be exemplified by the energy consumption of Dutch data centers. Statistics indicate that global internet traffic has tripled, and data center workloads have doubled since 2015. But thanks to improvements in energy efficiency, data centers' energy use has remained constant (Kamiya 2020). Some data center operators have also achieved remarkable improvements in decarbonization. The Dutch Data Center Association reported that 86% of its data centers operate on renewable sources, even reaching 99% in the Amsterdam region where three quarters of the nation's operators are located (DDA 2020).[8]

While these numbers are laudable, they hide other environmental frictions. For one, data centers rely on the common grid, and their energy consumption places a large burden on an electrical infrastructure already at its limit in regions like Amsterdam.[9] Furthermore, there are fears that the staggering amount of cooling water that Dutch data centers consume could jeopardize water supplies (van Kessel 2021b). This is a major risk factor in a country where groundwater quality is deteriorating and water shortages are increasingly common (van Engelenburg et al. 2021). Furthermore, ambitious plans to reuse data centers' residual heat to warm homes and offices frequently fizzle out. Data centers tend to opt to be physically close to electricity sources and cable landing points. Such locations are often unsuitable for heat networks.[10] And thus, while waste heat reuse is often

8 Dutch data centers purchase their green energy via certificates, an industrial practice that Pasek (2019) criticizes as negating the environmental impact while driving up the grid demand and therefore failing to reduce carbon emission.

9 Data centers constitute 4.2% of the country's entire electricity usage, according to 2019 estimates from the DDA and Statistics Netherlands.

10 In addition, the relatively low-temperature heat that data centers generate also makes them uninteresting for many heat networks.

touted as one of data centers' great contributions to a more sustainable resource economy, hyperscale centers such as the one built by Google in the port of Eemshaven simply dissipate heat into thin air (van Kessel 2021a).

These environmental and infrastructural complexities pose methodological hurdles. Proper impact assessment not only needs to be multilateral by considering greenhouse gas emissions and energy consumption as well as water and land use, as Obringer et al. (2021) have argued. It also needs to handle temporal intricacies that necessitate constant adjustments to the data, and it must deal with spatial challenges, such as national and regional differences. Calculating for collateral impact on other, underlying infrastructures—electricity or water supply, for instance— demands the consideration of even more comprehensive factors. Ultimately, these complications only underscore the necessity of interdisciplinary research and of relational thinking regarding the character of digital infrastructures beyond what is usually considered to be "the media" or "the internet."

But alongside questions of infrastructure on the macro scale, microscale computational processes also contribute significantly to the total environmental effects of video streaming, as we show in the following.

The Material Effects of Video Compression

Compared to sound or still images, video requires an enormous amount of data to look reasonably good. Such large amounts of data are impractical to store and distribute, because storage space and network bandwidth is limited. Shrinking these data means they can be delivered to end users much more easily. Lossy compression—that is, compression in which some of the original information is removed permanently—ensures that a video recording you made with your phone is 30 rather than 3000 MB big. The algorithms that achieve this are a vital element of streaming infrastructure.

Scientific discourses on video compression frequently feature two divergent notions of efficiency. A brief discussion of these terms is useful in articulating the role of compression in what Allison Carruth (2014) calls "the micropolitics of energy." Together, the concepts of *computational* efficiency and *compression* efficiency can help us interrogate the chain of relationships that connects calculations inside a processor to large network and electricity infrastructures, and to an even larger political economy of video streaming and global hardware supply chains.

The basic principle of lossy video compression is that much of the visual information in a moving image can be discarded without becoming

noticeable to humans. Specialized algorithms are used to identify and then eliminate these redundant data. On the most elementary level, this process boils down to adding and multiplying numbers. Algorithmically, any given compression method might be realized with more or fewer operations in (more or less) efficient ways. An algorithm that solves a problem with fewer operations is said to be more computationally efficient. A good example of this is a mathematical tool known as the discrete cosine transform (DCT). The DCT and its derivatives are used to reduce the amount of data in practically every major digital audiovisual format, from JPEG images and online video to DVD, Blu-ray disks, and digital TV broadcasting standards. The DCT can be computed in many ways, some of which are faster than others. Algorithms can be sped up with various mathematical shortcuts that exploit the structure of the processed data, take advantage of certain properties of trigonometric functions, or utilize knowledge of the processing hardware—for example, how much longer a specific electronic circuit needs to multiply two numbers as opposed to adding them.

The savings among different algorithms can be minuscule in a relative sense. For instance, the 2D Arai-Agui-Nakajima DCT algorithm from 1988 requires 464 additions, the Feig Fast DCT (1990) requires 462, and the Generalized Chen Transform (1994) requires 608 additions but no multiplications (Kuhr 2001). But these algorithms are run thousands of times for every single frame of video that flickers across our screens. With trillions of calculations performed daily on the scale of visual culture, a difference of two additions per block of data translates into enormous savings in computation time. In this way, the microtemporalities of compression scale up to tangible and environmentally significant fluctuations on the level of infrastructure in the form of increased or decreased electricity demand.

Aside from computational efficiency, the interrelated notion of *compression* efficiency is equally environmentally significant. Compression efficiency refers to how much smaller an algorithm can make a video file at a given image quality. Streaming services have an incentive to strive for the highest possible compression efficiency, because smaller video files can be delivered to end devices faster and counteract what Neta Alexander (2017, 8) has called "digital dams"—the experiences of network latency, delay, and buffering.

New compression algorithms are constantly being developed, and the process of standardizing, implementing, and promoting them carries significant vested interests. Companies like Netflix are continuously optimizing compression efficiency, re-encoding their catalogs as more efficient techniques emerge, and performing adjustments to encoding

parameters to decrease file sizes and increase perceived image quality.[11] As with computational efficiency, there are numerous strategies for increasing compression efficiency. For example, if the encoding algorithm is taught to "understand" the notion of film editing, it can recognize cuts in moving images and operate with individual shots, thus compressing motion more efficiently. The result is a video file with a smaller size, which means less buffering and data consumption.

However, from an environmental disposition, the salient point is that more "efficient" compression schemes are also more complex and therefore consume more energy (Lin, Liu, and Liao 2010; Sharrab and Sarhan 2013; Ejembi and Bhatti 2014; Monteiro et al. 2015; Uitto 2016; Kränzler, Herglotz, and Kaup 2020). The encoding device draws more electricity in order to compress data more heavily. When the file is decompressed during playback, the decoding also generally requires more power from a television set and drains the batteries of a mobile device faster. These batteries then need to be recharged more often and their capacity diminishes more rapidly, decreasing the device's lifespan and accelerating the rate at which electronic waste is produced. The speed of video streaming is paid with environmental costs that ultimately contribute to the warming of the planet. This is "the materiality of media heat" (Starosielski 2014) at work—the concrete effect of video compression on the physical world.

Data centers play a comparatively minor role with respect to compression and the accompanying energy consumption, despite their environmental costs mentioned above. Data centers and content distribution networks encode and store the video files that eventually get delivered to end users, but the computational work of decompression is performed by the billions of devices at the end of the delivery chain. This is one of the reasons why end-user devices are responsible for about half of the energy consumption of all digital services (DDA 2020; Malmodin and Lundén 2018).[12] Our television sets, laptops, smartphones, gaming consoles, and set-top boxes consume the largest proportion of electricity required to view video. And with each more complex generation of video compression standards, they consume progressively more than the rest of streaming infrastructure.

The streaming industry capitalizes on the growing processing power of these devices. By leveraging computationally demanding compression algorithms, streaming providers ensure that the data centers and cable

11 For concrete examples, see Sole et al. (2018) and Mavlankar et al. (2020).

12 Some recent research claims that data centers account for less than 1% of video streaming's total emissions and energy (Carbon Trust 2021).

and cellular networks they rely on handle ever smaller file sizes. But the highly compressed files are more energy-intensive to decode and, consequently, magnify the end users' overall share in energy demand.[13] Simply put, streaming providers, network operators, and data centers all benefit from the increased speed and lower bandwidth demands of a smaller file, but the users have to compensate by expending more energy to compute the equations needed to play that file back.

Together with intensifying calls for environmentally aware consumption, such as the EU's Green Deal or the UN's Sustainable Development Goals, streaming users are increasingly being prompted to assume responsibility for their rising energy use. Simultaneously, the burden of sustainable action gradually seems to be moving away from streaming providers. While vaunting a largely decarbonized or carbon-offset electricity supply on their own end (e.g., Netflix 2020), streaming services can point the finger to hardware manufacturers and divert attention to the need for more energy-efficient technology (e.g., Carbon Trust 2021, 70). Ultimately, this reaffirms Julia Velkova's conclusion that "data centre operators [and streaming services] do not offset the environmental problems that the industry generates, but rather reshape the discourse around it" (2016, 8).

Such deflections may make it seem as though streaming services and hardware manufacturers were operating in different industries. In truth, they are closely interconnected, as large streaming providers invest considerable effort into the development of new compression standards, and new standards frequently necessitate new hardware. The case of AV1, a recent compression codec geared toward ultra-high resolution video, is useful to demonstrate the relationship between standards-making, infrastructure, and electronics supply chains. AV1 was created by the Alliance for Open Media (AOM), an industry consortium developing new, more "efficient" compression standards whose members include, among other tech giants, Amazon and Netflix. Netflix, YouTube, and other major video platforms began streaming videos in AV1 in 2020. Like most high-complexity codecs, AV1 is very energy-inefficient and impractical to decode with software and thus requires specialized hardware with a suitable chip. Google, another

13 There are some established and emerging strategies that counteract compression standards' growing hunger for energy, such as fast algorithms, efficient display technology, code optimizations, low-complexity enhancements to existing codecs, or coding practices that consider the energy cost of decompression already during encoding (Herglotz, Heindel, and Kaup 2019; Corrêa et al. 2018). But these measures are unlikely to offset the energy needed to power increasingly bright screens with exponentially swelling resolutions, frame rates, and bit depths, as well as the surge in the sheer number of screen devices.

member of AOM, has reportedly mandated that all new television sets with the Android operating system support the codec (Rahman 2021), further underscoring how mutually intertwined software and hardware are. The new compression standard thus not only transforms how audiovisual data is processed on a computational level. It also reinforces the consumptive cycle of material extraction, electronics production, obsolescence, and waste. Despite their public commitments to sustainability, tech companies and streaming providers thereby contribute to an ultimately unsustainable electronics supply chain (c.f. Gabrys 2011; Maxwell and Miller 2012; Cubitt 2017).

Conclusion

From an infrastructural approach, it becomes clear that inquiries into the environmental effects of video streaming fall short if terms like "Netflix" or "data center" are considered self-contained entities. Digital infrastructures are highly relational; they consist of a multitude of interacting elements. As second order systems, they are based on already existing infrastructures, and they are unruly. To comprehend the infrastructure of video streaming and its environmental impact, we therefore need to take a wide range of elements and relations at varying scales into account: cable networks, compression algorithms, telecom companies, pay television operators, browser and operating system developers, industry consortia, and national energy policies, but also more obscure actors such as chip vendors, set-top box firmware integrators, and others.

Naturally, the complexity of such an assessment demands interdisciplinary research. This chapter has indicated how media studies and science and technology studies can productively inform critical inquiries into data. As humanities scholars, we can contribute by, for example, keeping track of trends in the media industry, observing the development and standardization of new forms of compression, and calling critical attention to discourses and practices that transfer, manipulate, and redistribute environmental responsibilities.

Currently, the compression standards development process at AOM is primarily driven by cost considerations and the media companies' aversion to the complex and costly licensing structures that the MPEG standards family was encumbered with. While open-source standards are a positive development, sustainability—not cost—should be the principal value and fundamental driving force in standards-making and governance.

And despite our criticism of corporate strategies in this chapter, we want to stress that our own behavior as scholars and consumers matters and has a significant impact as well. Not only can we make sure to use electricity from renewable resources, but we can also stream less, reject "single-use films" (Veléz-Serna 2021), demand more energy-efficient standards and electronic devices, or, even better, resist the manufactured impulse to purchase ever larger and brighter screens.[14]

References

Alexander, Neta. 2017. "Rage Against the Machine: Buffering, Noise, and Perpetual Anxiety in the Age of Connected Viewing." *Cinema Journal* 56, no. 2: 1–24. https://doi.org/10.1353/cj.2017.0000.

Aslan, Joshua, Kieren Mayers, Jonathan G. Koomey, and Chris France. 2018. "Electricity Intensity of Internet Data Transmission: Untangling the Estimates." *Journal of Industrial Ecology* 22, no. 4: 785–98. https://doi.org.10.1111/jiec.12630.

Blanchette, Jean-François. 2011. "A Material History of Bits." *Journal of the American Society for Information Science and Technology* 62, no. 6: 1042–57. https://doi.org/10.1002/asi.21542.

Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2009. "Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by Jeremy Hunsinger, Lisbeth Klastrup, and Matthew Allen, 97–117. Dordrecht: Springer.

Brodie, Patrick. 2020. "Stuck in Mud in the Fields of Athenry: Apple, Territory, and Popular Politics." *Culture Machine* 19. https://culturemachine.net/vol-19-media-populism/stuck-in-mud-in-the-fields-of-athenry-patrick-brodie.

Carbon Trust. 2021. "Carbon Impact of Video Streaming." https://prod-drupal-files.storage.googleapis.com/documents/resource/public/Carbon-impact-of-video-streaming.pdf.

Carruth, Allison. 2014. "The Digital Cloud and the Micropolitics of Energy." *Public Culture* 26, no. 2: 339–64. https://doi.org/10.1215/08992363-2392093.

Corrêa, Douglas, Guilherme Correa, Daniel Palomino, and Bruno Zatt. 2018. "OTED: Encoding Optimization Technique Targeting Energy-Efficient HEVC Decoding."

14 Laura Marks et al. (2020) recently published a list of suggestions on how media scholars can mitigate their carbon footprint, and Anna Pasek (2020) has some recommendations for low-carbon academia.

In *2018 IEEE International Symposium on Circuits and Systems* (*ISCAS*), 1–5. https://doi.org/10.1109/ISCAS.2018.8351211.

Cubitt, Sean. 2017. *Finite Media: Environmental Implications of Digital Technologies*. Durham, NC: Duke University Press.

Daigle, Thomas. 2020. "'Completely Unsustainable': How Streaming and Other Data Demands Take a Toll on the Environment." *CBC News*, January 2, 2020. https://www.cbc.ca/news/science/data-centres-energy-consumption-1.5391269.

DDA (Dutch Data Center Association). 2020. "State of the Dutch Data Centers 2020." https://www.dutchdatacenters.nl/en/publications/state-of-the-dutch-data-centers-2020.

Edwards, Paul N. 2021. "Platforms Are Infrastructures on Fire." In *Your Computer Is on Fire*, edited by Thomas S. Mullaney, Benjamin Peters, Mar Hicks, and Kavita Philip, 314–36. Cambridge, MA: MIT Press.

Ejembi, Oche, and Saleem N. Bhatti. 2014. "Help Save The Planet: Please Do Adjust Your Picture." In *Proceedings of the 22nd ACM International Conference on Multimedia*, 427–36. New York: Association for Computing Machinery. https://doi.org/10.1145/2647868.2654897.

European Commission. 2020. "Commission Calls on Streaming Services, Operators and Users to Prevent Network Congestion, Discusses Issue with European Regulators." *Daily News*, March 19, 2020. https://ec.europa.eu/commission/presscorner/detail/en/mex_20_489.

European Environment Agency. 2021. "CO2 Performance of New Passenger Cars in Europe." June 1, 2021. https://www.eea.europa.eu/data-and-maps/indicators/average-co2-emissions-from-motor-vehicles-1/assessment.

Feldmann, Anja, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner et al. 2021. "A Year in Lockdown: How the Waves of COVID-19 Impact Internet Traffic." *Communications of the ACM* 64, no. 7: 101–8. https://doi.org/10.1145/3465212.

Gabrys, Jennifer. 2011. *Digital Rubbish: A Natural History of Electronics*. Ann Arbor, MI: University of Michigan Press.

Henke, Christopher R, and Benjamin Sims. 2020. *Repairing Infrastructures: The Maintenance of Materiality and Power*. Cambridge, MA: MIT Press.

Herglotz, Christian, Andréas Heindel, and André Kaup. 2019. "Decoding-Energy-Rate-Distortion Optimization for Video Coding." *IEEE Transactions on Circuits and Systems for Video Technology* 29, no. 1: 171–82. https://doi.org/10.1109/TCSVT.2017.2771819.

Hogan, Mél. 2015. "Data Flows and Water Woes: The Utah Data Center." *Big Data & Society* 2, no. 2: 1–12. https://doi-org.proxy.library.uu.nl/10.1177/2053951715592429.

Holt, Jennifer, and Kevin Sanson. 2013. *Connected Viewing: Selling, Streaming, and Sharing Media in the Digital Age*. New York: Routledge.

Johnson, Alix. 2019. "Emplacing Data within Imperial Histories: Imagining Iceland as Data Centers' 'Natural' Home." *Culture Machine* 18. https://culturemachine.net/vol-18-the-nature-of-data-centers/emplacing-data.

Kamiya, George. 2020. "The Carbon Footprint of Streaming Video: Fact-Checking the Headlines – Analysis." *IEA*, December 11, 2020. https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines.

Kaufman, Mark. 2020. "Ignore the Claim That Streaming Netflix 'Makes Climate Change Worse.'" *Mashable*, January 9, 2020. https://mashable.com/article/streaming-versus-driving-carbon-emissions.

Kittler, Friedrich. 1995. "There is No Software." *C-Theory* 32. https://journals.uvic.ca/index.php/ctheory/article/view/14655/5522.

Kränzler, Matthias, Christian Herglotz, and André Kaup. 2020. "A Comparative Analysis of the Time and Energy Demand of Versatile Video Coding and High Efficiency Video Coding Reference Decoders." In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing* (*MMSP*), 1–6. https://doi.org/10.1109/MMSP48831.2020.9287098.

Kuhr, Stefan. 2001. "Implementation of a JPEG Decoder for a 16-Bit Microcontroller." MA thesis, Stuttgart University of Applied Sciences. https://www.stefan-kuhr.de/download/thesis2side.pdf.

Larkin, Brian. 2013. "The Politics and Poetics of Infrastructure." *Annual Review of Anthropology* 42, no. 1: 327–43. https://doi.org/10.1146/annurev-anthro-092412-155522.

Lin, Chu-Hsing, Jung-Chun Liu, and Chun-Wei Liao. 2010. "Energy Analysis of Multimedia Video Decoding on Mobile Handheld Devices." *Computer Standards and Interfaces* 32, no. 1: 10–17. https://doi.org/10.1109/MUE.2007.125.

Lobato, Ramon. 2019. *Netflix Nations: The Geography of Digital Distribution.* New York: NYU Press.

Malmodin, Jens, and Dag Lundén. 2018. "The Energy and Carbon Footprint of the Global ICT and EandM Sectors 2010–15." *Sustainability* 10, no. 9: 3027. https://doi.org/10.3390/su10093027.

Marks, Laura U., Joseph Clark, Jason Livingston, Denise Oleksijczuk, and Lucas Hilderbrand. 2020. "Streaming Media's Environmental Impact." *Media+Environment* 2, no. 1. https://doi.org/10.1525/001c.17242.

Marks, Laura U., Stephen Makonin, Alejandro Rodriguez-Silva, and Radek Przedpełski. 2021. "Tackling the Carbon Footprint of Streaming Media: Final Report." https://www.sfu.ca/content/dam/sfu/sca/Streaming-Carbon-Footprint/Tackling%20the%20Carbon%20Footprint%20of%20Streaming%20Media%20report,%20bibliography%20updated%20September%202022.pdf.

Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. 2020. "Recalibrating Global Data Center Energy-Use Estimates." *Science* 367, no. 6481: 984–86. https://doi.org/10.1126/science.aba3758.

Mavlankar, Aditya, Liwei Guo, Anush Moorthy, and Anne Aaron. 2020. "Optimized Shot-Based Encodes for 4K: Now Streaming!" *Netflix Technology Blog. Medium*, August 28, 2020. https://netflixtechblog.com/optimized-shot-based-encodes-for-4k-now-streaming-47b516b10bbb.

Maxwell, Richard, and Toby Miller. 2012. *Greening the Media*. Oxford: Oxford University Press.

Mayer, Vicki. 2020. "From Peat to Google Power: Communications Infrastructures and Structures of Feeling in Groningen." *European Journal of Cultural Studies* 24, no. 4: 901–15. https://doi.org/10.1177/1367549420935898.

Monteiro, Eduarda, Mateus Grellert, Sergio Bampi, and Bruno Zatt. 2015. "Rate-Distortion and Energy Performance of HEVC and H.264/AVC Encoders: A Comparative Analysis." In *2015 IEEE International Symposium on Circuits and Systems* (*ISCAS*), 1278–81. https://doi.org/10.1109/ISCAS.2015.7168874.

Netflix. 2020. "Environmental Social Governance. 2019 Sustainability Accounting Standards Board (SASB) Report." https://s22.q4cdn.com/959853165/files/doc_downloads/2020/02/0220_Netflix_EnvironmentalSocialGovernanceReport_FINAL.pdf.

Obringer, Renee, Benjamin Rachunok, Debora Maia-Silva, Maryam Arbabzadeh, Roshanak Nateghi, and Kaveh Madani. 2021. "The Overlooked Environmental Footprint of Increasing Internet Use." *Resources, Conservation and Recycling* 167 (April): 105389. https://doi.org/10.1016/j.resconrec.2020.105389.

Parks, Lisa, and Nicole Starosielski. 2015. "Introduction." In *Signal Traffic: Critical Studies of Media Infrastructures*, edited by Lisa Parks and Nicole Starosielski, 1–27. Urbana-Champaign, IL: University of Illinois Press.

Pasek, Anne. 2019. "Managing Carbon and Data Flows: Fungible Forms of Mediation in the Cloud." *Culture Machine* 18. https://culturemachine.net/vol-18-the-nature-of-data-centers/managing-carbon.

Pasek, Anne. 2020. "Low-Carbon Research: Building a Greener and More Inclusive Academy." *Engaging Science, Technology, and Society* 6: 34–38. https://doi.org/10.17351/ests2020.363.

Rahman, Mishaal. 2021. "Google May Require New Android TV Devices Support AV1 Video Decoding." *Xda-Developers* (blog). January 16, 2021. https://www.xda-developers.com/google-requires-new-android-tv-av1-video-decoding.

Rohde, Friederike, Maike Gossen, Josephin Wagner, and Tilman Santarius. 2021. "Sustainability Challenges of Artificial Intelligence and Policy Implications." *Ökologisches Wirtschaften – Fachzeitschrift* 36 (01): 36–40. https://doi.org/10.14512/OEWO360136.

Sharrab, Yousef O, and Nabil J. Sarhan. 2013. "Aggregate Power Consumption Modeling of Live Video Streaming Systems." In *Proceedings of the 4th ACM*

Multimedia Systems Conference, 60–71. New York: Association for Computing Machinery. https://doi.org/10.1145/2483977.2483983.

Sole, Joel, Liwei Guo, Andrey Norkin, Mariana Afonso, Kyle Swanson, and Anne Aaron. 2018. "Performance Comparison of Video Coding Standards: An Adaptive Streaming Perspective." *Netflix Technology Blog. Medium*, December 15, 2018. https://medium.com/netflix-techblog/performance-comparison-of-video-coding-standards-an-adaptive-streaming-perspective-d45d0183ca95.

Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7, no. 1: 111–34. https://doi.org/10.1287/isre.7.1.111.

Starosielski, Nicole. 2014. "The Materiality of Media Heat." *International Journal of Communication* 8: 2504–8.

Starosielski, Nicole. 2015. *The Undersea Network*. Durham, NC: Duke University Press.

Sterne, Jonathan. 2012. *MP3: The Meaning of a Format*. Durham, NC: Duke University Press.

Tarnoff, Ben. 2020. "COVID-19 and the Cloud." *Progressive International*. May 21, 2020. https://progressive.international/blueprint/d0b51aca-6c19-4216-b836-1974b74ee21f-ben-tarnoff-COVID-19-and-the-cloud/en.

Uitto, Mikko. 2016. "Energy Consumption Evaluation of H.264 and HEVC Video Encoders in High-Resolution Live Streaming." In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications* (*WiMob*), 1–7. https://doi.org/10.1109/WiMOB.2016.7763234.

van Engelenburg, Jolijn, Erik van Slobbe, Adriaan J. Teuling, Remko Uijlenhoet, and Petra Hellegers. 2021. "Sustainability Characteristics of Drinking Water Supply in the Netherlands." *Drinking Water Engineering and Science* 14, no. 1: 1–43. https://doi.org/10.5194/dwes-14-1-2021.

van Kessel, Daniël. 2021a. "'Groene' restwarmte datacenters is vooral kille PR." *Follow the Money*, April 10, 2021. https://www.ftm.nl/artikelen/restwarmte-datacenters-big-tech.

van Kessel, Daniël. 2021b. "Google en Microsoft drinken onze waterbronnen leeg. Wat als er droogte komt?" *Follow the Money*, April 1, 2021. https://www.ftm.nl/artikelen/big-tech-datacenters-watervoorziening.

Veléz-Serna, María A. 2021. "Culture Beyond Extractivism: What Might a Post-Growth Cinema Look Like?" *Enough!* (blog). February 9, 2021. https://www.enough.scot/2021/02/09/culture-beyond-extractivism-what-might-a-post-growth-cinema-look-like-by-maria-antonia-velez-serna/.

Velkova, Julia. 2016. "Data That Warms: Waste Heat, Infrastructural Convergence and the Computation Traffic Commodity." *Big Data & Society* 3, no. 2: 1–10. https://doi.org/10.1177/2053951716684144.

Volmar, Axel. 2020. "Reformatting Media Studies: Toward a Theoretical Framework for Format Studies." In *Format Matters: Standards, Practices, and Politics in Media Cultures*, edited by Marek Jancovic, Axel Volmar, and Alexandra Schneider, 27–45. Lüneburg: meson press.

Vonderau, Asta. 2019. "Storing Data, Infrastructuring the Air: Thermocultures of the Cloud." *Culture Machine* 18. https://culturemachine.net/vol-18-the-nature-of-data-centers/storing-data.

About the Authors

Marek Jancovic is Assistant Professor of Media Studies at the Vrije Universiteit Amsterdam. His current research is centered around the materialities of the moving image, film preservation practices, sustainable media, and format studies.
> m.jancovic@vu.nl

Judith Keilbach is Associate Professor in the Media and Culture Studies Department at Utrecht University. Her research interests include media industries and infrastructures, the transformation of television, and sustainable media.
> j.keilbach@uu.nl

5. Out of the Bin, into the Open

Looking at the Mediating and Performing Material Afterlives of Data

Tamalone van den Eijnden

Abstract

With the notion of a material afterlife of data, this chapter addresses the issue of unregulated electronic waste recycling as an integral aspect of our digital cultures and the ways we interface with digital data. As such, this chapter not only foregrounds important issues of social and environmental justice as part of digital cultures, but it also counters the historical cultural tendency in Western thinking that privileges disembodied information over materiality. With an orientation toward the ways materiality mediates and performs, I take a new materialist approach in my analysis of the documentary *Welcome to Sodom* (2018). I conclude by proposing that the fire representations in the documentary are a productive imaginary for understanding the material entanglements of the afterlife of data.

Keywords: electronic waste, recycling infrastructures, environmental justice, imaginaries, New Materialism, Welcome to Sodom

Deleting something on an Apple computer provides a satisfactory experience. You softly drag an item to the trash, an elegantly shaped, stainless white bin with a few paper balls, and hear the sound of dry paper folding. In line with anthropologist Mary Douglas's understanding of elimination, perhaps this is "not a negative movement, but a positive effort to organize the environment" ([1966] 2003, 2) in the user's experience. The item is dropped and you immediately feel space opening up in your head…[1]

1 A similar satisfactory experience has been described by Italo Calivo in the context of taking out his household trash to the container in "La Poubelle Agréée" (1994, 93–126); this has been also described by Elizabeth Spelman in "On Taking Out the Trash" (2016, 86–96).

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH05

The aestheticized experience of deleting is not unique to an Apple computer. The bin icon, a closed-off container with a lid, suggests that junk can easily be disposed of and disappear. The idea that we can delete data by simply dragging and dropping them into virtual bins carries and produces three assumptions: 1) data are immaterial, 2) they disappear out of sight once they are dropped in the bin (and the bin is emptied), and 3) after disposal they cease to exist.[2] This principle of the digital white bin is also relevant in contexts that, at first sight, seem different from digital data.

In this chapter, I will use the principle of the digital white bin to examine electronics and their recycling infrastructures. I distinguish two ways in which one can think of the relationship between the logic of the digital white bin and electronics. First, there is a metaphoric relationship: there are aspects of the digital white bin—such as the three assumptions I identified above—that resonate with important principles of electronic waste recycling. Second, I will make the case for understanding the relation also in terms of a synecdoche, another literary device in which a part refers to the whole or the other way around.[3] Understanding the relationship between digital data and electronics as synecdochic foregrounds the material entanglements of digital data culture and electronic waste infrastructures. Such an understanding counters Western cultural assumptions that privilege data's virtuality over material realities.

In her seminal book *Digital Rubbish: A Natural History of Electronics* (2011), scholar of media, culture, and environment Jennifer Gabrys argues that data and information cannot be separated from the various ways their production and recycling produce waste, thus challenging the absolute distinction between data and electronics. She points out that, while any "discussion of information overload may seem remote from the unwieldy and extensive remainders of electronic waste in the form of abandoned computers and other discarded electronics, it is, in fact, an integral part of the processes of electronic materialization" (38). I follow Gabrys's proposition to understand electronics as part of data in general and consider the discussion on waste a productive entry for thinking about the materiality of information. As

2 These three assumptions resonate with scholarly accounts on our relation to waste more generally. For example, Valenzuela and Böhm (2017) do a Marxist and Lacanian reading of the "sustainable" rhetoric of Apple (23–60). Here, the desire of a condition without waste becomes very evident. Below, I will also engage with the accounts of Slater (1971), Graeber (2012), Morton (2013), and Doeland (2019), who also write about the undesirability of being confronted with waste.

3 An example of a synecdoche would be if a bin were taken to refer to a waste management system as a whole.

such, this chapter addresses data in terms of the socio-environmental infrastructures and ecologies of the electronic devices that allow us to interface with data.

Electronic waste, or devices that no longer allow us to interface with data in the desired and expected ways, are often associated with a vocabulary of death. For example, electronics are said to have a "product lifetime" and "product life cycle management." Electronics that are no longer available or receiving updates are dubbed an "end-of-life product." Yet, as material entities, electronics do not cease to exist after disposal, something I refer to as "the material afterlife of data." With the concept of afterlife, I want to draw attention to what happens after disposal. As I will argue, this "afterness" is not only a matter of temporal chronology but also about a relation to reality that, for many people living in Western cities, is experienced as retroactively mediated even if it is always already there on an ontological level. The life of afterlife is a way of conceptualizing the materiality of data as not simply static or dead but rather as performing a certain liveliness within organic and non-organic bodies. With this orientation toward the ways matter dynamically performs, I closely follow Jane Bennett's (2018) understanding of "vibrant matter," and I take a new materialist approach for my conceptual investigation of the material afterlife of data. Specifically, I will examine how this materiality becomes relevant on an epistemological, aesthetic, and ethical level.

The documentary *Welcome to Sodom: Your Smartphone is Already Here* (2018) by Christian Krönes, Florian Weigensamer, and Roland Schrotthofer will allow me to consider together the epistemological, aesthetic, and ethical aspects of the material afterlife of data. *Welcome to Sodom* portrays the lives of people in the city of Agbogbloshie, an electronic dumpsite near Accra in Ghana. "Sodom" is how the people that make their living in Agbogbloshie refer to the place that is a continuous open fireplace, where electronic waste is burned in an uncontained environment and without protection to retrieve valuable metals such as copper. Using this documentary, I develop alternative imageries to the bin icon. The focus on imagery makes clear that the material afterlife of data is not only an epistemological question, but it also requires a critical analysis of the aesthetics at stake, which further provoke ethical questions. As an object that mediates between me and where my electronic waste ends up (e.g., my smartphone, in keeping with the subtitle of the film), the documentary draws attention to the material infrastructural realities that are produced by geopolitical infrastructures of electronic waste recycling. It also highlights the impossibility of innocence in my own viewing position, which is situated in an urban landscape that

mediates an experience of cleanliness, thanks in part to the outsourcing of electronic waste.

The chapter is divided into three parts structured according to the principle of the white bin. First, I counter the first assumption concerning the immateriality of data on an epistemological level by reviewing theories arguing that data should be conceived of as material. Here, I show the relevance of taking what Starosielski (2019) terms an "elemental approach" to understanding how data stands in a synecdochic relationship to the object matter of electronic waste. My turn to the literary device of the synecdoche already indicates the relevance of understanding the materiality of data not only as an epistemological question, but also an aesthetic one. Secondly, I further analyze the aesthetic dimension by problematizing the second assumption that waste disappears out of sight once it is deposed in the bin. I do this by showing how infrastructures of waste recycling mediate unequal viewing positions for seeing what happens with electronic waste, which makes the impression of "out of sight" a privileged position. This, in turn, shows how the aesthetics of visibility provoke ethical questions. Lastly, the ethical dimension is further conceptualized by showing the limitations of the third assumption, which conflates disposal with an end of existence. To this end, I analyze the rhetoric and visual imagery of fire in the documentary to address the ways in which electronic waste performs lively material realities that escape total human control. The fire imagery, I argue, opens an understanding of the materiality of data that captures what is at stake in our digital and wasteful material cultures more accurately than bin-imaginaries.

Towards Synecdochic Epistemologies of Understanding Materiality

Attempts to radically materialize data offer powerful alternatives to a tendency in Western thinking that privileges the immaterial over the material. Marianne van den Boomen et al. use the term "technological mysticism" to refer to a "special brand of technological determinism" in the early 90s, "in which digitality and software are considered to be ontologically immaterial determinants of new media" (2008, 8). Similarly, while critically engaging with some of the aspirations and thought experiments of cybernetic and transhumanist thinkers, postmodern literary critic Katherine Hayles has observed that in many of their accounts, "information lost its body" (1999, 2–24). Hayles particularly criticizes the assumption in the thinking of

Hans Moravec, Marvin Minsky, Nobert Wiener, and Claude Shannon that information retains its meaning no matter which material device holds the information. We may understand how this assumption stands in a longer tradition of Western thinking with the help of feminist theorist Elizabeth Grosz (1994), according to whom the body "has remained a conceptual blind spot" in Western mainstream and feminist thinking about subjectivity, for which Descartes's dualism was foundational. Grosz takes issue with prevalent binary approaches that separate mind and body, privileging the mind over the body, the latter of which was understood as "nonhistorical, [...], passive, inert" (3). Yet, since at least the 1990s, there have been important academic contributions that theorized the body of information.

In *The Stuff of Bits: An Essay on the Materialities of Information* (2017), computer scientist Paul Dourish brings together insights from computer science and social science and draws attention to the inherent materiality of information. The focus of his book is on the ways in which representation of data is a material phenomenon (4). He argues "that the material arrangements of information," which include the ways in which data is represented and made available to users and operators, affect the ways in which the information can be experienced (4). Elements of such a materiality are the "formats that constrain, enable, limit, and shape the ways in which those representations can be created, transmitted, stored, manipulated, and put to use" (6). This complicates naïve ideas of the translatability of information regardless of the material body that Hayles famously criticized. In this way, Dourish goes beyond what Grosz identified as a "blind spot" in Western thinking. His work typifies an approach that looks at the materiality of information.

Such a materiality of data is different from what van den Boomen et al. mean by data as "in-material," by which they point to the reality that data's virtuality is only possible thanks to the materiality of the electronics that allow us to interface with data (2008, 9). A scholar who foregrounds the materiality of electronics as an important constituent of digital information is Gabrys (2011), whose work I have already cited above. A similar approach can be found in the work of Jussi Parikka's *Geology of Media* (2015). To Parikka, media not only help to understand geology and climate but also make up part of it as discarded media and technological artifacts sediment into geological layers (2015, 60). More recently, in *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (2021), Crawford considers artificial intelligence (AI) in terms of different material aspects, such as the materials from which it is produced and the labor that goes into it. The dimension of social exploitation that goes with the environmental one

resonates through all the accounts above and is also addressed by other scholars such as Christian Fuchs, who, in his critical introduction to social media, also acknowledges how the hardware that enables our digital cultures involves labor exploitation in various Asian and African countries (2014, 120).

Media, Culture, and Communication theorist Nicole Starosielski considers work that is characterized by an interest in the substances of media the "elemental turn" within media studies (2019, 1).[4] She thereby draws on the various meanings of the word element, such as the Greek and periodic elements and the ecological sense of the term, as in "being on one's element" (1). This elemental orientation often goes with an ecological concern that also offers a foundation for the politicization of the substances of media (2). An elemental approach to media studies, then, brings into focus "a network of infrastructural and ecological phenomena" (3). Recognizing the importance of the work of Gabrys (2011) and Parikka (2012), Starosielski nonetheless concludes that research on electronic waste "remains an understudied area" (3). In this chapter, I want to contribute further to this field of research by examining more specifically what such an elemental turn towards data studies could bring. To this end, I take an elemental approach that both relates to my interest in the substances of electronic waste and also brings the element of fire to the conversation.

Considering that the Greek elements of water, air, and earth permeate the ways we talk and think about data, it is quite striking that the element of fire has been largely omitted so far. Water has become a relevant metaphor for thinking about data, as is evident in expressions such as "data streams" (Hwang and Levy 2015). And in her academic work Melody Jue, (2020) proposes rethinking media studies from the perspective of the ocean. Air as an element can be traced in data vocabulary through metaphors such as "the cloud," (Hwang and Levy 2015), an aspect that also returns in the scholarly work of John Durham Peters (2015), Eva Horn (2018), and Derek McCormack (2018). The element of earth finds its way into thinking about data in terms such as "data mining" (Hwang and Levy 2015). Moreover, the critical approaches I outlined above—Gabrys (2011), Parikka (2015) and Crawford (2021)—seek to materialize data with earthly imagery, even though this imagery operates according to the principle of different literary devices.

Crawford chooses what can be understood as a metaphorical perspective. To "ground" lofty ideas on AI in material realities, she employs the "earthly" perspective of the atlas. By using the concept of an atlas to look at the

4 With the term "elemental," Starosielski draws on scholars who have already used the term to define their own writing; her intervention is to use it to describe an emerging field.

phenomenon of AI, she emphasizes that her account of AI is a "collection of disparate parts" that form a "particular viewpoint of the world" and also how AI is enmeshed with material and capitalist forces (2021, 9–11). Understanding looking at AI *like* an atlas illuminates shared principles between a material account of AI and an atlas. Yet, the atlas is not part of AI, nor is AI part of the atlas. Such a metaphoric perspective still suggests a certain distance from the actual materiality at stake. By contrast, Parikka's idea of a geology *of* media provides a synecdochic way of looking at the materiality of media.

To Parikka, geology is not only a way of looking at media; it also refers to the ways media materially sediment into geological layers, forming landscapes of what may be understood according to Parikka as "media-natures" (13). A synecdochical understanding of media as part of geology stages—more centrally than a metaphoric approach—the ways in which the materiality of information engenders geo-realities. Following this line of thinking as I analyze the material afterlife of data, I examine the afterlife of electronics as part of our digital cultures, including the infrastructures of waste management and the practices of separating materials by means of open fires. Consequently, in my analysis of *Welcome to Sodom*, I largely leave aside metaphorical allusions to the "Sodom" from Abrahamic religious texts. Taking a synecdochic approach to show data's materiality illustrates how this is not merely an epistemological question but rather an issue that requires an aesthetic analysis sensitive to how we look at the material afterlives of data.

The Aesthetics of Mediating Infrastructures

An important aspect of the aesthetics of disposal, introduced as the second assumption of the digital white bin, is the idea that what is discarded into the bin disappears out of sight. When we are welcomed to "Sodom," we are invited to look at one aspect of the materiality of data that did not disappear, namely at how electronics are recycled. Yet importantly, we do not physically go to Agbogbloshie to look, since our looking is mediated through the documentary. "Your smartphone is already here—*while you are sitting at a safe distance on your couch*" might as well have been the extended subtitle of the documentary. *Welcome to Sodom* shows how the spatial distance between me and "my" discarded phone plays out on several levels, mediating different inequalities between me and what I am looking at.

Mediation, in its broadest sense, refers to the way things are made to be seen. Such a definition would allow also us to understand urban infrastructures as mediating factors in the way a city is made to be seen. Media and performance scholars Nanna Verhoeff, Michiel de Lange, and Sigrid Merx (2019) propose understanding the city in terms of its mediality, which entails "understanding city life [...] *as media* that communicate, inform and connect." With reference to Kevin Lynch, they show how the city has been understood as a text that can be read and that "informs people about accessibility and navigability," among other things. Following this line of thinking, infrastructures of waste management and recycling mediate experiences concerning cleanliness, organization, and welfare.

How waste recycling infrastructures mediate a particular experience may be further understood with anthropologist David Graeber (2012), who analyzes some of the ideological implications of urban waste management infrastructures. Graeber observes that waste incineration, similarly to crematoria, factories, and hospitals, are mostly located outside the city in accordance with Western cultural preferences of avoiding questions of the beginning and end of life, whether they concern humans or commodities (277–78). In this logic according to Graeber, commodities "are imagined as having magically appeared, proceeding to 'circulate' [...] and then, finally, disappear into that same abyss from whence they came" (279). Following this argument, Western cities are built to mediate a reality that avoids this confrontation, which, most relevantly for this chapter, also means a peripheralization of infrastructures of electronic waste recycling.

For large parts of Western urban populations, this may then evoke the impression of an experience in which waste ceases to exist as a pressing material reality. Such obscuring infrastructures allow for a perceived reality that sociologist Philip Slater (1971) has dubbed "toilet assumption," which is "the notion that unwanted matter, unwanted difficulties, unwanted complexities and obstacles will disappear if they are removed from our immediate field of vision" (33). As philosopher Lisa Doeland (2019) has pointed out (5), the more recent iteration of this idea is philosopher Timothy Morton's (2013) "ontological u-bend" of the toilet that magically makes waste disappear "into a different dimension" (31, 115). Thus, urban infrastructures that push waste out of sight/site mediate an experience of a wasteless city.

Since waste does not actually disappear after dumping it, the wasteless city is not an experience for everybody, and thus "welcome to Sodom." In 2018, the documentary tells us that 250,000 tons of electric waste are illegally shipped to Ghana annually. This has made Agbogbloshie one of the most toxic places on earth. It is victim to what Rob Nixon (2011) has termed "slow

violence," as the dumping of toxic waste is "a violence that occurs gradually and out of sight, a violence of delayed destruction that is dispersed across time and space, an attritional violence that is typically not viewed as violence at all" (2). As Nixon also points out, this violence occurs in a geopolitical context in which many Western countries ship their waste material to countries with less strict environmental laws (Nixon 2011, 1; Gabrys 2011, 95; Tong and Wang 2012, 98–101). This dynamic, by which pollution and the burden to recycle is pushed to the territories of Indigenous Land and former colonies is, with Liboiron (2021), "an enactment of ongoing colonial relations to Land" (6). Environmental degradation, with destructive consequences for soil fertility and human and animal health, is more likely to take place in areas that historically suffered from European exploitation and where the people bear the least responsibility for "the mess." In the context of climate change, environmental studies scholar Sherilyn MacGregor (2019) already notes that the idea that we are "in this mess together" does not hold (57). The documentary makes it very visible how only some must deal with the emissions of toxins, as illegal recycling practices only take place in countries with more vulnerable geopolitical positions. In this way, infrastructures of waste recycling not only benefit from existing inequalities, but they also further entrench these structures as they exacerbate the disadvantages of impoverished areas.

Agbogbloshie should therefore not be seen as an isolated place but rather as the material result of the mediation of waste-free cities elsewhere. I can discard my phone without having to look at the waste and all its repercussions due to infrastructures that ship my phone to Ghana. Being able to not see the waste is a luxury that produces an infrastructure of waste recycling with unequal viewing positions. I am able to watch the documentary at such a safe distance that I can see the flames of the fires without feeling the biting smoke in my eyes. But there is also a distance in time, a sense of posteriority that allows me to look back at a dangerous situation after the fact. In my viewing experience, the material afterlife of data is something I can observe without being immediately physically affected by it. These privileges of looking are not merely questions of comfort, as they further produce an aesthetics that is not neutral but instead thoroughly entangled with sociocultural issues of representation.

If the geopolitical recycling infrastructures mediated unequal cityscapes, where not to see waste is a privilege, looking at the materiality of data through the mediation of a documentary privileges me again, but now as a viewer. The aesthetics of "who looks at whom" are not neutral but are instead entangled in complex power dynamics that cannot be separated

Fig. 4. A man burning electronics (*Welcome to Sodom* 2018).

from histories of colonialism and exploitation; this has been pointed out by many scholars, perhaps most famously by Edward Said ([1978] 2003). In the case of looking, as a white European, at *Welcome to Sodom* through the camera and eyes of white European documentary filmmakers, these "looking-relations" further highlight the inequality that is aesthetically mediated by the infrastructure of waste recycling with unequal viewing positions. Thus, thinking about the materiality of data is not merely an epistemological question of understanding what happens with electronic waste; it also makes me part of an aesthetic regime in which I as a viewer cannot remain innocent. Even though the materiality of data is pushed far outside the border of many European cities, it does not actually disappear out of sight. Instead, it produces infrastructures from which problematically unequal viewing positions come forward. The material implications of these unequal viewing positions raise ethical questions.

The Ethics of a Fiery and Lively Performance

The third assumption produced by the bin icon is the idea that data ceases to exist after dispoal. Shipping electronic waste outside of Europe rearticulates this assumption, as if electronics will stop "being" waste there. Yet the contrary is true, as can be seen in the documentary *Welcome to Sodom*, which gives insight into the ways that electronic waste performs lively

material realities. The performative dimension of waste becomes especially clear through the imagery of fire, which is represented not only as a tool for recycling but also as a force to be reckoned with. By analyzing the destructive force of fire, I foreground the ethical dimension of the material afterlife of data, as it raises various questions concerning justice.

The documentary is full of fire imagery, ranging from perfectly composed close-ups of burning matter to more distant fires and thick gray smoke trails in the background, which are as much part of the landscape as the omnipresent litter. In the opening scene of the documentary, we hear the voiceover of a child telling a mythical story about how a paradisiacal piece of land was set on fire by the gods as punishment for the uncaring behaviors of the people living there. A similar message is already part of the title of the documentary, which in fact quotes from the song "Welcome to Sodom" by Agbogbloshie's local rapper. "Sodom," the documentary tells us, is the way people living in Agbogbloshie refer to their place. It is a reference to the Sodom that is mentioned in religious texts of Judaism, Islam, and Christianity, a city set on fire by God as a punishment for sinful behavior. The ubiquity of the reference to the fire as punishment shows how the people understand the fire as a curse. Their livelihood depends on fire and is simultaneously endangered by it. It also shows that the cursed fire is a way for them to make sense of their work and existence in Agbogbloshie.

Besides this entanglement between fire and place, the documentary also shows us that people understand their own identity and work as enmeshed with Agbogbloshie's fires. "I am a man of fire," we hear the voice over of a young man saying, and he also relates this statement to his reluctance to jump into a fragile boat to cross the Mediterranean (00:35:00 min).[5] This "man of fire" identity comes with a certain pride in his craft. "I know about fire, I burn everything. Cables, screens, computers. I burn everything and get the copper out [...] For me, it's a good thing. It separates the metals from plastic. The fire always creates something new, fresh copper" (00:24:00–00:25:00). Fire is his identity, his skill, his recycling work. Fire is not just a tool for separating materials; fire is the *process* by which materials are separated. Thus, fire is the curse of the place, identity of the people, a skill to be mastered, and the process by which recycling is done. As such, fire is as much part of electronics (and by extension the material afterlife of data) as media are part of geology in Parikka's theory. Therefore, fire stands in a synecdochic relation with electronics.

5 All quotes from *Welcome to Sodom* are my own translations from the German version of the documentary.

Understanding fire and its potential to accelerate transformation as a synecdochical part of electronics foregrounds the performative quality of matter and opposes the idea that old electronics are dead, passive, and inert. This performative quality materializes as a threat. The voiceover, belonging to the same young man who took pride from his work with fire, also acknowledges, "They say that the fire is a monster […]. It is also said that the fire makes the heart of the people black. No one can see it, but it turns you into a dark creature, a ghost, non-human […] It gets into your body and makes you crazy, makes you sick" (0.24–25 min). Similarly recognizing the dangerous game of power and control, another young man says more generally "Sodom is a beast. Sometimes you kill the beast, and sometimes the beast kills you" (01:25:00 min).

In evoking imageries of monsters and beasts, the young men are addressing the ways in which the fire aggressively escapes human control. Fire gets into the body and changes it so profoundly that it makes the person lose their mind, changing them into "a ghost, not human." Even though humans initiate and can direct the material transformations of electronics, the fire as monster or beast is a reminder that this control is not complete and that electronic waste keeps performing in non-human and dehumanizing ways. As has been noted by Gabrys (2011) already, "[w]hile recycling appears to be a way to rid ourselves of remainder, to incorporate neatly all that is leftover, it in fact performs a deferral and inevitable return to the death of objects" (137). Similarly, Doeland (2020) writes that waste does not go "full circle, but in uncanny loops" (22). In burning electronic waste, matter performs in uncontrollable ways, allowing plastic and heavy metal to leak, sediment, and evaporate "outside the circle" of zero-waste recycling into the human body and other environments. As Jussi Parikka puts it, together with Garnet Hertz, "media do not die; media persists as electronic waste, toxic residue, and its own sort of fossil layer of disused gadgets and electronics" (Parikka 2015, 141). Contrary to the assumption that waste ceases to exist after disposal, matter thus reveals itself not as being dead but rather lively in how it molds new realities, or "vibrant," following Jane Bennett (2010).

As electronic waste materially performs in uncontrollable ways, it also narrates its own stories. With electronic waste creeping into new bodies and environments, the context, materiality, and even meaning of the electronics change. Gabrys points to this when she writes:

> In this sense, a dump is not just about waste, it is also about understanding our cultural and material metabolism. A dump registers the speed and voracity of consumption, the transience of objects and our relation with them, and the enduring materiality of those objects. (2011, 16)

While a phone in Europe "contains" my personal data, on a dumpsite in Ghana, the phone tells of a capitalist and consumerist lifestyle in certain parts of the world, one that a person doing the recycling inhales via evaporation. It becomes a "dark creature" that might kill and, as such, testifies to the slow violence of exploitative and unhealthy working conditions. Leaking into the marshlands of Agbogbloshie, it hardens and deposits into a geological layer that may be thought of as the Anthropocene.[6]

Taking the open fires of "Sodom" as synecdoche of the materiality of data offers an unsettling perspective on electronic waste. It provides a lens that highlights the troubled view in which matter is precisely not settled in a closed container, performing its own realities. It is also unsettling from an ethical perspective, as it clearly brings into relief the necessity for a justice-driven approach. Who is accountable for the repercussions of slow violence in human animal health and suffering? Who is responsible for the environmental degradation? How can we imagine a world with infrastructures of electronic waste recycling that are less damaging?

Concluding with Fire

According to Amitav Ghosh, the "climate crisis is also a crisis of culture, and thus of the imagination" (2016, 9). The crisis Ghosh is writing about results from a dissonance between the stories we tell ourselves about the world (as calculable and with the capacity for ever-increasing luxury) and the reality of climate change. As I have shown in this chapter, we also face a crisis of imagination when it comes to our digital cultures. Taking an elemental media studies approach to conceptualizing the afterlives of data, I have centered my analysis around the material afterlives of electronics as represented in the documentary *Welcome to Sodom*I have drawn attention to the crisis of how we

6 I am using the concept of "Anthropocene" here because this is the concept that Parikka uses in *A Geology of Media* (2015). It is a concept that has been popularized by meteorologist and atmospheric chemist Paul J. Crutzen along with Eugene Stoermer around the beginning of the century and sparked much research interests across disciplines. This scholarly attention also includes productive critique of the concept and proposals for alternatives that take a more political approach to the seemingly neutral and universalizing category of "anthropos." For example, Jason W. Moore (2015) proposes the Capitalocene, a concept that Haraway (2016) suggests alongside the Chthulucene. Additionally, the concept of the Plantationocene has been proposed by scholars such as Haraway, Ishikawa, Gilbert, Olwig, Tsing, and Bubandt (2016) along with many relevant critiques calling for decolonizing the Anthropocene, such as those formulated by Yusoff (2018) and Davis and Todd (2017). All these critiques are highly relevant for the specific manifestation of the "Anthropocene" in the case of Agbogbloshie.

look at the material afterlives of data by showing the unequal viewing positions on the material afterlife of data produced by electronic waste recycling infrastructures, which benefit from historical geopolitical inequalities. The open fires of "Sodom," I proposed, could serve as an alternative to the convenient bin imaginary in understanding the material implications of our digital cultures.

On an epistemological level, I have shown that understanding the materiality of data offers insight into the implications and limitations of our digital culture and thereby importantly counters a historical bias in Western thinking. On an aesthetic level, understanding the materiality of data is not only an exercise of abstract thinking, but it is also about perceiving, visualizing, and imagining. I have particularly emphasized the productivity of a synecdochic perspective on the issue through the imagery of fire. As part of the materiality of data, fire also highlights an important part of said materiality, namely the ways in which it is uncontained, lively, and threatening. As such, it offers a more accurate image for what is happening with electronic waste than the bin icon. Imagining the material afterlife of data in terms of an open fire instead of the bin icon is not yet an answer to the "crisis of imagination" that Ghosh writes about. Yet it does bring into focus a crisis that is "infrastructured" out of sight in many Western cities, namely the illegal dumping and unregulated electronic waste recycling that is visualized in *Welcome to Sodom*. As such, imagining the materiality of data in terms of the open fire exposes the ethical dimension of this materiality. As I have shown, the viciously lively performance of matter provokes questions concerning human, animal, and environmental justice. Understanding these justice-related issues as an intrinsic aspect of a particular Western digital culture is a necessary first step to reimaging and reconfiguring current infrastructures of electronic waste recycling. Finding ways to counter the crisis of imagination is not something that is done easily. Yet, this chapter is perhaps able to outline a first step—that is, visualizing the crisis along with all the problematic histories that come with such a task. The image of the open fire, I want to propose, synecdochally illuminates the material afterlife of data and therefore of digital culture more generally. It shows both "a positive effort to organize the environment," again following Douglas, and the incapacity to do so as material forces escape human control.

Acknowledgements

I could not have written this chapter without the relentless patience and support of Karin van Es, Nanna Verhoeff, and Chris van der Vegt. I am deeply grateful for all their help that went beyond the work of an editorial team.

References

Bennett, Jane. 2010. *Vibrant Matter: A Political Ecology of Things*. Durham, NC: Duke University Press.

Calivo, Italo. 1994. "La Poubelle Agréée." *The Road to San Giovanni*. New York: Vintage.

Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven, CT: Yale University Press.

Crutzen, Paul J., and Eugene F. Stoermer. 2021. "The 'Anthropocene' (2000)." In *Paul J. Crutzen and the Anthropocene: A New Epoch in Earth's History*, edited by Susanne Brenner, Gregor Lax, Paul J. Crutzen, Ulrich Pöschl, Jos Lelieveld, and Hans Gunther Brauch. Cham: Springer. https://doi.org/10.1007/978-3-030-82202-6_2.

Davis, Heather, and Zoe Todd. 2017. "On the Importance of a Date, or, Decolonizing the Anthropocene." *ACME* 16, no. 4: 761–80.

de Lange, Michiel, Sigrid Merx, and Nanna Verhoeff. 2019. "Urban Interfaces: Between Object, Concept, and Cultural Practice." Introduction to *Urban Interfaces: Media, Art & Performance in Public Spaces*, edited by Nanna Verhoeff, Sigrid Merx, and Michiel de Lange. *Leonardo Electronic Almanac* 22, no. 4. http://www.leoalmanac.org/urban-interfaces-between-object-concept-and-cultural-practice-nanna-verhoeff-sigrid-merx-michiel-de-lange.

Doeland, Lisa. 2019. "At Home in an Unhomely World: Living with Waste." *Detritus* 6: 4–10. https://doi.org/10.31025/2611-4135/2019.13820.

Doeland, Lisa. 2020. "Turning to the Specter of Waste – A Hauntological Approach." In *Perspectives on Waste from the Social Science and the Humanities: Opening the Bin,* edited by Richard Ek and Nils Johansson. Newcastle upon Tyne: Cambridge Scholars Publishing. https://www.academia.edu/39813658/Turning_to_the_Specter_of_Waste_A_Hauntological_Approach.

Douglas, Mary. 2003. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London: Routledge.

Dourish, Paul. 2017. *The Stuff of Bits: An Essay on the Materialities of Information*. Cambridge, MA: MIT Press.

Fuchs, Christian. 2014. *Social Media: A Critical Introduction*. London: SAGE Publications.

Gabrys, Jennifer. 2011. *Digital Rubbish: A Natural History of Electronics*. Ann Arbor, MI: University of Michigan Press. https://doi.org/10.1353/book.22906.

Ghosh, Amitav. 2017. *The Great Derangement: Climate Change and the Unthinkable*. Chicago: University of Chicago Press.

Graeber, David. 2012. "Afterword: The Apocalypse of Objects – Degradation, Redemption and Transcendence in the World of Consumer Goods." In *Economies of Recycling: The Global Transformation of Materials, Values and Social Relations*, edited by Catherine Alexander and Joshua Reno, 277–90. London: Zed Books.

Grosz, Elizabeth. 1994. *Volatile Bodies: Toward a Corporeal Feminism. Theories of Representation and Difference.* Bloomington, IN: Indiana University Press.

Haraway, Donna J, Noboru Ishikawa, Scott F. Gilbert, Kenneth Olwig, Anna L. Tsing, and Nils Bubandt. 2016. "Anthropologists Are Talking – About the Anthropocene." *Ethnos* 81, no. 3: 535–64. https://doi.org/10.1080/00141844.2015.1105838.

Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=488099.

Haraway, Donna J. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham, NC: Duke University Press.

Horn, Eva. 2018. "Air as Medium." *Grey Room* 73 (Fall): 6–25. https://doi.org/10.1162/grey_a_00254.

Hwang, Tim, and Karen Levy. "'The Cloud' and Other Dangerous Metaphors." *The Atlantic*, 20 Jan. 2015. https://www.theatlantic.com/technology/archive/2015/01/the-cloud-and-other-dangerous-metaphors/384518/.

Jue, Melody. 2020. *Wild Blue Media: Thinking through Seawater*. Durham, NC: Duke University Press.

Liboiron, Max. 2021. *Pollution Is Colonialism*. Durham, NC: Duke University Press.

MacGregor, Sherilyn. 2019. "Zooming In, Calling Out: (M)anthropogenic Climate Change Through the Lens of Gender." In *Climate Futures: Re-Imagining Global Climate Justice*, edited by Bhavnani Kum-Kum, John Foran, Priya A. Kurian, and Debashish Munshi, 57–63. London: Zed Books.

McCormack, Derek P. 2018. *Atmospheric Things: On the Allure of Elemental Envelopment*. Durham, NC: Duke University Press. https://doi.org/10.1215/9780822371731.

Moore, Jason W. 2015. *Capitalism in the Web of Life: Ecology and the Accumulation of Capital*. New York: Verso.

Nixon, Rob. 2011. *Slow Violence and the Environmentalism of the Poor*. Cambridge, MA: Harvard University Press.

Parikka, Jussi. 2015. *Geology of Media*. Minneapolis: University of Minnesota Press.

Peters, John Durham. 2015. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226253978.001.0001.

Said, Edward W. 2003. *Orientalism*. London: Penguin.

Slater, Philip Elliot. 1971. *The Pursuit of Loneliness: American Culture at the Breaking Point*. Boston: Beacon Press.

Spelman, Elizabeth. 2016. *Trash Talks: Revelations in the Rubbish.* Oxford: Oxford University Press.

Starosielski, Nicole. 2019. "The Elements of Media Studies." *Media+Environment* 1, no. 1: 1–6. https://doi.org/10.1525/001c.10780.

Tong, Xin, and Jici Wang. 2012. "The Shadow of the Global Network: E-Waste Flows to China." In *Economies of Recycling: The Global Transformation of Materials, Values and Social Relations*, edited by Catherine Alexander and Joshua Reno. 277–90. London: Zed Books.

Valenzuela, Francisco, and Steffen Böhm. 2017. "Against Wasted Politics: A Critique of the Circular Economy." *Ephemera: Theory and Politics in Organization* 17, no. 1: 23–60.

van den Boomen, Marianne, Sybille Lammes, Ann-Sophie Lehmann, Joost Raessens, and Mirko Tobias Schäfer. 2018. "Introduction: From the Virtual to Matters of Fact and Concern." In *Digital Material Book Subtitle: Tracing New Media in Everyday Life and Technology*, edited by Marianne van den Boomen, Sybille Lammes, Ann-Sophie Lehmann, Joost Raessens, and Mirko Tobias Schäfer. Amsterdam: Amsterdam University Press.

Weigensamer, Florian, Christian Krönes, and Roland Schrotthofer, directors and writers. 2018. *Welcome to Sodom: Dein Smartphone ist schon hier.*

Weigensamer, Florian, and Christian Krönes. 2018. "Welcome to Sodom." *Kult.kino*. https://www.kultkino.ch/film/welcometosodom/.

Yusoff, Kathryn. 2018. *A Billion Black Anthropocenes or None*. Minneapolis, MN: University of Minnesota Press.

About the Author

Tamalone van den Eijnden is a PhD candidate at the University of Twente, Department of Technology, Policy and Society. She works within the BIOTraCes Project where she studies the social-ecological phenomenon of community-led farming from a feminist science and technology studies perspective. Her background is in Media, Art and Performance Studies, with a research focus on imagining futures in the larger context of climate change.
> t.g.a.vandeneijnden@utwente.nl

Part 2

Justice

6. Data as Boundary Objects, Datafication as Boundary Work

Koen Leurs

Abstract
In this chapter, I argue the twin concepts of the "boundary object" and "boundary work" (Star 2010) enable researchers to tease out how the datafication of governance and bureaucracy results in inclusion and exclusion. The concept of boundary work enables us to ask how, for whom, for what purposes, and in what circumstances data are created, collected, categorized, used, and processed. The concept of the boundary object invites us to scrutinize the tangible records of datafication such as categories, units, numbers, and symbols by asking what forms of inclusion and exclusion they maintain or challenge. In the chapter, I explore the analytic potential of these concepts by comparing historical analogous and contemporary digitized bureaucratic governance of human mobility. The case of historical Surinamese slave registers and the contemporary Dutch passport show the urgency of questioning taken-for-granted intersectional power relations between boundaries and datafication. Future research may explore further how datafication benefits some, hurts others, materializes in particular data objects, and reflects particular situated (historical) contexts.

Keywords: datafication, boundary objects, boundary work, governance, slave registers, passports

Throughout history, governance has relied on the datafication of administrative structures and processes. Consider, for example, analogous and digital data used to control the mobility of human subjects. Think about state census records, ID cards, and biometric passports (e.g., Torpey, 2000). Focusing on examples of historical and contemporary governance of mobility, in

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH06

this chapter I will describe how we can apply the twin concepts of the "boundary object" and "boundary work" (Star 2010) to study datafication as a process of inclusion and exclusion. The concept of boundary work offers a means to study how humans or cultural, political, and natural processes are categorized into separable, bounded units. Here, I take boundary work to unravel what happens over the course of the process of abstracting the complexities of the world into representational forms of differentiation like numbers, categories, and symbols. The concept of boundary objects can be used to shed light on the concrete manifestations, outcomes, and protocols of boundary work. In the context of mobility governance, boundary objects include birth certificates, travel passes, or visas. With these twin concepts, we can grasp how datafication procedures and outcomes are inherently discriminatory (e.g., constructing hierarchical distinctions between nature and culture; between self and other; between genders, "races," nationalities, citizenship, among others).

To demonstrate the analytic potential of the two concepts, I will discuss an example of historical analogous and contemporary digitized bureaucratic governance. First, I will sketch the historical Surinamese slave registers as boundary objects. These registers reveal how non-white bodies were recorded and categorized. Contextualizing these registers offers insight into the boundary work conducted under the colonial regime, dominated by parties including powerful government institutions, wealthy corporations, and privileged societal groups. Secondly, I take the Dutch passport as a contemporary boundary object and trace the boundary work behind the passport by looking at the actors involved and decision-making processes. By having these cases speak back to feminist, postcolonial, and critical-race theories, we can draw out the common threads and recognize how data and datafication reflect and reproduce power relations and inequalities. The comparative framework is beneficial for making strange the relations between datafication and boundaries, which are commonly taken for granted. The chapter is structured as follows: Definitions and genealogies of boundary objects and boundary work are considered below, and then the two cases are presented. I end the chapter with a conclusion and reflection on the potential scope of the analytic framework presented.

Conceptual Stepping Stones: Boundary Objects, Boundary Works

The concepts of the boundary object and boundary work are heuristic lenses articulated in the field of Science and Technology Studies (STS). STS

scholars study "systems, assumptions and exclusions" (Lupton 2019, 263) as fundamentally relational processes. For this, STS offers conceptual and methodological tools to understand what political, cultural, ethical, social, and economic choices are folded into socio-technical systems. In particular, I will draw from Susan Leigh Star's work (1954–2010). Her personal and intellectual legacy combined offers alternative feminist and anti-racist analytical frameworks to understand, question, and intervene in how power emerges from the relationships between humans and technological agents. Star, who was "half-Jewish, one-quarter Scottish, and one-quarter English" (Bowker and Star 1999, 11) embodied the feminist slogan "the personal is political." She helped set the agenda to study the everyday in relation to technological concerns. In her words, she pursued "Lived experience, technologies (both everyday and those at some remove) … and silences" (Star 2007, 227). This latter point can inspire us to address who is included and can speak in relation to who is silenced and is excluded because of datafication.

However, despite the broad embrace of boundary work research across fields, as Maria do Mar Pereira argues, inequality has remained largely undertheorized and empirically understudied in boundary work research: "there is still much work to be done in integrating into our theorizing of boundary-work a central and systematic consideration of how that work is gendered, racialized, and structured by other axes of social inequality" (2019, 340). For this purpose, she argues to connect boundary work and objects with feminist, critical-race, and postcolonial theories to study how and for whom boundary work does and "does not work" (2019, 338). Before articulating these new connections, I will discuss the emergence and travel of the two concepts.

The concept of the boundary object was proposed by Star together with James R. Griesemer in a 1989 paper, which examined the collaboration between various actors in Berkeley's Museum of Vertebrate Zoology (USA). The paper analyzes how amateurs and professionals collectively aimed to preserve and thereby represent natural history. The authors observe how the various actors involved in the museum established a modus operandi based on boundary objects such as repositories of knowledge, ideal types, and standardized forms. As such, a multiplicity of actors and certain degrees of structure or dynamism in arrangements are key characteristics of boundary objects, as examples like field notes, maps, and specimens demonstrate: "Their boundary nature is reflected by the fact that they are simultaneously concrete and abstract, specific, and general, conventionalized, and customized. They are often internally heterogeneous" (Star and Griesemer 1989, 408).

Seen in this way, when we consider objects or whatever kind of material, we should not assume or attempt to find an essence or fixed "thing-ness" but rather acknowledge that objects are "something people (or, in computer science, other objects and programs) act toward and with" (Star 2010, 603). As the "stuff of action," objects are arrangements, including materials and political-socio-cultural-economic processes that "form the boundaries between groups through flexibility and shared structure" (Star 2010, 603). The boundary object is the result of "information needs" and the desire for "making an orderly array out of natural variety" held by actors (Star and Griesemer 1989, 393). We should add here that these needs cannot be seen in isolation from questions of power and control, as its dominant actors have the chance to initiate boundary work.

Boundary work, then, revolves around arranging people and non-human entities into standardized systems. This is a convoluted process, which raises questions about the role and workings of actors including gatekeepers and about who ultimately gets to define and allocate boundary objects such as categories or standards. Thus, the construction of standards reflects a certain outsourcing of morality to objects, which "is at the core of many social justice issues concerning standardization" (Star 2010, 614). The gap between what can be made to fit into pre-conceived standards, ranks, or categories of legibility and elements that cannot be made to fit presents an important entry point for grasping the politics of boundary work. For example, those subjects in efforts of administration and categorization who end up being slotted into residual categories such as "not elsewhere categorized," "none of the above," or "not otherwise specified" (Star 2010, 614) are often outsiders of the given system. Seen through the prism of boundary work, we can scrutinize how datafication and ordering the complexity of the world into neat categories demonstrate that boundary objects are never total, complete, or stable.

Across the humanities and social sciences, the concepts of boundary objects and boundary work have become widespread. These are commonly used to address how particular categorizations have become accepted or rejected as well as to pay attention to the role of various actors involved in these processes. Let us consider exemplary studies and cases relevant to researching data from the perspective of media, art, and performance studies and adjacent fields. In critical data studies, boundary objects and work have been applied to study the politics behind the categorization of data in open data platforms (Seoana and Hornidge 2020). In game studies, they have shown to be generative to reflect on the development and limitations of the field (Gekker 2020). In the field of journalism, the term boundary work is used to

study evolving arrangements of authority, knowledge production, and the roles of gatekeepers in the face of citizen journalism and digital activism: "For journalism, boundary work is a constant process, with visible consequences, in which actors, practices, texts, and institutional structures gain (or lose) status as legitimate forms" (Carlson 2019, 1). In the field of communication studies, the lens has been deployed as an interpretative framework to scrutinize ideological rhetorical work. For example, Mark Ward has studied the organizational and technical texts of the Nazi SS underpinning its genocidal project as boundary objects, which include "formatted documents, official stationery, preprinted forms, filing codes, organizational nomenclature and bureaucratic catchphrases" (2013, xv). In the field of critical sociology, the concepts have been deployed to study the impact of digitization and datafication on workers' digital labor (e.g., Zhao 2020), while in cultural geography, they are embraced to study the spatial distribution, appropriation, and situatedness of technologies (e.g., Mahony 2021).

There are various ways to connect boundary work and boundary objects to the study of inequality, exclusion, and discrimination. Feminist technoscience scholar Donna Haraway's (1991) figure of the cyborg can be taken to inverse the boundary object: the cyborg provides impetus to reconsider and blur the boundaries between object/subject, nature/culture, and human/technology and question established categories of gender, sexuality, and race. Critical race scholars point out how the physical and digital skin act as a material and semiotic surface—or boundary object—of inclusion and exclusion, where relational power relations are formed "marking exteriority and alterity, demarcating one object from another" (Thakor 2019, 198). In a similar vein, the postcolonial thinker and filmmaker Trinh Minh proposed the concept of the "boundary event" to consider the performativity of borders, which allow particular traveling subjects to pass and halt others. Finally, Sasha Costanza-Chock's "design justice" lens allows for scrutiny of the consequences of boundary work from a decarceral, decolonial perspective; this framework of analysis addresses how the design of socio-technical systems influences the distribution of benefits and burden between various groups of people (2020).

These approaches to researching boundaries share an intersectional approach. In order to consistently draw out how datafication impacts various interrelated forms of inequality, analysis of boundaries should be combined with this Black feminist theory and methodology. Intersectionality allows scholars to move from single-axis analysis (which, for example, would single out how datafication relates to racial discrimination) to the scrutiny of how race, sexuality, gender, ability, and class interact and reinforce each other.

For example, it promotes studying how datafication replicates a particular normative relation between gender, race, and class, which Patricia Hill Collins describes as the "matrix of domination" (Collins 2000).

Researching Historical and Contemporary Data Boundaries

In this section, I illustrate how we can work with boundary work and boundary objects as lenses for a critical analysis of data as media/data as performance. I first offer a case study of the setup and historical development of Surinamese slave registers to address the boundary work of Dutch colonial data collection. Secondly, I discuss important moments in the historical development of the Dutch passport to tease out how normative boundaries have been constructed. Both case studies encompass three types of data: indexical, attribute, and metadata. Indexical data refers to data that "enable identification and linking" of individuals and attribute data concerns such as "age, sex, height, weight, eye colour," and so on, while metadata are "data about data," such as column headers and definitions of data (Kitchin 2014, 8–9). I will discuss in particular how both function(ed) as boundary objects as a result of boundary work processes by asking the following five explorative questions: 1) Who was involved in the process? 2) Who was harmed in the process? 3) Who benefitted in this process? 4) How has this process materialized in concrete (data) object? 5) How are bounding processes shaped by particular socio-historical contexts?

Surinamese Slave Registers

> While resorting to a formal archive for research on slavery, the realization can hit that the archive is a violent place. People who have been treated like cattle, or objects, similar to furniture, are listed as numbers and amounts in rows of bookkeeping records. How does one begin to unpack this? (Jouwe 2021, 324).

The slave register in what is now known as the country of Suriname in South America was established by Royal Dutch Decree in 1826. "In these books all private slave owners and slave-owning plantations were registered, together with the names and additional information on all the people they owned" (van Galen and Hassankhan 2018, 504). By arguing that these nineteenth-century administrative technologies datafied enslaved people in Suriname for the purpose of identification, categorization, and rule, this

first case study offers a preliminary pre-history of analogues of administrative datafication. Here, I draw inspiration from the perspective of "deep time," an archeological approach to media history that, following Siegfried Zielinski, does not seek to find "the old in the new" but rather aims to "find something new in the old" (2006, 3). By seeing such analogue record keeping systems as pre-digital forms of datafication, we can consider historical parallels, continuities, and ruptures with the mechanisms, protocols, and techniques that undergird contemporary migration governance systems. More specifically, by addressing the slave registers from the perspective of deep time and pre-digital datafication, we can trace how the "deep pressure points" of colonialism (Stoler 2016, 5–6) and the "wake" of slavery (Sharpe 2015) built norms of anti-Blackness and white supremacy into seemingly objective and efficient administrative technologies.

The bureaucratic administration of enslaved subjects was initiated by Dutch rulers, which included private companies (Dutch West India Company, Suriname Company [Sociëteit van Suriname]) and public agencies (Fatah-Black 2013). To optimize efficiency, extraction, and profitability, the Dutch colonial rulers devised a complex administrative system, which included myriad analogue, pen-and-paper-based procedures of datafication. Boundary work alongside mapping and visualization of sugar, cotton, and coffee plantations included the design and maintenance of "tabular media" like cargo lists, insurance policies, and slave registers, which supported the "remediation of black bodies as commodities" (Wernimont 2020, 145).

Founded as a plantation colony by British settlers in 1650, Suriname was taken over by the Dutch in 1667. Under Dutch rule, the Surinamese plantation economy thrived in the eighteenth century, maximizing profit using forced labor, in particular of enslaved people from Africa. An estimated 213,000 people were brought to Suriname as part of the slave trade (van Galen and Hassankhan 2018). During the Napoleonic wars, Suriname again came under British rule, and trans-Atlantic slave trade was abolished by the British in 1808. In 1816, the colony once again became Dutch. The Atlantic slave trade ban was ratified in 1814 by the Dutch, but slavery did not end until the abolition of slavery in Suriname on July 1, 1863. The slave register was established to end elicit slave trade and smuggle. Starting in 1826, owners of enslaved people had to register their "properties" with a civil servant in Paramaribo or Nieuw Rotterdam. Afterwards, every "mutation" ("mutatie) had to be filed, including births, deaths, acquisitions, departures and gifting, trade, and sales of enslaved people, creating a near complete closed registration system. Two centuries later, forty-three books currently remain containing 15,000 folios. The early records are the least well preserved, but more than

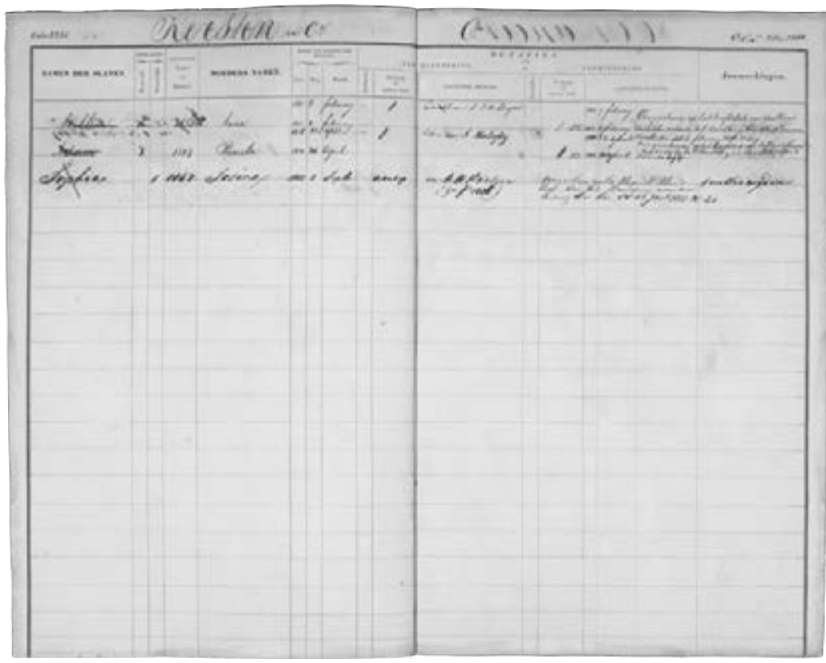


Fig. 5. Surinamese slave register folio 2320, NT00461.15 15 (Nationaal Archief).

90% of the records covering the period of 1851 to 1863 have been preserved. Digital scans of the registers can be accessed and studied through the website of the Dutch National Archive (https://www.nationaalarchief.nl), and original files are held at the National Archive in Suriname.

The slave registers act as boundary objects, as they show how boundary work is materialized through a process of datafication. This analogue, paper-and-pencil-based datafication of enslaved people will be argued to have operated based on an intersectional matrix of domination, which reflects boundary work along the lines of race, gender, sexuality, and other categories in the name of colonial, hetero, capitalist, and white supremacy. The registers are devised from the standpoint of the dominating party (white European rulers), who sought to administer their property of Black enslaved people to yield profit. The registering of enslaved people is boundary work, and the slave register is in essence a boundary object, creating and reinforcing differential standings between plantation owners and enslaved people.

The institutionalization and normalization of enslaving subjects, for the purpose of extracting profit, was partly made possible by abstracting, dehumanizing, and de-individualizing particular bodies through paper-based coding, categorization, and labeling in slave registers—all processes

of pre-digital datafication. When studying the scanned records in detail, we can see how enslaved people became administered. In figure 5, we see from the top left the folio number and the name of the owner (in this case Kersten and Co). Categories of registration include "names of slaves" ("namen der slaven"), which is already an important indication of how humans became stripped of their individuality. Only first names, given by the owners, were recorded. To ensure identification and avoid overlap, we see in the registry additions entered following first names, which might be a number, e.g., "Charles 2e" (Charles 2, born 1828, registered as male, owned by plantation Johanna Maria Coronie); a property such as "Santje groot" (large Santje, born 1808, registered as female, owned by plantation Hooijland divisie beneden Commewijne); or the abbreviation of a plantation.

Only after the abolition of slavery could freed people obtain and have recorded a last name and sometimes additional first names. Alongside names, gender ("geslacht") was to be indicated as binary male ("mannelijk") or ("vrouwelijk") as well as the year of birth, estimated or known ("geboortejaar. Gegist of bekend"). Per owner, men were registered in order of their age, followed by women and additional entries of people who joined during the registration period (children). The gendered division of labor (enslaved men were preferred for heavy plantation construction and maintenance work—which was seen as more prestigious; women were forced to pick crop—less prestigious) is discernible in the records, as more men than women died on plantations (van Galen, Quanjer, Rosenbaum-Feldbrügge, and Kraijo 2021).

In the years following initiation, details recorded in the slave registers grew. For example, the slave registers from 1848 also mention the mother for each registered person, which makes reconstructions of female family lines possible. The colonial materialization of knowing and ruling people by datafying boundaries between people based on assumed cisgender, heteronormativity, and binary genders is still dominant in the contemporary boundary work of classifying people. Furthermore, the registers have many silences and unknowns. For example, "n.o." (name unknown) is a residual category, and other residual categories such as "sent as contaminated" ("verzonden als besmet") were deployed to mark those enslaved people who had fallen ill with leprosy or parasitic worms. Their illnesses were not specified, but the label of "contaminated" marked them as an unproductive cohort for the workforce that had to be isolated, as they posed a risk to their owners and fellow enslaved people. Enslaved people were commonly registered as the property of pl./plante. (plantation). "Pé" was the marker indicating private ownership, a residual category commonly indicating

(and masking) forced domestic labor, which included the prevalent sexual exploitation by predominantly male owners. "Privé en N.ux." (private nomen uxoris) indicated a person was registered as the property of a male owner and his wife was in a residual category.

The register records are frozen moments that allow us to observe how confining social interactions between rulers and those ruled were at work in colonial-era Suriname. Slave registers as "formalizations" of boundaries (Haraway 1991, 302) between white and Black, European and non European, or ruler and enslaved people are instruments that constructed and enforced hierarchical divisions of the social world. The slave registers, from the point of view of the Dutch state and plantation owners, produced standardized boundaries, and this was for them a neutral "orderly repository" that ordered the natural world (Star and Griesemer 1989, 190). But for those Black women, children, and men not in charge over their own categorization, such administrative processes became "obligatory points of passage" (409) that stripped their individuality, context, and humanity. The slave registers were thus made to function as a rational, objective, and efficient means to legitimate the ownership and exploitation of enslaved bodies. These boundary objects thus supported "racial capitalism," which refers to the accumulation and extraction of economic value from the "racial identity" of another person (Leong 2013, 2152; Robinson 1983/2000).

Every "mutatie," such as a birth, death, registration, and deregistration, required an entry. Over time, the number of "manumissies" and "vrijlatingen" ("released enslaved people") increased, and people could be "freed" subject to government permission. The number of registered "gemanumitteerden" from 1831–1863 grew initially with (sexual) partners and children to a more diverse group of 6,781 registered "freed" subjects. Besides freeing, sales, deaths, trades, and gifting of enslaved people, the column for the "decrease" ("vermindering") of enslaved people has not been completed for many subjects. This lack of data is ambiguous and revealing, as there was a sizable number of enslaved people who managed to escape the confined space of the plantation and who went on to establish self-contained communities in the jungle (Maroons); these individuals would over time also seek to free fellow enslaved people from plantations through attacks (Fatah-Black 2013). The registers were overall a means for owners to keep inventories and oversight over their human capital, and this is also evident in the fact that after emancipation, those who could prove their ownership received 300 guilders per enslaved person. However, formerly enslaved people did not receive compensation, and many were instead forced to continue plantation work under the "state supervision" program for a decade.

With Simone Browne, we can recognize the Surinamese slave registers as boundary objects that were deployed as "simple, but violent instruments," which through boundary work "catalogued enslaved people as merchandise" as part of a broader "racializing surveillance of the slave system" (2015, 42). The colonial administration has continued effects on descendants of enslaved people, as is apparent, for example, in the recent public debate in the Netherlands on whether enslaved people who were given names by Dutch plantation owners should pay to have their assigned names officially changed in the Dutch population registry (Sneekes and Ankh Re 2021).

The Dutch Passport

The Dutch passport is not neutral but rather power ridden. As a document enabling the authentication of travelers at border sites, it reflects specific historical, political, economic, and design decisions (Torpey 2000). As "material evidence of exercising discrimination" (Keshavarz 2019, 3), the passport can be said to function as a powerful boundary object. I hypothesize that tracing the historical development of the Dutch passport offers insights into how normative intersectional configurations of gender, sexuality, race, nationality, and embodiment are enacted. Etymologically, the word "passport" is said to stem both from a document that enabled one to "pass the porte (gate) of a city" and it referred to a "pass par tout," a pass for everything (van Zoonen 2013, 83).

Borders as sites of control and containment create insiders and outsiders, and Trinh-Minh argues, "Every voyage can be said to involve a re-siting of boundaries" (2011, 27). How has mobility been encoded into the Dutch passport and for whom? What has the passport afforded and "disafforded" (Costanza-Chock 2020, 90) and for whom? Following the Henly Passport Index of visa restrictions, the contemporary Dutch passport ranks as the sixth passport in the world in terms of allowing visa-free visa on arrival access to 188 countries or territories (Henley and Partners 2021). Below, I will trace how the passport has been used to establish and normalize boundaries of race, gender, sexuality, age, and nationality through controlling the ownership of passports and abstracting people into particular categories. Colonialism, wars, commercial, and (most recently) health management incentives have accelerated the development and rollout of passports.

There is a long pre-history to the passport. As detailed in the Old Testament of the Bible, in 445 BCE, Nehemiah could assist in the rebuilding of Jerusalem because he could show letters from his king that granted safe passage. In 206 BCE, in the Chinese Han Dynasty documents, which included

identifiers such as an individual's height and age, were introduced to regulate movement in the imperial territories. The medieval Islamic Caliphate issued proof of taxes paid that allowed travelers to cross checkpoints during their travels (Mangion 2020). In several ways, these pre-modern passports, often issued on behalf of rulers or cities, resemble modern official documents granted by state bodies to national citizens. The concept of the worldwide passport, however, is a relatively recent invention. In what is now the Netherlands, a passport law was issued in 1813. From then on, distinguished men could obtain a "passe-port," an A4-size document signed in the name of the Dutch king and issued in French that facilitated mobility. This document contained data on the owners' appearance, including height, hair color, and eye color. For example, travel documents were granted to colonial officials to warrant safe passage between Europe and the West and East Indies (present-day Suriname, Indonesia, among others). Until the beginning of the twentieth century, Dutch state borders were generally open for all, and only Roma ("gipsy" or "zigeuner") travelers' mobility was actively controlled (Taylor 2014). The mobility of Roma travelers remains heavily contested until today, a reminder of how particular racialized groups of people are considered undesirable and in need of thorough scrutiny.

From a relatively small number of international colonial male travelers, the number of Dutch inhabitants owning passports grew during the First World War (1914–1915). The Netherlands sought to remain "neutral" and therefore kept its borders closed. At the initiative of the German, the Belgian–Dutch border was sealed with a two thousand volt charged barbed wire to avoid Belgian refugees escaping the war via the Netherlands. However, the Netherlands wanted to ensure transnational commercial activities for its nationals, and the passport proved an efficient boundary object to facilitate cross-border movement of eligible travelers. In the interbellum, there was a brief movement to abolish passports, but suspicion between countries resulted in the maintenance of passports.

The Second World War (1940–1945) resulted in the further institutionalization and broader adoption of identification papers. In May 1940, the Netherlands was occupied. From October 1940 onward, all Dutch inhabitants over the age of fifteen were instructed to carry an obligatory identity card (called "persoonsbewijs"). This identity card was designed by Jacobus Lambertus Lentz. This civil servant was previously in charge of the State Inspectorate of the Dutch Population Register. Right before World War II, he took the initiative to set up a system to register the identity of all Dutch inhabitants. In March 1940, the Dutch government rejected this plan, stating that it "basically considers every citizen to be a potential criminal"

(Roest et al. 2014, 155). But soon thereafter, his plans were incorporated by the German occupants. The ID included a photo, full name, date and city of birth, full address, binary gender male/female, signature, special characteristics (markers such as a missing eye), date of registration, two prints of the right index finger, stamps, and a registration number. This pass was increasingly used as a tool to expedite the identification and genocide of Jewish Dutch people. Starting in late 1940, Jewish Dutch inhabitants had to register themselves separately. Their identity cards were stamped with the letter "J" on the front and back. Beginning on May 3, 1942, Jews also had to wear a yellow star on their clothing. To protect against the creation of fake documents by the resistance fighters, receipts of all distributed ID cards were kept in a unique "national centralized population registry" housed at the Kleykamp villa in the Hague. Lentz's "house of cards" was eventually bombed by allied forces in 1944, destroying an estimated 17–40% of the register (Ketelaar 2020, 33b).

The post–World War II period is characterized by boundary work around evolving norms of gender and heteronormativity. Until 1956, a married woman was legally incompetent under Dutch law. Married women did not have an official state-issued identity, as their public identities were tied to their husbands; a passport was issued to the husband, as if a married woman's body was his property and responsibility. Until June 26, 2012, children could not obtain their own passports, and they were instead registered in their parents' passports as appendices. Until very recently, the Dutch passport also indicated the normative body is a binary-gender body. Although the non-binary gender X was technically made possible twenty-five years ago in 1996, resulting from International Civil Aviation Organization (ICAO) guidelines for machine-readable travel documents. In 2018, the Netherlands registered a (non-binary) gender X in a passport for the first time. Unlike other geographical contexts, this is not possible by self-determination but instead through a complex legal procedure, which involves suing the local government where one was born. It is important to note that non-binary gender markers in passports continue to raise suspicion at border crossings, and as such, non-normative documentation is argued to be "reinforcing Fortress Europe 2.0" (Quinan and Hunt 2021).

In recent decades, the passport is increasingly perceived as an (exclusive) marker of national belonging and allegiance. The Dutch passport is commonly invoked by Dutch politicians and policymakers in discussions about belonging and integration. These discussions are particularly heated regarding the implications of people in the Netherlands who hold dual citizenship and therefore have two passports. The Dutch passport, which recognized

refugee migrants who function well in the Dutch system can obtain upon meeting the requirements of integration procedures, is seen as the "crown on participation and integration into Dutch society" (van Zoonen 2013, 83), and as a result, politicians publicly called for former migrants to renounce their non-Dutch nationality if they have dual citizenship. But politicians argue that Dutch expats, on the other hand, should be allowed to keep their Dutch passport when migrating to another country, because "we are proud of those Dutch nationals who export our knowledge and expertise" (83).

In summer 2022, two years into the global COVID-19 health pandemic, the "Corona passport" remains a rallying point in polarizing discussions in the Netherlands and beyond. In the Netherlands, this immunity passport functions as an app that demonstrates an individual's vaccination status, proof of antibodies to COVID-19, or recent negative COVID-19 test through a QR code. In periods of government-imposed "lockdowns," it is used to authenticate holders' health statuses, thereby constructing boundaries between those abiding by state-sanctioned health regimes—who may use the app to travel internationally or enter restaurants, bars, and sport venues—and those who are not. This initiative builds on longstanding vaccination certification schemes supported by the World Health Organization, such as confirmation of vaccination against yellow fever required for entry into several countries (Osama, Rarzai, and Majeed 2021). Further research is needed to understand the broader ramifications of such new, digitally augmented passports. Through the lens of boundary objects, we can consider how these and future state-sanctioned passports are unstable, power-ridden, and paradoxical objects that play distinctive roles in processes of controlling and containing people along intersectional axes of difference.

Conclusions

When addressing datafication (analogue or digital), the notions of boundary work and boundary objects invite critical contextual and historical reflection on how people are abstracted into digital data objects. The concept of boundary work enables us to ask how, for whom, for what purposes, and in what circumstances data are created, collected, categorized, used, and processed. The concept of the boundary object allows us to scrutinize the tangible records of datafication by asking what forms of inclusion and exclusion they maintain or challenge. The twin concepts offer means to establish a much-needed situated understanding of the modalities and consequences of administrating bodies that utilize numbering and categorizing. A historical

comparative analysis of bounding through datafication allows us to make strange what has come to be considered common sense over the course of time. The apparent neutrality, objectivity, and taken-for-grantedness of administrative records can be challenged when considered in tandem with an intersectional understanding of power hierarchies as mutually constituted along axes of difference such as gender, race, sexuality, age, nationality, and health status. The cases under study—which I could only sketch in broad strokes within the constraints of this chapter —merit further in-depth scrutiny. Future research should address historical analogue and contemporary digital bounding datafication procedures from the perspective of non-state actors and stakeholders. In particular, there is an urgent need to uncover and amplify the bottom-up experiences of people subject to boundary work.

Acknowledgements

I am grateful for the patient and dedicated editorial support of Karin van Es and Nanna Verhoeff and the assistance of Chris van der Vegt. The chapter builds on ongoing work with Philipp Seuferling excavating the interrelated histories of migration and technological innovation. It also draws on input from co-teachers Gerwin van Schie and Ruud Dielen and the questions received from students who took the 2022 "Digital identity" bachelor course.

References

Bowker, Geoffrey, and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press.

Browne, Simone. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.

Carlson, Matt. 2019. "Boundary work." In *The International Encyclopedia of Journalism Studies*, edited by Tim P. Vos and Folkder Hanusch, 1–6. Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781118841570.iejs0035.

Costanza-Chock, Sacha. 2020. *Design Justice. Community-Led Practices to Build the Worlds We Need.* Cambridge, MA: MIT Press.

Fatah-Black, Karwan Jalal. 2013. "Suriname and the Atlantic World, 1650–1800." PhD diss., Leiden University. https://hdl.handle.net/1887/21912.

Gekker, Alex. 2020. "Against Game Studies." *Media and Communication* 9, no. 1: 73–83.

Haraway, Donna J. 1991. *Simians, Cyborgs and Women: The Reinvention of Nature.* New York: Routledge.

Henley and Partners. 2021. "The Henley Passport Index." https://www.henleyglobal.com/passport-index/ranking.

Hill Collins, Patricia. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Jouwe, Nancy. 2021. "Rehumanizing Acts. An Outlook on (the Meaning of) Dutch Slavery Research." In *Mix & Stir. New Outlooks on Contemporary Art from Global Perspectives*, edited by Helen Westgeest & Kitty Zijlmans, 321-27. Amsterdam: Valiz.

Keshavarz, Mahmoud. 2019. *The Design Politics of the Passport: Materiality, Immobility and Dissent.* London: Bloomsbury.

Ketelaar, Eric. 2020. *Archiving People. A Social History of Dutch Archives.* 's-Gravenhage: Stichting Archiefpublicaties.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Leong, Nancy. 2013. "Racial Capitalism." *Harvard Law Review* 126, no. 8: 2151–226.

Lupton, Deborah. 2019. "Introduction / Infrastructure." In *DigitalSTS. A Field Guide for Science and Technology Studies*, edited by Janet Vertesi and David Ribes, 263–66. Princeton, NJ: Princeton University Press.

Mahony, Martin. 2021. "Geographies of Science and Technology 1: Boundaries and Crossings." *Progress in Human Geography* 45, no. 3: 586–95. https://doi.org/10.1177/0309132520969824.

Mangion, Nicola. 2020. "The Passport Throughout History – The Evolution of a Document. The Investment Migration Knowledge Platform." *Investment Migration Insider*, June 2, 2020. https://www.imidaily.com/editors-picks/the-passport-throughout-history-the-evolution-of-a-document.

Minh-ha, Trinh. 2011. *Elsewhere, within Here. Immigration, Refugeeism and the Boundary Event.* London: Routledge.

Osama, Tasnime, Mohammed S. Razai, and Azeem Majeed. 2021. "COVID-19 Vaccine Passports: Access, Equity, and Ethics." *British Medical Journal* 373. https://doi.org/10.1136/bmj.n861.

Pereira, Maria do Mar. 2019. "Boundary-Work That Does Not Work: Social Inequalities and the Non-Performativity of Scientific Boundary-Work." *Science, Technology, & Human Values* 44, no.2: 338–65. https://doi.org/10.1177/0162243918795043.

Quinan, Christine L, and Mina Hunt. 2021. "Non-Binary Gender Markers: Mobility, Migration, and Media Reception in Europe and Beyond." *European Journal of Women's Studies* (June). https://doi.org/10.1177/13505068211024891.

Robinson, Cedric. (1983) 2000. *Black Marxism. The Making of the Black Radical Tradition*. Chapel Hill, NC: University of North Carolina Press.

Roest, Friso, Johan van Someren, Miek Wijnberg, Kees Boersma, and Pieter Wagenaar. 2014. "Policy Windows for Surveillance: The Phased Introduction of the Identification Card in the Netherlands Since the Early Twentieth Century." In *Histories of State Surveillance in Europe and Beyond*, edited by Kees Boersma et al., 150–69. London: Routledge.

Seoane, Maximiliano Vila, and Anna-Katharina Hornidge. 2020. "The Politics of Data Portals in Inter- and Transdisciplinary Research." *Journal of Information Systems and Technology Management* 17. https://doi.org/10.4301/S1807-1775202017008.

Sharpe, Christina. 2015. *In the Wake: On Blackness and Being*. Durham, NC: Duke University Press.

Sneekes, Mirelle, and Ankh Re. 2021. "Auset Ankh Re: 'Ik vertik het om te betalen voor het veranderen van een naam die mijn familie heeft moeten aannemen.'" *Volkskrant,* September 8, 2021. https://www.volkskrant.nl/nieuws-achtergrond/auset-ankh-re-ik-vertik-het-om-te-betalen-voor-het-veranderen-van-een-naam-die-mijn-familie-heeft-moeten-aannemen~bafbod53.

Star, Susan Leigh. 2007. "5 Answers." In *Philosophy of Technology: 5 Questions,* edited by Jan Kyrre Berg Olsen Friis and Evan Selinger, 223–31. Copenhagen: Automatic Press/VIP.

Star, Susan Leigh. 2010. "This Is Not a Boundary Object: Reflections on the Origin of a Concept." *Science, Technology, & Human Values* 35, no. 5: 601–17.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39." *Social Studies of Science* 19, no. 3: 387–420.

Stoler, Ann L. 2016. *Duress. Imperial Durabilities in Our Times*. Durham, NC: Duke University Press

Taylor, Becky. 2014. *Another Darkness, Another Dawn: A History of Gypsies, Roma and Travellers.* London: Reaktion Books.

Torpey, John. 2000. *The Invention of the Passport.* Cambridge, UK: Cambridge University Press.

van Galen, Cornelis W., and Maurits S. Hassankhan. 2018. "A Research Note on the Slave Registers of Suriname, 1830–1865." *The History of the Family* 23, no. 3: 503–20. https://doi.org/10.1080/1081602X.2018.1507917.

van Galen, Cornelis W., Björn Quanjer, Matthias Rosenbaum-Feldbrügge, and Matthijs Kraijo, 2021. "Endless Digging and Endless Picking: Sex Ratios and Gendered Labour in Surinamese Plantations, 1830–63." *Historical Life Course Studies* 10: 46–52. https://doi.org/10.51964/hlcs9566.

Ward, Mark. 2013. *Deadly Documents. Technical Communication, Organizational Discourse, and the Holocaust: Lessons from the Rhetorical Work of Everyday Texts*. Amityville, NY: Baywood.

Wernimont, Jacqueline. 2020. *Numbered Lives. Life and Death in Quantum Media.* Cambridge, MA: MIT Press.

Zielinski, Siegfried. 2006. *Deep Time of the Media. Toward an Archaeology of Hearing and Seeing by Technical Means.* Cambridge, MA: MIT Press.

Zhao, Xinyu. 2020. "Digital Labour in Transnational Mobility: Chinese International Students' Online Boundary Work in Daigou." *New Media & Society* 23, no. 9: 2554–74. https://doi.org/10.1177/1461444820934096.

About the Author

Koen Leurs is Associate Professor in the Graduate Gender Program, Department of Media and Culture at Utrecht University. His research addresses the interrelationships between migration and digitization/datafication. He leads the "Co-designing a fair digital asylum procedure" project ('21–'22), and he is currently writing *Digital Migration Studies* (SAGE, 2023).
> k.h.a.leurs@uu.nl

7. The Datafication of Racialization and the Pursuit of Equality

The Case of the "Barometer Culturele Diversiteit"

Gerwin van Schie

Abstract
In this chapter, I show how the investigation of racialization in datafied applications can be done through an instrumental, epistemological, and ontological approach to datafication and that the results of each approach do not necessarily match. By analyzing the attempted implementation of a tool aimed at measuring the composition of personnel in terms of migration background called *Barometer Culturele Diversiteit* (BCD) at Utrecht University in the Netherlands, I show how the tool is using ideas about race (instrumental), shaping knowledge through colonial politics (epistemological), and producing race (ontological) simultaneously. Aided by this analysis, I will advocate for an understanding of the use of race-ethnic data for affirmative purposes in terms of strategic essentialism, making epistemic imperfection regarding race warranted only in antiracist data systems working toward their own obsolescence.

Keywords: datafication, racialization, strategic essentialism, infrastructural inversion, critical data studies, postcolonial studies

In 2019, the Sociaal-Economische Raad (SER), an important socio-economic advisory board for the Dutch government, concluded that the speed at which the representation of women and cultural minorities in top positions in companies and public sector organizations in the Netherlands was not increasing fast enough (SER 2019).[1] In their report, they suggested creating

1 SER = Social and Economic Council of the Netherlands

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH07

tools for measuring and monitoring the representation of disadvantaged social groups in order to track if policy changes have positive effects (SER 2019, 41). As a result, in July 2020, *Centraal Bureau voor de Statistiek* (CBS) made available a new tool called *Barometer Culturele Diversiteit* (BCD), or the Cultural Diversity Barometer (Koolmees 2020).[2] While companies and organizations usually have enough information to track the representation of men and women reliably, this is more difficult when it comes to numbers on race, ethnicity, or migration background (*migratieachtergrond*) due to legal restrictions, privacy concerns, and ambiguities in terminology. Furthermore, there is a profound and well-reported cultural uneasiness and ignorance concerning race-ethnic issues in the Netherlands (for a collection of studies on this topic, see Essed and Hoving 2014), an attitude that critical race scholar Gloria Wekker (2016) articulates as "white innocence." Due to unspoken racial undertones in Dutch discourse on ethnicity, I follow Yanow, van der Haar and Völke (2016) in referring to Dutch governmental "ethnic" categories as "race-ethnic" categories. The pervasive idea of Dutch institutions being "color-blind" clashes with the lived reality of many employees working at Dutch universities (see Essed 1999). BCD is, in part, meant as a solution to tackle many of these issues, as it delegates the statistical and categorization process to CBS and gives the organizations that order its results the possibility to focus their effort and attention on initiatives that improve the position of minority groups in their organization.

While this process might seem straightforward, this initiative became rather controversial at Dutch universities. It was met with criticism regarding its methods and politics (see Heck 2021). Furthermore, around the same time as this public discussion, CBS decided to follow the advice of the Scientific Council for Government Policy (*Wetenschappelijke Raad voor Regeringsbeleid*, from now on WRR) to abolish the adjectives "Western" and "non-Western" from their reports and communication (Bovens et al. 2021), including initiatives like BCD. As a result of the heated discussions and the upcoming changes in the labeling policies of CBS, several Dutch universities suspended their participation in the BCD initiative. The case of BCD raises many important questions concerning race-ethnic classification in data systems and associated ethical, epistemological, cultural, and political issues. First, there are questions concerning the necessity for datafied instruments given the available alternatives. Second, the BCD brings up important questions concerning the epistemic accuracy of categorization and the politics of labeling. Third, since categorization and labeling are also

2 CBS = Statistics Netherlands

fundamental processes producing stigmatization and discrimination, we need to consider how we can distinguish between racist categorization and anti-racist categorization and labeling. Furthermore, how do we prevent categorization and labeling processes aimed at decreasing inequality from becoming a source for (or the continuation of) inequality over time? These are the questions central to this investigation and for which I propose a three-part methodological approach. My aim here is not to answer all the above-mentioned questions but rather to offer an analytic framework that allows for the investigation of the meaning-making process that underlies the use of identity characteristics in governmental data systems. By highlighting three different ways in which data can be understood through an instrumental, epistemological, and ontological approach, it becomes possible to reflect on whether race-ethnic labels and categories work in racializing and/or empowering individuals in particular situations and why. Outcomes of such an analysis, in turn, can enable political discussions about strategies and policies aimed at creating inclusive and diverse working populations in companies, organizations, and government agencies.

Three Approaches to the Datafication of Racialization

Racialization should not be confused with racism. While racism signifies an ideology that informs negative attitudes and behavior toward people deemed of a different race (Todorov 1993), racialization designates a meaning-making process. In the context of this chapter, racialization can be seen as a precursor for racism. As a process, racialization is continually producing racial formations, which I understand as geographically and historically situated collections of people, things, and practices grouped on the basis of their perceived race, ethnicity, or nationality. Such an understanding contests both the essentialist views on race as something objective, biological, and concrete and the social constructivist view on race as an "illusion" born in social relations and discourse (Omi and Winant 2015, 109). Historians Michael Omi and Howard Winant (2015) state that, while the concept of race invokes "seemingly biologically based human characteristics (so-called phenotypes), selection of these particular human features for purposes of racial signification is always and necessarily a social and historical process" (110). While I fully subscribe to the thesis that the selection of human features for racial signification is socially and historically situated, we also need to take into account the technologies that mediate these very processes in today's datafied society.

Therefore, the starting points for my analysis of BCD are the categories and labels a system uses and the way in which they relate to social, cultural, and technological factors. This method, called an "infrastructural inversion," is explained as a way of recognizing "the depths of interdependence of technical networks and standards, on the one hand, and the real work of politics and knowledge production on the other" (Bowker and Star 1999, 34). Therefore, an infrastructural inversion is a means to investigate not only the political and epistemological underpinnings of definitions and standards but also the systems and institutions in which they materialize and come into effect. By tracing categories back from the moment when we encounter them to the moment they were assigned—the *inversion*—it becomes possible to make visible "foundational though invisible patterns" (Loukissas 2019, 72). This way, I show how racialization is not made possible by a singular system or organization but rather happens throughout the Dutch governmental data infrastructure. Practically, this means that I start my investigation with the Barometer Culturele Diversiteit of CBS, after which I systematically trace its data sources and the categories and definitions used in those sources.

The Instrumental Approach

In an instrumentalist perspective on datafication, data about identity characteristics like ethnicity, race, or nationality can also quite literally become instrumental. The selection, measurement, and processing of people's characteristics is seen as a neutral endeavor, since instrumentalism generally also favors a rather empiricist understanding of knowledge production. The assumption here is that, even in matters of ethnicity or race, there is some objective truth out there in the world that can be known when the right tools are used. In this world view, knowledge about skin color, countries of origin, and religion can be recorded as facts and from that moment onwards be used to cross-reference with other demographic and statistical characteristics. This is not generally seen as a political or cultural matter but rather as an objective and neutral scientific endeavor. However, when the process described in this paragraph is compared with the aforementioned definition of racialization, the similarities are striking. Here, "the extension of racial meaning to a previously racially unclassified relationship, social practice or group" (Omi and Winant 2015, 64) and the selection of "particular human features for purposes of racial signification" (Omi and Winant 2015, 110) is done in somewhat more neutral and scientific-sounding terms. Therefore, in a context in which information

about race, ethnicity, and nationality is readily available, instrumentalism makes governmental data projects very prone to racializing features. In such contexts, race quite literally starts to function *as technology* (Chun 2009). Through the data-as-instrument approach, we investigate data systems in terms of how they instrumentalize race and for what purposes. Additionally, we should pay attention to whether the chosen approach does what it is supposed to do and how it engages with existing societal power relations.

An Epistemological Approach

The political role of technologies in processes of knowledge production has been one of the core subjects in the field of science and technology studies (STS). Scholars like Bruno Latour (2005), Sandra Harding (1991), and Donna Haraway (1991) have all pointed to the socially constructed and non-neutral nature of knowledge production. Feminist critiques of scientific objectivity can be particularly helpful in addressing the politics and implications of knowledge producing assemblages such as data systems. Feminist STS scholar Sandra Harding distinguishes two possible modes of critique in the fight against epistemic inequalities in her influential work *Whose Science? Whose Knowledge? Thinking from Women's Lives* (1991). The first approach, which she calls "feminist empiricism," sees social biases in the outcomes of research as bad scientific practice (111–18). When prejudice ends up in scientific work, this is rarely the result of structural or institutional issues but can usually be traced back to human conduct. The second approach Harding recognizes and advocates uses a different understanding of how knowledge relates to the world. Such "feminist standpoint epistemology" suggests that actors and agents involved in knowledge production practices should be attentive to power relations and should incorporate in their work who benefits from a particular perspective, and, importantly, who does not (Harding 1991, 119–37). In this view, objectivity does not so much lie in the information that is extracted from the world but rather in the way in which researchers communicate the perspectives that are used in their research. To acknowledge one's position (such as in my case, a white, European, middle-class, heterosexual, cisgender male) is seen as a form of "strong objectivity" (Harding 1991, 149). It considered much better than pretending that I somehow produce knowledge in a neutral and impartial way. As I show in the following case study, data systems rarely explicitly communicate the perspective and contexts from which they operate and would therefore be considered a form of "weak objectivity." Not communicating a perspective, however, is not equated with lacking one. On the contrary, data systems

generate knowledge from a particular vantage point, and feminist standpoint theory enables us to recognize that perspective. By taking a close look at a system's input, possibilities, visualizations, and outputs, it is possible to reconstruct a perspective as well as the normative assumptions embedded in a system. Such an evaluation not only includes what is present in a system but also what is missing. By focusing on the technologies and processes of knowledge production incorporated in data systems, it becomes possible to uncover structural and institutional discriminatory and racializing practices instead of simply pointing fingers at racist programmers and or bureaucrats.

An Ontological Approach

One limitation of thinking about data-as-knowledge is that it is largely unable to account for the ways in which data systems engage more directly with our everyday lives. In the case of automated systems, the reflexive step in which knowledge is interpreted and choices are made is often delegated to computers as well. The idea of computer logic bleeding into our lifeworld therefore invites thinking about datafied racialization in more than an epistemological way. Data systems not only report on our world; in a way, they perform it. They are not just representational but also operational (Loukissas 2019). Investigating *datafied racialization* through this lens means asking questions about how race comes into being or is performed in relation to data technologies. Race-ethnic categories like the recently abolished "allochtoon," "autochtoon," "Western," and "non-Western" function as Foucauldian "régimes of truth" that are not only enacted through governmental policy but also in scientific, political, and public discourse (Rath 1991; Prins 2000). If categories like "niet-Westers" work in an institutionalized practice, they will be accepted as real; it is their institutionalization that makes categories more than mere representations. They become actors that shape and define us as people and as well as the world we live in. Considering datafication in processes of governance in terms of ontological properties helps in situating particular racial formations and technologies as inherently connected. What we end up with, then, is a relational ontology of race in which racial formations are always mediated by specific data technologies; each technology produces its own racial ontology based on its historical and sociotechnical context *and* technological possibilities. It should be noted that this stance does not equate to an ontologization of race, i.e., considering race a reality that is merely made legible by technologies, but rather makes race into an object that does not exist in its own right; race needs actors or data systems, in the case of this chapter, that connect and produce its meanings and materiality

(see M'charek 2013). The job of a researcher is not to separate fact from fiction but rather to understand how the two are connected. When systems are no longer merely treated as knowledge-producing actors but rather as world-making agents, we can be more conscious of the fact that race-ethnic categories should only be used when we are aware of the ontologies that will be, at least temporarily, produced during the design process of a system. Such awareness can aid in assessing whether race-ethnic categories in data systems are justified and whether they aid in creating a more equal and just society. In the next paragraph, I will use the case of BCD as an example to show how to conduct such an analysis.

Case Study: The "Barometer Culturele Diversiteit"

The "Barometer Culturele Diversiteit" was created to solve legal and methodological problems that Dutch organizations interested in striving for race-ethnic inclusion and equality encounter when trying to produce knowledge about their race-ethnic make-up and attempting to set goals for the future. In the institutional context of the Netherlands, the CBS is a logical partner for producing race-ethnic statistics, as it both meets the legal requirements for the collection of demographic data and already possesses most of the necessary information. The only missing information needed to generate the BCD for a specific organization is a list of the people that it employs and potentially additional information regarding, for example, these individual's department, income bracket, and/or specific function. In the following section, I will discuss the details of the CBS work process in applying BCD, taking Utrecht University as an example. Most of the specific information about the Utrecht University case was taken from a document the organization sent to its personnel and made available on its intranet server (see Universiteit Utrecht 2021). I will begin my discussion about the instrumentality of BCD by examining the different stages in the process, from data collection to policy initiatives.

BCD as Instrument

The European General Data Protection Regulation (GDPR.EU 2029) requires data collectors to request consent from the people they collect data about. Utrecht University chose to ask all its Dutch personnel permission via an opt-out system, meaning that only the people who explicitly asked to be excluded from the dataset in a given timeframe were excluded; a non-responding

person was considered to have tacitly agreed to their data being used for the BCD (Universiteit Utrecht 2021).[3] After the deadline for the opt out process has expired, information about the remaining people is gathered after subtracting refusals. This data consists of the variables in which an organization is interested. In the case of Utrecht University, these were:

- Date of birth
- Sex
- Job profile
- Income bracket
- Type of contract (permanent or temporary)
- Department

Furthermore, the organization needs to add some identifying information that will help CBS couple the information above with their own database:

- Zip code
- House number
- House number addition (if available)

When all this data is gathered, the files are uploaded to CBS by the client or downloaded by CBS from the client via a secure internet connection (CBS 2021b). The next parts of the process happen at the secure servers of CBS.

When CBS receives data from an organization, it starts its process by first pseudonymizing all data and matching the data with information from the Dutch national registry (*Basisregistratie Personen*, from now on BRP) (CBS 2021a). This way, we are able to combine the identifying information with the migration background of the people in the organization under scrutiny. After this process of data coupling, all data is aggregated on the level of the requested variables and on the level of the organization. In short, this means that statistics are produced about the division of migration backgrounds into three categories, which I discuss in greater detail in the next paragraph on epistemology. For now, it suffices to know that the categories are named: "*Nederlands*" (Dutch), "*Westers*" (Western), and "*niet-Westers*" (non-Western).

3 The justification given by Utrecht University for the choice of an opt-out system was the consideration that BCD should be considered a "legitimate interest" except when people find their private interests outweigh that of the organization. Utrecht University explains the legitimate interest it is pursuing as the "public value of equal opportunity for (future) employees with a migration background" and that the opt-out system reflects these values best (Universiteit Utrecht 2021).

These categories are used to produce statistics about Utrecht University as a whole, its individual departments, its different income brackets, and its different job profiles. To make sure that no one can be recognized individually, CBS only provides information on subcategories that consist of at least 250 people (Universiteit Utrecht 2021). If a requested subcategory is not large enough, it will be merged with one or more other categories until it matches or exceeds 250 people. An additional privacy safeguard is that information shared for the purposes of the BCD will never be used for any other purposes, unlike some other CBS data.

After processing, all personal information is deleted and only the aggregated results are processed further in the form of statistics about the race-ethnic make-up of the total organization and the requested subpopulations. With the processed statistical information, CBS creates tables and charts that visually show the division of race-ethnic representation in the Netherlands as a whole, in different industries, and in individual organizations (see for example Figure 6). By comparing different populations, an organization can be compared to other organizations in the same field or to the Dutch workforce as a whole. And when BCD is repeated after several years, developments within an organization can be monitored.

Within the instrumental approach, the BCD tool should not be confused with policy itself, as it merely measures the current state of race-ethnic representation in a particular population. Implementing BCD does not mean that any change should be expected. To increase the prospects of marginalized groups, organizations need policy initiatives that address the power structures that caused the situation to be beneficial for certain groups, e.g., white men, and detrimental to others. If the causes for unequal representation of groups are unknown, BCD might prompt further, often qualitative, research into the apparent mismatch between specific marginalized groups and the organizational culture. This was demonstrated in one of the early BCD tests when it was used to investigate the race-ethnic make-up of the Dutch national police (CBS 2017). When looking at the differences between different ranks, it became clear that people with a "non-Western" migration background were well represented at the different police academies, but that they tended to quit working for the police in their first few years of service at a disproportionate rate compared to people with a Dutch migration background (Politie Nederland 2021). The Dutch police is currently investigating the causes for the apparent discrepancy. This situation shows how BCD is not a solution to a problem but rather a starting point. In the next section, I will discuss how BCD is a very particular way of knowing and not the only option.

Fig. 6. Interface of the BCD dashboard showing data about race-ethnic representation in the Dutch labor force in 2019 (Cultureel Bureau van de Statistiek [CBS]).[4]

Situating Knowledge in BCD

Through an epistemological approach, we can see that BCD is presenting not a general but rather a very specific view on "cultural diversity." First, in the current configuration, BCD only counts diversity in terms of migration background. Marginalizations based on sexuality, gender, (dis)ability, class, or other characteristics are not considered. This practice silences intersecting marginalizations that might have different implications than simply the sum of its parts (see Crenshaw 1990). Since BCD only takes into account migration background, the name BCD seems awkwardly chosen.

Second, the way in which migration background is conceptualized in BCD is very specific to the Netherlands. In the Netherlands, a migration

4 See https://dashboards.cbs.nl/v3/barometerculturelediversiteit.

background is determined based on the country of the place of birth of a person's parents. However, this is only one of the ways in which a person might be considered to deviate from the race-ethnic norm in an organization. Diversity in terms of accent or skin colors cannot be quantified in this way. BCD is therefore clearly limited by the informational standard that CBS has available. This does not mean that the information it produces is necessarily wrong or biased but merely that it is constrained in particular ways. More qualitative forms of knowledge production do not have similar constraints. Take, for example, the previously mentioned case of the Dutch police. It has been well-reported that, since the 1980s, the police has problems with the recruitment and retention of police officers with a migration background (see de Ruijter 1998; Broekhuizen, Raven, and Driessen 2007). The reasons for these processes, which are considered to be a combination of workplace discrimination, microaggressions, and institutional racism, have been investigated well but without reference to any datafied methods (see Broekhuizen, Raven, and Driessen 2007; Mutsaers 2014; Cankaya 2017; Gowricharn and Çankaya 2017). It seems that administrators and policy makers in this case mistakenly value datafied "solutions" over different methods of knowing, only to find similar results. Ironically, datafied methodologies can rarely point at the "why" of a certain social phenomenon and results and therefore often still require more qualitative research.

A third epistemological issue: the three migration backgrounds are officially clustered along the lines of "cultural and economic similarities with the Netherlands" (CBS 2021c) when establishing the categories "*Nederlands*" (Dutch), "*Westers*" (Western), and "*niet-Westers*" (non-Western). These categories are the continuations of the racialized categories "autochtoon," "Westerse allochtoon," and "niet-Westerse allochtoon" that were the informational standards since the early seventies. These labels split the clusters of migration backgrounds roughly between affluent Christian and racially white nations (most nations in North America and Europe, plus Australia, New Zealand, Japan, and Indonesia), and their Others (All African, South-American, Middle Eastern, and Asian nations, plus Turkey, Surinam, and the Dutch Antilles). The most obvious anomalies in this categorization scheme (such as the former Dutch colony Indonesia being considered western, but Surinam, another former Dutch colony, being considered non-Western) can only be explained by considering Dutch colonial history and both its historical and more recent processes of racialization (Yanow, van der Haar, and Völke 2016). While the labels of these categories were changed around 2017 as the result of a critical report by WRR on the division between allochtoon and autochtoon (see Bovens et al. 2016), the functionality of the categories

remained in use in CBS statistics and third-party data systems based on those statistics (see van Schie, Smit, and López Coombs 2020). BCD, too, is a system that uses historically racialized categories that are merely labeled differently. Looking at BCD as situated knowledge, we therefore can conclude that it still produces knowledge from the same perspective as before and consequently reinforces the same norm. Thinking in terms of migration background through categories such as "Westers," and "niet-Westers" has been naturalized and institutionalized to such an extent in the Netherlands that it is the standard way of envisioning any social problem that requires knowledge about race or ethnicity.

Racial Ontologies in BCD

Through an ontological approach, we can see that there are reasons perhaps warranting the use of racialized categories in systems monitoring "diversity." The clustered migration backgrounds signified by the orientalist and colonial designations of "Western" and "non-Western" distinguish quite precisely the ideological demarcations between non-stigmatized migration backgrounds and stigmatized migration backgrounds in the context of the Netherlands. While it might seem counterintuitive, these clusters of categories, ideally with new labels, are rather fitting for the task for which the BCD was selected. In situations such as these, the political goals could temporarily trump epistemic precision, a practice referred to in feminist and postcolonial discourses as "strategic essentialism" (Eide 2016). In the words of postcolonial scholar Gayatri Spivak: "You pick up the universal that will give you the power to fight against the other side, and what you are throwing away by doing that is your theoretical purity" (Spivak in Spivak and Harasym 2014, 12). Such a pragmatic use of essentialism should, however, be treated with care and, only temporarily, to prevent the reification and naturalization of categories one eventually wants to eliminate.

The ontological dimension of this strategy does not lie in the supposed reality of the essentialist categories but rather in the future one is attempting to create. Ideally, Dutch universities will eventually become the meritocracies they are already pretending to be. Meanwhile, it is important to realize that affirmative action initiatives relying on stigmatized categories do not operate in a similar manner as surveillance systems relying on the same categories; the former is correcting a societal imbalance, while the latter is exploiting it in favor of people who are already benefiting from existing inequality. I therefore argue that whether we should consider the use of

race-ethnic categories a form of racialization depends heavily on the future it is helping to realize. While BCD produces race-ethnicity in a way that follows the colonial and racist logic inherited by the categories it uses, it does so in a way that aims to paint a picture of the current situation so that affirmative action can be taken from the results. Nevertheless, it remains important to realize that the results of BCD imply that the representation of people with different migration backgrounds should eventually reflect numbers that correspond with the values of institutions that truly value inclusion and diversity; a responsible use of the results implies both change in organizational culture and the implementation of policies and initiatives to enact that change.

Conclusions

In this chapter, I have shown how the investigation of racialization in datafied applications can be done through an instrumental, epistemological, and ontological approach to datafication, and that the results of each approach do not necessarily match. By analyzing the attempted implementation of BCD at Utrecht University, we find that, with the instrumental approach, there seem to be no inconsistencies in the relation of BCD to the goals set by the Dutch government, in the choice of data, and in the phenomenon investigated. On a practical level, the consent request is the only issue that needs a more accountable method concerning the decisions made. We need a better framework to decide when consent is necessary and desirable and when the legal requirement of consent might be outweighed by the public value of equality in opportunity. The institutional availability of race-ethnic data in the Dutch context can be understood as a testament to instrumentalism being the main mode in which both technological *and* race-ethnic matters are considered. Instrumentalism regarding technology invariably leads to instrumentalism regarding race; rather than an accident, racialization becomes a feature and add-on or plug-in, which can be turned on when programmers deem the use of this information "effective." The availability of race-ethnic data makes it possible to look at societal problems as if they are an engineering problem: an engineering problem that naturally requires a technological solution. Such technological determinism, however, often causes organizations and companies to overlook non-technological solutions to societal problems, even in situations in which it is still very unclear whether apps or data systems will eventually live up to the (often very high) expectations.

With an epistemological approach, we see how BCD has a necessarily situated and partial perspective on diversity that is only focused on migration background in the way it is captured in already available data. Until 2021, the format of the available data reflected discriminatory and racialized categories that have since been abolished. The results of BCD can, therefore, only account for diversity in organizations in a limited way and without any further explanations about the reasons for a particular division in representation between different groups. Qualitative research providing more detail and explanations and that has already been performed in the recent past, in the case of some organizations, seems to be valued less or considered less objective without proper substantiation.

With the ontological approach, I demonstrated that, while the labels and categories used in BCD reflect discriminatory and racialized discourses concerning people with a migration background, the way they are employed works to address and combat the problems to which those same categories have historically contributed. This does not mean that using racialized discourse within affirmative action initiatives is always warranted but rather that strategic essentialism can be an option in achieving long-term goals at the cost of short-term epistemic imperfection. How to leverage in the long and short term, or instrumentalism, epistemology, and ontology, will be different in each datafied application and in each situation caused by institutional inequality. Only by being able to distinguish between the different forms and problematize their implications will we be able to have fruitful political discussions about how to create an equal and just datafied society. Only then can we provide the conditions for people of all migration backgrounds, genders, sexualities, colors, and other axes of difference to take part in this society.

References

Bovens, Mark, Meike Bokhorst, Roel Jennissen, and Godfried Engbersen. 2016. "Migratie en classificatie: Naar een meervoudig migratie-idioom." *WRR-Verkenningen* 34. The Hague: Wetenschappelijke Raad voor het Regeringsbeleid.

Bovens, Mark, Roel Jennissen, Meike Bokhorst, and Godfried Engbersen. 2021. "Afscheid van westers en niet-westers: Naar meervoudige indelingen van herkomstgroepen." *WRR-policy brief* 8. The Hague: WRR.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Broekhuizen, Jolijn, Judith Raven, and Franciscus Maximiliaan Hendricus Driessen. 2007. "Positie en expertise van de allochtone politiemedewerker. Op weg naar een volwaardige plaats binnen de Politie." *Politiewetenschap* 37. Utrecht: Bureau Driessen.

Cankaya, Sinan. 2017. "The Racialization of Ethnic Minority Police Officers and Researchers: On Positionality and (Auto)Ethnographic Fieldwork." *European Journal of Policing Studies* 5, no. 1: 119–32.

CBS. 2017. "Migratieachtergrond werknemers Politie Nederland 2017: Pilot Barometer Culturele Diversiteit." *CBS*. https://www.cbs.nl/-/media/_excel/2019/05/190117_eigen_data_politie_nederland.xlsx.

CBS. 2021a. "Barometer Culturele Diversiteit – ingezoomde variant." *Centraal Bureau voor de Statistiek*. https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/barometer-culturele-diversiteit-ingezoomde-variant.

CBS. 2021b. "Data uitwisselen met het CBS." *Centraal Bureau voor de Statistiek*. https://www.cbs.nl/nl-nl/deelnemers-enquetes/deelnemers-enquetes/bedrijven/meer-over-cbs-enquetes/data-uitwisselen-met-het-cbs.

CBS. 2021c. "Persoon met een niet-westerse migratieachtergrond." *Centraal Bureau voor de Statistiek*. https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/persoon-met-een-niet-westerse-migratieachtergrond.

Chun, Wendy Hui Kyong. 2009. "Introduction: Race and/as Technology; or, How to Do Things to Race." *Camera Obscura: Feminism, Culture, and Media Studies* 24, no. 1 (70): 7–35. https://doi.org/10.1215/02705346-2008-013.

Crenshaw, Kimberle. 1990. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review* 43, no. 6: 1241–300.

de Ruijter, Ad. 1998. *Pluriform in uniform: diversiteit bij de politie Amsterdam-Amstelland, een overzicht van 1987–1998*. Lelystad: Stichting IVIO.

Eide, Elisabeth. 2016. "Strategic Essentialism." In *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*, 1–2. Hoboken, NJ: John Wiley and Sons. https://doi.org/10.1002/9781118663219.wbegss554.

Essed, Philomena. 1999. "Ethnicity and Diversity in Dutch Academia." *Social Identities* 5, no. 2: 211–25. https://doi.org/10.1080/13504639951563.

Essed, Philomena, and Isabel Hoving. 2014. *Dutch Racism*. Amsterdam: Rodopi.

GDPR.EU. 2019. "What Are the GDPR Consent Requirements?" January 25, 2019. https://gdpr.eu/gdpr-consent-requirements.

Gowricharn, Ruben, and Sinan Çankaya. 2017. "Policing the Nation: Acculturation and Street-Level Bureaucrats in Professional Life." *Sociology* 51, no. 5: 1101–17. https://doi.org/10.1177/0038038515601781.

Grondwet. 2008. *Artikel 1*. http://wetten.overheid.nl/BWBR0001840/2008-07-15.

Haraway, Donna J. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3: 575–99. https://doi.org/10.2307/3178066.

Harding, Sandra. 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, NY: Cornell University Press.

Heck, Wilmer. 2021. "Ophef over diversiteitsproject dat kansengelijkheid op de arbeidsmarkt moet vergroten." *NRC*, April 14, 2021. https://www.nrc.nl/nieuws/2021/04/14/diversiteitsproject-cbs-stilgelegd-a4039885.

Koolmees, Wouter. Kamerbrief. 2020. "De barometer culturele diversiteit komt per 1 juli 2020 beschikbaar." May 14, 2020. https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2020Z08725&did=2020D18653.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Loukissas, Yanni Alexander. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

M'charek, Amade. 2013. "Beyond Fact or Fiction: On the Materiality of Race in Practice." *Cultural Anthropology* 28, no. 3: 420–42. https://doi.org/10.1111/cuan.12012.

Mutsaers, Paul. 2014. "'All of Me': Psychologizing Turkish-Dutch Police Officers in the Netherlands." *Anthropology of Work Review* 35, no. 2: 72–83. https://doi.org/10.1111/awr.12041.

Omi, Michael, and Howard Winant. 2015. *Racial Formation in the United States*. 3rd ed. New York and London: Routledge and Kegan Paul.

Politie Nederland. 2021. "Diversiteit politie neemt geleidelijk toe." July 6, 2021. https://www.politie.nl/nieuws/2021/juni/7/diversiteit-politie-neemt-geleidelijk-toe.html.

Prins, Baukje. 2000. *Voorbij de onschuld: Het debat over de multiculturele samenleving*. Amsterdam: Van Gennep.

Rath, Jan. 1991. *Minorisering: De sociale constructie van "minderheden."* Amsterdam: Sua.

SER. 2019. "Diversiteit in de top – Tijd voor versnelling." The Hague: Sociaal-Economische Raad.

Spivak, Gayatri Chakravorty, and Sarah Harasym. 2014. *The Post-Colonial Critic: Interviews, Strategies, Dialogues*. New York and London: Routledge.

Todorov, Tzvetan. 1993. "Race and Racism." In *Theories of Race and Racism: A Reader*, edited by Les Back and John Solomos, 64–70. London: Macmillan Education UK. https://doi.org/10.1007/978-1-349-22911-6_2.

Universiteit Utrecht. 2021. "Deelname Universiteit Utrecht Aan Barometer Culturele Diversiteit." April 16. https://intranet.uu.nl/kennisbank/deelname-universiteit-utrecht-aan-barometer-culturele-diversiteit.

van Schie, Gerwin, Alex Smit, and Nicolás López Coombs. 2020. "Racing through the Dutch Governmental Data Assemblage: A Postcolonial Data Studies Approach." *Global Perspectives* 1, no. 1. https://doi.org/10.1525/gp.2020.12779.

Wekker, Gloria. 2016. *White Innocence: Paradoxes of Colonialism and Race*. Durham, NC: Duke University Press.

Wet Op Het Centraal Bureau Voor de Statistiek. 2004. *Artikel 35*. http://wetten.overheid.nl/BWBR0015926/2017-01-01/0/Hoofdstuk5/Paragraaf1/Artikel35/informatie.

Yanow, Dvora, Marleen van der Haar, and Karlijn Völke. 2016. "Troubled Taxonomies and the Calculating State: Everyday Categorizing and 'Race-Ethnicity' – The Netherlands Case." *The Journal of Race, Ethnicity, and Politics* 1, no. 2: 187–226. https://doi.org/10.1017/rep.2016.7.

About the Author

Gerwin van Schie is Assistant Professor in the Media and Culture Department of Utrecht University, interested in processes of datafied racialization in the Netherlands. He dedicates his chapter, written without funding, to action group 0.7, which is fighting for better working conditions for temporary and part-time staff at Dutch universities.
> g.a.vanschie@uu.nl

8. Caged by Data

Exposing the Politics of Facial Recognition Through Zach Blas's Face Cages

Rosa Wevers

Abstract
With the emergence of facial recognition software, faces are continuously digitized and analyzed through machine vision. While facial recognition appears as an objective and unobtrusive security tool, feminist data scholars have shown that this technology is entangled with structures of power. This chapter explores how critical artistic responses to facial recognition have the potential to activate feminist critiques on the politics of facial recognition in nonverbal, material, and affective ways. Taking Zach Blas's *Face Cages* as a case study, the chapter analyzes how the art project uses strategies of defamiliarization to instigate critical reflection and activate an understanding of biometric datafication as a process of capture, which entails a violent reduction of lived experiences of identity and embodiment into biometric capta.

Keywords: facial recognition, art, Feminist Data Studies, capta, capture

In a dark room at the Sonic Acts Festival in Amsterdam, three faces appear on screens. At first sight, they seem to be photographs of faces taken against a dark background. But when I look more closely, I realize that the faces are slowly moving. I see eyes blinking, heads that totter slightly from left to right, and curly hairs in motion. Any other possible movement of the face is hindered by the metal masks that the people portrayed are wearing. The masks consist of metal bars in geometric lines that follow the shape of the face and connect the face's nodal points, defining facial features such as the distance between one's eyes. The metal bars press hard into the skin and encapsulate the face as a cage.

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH08

Fig. 7. *Face Cage 3* by Zach Blas, endurance performance with micha cárdenas, 2014. (Courtesy of the artist).

The geometric lines of the masks may remind us of a phenomenon that is gaining visibility in our visual culture today: the digital face prints that are used in facial recognition technology. When facial recognition software scans a face, the information about the face's nodal points is converted to a face print that consists of digital code. These face prints — or "data masks"— can

be compared to a database, on the basis of which someone's identity can be checked. However, rather than the distant and unobtrusive character that is dominantly attributed to these digital face masks (Introna and Wood 2004, 178), the metal masks of the art installation Face Cages (2014–2016) tell a different story. By squeezing the flesh of faces to fit within the frame of the metal cage, a violent image appears that I find much harder to look at than the digital green and blue lines of facial recognition software. In her review of the Sonic Acts exhibition in the art magazine *Metropolis M*, Lotte van Geijn describes a similar feeling of discomfort. She writes that the unpleasant performance creates an image that reminds her of torture devices and causes her to feel "anything except safe" (2017). Both I and van Geijn were moved by the Face Cages; the installation mobilized affect, which can be understood as "a social, pre-personal and pre-subjective dimension—[…] that which forces us to feel" (Quinan and Thiele 2020, 1). Through this embodied sensation, a process of reflection begins, because we are forced to make sense of this experience (Hengel 2018, 134).

This moving image of the caged faces directs the spectator's attention to the embodied and lived experiences of the datafication of faces: Which meanings about faces and identities are produced when faces are reduced to machine-readable code, and what gets lost in transformation? What relations of power are involved when faces are datafied and how does it produce processes of in- and exclusion? By evoking such questions, *Face Cages* involves the viewer in a process of critical reflection on the datafication of faces, and how this process is embedded in structures of power. Media scholars Ulises Mejias and Nick Couldry (2019) understand datafication as a set of processes in which elements of human life become quantified into digital code and in which value is generated from that data (e.g., for surveillance or economic purposes). As they state, datafication always comes with abstraction because social meaning is transformed into "streams of numbers that can be counted" (2019, 3). In the context of algorithmic facial recognition, which is used for security purposes (among others) such as international border control, this abstraction concerns the face.

In this chapter, I take up the art installation *Face Cages* as a central case study to investigate how artistic practices contribute to critical feminist debates on facial recognition technology. Using the strategy of "defamiliarization" (Stark and Crawford 2019) as an analytical lens, I analyze how Zach Blas's *Face Cages* not only mediates feminist critiques in nonverbal, material, and affective ways but also activates new ways of conceptualizing and making sense of biometric data and the datafication of faces, namely as "biometric capture" (Blas and Gaboury 2016) and "capta" (Kitchin 2014).

My inquiry draws upon methodological approaches from cultural analysis and feminist data studies. I combine visual analysis with a close reading of *Face Cages* alongside feminist data studies scholarship that examines and exposes the discriminatory logics of facial recognition algorithms.

Feminist Approaches to Data and Facial Recognition

Processes of datafication are not neutral, nor do they exist in a vacuum. Rather, data practices and visualizations are situated and implicated in intersecting structures of power (D'Ignazio and Klein 2020; Luka and Leurs 2020). Feminist approaches to data, which analyze how power operates and how it creates specific positions of in- and exclusion, have proven to be fruitful in deconstructing a deeply rooted belief in the objectivity of data practices. When it comes to facial recognition technology, such approaches to data allow us to analyze the cultural norms and prejudices that are part of the system's logic. The digital lines and numbers that appear in a biometric scan suggest that this is a neutral and objective registration of a face in which the face is presumed to be a unique and stable "anchor" of identity (Currah and Mulqueen 2011; Wevers 2018). In other words, facial recognition operates from the expectation that faces and identities are static sources of information that are "legible" to the algorithm. These systems scan faces and categorize them into identity categories; their programmers claim that they can identify "gender," "race," and "age," and those categories are defined through binary frameworks, erasing ways of being in the world that do not fit one single category or refuse those categories in their self-identification (Browne 2015; Magnet 2011; Quinan 2017). As design scholar Os Keyes argues regarding the recognition of gender in facial recognition, these systems "impos[e] their views on gender on unwitting users and research subjects" (2018, 17) and deny the role of self-identification and self-knowledge, which makes these systems structurally trans*-exclusive. Additionally, facial recognition systems disproportionality misrecognize or fail to recognize, for example, people of color, people with disabilities, and individuals who are situated at the intersection of those categories (Buolamwini and Gebru 2018; Magnet 2011; Quinan 2017). These structural failures reveal the instability of faces as "anchors of identity," which is an assumption deeply ingrained in facial recognition systems.[1]

1 This assumption is also apparent in recent studies by Stanford University that used facial recognition techniques to make claims about people's sexual orientation on the basis of their physical appearance. These studies were highly criticized by LGBTQI+ and human rights

In addition to issues concerning the structural failure of the technology, marginalized and minoritized subjects and communities are disproportionately targeted by biometric surveillance. Facial recognition systems are used to profile, police, and criminalize marginalized and minoritized groups (Blas and Gaboury 2016; Browne 2015; Magnet 2011; M'charek, Schramm, and Skinner 2014). While facial recognition systems are frequently installed under the guise of "objectivity," substituting profiling by human security guards, the decision of which groups to subject to facial recognition surveillance and subsequently whose information to save in databases for further profiling is deeply political (Wevers 2018).

Feminist approaches to datafied practices such as digital facial recognition offer important insights in the politics that surround data. However, as they operate on the level of theory, they are also complex and abstract, especially for non-expert audiences. In visual, performative, material, or sonic ways, artistic critiques to facial recognition offer a different entry point into these discussions, which offers potential for engaging non-expert audiences into these conversations.

Artistic Interventions into Datafication

An engagement with algorithms and datafication is prominent in the field of cultural critique in the arts (Alacovska, Booth, and Fieseler 2020; Stark and Crawford 2019). Facial recognition software is a popular topic of inquiry among such artistic critiques on datafication (Vries 2019). Many contemporary art projects, such as *How do you see me?* by Heather Dewey-Hagborg, expose the logics and politics of facial recognition technology by making visible the inner operations of these systems. Other works, such as Zach Blas's *Facial Weaponization Suite*, take the form of anti-surveillance projects that propose tactics of masking and camouflage to hide individuals from biometric recognition. Facial recognition is also used as an artistic tool, often as a way to create interaction with the spectator as a starting point for critical reflection. Projects including *The Biometric Mirror* by Lucy McRae and *Face to Face* by Ningli Zhu use facial recognition to make the spectator part of the artwork. Each of these artistic projects engender and embody a cultural critique of data in their own way.

In their analysis of the role of art in discussions on data ethics, critical data, and media studies, scholars Kate Crawford and Luke Stark (2019) argue

organizations, among others, that expressed their concerns on how this would impact the safety and privacy of LGBTQI+ communities.

that many artists working on and with data deploy strategies of defamiliarization to engage audiences into these debates. By introducing an experience of unfamiliarity, strangeness and discomfort, artworks can create a critical distance between the spectator and digital technologies that prompts critical reflection. Media scholar Loukissas defines critical reflection as "a process by which the interwoven social and technical dynamics of data are made visible and accessible to judgment" (2019, 162). Critical reflection thus entails critically attending to the hidden "attachments, values, absences, and biases in data" (Loukissas 2019, 162) and processes of datafication and is made possible through artworks that expose these otherwise invisible elements.

Artistic strategies of defamiliarization that enable such critical reflection include opening up black-boxed digital technologies, evoking strong emotional responses in the viewer, showing the moments when systems fail (for instance, by redesigning systems to turn them against themselves), and making normalized elements of datafication seem strange (Stark and Crawford 2019). The defamiliarizing and destabilizing potential of art seems to be especially productive when artists work with the very digital technologies that they aim to criticize. Such close engagement with digital technologies directs the spectator's attention to their destructive and structurally exclusionary elements and can function as an awareness strategy against "technological carelessness" (Alacovska, Booth, and Fieseler 2020, 31; Alacovska 2020).

Due to the limited scope of this chapter, and to do justice to the specific ways in which an artwork can mobilize critical perspectives and activate new conceptualizations of data, I now zoom in on Zach Blas's *Face Cages* as an exemplary case study. I find this installation especially significant, because it exposes the oppressive dynamics of facial recognition in material, visual, and affective ways, and because it was made by using the very technology that the art project puts into question. The project thereby allows us to understand how facial recognition is a process in which the complexities of bodies and identities are abstracted and reduced to binary code and how this process is embedded within intersecting structures of power. Before I turn to my analysis of the affective dimensions of *Face Cages* and its intervention into discussions within feminist data studies, I introduce the artwork and provide a visual analysis using semiotics (Barthes 1997) by discussing its most important visual elements.

Face Cages

Face Cages is a mixed media art installation that consists of four metal masks and accompanying videos that present "a dramatization of the abstract

violence of the biometric diagram" (Blas n.d.).[2] In these videos, we see an endurance performance in which the masks are worn by four queer-identifying artists: micha cárdenas, Elle Mehrmand, Paul Mpagi Sepuya, and Zach Blas himself. In theory, the masks should fit the performers' faces perfectly, as they were constructed from their personal biometric information that Blas measured using facial recognition software. However, once materialized into three-dimensional metal form, the personalized masks turn out to be extremely painful to wear, which is the direct result of biometrics' processes of abstraction and reduction. *Face Cages* is part of Blas's ongoing investigation of the implications of biometric technologies for non-normative and marginalized subjects, which includes criminalization, discrimination, and violence.[3] The installation reflects Blas's interdisciplinary artistic approach, which is characterized by a combination of moving image, computation, performance, theory— with a focus on feminist and queer theory— and science fiction.

The metal masks of *Face Cages* have a violent connotation that evokes associations with prison bars and cages. The project thus provides a counter-image to the digital, unobtrusive, and scientific connotations typically associated with the digital face masks used for biometric recognition. It is a suggestion that is strengthened by the green or blue light that usually appears during a biometric scan and which presumably implies "a scientific, clean moment of technological identification" (Magnet 2011, 134). The masks of *Face Cages* show a gradation of intensity: the more metal bars, the more clearly they recall the facial torture devices that were used in medieval Europe and during periods of slavery in the United States.[4] In addition to torture devices, the metal masks also remind one of nineteenth century anthropometric instruments that were used to measure human skulls with the purpose of classifying them into different categories of criminality or hierarchically organized racial groups, with whiteness as the norm. The theories of difference that were built on these anthropological measuring practices of the face have shaped the project of colonialization and functioned as a scientific justification for colonialist oppression and violence (Gould 1996; M'charek 2020; Pugliese 2005). As scholars have shown, anthropometric knowledge is still used for the development of facial recognition technologies today, despite its colonial and racist history (see for example Magnet 2011; Browne 2015).

2 At Sonic Acts Festival (2017), three out of the four masks were exhibited.

3 Theoretically, *Face Cages* is heavily informed by Shoshana Magnet's work (2011) on the gendered and racialized failure of biometric technologies.

4 Grada Kilomba's book *Plantation Memories: Episodes of Everyday Racism* (2008) offers an analysis of the history and effects of this facial torture.

Defamiliarizing Facial Recognition

The violent and severe connotations that *Face Cages* evokes reveal the intrusiveness of facial recognition systems and open up new ways of conceptualizing and making sense of the datafication of faces. In the following, I deploy defamiliarization (Stark and Crawford 2019) as an analytical lens to investigate how *Face Cages* evokes critical reflection on the datafication of faces and how this activates different conceptualizations of facial recognition and biometric data that are sensitive to their implication in power structures. In addition to this theoretical intervention, I analyze how *Face Cages*—through a defamiliarization on the level of affect—mediates theoretical feminist critiques on data in emotional and embodied ways.

In *Face Cages*, defamiliarization is at work both on a material and on a visual level. By materializing his subjects' biometric data into metal cages, the artist makes visible what usually stays obscured and black-boxed, namely the violence implicit in the abstraction that comes with datafication. The faces caged in metal grids function as a counter-image that disrupts dominant representations of biometric recognition as efficient, unobtrusive, and objective. Turned into hard metal, the biometric masks appear as cages that evoke the spectator to consider facial recognition as a form of "capture" that fixates predefined notions of identity onto the body.

Inspired by communication scholar Philip Agre (1994), Blas understands biometric capture as a process in which bodies and identities are read through predefined "grammars" that function as a framework through which the face becomes codified (Blas and Gaboury 2016). Biometric technologies can only start processes of identification and verification when someone's face is first reorganized in a template that is legible to the biometric apparatus. In other words, somatic information needs to be transformed by algorithms into "a machine-readable identifier" (van der Ploeg 2009, 86—87). What the notion of biometric capture that is made present in *Face Cages* allows us to see, is how this process goes beyond a merely visual practice of scanning the face but entails a transformation of the subject into binary data. Thus, rather than a passive registration of bodily information, capture, which has a connotation of imprisonment and conquest, points at to an active force of control (Blas and Gaboury 2016).

Through the defamiliarizing image of the materialized biometric mask, *Face Cages* invites critical reflection on how we conceive of biometric data. As human geographer Rob Kitchin (2014) has shown, "data" originates from the Latin word "dare," which means "to give." The term thereby suggests that data is a simple given, which implicitly obscures the fact that data is

always already interpreted and abstracted. As an alternative to data, Kitchin (2014) and digital humanities scholar Johanna Drucker (2011) propose using the term "capta," meaning "to take." The notion of capta makes explicit that data are always partial, situated, and interpreted rather than objective and neutral representations (Drucker 2011, 7). *Face Cages* activates this notion in the context of the datafication of faces. The project exposes how biometric capta are not "already out there" but rather need to be scanned and turned into digital code in order to be meaningful for the biometric apparatus. In this process of datafication, the face is fragmented: only the face's nodal points are deemed relevant for recognition whilst other dimensions of the face are ignored. *Face Cages*' violent aesthetic of the metal bars pressing into facial skin makes the violence implicit in biometric capta and capture visible and allows us to draw connections between current forms of biometric capture and the technologies' colonialist and racist histories.

In representing a dramatization of biometric capture, *Face Cages* creates a critical distance in the viewer toward facial recognition systems and evokes questions about power: By whom were these systems designed? For what purposes? Whom does it serve, and whom does it harm? The installation thereby invites spectators to examine critically the intersecting operations of power in facial recognition systems, which is an approach that is similar to D'Ignazio and Klein's propositions for practicing "data feminism" but molded into a different form (2020). Such an approach includes, among other strategies, "asking who questions about data science" (2020, 26), gaining insight in the ways in which data practices are intertwined with structures of power, challenging classifications, and asserting that data are neither neutral nor objective. As spectators, we see four performers, each with their own positionalities along the axes of gender, ethnicity, sexuality, and nationality, who are all vulnerable to experiencing structural exclusion via biometric technologies and/or to becoming the target of profiling because of these positionalities. By visually foregrounding these non-normative subjectivities in relation to the violent aesthetic of the metal face masks, Face Cages emphasizes how facial recognition is implicated in structures of power and produces vulnerabilities and exclusions.[5]

5 The choice to work with these performers also raises a complex ethical question, as these artists were subjected to a form of biometric violence during the performance that they were already potentially subjected to on a daily basis. When asked about this, the artist explained that he has thoroughly discussed this with the participating performers, who then agreed, because *Face Cages* was created parallel to *Facial Weaponization Suite*, which is another of Blas's works that offers the possibility to resist biometric recognition. Together, the works present a dystopic and utopic perspective on biometric capture.

Through the lens of defamiliarization, we can also analyze how Face Cages involves the spectator on an emotional level in feminist critiques on facial recognition. As a growing body of scholarly texts points out, "there are clear links between perception, embodiedness and empathy. The perception of images involves seeing and reading, but also, importantly, it involves feeling" (Shinkle 2013, 78). When art produces affect in spectators, they relate to the artworks' central issue not only on a cognitive but also on a corporeal level. In providing an embodied and emotional entry point into discussions on datafication, artworks such as *Face Cages* can communicate expert knowledge from feminist and critical data studies in nonverbal ways (Alacovska, Booth, and Fieseler 2020, 23; Duxbury 2010) and play an important role in making critical approaches to data "stick."

As I described above, when I encountered the work, I was immediately immersed in it, while also experiencing it as extremely uncomfortable to watch. The videos of caged faces confronted me with four different experiences of biometric capture and the violence that is involved in condensing "complex relationships and situated knowledges into a single digital map of the body" (Magnet 2011, 29). As the performers wearing the masks are visibly in pain, the endurance performance disturbs the unobtrusive and distant character of facial recognition. Through this form of defamiliarization, *Face Cages* makes tangible "how artificial intelligence and data systems are embodied—not abstract—elements of everyday existence" (Stark and Crawford 2019, 446). In *Face Cages*, feminist critiques on facial recognition are thus taken outside the realm of abstract theory and concretized in a video performance that makes these critiques visible and sensible while also engendering new critical questions itself.

Conclusion

In the video installation *Face Cages*, the dominant image of the "unobtrusive" digital biometric scan is defamiliarized and disturbed through a new image of metal pressing into skin. This moving image makes tangible how facial recognition captures people into normative grammars of identity and how it produces quantifiable biometric capta that erases performances and experiences of identity that do not fit those norms. The artwork was created by using technology to form a critique on this technology, which is an artistic strategy that appears to be effective in drawing attention to the destructive elements of digital technologies.

Face Cages forms an exemplary case that shows how art can make debates on the politics of facial recognition visible and tangible. Such interventions are especially relevant for feminist approaches to data when they draw attention to the entanglement of data and power. As I have sought to show in this chapter, by creating a critical distance in the viewer through strategies of defamiliarization, critical art can contribute to theoretical conceptualizations of data and datafication. Moreover, by communicating expert knowledge in visual and affective ways (Alacovska, Booth, and Fieseler 2020; Duxbury 2010), art projects such as *Face Cages* have the potential to involve non-expert audiences in critical data studies debates.

Acknowledgements

The author has received funding for the research leading to this chapter from the Dutch Scientific Research Council (NWO) for the project "Opening Up the Black Box: Artistic Explorations of Technological Surveillance." The author would like to thank Zach Blas for granting permission to reprint images of his work.

References

Agre, Philip E. 1994. "Surveillance and Capture: Two Models of Privacy." *The Information Society* 10: 101–27. https://doi.org/10.1080/01972243.1994.9960162.

Alacovska, Ana, Peter Booth, and Christian Fieseler. 2020. "The Role of the Arts in the Digital Transformation." *Artsformation Report Series*. https://doi.org/10.2139/ssrn.3715612.

Barthes, Roland. 1997. *Image, Music, Text*. London: Fontana Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity Press.

Blas, Zach. n.d. *Face Cages*. https://zachblas.info/works/face-cages.

Blas, Zach, and Jacob Gaboury. 2016. "Biometrics and Opacity: A Conversation." *Camera Obscura: Feminism, Culture, and Media Studies* 31, no. 2 (92): 155–65. https://doi.org/10.1215/02705346-3592510.

Browne, Simone. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15. http://proceedings.mlr.press/v81/buolamwini18a.html.

Currah, Paisley, and Tara Mulqueen. 2011. "Securitizing Gender: Identity, Biometrics, and Transgender Bodies at the Airport." *Social Research* 78, no. 2: 557–82.

Denzin, Norman K. 2003. *Performance Ethnography: Critical Pedagogy and the Politics of Culture*. London: SAGE Publications.

de Vries, Patricia. 2019. *Algorithmic Anxiety in Contemporary Art: A Kierkegaardian Inquiry into the Imaginary of Possibility*. Amsterdam: Institute of Network Cultures.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT press.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5, no. 1: 1–21.

Duxbury, Lesley. 2010. "A Change in the Climate: New Interpretations and Perceptions of Climate Change through Artistic Interventions and Representations." *Weather, Climate, and Society* 2, no. 4: 294–99. https://doi.org/10.1175/2010WCAS1053.1.

Gould, Stephen Jay. 1996. *The Mismeasure of Man*. Revised ed. New York: Norton.

Introna, Lucas D., and David Wood. 2004. "Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems." *Surveillance & Society* 2: 177–98.

Keyes, Os. 2018. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." *Proceedings of the ACM on Human–Computer Interaction* 2 (CSCW). https://doi.org/10.1145/3274357.

Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1, no. 1: 1–12.

Luka, Mary Elizabeth, and Koen Leurs. 2020. "Feminist Data Studies." In *The International Encyclopedia of Gender, Media, and Communication*, 1–14. https://doi.org/10.1002/9781119429128.iegmc062.

Magnet, Soshana Amielle. 2011. *When Biometrics Fail. Gender, Race, and the Technology of Identity*. Durham, NC: Duke University Press.

M'charek, Amade. 2020. "Tentacular Faces: Race and the Return of the Phenotype in Forensic Identification." *American Anthropologist* 122, no. 2: 369–80. https://doi.org/10.1111/aman.13385.

M'charek, Amade, Katharina Schramm, and David Skinner. 2014. "Topologies of Race: Doing Territory, Population and Identity in Europe." *Science, Technology, & Human Values* 39, 4: 468–87. https://doi.org/10.1177/0162243913509493.

Mejias, Ulises A., and Nick Couldry. 2019. "Datafication." *Internet Policy Review* 8, no. 4. https://doi.org/10.14763/2019.4.1428.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Pugliese, Joseph. 2005. "In Silico Race and the Heteronomy of Biometric Proxies: Biometrics in the Context of Civilian Life, Border Security and Counter-Terrorism Laws." *Australian Feminist Law Journal* 23, no. 1: 1–32. https://doi.org/10.1080/13200968.2005.10854342.

Quinan, Christine. 2017. "Gender (in) Securities: Surveillance and Transgender Bodies in a Post-9/11 Era of Neoliberalism." In *Security/Mobility. Politics of Movement*, edited by Matthias Leese and Stef Wittendorp, 153–69. Manchester: Manchester University Press.

Quinan, Christine, and Kathrin Thiele. 2020. "Biopolitics, Necropolitics, Cosmopolitics – Feminist and Queer Interventions: An Introduction." *Journal of Gender Studies* 29, no. 1: 1–8. https://doi.org/10.1080/09589236.2020.1693173.

Shinkle, Eugénie. 2013. "Uneasy Bodies: Affect, Embodied Perception, and Contemporary Fashion Photography." In *Carnal Aesthetics. Transgressive Imagery and Feminist Politics*, edited by Bettina Papenburg and Marta Zarzycka, 73–88. London and New York: I.B. Tauris.

Stark, Luke, and Kate Crawford. 2019. "The Work of Art in the Age of Artificial Intelligence: What Artists Can Teach Us about the Ethics of Data Practice." *Surveillance & Society* 17, no. 3–4: 442–55. https://doi.org/10.24908/ss.v17i3/4.10821.

van den Hengel, Louis. 2018. "The Arena of Affect: Marina Abramović and the Politics of Emotion." In *Doing Gender in Media, Art & Culture: A Comprehensive Guide to Gender Studies*, edited by Rosemarie Buikema, Liedeke Plate, and Kathrin Thiele, 2nd ed., 123–35. New York: Routledge/Taylor and Francis Group.

van Geijn, Lotte. 2017. "The Noise of Being, Sonic Acts Festival 2017, de Tentoonstellingen." *Metropolis M*, February 6, 2017. https://www.metropolism.com/nl/reviews/30478_the_noise_of_being.

Wevers, Rosa. 2018. "Unmasking Biometrics' Biases: Facing Gender, Race, Class and Ability in Biometric Data Collection." *Tijdschrift voor Mediageschiedenis* 21, no. 2: 89–105. https://doi.org/http://doi.org/10.18146/2213-7653.2018.368.

About the Author

Rosa Wevers works as a PhD candidate and teacher in Gender Studies at Utrecht University. In her NWO-funded PhD research, she analyzes how contemporary art exhibitions confront visitors with critical perspectives on surveillance and engage them in strategies of resistance. In 2021, she curated the exhibition "Face Value" in collaboration with IMPAKT and the NFF.
> r.l.l.wevers@uu.nl

9. Dirty Computers versus the New Jim Code

Janelle Monáe's Datafied Performance

Dan Hassler-Forest

Abstract
While digital data networks provide new opportunities for anti-racist creative production and activism, they also constitute a powerful surveillance network that reproduces and even exacerbates racist social structures. This article focuses on the creative work of musician, performing artist, and activist Janelle Monáe, whose creative work across digital media platforms has developed Afrofuturist storyworlds that reflect this dialectic. By using androids and "dirty computers" as signifiers for processes of racialized, gendered, and sexual exclusion throughout her musical career, her work brings into sharper focus how digital data networks constitute what Ruha Benjamin has described as the "New Jim Code." At the same time, her fully datafied performance in VR space as a transmedial extension of the television series *Lovecraft Country* shows how these same data systems can be used to creatively resist and potentially transform our understanding of these ubiquitous networks.

Keywords: New Jim Code, Janelle Monáe, Afrofuturism, anti-racist artistic practice, transmedia

Within a colorful environment atop a constantly shifting terrain of strikingly colored fractal structures, a towering virtual version of pop star and actress Janelle Monáe appears like an oversized goddess (figure 8). As the various avatars of human audience members scramble to find the best spot to

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH09

Fig. 8. Janelle Monáe in 'Music of the Cosmos' (The Mill).

witness her live performance in Oculus VR, the singer first speaks some reassuring words of welcome:

> Yes, you're in virtual reality, and you are in the future, and you will have a liberating, transformative experience. This experience is meant to remind you of your power, of who you are, and what you are capable of doing and being. This is Black joy. (as cited in Daw 2020)

Thus began the special Virtual Reality (VR) event "Music of the Cosmos," the third and last of a series of VR experiences that accompanied the first (and only) season of the HBO series *Lovecraft Country*. During this event, audience members in attendance were treated to the live performance of three tracks, preceded by short explanations of why those particular songs were chosen for an occasion that marked the ending of a unique TV show about the complex relationship between American genre fiction and anti-Black racism. In this way, the artist's presence as a datafied performer in virtual space connected the topics addressed in the series to social debates regarding race, gender, and sexuality. At the same time, the virtual nature of the performance reversed the racialized and gendered power of data, demonstrating how it can also be used as a powerful tool of resistance.

This provocative articulation of performative data offers the perfect entry point into a discussion of the ways in which contemporary Black artists have intervened in debates about our increasingly datafied society

(Schäfer and van Es 2017). As a musical artist prominently associated with the Afrofuturist movement, Janelle Monáe has devoted her musical career to exploring the ways in which technology can be used as a symbolic reflection of racialized and gendered oppressions. In her initial series of concept albums that appeared from 2008 to 2013, she developed an Afrofuturist narrative in which she played the role of an android alter ego named Cindi Mayweather, whose adventures in a twenty-eighth century dystopia were a symbolic reflection of a historical reality in which Black women have been systematically objectified and dehumanized. By drawing attention to the multiple oppressions inflicted upon an android doppelganger, her albums and music videos illustrated above all how technology serves "a racial order of things in which humanity can be affirmed only through degraded categories created for use, exploitation, dispossession, and capitalist accumulation" (Atanasoski and Vora 2019, 13).

More recently, Monáe has shifted from the mechanical figure of the android to the concept of the "dirty computer" as her primary metaphorical vehicle. Where androids have often been used as fruitful vessels for critiques of computer technology as a physical infrastructure, the performances that inform and surround her transmedial *Dirty Computer* project (2018) intervene in the ways we understand the more abstract and disembodied data flows that define the workings of our network society (Castells 1996). In this chapter, I use Monáe's work to illustrate the tension between two separate but interwoven aspects of data as media/performance: first, how the social and technological power of data has historically reinforced racist power structures, codified into assemblages of hardware and software that constitute elaborate networks of surveillance and oppression; and second, how Black performing artists can also use data as a concept to illuminate, expose, and critique these networks through datafied performance across media. In Monáe's work as a musician and performing artist, this work is both metaphorical—reflecting on the meaning of these power dynamics—and literal—using purely digital performance to claim a space within those networks. The first part of this analysis will be focused on data-as-metaphor in the *Dirty Computer* project, and the second on data-as-performance in the aforementioned VR-event.

Both parts are approached from within an overarching cultural and historical framework that Ruha Benjamin has dubbed the "New Jim Code": an understanding of new digital technologies and infrastructures that "reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era" (2019, 3). Benjamin's term constitutes a datafied

play on critical race theorist Michelle Alexander's influential use of the phrase "the New Jim Crow" to identify the post-1960s era of racialized mass incarceration (2020). Alexander's observation that we have merely redesigned racial caste in the "post-racial" age is directly reflected by the choices that have shaped our technologies (Alexander 2020, 2). The New Jim Code captures how our digital platforms, infrastructures, and data strengthen a social system that remains grounded in structures of racist oppression.

But even as the New Jim Code continues to manifest itself as an extension of this system of power, these same technologies can also be used as tools of resistance. From online activism organized around hashtags like #BlackLivesMatter and #OscarsSoWhite to the proliferation of Black online communities, digital platforms are best described as sites of struggle. Across these sites, Black artists like Janelle Monáe have combined an ongoing critique of the workings of the New Jim Code with elaborate and effusive celebrations of Black arts, cultures, and creativity: "as celebrations of self in defiance of norms that can be imposed by both external and internal forces" (Brock 2020, 131).

Located within the broader field of Black studies, my approach throughout the chapter is more specifically informed by the emergence of *race critical code studies*: a theoretical framework grounded in the notion that technology and race are mutually constitutive concepts (Benjamin 2019, 44–46). Race critical code studies sees race as a technology, in the sense that racism is considered "as not just an ideology or history, but as a set of technologies that generate patterns of social relations, and these become Black-boxed as natural, inevitable, *automatic*" (Benjamin 2019, 44–45). By the same token, this means that technology and data sets are always already racialized in that they structurally reiterate "use, value, and productivity as mechanisms of hierarchical differentiation and exploitation within racial capitalism" (Atanasoski and Vora 2019, 15).

From this perspective, I also approach "data" in a broad sense as a central component within common infrastructures of information, both in the contemporary network society of digital data systems and in their non-digital data predecessors. In both cases, data-driven infrastructures constitute crucial platforms on which Black feminism's central dialectic of oppression and resistance is played out (Collins 2000, 15): they constitute a terrain of political struggle where technological data systems can be used both as a tool of racist oppression and as a site of antiracist resistance. Janelle Monáe's *Dirty Computer* project vividly illustrates both these dimensions.

Dirty Computer: Data as Memory, Identity, and Power

Janelle Monáe released her fourth album *Dirty Computer* in April 2018, alongside a forty-eight-minute film—or, as she prefers to call it, an "emotion picture"—that provides a narrative frame for most of the songs on the album. Like Monáe's previous work centered on her android alter ego, the narrative elements are once again science-fictional: the emotion picture depicts a near-future dystopia in which the fascist movement "New Dawn" has come to power. In this nightmarish police state, those who deviate from the norm are classified as "dirty computers" and undergo an extreme form of technologically mediated psychological conditioning. The voice-over that opens the film introduces this world and its organization as follows:

> They started calling us computers. People began vanishing, and the cleaning began. You were dirty if you looked different. You were dirty if you refused to live the way they dictated. You were dirty if you showed any form of opposition at all. And if you were dirty, it was only a matter of time…

The rest of the emotion picture proceeds to illustrate how this "cleaning" takes place in practice. Monáe plays a captured "dirty computer" identified only as Jane 57821, who has been taken captive by New Dawn. We follow Jane 57821 through New Dawn's facility, where she is processed as a "dirty computer" whose memories can be accessed and deleted like data on a computer. Once the two bored-looking technicians manning the control booth launch the software on an interface called the "Memmotron," the film's narrative structure takes shape: the technicians use the interface to view Jane's memories, which we then access in audiovisual form before each one is casually deleted. The body of the film is therefore made up of eight musical sequences based on musical tracks from the album, each of which is presented as a specific memory that is subsequently erased. This structure allows for a tremendous diversity of styles across these various memories, which together sketch out a slender but effective narrative of polyamorous pansexuality, political resistance, and transgressive experimentation. At the same time, it has the practical advantage of facilitating the use of the individual segments as more traditional stand-alone music videos that could be used to promote the *Dirty Computer* album and tour.

As we cycle through the data stored in the character's mind alongside the technicians operating the interface, the songs at first represent memories of Jane 57821's recent life, while later sequences operate more like dreams or visions. These musical segments thereby come to represent a part of the

character's consciousness that is captured by technology and transformed into data that can be stored, replayed, manipulated, and deleted by those wielding the technological apparatus. While the Memmotron interface is supposed to depict the subject's memories, later sequences in the process show parts of Jane's consciousness that confuse the technicians, as they are clearly unclassifiable as such. But irrespective of the nature of the thoughts, memories, and dreams harvested by the Memmotron interface, the NeverMind technology reduces it all to data: ones and zeroes that can be used by those in power as a tool of oppression as we watch the technicians delete all the data they find, irrespective of the contents.

This depiction of technology as an extension of social and political power is a common motif in science fiction, from the telescreens of George Orwell's *1984* (1949) to the endless varieties of digital media run amok in the popular television series *Black Mirror* (2011–). But in the emphatically racialized way it is employed in *Dirty Computer*, the use of invasive surveillance technology to monitor and condition the interior lives of Black subjects operates in the same register as the "New Jim Code": "the employment of new technologies that reflect and reproduce existing inequities but are promoted and perceived as more progressive than the discriminatory systems of a previous era" (Benjamin 2019, 5–6). As put forth in influential books like Benjamin's *Race After Technology: Abolitionist Tools for the New Jim Code* and Safiya Noble's *Algorithms of Oppression: How Search Engines Reinforce Racism*, the neutral-seeming commercial platforms that prey upon our data profiles are meanwhile "encoding race, ethnicity, and gender as immutable characteristics that can be measured, bought, and sold" (Benjamin 2019, 21).

By extrapolating the racial oppression that is already baked into our data-fied society's New Jim Code, *Dirty Computer* is a performative intervention that shines a bright light on racist habits that commonly "enter through the back door of tech design" (Benjamin 2019, 160). It shows how ownership of data-harvesting technology is connected to racial capitalism's grossly asymmetrical organization of power (Robinson 2000, 4), while New Dawn's use of automated drones for capture, interrogation, and surveillance dramatizes how these technologies also strengthen the more directly coercive and punitive state forces of state-sanctioned violence. At the same time, the many moments of collective joy contained within Jane 57821's memories also show how the New Jim Code is endlessly countered by resistant forms of queer Black activism that are calibrated "to discern a multiplicity of interlocking identity components and the way they affect the social" (Muñoz 8).

Dirty Computer's narrative of resistance to the New Jim Code and its accumulation of data as a means of control draw on the contemporary

practice of Black cyberculture's digital ratchetry: digital practices "born of everyday banal, sensual, forward, and 'deviant' political behavior that is rooted in Black culture and discourse" (Brock 2020, 126). The emotion picture's allegorical frame story illustrates the tremendous value that lies in shared expressions of Black joy—even if the digital platforms on which we share those moments are owned and operated for profit by companies with a vested interest in maintaining racial capitalism's multiple forms of oppression (Srnicek 2017). But the collection of data as a form of collective resistance to racial capitalism's racialized and gendered forms of dehumanization has a much longer history that frames this ratchetry as one of many contemporary extensions of the Black radical tradition (Robinson 2000). This transformative media work is what connects Monáe's intervention to the one contained within the *Lovecraft Country* TV series that her VR performance helped promote.

Lovecraft Country: Anti-Black Data from Jim Crow to the New Jim Code

Lovecraft Country's narrative begins by focusing on specific forms of pre-digital data collection that give Black people important tools for survival in an anti-Black environment. Adapted from the 2016 novel by white author Matt Ruff, the series begins by introducing a character who edits and researches a travel guide listing venues in the Jim Crow-era Southern US that are safe for Black people to visit.[1] A scene in episode 1 "Sundown" illustrates the importance of reliable data collection, as the main characters attempt to have lunch at a diner that has changed ownership since the guide's last edition was published, and the trio just barely survives the violent encounter with racist whites that ensues. A similar scene, in which series protagonist Atticus Freeman is pulled over by a racist police officer who informs him about "sundown towns"—all-white communities in which Black people were instructed to leave before the sun had set—underlines how detailed data about geographical boundaries, social norms, and precise time measurement were all vital to Black travelers. Again, reliable and current data are thereby shown to be essential in a white-centered society where Black lives are considered disposable and where data collection on safe routes and locations can be seen as an early form of Black ratchetry.

1 The travel guide is based on Victor Hugo Green's annual publication of *The Negro Motorist Green Book*, which also inspired the Oscar-winning 2018 film *Green Book* (Brock 2020, 2–5).

Lovecraft Country dramatizes these historical instances of Black resistance through data collection from the twenty-first century context of #BlackLivesMatter and other new forms of antiracist activism. The series thereby emphasizes above all the continuities between past and present forms of institutional racism. As Saidiya Hartman wrote on the fundamental importance of historical narratives about racial histories:

> The past is neither inert nor given. The stories we tell about *what happened then*, the correspondences we discern between today and times past, and the ethical and political stakes of these stories redound in the present. If slavery feels proximate rather than remote and freedom seems increasingly elusive, this has everything to do with our own dark times. If the ghost of slavery still haunts our present, it is because we are still looking for an exit from the prison. (Hartman 2007, 133)

The fictionalized depictions of specific ways in which American society was shaped by anti-Blackness therefore provide important data for contemporary artists, activists, and citizens. With popular films and television series like *Black Panther* (2018), *Watchmen* (2019), *Antebellum* (2020), *Them* (2021), and *The Underground Railroad* (2021) each offering a strong focus on specific forms of systemic racism, popular narrative media are contributing to the #BlackLivesMatter movement by filling in vital gaps in white society's collective cultural memory. The abundance of such gaps in the widely available data on Black and diasporic histories, both in the US and in Europe, demonstrates painfully how one key pattern of suppression is indeed that of omission (Collins 2000, 8). The documenting, sharing, and reproducing of data documenting these systems of racial oppression has been central to the Black radical tradition, from the hidden codes and symbols that were shared among enslaved Africans to the resistant use of data by Black social media users in the twenty-first century (Taylor 2016, 217–18).

But as with the *Dirty Computer* example discussed above, this is more than merely the documentation of racist histories—it is also the transformation of this historical legacy into new histories that have the power to shape new theory (Robinson 2000, 307). Considering, then, that the liberal humanist tradition has traditionally defined Blackness as the negative inverse of the white liberal subject, this perception of Blackness has been defined by *plasticity*:

> Plasticity is a mode of transmogrification whereby the fleshy being of blackness is experimented with as if it were infinitely malleable lexical

> and biological matter, such that blackness is produced as sub/super/human at once, a form where form shall not hold: potentially "everything and nothing" at the register of ontology. (Jackson 2020, 3)

Lovecraft Country intervenes in this conception of Black plasticity by showing it while at the same time disrupting it. Both the novel and the series do this in the first place by engaging directly with the overwhelming whiteness of genre fiction (Carrington 2016, 16–17): not only does the show deviate from the traditional norms governing science fiction, fantasy, and horror productions by featuring a majority-Black cast, but on another level, the characters acknowledge and negotiate their own ambivalence in relation to a genre that suffers from a tremendous imagination gap when it comes to the depiction of Blackness (Thomas 2019, 5).

This longstanding imagination gap derives from the fact that fantastic fiction has revolved around the figure of the Dark Other as racialized antagonist, forever associating Blackness with "the monstrous Thing that is root cause of *hesitation*, *ambivalence*, and the *uncanny*" (Thomas 2019, 23). The implicit message—that Black readers of such texts are themselves the real monsters—has a lengthy tradition in western fantasy, thereby obviously contributing to a "nameless and lingering fear of dark people in the present" (Thomas 2019, 20). Few fantastic authors have been more influential in this regard than H.P. Lovecraft, the notorious racist, misogynist, and antisemite whose enduringly popular weird fiction introduced the Cthulhu mythos alongside a worldview in which Black people were invariably depicted as subhuman savages (Rieder 2008, 45).

Lovecraft Country wrestles with this legacy of anti-Blackness in fantastic fiction on two levels: first by presenting itself as a revisionist adaptation of Lovecraft's cultural legacy, and second by featuring Black characters like Atticus who are fans of stories that have such painful flaws that Atticus's uncle George remarks, "They stab me in the heart" (Ruff 2016, "Lovecraft Country"). In both ways, the series pushes back against a White supremacist normative framework that bases its ideas about the future on the sins of the past. Translated to the era of the datafied society, this means pushing back against the cultural logic of the New Jim Code and the countless predictive algorithms that harden the dehumanization of Black people by rendering it "objective" via mathematical codes (McKittrick 2021, 113–14). Thus, shifting Blackness from the Dark Fantastic's periphery to its center comes to represent a powerful act of "counter-coding" designed to collate these lesser-known stories as "certainties that underlie the brutal statistics, traits and mathematics of dysselection" (McKittrick 2021, 115).

As a popular media text that pushes back against the predictive algorithms that have effectively automated white supremacist systems of social and technological power, *Lovecraft Country* counter-codes the genre's dominant narrative, using a transformative Black imagination to expose and transform existing realities that "had long been centered on reinforcing extant power structures" (Zamalin 2019, 16). By focusing within the series on the resistant sharing of data and the counter-coding of monstrous narratives that strengthen white supremacist social structures, the show illustrates how the seemingly neutral predictive algorithms and other digital data streams are in many ways merely the latest incarnations of a public infrastructure that is fundamentally hostile to Black people. And like so many narratives that derive from the Black radical tradition, it combines this critique of racialized data platforms with a vivid emphasis on the resilience of alternative systems of data collection and distribution, showing how to navigate this world "as a laborious aesthetics of freedom-making" (McKittrick 2021, 68).

Music of the Cosmos: Black Performative Data Archives

Both Janelle Monáe's performance-driven resistance to extractive and exploitative data systems and *Lovecraft Country*'s narrative focus on historical precursors to the New Jim Code come together in "Music of the Cosmos," the VR-set live performance that followed directly after the show's final episode.[2] This was the third installment in a series of VR-based events produced to support and promote the show, titled *Lovecraft Country: Sanctum*. Media and technology company The Mill, which produced the series of events, describes them as follows:

> *Lovecraft Country: Sanctum* is a three-part immersive journey to another dimension. "Travelers" visiting the Sanctum Dimension experienced three different events that are a complement to *Lovecraft Country*'s themes of socially relevant storytelling and cosmic horror, featuring original content performed by the cast from the series. In the first event, guests explored a mesmeric sculpture garden, with original artworks crafted by Black Afrofuturist artists. In the second, visitors entered an otherworldly interactive theatre that defies the laws of physics. The third event featured an unforgettable musical performance by one of today's biggest music

2 The first season's last episode would also be the series' last installment, as HBO decided not to renew the series beyond its initial ten episodes.

> stars, Janelle Monáe. Throughout this social VR experience, guests are [sic] able to interact with other Travelers, as well as unlock hidden rooms and challenges that took them further into the dark unknown. (The Mill, n.d.)

Much of the wording here typifies the (self-)promotional nature of this campaign, which was obviously commissioned first and foremost to help publicize a high-profile new show on a premium cable network. But those commercial PR-based imperatives do not cancel out the creative choices that still meaningfully contribute to antiracist activism. By drawing on the strengths of data platforms to incorporate resistant Black digital practices (Brock 2020, 30), the *Sanctum* VR events provided transmedia extensions of *Lovecraft Country*'s counter-coding of narratives, histories, and technologies that reinforce white supremacy.

The first event, "Garden of Eden," had taken the form of a virtual art show, transforming physical installations by Black artists David Alabo and Devan Shimoyama into digital reproductions that the audience could move around and interact with, while Afrofuturist designer Yung Yemi created a towering virtual sculpture that visitors could go inside and explore. The VR installations deliberately made use of affordable mobile hardware (the Oculus Quest technology) alongside the existing VRChat platform for social interactions, thereby making the event as accessible as possible—even if there were only a limited number of admissions available for each event, and these were primarily targeted at high-profile online influencers. The designs of the space and installations deliberately used the kind of brutalist aesthetic that is maximally functional within a digital platform and its limited affordances for transferring and rendering massively complex data.

The second event, "An American Dream," presented a virtual dialogue between Black author and intellectual James Baldwin and a Black woman "living in the present" (The Mill, n.d.). Actress Jurnee Smollett, who played one of the main characters on the show, responded in this immersive theatrical performance to words made famous during Baldwin's legendary 1965 television debate with conservative firebrand William F. Buckley. The resulting performance established a dialogue between past and present that made palpable both the continuities and differences in racial perspectives across these different periods.

The "Music of the Cosmos" event did this work by combining a well-known pop artist's datafied performance with an interactive environment that gave visitors access to a space called "the Reliquary": a "mystery room," hosted by actor Michael Kenneth Williams, where visitors could interact with a wide variety of artifacts with accompanying audio about the histories

of racism that formed the series' primary context. Much like previous notable transmedia campaigns that simultaneously promoted the brand and deepened the storyworld, the *Sanctum* events thereby contribute to the franchise's central text by offering supplemental information that deepens fans' engagement with the narrative and its various layers of meaning (Hassler-Forest 2016, 145). In the case of *Lovecraft Country: Sanctum*, these supplemental data deviate from the usual focus on narrative, focusing instead on the racial organization of power and meaning that the series attempts to reveal and disrupt.

This disruption is given another dimension by Monáe's musical performance, appearing as a towering digital avatar whose actions manipulate the appearance of the entire virtual space (The Mill, n.d.). As in the preceding *Sanctum* events, the overarching goal was again to create a performative dialogue between past and present that illuminates the social and cultural powers of anti-Blackness and the creative responses this oppression engenders. In this case, Monáe's performance of the three tracks "Django Jane," "Americans," and "Come Alive (War of the Roses)" was staged and introduced as a virtual recreation of a 1950s Chicago block party, as featured in the season's pilot episode. Like these informal neighborhood expressions of Black culture, creativity, and community, Monáe's innovative virtual performance represented a powerful way in which media and performance could be used to illustrate the tension that exists between data and social power in a datafied society.

Conclusion

In this chapter, I have offered a reflection on the complex relationship between data and race in the era of "the New Jim Code." In a social and technological context in which digital data platforms play a vital role in anti-Black racist oppression, media texts and performances illuminate how deeply entwined data systems and social power have become. At the same time, the dialectic of oppression and resistance yields forms of activist resistance that use ubiquitous datafication to antiracist ends.

To illustrate this dialectical tension, I have looked at how pop star, actor, and activist Janelle Monáe has used figures like the android and the "dirty computer" to give dramatic form to these datafied forms of oppression, while appropriating its forms and vocabulary in powerful acts of counter-coding. By the same token, the TV series *Lovecraft Country* counter-codes the cultural legacy of fantastic fiction and its reliance on the Dark Other as monstrous

antagonist, reversing the horror genre's racial dynamic to foreground its enduring resonance while at the same time emphasizing the New Jim Code's non-digital antecedents. These two texts then came together in Monáe's VR performance, which deepens and extends *Lovecraft Country*'s forms of ratchetry, yielding a VR performance designed to educate its audience not just about the power inherent in datafied power, but even more crucially, about how its power can be harnessed by inspired antiracist artists across media.

References

Alexander, Michelle. 2020. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness – 10th Anniversary Edition.* New York: The New Press.

Atanasoski, Neda, and Kalinda Vora. 2019. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke University Press.

Benjamin, Ruha. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity Press.

Bonneuil, Christophe, and Jean-Baptiste Fressoz. 2016. *The Shock of the Anthropocene: The Earth, History, and Us*. Translated by David Fernbach. London: Verso Books.

Braidotti, Rosi. 2019. *Posthuman Knowledge*. Cambridge, UK: Polity Press.

Brock, André Jr. 2020. *Distributed Blackness: African American Cybercultures*. New York: NYU Press.

Carrington, André M. 2016. *Speculative Blackness: The Future of Race in Science Fiction*. Minneapolis, MN: University of Minnesota Press.

Castells, Manuel. 2009. *The Rise of the Network Society*. 2nd Edition. Hoboken, NJ: Wiley-Blackwell.

Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Daw, Stephen. 2020. "Janelle Monae Transforms into a 3D Avatar for VR 'Lovecraft Country' Performance." *Hollywood Reporter*, October 20, 2020. https://www.hollywoodreporter.com/news/general-news/janelle-monae-transforms-into-a-3d-avatar-for-vr-lovecraft-country-performance-4079671..

Haraway, Donna J. 2016. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham, NC: Duke University Press.

Hartman, Saidiya. 2007. *Lose Your Mother: A Journey Along the Atlantic Slave Route*. New York: Farrar, Strauss and Giroux.

Hassler-Forest, Dan. 2016. *Science Fiction, Fantasy, and Politics: Transmedia World-Building Beyond Capitalism*. London: Rowman and Littlefield International.

Jackson, Zakiyyah Iman. 2020. *Becoming Human: Matter and Meaning in an Antiblack World.* New York: NYU Press.

McKittrick, Katherine. 2021. *Dear Science and Other Stories.* Durham, NC: Duke University Press.

Mill, The. n.d. "HBO's 'Lovecraft Country' Social VR Event." https://www.themill.com/work/case-study/hbo-lovecraft-country-sanctum-virtual-event.

Nash, Jennifer C. 2019. *Black Feminism Reimagined: After Intersectionality.* Durham, NC: Duke University Press.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Rieder, John. 2008. *Colonialism and the Emergence of Science Fiction.* Middletown, CT: Wesleyan University Press.

Robinson, Cedric J. 2000. *Black Marxism: The Making of the Black Radical Tradition*. Chapel Hill, NC: University of North Carolina Press.

Ruff, Matt. 2016. *Lovecraft Country*. New York: Harper Perennial. Kobo.

Schäfer, Mirko Tobias, and Karin van Es. 2017. *The Datafied Society: Studying Culture through Data*. Amsterdam: Amsterdam University Press.

Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge, UK: Polity Press.

Taylor, Keeanga-Yamahtta. 2016. *From #BlackLivesMatter to Black Liberation*. Chicago: Haymarket Books.

Thomas, Ebony Elizabeth. 2019. *The Dark Fantastic: Race and the Imagination from Harry Potter to the Hunger Games*. New York: NYU Press.

Yusoff, Kathryn. 2018. *A Billion Black Anthropocenes or None*. Minneapolis, MN: Minnesota University Press.

Zamalin, Alex. 2019. *Black Utopia: The History of an Idea from Black Nationalism to Afrofuturism*. New York: Columbia University Press.

About the Author

Dan Hassler-Forest works as Assistant Professor of Media and Culture Studies at Utrecht University. He has published books and articles on superhero movies, comics, transmedia storytelling, critical theory, and Afrofuturism. He has recently completed two books on race and global media, focused on the creative work of performing artist Janelle Monáe.
> d.a.hassler-forest@uu.nl

Part 3

Knowledges

10. How Eva Louise Young (1861–1939) Found Me

On the Performance of Metadata in Knowledge Production

Iris van der Tuin

Abstract
Human knowers in academic settings today are caught up in computational procedures. Such procedures have constraining and surprising effects on the "findability" of scholars and scholarly works. This chapter argues that, and shows how, digital literacy is beneficial for epistemological and methodological reflection and creativity during the research process. Unraveling the intricacies of the chapter's author meeting a "forgotten" philosopher—Eva Louise Young (1861–1939)—in a situated human–computer interaction meant acquiring the competence of being critical of, and creative with, Google's functioning.[1] It meant learning that, in today's algorithmic condition, canonization and knowledge production are complicated *posthuman* entanglements. Literacy here means combining tool criticism and creativity from media studies with bioinformatical practices of data and information storage, labeling, and retrieval in dynamic settings.

Keywords: algorithmic functioning, creativity, digital literacy, doing research, findability, human–computer interaction

In October 2016, I met the British philosopher Eva Louise Young online. Young was born in 1861 in the Punjab in the former British colony of India,

1 I presented this case study first in a keynote at the *8th Annual Conference on the New Materialisms* in Paris in 2017 and included the case in my 2018 inaugural lecture as well (see van der Tuin 2018).

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH10

and she died in Letchworth Garden City, England in 1939. Back in 2016, I was searching for literature in the field of posthumanist theory via Google Books on my laptop, and Young's book, *A Philosophy of Reality* from 1930, appeared as one of the search results. Nothing about the scanned pages of the book that appeared on my screen after several mouse clicks made it explicit that its author, E. L. Young, is a woman, yet Google Books' sidebar made mention of "Eva Louise." While I was not looking for a publication in metaphysics, I allowed myself to be distracted by this record, dynamically sourced from library metadata in response to my search terms, search history, and user profile. It struck me that I was unfamiliar with the author (I thus immediately critiqued processes of philosophical canonization), but I also realized that she had the potential of becoming important to my feminist research (I was thus immediately creative with the search result in the hope of broadening posthumanism's knowledge base). Using the internet, I was quickly able to uncover several additional facts: Young wrote only a single philosophical monograph, and beyond the 1930s, her work has hardly been referenced. Furthermore, reviews of the book in the 1930s were predominantly negative, and many of the reviewers were mistaken about her gender. I also noticed the difficulty of finding information about Eva Louise Young online; simple Google searches generated few results. In sum, Young has been effectively erased from history (i.e., *eclipsed from view*). The photograph that I found later in the Garden City Collection nicely illustrates this point (see figure 9). So, why did, and do, I attach such importance to an obscure search result?

Young surfaced as a piece of information on the search results page displayed by Google Books in response to one of my queries – importantly, *not* a query directly pertaining to Young herself or even to her monograph *A Philosophy of Reality*. She surfaced in the thick of the non-exhaustive workings of a situated human–computer interaction, the workings of which interest me in this chapter on how metadata are active participants in processes of canonization and knowing today. How did the British philosopher Eva Louise Young—who died decades before the internet began to influence the philosophical profession—find me online seventy-seven years after her death? What does a provisional or perhaps *speculative* answer to this question reveal about the doing of research in the twenty-first century? Key to the discussion of how Young found me online is the question of what enabled her to do so—that is, under what operational logics and socio-technical conditions could this long-dead philosopher grab my attention for a forward citation in this chapter? I thus play with logics of "(un)findability"—here utilized conceptually for understanding search processes that are more

Fig. 9. LBM3056.43.47 – Digital copy of a photograph of Eva Young, her brother, and Lord Lytton. (Courtesy of Garden City Collections).

complex than those conducted linearly by a goal-oriented user of an online search platform.

Taking the everyday reality of scholars, educators, and students doing research online as its starting point, this chapter acknowledges that, in the "algorithmic condition" (Colman et al. 2018), the Internet is easily and frequently accessed via one's laptop, tablet, or mobile phone even while reading a paper copy of an article or a book. The internet, here, is the global and lively memory bank that we use for storing, accessing, and transmitting affects, data, information, and knowledge, both professionally and privately. Importantly, the internet is also used for the *building* of not only sets of affect icons, data warehouses, information systems, and knowledge centers, but also of the affective and data relations, pieces of information, and knowledge claims themselves. This chapter positions itself after the turn to artificial intelligence (AI) in our knowledge-producing endeavors as they are globally conceived and locally enacted. I unravel specifically how we, as humans today, produce knowledge through the screen and with so-called "metadata." I propose how to be critical of, but more importantly also creative with, the computational procedures that both impact us and—through our search and click behavior—in which we participate.

Metadata are data about other data, such as an author's name on the cover of a book or specific facts about the size and date of a digital file. This kind of data is always already part of a classificatory structure that

gives meaning to both the classes (name, size, date) and, through the structure itself, their contents. It is obvious that by naming, classifying, and managing classes XYZ and potentially ignoring classes ABC, the power to define and value are intrinsic to metadata practices (Acker 2021). In this chapter, I will argue that and show how metadata are curiously *active* in processes of knowing through data labeling, categorization, and prioritization. This machine-learning activity implies that we may also decide to be *creative* with the ongoing performance of metadata along with developing a critical stance on it. Adding metadata to the scan of a philosophical treatise, for instance, does not just make that treatise potentially findable online. The very processes of labeling, categorization, and prioritization are at once generative and restrictive of what can be brought forth in the algorithmic condition. Inclusion in and exclusion from a scholarly canon or some concrete knowledge claims (and not others) were not just brought forth by independently functioning machine-learning algorithms. Rather, they came about in interactions between non-human agents and human knowers via the screens of laptops, tablets, or mobile phones. This makes for a complex situation that cannot be easily understood or equivocally judged.

Today's knowledge is being produced not by humans alone but rather by humans, such as we ourselves, in a complicated relation of *entanglement* with non-human agents, such as the algorithms sourcing, using, and presenting metadata on online platforms. Humans and non-human agents today form an intricate "cognitive assemblage" with predictable, emergent, and surprising epistemic effects (Hayles 2017). One could say that the argument I am making about knowledge production in contemporary media-technological landscapes is a "posthuman" one (cf. Thylstrup 2018, 21). Some knowledges constructed through algorithm-driven platforms such as Google Books are exciting new opportunities for further research that extends social-constructivist "tool criticism" of the regulative and regulated nature of Google's search engine; this function is entwined with the user in what could perhaps be called "tool *creativity*," a perspective affirming that algorithmic functioning may at times be more playful than just predictive and predictively consensual.[2] In a situated human–computer interaction, a surprising find (a long-dead and forgotten female philosopher, for instance) may surface as the result of connecting metadata from the lively memory bank that is the internet. After this, a scholar *recognizing* the find

2 For tool criticism, see: Koolen et al. 2017 and van Es et al. 2021. For consensuality and Google's search engine, see: van Dijck 2010.

as potentially interesting does her research—again, often on Google— and thus in turn both user profiles and (potentially) the philosophical canon are affected.[3] Importantly, a "philosophical canon" is also a regulative and regulated device that needs both criticism and creative use in order for research to leap into the future.

Researching (in) the Algorithmic Condition

Researching a specific case of the performance of metadata is important in the light of present-day calls for "situated knowledges" (Haraway 1988) in the study of data use and computational machine learning. Why? Because when we as scholars and students refrain from reflecting on how our research objects appeared to us and ignore that, more often than not, this happens online, we are at risk of repeating exclusionary practices in terms of both the research itself (we should not ignore that our interest was raised in a media-technoscape) and in terms of who or what we cite as our academic and non-academic "influencers" or collaborators. Library and information scholar Anna Lauren Hoffmann argues that "the problem here isn't only one of biased datasets or unfair algorithms and of unintended consequences. It's also indicative of a more persistent problem of researchers actively reproducing ideas that damage vulnerable communities and reinforce current injustices" (in Crawford 2021, 117). In her book *Atlas of AI*, researcher of artificial intelligence Kate Crawford argues that the call for a responsible use of data and reliance on algorithms has a history that goes back to at least the 1970s with computer scientist Joseph Weizenbaum arguing against a techno-determinist perspective and in favor of an embodied (i.e., situated) perspective in his 1976 monograph *Computer Power and Human Reason: From Judgment to Calculation*:

> The lesson, therefore, is that the scientist and technologist must, by acts of will and of the imagination, actively strive to reduce such psychological distances, to counter the forces that tend to remove him [sic] from the consequences of his actions. He must – it is as simple as this – think of what he is actually doing. (as cited in Crawford 2021, 118)

Taking situatedness onboard, I will now ask what it means to do research in the algorithmic condition, thereby in fact researching this condition.

3 For recognizing surprise, see: Darbellay et al. 2014.

In today's world of networked ICTs and dynamic machine learning, the constitution of the researcher (the subject doing the knowing) and the researched (the object about which the scholar or the student is curious) has fundamentally shifted compared to the previous "postmodern condition" of stand-alone desktop computers and statically wired networks (Lyotard [1979] 1984). As researchers, we must now demonstrate a certain literacy around digital technology (Bühlmann et al. 2017; Erstad 2010; Johannesen et al. 2014; Koolen et al. 2019; van Dijck 2010) in order to function well, by which I mean that we must be able to use algorithm-driven platforms reflectively; we must also be critical toward the implied construction of truth and toward media technologies participating in knowledge production and in processes of in- and exclusion. We live in a time in which the thoroughly entangled nature of our human knowing with algorithmically driven search engines increasingly leads to situations that are characterized by ignorance, indifference, or the "user unconscious" (Clough 2018). Who are we as "posthumans" and in what kind of world do we live when this world, its objects, and our data-subjectivities come into being according to computational procedures that are generally invisible *and yet* have far-reaching epistemic effects?

As contemporary researchers making daily use of our laptops, tablets, and mobile phones, we edge our way through the dynamic archives that are consulted via the internet of algorithmic media. If we follow thinkers such as the French philosopher Michel Serres ([2012] 2015), then we become subjects in this world by engaging with/in such media. We become "Thumbelinas" or "Tom Thumbs" in the process of working with, and adding the power of definition and value to, pieces of information and click-worthy visuals. The algorithmic media themselves and the digital or digitized artifacts that emerge in such human–computer interaction are constituted as objects through networked acts. In our times, archives are still organized hierarchically—think of university libraries or Wikipedia—and they propagate familiar ways of structuring and representing knowledge. We must be critical toward this, as everything familiar is gendered, racialized, sexualized, etc. Yet by virtue of their "on-demand" nature, our online archives are also inclusive (albeit in a rudimentary form), and thus there is room for unexpected creativity. What I encounter online depends on the inputted query, my search history, the way in which I deal with issues of privacy, and the extent to which I allow the algorithms and computational procedures of the various platforms and search engines to access my searching behavior for user profiling. After all: "every swipe [is] a record in a database [...] [and] every choice we make is recorded"

(Witten and Frank 2005, 4). It is precisely for this reason that my search results differ from yours and that bias in, and responsible play with, truth and value are at stake.[4]

On Metadata Participating Algorithmically in the Research Apparatus

The future of Young's inclusion in the philosophical canon was—in media theorist Wendy Hui Kyong Chun's language of "programmability"—predicted and shaped based upon past data primarily through Google's tracking of my IP address's search terms. The service provided by American multinational tech company Google, i.e., Google Books, brought Young to me "through the data traces produced by [my] mappings" online (Chun 2011, 8). One could say that the interactive and ideological interfaces of Google "have been key to creating [the] 'informed' individual [here: Iris van der Tuin] who can overcome the chaos of global capitalism by mapping their relation to the totality of the global capitalist system" (8). However, speaking in an emancipative sense—and perhaps a little naively so—the rare event of attributing the book of metaphysics, *A Philosophy of Reality*, specifically to the female philosopher Eva Louise Young assured that, at the same time as corporate-run and market-driven interfacing takes place, "our computers execute in unforeseen ways, the future opens to the unexpected. Because of this, any programmed vision will always be inadequate, will always give way to another future" (9; cf. Gauthier 2016; Verhoeff and van der Tuin 2020). My naive enthusiasm about *the woman* E. L. Young can perhaps be traced to what has been described as the desire to be involved in change as the pleasurable fantasy that one is a "change agent" online: "we click, we change," summarizes Chun (2011, 69).[5] The desire to perhaps change the genealogy of posthumanism as a theoretical landscape (cf. Braidotti and Hlavajova 2018; Braidotti et al. 2022) by adding a yet unknown female philosopher to it could only happen by ignoring the machine reading and writing—the *computation*—that is invisibly performed in order for any user interface to function at all. This blindness creates the fiction of user control and authorship rather than acknowledging interpellation in human–computer

4 N. Katherine Hayles (2017, 32) argues that "the pockets within which technical systems operate autonomously are growing larger and more numerous." Among the examples of increasing autonomy that she gives are digital search engines.

5 Cf. Tara McPherson and Alexander Galloway in Chun 2011 (69, 205 n. 38).

interaction as involving all of the following: oppression, liberation, and diffraction.[6]

When stumbling upon Eva Louise Young's *A Philosophy of Reality*, published in 1930 by Manchester University Press, I thought: "Eva Louise Young?" Google Books' use of metadata from library and other databases provides her full given names, whereas the book itself portrays the author in a gender-neutral fashion as E. L. Young. The latter representation will have led to most book reviews of *A Philosophy of Reality*, published in the 1930s, representing Young as he/him/his (cf. de Beauvoir [1949] 2010). The early reception of *A Philosophy of Reality* was, to say the least, a heavily gendered process with reviewers questioning Young's rhetoric and style as soon as her femaleness was known to the reviewer. But rhetoric and style may very well have been consciously chosen, and her naming strategy may very well have been gender-aware! The point is that I would not have stumbled upon Young in October 2016 without the participation of Google's algorithm. I would have ignored E. L. Young's *A Philosophy of Reality* in a physical library or secondhand bookstore, as I was not looking for a book of metaphysics and because, as a scholar, I am not primarily interested in contributing to research about or in researching with male philosophers. Further, Young has not been canonized; information on Eva Louise Young is not readily available on the Internet or anywhere else. I really needed Google Books to "gender" E. L. Young and to "rank" the presentation of her work to me.

How Metadata Intervene in Processes of Knowledge Production

Google Books is the still-growing result of Google's book scanning activities at university libraries and academic publishers based mainly in the US, Europe, and Japan. Google Books, supported by Google as a larger company, Silicon Valley as an industrial area in California, and the internet globally, allows its users to search the full text of millions of publications that have been 3D-image scanned, converted to text using optical character recognition (OCR), and stored in their digital database. The outcomes of the digitization process for Google Books have been much debated by scholars and journalists alike for the initially low quality of its manuscript images, the poor functionality of OCR, and errors in the associated metadata (James and

6 Diffraction, here, stands for non-linear patterning, and oppression and liberation for predictable exclusive and inclusive linearities. For diffraction in both quantum physics and cultural inquiry, see Barad 2003 and 2007.

Weiss 2012, 16). Google Books is an archive that, as it is supported by Google in particular, is built on a logic of *feedback* (not just "access") between users, machines, and engineers and of controlling the informational *process* (not just the information itself) (Thylstrup 2018, 39). Google Books' descriptive metadata come from a variety of providers (41). As becomes clear in librarians Ryan James and Andrew Weiss's article "An Assessment of Google Books' Metadata," "[t]his process, presumably, involves using humans to generate the metadata" (James and Weiss 2012, 16). Should metadata be missing, however, Google guesses the necessary information for database-completion purposes (16).[7]

Debates about Google Books have mostly centered upon the negative impacts of errors on "the somewhat indeterminate concept of 'findability'" (19). This pertains to Google Books' practice of combining full-text searches with metadata that are affected—supposedly in approximately 37% of all items—by mechanical inaccuracies, not only typographic but also errors affecting meaning, such as misattributions. James and Weiss state: "We do not know the inner workings of the proprietary algorithms Google Books uses to order the search results list, but we can see that metadata are featured prominently on the search results list" (21). While this may very well be the case empirically speaking, I however must dare to disagree with the negative tone of the discussion (cf. Thylstrup 2018, 30, 37–38) given the way in which Eva Louise Young found *me* instead. The featured metadata "Eva Louise Young" generated a leap into the future of posthuman theory, as opposed to *A Philosophy of Reality* simply representing another potentially faulty search result, albeit one that was truthful to the past of a rare book publication.[8] What if we conceptualize findability not as uncertain or as vague, but as coming into being (*emerging*) in an "apparatus" in the sense of feminist science and technology studies scholars Katie King (in Haraway 1988, 595; 1994), Donna Haraway (1988), and Karen Barad (2007)? What if we try to both capture the moment—a Bergsonian "interval," as I will suggest below—of having been found by Eva Louise Young, retrospectively trace what happened in that timespan, and conceptualize the ingredients of the philosophical impetus of this moment's effect on the feminist genealogy of posthumanism?

7 I do not know how this is being done (by employees or algorithmically, or by a combination of both). See also Crawford (2021) on metadata not being the pinnacle of cleanliness (by decontextualization) but a more complex situation instead.

8 Results supposedly refer – from Latin *referre* "carry back," from *re-* "back" + *ferre* "bring" – to an original, flawless, non-digital publication.

Following a logic of authenticity, comparing Google Books' scanned cover of *A Philosophy of Reality* with the book's digital record in Harvard University Library's HOLLIS catalog seems to confirm that the 3D scanning of the book was initially performed in Cambridge, Massachusetts. HOLLIS's metadata—i.e., the book's WorldCat record and its MARC view—present the "E. L. Young" from the book's cover as Eva Louise Young, born in 1861. "MARC" stands for Machine Readable Cataloging, so it is likely that these data (among other data) were used by the Google Books algorithm to fill my ranked search results list in October 2016. But how do we move from the paradigm of spatiotemporally linear (un)findability to conceptualizing findability as generatively coming into being via an apparatus of canonization (here: philosophical) and knowledge production? How do we develop a method for retracing such emergence, "a method attuned to the entanglement of the apparatuses of production, one that enables genealogical analyses of how boundaries are produced rather than presuming sets of well-worn binaries in advance" (Barad 2007, 29–30)? For that, we must turn to the phenomenon of "nanopublication" and the concept and method of "quantum attribution." These practices stem from the field of bioinformatics, a field that equally grapples with data and information storage, labeling, and retrieval in dynamic settings and that may therefore help us be creative with tools.

On Affecting Canonization through Nanopublication and Quantum Attribution

In its most basic form, Google Books' descriptive metadata listed "*A Philosophy of Reality* by Eva Louise Young" in the left sidebar on my screen. This comes down to the assertion: "Eva Louise Young is the author of *A Philosophy of Reality*." This ostensibly insignificant fact has proven extremely meaningful, as at least one of its effects has been the research for this chapter, with another effect the creative addition—in the sense of French philosophers Gilles Deleuze and Claire Parnet's ([1977] 1987) "creative AND"—of E. L. Young to the feminist genealogy of posthumanism. It can be argued that the assertion of Eva Louise Young as the author of *A Philosophy of Reality* was published online only when I stumbled upon the scanned book while using Google Books. Such a simple, published assertion, dynamically generated in an instance of human–computer interaction, could be called a "nanopublication." The phenomenon of nanopublication, in the field of the history of philosophy, consists of publishing historical facts, philosophical facts, and connecting facts, with all such facts being of a simple nature and containing

a subject, a relation, and an object.[9] Here are some nanoassertions about Eva Louise Young and her monograph, *A Philosophy of Reality*:

– Historical fact: Eva Louise Young was a teacher, a gardener, an Esperantist, and the author of *A Philosophy of Reality*;
– Philosophical fact: Matter and mind are the public and private portions of one real, continuous, and comprehensible universe;
– Connecting fact: Eva Louise Young claims that things are what they seem.

These and other nanoassertions can be proven true or false by further research, so they are part of the positivist paradigm of spatiotemporally linear (un)findability. This is confirmed by the apparatus I have been using for these reflections: a website that facilitates nanopublication in the field of the history of early modern philosophy. Let me explain the positivist paradigm first and then move on to a paradigm that accommodates the dynamic non-linearity of the internet of algorithmic media.

The website *Early Modern Thought Online* (*EMTO*) *Nanopub* produces and stores a system of (cross)references to and from databases of libraries and archives linked to an individual researcher's name. The databases in this context comply with the hierarchical model—that is, they statically organize data into tree-like structures. In the vein of archival positivism, the philosophical apparatus of nanopublication would therefore produce descriptions that look something like this: "Iris van der Tuin has confirmed by way of the digitized *Pageant of Letchworth 1903–1914* by A. W. Brunt, first published in 1942 and now available through the website of The Letchworth Garden City Society, that Eva Louise Young was a teacher, a gardener, an Esperantist, and the author of *A Philosophy of Reality*."[10] Given that Young found me via the internet, it is paramount to abandon archival positivism—bound up as this epistemological stance is with the access paradigm of offline hierarchical archives—and to work instead toward a stance that can accommodate a logic of feedback and informational processing in an entangled apparatus of knowledge production. So, again: how to proceed?

The discussion about nanopublication as a phenomenon was, in fact, initiated in the field of bioinformatics under the interchangeable labels "microattribution," "precise citation," and "quantum attribution."[11] In this context,

9 http://emto-nanopub.referata.com/wiki/EMTO_Nanopub. See also http://nanopub.org/wordpress.

10 http://lgcs.org.uk/pageant/index.htm.

11 https://en.wikipedia.org/w/index.php?title=Microattribution.

the discussion exceeds spatiotemporal linearity given that bioinformatical data and information are always on the move. The concept and method of quantum attribution afford vertically static, horizontally dynamic, and transversally contingent acknowledgements of database entries and tagged archival records as situated scholarly contributions (cf. Verhoeff and van der Tuin 2020; van der Tuin forthcoming). In the words of communications and digital media scholar Nanna Bonde Thylstrup (2018, 22): today's knowledge is being produced online "where vertical hierarchies and horizontal networks entwine in a new political mesh" that she calls "networked assemblages." We must specify this statement given that the nanoscale—as the scale of quantum effects in both biology and in informational practices online—has afforded not only the nanopublication of specific entities and events that can be identified (DNA sequences, their variation, and their consequences [e.g., the spread of disease]) but also nanopublication of pure events such as DNA sequencing and (un)controlled genetic mutation (Patrinos et al. 2012). A cartographical "Janus face" thus emerges on this scale. On the one hand, the focus on genetic mutation and variation demonstrates how bioinformatics and therefore the method of quantum attribution are inescapably entangled with biopolitics and other racialized forms of population control. These practices need our critical response. On the other hand, correlational and causal relations do appear non-linearly as well, and they are important for the understanding and taking advantage of creativity in our research on the algorithmic condition (cf. Barad in Dolphijn and van der Tuin 2012, 55).

The bioinformatic nanoassertion, defined as "the smallest unit of publishable information that can be linked to its contributor via their unique scientific identity, and which can be cited and evaluated in terms of its impact upon the research community" (Patrinos et al. 2012, 1506), would first be stored in an open access database, and then a "microattribution analysis article" would "summarize the features of all variants at a particular locus, such as phenotypes, clinical findings, allele frequencies, and so on" (1506); all contributors of nanoassertions would thus be considered co-authors. To offer dynamic affordances, the bioinformaticians extend the simple subject-predicate-object structure of the good old positivist nanopublication to "include supporting information such as the nature of the data source, experimental conditions, and other contextual or 'credibility' features that the authors consider essential evidence for the assertion" (1506). Here we see that entanglement and, particularly, *movement* are accommodated in the data used and information provided—including sequence as well as sequencing, and index as well as indexing—and in the apparatus; stable links to obscure(d) data sources are exchanged for a method that facilitates data

mining and includes situated information about experimental conditions and context. Given that the neoliberal university and world-historical relations of race lurk behind the scenes of any bioinformatics project, it is a matter of research ethics to make explicit the situated apparatus in which (patented) DNA sequences came into being, are sequenced, and move around the globe, including how and where the sequencing happens (cf. Harvey 2016; Jamison 2016). Yet there is much to learn from bioinformatics for consideration and application in the fields of philosophy, media studies, and cultural inquiry alike, as they are being practiced in the algorithmic condition and take networked ICTs and dynamic machine learning as their starting points.

Lessons for Being Creative with Tools

It was the sudden appearance in Google Books of a datum (E. L. Young) as embodied (Eva Louise Young)—an effect of the behind-the-scenes quantum attribution of the book *A Philosophy of Reality* to the latter female philosopher—that produced the rare phenomenon of "posthuman interpellation" (cf. van der Tuin 2014) that halted me in my studies and initiated the unintended research for this chapter.

The diffraction that occurred amid my research with/in the internet in October 2016 caused Eva Louise Young to have perhaps always already been part of the feminist posthumanist genealogy. Now I understand the twenty-first century philosophical apparatus as simply hinting at the type of content, the underlying structure, and some of the affordances of media theorist Wolfgang Ernst's "dynarchive," in which what is archived remains mobile, as with the storage medium itself. Whereas here the "archive" stands for indirect/sequential access and the "anarchive" for direct/random access, Ernst (2014) opens to a transversal approach that is neither oppressive nor automatically liberating. In a dynarchive, there is a computational effect; past-based predictions about the future are made, and they take effect both in real-time and for future past-based predictions about the future (machine learning). And alongside the computational effect, in the words of Chun, "new media's modes of repetition and transmission [...] open up gaps for a future beyond predictions based on the past" (Chun 2011, 2).

This "opening-up" is that to which I previously alluded as the Bergsonian temporal interval. Reading the work of French philosopher Henri Bergson from his monograph *Time and Free Will: An Essay on the Immediate Data of Consciousness* ([1889] 1913) into Chun's words is not unsubstantiated, because Chun argues that in order to grasp "software's dynamic porousness

[it] is often conceptually transformed into well-defined layers. Software's temporality, in other words, is converted in part to spatiality, process in time conceived in terms of a process in space" (Chun 2011, 3). Theorist Stephen Crocker, author of *Bergson and the Metaphysics of Media*, says that "Bergson tries to surpass the simple opposition of discrete and continuous forms of organization to understand the medium in which they participate, which he calls "the Past in General" (Crocker 2013, 12).

This "past in general" inserts creative multiplicity into past-based predictions about the future, as the prediction *immediately* feeds back into the past and starts affecting predictions unfolding in real time.[12] The philosophical apparatus of which the appearance of "*A Philosophy of Reality* by Eva Louise Young" was part is temporal given that everything with regard to big data happens, in Chun's terms, "*in media res/race*," just as we have seen in the preceding discussion on the entanglement of quantum attribution and bioinformatics in our media-technoscape.

Conclusion

I have now researched my encounter with Eva Louise Young and the ways in which human–computer interaction intervened in the feminist genealogy and canonization process of posthumanism for about five years. By now I have given a few lectures (among others, an inaugural lecture at Utrecht University) in which Young has been put forward as part of that genealogy. I have brought the logics of (un)findability back to its fundamentals, which means I have understood, described, and analyzed the active role of metadata in that algorithmically driven and manually supported process, in which hardly anything is stable. Google Books has, in a successful and inimitable way, brought together metadata from all kinds of databases to finally come to the assertion that it was *Eva Louise* Young who wrote *A Philosophy of Reality* in 1930. The way in which Google Books operates must be understood in an interdisciplinary manner: the nanopublications from the history of philosophy should be supplemented with insights into micro-attributions,

12 Here, close affinities with Mark Hansen's recent work present themselves. Cf. the doubleness of twenty-first century media: "*at one and the same time*, twenty-first-century media *broker human access* to a domain of sensibility that has remained largely invisible (though certainly not inoperative) until now, *and*, it *adds to* this domain of sensibility since every individual act of access is itself a new datum of sensation that will expand the world incrementally but in a way that intensifies worldly sensibility" (Hansen 2015, 6; original emphasis).

precise citations, and quantum attributions from bioinformatics around a question that is central to the field of media studies.

What fascinates me is how online encounters produce facts in a process that cannot be described exhaustively, both because of its speed and because the description itself of the encounter is part of, and data for, that process. After all, we no longer write without Wi-Fi or 4- or 5G, and a click on *A Philosophy of Reality* immediately feeds the algorithm, thus adding weight to Google Books' unique record; this record is dynamically sourced from library and other metadata that are both authoritative and often erroneous. How do we handle such doubling of data? How do we apply "doubled vision," both programmed and human, to use ICT differently for a critical intervention in "the cycles of continuous reciprocal causality" so that we do more than passively respond to the pressures of accelerating information flow of which we are part (Hayles 2012, 102)?[13] How do we become aware of the fact that we, as technology users, media consumers, scholars, educators, and students, can potentially *benefit* from "[leaving] the potentialities open and [...] suggest[ing] alternative and more complex architectures of knowledge" (Pasquinelli 2018, 256)?[14] What posthuman, critical, and creative data-subjectivities ask of us is to position ourselves as digitally literate, because there is no escape from computational media-technologies. We will have to specify and mobilize that literacy by critically unpacking the operational logics and socio-technical conditions involved and by creatively jumping upon the surprises they bring forth, just as I have done in this chapter for one specific case.

References

Acker, Amelia. 2021. "Metadata." In *Uncertain Archives: Critical Keywords for Big Data*, edited by Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel, 321–29. Cambridge, MA: MIT Press.

Barad, Karen. 2003. "Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter." *Signs: Journal of Women in Culture and Society* 28, no. 3: 801–31. https://doi.org/10.1086/345321.

13 I draw on a proposal by posthuman theorist of literature and new media N. Katherine Hayles (2012), and "doubled vision" was borrowed from the early feminist theorist Joan Kelly (1979).

14 As Pasquinelli quotes, after all: "With the temporality put to work by microprocessors, enormous quantities of data and problems can be processed in minuscule periods of time, in such a way that the new machinic subjectivities keep on jumping ahead of the challenges and stakes with which they are confronted" (Félix Guattari in Pasquinelli 2018, 256).

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.

Bergson, Henri. [1889] 1913. *Time and Free Will: An Essay on the Immediate Data of Consciousness*. Translated by F.L. Pogson. 3rd ed. London: George Allen.

Braidotti, Rosi and Maria Hlavajova, eds. 2018. *Posthuman Glossary*. London: Bloomsbury Academic.

Braidotti, Rosi, Emily Jones and Goda Klumbyte, eds. 2022. *More Posthuman Glossary*. London: Bloomsbury Academic.

Bühlmann, Vera, Felicity Colman, and Iris van der Tuin. 2017. "New Materialisms, Novel Mentalities, Quantum Literacy." *the minnesota review* 88, no. 1: 47–58. https://doi.org/10.1215/00265667-3787378.

Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge, MA: MIT Press.

Clough, Patricia Ticineto. 2018. *The User Unconscious: On Affect, Media, and Measure*. Minneapolis, MN: University of Minnesota Press.

Colman, Felicity, Vera Bühlmann, Aislinn O'Donnell, and Iris van der Tuin. 2018. *Ethics of Coding: A Report on the Algorithmic Condition* [*EoC*]. Brussels: European Commission. https://cordis.europa.eu/project/rcn/207025_en.html.

Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Crocker, Stephen. 2013. *Bergson and the Metaphysics of Media*. London: Palgrave Macmillan.

Darbellay, Frédéric, Zoe Moody, Ayuko Sedooka, and Gabriela Steffen. 2014. "Interdisciplinary Research Boosted by Serendipity." *Creativity Research Journal* 26, no. 1: 1–10. https://doi.org/10.1080/10400419.2014.873653.

De Beauvoir, Simone. [1949] 2010. *The Second Sex*. Translated by Constance Borde and Sheila Malovany-Chevallier. New York: Alfred A. Knopf.

Deleuze, Gilles, and Claire Parnet. [1977] 1987. *Dialogues*. Translated by Hugh Tomlinson and Barbara Habberjam. London: The Athlone Press.

Dolphijn, Rick, and Iris van der Tuin. 2012. *New Materialism: Interviews and Cartographies*. Ann Arbor, MI: Open Humanities Press.

Ernst, Wolfgang. 2014. "Between the Archive and the Anarchivable." *Mnemoscape* 1. https://www.mnemoscape.org/single-post/2014/09/04/Between-the-Archive-and-the-Anarchivable-by-Wolfgang-Ernst.

Erstad, Ola. 2010. "Educating the Digital Generation: Exploring Medial Literacy for the 21st Century." *Nordic Journal of Digital Literacy* 5, no. 1: 56–72. https://doi.org/10.18261/ISSN1891-943X-2015-Jubileumsnummer-07.

Gauthier, David. 2016. "On Commands and Executions: Tyrants, Spectres and Vagabonds." In *Executing Practices*, edited by Helen Pritchard, Eric Snodgrass, and Magda Tyżlik-Carver, 63–76. Brooklyn, NY: Autonomedia.

Hansen, Mark B. N. 2015. *Feed-Forward: On the Future of Twenty-First-Century Media.* Chicago: University of Chicago Press.

Haraway, Donna J. 1988. "Situated Knowledges: *The Science Question in Feminism* and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3: 575–99. https://doi.org/10.2307/3178066.

Harvey, Sandra. 2016. "The HeLa Bomb and the Science of Unveiling." *Catalyst: Feminism, Theory, and Technoscience* 2, no. 2. https://doi.org/10.28968/cftt.v2i2.28803.

Hayles, N. Katherine. 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.

Hayles, N. Katherine. 2017. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: University of Chicago Press.

James, Ryan, and Andrew Weiss. 2012. "An Assessment of Google Books' Metadata." *Journal of Library Metadata* 12, no. 1: 15–22. https://doi.org/10.1080/19386389.2012.652566.

Jamison, Matthew. 2016. "Becoming-HeLa: Towards an Ethico-Onto-Epistemology of Grief." Unpublished MA thesis, Utrecht University.

Johannesen, Monica, Tonje Hilde Giæver, and Leikny Øgrim. 2014. "Notion in Motion: Teachers' Digital Competence." *Nordic Journal of Digital Literacy* 9, no. 4: 300–12. https://doi.org/10.18261/ISSN1891-943X-2014-04-05.

Kelly, Joan. 1979. "The Doubled Vision of Feminist Theory: A Postscript to the 'Women and Power' Conference." *Feminist Studies* 5, no. 1: 216–27. https://doi.org/10.2307/3177556.

King, Katie. 1994. *Theory in its Feminist Travels: Conversations in U.S. Women's Movements*. Bloomington, IN: Indiana University Press.

Koolen, Marijn, Jasmijn Van Gorp, and Jacco van Ossenbruggen. 2019. "Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice." *Digital Scholarship in the Humanities* 34, no. 2: 368–85. https://doi.org/10.1093/llc/fqy048.

Lyotard, Jean-François. [1979] 1984. *The Postmodern Condition: A Report on Knowledge*. Translated by Geoff Bennington and Brian Massumi. Manchester: Manchester University Press.

Pasquinelli, Matteo. 2018. "Metadata Society." In *The Posthuman Glossary*, edited by Rosi Braidotti en Maria Hlavajova, 253–56. London: Bloomsbury Academic.

Patrinos, George P., David N. Cooper, Erik van Mulligen, Vassiliki Gkantouna, Giannis Tzimas, Zuotian Tatum, Erik Schultes, Marco Roos, and Barend Mons. 2012. "Microattribution and Nanopublication as Means to Incentivize the Placement of Human Genome Variation Data into the Public Domain." *Human Mutation* 33, no. 11: 1503–12. https://doi.org/10.1002/humu.22144.

Serres, Michel. [2012] 2015. *Thumbelina: The Culture and Technology of Millennials*. Translated by Daniel W. Smith. London: Rowman and Littlefield International.

Thylstrup, Nanna Bonde. 2018. *The Politics of Mass Digitization*. Cambridge, MA: MIT Press.

van der Tuin, Iris. 2014. "Diffraction as a Methodology for Feminist Onto-Epistemology: On Encountering Chantal Chawaf and Posthuman Interpellation." *Parallax* 20, no. 3: 231–44. https://doi.org/10.1080/13534645.2014.927631.

van der Tuin, Iris. 2018. "Vindingrijk: Kritiek en creativiteit in cultuurwetenschappelijk onderzoek" [Resourceful: Critique and Creativity in Cultural Inquiry]. Inaugural Lecture, Utrecht University. http://issuu.com/humanitiesuu/docs/inaugural-lecture-iris-van-der-tuin?e=11895952/66165689.

van der Tuin, Iris. Forthcoming. "The Triple Mechanics of Susanne K. Langer's Card-Index System: Horizontal, Vertical, and Transversal Aspects." In *The Bloomsbury Handbook of Susanne K. Langer*, edited by Lona Gaikis. London: Bloomsbury Academic.

van Dijck, José. 2010. "Search Engines and the Production of Academic Knowledge." *International Journal of Cultural Studies* 13, no. 6: 574–92. https://doi.org/10.1177/1367877910376582

van Es, Karin, Mirko Tobias Schäfer, and Maranke Wieringa. 2021. "Tool Criticism and the Computational Turn: A 'Methodological Moment' in Media and Communication Studies." *M&K. Medien und Kommunikationswissenschaft* 64, no. 1: 46–64. https://doi.org/10.5771/1615-634X-2021-1.

Verhoeff, Nanna, and Iris van der Tuin. 2020. "Interfaces (for) Diffracting Technobodies: A Science-Humanities-Design Perspective for an Algorithmic Somatechnics." *Somatechnics: Journal of Bodies – Technologies – Power* 10, no. 3: 374–96. https://doi.org/10.3366/soma.2020.0328.

Witten, Ian H., Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. Amsterdam: Elsevier.

Young, E. L. 1930. *A Philosophy of Reality*. Manchester: Manchester University Press.

About the Author

Iris van der Tuin is Professor of Theory of Cultural Inquiry in the Department of Philosophy and Religious Studies at Utrecht University, where she is also university-wide Dean for Interdisciplinary Education. Her research is part of the group "Transmission in Motion" of the Institute for Cultural Inquiry.
> i.vandertuin@uu.nl

11. Interstitial Data

Tracing Metadata in Archival Search Systems

Jasmijn Van Gorp

Abstract
Metadata do not merely give explicit information about records in the archive but can also be considered a source of information about the (historical) context in which they are created. This chapter combines the insights of critical data studies and archival studies to formulate a hands-on approach to tracing metadata in archival search systems. The approach, which builds further on Loukissas's local reading strategies, consists of two distinct phases: an exploration phase to trace and select and an analysis phase to trace and compare. The author concludes that a lot of data necessary to understanding metadata in search systems is hidden—different forms of what can be considered "interstitial data."

Keywords: archive, metadata, television, search system, tracing, local reading

Metadata, or data about data, provide the essential context of a record in an archival collection (Kitchin 2014, 4). Similarly, Pomerantz (2015) refers to metadata as a map that represents the complexity of an object in a simpler form. Metadata can be automatically generated and/or created by humans, such as archivists or users (cf. Noordegraaf 2015). Without metadata, it would be virtually impossible to find a record in large-scale archival collections. First, they support the identification of any given record "at a glance." Second, metadata are based on a logic and classification system that provides information of a given record within a collection. Metadata also make records retrievable through descriptions of the records.

However, metadata do not merely give explicit information about records. Within critical data studies and archive studies, it is common to consider metadata also a source of information about the (historical) context in which

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH11

they are created. In their discussion of new media art archives, Rinehart and Ippolito (2014) show how metadata can shape the historical record. Each metadata standard "frames": it provides a point of view and determines "what we choose to remember and what we choose to forget" (Rinehart and Ippolito 2014, 60–62). Loukissas (2019) elaborates on how (meta)data can also serve as cultural markers of past collection practices. Data, he contends, can be "locally inscribed." He proposes an active understanding of context and considers it to be assembled through a combination of social, technological, and spatial practices. Within archive studies, a similar perspective has been advocated by, among others, Eric Ketelaar (2001), who shows how archives reveal the context in which they are created through "tacit narratives." With "tacit narratives," Ketelaar means all practices and technologies that leave traces. Traces can be found not only in metadata that is available but also, or even more so, in what is not available, since "archiving also entails what should and what should not be kept" (Ketelaar 2001, 136).

It is this intriguing interplay between availability and unavailability that I will further investigate in this chapter. Using a case study that enforces what is at the margins, the "forgettable" broadcast genre of interstitials, I investigate the different traces of metadata.[1] To this end, I test and extend a proposed reading by Loukissas (2019, 62–69) by discerning two phases of "tracing metadata" as comprising a hands-on approach for studying metadata in archival search systems. While Loukissas focuses on all kinds of collections data in data infrastructures, I focus specifically on the function of metadata in the scholarly practice of searching in archival search systems. The two goals of this chapter are to investigate the extent to which traces of archival practices can be revealed through an analysis of metadata and, at a practical level, to help train students in critical reflection on the role of metadata when making use of archival search systems. This chapter concludes that talking about "interstitial data" is fruitful for encapsulating metadata's various forms and related practices.

Trace and Select

Most archives have online search systems that are based on semantic searches with words matching the metadata of the records. The user defines

1 I borrow the concept of "forgettable television" from Polan (2013) who conceptualized it as "programming designed to be forgotten at virtually the very moment of its original viewing—as virtually most TV was" (347).

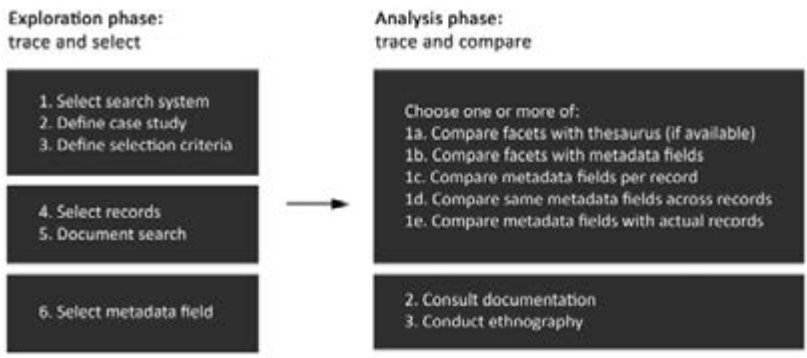


Fig. 10. The two phases of tracing metadata in archival search systems (image created by the author).

a query or conducts a "keyword search" (cf. Althaus and Phalen 2010), after which the system retrieves matching records in the search system based on their metadata and ranks the records in a result list. The metadata, as present in these search systems, are the object of my analysis. For the analysis of metadata, I outline a tracing method that consists of two distinct phases: an exploration (or "preparatory") phase and an analysis phase. In the first phase, I explore a search system using keyword searches and trace my journey in the search system. The goal of this exploration phase is to select records for the analysis phase and to come into grips with the search system and with the relevant metadata fields.

The first step of the exploration phase is to select a search system. I selected a search system that I know very well, as I co-developed it: the CLARIAH Media Suite. The Media Suite is a new research infrastructure for audiovisual data in the Netherlands, aimed specifically at researchers (Melgar et al. 2018). It provides access to an extensive collection of the Netherlands Institute of Sound and Vision, containing almost two million television and radio broadcasts from public broadcasters. A second step is to define a case study with which to build a corpus. For my case study, I focus on what is seen as a "forgotten part" of television history: interstitials. Interstitials are the broadcasts that air in between the actual programs, such as trailers, commercials, and idents (Ellis 2012). As Johnson (2012) explains, archival access to interstitials is limited because they are often not saved, except in the case of full-day recordings, which are comparatively rare.

Once a search system and a case study are defined, an important step is defining a selection strategy for records and making this explicit (cf.

Creeber 2004). To achieve a balanced selection, I applied four search criteria. I wanted to select (1) various types of interstitials, (2) across a broad historical spectrum, which preferably have (3) access to the video and (4) are found by applying various search strategies, i.e., using different queries, filters, and settings for each record. I ended up with a selection of nine records: one commercial for STER for Duyvis tiger nuts; one introductory announcement of an old news program for the hearing impaired; one film announcement for a Saturday night film as part of the series *Jiskefet*; five television announcements of the fairy tale *Klaas Vaak*, with one original clip and four rebroadcasts of the same clip within a children's television program; and one full-day of recording of a Sunday on the channel NPO 1.

Tracing the journey in this exploration phase can be understood in the reflective sense or as "critical tracing" (van der Tuin and Verhoeff 2021, 197): walking back your own footsteps and reconsidering what you have found and why, by means of a critical reflection on your search and selection process. Systematic documentation or "logging" in research journals should be a standard phase of any research project (Borg 2001). Therefore, it is important to keep track of this exploratory phase by writing down one's keywords, filters, and settings. Many search systems have automated tools for this, such as history tracks and bookmarks. The Media Suite has bookmarks and URLs that reveal the tools used and the search terms.[2] In the method of tracing metadata, documenting helps the user not only to grasp semantic searches in search systems but also to find relevant metadata fields to analyze in the second phase. I consulted my research journal and noticed four metadata fields that were used extensively, triggering some questions: the descriptive metadata fields labeled "genre," "title," and "description" and the administrative metadata field labeled "date."[3] Based on my research journal, I also reconstructed the search paths for each item to have it as reference for my metadata analysis (see box 1 for an example).

2 In item 1, for instance, it is "searchTerm": "\"aankondiging\"": https://mediasuite.clariah.nl/tool/resource-viewer?id=2101608040033514831andcid=daan-catalogue-aggrandst=aankondiging#. The URL not only contains the search term (st=aankondiging), but also the Media Suite Item ID (2101608040033514831), the collection (daan-catalogue-aggr), and the tool used (resource-viewer). The URL helps both to trace back your steps and to analyze them.

3 While descriptive metadata describe the content and provide context (title, author, publisher, subject, description, etc.), administrative metadata inform us when and how the dataset was created, on technical aspects, and who owns and can use the data. Pomerantz (2015) also identifies three other types: structural metadata, preservation metadata, and use metadata.

CLARIAH Media Suite > Tool "Search" > Clear search > Add new Facet Broadcast type > select facet broadcast type: "Collection band" (<verzamelband) > 48 items in result list > select facet Genre (series): reclame > 6 items in result list of which 5 items 'commercials database' and 1 item 'ster reclame' > select Item *Ster reclame (Ster commercial)*

Main criteria: type (commercial), search strategy (add facet broadcast type)

URL selected item: https://mediasuite.clariah.nl/tool/resource-viewer?id=210160804003394513 1andcid=daan-catalogue-aggr

Box 1. Example of a search path, criteria and URL as taken from my research journal.

Trace and Compare

In the second phase of the tracing method, the analysis phase, I looked at traces in the material sense. These concern materializations of elements that might be invisible or "traces that mark what has been 'there'" (van der Tuin and Verhoeff 2021, 196). Both Loukissas (2019) and Ketelaar (2001) refer to all kinds of processes that leave traces. First, the data handlers, the archivists, and their collection practices leave traces. Second and relatedly, traces are also left by technological processes that transfer data from the physical sources (paper, reel, etc.) to the interface, such as digitization and indexing. I also add here the technical practices of standardization and normalization that are characteristic of data infrastructures.[4] While "standardization" refers to setting up one standard to which all other practices have to adhere, "normalization" means having data conform to a certain format so that it becomes machine legible.[5] As these definitions reveal, both practices are difficult to discern for humanities scholars with no computer science background (like myself) but are in essence aimed at increasing technical interoperability when collections of different archives are aggregated. Third, and often forgotten, is the user (you) who searches the search system. The user also leaves traces, which loop back to tracing in the exploration phase to show how the user explored the archive in the first phase.

4 Rinehart and Ippolito (2014, 60) explain that metadata standards indeed contain traces of its historical context, but also discuss the benefits of these standards for data infrastructures: platform independence, portability, accessibility, extensibility, and longevity.

5 I compiled my own working definition of "normalization" by reading Hoekstra et al. (2019) and Loukissas (2019) and of "standardization" by reading Van Zundert (2013).

For my analysis phase, I follow one of the local reading strategies proposed by Loukissas (2019, 62–69), who points to six features that shape data.[6] A first element he discerns is *classification*. In classification systems, the world is divided into segments. These segments are used to help in administrative or knowledge production (Bowker and Star 1999). In my case study, I translate local classification to the object at hand, the metadata of television broadcasts, and view genre as a classification system. Second and third, Loukissas looks at *schemata*, the ways of recording metadata, and *constraints*, the conditions that apply for inscribing data such as technical limitations. The last feature type he offers is *errors*. Data are often cleaned or filtered, but ideally the mistakes are kept, as they are relevant as traces (Loukissas 2019, 67). OCR mistakes are commonly known, but the question for my research would be whether I can discern mistakes in audiovisual data collections. Another element, which I consider to be key, is *absences*. As archives define what is kept, shown, or put on display, they also leave out data and metadata. With the last element, *rituals*, Loukissas refers to cultural practices that can be seen, in my case, as the archivists' practices.

These material traces can be revealed by applying "a comparative lens," a useful framework again provided by Loukissas (2019). Comparison is at the heart of most research, Berger (2016, 21) contends. While comparison is usually meant in either a diachronic or synchronic sense, I compare items within the same search system. Hoekstra and Koolen (2019) refer in this context to datascopes, different representations of the same item within archives and infrastructures. My aim, then, is to look for patterns across the six elements as I compare different representations of records in metadata.

To help identify and understand the six elements, I use the documentation of the Netherlands Institute for Sound and Vision: a metadata handbook as used by archivists, the metadata translations for the Media Suite, and the Media Suite's user manual. In addition, I send emails to archivists and developers of the Media Suite when I have questions about some of the metadata fields. I choose to stay as close as possible to students' situations and investigate the limitations of doing this kind of text-driven research with only a limited amount of help from practitioners.

6 Two decades earlier, Ketelaar (2001) discusses similar elements but in more general terms. He also discusses absences, classification, and rituals/practices but does not specify constraints, schemata, and errors.

Metadata Field Genre

In my first comparison, I compared the facets with the controlled vocabulary list and with the key words of the selected items. I focused on the main category for television collections, which is the television genre. Genre is characteristic for both film and television studies and refers—to use a simple definition—to a combination of style and content elements.[7]

First, I wanted to generate an overview of genres as shown in the facets, the filter mechanisms to narrow down the search list, and an aggregation of the key words of all records matching the query. When searching on the full collection with an empty query, the facets show only three types: promos (315), promotions (162), and commercials (24,589). The more general category of interstitials is missing, as well as announcements, idents, logos, trailers, and teasers. This implies that users are not steered toward most interstitial types when using facets.

A similar picture is painted by the thesaurus *Common Thesaurus for Audiovisual Archives* (GTAA), which provides the controlled vocabulary for the audiovisual archives in the Netherlands. Archivists use a thesaurus for disambiguation and for tagging records with key words that are used to optimize retrieval. The GTAA, which is part of the Media Suite's resource viewer, mentions three different terms for "commercials": the Dutch term with capitalization, the Dutch term without capitalization, and the English term as it is also used colloquially in Dutch. I find similar synonyms for the English word "promos." In other words, the GTAA—in its core function—does not normalize the data: it has multiple categories for the same genre. And while "leaders" and "trailers" are not categories in the genre facet, they are mentioned in the GTAA.

The GTAA also uncovers rituals, as it provides for short definitions of the different terms as seen by archivists that are given especially for disambiguation purposes. The definitions show that the common denominator for commercials, promos, and leaders are "(very) short" and "small," by means of the diminutive Dutch suffix "pje" (e.g., filmpje). Promotions are defined as the opposite of commercials, which indicates informal schemata to define interstitials against the "main" genre, commercials. The definitions in the GTAA point to the importance of commercials and promotions for the archive, which is enforced by the fact that the GTAA contains all variations of commercials and promos but not of leaders and trailers. Lacking in facets and GTAA are interstitials and idents. This may be due to the use

7 For advanced theories on television genre, I refer to Mittell (2004).

of English language and the lack of an apt Dutch translation for interstitials and idents, expanding the "semantic gap" to a language gap. Absence, then, is also related to language constraints.

If I then look at the records of my nine items, I notice that only two items have a genre that refers to the interstitial: the commercial is tagged with "Reclame," and the original leader for the television broadcast is classified as a "Promo" (and not as leader). The other six are classified with the genre of the broadcasts that they encompass, such as "comedy," "entertainment," or "news." The genre metadata field shows that interstitials are chiefly not classified as interstitials, which might be explained by the constraint that each broadcast can have only three genre tags, a requirement that may push interstitial genres into the absent place. In addition, it also shows that the longer an interstitial's duration is, the more likely it is to be considered a broadcast type that is worth mentioning and archiving.

Leaders:	/
Reclame:	/
commercials:	*radio and television commercials*
reclame:	*advertisements, mostly (ultra)short production, of which the main goal is to sell a specific product or service*
Trailers:	*short promo videos [filmpjes] for television programs and feature films, not for trailers as vehicles*
Promo:	*mostly (ultra)short production, of which the main goal is to get attention for a radio or television program*
Promos:	*promotion spots for radio and television programs*
Promotion:	*production that is aimed at convincing the viewer or listener, without a commercial intent*

Box 2: GTAA entries with definitions (if available) in italics. Translated from Dutch by the author except for "reclame," which is the Dutch word for commercial.

Metadata Field Title

Titles are the first relevant check of records for users of search systems. The full metadata in the Media Suite shows 27 different title fields, which offers me a plethora of options for comparing title fields with each other.

The interstitial type is mentioned in title fields of only one out of nine records. Not surprisingly, it is the commercial that also has "commercial" in

its title. In the generic title aimed at media professionals, the interstitial type is even mentioned three times: "STER COMMERCIAL; Ster commercial (1993) Duyvis: Tiger nuts – Gerard Cox in tiger outfit is advertising Duyvis tiger nuts." The genre is mentioned on different occasions in the title and the first time in all capitals so that it could be easily detected. It also shows a technical ritual of the title fields by the Media Suite title field: all title fields are squeezed into this generic metadata field for title. It is only through comparison with the other title fields that I realized that this main title is a composite.

The schemata and annotation rituals differ per title field. Since the archive also has a museum, some items also have a "museum" title field. The museum title of the commercial is "Gerard Cox in tiger costume is advertising Duyvis tiger nuts," starting with the name of a celebrity rather than the genre (commercial) or the section name (STER), which is mentioned in the general title. The interstitial type "commercial" gets a lot of attention in the program's title when aimed at media professionals, but it is played down when aimed at the museum audience, which is engaged by using the celebrity's name. It shows an implicit assumption about the kinds of information museum audiences are looking for.

While title is an important metadata field, it does not provide many traces in the case of the other eight records. This absence of interstitial types in the title field might be indicative of the ritual that interstitials should not be put "on display" in the title field. In its title, the announcement of *Klaas Vaak* is just named "clip," a generic term that could equally be referring to a regular clip. The title of the weekly recording, which is a full overview of all programs in the order of broadcast, only mentions the programs and leaves out all the interstitials. In the oldest item, the news broadcast, the title field does not mention the leaders, but the tape fields do. The reels contain the news items and have separate titles, of which two are named "leader" followed by a code: FHD. An archivist told me that FHD means it is digitized from film. The video's materiality and the way it is saved are shown in the reel title, a hidden datum resulting from a ritual. Absence can be taken quite literally, as only six out of twelve reels play out. The titles of the tapes, even when the play-out or video is missing or broken, show what was there and make the invisible visible again.

Metadata Field Description

Like the metadata field title, the "abstract" or description of a record is frequently used to understand the content of a document (Althaus and

Phalen 2010). In the case of the Media Suite, I compare the metadata field description across the nine items and the description of each item with the actual record, the video content.

The first question is whether the records are described (and thus classified) as interstitials. In line with my previous findings, the commercial is described extensively. While the title of the *Klaas Vaak* clip labels it with the generic label "clip," the description starts with "Promoclip," making it easier to identify it as an interstitial. None of the description fields of the other items mentions the interstitial type. The documentation manual shows that it is the ritual not to mention "leaders" and "credits" in descriptions.[8] Other metadata fields show information on the provenance of descriptions, such as the names of the annotators as well as the source of the description. In five records, the annotation field shows that the description was not made by watching the video but by consulting "information," such as a broadcast magazine. In the case of the weekly recording, the broadcast times are set crudely to 19:15–21:15, followed by the name of the reel ('\nDs783/V7483'). The ritual leaves out all the interstitials, as the time schedule is taken from a broadcast magazine and is not based on the video itself.

It is particularly instructive to compare descriptions of duplicates that Loukissas (2019, 60) considers to be "key to learning about heterogeneity of data infrastructures." The *Klaas Vaak* items are very clear examples of duplicates. The four times in which the *Klaas Vaak* announcement was mentioned in the description field were standardized: the same sentence appeared in all four descriptions. In addition, in all four occurrences in the video, the same voice-over announces the clip with the exact same phrase. This shows that standardization is related to schemata—that is, a standard—and most likely also relates to automatically generated metadata. Interestingly, the announcement of the film is also mentioned in the description and follows the same order of words as the *Klaas Vaak* items. However, this time, it coincides with an error. In the video, a voice-over announces a film in a 1970s timbre, which is ambiguous, as the program is broadcast in 2006. When I recognized a famous, contemporary Dutch actor in the video, I realized this was a parody of an announcement. This shows that the description obscured the fact that it is a parody: the archivist did not use the word "parody" but described the item in the very same manner that a "real" announcement would be described. Therefore, the schemata of archival descriptions have transformed the fictional announcement into

8 It specifies that if credits are interesting for re-use, these may be mentioned in the annotation field.

a real announcement, which can be considered an error but equally an indicator for the ritual of description practices. It shows that Loukissas's elements of constraints, errors, rituals, and schemata can also reinforce each other.

Absences are an element that comes explicitly to the fore in the descriptions. In the case of the weekly recording, a two-hour recording of all programs on a Sunday, I tried to find the duplicates and the separate broadcasts, but to no avail. This full-week recording collection that is saved for media history gives insight into missing programs, which is very valuable for media history. In addition, the video shows that the recording contains twelve idents, five announcements, and twelve commercials. All twenty-nine interstitials were left out of the description. However, they are mentioned in the metadata field summary, the description field to provide information on the series to which a broadcast belongs: "The programs are interrupted by commercial breaks (STER), promos, trailers and *Postbus51*-commercials." Again, commercials are mentioned first, and leaders and idents are not mentioned at all.

Metadata Field Date

While I did not expect there to be interesting traces in the pure administrative metadata, I found the date field in the CLARIAH Media Suite to be a treasure trove of traces. In total, there are 120 different date fields. I wanted to check whether the date fields might enable me to reconstruct the entire trajectory of physical objects in the archive to items in the catalog and the Media Suite, which Ketelaar (2001, 138) calls the "semantic genealogy of the record." To this end, I lined up a selection of date fields per record. I discuss the fields in depth for one example, namely, the commercial (box 3).

The date metadata fields are very technical and almost impossible to understand without the direct help of practitioners. At first, I surmised that the broadcast had been put on a VHS tape and digibeta in 1982, but then the editors of this book asked me whether that is possible. I enquired with the archivists of Sound and Vision, who were also inclined to think that this is an error or, more precisely, a randomly chosen date, as the commercial was put on a "collection band" (*verzamelband*). At that point, I decided to show my other interpretations of the date fields of the commercial to the archivists of Sound and Vision. I came to realize that I challenged myself to pick the commercial as example, as it was put on a collection band, and it stems from the early 1990s. In general, the more recent a broadcast is, the

fewer data transformations are inflicted on its metadata and the easier it is to reconstruct the path.

The commercial was broadcast in 1993. It has a specific publication date (January 1, 1993), which is a trace of a constraint that this date field cannot be left blank and should be set on the first of January of the year. In the metadata field annotation, it is mentioned that the exact date is unknown, confirming the constraint. All dates are incorporated in the field "sort date," which is algorithmically calculated as the most likely publication date when fields are left empty. There are various fields called "Date created." One of these was close to Museum Genre, so I interpreted it as the date the metadata field was created specifically for the museum, which was confirmed by the archivists as the most likely interpretation.

The dates in the metadata also show when the digital file was entered into the system by means of the field "Asset Date." "Asset Date" shows that the commercial's metadata record was ingested by the system in 2008, at about 9:30 in the evening. All other records were ingested in 2008 or 2009. The "Last Updated" date refers to all kinds of updates but most likely to the conversion of the old iMMix catalog to the DAAN catalog; it ranges between 2017 and 2021. The first moment of conversion is most likely "Date created [logtrack]" in 2017. "Asset items created" and "Date last updated" have the same annotation system, namely, the unix timestamp dates that are in schemata, which is incomprehensible for humans. An archivist pointed me to a converter that translates this into a humanly readable date. In this sense, it is an invisible date that only becomes visible through an archivist's expert knowledge.[9]

Carrier date of the vhs:	1982—01—01
Carrier date of the digibeta:	1982—01—01
Broadcast date:	1993—01—01
Sort date:	1993—01—01
Date created [museum]:	116216280000 *(2006—10—30)*
Asset item date:	2008—04—11T21:30:09Z *(2008—04—11 at 09:30:09pm)*
Date created [asset item]:	1207949409000 *(2008—04—11)*
Date last updated [logtrack]:	1488757329517 *(06—03—2017 at 00:42:09am)*
Date last updated [conversion]:	1606473164285 *(2020—11—27)*

Box 3: Date metadata fields for the selected commercial (converted dates in Italics).

9 Converter: https://www.epochconverter.com/

Conclusion

In this chapter, I have proposed a metadata tracing method that expands on one of Loukissas's local reading methods. The genres of the interstitial are quite invisible in the facets used in search systems to come to selection and therefore proved themselves to constitute an interesting case study for investigating the interplay between available and unavailable data. First and foremost, it is thanks to the efforts of the archive that most of my interstitials were saved in the first place—most of the records were saved on collection bands. These records, then, were found because I articulated more advanced search strategies. This implies that it is important to spend quite some time on the exploration phase but also that my fifth selection criterion—using different search strategies—is a good one to follow, especially when one's chosen case study is a bit off the beaten path. In terms of (training) critical reflection on search systems, I realized how much my search and my selection are related to oft-hidden decisions that I tried to unravel.

Yet I came to realize that the "interstitiality" of interstitials reaches beyond the case study itself and points at different forms of what can be considered "interstitial data," data that is hidden at first glance. First, the concept refers to all data that comes to the fore through comparison. The videos shed light on rituals while revealing what was missing in the archive. The duplicates of records were the most illuminating records for the investigation of rituals, while they are also very difficult to locate in large collections. That is why it is important to spend considerable time on the exploration phase, as it allows one to find striking and interesting examples for the analysis phase. It is in the combination of the different representations that even more data and tacit narratives arise.

Second, interstitial data also points at the data that are considered less important for re-use and which therefore disappear from the displays of search systems. The commercial is "on display" in all possible ways: in the facets, the GTAA, the title, and the description. This shows how the archive, at least when it comes to interstitials, is built up for commercials. Commercials are clearly regarded as the most reusable category of interstitials. The "title" field can be considered the "chief" metadata field and is left almost empty in case of interstitials. The GTAA sheds light on the rituals of archivists by providing other types of spellings and definitions. The "description" field is the place where all Loukissas's elements come together and reinforce each other. Tacit narratives were also found in an unexpected metadata field, the date field. This field is

key to the investigation of the provenance of an item and its tacit narratives, also in cases in which the title and description fields are missing or incomplete. My analysis shows that classification, absences, and rituals run through all metadata fields together with constraints and schemata. The six elements helped me to get a better grasp of the mechanisms of local data. Probably due to the limited number of records I analyzed (only nine), I could barely identify errors.

My small-scale analysis enabled me to see some hidden data, but I also came to realize that some invisible data remains invisible. The older the record in the archive, the more difficult it becomes to gather all the information necessary to understand the (different transformations of) metadata. It also costs the precious time of multiple practitioners to unravel past practices. Knowledge about archiving practices has been increasingly preserved in documentation, but this documentation is inherently partial. Interstitial data therefore also refers to invisible data that remains invisible. It is also related to my own standpoint or skills, as a television scholar who knows the search system very well but is neither a trained information scientist nor a trained computer scientist—I was not able to look under the hood of the search system myself.

This raises a question, which is particularly relevant to students: To what extent it is possible to conduct this research without the help of practitioners? I could quite easily understand the title fields, the GTAA, and the description fields, as these are lengthy texts that can be analyzed through comparison. The description fields are an especially interesting metadata field. However, thesauri such as the GTAA are often not available within search systems, while it really helped me to understand the rituals. My advice is to consult the documentation of search systems and see whether they made use of e.g., Europeana or Unesco thesauri. The field that is impossible to understand without help of practitioners is the date field. It was the date field that triggered me the most, and my next plan is therefore to dig into date fields together with practitioners and data scientists to investigate whether we can come a step closer to reconstructing the genealogy from inception and broadcast, through digitization and ingestion, to searchable entry in the Media Suite. This research shows that there is still so much more to investigate for the sake of bringing to the fore historical knowledge about the collections and their operating practices. As Loukissas writes, local reading of obscure data "stimulates curiosity" (67), and that is exactly what it did to me.

Fig. 11. Follow this QR code to the playlist "Finding Interstitials in the Television Archive," curated by the author.

Acknowledgements

The author would like to thank the archivists and developers of Netherlands Institute for Sound and Vision for their help in understanding metadata and the editors for their input and feedback. This research has been made possible by the NWO infrastructure project CLARIAH-PLUS and the CLICK-NL project Re-Frame.

References

Althaus, Scott L., Jill A. Edy, and Patricia F. Phalen. 2002. "Using the Vanderbilt Television Abstracts to Track Broadcast News Content: Possibilities and Pitfalls." *Journal of Broadcasting & Electronic Media* 46, no. 3: 473–92. https://doi.org/10.1207/s15506878jobem4603_9.

Berger, Arthur A. 2016. *Media and Communication Research Methods: An Introduction to Qualitative and Quantitative Approaches*. 4th edition. Los Angeles: SAGE Publications.

Borg, Simon. 2001. "The Research Journal: A Tool for Promoting and Understanding Researcher Development." *Language Teaching Research* 5, no. 2: 156–77.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Inside Technology. Cambridge, MA: MIT Press.

Creeber, Glen. 2004. "Introduction." In *Fifty Key Television Programmes*, xiii–xvi. London: Arnold.

de Leeuw, Sonja. 2012. "European Television History Online: History and Challenges." *VIEW Journal of European Television History and Culture* 1, no. 1: 3–11. https://doi.org/10.18146/2213-0969.2012.jethc002.

Ellis, John. 2011. "Interstitials: How the 'Bits in Between' Define the Programmes." In *Ephemeral Media: Transitory Screen Culture from Television to YouTube,* edited by Paul Grainge, 59–69. London: Palgrave Macmillan.

Hoekstra, Rik, and Marijn Koolen. 2019. "Data Scopes for Digital History Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2: 79–94. https://doi.org/10.1080/01615440.2018.1484676.

Johnson, Catherine. 2012. *Branding Television.* London: Routledge.

Ketelaar, Eric. 2001. "Tacit Narratives: The Meanings of Archives." *Archival Science* 1, no. 2 (June): 131–41. https://doi.org/10.1007/BF02435644.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications.

Loukissas, Yanni Alexander. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society.* Cambridge, MA: MIT Press.

Melgar Estrada, Liliana, Roeland Ordelman, Jasmijn Van Gorp, and Julia Noordegraaf. 2018. "Media Suite: Unlocking Audiovisual Archives for Mixed Media Scholarly Research." *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8–10 October 2018.* http://www.ep.liu.se/ecp/159/014/ecp18159014.pdf

Mittel, Jason. 2004. *Genre and Television: From Cop Shows to Cartoons in American Culture*. New York: Routledge.

Noordegraaf, Julia. 2015. "Crowdsourcing Television's Past: The State of Knowledge in Digital Archives." *Tijdschrift voor Mediageschiedenis* 14, no. 2: 108–20. http://doi.org/10.18146/tmg.139.

Polan, Dana. 2013. "It's Fun to Eat: Forgotten Television." In *How to Watch Television*, edited by Ethan Thompson and Jason Mittel, 347–54. New York: NYU Press.

Pomerantz, Jeffrey. 2015. *Metadata.* Cambridge, MA: MIT Press.

Rinehart, Richard, and Jon Ippolito. 2014. *Re-collection: Art, New media, and Social Memory.* Cambridge, MA: MIT Press.

van der Tuin, Iris, and Nanna Verhoeff. 2021. *Critical Concepts for the Creative Humanities.* Lanham, MD: Rowman and Littlefield.

van Zundert, Joris. 2012. "If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities." *Historical Social Research / Historische Sozialforschung* 37, no.3: 165–86. https://doi.org/10.12759/hsr.37.2012.3.165-186.

About the Author

Jasmijn Van Gorp is Assistant Professor in the Department of Media and Culture Studies at Utrecht University. Her research is situated at the intersection of audiovisual data and the digital humanities. Since 2019, she has co-led the "Audiovisual Data and Media Studies" work package of the Dutch digital research infrastructure CLARIAH, which develops the Media Suite.
> j.vangorp@uu.nl

12. Data and Algorithms in Transition

A Diachronic Affordance Analysis Perspective

Stefan Werning

Abstract
The chapter investigates how ongoing feature changes in contemporary software reframe our understanding of data and algorithms, interprets these changes as shifting "rules of play," and introduces the notion of diachronic affordance analysis to conceptualize their rhetorical implications over time. First, it elaborates on the perceived "gameness" of software applications like online social networks, which often intensifies over time through continuous datafication via concepts like procedural rhetoric and the "implied player." Second, as most contemporary software is defined by constant "tweaking," it suggests considering the rhetorical implications of software change itself rather than focusing on the assemblage of features at any contingent moment. Finally, to demonstrate the approach, important developments between 2017 and 2020 in the mobile app *Samsung Health* serve as a case study.

Keywords: software affordances, diachronic affordance analysis, procedural rhetoric, implied player/user

As software applications evolve over time, they usually encourage users to turn more and more aspects of their professional and everyday lives into metrics and, often, to increase these metrics like high scores in a game. For example, author Nicholas Carr pointedly describes the experience of writing for the then-new *Kindle Unlimited* distribution program as a "zero-sum game that pits writer against writer."[1] Similarly, many non-commercial social media users intuitively experience and approach their online presence as

1 See http://www.roughtype.com/?p=6290.

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH12

a quasi-game and develop routines to evaluate and optimize their "performance" accordingly. But how can we make this perceived "game-ness" more explicit to understand better the cultural logic of software development?

In this chapter, I investigate how constant feature changes in contemporary software, particularly mobile applications, continually reframe our understanding of data and algorithms through the lens of game studies.[2] The argument is subdivided in two phases. First, I propose to conceptualize the ongoing transformation of software applications in terms of "rules of play," drawing on Brock and Shepherd's use of procedural enthymemes—traditionally applied to study meaning making in games—to explain the work of algorithms (2016). Second, I explain how these rules are continuously changing, as well as how these changes, more than any individual configuration, can inform our understanding of software. To interpret the rhetorical implications of affordance adjustments over time, the notion of *diachronic affordance analysis* is introduced. That is, the chapter argues that we need to be mindful of how software evolves—e.g., by quantifying reading flow and paying authors per page like the *Kindle Unlimited* mentioned above—to understand how users such as Nicholas Carr, who gradually internalize the tool over time, make sense of it. Both the conceptual proposition and corresponding methodology will be exemplified using the mobile application *Samsung Health* as a case study to show how changes in the user interface and algorithmically determined functionality have contributed to reframing Samsung's stance towards the datafication of personal health over time.[3] With this approach, the chapter is intended to be relevant both as a methodological inspiration on how to analyze evolving software applications (e.g., within software studies or critical data studies) and as a guide for game scholars on how their methods and concepts apply to the broader datafication of contemporary digital societies.

Studying Software Affordances as Quasi-Textual Properties

For this argument, I propose studying algorithmically determined functionality—rather than specific types of algorithms like recommendation

2 Below, I refer to algorithmics primarily in terms of algorithmically defined software functionality (like determining how users can compare metrics or interact with others in apps like *Samsung Health*) rather than in terms of implementation on the level of code (like different approaches toward sorting or recommendation routines).

3 For the Android version of *Samsung Health* on the Google Play store, see https://play.google.com/store/apps/details?id=com.sec.android.app.shealth&hl=en&gl=US.

systems or layout algorithms for data visualization—in terms of software affordances (Curinga 2014). In media and culture studies, affordances have been defined as a "relational property" (Bucher and Helmond 2018, 235), i.e., as "socio-technical" (248) characteristics of software platforms understood as "environment[s]" (243). Matthew Curinga (2014) expands on that framing by intuitively describing "software as text" (n.p.) but without making explicit how to study the textual properties of software design. Yanni Alexander Loukissas (2019) uses similar terminology, arguing that "if data can be considered as texts […] interfaces are contexts: the settings in which data are meant to be fully understood" (125). This analogy suggests that the relationship between user and software/algorithm can be understood in similar terms as the co-creative relationship between reader and text. Just as a written text makes certain interpretations more or less plausible, the "interface as discourse" (Stanfill 2015) communicates "norms of use" by "mak[ing] certain uses easier or harder" (1061).

To further operationalize this perspective and formulate a heuristic specifically suitable for the diachronic perspective below, I suggest approaching (critical) affordance analysis as an equivalent of textual analysis, which implies conducting a "close reading" (Looy and Baetens 2003, 8) of the rhetorical implications of software functionality. The following six criteria may be useful to guide a "close reading" of software affordances:

First, characteristic omissions may communicate norms and values in the same way as adding features. For example, when Facebook introduced five predefined emotional responses for users to connote content posted on the platform in 2016, both tech journalists and (many) users were keenly aware that these did not include the much-discussed "dislike button" that many had expected, and which the company had been experimenting with.[4]

Second, since a close reading prioritizes uncommon phrasings rather than interpreting a text line by line, unusual design choices are a good place to start. According to Matthew Curinga, "often, the most powerful interpretations push the boundaries of [the] rules [of software design]" (n.p.).

Third, like exploring paradigmatic semantic relations in literary texts, i.e., comparing word choices against potential alternative phrasings, software design choices should be interpreted based on hypothetical alternatives. For example, Google's decision to keep and remodel its "I'm feeling lucky" button appears particularly meaningful given that the button has been

4 See https://www.businessinsider.com/why-facebook-didnt-make-dislike-button-2016-2.

part of the launch page since the early days and could have been removed long ago because it no longer served a clear function.[5]

Fourth, software design choices constitute an "'ideal user,' i.e., 'script[…]' a discursive material configuration of ideal use" (Docherty 2020, 1); following the analogy above, this process can be compared to how a text, according to Umberto Eco, constitutes a "model reader" (Pisanty 2015).

Fifth, much as a text that is not phrased carefully can elicit unintended interpretations, this also occurs as a consequence of software affordance design. As an example, consider apps like *Forest*, which aim to promote productivity by actively discouraging smartphone use as a distraction; because they simultaneously include social features like screenshot sharing to grow their user base, these design choices contradict the core functionality.

Sixth and finally, just like literature contributes to altering and/or expanding contemporary social imaginaries (e.g., Fluck 1983), software not only contributes to socio-cultural transformations but simultaneously shapes the users' algorithmic imaginary (Bucher 2017), i.e., assumptions about how algorithms "function" on both a technical and on a social level. Alexis Papazoglou (2019) provides a pertinent example by examining Facebook's then-new "Why am I seeing this post?" feature, arguing that it increases "algorithmic transparency" but also "could affect how we see ourselves," as it creates a feedback loop by mirroring our user activity through the lens of "Facebook's algorithm" (n.p.).

After outlining how to analyze the quasi-textual properties of software, the following section adds to this definition by emphasizing how many software applications, not just games or examples of overt gamification, can be understood as quasi-games that afford different "playing styles."

Re-Framing Software Affordances in Terms of Games

With his notion of "expressive processing," Noah Wardrip-Fruin (2009) posits that "data" and "process" (10) are always connected in software, particularly in digital games. He suggests that by observing processes like "AI techniques in the context of a relatively easy to evaluate area such as computer games […] we can use that understanding to judge proposals for using similar techniques in higher-stakes social contexts (e.g., areas such as surveillance)"

5 See https://www.thesearchengineguys.com/googles-im-feeling-lucky-button-has-received-a-remodeling.

(5). Similarly, authors like Alexander Galloway (2004) have drawn attention to how playing a digital game like *Civilization* involves "learning, internalizing and becoming intimate with a massive, multipartite global algorithm." Both authors do not use the notion of software affordances, but the connection they establish between games and other broadly defined types of software is very plausible. However, while both Wardrip-Fruin and Galloway suggest interpreting games as software, I instead propose to interpret algorithmic systems as games, specifically from the perspective of the users.

The ongoing popularity of gamification has led to an—often uncritical—incorporation of basic game mechanics like virtual currencies or leaderboards into a wide range of software applications.[6] However, non-game software also arguably encourages playful forms of use or even implicitly uses design metaphors from games. One such example is the snap-streaks feature that "challenges" Snapchat users to create unbroken chains of messages by responding within a narrow time frame, a feature that is clearly reminiscent of "combo" mechanics popular in digital games and can be just as "addictive" as reaching a high score in a game.[7] This "gameness" (Malaby 2007) of (specific types of) software becomes particularly evident from a long-term perspective; similar to how players of service games react to—also partly black-boxed—changes to the "meta" over time, social media creators adapt their strategies to changes in recommendation algorithms, using terminology reminiscent of games like "survive" and "outsmart" to describe the process.[8,9] Based on this premise, the aforementioned notion of the "ideal user" (Docherty 2020) appears comparable to the "implied player" (Aarseth 2014), a term Espen Aarseth uses to describe the strategies and tactics "suggested" by the rules and constraints of a game, i.e., forms of player behavior that prove successful and thus opportune. Docherty emphasizes how Facebook, both internally and externally, frames forms of use that are conducive to their goals as "healthy" (n.p.). Aarseth defines the implied player in similar, albeit fuzzier, terms as "a role made for the player by the game, a set of expectations that the player must fulfill for the game to 'exercise its effect,'" which have "a concrete, material existence" (both 132) enforced by the game's algorithmic composition.

6 See https://techcrunch.com/tag/gamification.

7 See https://www.businessinsider.com/teens-explain-snapchat-streaks-why-theyre-so-addictive-and-important-to-friendships-2017-4.

8 See https://www.pcgamesn.com/path-of-exile/expansion-expeditions-new-gems.

9 See https://sproutsocial.com/insights/instagram-algorithm.

This re-framing of algorithms "as games" can be helpful as a conceptual middle ground between the more traditional perspectives of technological determinism and the social construction of technology. Increasingly ostracized in many academic discourses, "technological determinism persists in the actions taken and justifications given by many actors" (Wyatt 2008, 167) beyond academia, and re-framing the user as a "player" who partially co-creates the game can offer a more "contemporary form" of "hybrid, or 'weak technical determinism'" (Curinga 2014, n.p.) suitable for the analysis of mobile applications and other consumer-facing software technologies characterized by constant updates and "tweaking" (Bogost 2016). Below, these changes will instead be interpreted as changing "rules of play," which lead to a spectrum of likely changes in "player behavior." As a case study, I use the mobile application *Samsung Health*.

The Rhetorical Dimension of Software Affordance Changes

Existing critical affordance analyses of software applications, ranging from blogging software (Hopkins 2013) and online social networks like Facebook (Curinga 2014) to civic tech organizations like mySociety (Baack 2018), usually focus disproportionately on the contingent moment of observation. Yet software is increasingly characterized by constant change. In her discourse analysis of the 1968 Garmisch conference, which is often considered the origin of software engineering, Federica Frabetti (2015) emphasizes "the pace of software growth" (73). This already encouraged developers in the 1960s to take disproportionate shortcuts and insisted that "society need[ed] to take responsibility for an incalculable risk" (75). These rapid changes have only become more prominent, to the point where constant change arguably constitutes an important aspect of the "social epistemologies," i.e., "the way in which we use and develop knowledges in everyday life" (Berry 2012, 381), of software itself. Companies like Salesforce (2007) initiated this shift and pioneered the platformization of software by selling subscriptions rather than physical products.[10] Ian Bogost (2016) has argued that the constant "tweaking" of an iconic algorithm like Facebook's Edgerank imbues it with quasi-religious connotations, i.e., "raises its station, fetishizes it, treats it as a totem" (n.p.). More recently, YouTube's controversial changes to its dislike functionality in November 2021 made it particularly evident that content

10 See for example https://techcrunch.com/2019/03/22/how-salesforce-paved-the-way-for-the-saas-platform-approach.

creators but also viewers are becoming increasingly aware of how companies communicate norms and values (e.g., suggesting to create an "inclusive and respectful environment") through affordance changes.[11] Marshall McLuhan (1994) famously argued that if "social rules change suddenly, then previously accepted social manners and rituals may suddenly assume the stark outlines and the arbitrary patterns of a game" (238–39), and many YouTube users indeed considered the seemingly arbitrary implementation an abrupt change of the "rules of play" on the platform.

Investigating patterns of change is also important to better understand and contextualize the data processed by a given software application because, as Loukissas (2019) reminds us, "data and algorithms are inextricably entangled" (103). In other words, they can only be meaningfully investigated in conjunction with each other. For example, by discussing how "algorithms can be racist and sexist," Rebecca Heilweil (2020) illustrates the difficulty in separating between algorithm and data, even though commonly used terms like "algorithmic bias" (n.p.) suggest that the root of the problem lies in the algorithm as "text." Yet many instances of algorithmic bias reported in recent years are primarily caused by insufficiently diverse training datasets used to improve machine learning applications. This entails that a change in training data or the availability of new training data also need to be taken into consideration along with affordance changes of algorithmic systems.

Reflecting on Twitter, Taina Bucher and Anne Helmond (2018) already address two important changes in its functionality: the "turn to hearts" and the corresponding icon change as well as "enabling a new timeline ordering" (244). However, the authors primarily focus on how the new status quo can be interpreted using several variations of the affordance concept rather than conceptualizing patterns of change themselves. To address this gap, diachronic affordance analysis focuses on tracing changes in the algorithmic behavior of software over time as procedural rhetorical operations in themselves, which create meaning by re-writing the implied "rules of play." As an analytical method, diachronic affordance analysis is driven by a research question and relevant concepts that help to identify and select the most relevant affordances for analysis. Here, the method will be exemplified by considering several important moments of change in the mobile application *Samsung Health*, the discursive context in which they take place, and how they reflect Samsung's stance on the datafication of personal health.

11 See https://blog.youtube/news-and-events/update-to-youtube.

Samsung Health, or: The Instrumentalization of Personal Health in Platform Politics

Launched in 2012 under the name *S Health*, the app primarily enabled users to monitor weight, blood pressure, and blood sugar levels by synchronizing with devices by Lifescan, Omron, and AandD via Bluetooth or USB. Much like its competitor *Apple Health*, *Samsung Health* (as it was rebranded in 2017) has become increasingly integrated into the functionality and algorithmic imaginary (Bucher 2017) of the smartphones it runs on (for example, it was launched simultaneously with and preinstalled on Galaxy S3 smartphones). As such, the evolution of the app and its "rules of play" reflect different phases in Samsung's platformization strategy, specifically in competition with Apple. The diachronic affordance analysis demonstrated in this short chapter primarily refers to articles from the Samsung Newsroom website and a few user reviews as material; these sources are indexed (see the section "Primary Sources"] and referenced by their indices below. When analyzing the rhetorical import of changes in the software affordances, changes pertaining to different types of data, including health-related data, user profile information, and relevant metadata (for example, incorporating *Samsung Health* into the Samsung Rewards program [S-17-3], i.e., effectively translating in-app activity into this external "virtual currency") will be particularly pertinent. Rather than documenting the addition, modification, and removal of features chronologically, the analysis below focuses on several rhetorically significant patterns of change and the corresponding cultural implications.

How Social Metadata Contribute to the "Gameness" of Samsung Health

It is important to point out that *Samsung Health*—like many "quantified self" applications—contains elements of gamification, but gamification is not part of its core functionality nor does it fit the definition of a digital game. This section focuses on how new metrics increase its gameness, but below I will elaborate on why it makes sense to consider *Samsung Health* "as a game," regardless of leaderboards and challenges.

In 2017, *Samsung Health* began systematically generating social metadata through its competitive "Together" feature, which incentivized users to compare fitness levels through a steps leaderboard and one-on-one challenges [S-17-3]. This step in the increasing datafication of *Samsung Health* makes it particularly plausible to re-frame the application "as a game" as suggested above, because it affords setting and especially measuring user-defined

Fig. 12. Samsung Health "Together" feature (Samsung Newsroom).

"goals." Marshall McLuhan points out the inherent momentum of numbers, because, understood as "media," they exhibit a "dynamic drive toward growth and aggrandizement" (McLuhan 1994, 106); that is, numerically expressing any given phenomenon implies a tendency "toward unlimited growth."[12] Samsung's advertising of this feature change confirms McLuhan's proposition that "the joy in the multiplication of numbers" stems from "the pleasure of being among the masses" (both 107) by being immersed in a community, e.g., by listing the aggregate achievements of the *Samsung Health* user base, including walking a total distance of 59 billion km or collectively burning 3 billion kcal [S-21-1]. The ongoing refinement of social metadata in the app only explicates and intensifies this inherent "gameness" of software applications like *Samsung Health*. For instance, an early 2021 update introduced a "group challenge feature" that uses game-related iconography like a crown and a racetrack to incentivize the creation of new challenges, which in turn catalyze the creation of new metadata. These challenges are "hidden" behind titles like "who buys coffee?" that attribute social meaning to the numerical comparison. In this case, metadata like the time passed since a user's last in-app activity are designed to nudge users/players toward competitive behavior, albeit unnoticeably, since they appear in a small, light gray font in the user interface.

12 While this characteristic applies to many game genres like strategy or role-playing games, it is literally at the core of the more recent micro-genre of "incremental" or "idle games"; see for example https://pixl.nmsu.edu/files/2018/02/2018-chi-idle.pdf.

Other metadata created by the service include synthesizing aggregate "score[s]" for health-related phenomena, e.g., via the new sleep tracking features in September 2020 [S-20-9], which combine metrics like "breathing, vitals, and REM cycles" into a new data point via a proprietary algorithm. Devin Gaffney and Cornelius Puschmann (2012) point out that these aggregate "scores," popularized by the now discontinued "Klout score," appear game-like due to the lack of "algorithmic transparency" (2). That is, because the influence of individual metrics is partially but not completely transparent like in a digital game, these systems afford playful usage practices, such as tweaking individual parameters to infer their importance and gradually refining one's understanding (or, in game terminology, one's mastery) of the algorithmic model. Since 2018, the social metadata are also gradually being monetized, as the 6.0 update introduced a "discover" feature that uses the previously aggregated data on user "interests and fitness level" to display external content and "partner applications" and to allow partners to sell fitness accessories and clothing "within the platform" [S-18-10; translated from German by the author]. While this section focused on how new metrics increase gameness by suggesting new types of "scores" to compare, the focus below shifts toward the logic of incremental technological sophistication, which applies both to smartphones and video game hardware.

The Teleological Impulse of Software and Game Technologies

From a diachronic perspective, the *Samsung Health* app and its data-processing features follow the same logic of escalation as digital entertainment media, specifically gaming hardware. For instance, the on-demand electrocardiogram (ECG) added to Galaxy Watch devices in September 2020 was advertised as a "next-generation feature" [S-20-9], a framing reminiscent of the recently launched new game console generation. This incremental addition of new parameters to track, for example the concurrently implemented feature to "track oxygen saturation on Galaxy Watch3," maintains a constantly reinforced sense of "completeness" and an "encyclopedic impulse" (Clark 1992) as users are repeatedly reassured that the app will eventually offer a perfect datafication of their vital functions. Again, this paradigm becomes particularly plausible because it similarly applies to contemporary developments in gaming such as "games-as-a-service" (Dubois and Weststar 2021), which are continuously updated with new properties and mechanics to customize their virtual characters.

A corresponding development is the increasing focus on small-scale but real-time data evaluation, for example via the "new digital running

coach" that was introduced with the same September 2020 update and which arguably promotes a neoliberal "dashboard" (Batty 2015) approach to personal health. Batty shows how "the idea of monitoring human systems [via real-time dashboards] is intrinsic to modern medicine" and was later expanded to "monitoring human organizations" (29) like cities, inherently understanding them in similar terms as an "organism." The characteristic recombination of seemingly disjointed data points like "social class and density as well as pollution, twitter feeds etc." (31) in dashboards also applies to the constant addition of new real-time metrics in *Samsung Health*. After examining how both gamers and software users expect the ongoing inclusion and interconnection of new data points, the next section addresses specific types of "ideal users" as well as "implied players."

Defining "Ideal Users" Based on Combinations of Data Points

The "running coach" also demonstrates how, rather than limiting itself to generic "health" data, *Samsung Health* increasingly introduces and recombines metrics specific to "ideal types" of users with distinct interests like running, meditation, or injury prevention. These can be interpreted following Aarseth (2014) as implied players, who, similar "to [Hans-Georg] Gadamer's notion of the unfree player subject," constitute a "a boundary imposed on the player-subject by the game" (132) by tweaking the rules to encourage certain types of interaction (by adding further nuances) while discouraging others (by making them harder or more time-consuming to do). Constant feature change turns this mutual conditioning into an actual feedback loop; for example, the "launch screen [in the 6.0 update] was significantly simplified" in accordance with "the requirements and habits of the users" [S-18-10; translated from German by the author], i.e., often-used features and data points are positioned even more prominently and reinforce existing usage habits and "types." The notion of archetypal usage scenarios also affects the interrelatedness of data and algorithms, as new algorithmic features like "trip detection" can use available data (in this case the movement speed via GPS combined and/or the pedometer information) to infer standardized usage contexts, in this case e.g., to check only for tripping if the user is found to be "running" (rather than walking or meditating, for example). Thus, while users often do not reflect on how the software-as-game affords different ideal types of use, companies like Samsung gradually solidify existing "player types" as categories by adding new functionality that adds further nuance or gratification for users/players following these pre-existing paths. Until now, the analysis has focused

on how implications of affordance change for *Samsung Health* itself; the following sections widen the scope to discuss how recent changes position the application more broadly within pertinent societal discourses.

Intervening in Societal Debates via Affordance Changes

One of the most evident instances of rhetoric via affordance change was the "response" to the COVID-19 pandemic, notably by incorporating *Samsung Health* into Samsung Smart TVs in May 2020, two months after the first period of worldwide lockdowns. The television set has long been understood as the center of the "home." As David Morley (2004) notes, "the concept of home [has been] destabilized, both by new patterns of physical mobility and by new communication technologies" (303), and while some of these boundaries, specifically between work and leisure, have become even more permeable, the pandemic also clearly rearticulated the home as the locus of family life. The lockdowns reasserted TV's place in the home and the family, not least because it is usually connected to gaming consoles and runs Smart TV apps as well. Consequently, data-related changes in *Samsung Health* emphasize the family, for example via "individual accounts for yourself as well as your family members" to provide "personalized recommendations on workouts" [S-20-5] and more. The new affordance of being displayed, according to Samsung, "on the biggest screen in the household" (i.e., occupying a central space in the users' lives, especially under conditions of working from home and home schooling) also facilitated new ways of receiving metadata like "routines," i.e., reminders to perform workouts or relaxation exercises at specified times during the day.

Samsung itself did not explicitly address the pandemic, only stating that "given the current climate, we hope that the launch of *Samsung Health* makes it easier for our consumers to prioritize their physical and mental wellbeing on a daily basis."[13] Thus, the affordance changes can be understood as "filling in the gaps" in Samsung's official corporate communication via the media modality of user experience design. At the same time, extending the dashboard approach into the family, such as via health "goals" that can be expressed numerically (e.g., steps per day or number of meditation sessions per week) and shared between family members, expands the influence of "computing as a neoliberal governmental technology" (Chun 2011, 6) in the household. Apart from non-verbally "responding" to the unprecedented

13 See https://news.samsung.com/us/samsung-health-now-available-2020-samsung-smart-tvs-fitness-wellness-platform.

pandemic as it unfolded, *Samsung Health* recently used content updates even more granularly to intervene in societal debates tied to specific, distinct, or recurring events like the Christmas holidays. For example, the aforementioned group challenge feature was explicitly associated with the users' "New Year's Resolution" in the corresponding announcement [S-21-1], suggesting that the evolving data manipulation affordances built into the software can and should be interpreted as part of the users' everyday life and cultural environment.

This section addressed how software companies can respond to and intervene in societal debates through affordance changes, which are often more imperceptible than verbal or even visual corporate rhetoric and thus offer rhetorical opportunities because many users are not yet trained to "decode" them (e.g., in comparison with decades of advertising literacy education). The final section below tentatively incorporates user reviews, which can offer a glimpse into how users actually interpret specific affordance changes and develop procedural literacy in the process.

Considering User Reviews to Validate Hypotheses

Interpreting affordance analysis as a textual analysis of software implies that it can primarily identify likely interpretations on the basis of aesthetic choices; to assess the plausibility of these interpretations, analyzing discursive patterns in user reviews from the online app stores can be a suitable next step, even though these reviews can only offer anecdotal evidence. For example, user reviews can provide insights into how affordance changes shape the perceived algorithmic imaginary (Bucher 2017) of *Samsung Health*, with users explicitly addressing how adding or changing features affects their user behavior or "playing style." For example, user reviews often propose feature additions and changes to address usability concerns common in digital games. One highly evocative and controversial affordance change was the removal of the weight management, caffeine, and calorie tracking in July/August 2020. This feature removal was not explicitly communicated and justified by Samsung, which led to confusion and irritation within the user community. Responses indicate that users feel that their "investment" (both financial and emotional) in Samsung devices was devalued by this decision, arguing that it turns "existing Galaxy Smartwatches into an expensive step counter" and that the "cheaper Fitbit beats you [i.e., Samsung] now."[14] This suggests data are (justly) interpreted as assets in the ongoing platform competition, but, as users feel tied to platforms

14 See for example https://eu.community.samsung.com/t5/mobile-apps-services/samsung-health-app-weight-management-you-killed-it-care-to/td-p/1917065.

like *Samsung Health*, they expect them to "play that game" on their behalf as effectively as possible. For example, one user argues that this "unnecessary and inexcusable change has made the Samsung ecosystem useless to [them]"; another even explicitly mentions the game metaphor, defending Samsung's community managers by arguing that they are "just minions in this game." The five directions for diachronic affordance analysis outlined above do not claim to constitute a complete methodology, but they can be adapted to analyze how other types of non-game software change the implied "rules of play" and thereby continually readjust the procedural rhetoric of the respective application. The final section below offers some considerations for that purpose.

Outlook

As shown above, changes in software affordances readjust the framing of personal health (as well as related concepts) over time and give users new rules to play by. These involve framing health as an inherently social issue by adding social metadata but also by associating Samsung as a technology company with health insurances; the integration of Samsung Rewards operates similarly to incentive programs offered by insurances, i.e., providing benefits for using health-related in-app features, albeit within the Samsung ecosystem rather than society at large. As these changes occur gradually, they are often imperceptible to individual users, which can make them more influential. It should be noted in a few cases that interoperability qua data was seemingly counterintuitively limited, for instance by removing the integration with other apps through "connected services" [AA-18-8] in September 2018. These changes likely have pragmatic reasons, but they might nonetheless elicit "unintended interpretations" as suggested above, e.g., making the company appear "less open and more restrictive."[15]

Due to its scope, this chapter can primarily demonstrate the benefits of diachronic affordance analysis and of framing algorithmic systems as games using a limited case study; therefore, it appears useful to end on a few methodological suggestions. For example, the user-as-player analogy can be more systematically operationalized by elaborating on the user's explicit or implicit goals, routines, and strategies. Noah Wardrip-Fruin (2010) suggests using the "MDA framework" to conceptualize both games and other types of "operational logics" (17); this could help in differentiating between interfaces

15 See for example https://www.sammobile.com/2018/08/27/samsung-health-syncing-data-third-party-apps.

and "rules" (mechanics), emergent routines and feedback loops users adopt in response to these "material constraints" (dynamics), and more interpretive, self-reflexive observations based on long-term habitual use (aesthetics). For a larger-scale analysis, it would be useful to chronologically organize changes according to category (UI, social functionality, connection to devices like wearables or smart TVs, etc.) and include contemporary tech blog coverage, ideally multiple sources per update to identify potential interpretations of affordance changes from different angles.[16]

Furthermore, the method outlined above can be tweaked to accommodate other types of software. For example, I have demonstrated earlier how to use the timeline tool *Timeflow* to visually explore affordance changes, which is particularly useful for larger datasets and/or for specifically comparing affordance changes in different categories; these can be color-coded in *Timeflow* (Werning 2019). Using the online archive *Wayback Machine* offers additional opportunities for visual comparison, e.g., in adapting the method to study web applications (such as the *Coronadashboard* of the Dutch government), as it allows for contrasting different versions of the application's launch page over time. In this way, instances of priming (e.g., through the order and visual composition of data points on the page), the verbal and audiovisual framing of the implied user (Docherty 2020), and preferred "playing styles" can be compared systematically.[17]

References

Aarseth, Espen. 2014. "I Fought the Law: Transgressive Play and The Implied Player." In *From Literature to Cultural Literacy*, edited by Naomi Segal and Daniela Koleva, 180–88. London: Palgrave Macmillan. https://doi.org/10.1057/9781137429704_13.

Baack, Stefan. 2018. "Civic Tech at MySociety: How the Imagined Affordances of Data Shape Data Activism." *Krisis: Journal for Contemporary Philosophy*, no. 1. https://archive.krisis.eu/civic-tech-at-mysociety-how-the-imagined-affordances-of-data-shape-data-activism.

Batty, Michael. 2015. "A Perspective on City Dashboards." *Regional Studies, Regional Science* 2, no. 1: 29–32. https://doi.org/10.1080/21681376.2014.987540.

Berry, David M. 2012. "The Social Epistemologies of Software." *Social Epistemology* 26, no. 3–4: 379–98.

16 See the official coverage of the 6.0 update in late 2018 [S-18-10] and the corresponding feature preview by Android Authority [AA-18-8] as proof-of-concept below.

17 See https://web.archive.org/web/20200801000000*/https://coronadashboard.rijksoverheid.nl.

Bogost, Ian. 2016. "Go Tweak Yourself, Facebook." *The Atlantic*, April 28, 2016. http://www.theatlantic.com/technology/archive/2016/04/go-tweak-yourself-facebook/480258.

Brock, Kevin, and Dawn Shepherd. 2016. "Understanding How Algorithms Work Persuasively Through the Procedural Enthymeme." *Computers and Composition* 42 (December): 17–27. https://doi.org/10.1016/j.compcom.2016.08.007.

Bucher, Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information Communication and Society* 20, no. 1: 30–44. https://doi.org/10.1080/1369118X.2016.1154086.

Bucher, Taina, and Anne Helmond. 2018. "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Thomas Poell, and Alice Marwick, 233–53. London and New York: SAGE Publications.

Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge, MA: MIT Press.

Clark, Hilary A. 1992. "Encyclopedic Discourse." *SubStance* 21, no. 1: 95–110. https://doi.org/10.2307/3685349.

Curinga, Matthew X. 2014. "Critical Analysis of Interactive Media with Software Affordances." *First Monday* 19, no. 9. http://firstmonday.org/ojs/index.php/fm/article/view/4757/4116.

Docherty, Niall. 2020. "Facebook's Ideal User: Healthy Habits, Social Capital, and the Politics of Well-Being Online." *Social Media and Society* 6, no. 2. https://doi.org/10.1177/2056305120915606.

Dubois, Louis Etienne, and Johanna Weststar. 2021. "Games-as-a-Service: Conflicted Identities on the New Front-Line of Video Game Development." *New Media & Society* 24, no. 10: 2332–53. https://doi.org/10.1177/1461444821995815.

Fluck, Winfried. 1983. "Literature as Symbolic Action." *Amerikastudien* 28, no. 3: 361–71. http://www.jfki.fu-berlin.de/en/v/publications_fluck/1980/fluck_symbolic_action/Fluck_Symbolic_Action.pdf.

Frabetti, Federica. 2015. *Software Theory: A Cultural and Philosophical Study*. London: Rowman and Littlefield.

Gaffney, Devin, and Cornelius Puschmann. 2012. "Game or Measurement? Algorithmic Transparency and the Klout Score." In *#influence12: Symposium and Workshop on Measuring Influence on Social Media*. Halifax, Canada. http://devingaffney.com/files/influence12_submission_23.pdf.

Galloway, Alexander. 2004. "Playing the Code: Allegories of Control in Civilization." *Radical Philosophy* 128: 33–40. https://www.radicalphilosophy.com/wp-content/files_mf/rp128_article3_playingthecode_galloway.pdf.

Heilweil, Rebecca. 2020. "Why Algorithms Can Be Racist and Sexist." *Recode*, February 18, 2020. vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency.

Hopkins, Julian. 2013. "Assembling Affordances: Towards a Theory of Relational Affordances." In *AoIR Selected Papers of Internet Research*. Vol. 3. Association of Internet Researchers. https://spir.aoir.org/index.php/spir/article/view/783.

Leonardi, Paul M. 2010. "Digital Materiality? How Artifacts without Matter, Matter." *First Monday* 15, no. 6. http://firstmonday.org/ojs/index.php/fm/article/view/3036.

Loukissas, Yanni Alexander. 2019. *All Data Are Local. Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

Malaby, Thomas M. 2007. "Contriving Constraints (The Gameness of *Second Life* and the Persistence of Scarcity)." *Innovations: Technology, Governance, Globalization* 2, no. 3: 62–67. https://doi.org/10.1162/itgg.2007.2.3.62.

McLuhan, Marshall. 1994. *Understanding Media: The Extensions of Man* (*MIT Press Edition*). Cambridge, MA: MIT Press.

Morley, David. 2004. "At Home with Television." In *Television after TV: Essays on a Medium in Transition*, edited by Lynn Spigel and Jan Olsson, 303–23. Durham, NC: Duke University Press.

Papazoglou, Alexis. 2019. "Understanding Facebook's Algorithm Could Change How You See Yourself." In *Social Media and the Self. An Open Reader*, edited by Jefferson Pooley. Allentown, PA: mediastudies.press. https://doi.org/10.32376/3f8575cb.6e843c25.

Pisanty, Valentina. 2015. "From the Model Reader to the Limits of Interpretation." *Semiotica* 206: 37–61.

Stanfill, Mel. 2015. "The Interface as Discourse: The Production of Norms through Web Design." *New Media & Society* 17, no. 7: 1059–74. https://doi.org/10.1177/1461444814520873.

van Looy, Jan, and Jan Baetens, eds. 2003. *Close Reading New Media: Analyzing Electronic Literature*. Leuven: Leuven University Press.

Wardrip-Fruin, Noah. 2009. *Expressive Processing: Digital Fictions, Computer Games, and Software Studies.* Cambridge, MA: MIT Press.

Werning, Stefan. 2019. "Disrupting Video Game Distribution: A Diachronic Affordance Analysis of Steam's Platformization Strategy." *Nordic Journal of Media Studies* 1, no. 1: 103–24.

Wu, Teresa, Maurine Kwende, and Nada Dabbagh. 2020. "Affordance Analysis of WordPress." In *Innovate Learning Summit 2020*, edited by Theo J. Bastiaens and Gary H. Marks, 43–57. Waynesville, NC: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/218780.

Wyatt, Sally. 2008. "Technological Determinism Is Dead: Long Live Technological Determinism." In *The Handbook of Science and Technology Studies*, edited by Edward J. Hackett, Olga Amsterdamska, Michael E. Lynch, and Judy Wajcman, 3rd ed., 165–80. Cambridge, MA: MIT Press. https://www.dhi.ac.uk/san/waysofbeing/data/data-crone-wyatt-2007b.pdf#page=181.

About the Author

Stefan Werning is Associate Professor in the Department of Media and Culture Studies at Utrecht University, where he organizes the annual Game Research summer school. Stefan has been a visiting scholar and fellow in Comparative Media Studies at MIT and has worked in the gaming industry at Codemasters and Nintendo of Europe.
> s.werning@uu.nl

13. Schooled by Dashboards?

Learning Platforms' Performance-Centered Pedagogy and Its Impact on Teaching

Niels Kerssens

Abstract

Personalized learning is rapidly becoming a reality in classrooms worldwide through platformization. At the classroom level, digital platforms shape learning toward personal needs through pedagogies encoded into their design—their algorithms, but also into dashboard interfaces teachers increasingly employ as part of their educational toolkit. This study investigates how dashboards can impact teaching in primary school classrooms by examining how their data visualizations configure particular views on learning, which educators increasingly depend on to make pedagogical decisions. It will address two research questions: What are the pedagogical underpinnings of learning dashboards integrated in personalized learning technologies? How may pedagogies encoded into these dashboards affect teaching? To answer these questions, the chapter will start by setting out a theoretical perspective on platform pedagogy. Subsequently, it will describe teaching and learning relationships encoded in the teaching dashboard of the Dutch adaptive learning platform Snappet and argue that its pedagogy of performativity may disempower teachers' control over learning. The concluding section will discuss actions needed to strengthen schools' and teachers' control over the pedagogical dimensions of learning platforms.

Keywords: platform pedagogies, performance-centered pedagogy, educational technology, Snappet

Personalized learning—"customizing instruction based on analytics" (Friesen 2018)—manifests through platformization in public primary

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH13

education worldwide (Kerssens and van Dijck 2021). At a global scale, schools have started to implement personalized learning through AI-based adaptive learning platforms—such as *Knewton* in the United States, *Smart Sparrow* in Australia, and *Snappet* in the Netherlands—on which young students engage in learning activities while they work on a laptop or a tablet in a classroom. These intelligent learning platforms use algorithmic analytics to tailor education to a student's learning needs. Teachers interact through these platforms' interfaces, or learning dashboards—"single display[s] that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into [...] visualizations" (Schwendimann et al. 2016)—which have become pivotal technologies initiating and informing teachers' pedagogical actions for personalized learning (Molenaar and Knoop-Van Campen 2019). And these technologies' persuasive narratives may impact how teachers think about and teach students (Jarke and Macgilchrist 2021).

This study investigates how dashboards can impact teaching in primary school classrooms by examining how their "extracted analytics"—"data that are [visually] presented for interpretation" (Admiraal et al. 2020)—configure particular views on learning, which educators increasingly depend on to make pedagogical decisions. Dashboards, I contend, are not pedagogically neutral but incorporate values about learning and teaching, or "platform pedagogies" (Sefton-Green and Pangrazio 2021). At stake is not a concern about platform pedagogies replacing teachers but rather about teachers' pedagogical actions being exercised through platform algorithms and interfaces. A growing dependence of teachers on dashboard information as the basis for pedagogical decision-making may reshape teaching practice through platform logics and values of "good education," which do not necessarily accord with the educational and pedagogic interests of public schools and teachers and which are often excluded from public debate and examination within the educational field (Zeide 2019).

This chapter addresses the following research questions. RQ1: What are the pedagogical underpinnings of learning dashboards integrated in adaptive learning technologies? RQ2: How may pedagogies encoded into these dashboards affect teaching? To answer these questions, the chapter will start by sketching a theoretical perspective on platform pedagogy. The following section will implement this perspective by analyzing the case-study of the dashboard of the Dutch adaptive learning platform Snappet. Snappet's dashboard pedagogy, I argue, is performance-centered and may impact teaching by nudging teachers' pedagogical actions toward performance judgements and performance optimization, hence demanding that teachers perform the role of manager. In the concluding section, I will

briefly discuss what actions are needed to protect the pedagogical values of schools and teachers and to strengthen public control over learning platforms' pedagogical dimensions.

Platform Pedagogies

Do educational platforms have pedagogies? The answer to such a question is nothing but self-evident, as it is still highly common for educational professionals and scholars to approach technologies as neutral tools for improving teaching and learning. In the educational sciences, attributing pedagogical power to technology is deeply inconsistent with accepted views of pedagogy. One of these established views has been articulated by educational scholars Chris Watkins and Peter Mortimore, who define pedagogy as "any conscious activity by one person designed to enhance learning in another" (1999, 3). This definition illustrates at least two key points: 1) pedagogy is about power and exercising control over learning; and 2) pedagogy is human-biased; it is about people-effecting control.

In the educational sciences, such a human-biased view of pedagogy is wedded to a widespread instrumental perception of educational technology (Kerssens and de Haan 2022). For example, in their study of the dashboard use by teachers in Dutch primary school classrooms, educational scholars Inge Molenaar and Carolien Knoop-van Campen draw upon distributed cognition theory to investigate whether dashboard analytics provide efficient means to pedagogical ends: "a 'new' instrument that supports the selection of effective pedagogical actions by teachers" (2019, 1). Through classroom observations and interviews with teachers, Molenaar and Knoop-van Campen demonstrate that "teachers make dashboard information actionable" through the interpretation of displayed information to arrive at effective pedagogical practices for the personalization of learning (1). To interpret dashboard information, Molenaar and Knoop-van Campen describe, teachers activate diverse forms of existing knowledge about students and the class. This at least shows that teachers do not allow themselves to be blindly guided by dashboard information, which challenges any understanding of platforms' pedagogical logic as deterministic. Moreover, they conclude that their study provides first indications of how "dashboards progressively impact teaching practice and more profound behavioral changes seem to follow as teachers become more proficient in using dashboards" (7). Although these are important conclusions, to understand such impact, it is equally important to truly account for technologies' role in this process. Through

an instrumental perspective on educational technologies as potentially effective means to teachers' pedagogical ends, we are unable to account for the pedagogical logics incorporated in educational platforms and for how teachers' pedagogical actions and forms of activated knowledge about students may already be channeled and shaped by platforms' particular pedagogical dimensions.

This chapter draws on a concept of "platform pedagogy" to make sense of the framing and configuration of teaching and learning by dashboards within educational platforms (Sefton-Green and Pangrazio 2021). Although platform pedagogy as a conceptual tool centers analytical focus on platforms' potential shaping of human conduct, it does not in any way perceive pedagogical logics of platforms as deterministic. In contrast, it directs analytical attention to what science and technology studies scholar Philip Agre labeled "grammars of action" (1995)—that is, frameworks through which a platform architecture "seeks to shape participation" (Perrotta et al. 2021). As programmed structures, digital education platforms do not determine but rather provide a framework for human action, which may train particular forms of behavior and participation (Sefton-Green and Pangrazio 2021). A notion of platform pedagogy, then, challenges humanist bias in established concepts of pedagogy as a theory of power over learning—not just humans but also platform algorithms and interfaces can mediate, structure, and shape teaching and learning relationships.

Algorithmic analytics arranging adaptivity in learning technology has already been criticized in terms of how its pedagogy of automatically adapting education to student needs contests both student and teacher autonomy. Critical scholar of educational technology Norm Friesen has argued that algorithms coded into personalized learning platforms underpin a behaviorist model of learning, which may usher in a revival of "new behaviorism" in primary school classrooms (Friesen 2018). Friesen perceives such behaviorist reform of education as a threat to student autonomy, since platform algorithms nudge students toward behaviors predefined by learning analytics, which he considers antithetical "to the most basic priorities and purposes of education: to cultivate in students a sense of ownership in their own learning" (2018, 1). Law scholar Elena Zeide, on the other hand, points out that algorithmic analytics challenge the pedagogical authority of teachers, who have little insight in pedagogical decision-making encoded into algorithmic processing (2019).

Yet teachers' pedagogical actions are most significantly mediated through the pedagogical dimensions of learning platforms' interfaces—that is, the "extracted analytics" by which dashboards present data visually for

interpretation (Admiraal et al. 2020). Dashboard analytics have not yet been critiqued in terms of their inscribed pedagogies as the behaviorist logic of algorithmic adaptivity has. Such critique is equally important, as dashboard interfaces are "permeated with normative and evaluative ideas about good (e.g., 'green') or bad (e.g., 'red') education" (Decuypere et al. 2021). How interface pedagogies mediate learning and teaching, however, is not transparent to teachers, since assumedly objective dashboards present a "realist epistemology" to their users (Kitchin, Lauriault, and McArdle 2015). Dashboards make it appear to teachers that they truthfully represent the sometimes messy reality of learning through modes of data display that make learning visible, knowable, and manageable in real-time. Yet dashboard analytics offer teachers only a limited and restricted view of the complex reality of learning, as data available for visual presentation are already selective. Moreover, analytics render visible for interpretation only "particular representations of that data" (Williamson 2016), often through "increased emphasis on metrics, indicators and measures" (Bartlett and Tkacz 2017, 8).

Sketching the implications of dashboards' particular metric views for public governance, critical media scholar Nathaniel Tkacz and tech journalist Jamie Bartlett argue that dashboards "encourage more intensified forms of monitoring and analysis," "change the empirical basis from which decisions are made," and are fundamental for giving shape to an "ambience of performance" across government in which user views "become more attuned to how whatever is measured is performing" (Bartlett and Tkacz 2017, 8). In public education, learning dashboards seem to introduce a new performance-centered technology at the level of classroom interactions, which may significantly affect the pedagogical decision-making of their key users: teachers. Almost twenty years ago, sociologist Stephen Ball argued that education had become increasingly subject to *performativity*—"a technology, a culture and a mode of regulation that employs judgements, comparisons and displays as means of incentive, control, attrition and change" (Ball 2003).[1] Importantly, Ball argued that performativity had come to play a pivotal role in the construction of new teacher subjects, "changing what it means to be a teacher" (217). The pervasion of dashboard technology in educational learning spaces and its integration with teaching raise important

1 Notably, the "performance" in Ball's performativity (2003) refers to the perceivable acts and output of labor as a site of control and optimization and differs from other well-known conceptualizations of "performativity" within the humanities as an ontological term that indicates the unfolding nature of a given phenomenon (Derrida 1988; Butler 1990; Barad 2003).

questions about what role this intelligent technology, like other AI-based educational technologies, play in the production of a performative classroom culture and in reshaping the role of the teacher.

In the following section, I will demonstrate how a pedagogy of performativity manifests in the teaching dashboard of the Dutch adaptive learning platform Snappet by descripting its encoded grammar of teaching and learning relationships (RQ1). For this analysis, I collected and examined screenshots of the user interface of Snappet's dashboard, which contain visualizations of learner data on which teachers base their pedagogical decision-making for personalized learning. To better understand dashboard functionalities and their role in classroom teaching, I consulted the user manual of the Snappet dashboard (Version 3.0) and held a brief (informal) interview with a primary school teacher using Snappet on a daily basis. The section concludes with a reflection on how dashboards' pedagogy of performance may affect teaching and reshape teachers' professional role (RQ2).

Snappet's Performance-Centered Pedagogy

Snappet is used by 2,800 elementary schools in the Netherlands, accounting for 45% of all primary schools (Molenaar 2021). Its pedagogy, as I will demonstrate, is structured around *performativity*—making learning visible and actionable through various modes of displaying learner performance. Through its exceptional focus on performance display, I argue, Snappet's dashboard pushes into teaching a pedagogical logic based on persistent evaluation and optimization of student performance. But before I demonstrate how such pedagogy manifests at the level of user interface, I will briefly trace its cultural roots to a pedagogical rationale of performance-based teaching and learning, which dominantly influenced the organization of classroom pedagogy in Dutch primary schools from 2007 onward, before the popularization of platform dashboards.

Performativity as a Culture and a Mode of Interface Display

For more than a decade, performativity as a culture—as a rationale of teaching and learning focused on improving learning performance—has been actively promoted and implemented by the Dutch government in their program of "performance-based working" (Kerssens and de Haan 2022). Performance-based working refers to schools that "work systematically and purposefully to maximize the performance of its students" (IoE 2010,

4). Its pedagogical rationale is rooted in the managerial philosophy of New Public Management (NPM), which from the early 1990s onward affected educational reforms internationally (Gunter et al. 2016). From NPM, the Dutch performance-based approach inherited its view of datafication—the systematic registration, tracking, and analysis of data about learners and learning—as a key *instrument* for gaining insight into, controlling, and maximizing learning processes and learning performance (Kerssens and de Haan 2022). Performance-based learning wedded datafication to the objective of personalization, tailoring education to students' individual level of competence. Providing advanced technological possibilities for datafication and personalization, new digital platform technologies such as adaptive learning technologies, learning analytics, and dashboards landed comfortably in performance-based pedagogy, which was in search of tools for optimizing learning performance.

Performativity as a culture is formalized within Snappet's dashboard interface, which configures a performance-centered mode of display. Snappet's dashboard makes learning visible through color-coded information in various modes of visualizing learner performance, such as real-time progress, classification through a skill meter, performance growth graphs, and performance relative to target levels and peers. Teacher actions are mediated through performance displays in three dashboard tabs: "lessons," "tracking," and "reports." Under the "lessons" tab, teachers can prepare and start lessons, which students can view and practice on their device (tablet or laptop). The tab also offers teachers at-a-glance insight into the learning objectives for different domains (e.g., math) and subdomains (e.g., multiplication tables) and the performance of the student or class per learning objective compared to the target level (blue check mark means class is at target level; blue check mark with number in yellow means class at target level except for a few students, etc.). The target level indicates a future skill level that is expected to be attainable for the child concerned. It is automatically predicted using individual learning results per student and per subject through Snappet's implementation of the ELO algorithm and is established after approximately six hundred completed tasks per subject. As a result, children always work toward their own target level—meaning that the degree of difficulty of exercises differs per child—and any evaluation of student performance in the lessons tab always proceeds relative to students' past performance.

The "follow" tab displays student progress and performance to teachers on different learning goals per lesson, per group, and of individual students in real-time while students are doing exercises in Snappet.

The dashboard indicates how many problems students have solved and whether the problems were answered correctly. Circles indicate problems answered. Green indicates a correct answer, red an incorrect response, and combined green with red circles indicate a correct response on the second attempt. Teachers can also view how students perform on learning goals they have worked on, or are working on, compared to their previous performances. If students have completed at least twenty-five to thirty-five problems for a specific learning objective, a score is calculated and displayed in a bar with zero to four stars (zero stars, lowest-scoring learning goal or just started; four stars, highest-scoring learning goals). These stars are assigned based on the average performance of the student (Faber and Visscher 2016). A progress indicator (human icon in front of students' names) shows teachers which students are making progress (green icon), are not making progress (red icon), or are currently unknown (grey icon). These indicators enable teachers to see in real-time whether student performance is growing or declining compared to their earlier performances. If progress is lagging behind, teachers can invite students for an extended lesson.

The "reports" tab displays to teachers all information about the skill and growth of the individual student per selected subject and learning objective. A skill meter designed as a colored bar displays in a percentile score students' mastery of skill for a particular domain compared to the national average; this is calculated based on the performance of all other students in the Netherlands in the same year group who work with Snappet. A percentile score of 68% means that the student scores better than 68% of all students. Based on this score, the student is classified in a category ranging from "far below average" to "far above average." A growth chart visualizes the student's skill growth in a particular subject area over time (e.g., from grade 3 to grade 6). The chart as well visualizes the performance of students on different learning goals compared to other students in the Netherlands, enabling teachers to rapidly compare student scores with the average scores of other Dutch students from the same year group. Snappet, like almost all learning dashboards, has a strong emphasis on comparison and competition with peers, using comparison as "a representative frame of reference for evaluating their performance" (Jivet et al. 2018, 32). The graph also offers teachers a quick view of performance growth based on past learning data compared to expected growth for a coming period, displayed as a dotted line. If a student grows faster or less quickly than expected, teachers can manually adjust the target level.

Dashboard Performativity Reshaping Teaching

Performativity inscribed into Snappet's dashboard is ultimately about focus—it is about the dashboard encoding a pedagogical grammar in which performance serves as a central organizational principle, framing learning through variables that can be optimized to maximize student performance, while "the variables that have nothing to do with key performance [… is] what the dashboard screens out" (Mattern 2015, n.p.). By spotlighting performance as the true locus of teacher control and manipulation, Snappet's dashboard may encourage teachers to view student learning as something that is always susceptible to further enhancement and improvement in terms of efficiency and effectiveness. Continuously signaling teachers with information—on real-time progress, performance relative to target levels and peers, competence level—the dashboard provides them with actionable levers they can pull to tweak learning, pushing students to shift from red to green, from "below average" to "average."

Importantly, Snappet's pedagogy does not involve replacing teachers. On the contrary, Snappet's dashboard exercises power over learning *through* teachers. Teaching-by-dashboard, after all, means that educators base pedagogical actions for personalized learning on dashboard analytics and visualizations that are reshaping teaching into a practice increasingly "initiated and determined by systems of data analysis" (Knox, Williamson, and Bayne 2020, 34). Growing the dependence of teachers on dashboards, in effect, can mean that teachers are pushed toward behaviors that match with particular pedagogies formalized in these algorithms and interfaces, to the detriment of others. By encouraging certain pedagogical actions towards the optimization of learning—"driving out poor performance, inefficiencies and redundancies" (Ball 2008, 27)—Snappet's dashboard may strengthen and intensify an already established culture of performance-based work in Dutch classrooms.

Teacher dependency on dashboards, for that matter, may as well signal a more significant shift in the educator's role with dashboards and their performance-centered pedagogies "interpellating" teachers as managers (Jarke and Macgilchrist 2021). This shift may risk challenging teachers' pedagogical autonomy. As Ball described, teachers subject to a culture of performativity experienced "a potential 'splitting' between [their] own judgements about 'good practice' and students 'needs' and the rigours of performance" (2003, 221). Dashboards, then, may actually disempower teachers' control over learning, since their own pedagogical judgements and intuition are subjected to, and channeled by, dashboards' performance-centered pedagogies.

Conclusions: How to Strengthen Public Control over Platform Pedagogies

Personalized learning is rapidly becoming a reality in classrooms worldwide through platformization. Platforms shape learning at classroom level through pedagogies encoded in their design—in their algorithms, but also in dashboard interfaces teachers increasingly employ as part of their educational toolkit. Dashboards, as I showed, are not pedagogically neutral. The interface arrangement of Snappet's dashboard is based on a pedagogy of performance, which employs personalization as a means to a larger end of maximizing learning outcomes. Its pedagogy perfectly accords with a cultural rationale of performativity already deeply rooted in Dutch education. Yet Snappet facilitates this rationale's intensification at the level of classroom interactions, where it may push teachers to squeeze their pedagogical actions into the dashboard's straitjacket of performativity. Importantly, any platform's pedagogical logic is not to be seen as deterministic. Teachers interpret dashboard information to arrive at meaningful and effective pedagogical actions for adapting education to student needs. To better understand how performance-centered platform pedagogies affect teacher interpretations and pedagogical decision-making, more empirical research using forms of ethnographic observation and/or interviews with educational professionals is needed.

Nonetheless, teaching is unquestionably influenced by platforms' pedagogical dimensions. At stake is not so much the fact that platforms and their dashboards have pedagogies (all digital platforms have pedagogical dimensions), but that platforms, rather than schools and teachers, seem to be dictating pedagogy increasingly; platform algorithms and interfaces prescribe what "good education" is and what agency teachers should have to exert control over learning. Importantly, these platform pedagogies do not necessarily represent the educational and pedagogical values of public schools and teachers, and they are often not transparent to educational professionals. Platforms' and dashboards' underpinning pedagogical values are not central to public debate on platformization within the Dutch education sector and therefore do not constitute *public* values. Yet with the intensification of educational platformization in the past years, it has become even more urgent to critically inspect their pedagogical impact—and not only their privacy impact—and publicly weigh platforms' internal pedagogical logic and the values it represents (e.g., performance) against values that teachers and schools represent (e.g., teacher autonomy).

The pedagogical accountability of digital education platforms and pedagogical autonomy of teachers may be fostered through "pedagogical impact

assessments" (PIAs). PIAs can be carried out at schools through dialogical frameworks like the Data Ethics Decision Aid (DEDA) developed for reviewing the social impact of government data projects (Franzke, Muis, and Schäfer 2021). For education, PIAs can engage educational professionals in a dialogue about the pedagogical impact of platforms in use and considered for use, reflecting on their embedded theories and values of learning and teaching and the required teaching literacy. Such dialogue necessitates educational professionals who challenge established instrumental views of educational technology and bring issues about values into the discussion on the use and adoption of digital platform technologies. These impact assessments should then also work toward developing platform-compliant literacy conceptualizations. Many teachers view educational technologies as tools serving their pedagogical aims and view literacy as the ability to use these tools effectively. Yet digital education platforms are not pedagogically neutral instruments, and educators need to make sense of how they impact teaching and learning.

But as the pedagogic dimensions of platforms are encoded into user interfaces and algorithms, they are not transparent for teachers and thus not directly accessible for critical review. To strengthen the accountability of pedagogical decision-making processes encoded into platforms, PIAs should be based on scholarly investigation of platform pedagogies following Sefton-Green's and Pangrazio's research agenda (2021). To conclude, PIAs can make an important contribution to governing edtech as a public good and to helping teachers account for platform pedagogies. This makes it all the more important that their development and application proceed through democratic debate and inspection within the educational field and through cooperation between all stakeholders, including schools, educational professionals, educational scholars, and educational technology providers.

References

Admiraal, Wilfried, Jordi Vermeulen, and Jacqueline Bulterman-Bos. 2020. "Teaching with Learning Analytics: How to Connect Computer-Based Assessment Data with Classroom Instruction?" *Technology, Pedagogy and Education* 29, no. 5: 577–91. https://doi.org/10.1080/1475939X.2020.1825992.

Agre, Philip E. 1995. "From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing." *Computer Supported Cooperative Work* (*CSCW*) 3, no. 2: 167–95. https://doi.org/10.1007/BF00773446.

Ball, Stephen J. 2003. "The Teacher's Soul and the Terrors of Performativity." *Journal of Education Policy* 18, no. 2: 215–28.

Ball, Stephen J. 2008. *Education Plc: Understanding Private Sector Participation in Public Sector Education*. London: Routledge.

Barad, Karen. 2003. "Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter." *Signs: Journal of Women in Culture and Society* 28, no. 3: 801–31. https://doi.org/10.1086/345321.

Bartlett, Jamie, and Nathaniel Tkacz. 2017. "Governance by Dashboard: A Policy Paper."

Butler, Judith. 1990. *Gender Trouble.* New York: Routledge.

Decuypere, Mathias, Emiliano Grimaldi, and Paolo Landri. 2021. "Introduction: Critical Studies of Digital Education Platforms." *Critical Studies in Education* 62, no. 1: 1–16. https://doi.org/10.1080/17508487.2020.1866050.

Derrida, Jacques. 1988. *Limited Inc.* Evanston, IL: Northwestern University Press.

Faber, Janke Marjan, and Adrie J. Visscher. 2016. "De Effecten van Snappet: Effecten van een Adaptief Onderwijsplatform op Leerresultaten en Motivatie van Leerlingen." *Enschede: Universiteit Twente*.

Franzke, Aline Shakti, Iris Muis, and Mirko Tobias Schäfer. 2021. "Data Ethics Decision Aid (DEDA): A Dialogical Framework for Ethical Inquiry of AI and Data Projects in the Netherlands." *Ethics and Information Technology* 23: 551–67. https://doi.org/10.1007/s10676-020-09577-5.

Friesen, Norm. 2018. "Personalized Learning Technology and the New Behaviorism: Beyond Freedom and Dignity." https://www.academia.edu/36111979/Personal_learning_new_behaviorism2.docx.

Gunter, Helen M, Emiliano Grimaldi, David Hall, and Roberto Serpieri. 2016. *New Public Management and the Reform of Education: European Lessons for Policy and Practice*. London: Routledge.

IoE (Inspectorate of Education). 2010. *Opbrengstgericht Werken in het Basisonderwijs: Een Onderzoek naar Opbrengstgericht Werken bij Rekenen-Wiskunde in het Basisonderwijs*. Utrecht: Inspectie van het onderwijs.

Jarke, Juliane, and Felicitas Macgilchrist. 2021. "Dashboard Stories: How Narratives Told by Predictive Analytics Reconfigure Roles, Risk and Sociality in Education." *Big Data & Society* 8, no. 1. https://doi.org/10.1177/2053951721102556l.

Jivet, Ioana, Maren Scheffel, Marcus Specht, and Hendrik Drachsler. 2018. "License to Evaluate: Preparing Learning Analytics Dashboards for Educational Practice." *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, *Sydney*.

Kerssens, Niels, and Mariette de Haan. 2022. "The Tipping Point in the Platformization of Dutch Public Education? How to Approach Platformization from a Values-Based Perspective." In *Learning to Live with Datafication: Educational Case Studies and Initiatives from Across the World*, edited by Luci Pangrazio and Julian Sefton-Green. London: Routledge.

Kerssens, Niels, and José van Dijck. 2021. "The Platformization of Primary Education in the Netherlands." *Learning, Media and Technology* 46, no. 3: 250–63. https://doi.org/10.1080/17439884.2021.1876725.

Kitchin, Rob, Tracey P. Lauriault, and Gavin McArdle. 2015. "Knowing and Governing Cities through Urban Indicators, City Benchmarking and Real-Time Dashboards." *Regional Studies, Regional Science* 2, no. 1: 6–28. https://doi.org/10.1080/21681376.2014.983149.

Knox, Jeremy, Ben Williamson, and Sian Bayne. 2020. "Machine Behaviourism: Future Visions of 'Learnification' and 'Datafication' across Humans and Digital Technologies." *Learning, Media & Technology* 45, no. 1: 31–45. https://doi.org/10.1080/17439884.2019.1623251.

Mattern, Shannon. 2015. "Mission Control: A History of the Urban Dashboard." *Places Journal* (March).

Molenaar, Inge. 2021. "Personalization of Learning: Towards Hybrid Human-AI Learning Technologies." In *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. Paris: OECD Publishing.

Molenaar, Inge, and Carolien A.N. Knoop-van Campen. 2019. "How Teachers Make Dashboard Information Actionable." *IEEE Transactions on Learning Technologies* 12, no. 3: 347–55. https://doi.org/10.1109/TLT.2018.2851585.

Perrotta, Carlo, Kalervo N. Gulson, Ben Williamson, and Kevin Witzenberger. 2020. "Automation, APIs and the Distributed Labour of Platform Pedagogies in Google Classroom." *Critical Studies in Education* 62, no. 1: 1–17. https://doi.org/10.1080/17508487.2020.1855597.

Schwendimann, Beat A., Maria Jesus Rodriguez-Triana, Andrii Vozniuk, Luis P. Prieto, Mina Shirvani Boroujeni, Adrian Holzer, Denis Gillet, and Pierre Dillenbourg. 2016. "Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research." *IEEE Transactions on Learning Technologies* 10, no. 1: 30–41.

Sefton-Green, Julian, and Luci Pangrazio. 2021. "Platform Pedagogies –Toward a Research Agenda." *Algorithmic Rights and Protections for Children*. https://doi.org/10.1162/ba67f642.646d0673.

Watkins, Chris, and Peter Mortimore. 1993. "Pedagogy: What Do We Know?" In *Understanding Pedagogy and Its Impact on Learning*, edited by Peter Mortimore. London: Paul Chapman.

Williamson, Ben. 2016. "Digital Education Governance: Data Visualization, Predictive Analytics, and 'Real-Time' Policy Instruments." *Journal of Education Policy* 31, no. 2: 123–41. https://doi.org/10.1080/02680939.2015.1035758.

Zeide, Elana. 2019. "Robot Teaching, Pedagogy, and Policy." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das. Oxford: Oxford University Press.

About the Author

Niels Kerssens is Assistant Professor in the Department of Media and Culture Studies at Utrecht University. His research investigates the implications of platformization on education.
> n.kerssens@uu.nl

Part 4

Agendas

14. Creative Urban Methods for the Datafied City

Nanna Verhoeff, Sigrid Merx, and Michiel de Lange

Abstract
Datafied and smart cities produce some challenges for inclusive, resilient, and sustainable urban futures. How can creative methods contribute to thinking and designing ways to imagine and co-create datafied cities with and for participatory citizenship and values for inclusion and sustainability? This question is central to the agenda of the research group [urban interfaces] and their collaboration in interdicisplinary and transdisciplinary partnerships. Working with and around the concepts of participation, criticality and imagination, the group brings cultural inquiry into datafied cities together with a methodological inquiry into creative urban methods. In the following, we sketch this agenda and approach and some recent examples of what such creative methods may yield.

Keywords: Datafied Cities, Creative Methods, Civic Participation, Criticality, Imagination

Cities today are datafied cities. Digital data and algorithms—and their primary interfaces in the form of platforms, apps, wearables, and urban dashboards—shape almost every aspect of urban life: commercial transactions, public governance, mobility, and everyday interactions between people. The proliferation of (big) urban data spurs a research and policy agenda aiming to improve the management of so-called "smart cities." Less attention goes to the question of how to involve citizens in shaping the future of the datafied smart city (see for instance Powell 2021). This is especially urgent, as the power relations and the values embedded in urban infrastructures, systems, and interfaces have a major impact on how inclusive cities are. The logics of optimization and efficiency that

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH14

underpin datafied systems tend to benefit some but not others. Parking apps, for example, promote car mobility at the expense of other forms of travel. Algorithmic sorting has been shown to exacerbate existing divisions in society (O'Neil 2016), as is the case with predictive policing (van Schie and Oosterloo 2020) and tax office fraud detection systems (Oosterloo and van Schie 2018). Even something as apparently inconspicuous as a real estate website may ultimately serve to push out lower income home renters from old established neighborhoods while benefiting other tenants and homeowners (Loukissas 2019).

A key challenge that we want to respond to is how to make sure that datafication in practice does not promote the interests of the few but instead supports collective and diverse interests of citizens and their ability to participate in an inclusive urban society.[1] How can data strengthen civic participation and public values in the smart city? How can the use of data lead to more equitable outcomes for citizens? Moreover, the future of urban life itself is contingent on dealing with the climate crisis and whether we can co-exist with other species and our natural environment. Hence, we also feel it is important to ask: how might we develop and expand a notion of more-than-human "data justice" (Dencik, Hintz, and Cable 2016) that is not limited to human dwellers only? How can our cities become more sustainable—socially, ethically, and ecologically—by considering the diverse interests, stakes, and perspectives from other living organisms and species, aside from and beyond humans (Wakkery 2021), and how does this also concern (responses to) urban datafication? In other words: what can and should be the role of datafication in supporting sustainable and liveable cities, now and in the future?

Situating Urban Data

These challenges ask for critical and creative responses and approaches for thinking, debating, and engaging with data. Much of the research related to datafied cities is either solution-oriented and applied (e.g., much of the smart city scholarship in fields like engineering, policy, and computer science) or tends to assume a critical but thereby also disengaged position. Recently however, several scholars in critical data

1 In the Faculty of Humanities at Utrecht University, we organize the research group [urban interfaces], and we participate in the interdisciplinary Open Cities platform of Utrecht University's strategic theme Institutions for Open Societies and the focus area Governing the Digital Society.

studies have addressed data with more experimental and affirmative approaches (see Kitchin 2022, 127–41 for an overview), often in urban contexts. In this vein, our research group activities have taken on the question of how we can start "doing things with data" in ways that shed a critical light on urban datafication while at the same time allowing for creative and future-oriented speculation on how this could be different and better. For example, in line with the growing attention on informal urban practices as part of academic so-called action research, we ask how we can investigate the myriad daily urban data practices—practices on a "street level," so to speak. And consequently: how can we develop situated and hence cultural perspectives on data that allow us to move beyond the self-contained and purged datasets as the primary unit of analysis? Or: how can we build on the insight that data are always already situated and intertwined with various cultural and social practices, experiences, narratives, identities, systems of meaning, power dynamics, politics, and so on? And, specifically, what critical, yet productive, role can urban media, art, and performance play in teasing out and shedding new lights on those entanglements?

To address these questions, we work with concepts of *participation*, *criticality*, and *imagination*. These underpin our search for ways to foster more equitable citizen engagement with datafied urban life. Specifically, we ask how debates about societal frictions and controversies around data and algorithms can contribute to the development of urban imaginaries for more inclusive and sustainable futures. These are not only questions central to our research agenda but also inform our methodological inquiry into creative urban methods that bring together a cultural inquiry into datafied and algorithmic cities; citizen engagement; and the aims and strategies of critical, speculative, and value-based design.

Creative Urban Methods

To engage scholars, designers, and local citizens in shaping the present and future of the datafied city, we believe methods are needed that construct knowledge and awareness of the presence, (dis)functioning, and generative power and performativity of data in relation to urban realities and our own position within these datafied processes and infrastructures. Such methods can reveal the underlying layers of datafication and the actual powers that mobilize it and that it mobilizes (Karimnia 2019). Furthermore, we need methods that are inter- and transdisciplinary, given that the complexity of

the datafied city and how datafication shapes subjects and society at large cannot be understood or approached from an isolated disciplinary vantage point (Verloo and Bertolini 2020). Indeed, we need approaches that allow for practical and theoretical engagement with data and processes of datafication, with different stakeholders in specific, situated social environments as well as within interdisciplinary research and education contexts. A productive approach to the datafied city is ideally not only analytical (what are data, how does datafication manifest) and critical (what do these data "do," i.e., unpacking how data do not merely represent but also performatively produce reality) but also actionable (how can we act with, co-shape, or (re) design data and data systems). This combination could inform what Teli et al. in line with Kelty (2008) call recursive engagement: "the capability of a public of being able to take care of the infrastructure that allows its existence as a public" (2015, 20).

In this respect, we believe particularly in the potential of creative methods, with their focus on embodiment, exploration, experimentation, and intervention. Creative methods have shown to be productive for participatory, community-based, and action-based research, as they reflect the multiplicity of meanings that exist in social contexts, allowing for different stakeholders to participate in debate and collaborate in (practical) research (Hjorth et al. 2019; van der Vaart, van Hoven, and Huigen 2018). Creative methods can play an important role in bringing different perspectives and knowledges together, providing fresh and alternative approaches (Kara 2020; Dunn and Mellor 2017), as well as raising awareness and generating questions around complex subtleties (Eisner 2008). Moreover, creative methods value situational specificity (Kara 2020) and can provide access to emotional aspects of people's experiences not easily accessed by mainstream methods (Dunn and Mellor 2017).

Today, in the humanities (van der Tuin and Verhoeff 2022) as well as in urban planning, we observe an increasing interest in creative urban research methods, for instance in collaborative approaches to (smart) city making (Foth, Bryskov, and Ojala 2015; de Lange and de Waal 2019). These comprise methods such as data walking, performative mapping, experimental ethnography, interface analysis, action-based research, research by design, and critical making; these are methods that can be characterized as mapping methods, performative methods, and/or making methods (see also Verhoeff, Merx, and De Lange 2019). Below, we have included three short vignettes with specific projects to illustrate the situated character of these methods. These cases all share a perspective toward material, relational, performative, and affective structures of urban environments that

is sensitive to dynamics of change and has a phenomenological emphasis on embodied experiences of the (citizen/academic) researcher. Together, they demonstrate strategies for reflecting on and rethinking the datafied city. They show how creative methods can offer tools for alternative ways of collecting and presenting urban data and (co)creating or (re)imagining urban data or data infrastructures, which in turn might lead to finding new insights, raising further questions, and calling for alternative scenarios for the future of the datafied city.

Co-Creating Alternative Zero-Waste Imaginaries

Co-creating alternative "zero-waste" imaginaries is a research project that addresses the participation of citizens in imagining and shaping sustainable urban futures, focusing on issues and infrastructures of waste.[2] The project was initiated and led by two members of [urban interfaces] and aims to develop and test creative and transdisciplinary methods that can support the co-creation of alternative and more inclusive "zero-waste" imaginaries. The first test in November 2020 was a data walk. Small groups (students, researchers, designers, and local residents) explored "Het Werkspoorkwartier," a former industrial area in Utrecht, now transformed into a hub for creative making and circular entrepreneurship. Guided by a set of questions, participants were challenged to closely observe the environment and look for material glimpses of what a zero-waste future might look like. While walking, they collected objects, materials, images, and sounds—"relics of the future"—that were then used to create a map presenting different meanings of and perspectives on waste. A follow-up workshop in December 2020 built upon these analogue maps, inviting participants to combine them with visualizations of existing open data sets and citizen-science mappings of the area. Working through and with these different mappings, participants were invited to locate "fertile" starting points and pathways for potential zero-waste futures in the area and to creatively shape, share, and "sow," as the organizers called it, visions of these futures. Finally, participants discussed the relation between these visions and the "soil" that might be needed to ensure their growth and sustainability over time.

2 The project *Co-Creating Alternative "Zero-Waste" Imaginaries* started in 2021 at Utrecht University and is led by Corelia Baibarac-Duignan (University of Twente) and Tamalone van den Eijnden (University of Amsterdam) in collaboration with Creative Coding Utrecht. It received seed funding from the Transforming Cities Hub of the Focus Area Pathways to Sustainability.

Data-West 2021

Data-West 2021 was a local, public art project in the district of Woensel-West in the city of Eindhoven in the Netherlands that aimed to reconnect inhabitants of the neighborhood with their data. The [urban interfaces] research group was invited to contribute to the project as an academic partner through analysis and critical reflection. Art is used, here, to address and repair the loss of sensible and sensual access to, and ownership of, local data for urban citizens.[3] A group of young artists and designers from different disciplines, working with diverse media and methods (e.g., photography, film, audio, interactive installations, screens, animation, cooking, poetry), were invited to collaborate with local residents and stakeholders in the neighborhood to collect various forms of data and translate these into artistic data interfaces, ranging from multimedia, audio-visual experience design and installations to algorithmic wanderings, interactive animation, a recipe collection, and computer-generated poetry. The mission was to address local data as a social challenge and to use art and design to arrive at new insights, solutions, and connections. For this, the artists partnered with the local community, in all its diversity, to examine how data art can help to reclaim their local, environmental, and embodied bio-social data and to explore how data can regain expressive and experiential qualities and meanings. Art and design, here, are the methods to give data a "personal touch," as the organizers put it.[4] The project is both critical and creative, as it experiments with a variety of artistic methods to gather, reflect on, and mediate data. It is programmatic in how it makes a claim for participatory and on-site approaches to raise awareness about the situated and performative presence of data on a street level.

***Frictional Urban Interfaces*: A Pressure-Cooker Workshop**

In 2018, the [urban interfaces] research group organized a two-day pressure cooker workshop as part of the Research MA program Media, Arts and Performance at Utrecht University and in collaboration with Creative Coding Utrecht.[5]

3 The project *Data West* was organized by Gaia van Egmond, Arjanne Bode, and Lisette Aarnink of the social design collective Tante Netty, located in Woensel-West in Eindhoven. A first iteration was presented during the Dutch Design Week in 2020. Participating artists and designers in 2021 were: Cas de Rooij, Jannie Guo, Sandipan Nath, Studio Antwan, Julia Luteijn, and Tom Jacobs.

4 https://ddw.nl/en/programme/5858/data-west-2021.

5 See the documentation on https://urbaninterfaces.sites.uu.nl/workshops/2017-2018-workshop-critical-making-of-frictional-urban-interfaces and Shannon Mattern's description (2021, 49).

The workshop was attended by students, coders, and municipal workers. The challenge was to "reverse engineer" existing datasets in order to unearth the underlying frictions and contestations that went into the production of what often has the appearance of a comprehensive, objective dataset. The intended goal was to develop a speculative and imaginative interface that would narrativize these frictions around the datafication of urban life. Examples of frictions include urban data giving rise to new processes of social sorting (think of crime maps), the further encroachment of commercial interests onto urban public space (e.g., customer loyalty cards, personalized marketing, tracking), and the militarization of urban space (e.g., surveillance, facial recognition, risk analysis). The workshop was inspired by the critical making approach (Ratto and Hertz 2019) and the idea of data dramatization (Akten 2015). In an iterative set of steps, participants had to 1) imagine the underlying story in the making of an urban dataset, 2) consider friction as a way to bring drama into the reconstruction of the dataset, 3) develop the narrative setting of the friction, 4) imagine an urban interface to mediate the frictions imagined in the dataset and to allow people to engage with it, and finally 5) situate this interface in an urban intervention in order to make it public. This workshop has been a fruitful way to experiment in an interdisciplinary way with critical and creative interfaces that acted as discussion pieces by highlighting frictions.

References

de Lange, Michiel, and Martijn de Waal, eds. 2019. *The Hackable City: Digital Media and Collaborative City-Making in the Network Society*. London: Springer.

Dencik, Lina, Arne Hintz, and Jonathan Cable. 2016. "Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism." *Big Data & Society* 3, no. 2: 1–12. https://doi.org/10.1177/2053951716679678.

Dunn, Valerie, and Tom Mellor. 2017. "Creative, Participatory Projects with Young People: Reflections over Five Years." *Research for All* 1, no. 2: 284–99. https://doi.org/10.18546/RFA.01.2.05

Eisner, Elliot. 2008. "Art and Knowledge." In *Handbook of the Arts in Qualitative Research*, edited by J. Gary Knowles and Ardra L. Cole, 3–12. London: SAGE.

Foth, Marcus, Martin Brynskov, and Timo Ojala, eds. 2015. *Citizen's Right to the Digital City: Urban Interfaces, Activism, and Placemaking*. Singapore: Springer.

Foth, Marcus, and Glenda Amayo Caldwell. 2018. "More-than-Human Media Architecture." In *Proceedings of the 4th Media Architecture Biennale Conference, Beijing.*

Hjorth, Larissa, Anne M. Harris, Kat Jungnickel, and Gretchen Coombs. 2019. *Creative Practice Ethnographies*. Lanham, MD: Rowman and Littlefield.

Kara, Helen. 2020. *Creative Research Methods: A Practical Guide*. 2nd ed. Bristol: Policy Press.

Karimnia, Eleha. 2019. "Creative Methods for Investigating Urban Infrastructures." https://theatrum-mundi.org/library/creative-methods-for-investigating-urban-infrastructures/?pdf=1729.

Kelty, Christopher M. 2008. *Two Bits: The Cultural Significance of Free Software*. Durham, NC: Duke University Press.

Mattern, Shannon. 2021. *A City Is Not a Computer: Other Urban Intelligences*. Princeton, NJ: Princeton University Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Oosterloo, Serena, and Gerwin van Schie. 2018. "The Politics and Biases of the "Crime Anticipation System" of the Dutch Police." In *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems – Sheffield, United Kingdom, March 25, 2018*, edited by Jo Bates, Paul. D. Clough, Robert Jäschke, and Jahna Otterbacher, 30–41.

Powell, Alison B. 2021. *Undoing Optimization: Civic Action in Smart Cities*. New Haven, CT: Yale University Press.

Ratto, Matt, and Garnet Hertz. 2019. "Critical Making and Interdisciplinary Learning: Making as a Bridge between Art, Science, Engineering and Social Interventions." In *Critical Makers Reader: Unlearning Technology*, edited by Loes Bogers and Letizia Chiappini, 18–28. Amsterdam: Institute of Network Cultures.

Teli, Maurizio, Silvia Bordin, María Menéndez Blanco, Giusi Orabona, and Antonella De Angeli. 2015. "Public Design of Digital Commons in Urban Places: A Case Study." *International Journal of Human-Computer Studies* 81: 17–30. https://doi.org/10.1016/j.ijhcs.2015.02.003.

Vaart, Gwenda van der, Bettina van Hoven, and Paulus P.P.Huigen. 2018. "Creative and Arts-Based Research Methods in Academic Research. Lessons from a Participatory Research Project in the Netherlands." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 19, no. 2. https://doi.org/10.17169/fqs-19.2.2961.

van der Tuin, Iris, and Nanna Verhoeff. 2022. *Critical Concepts for the Creative Humanities*. Lanham, MD: Rowman and Littlefield. https://rowman.com/webdocs/CriticalConceptsCriticalHumanitiesOA.pdf.

van Schie, Gerwin, and Serena Oosterloo. 2020. "Predictive Policing in the Netherlands: A Critical Data Studies Approach." In *A Critical Approach to Police Science: New Perspectives in Post-Transitional Policing Studies*, edited by Veronika Nagy and Klára Kerezsi, 169–96. The Hague: Eleven International Publishers.

Verhoeff, Nanna, Sigrid Merx, and Michiel de Lange, eds. 2019. "Urban Interfaces: Media, Art, and Performance in Public Spaces." Special issue, *Leonardo Electronic Almanac* 22, no. 4. https://www.leoalmanac.org/urban-interfaces.

Verloo, Nanke, and Luca Bertolini, eds. 2020. *Seeing the City. Interdisciplinary Perspectives on the Study of the Urban*. Amsterdam: Amsterdam University Press.
Wakkary, Ron. 2021. *Things We Could Design: For More than Human-Centered Worlds*. Cambridge, MA: MIT Press.

About the Authors

Nanna Verhoeff is Professor of Screen Cultures and Society in the Department of Media and Culture Studies at Utrecht University. She initiated the research group [urban interfaces] and is co-lead of the Open Cities platform at Utrecht University. Her research on urban media combining perspectives from (digital) media and performance studies contributes to the interdisciplinary methodological development of the creative humanities.
>n.verhoeff@uu.nl

Sigrid Merx is an Associate Professor Theatre Studies in the Department of Media and Culture Studies at Utrecht University. Her research focuses on the dramaturgy of contemporary socially engaged performance practices, particularly in public spaces. She investigates how situated art and performance can act as productive sites to negotiate contemporary frictions, concerns, and debates in and about urban, public spaces.
> s.merx@uu.nl

Michiel de Lange is Assistant Professor at Utrecht University in the Department of Media and Culture Studies and co-PI in the project "Designing for Controversies in Responsible Smart Cities." He is co-founder of The Mobile City, a platform for the study of new media and urbanism; a core member of the [urban interfaces] research group at Utrecht University; and a researcher in the field of (mobile) media and urban culture.
> m.l.delange@uu.nl

15. Investigating the Datafied Society

Entrepreneurial Research as Approach

Mirko Tobias Schäfer, Karin van Es, and Iris Muis

Abstract
Humanities scholarship is essential in the present-day datafied society. This contribution discusses the interdisciplinary research platform Utrecht Data School (UDS) and its entrepreneurial research approach for investigating the impact of datafication and algorithmization on culture and society. This research approach is informed by close cooperation with external partners, including (local) government organizations, (public) media, companies, and NGOs and accelerates areas in which traditional academic research in the humanities have often said to fall short: societal engagement, knowledge transfer, and the application of research findings. However, as reflected on in the conclusion, it is not without its challenges.

Keywords: action research, societal impact, interdisciplinary, external partners, university

Humanities scholarship is essential in the present-day datafied society characterized by increasing datafication and the offloading of decision-making to inscrutable algorithms (van Dijck 2017; Underwood 2018). As Lisa Parks (2020, 644) points out, "digitization and datafication have extended and altered the kinds of knowledge media scholars need in order to research and teach in the field and prepare students for work and life beyond campus." Indeed, teaching and researching this transformation raises a series of challenges:

1. Relying on existing expertise and methods is not sufficient; we have to revisit our research questions and methods (van Es et al. 2021). Gaining

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH15

input from other disciplines such as law, ethics, sociology, and so forth is essential.
2. Access to data (and algorithms) is uneven. Moreover, its collection and analysis often require technical skill (boyd and Crawford 2012, 673–75).
3. Continued cuts to the humanities and increased grant competition have contributed to inequalities in the distribution of funds.
4. Studying datafication up close requires access to the development and use contexts of data and algorithms, but it often happens behind closed doors.

In response to these challenges, we established the interdisciplinary research platform Utrecht Data School (UDS). Here, we investigate how datafication and algorithmization impacts culture and society. This research is informed by our close cooperation with external partners, including (local) government organizations, (public) media, companies, and NGOs. These projects provide access to real-life data and generate qualitative data—pertaining data discourses and practices—acquired through ethnographic fieldwork.

Our action research approach enables us to connect our experience in analyzing cultural complexity, history, ethics, and media (practices) with digital methods and data analysis in the field (van Es and Schäfer 2017, 12). In working with external partners, we have found collaboration across disciplines essential. We therefore take an active role in the university's focus area Governing the Digital Society (bridging scholars in media studies, gender studies, law, ethics, public law, and governance with those in computer science and applied data science), various special interest groups, and AI labs. In these projects, we bring to the table our critical attitude toward media technologies, culture, and society. These transdisciplinary research processes are developed to facilitate mutual knowledge transfer and joint action between academics and professionals, between citizens and policy makers.

The "What" and "Why" of Entrepreneurial Research

UDS explores and develops tools for what has been termed "entrepreneurial research" (Schäfer 2018). It should not be mistaken for research *into* entrepreneurship (e.g., Ucbasaran et al. 2001; Perren and Ram 2004; Landström and Lohrke 2010) or academic entrepreneurship (e.g., Etzkowith 2003), which describes the utilization of research findings for commercial ends. Entrepreneurial research, by contrast, responds to demands in the field that

researchers want to investigate. We develop products and offer services that not only co-finance our research but also enable us to embed ourselves in the societal domains where we want to study the impact of datafication and algorithmization. The entrepreneurial practice accelerates areas in which traditional academic research in the humanities have often said to fall short: societal engagement, knowledge transfer, and the application of research findings (Schnapp 2017).

At Utrecht University and other institutions, the notion of open science is now intricately connected with socially engaged research and transdisciplinary efforts to respond to the great challenges of our time: digitization, pandemic, migration, and climate change (Miedema 2021). While traditional humanities research is often focused on conceptualizing societal phenomena and reflecting on them in the context of their respective disciplines' theoretical frameworks, entrepreneurial research is primarily engaged in describing and mapping societal phenomena, identifying practical problems, and producing applicable solutions. This requires building iterative research processes that are close to the application in order to engage in testing and optimization within the area of application. Utrecht Data School's practice shows multiple examples: an investigation of bias in search algorithms of job websites immediately led to changes (van Es et al. 2021), and an inquiry into algorithms and fundamental rights led to developing a Fundamental Rights and Algorithms Impact Assessment (FRAIA) and to courses for training government employees in applying the impact assessment. In these projects, we always involve students (e.g., as interns, research assistants, tutors, or thesis students) and recent graduates from our programs, who are hired as junior researchers.

As computer scientist Ben Shneiderman points out, research and societal impact can be even more effective by structurally combining applied and basic research (Shneiderman 2016). This approach has consequences for teaching, research, and public engagement. An example is our impact assessment for data projects, the Data Ethics Decision Aid (DEDA), which we started to develop in close cooperation with municipal data analysts starting in 2016. The paper we published on DEDA was not a mere concept, an idea of how to advance responsible data practices, but rather the result of developing, testing, optimizing, and implementing a workable impact assessment for data projects (see Franzke et al. 2021). Its development responded to municipalities' need for greater awareness of the ethical issues involved in data projects and the ability to communicate to stakeholders, representatives, and the public how data use in these projects has been deliberated.

DEDA fulfills a dual purpose by facilitating value-sensitive design and dialogical deliberation of data projects. It is currently used by various municipalities and educational organizations in the Netherlands. Importantly, and in line with our entrepreneurial approach, it is not just a practical tool but rather also enables participatory observation for studying organizations and their discourses on data and power (Siffels et al. 2022). These insights are used to engage as academics in a broader, critical examination of the issues raised by datafication. It thus contributes to scholarly discussions on data ethics.

Challenges Ahead

In conducting entrepreneurial research, maintaining academic independence is of utmost importance, which is why we have formulated criteria and guidelines for collaboration. Projects at UDS must align with our strategic research themes, promote academic freedom, and allow for academic publication. Most importantly, thanks to Utrecht University's funding of some basic operational expenses and to research time, we are financially independent of external partners. We therefore can, and sometimes do, decline projects. Having to rely solely on contract research would create unwanted financial dependencies. Furthermore, it would create an undesirable bias toward studying issues that emerge *in the field today*. So while entrepreneurial research is a useful approach to studying our contemporary datafied society, continued investment in basic research remains essential.

As UDS seeks to further strengthen the research-education-impact triangle through entrepreneurial research, there remain challenges ahead. Universities have yet to catch up with these new ways of working, since it requires new forms of support for financial and legal matters as well as job titles (e.g., project managers who also publish and teach), but it also demands rethinking how academics are recognized and rewarded (particularly in valuing teamwork over individualism). In the case of the latter, Utrecht University is already taking active steps as part of its Open Science Programme (Utrecht University 2021). Aside from the need for institutional transformation, this type of research also demands that we as researchers learn how to communicate with external parties about our research. Here, we must consider how findings can be translated into implementation and perspectives for action or policy. Rather than offer critical commentary from the sidelines, it offers an opportunity to co-shape a datafied society by supporting ethical and responsible data practices.

References

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5: 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Etzkowitz, Henry. 2003. "Research Groups as 'Quasi-Firms': The Invention of the Entrepreneurial University." *Research Policy* 32, no. 1: 109–21.

Franzke, Aline Shakti, Iris Muis, and Mirko Tobias Schäfer. 2021. "Data Ethics Decision Aid (DEDA): A Dialogical Framework for Ethical Inquiry of AI and Data Projects in the Netherlands." *Ethics and Information Technology* 23: 551–67.

Landström, Hans, and Franz Lohrke. 2010. *Historical Foundations of Entrepreneurship Research.* Cheltenham, UK: Edward Elgar Publishing.

Miedema, Frank. 2022. *Open Science: The Very Idea.* Dordrecht: Springer.

Parks, Lisa. 2020. "Field Mapping: What Is the 'Media' in Media Studies?" *Television and New Media* 21, no. 6: 642–49.

Pattnaik, Pinaki Nandan, and Satyendra C. Pandey. 2014. "University Spinoffs: What, Why, and How?" *Technology Innovation Management Review* 4, no. 12.

Perren, Lew, and Monder Ram. 2004. "Case-Study Method in Small Business and Entrepreneurial Research: Mapping Boundaries and Perspectives." *International Small Business Journal* 22, no. 1: 83–101.

Schnapp, Kai-Uwe. 2017. "Das Hamburger 'Projektbüro Angewandte Sozialforschung.'" *Zeitschrift für Politikwissenschaft* 27, no. 4: 529–47.

Shneiderman, Ben. 2016. *The New ABCs of Research: Achieving Breakthrough Collaborations.* Oxford: Oxford University Press.

Siffels, Lotje, David van den Berg, Mirko Tobias Schäfer, and Iris Muis. 2022. "Public Values and Technological Change: Mapping How Municipalities Grapple with Data Ethics." In *New Perspectives in Critical Data Studies: The Ambivalences of Data Power*, edited by Andreas Hepp, Juliane Jarke, and Leif Kramp, 243–65. Cham: Palgrave Macmillan.

Ucbasaran, Deniz, Paul Westhead, and Mike Wright. 2001. "The Focus of Entrepreneurial Research: Contextual and Process Issues." *Entrepreneurship Theory and Practice* 25, no. 4: 57–80.

Underwood, Ted. 2018. "Why an Age of Machine Learning Needs the Humanities." *Public Books*, December 5, 2018. https://www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities.

Utrecht University. 2021. "Recognition and Rewards Vision." Utrecht University. https://www.uu.nl/sites/default/files/UU-Recognition-and-Rewards-Vision.pdf.

van Dijck, José. 2017. "Foreword." In *The Datafied Society: Studying Culture through Data*, edited by Mirko Tobias Schäfer and Karin van Es, 11–12. Amsterdam: Amsterdam University Press.

van Es, Karin, Daniel Everts, and Iris Muis. 2021. "Gendered Language and Employment Web Sites: How Search Algorithms Can Cause Allocative Harm." *First Monday* 26, no. 8. https://doi.org/10.5210/fm.v26i8.11717.

van Es, Karin, Mirko Tobias Schäfer, and Maranke Wieringa. 2021. "Tool Criticism and the Computational Turn: A 'Methodological Moment' in Media and Communication Studies." *M&K. Medien and Kommunikationswissenschaft* 69, no. 1: 46–64.

van Es, Karin, and Mirko Tobias Schäfer. 2017. "Brave New World." In *The Datafied Society: Studying Culture through Data*, edited by Mirko Tobias Schäfer and Karin van Es, 13–24. Amsterdam: Amsterdam University Press.

About the Authors

Mirko Tobias Schäfer is Associate Professor in Utrecht University's research area *Governing the Digital Society* and the Department for Information and Computing Sciences. He is co-founder and Faculty of Science Lead of the Utrecht Data School.
> m.t.schaefer@uu.nl

Karin van Es is Associate Professor of Media and Culture Studies at Utrecht University and is project lead Humanities at Utrecht Data School. Her main research concerns the datafication and platformization of culture and society.
> k.f.vanes@uu.nl

Iris Muis has a background in law and international relations. Since 2015, she has been working at Utrecht Data School (UDS) of Utrecht University, specializing in data ethics and its implementation in public management. Iris also coordinates UDS's cooperation with external partners.
> i.m.muis@uu.nl

16. Big Data and the Global South

A Case for Dialogue

Bruce Mutsvairo

Abstract
This contribution identifies and critiques the role that dialogue could play as an important tool in bridging the gap between Northern and Southern imaginaries, perceptions and realities on the presence of data mining corporates in Africa, Latin America, the Arab World, Asia and other regions of the world that do not associate themselves with the highly industrialized West. Coming at a time when cynicism over the presence of Silicon Valley and other major data players in Africa and other non-Western regions of the world has been growing, a stakeholders' dialogue to openly discuss and conscientize all parties on how and why the presence of these major tech players may be deemed unwelcome by locals and other observers.

Keywords: Global South, dialogue, data, inequalities, technology

The data revolution, which has largely been restricted to the industrialized countries, is starting to find its way into the so-called developing world market. For example, several cities across the Global South are adopting the "smart city" concept, which is centered on the use of large-scale data analytics with the aim of improving quality-of-life standards while also achieving sustainable development. The South African city of Cape Town has gone a step further, adopting big data solutions to boost its wildfire assessment and emergency response systems. Confronted by a burdensome emerging techno-economic culture, which is enabling datafication (Schäfer and van Es 2017), surveillance capitalism (Zuboff 2018) and data colonialism (Couldry and Mejias 2019), we need cool heads and strong leadership. We need to realize how much we can achieve by listening to each other.

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH16

Dialogue should be initiated to understand the disparities. There are no quick solutions, but dialogue is a good way to start.

The purpose of this essay is to make a case for dialogue as a way forward in identifying workable ways to deal with global challenges emerging from long-standing economic and social disparities necessitated by persistent technological developments, including big data. The term "Global South" has steered a significant amount of debate in some academic quarters. I do not intend to go into this debate, because, for me, "Global South" is not about geographical determinism but is instead taking a cue from Mahler (2017; 2018). I see it as a concept that represents determination and solidarity among previously dominated groups to scale up efforts to earn social and political agency recognizing their shared struggles, knowledge, and independence. Equally, there is plenty of debate on how such an objective can be achieved and that's where dialogue comes in handy. In an era in which calls for decolonizing everything, including data (see Quinless 2022), are scattered everywhere in academia, deciding how it can be done is no easy task. If you ask me, that's where the idea of dialogue reigns supreme. A white colleague and friend who teaches African history was attacked by students at a Dutch university, because students were adamant a "white man should not teach African history." When I invited a white friend to a panel on "decolonizing technologies in Africa," a colleague demanded answers, asking me to "decolonize" by choosing from a pool of Black academics instead. These examples more or less explain why I have chosen to be a flagbearer of dialogue. Dialogue to me simply means having a conversation. It means that, despite how gifted or assured you are about your cause or idea, you need to listen to others. That sounds very idealistic. Maybe. But if you spend some time working in conflict-marred societies like myself, perhaps you understand why dialogue is crucial.

One recurring argument I have heard each time I talk about dialogue is that it does not always guarantee meaningful change. I always teasingly reply by saying: neither do violence or hate. The point is that there are academics and activists who do not seem to see the irony of using racism to fight racism by thinking that the best way to correct historical injustices is by seeing everything in terms of color. Hate cannot heal hate. Indeed, it takes time to achieve meaningful change. Sometimes absolute change is impossible to achieve. South Africa is the (relatively) peaceful nation it is today because of the Truth of Reconciliation Commission (TRC) introduced by Nelson Mandela in 1995. Of course, there will always be people who feel that perpetrators of violence did not deserve amnesty for the violations they committed during the apartheid era, but sometimes you need to make

brave and difficult decisions for the sake of progress. The TRC represents how much dialogue, regardless of circumstances, can achieve. By proposing dialogue, I am not in any way claiming that other measures, some of them quite radical, should be discounted. I am simply proposing something that is supported by my inherent belief that, when people listen to each other, there is so much that can be achieved. The remainder of this piece will discuss the state of affairs regarding datafication and data practices in the Global South, including the opportunities and challenges that the Global South faces and how dialogue may aid in (the equity of) these developments.

Opportunities and Challenges

Concepts such as "big data for development" are hogging the limelight, particularly in non-Western societies where policymakers are debating ways through which big data could help identify, gain, and deepen a better understanding of global development and humanitarian challenges. This way, it is hoped that the open data movement could play a decisive role in ensuring poorer nations' determination to meet the UN's Sustainable Development Goals by the 2030 deadline. The major problem, as highlighted by Cohen and Kharas (2018), is that finding top-notch, timely, easily accessible data comes at a cost in most poorer countries, which are nevertheless considered the main targets for Western developmental endeavors. This notable challenge hasn't stopped the massive private and public sector investments in big data and other technologies across the Global South. Despite its corporate connections, studies have shown big data could help improve social change in poorer countries (Bellagio 2014). More recent studies have, however, criticized such notions. For example, Brevini (2021) argues that many countries in the Global South should not embrace artificial intelligence and big data due to environmental concerns, as judiciously elaborated in her latest book.

Data concerns do not end there. There are a host of challenges that need to be unpacked if we are to see the real potential of big data in developing markets. These range from poor connectivity trends (Mutsvairo 2019) to privacy and ethical concerns (Simo 2015) regarding the possible abuse of people's data. While China's Huawei is a dominant force in the African telecommunications sector, its big-scale investments in big data and surveillance technologies across the continent is worrying activists, who fear autocratic regimes will target them to dissuade dissent. The Chinese government makes no apology for its expanding investments in datafied

technologies. Thanks to big data, high school students' behavior is monitored in class (Chan 2018), while a controversially intrusive credit-score ranking system (Campbell 2019) is taking shape in the Asian giant of over 1 billion people. From this standpoint, it is clear there is an urgent need to broadly promote transnational discussions and dialogues on data-related issues so that monopolization and accountability issues can be prioritized. Many of the aforementioned societies are extremely unequal, which has led to fears that the presence of, or investments in, big data could further perpetuate the stridently growing inequalities.

Simultaneously, the number of universities in the Global South offering postgraduate degrees in data science, big data, or business analytics has increased over the last five to ten years. These include the Asian Institute of Management in the Philippines, Goa Institute of Management in India, the University of Malaya in Malaysia, Princess Nourah bint Abdulrahman University in Saudi Arabia, Covenant University in Nigeria, University of Pretoria in South Africa, Egade Business School in Mexico, among several others. Deeper partnerships with institutions in the Global North could stimulate knowledge on different data cultures among institutions and countries, possibly leading to equitable benefits from intricate advances in technology.

Ways Forward

Dialogue is the best way forward, because data actors including academics working in this sub-field have plenty to benefit from by listening to each other. Why would an edited book collection claiming to uncover global trends on data or technology only cover those from Europe and North America, for example? It points to the missing dialogue between scholars working in different regions of the world. Dialogue seems to be missing among private sector and governmental actors, too. Consider the fact that some of the countries that are being considered prime destinations of big data investments do not even have the requisite infrastructures conducive for such technocentric innovation. Some of these countries are even struggling to feed their own citizens, yet they seem to see it fit to receive data investments. Their priority should be on providing food, and if big data could help in that endeavor, then it should surely be promoted. Not to be outdone is certainly the lack of data science and other data-related training across the Global South. Only a few institutions of higher learning, some of which have been identified in this piece, are offering

university-level education on data related issues. It is important to invest in educating the local population first if long-term attributes of big data are to be comprehended, as the importance of homegrown skilled labor cannot be underestimated.

The London School of Economics' (LSE) joint MA in Global Media and Communications with the University of Cape Town is an example of how universities located in different economic regions of the world could collaborate and promote sustainable dialogue on matters related to our field. As tomorrow's leaders, students enrolled on such programs will certainly continue to question why engineers rush to develop new technologies before ensuring that the societal problems presented by old and current technologies, including the digital divide and digital inequalities, are stridently dealt with. We cannot expect societies to become equal tomorrow. They will probably never become completely equal. But the enormity of the challenges technology poses requires a collective approach, and that's where dialogue matters. If we want to have affordable and accessible internet in poor nations, we need to sit together and have a conversation on how that can be achieved. If data can transform societies, it is important to underscore its relevance to the common man and woman, and it only takes a conversation to see how that endeavor can be achieved.

References

Bellagio Big Data Workshop Participants. 2014. "Big Data and Positive Social Change in the Developing World: A White Paper for Practitioners and Researchers." *Oxford Internet Institute*. https://www.rockefellerfoundation.org/report/big-data-and-positive-social-change-in-the-developing-world.

Brevini, Benedetta. 2021. *Is AI Good for the Planet?* Cambridge, UK: Polity.

Campbell, Charlie. 2019." How China Is Using 'Social Credit Scores' to Reward and Punish Its Citizens." *Time*. https://time.com/collection/davos-2019/5502592/china-social-credit-score.

Chan, Tara Francis. 2018. "A School in China is Monitoring Students with Facial-Recognition Technology That Scans the Classroom Every 30 Seconds." *Business Insider*, May 21, 2018. https://www.businessinsider.com/china-school-facial-recognition-technology-2018-5.

Cohen, Jennifer L., and Homi Kharas. 2018. "Using Big Data and Artificial Intelligence to Accelerate Global Development." *Brookings Institute Report*, November 15, 2018. https://www.brookings.edu/research/using-big-data-and-artificial-intelligence-to-accelerate-global-development/.

Couldry, Nick, and Ulises Ali Mejias. 2019. *The Costs of Connection: How Data Colonizes Human Life and Appropriates it for Capitalism.* Stanford, CA: Stanford University Press.

Mahler, Anne Garland. 2018. *From the Tricontinental to the Global South: Race, Radicalism, and Transnational Solidarity.* Durham, NC: Duke University Press.

Mahler, Anne Garland. 2017. "Global South." In *Oxford Bibliographies in Literary and Critical Theory.* Edited by Eugene O'Brien. Oxford: Oxford University Press.

Mutsvairo, Bruce. 2019. "Challenges Facing the Development of Data Journalism in Non-Western Societies." *Digital Journalism* 7, no. 9: 1–6.

Quinless, Jacqueline M. 2022. *Decolonizing Data: Unsettling Conversations about Social Research Methods.* Toronto: University of Toronto Press.

Schäfer, Mirko Tobias, and Karin van Es. 2017. *The Datafied Society: Studying Culture through Data.* Amsterdam: Amsterdam University Press.

Simo, Hervais. 2015. "Big Data: Opportunities and Privacy Challenges." Working paper. https://arxiv.org/pdf/1502.00823.pdf.

Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: Public Affairs.

About the Author

Bruce Mutsvairo is Professor in Media, Politics and the Global South in the Department of Media and Culture Studies at Utrecht University, where he studies developments in data, democracy, and digital media across non-Western ecosystems.
> b.mutsvairo@uu.nl

17. Situating the Marketization of Data

Anne Helmond and Fernando van der Vlist

Abstract

Data are neither inherently valuable, nor do all data have the same value. This contribution argues how data are *made* useful and valuable to specific actors and for specific purposes. It draws attention to the material politics of data flows and valuation, and to the many different actors and stakeholders who build the technological conduits and pipelines that facilitate the circulation and use of data. Therefore, it highlights the need to study the *infrastructural layer* of the global data market, as well as the central role of *intermediaries* who build and uphold these infrastructures for the exchange and use of data for different purposes. Both are important to situate the processes of datafication and data marketization in specific empirical settings.

Keywords: data markets, data intermediaries, marketization, platform infrastructure, partnerships, digital marketing and advertising

Many critical media researchers, technology journalists, and activists have warned about the potential risks and harms of data aggregation and abuses of data by "bad actors," companies, law enforcement agencies, or states. While social media platforms, mobile apps, advertising companies, and data brokers emphasize that the pseudonymous data they collect cannot be traced back to real persons, recent cases have shown how easy it still is to do this.

VICE reported that location data and mobile device data were purchased from a commercial data broker to track the locations of a priest, outing him as gay through his use of the gay/bi/trans/queer dating app Grindr (Cox 2021b). Similarly, government agencies in the US have been known to buy mobile location data from commercial data brokers without warrants for various law enforcement purposes (Guariglia 2020). This commercially available

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH17

data contains unique identifiers, such as mobile device or advertising IDs, that cannot be linked to individuals directly. However, there are many actors in the global data market who offer tools, products, and services to help link and de-anonymize this type of data. This has led to an enormous industry of companies that connect pseudonymous identifiers to a wealth of information obtained from disparate sources, including people's real names, e-mail and home addresses, phone numbers, or credit data. When linked, this information can be used to identify and target individuals or groups of people, thus "shattering" their anonymity (Cox 2021a).

While many of these so-called "data brokers" operate in the shadows, we have learned a lot about them over the years from the critical investigations of many researchers, journalists, and activists (e.g., Beer 2018; Braun 2013; Crain 2021; Christl and Spiekermann, 2016; Lechardoy et al. 2020; Mellet and Beauvisage 2020; Nadler et al. 2018; Zuboff 2019). Additionally, we have done empirical and historical research ourselves into the role of business partners and software infrastructure development in the data economy, which we summarize below (Helmond, Nieborg and van der Vlist 2019; van der Vlist and Helmond 2021; van der Vlist et al. 2022). This research has surfaced some of the key actors, techniques, and technological systems, as well as the material conditions and relations of data as they fuel the advertising-based business models; data-driven business operations; and AI-based tools, products, and services of the contemporary internet. We have shown how the collection, processing, circulation, and use of data impact power relations and raise issues and concerns around the critical political economy of data and data flows.

We have been devising ways of situating data not only in terms of their production contexts but also in terms of their subsequent aggregation, processing, circulation, and use by many different types of users—and often for purposes other than originally intended. Despite what many believe, data are neither inherently valuable nor does all data have the same value. Instead, data are *made* useful and become valuable to specific actors and for specific purposes. Therefore, we draw attention to the materiality and politics of data flows and data valuation and to the many different (intermediary) actors and stakeholders who build the technological conduits and pipelines—or infrastructures—that facilitate the circulation and use of data. By situating the marketization of data in terms of the constitutive actors and infrastructures, we can thus put the opaque global data market in place and in context.

Firstly, our approach highlights the *infrastructural layer* of the global data market. These infrastructures for the exchange and use of data are built by

developers who use application programming interfaces (APIs) to develop data integrations and software applications "on top" of digital platforms. In the global data market, API-based connections between software systems function as the pipelines that enable the circulation and use of data and services between different software platforms and companies. These conduits, once they are built, give other companies and partners the ability to connect, control, and activate data in their own tools, products, and services. We have shown in a large-scale empirical study how this technological infrastructure of API-based integrations between thousands of companies worldwide both provides and *governs* the material conduits for contemporary "programmatic advertising," a multibillion-dollar market that relies on the global data market. With this infrastructure in place, ads and audience commodities are automatically ("programmatically") traded on ad exchanges and served across many media distribution channels and geographic regions in mere milliseconds through real-time bidding auctions. This whole process unfolds in the background each time a consumer opens a web page or uses an app. However, this digital advertising infrastructure also comes with serious risks and harms to society and can be "weaponized by political and anti-democratic actors" to influence political decisions (Nadler et al. 2018), to discriminate, or to otherwise violate people's digital rights (e.g., EDRi 2021).

By identifying who has integrated with, or has built on top of a platform's APIs, we can trace the channels that exist to circulate and use data. Many of these channels are interlinked to enable automated uses and exchanges of data on a large scale, across countries and continents, including through tracking and targeted advertising, and remain an opaque infrastructure for most consumers.

Furthermore, by closely investigating how APIs are designed and structured, we have examined in detail how digital platforms datafy people and their activities as data entities that can be identified and targeted through their associated data fields (e.g., name, birthday, home address, interests, etc.) and connections (e.g., friends, groups, likes, videos, etc.). We traced how data entities such as the "user" have changed and evolved over the years and discovered that Facebook removed sensitive data fields because of ongoing social and regulatory pressures from civil rights organizations and journalists. After Facebook removed data fields related to a user's dating preferences, relationship status, political and religious interests, and friend lists from its Graph API, we found that it kept these data fields available in its Marketing API for advertising and marketing developers for many more years.

APIs not only enable third parties to build applications and services but also provide a powerful means of "infrastructural control" for platform owners to govern, with increasing precision, who is and is not allowed to access data and under which requirements. For example, we observed how Facebook's popular Graph API evolved from a simple interface for data retrieval in the mid-2000s into an increasingly complex and layered "governance arrangement" of (data) access controls, application permissions, app review guidelines, and terms and policies. Additionally, while some "open" APIs are openly available to everyone, the APIs required for digital marketing and advertising are typically governed through special partner programs. Only selected and approved business partners are allowed to access or use platforms' data or to integrate with a platform's technological infrastructure for business purposes. This partnership strategy has been vital for platforms to be embedded in markets and industries other than their own and has led to a complex global data market comprising many interconnected actors and infrastructures.

Secondly, our approach highlights the central role of *intermediaries*, or those who build and uphold these infrastructures of data and automation for different purposes. These are the actors and stakeholders who are doing the practical work of connecting, aggregating, and modeling data from multiple sources (e.g., social media, mobile devices and apps, etc.) and make them available for further uses and users (Beer 2018). Data brokers, data marketplaces, data analytics companies, advertising networks and exchanges, and data management platforms (which enrich advertising bids with tracking data) are all examples of intermediaries in the global data market. Many of these intermediaries are also key "nodes" in the conduits of the global advertising market, enabling others to use their data for digital marketing and advertising purposes. It is the intersection of the global advertising market and the data market that enabled the "inevitable weaponization" of location and app data from Grindr (Cox 2021b).

Additionally, it is important to study these intermediaries to learn how they actually make data useful and valuable. Research on data "assetization" reminds us how it is not the data themselves, but rather the "users" and their "engagement" that are turned into assets. As Birch, Cochrane, and Ward observe, "'users', 'usage', and 'access to users' end up as the legible techno-economic objects that Big Tech can value as future revenue streams through different monetization strategies" (2021, 11; cf. van Doorn and Badger 2020). When advertising online, data generally do not leave "data silos"

like Facebook or Google (as they are called in the industry); rather, these platforms provide interfaces that give customers "access to users."

Popular "identity resolution" services from data partners such as Salesforce, LiveRamp, FullContact, Lotame, and many other companies have become key intermediaries in the global data market. They not only connect and aggregate (audience) data from multiple sources but crucially also make these data available for further uses and users "across the ecosystem." These services typically enrich data sources with additional attributes, such as email addresses, mobile advertising IDs, postal addresses, phone numbers, online or offline purchases, or voting data, to enrich or verify persistent profiles for real persons. Indeed, these services also enabled the de-anonymization of purchased location data that ousted the gay priest using Grindr.

It is common practice for digital platforms or apps to share user data with third parties, including advertising partners, service partners, and social media partners. These data-sharing practices are typically documented in privacy policies. Grindr, for instance, shares device IDs, advertising IDs, and location data with its advertising partners (Grindr 2022). Many of these advertising partners in turn also state in their policies that they share data with third parties, including with their own partners. In short, these partnership strategies are critical to the global data market's functioning and risks. While Grindr's and other companies' policies state that no personally identifiable account information is shared, we know that "identity resolution" services may be used to piece the information together nonetheless and render it personal data, subject to strict regulations like the GDPR.

Even though many of the described data aggregation and sharing practices are forbidden without a user's consent under recent consumer and privacy laws, especially the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), there are plenty of examples of advertising technology companies who have been breaching these legislations as well as complaints from privacy organizations about a lack of and slowness of enforcement efforts by regulatory bodies (Burgess 2022; ICCL 2021; Lomas 2022). The critical industry practice of real-time auction bidding has been found to violate the GDPR because of the industry's inability to trace personal data "behind the scenes" when it passes through the invisible infrastructures of the data market (Ryan and Santos 2022; Veale and Zuiderveen Borgesius 2022). Meanwhile, Apple and Google have begun deprecating the use of third-party cookies and mobile device identifiers (i.e., Identifier for Advertisers [IDFA) on iOS and advertising ID [AdID] on

Android) in the name of consumer privacy. This process is directly impacting the larger ecosystems of actors and technologies relying on Apple and Google's platforms. It has also increased the use of "first-party data" and identity resolution services and has led to the creation of new and competing types of identifiers in the industry (van der Vlist and Helmond 2021).

Taken together, critical perspectives on the technological infrastructures and intermediaries of the global data market enable critical empirical contributions that help us understand the many roles, risks, and harms of data in society. It offers new ways of situating the processes of datafication and marketization in empirical settings. Furthermore, it provides important insights and evidence to help stakeholders, policymakers, and regulators worldwide grapple with the challenges of governing data markets.

References

Beer, David. 2018. "Envisioning the Power of Data Analytics." *Information, Communication & Society* 21, no. 3: 465–79. https://doi.org/10.1080/1369118X.2017.1289232.

Birch, Kean, D.T. Cochrane, and Callum Ward. 2021. "Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech." *Big Data & Society* 8, no. 1: 1–15. https://doi.org/10.1177/20539517211017308.

Braun, Joshua. 2013. "Transparent Intermediaries: Building the Infrastructures of Connected Viewing." In *Connected Viewing: Selling, Streaming, and Sharing Media in the Digital Age*, edited by Jennifer Holt and Kevin Sanson, 134–53. New York: Routledge. https://doi.org/10.4324/9780203067994.

Burgess, Matt. 2022. "How GDPR Is Failing." *WIRED*, May 23, 2022. https://www.wired.com/story/gdpr-2022.

Christl, Wolfie, and Sarah Spiekermann. 2016. *Networks of Control: A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*. Vienna: Facultas Universitätsverlag.

Cox, Joseph. 2021a. "Inside the Industry That Unmasks People at Scale." *Motherboard* (*VICE*), July 14, 2021. https://www.vice.com/en/article/epnmvz/industry-unmasks-at-scale-maid-to-pii.

Cox, Joseph. 2021b. "The Inevitable Weaponization of App Data is Here." *Motherboard* (*VICE*), July 21, 2021. https://www.vice.com/en/article/pkbxp8/grindr-location-data-priest-weaponization-app.

Crain, Matthew. 2021. *Profit Over Privacy: How Surveillance Advertising Conquered the Internet*. Minneapolis, MN: University of Minnesota Press.

EDRi (European Digital Rights). 2021. "How Online Ads Discriminate." https://edri.org/our-work/how-online-ads-discriminate.

Grindr. 2022. "Privacy and Cookie Policy." https://www.grindr.com/privacy-policy/third-parties.

Guariglia, Matthew. 2020. "Law Enforcement Purchasing Commercially-Available Geolocation Data is Unconstitutional." *Electronic Frontier Foundation*, December 3, 2020. https://www.eff.org/nl/deeplinks/2020/12/law-enforcement-purchasing-commercially-available-geolocation-data.

Helmond, Anne, David B. Nieborg, and Fernando van der Vlist. 2019. "Facebook's Evolution: Development of a Platform-As-Infrastructure." *Internet Histories: Digital Technology, Culture and Society* 3, no. 2: 123–46. https://doi.org/10.1080/24701475.2019.1593667.

ICCL (Irish Council for Civil Liberties). 2021. "Europe's Enforcement Paralysis: ICCL's 2021 GDPR Report." https://www.iccl.ie/digital-data/2021-gdpr-report.

Lechardoy, L., A. Sokolyanskaya, and F. Lupiáñez-Villanueva. 2020. "Transparency in the Business-to Business Commercial Relations in the Online Advertising Market." *Observatory on the Online Platform Economy* 3 (December 25, 2020). https://platformobservatory.eu/news/analytical-paper-transparency-in-the-business-to-business-commercial-relations-in-the-online-advertising-market.

Lomas, Natasha. 2022. "Behavioral Ad Industry Gets Hard Reform Deadline after IAB's TCF Found to Breach Europe's GDPR." *TechCrunch*, February 2, 2022. https://social.techcrunch.com/2022/02/02/iab-tcf-gdpr-breaches.

Mellet, Kevin, and Thomas Beauvisage. 2020. "Cookie Monsters: Anatomy of a Digital Market Infrastructure." *Consumption Markets and Culture* 23, no. 2: 110–29. https://doi.org/10.1080/10253866.2019.1661246.

Nadler, Anthony, Matthew Crain, and Joan Donovan. 2018. "Weaponizing the Digital Influence Machine." *Data & Society*, October 17, 2018. New York: Data and Society Research Institute. https://datasociety.net/library/weaponizing-the-digital-influence-machine.

Ryan, Johnny, and Cristiana Santos. 2022. "An Unending Data Breach Immune to Audit? Can the TCF and RTB Be Reconciled with the GDPR?" SSRN Scholarly Paper, March 23, 2022. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4064729.

van der Vlist, Fernando N., and Anne Helmond. 2021. "How Partners Mediate Platform Power: Mapping Business and Data Partnerships in the Social Media Ecosystem." *Big Data & Society* 8, no. 1: 1–16. https://doi.org/10.1177/20539517211025061.

van der Vlist, Fernando N., Anne Helmond, Marcus Burkhardt, and Tatjana Seitz. 2022. "API Governance: The Case of Facebook's Evolution." *Social Media + Society* 8, no. 2: 1–24. https://doi.org/10.1177/20563051221086228.

van Doorn, Niels, and Adam Badger. 2020. "Platform Capitalism's Hidden Abode: Producing Data Assets in the Gig Economy." *Antipode* 52, no. 5: 1475–95. https://doi.org/10.1111/anti.12641.

Veale, Micheal, and Frederik Zuiderveen Borgesius. 2022. "Adtech and Real-Time Bidding under European Data Protection Law." *German Law Journal* 23, no. 2: 226–56. https://doi.org/10.1017/glj.2022.18.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

About the Authors

Anne Helmond is Associate Professor of Media, Data and Society at Utrecht University. She is part of the focus area "Governing the Digital Society," where she examines the processes of platformization, algorithmization, and datafication from an empirical and historical perspective.
> a.helmond@uu.nl

Fernando van der Vlist is a postdoctoral researcher of media and digital society at Utrecht University and with the Collaborative Research Centre "Media of Cooperation" at the University of Siegen (funded by the German Research Foundation, DFG). He is part of Utrecht University's focus area "Governing the Digital Society" and teaches at the Media and Culture Studies department.
> f.n.vandervlist@uu.nl

Index of Concepts