Collecting Diversity
Data and Citizen Science at the Museum für Naturkunde Berlin

Chiara Garbellotto, Tahani Nadim

Natural history museums are purveyors of diversity. Indeed, by way of taxonomy—the description, identification and naming of species—natural history museums actively produce the diversity they disseminate. Natural history museums generally embrace the position that diversity, or, more precisely, biodiversity, is vital for the health of the planet. With the alarming decline of biodiversity worldwide, natural history collections are experiencing renewed attention. A recent report for the International Council of Museums (ICOM) emphasises their central role in the 'conservation of global biodiversity': 'Understanding what species live where is a foundation of understanding biodiversity and nature conservation. Specimen labels provide basic information on what species occur where, or at least where they once occurred' (McGhie 2019: 15).

The passage from the ICOM report outlines two interdependent issues that sit at the heart of this chapter. First, biodiversity draws its meaning and practice from both biological and sociological categories. Sometimes biodiversity is used to denote the diversity of life in all its biological, cultural and social forms, which accords with the definition of biodiversity by the Convention on Biological Diversity (CBD). At other times, it refers exclusively to the diversity of non-human species. The ambiguity of the term, which indicates both diversity and biodiversity, captures the interpenetration of the biological and the social, nature and society. The criss-crossing of these concepts is particularly central for understanding the duality of bio/diversity. As Subramaniam observes, 'western societies (at least since Darwin) have struggled with the biological questions of variation alongside political questions of diversity and difference' (Subramaniam 2014: 10).

The second issue relates to the labels and other data points typically attached to museum objects providing proper names, taxonomic designation and provenance. For biodiversity sciences, such data are crucial for creating and managing biodiversity, especially amid increasing digitisation and the use of data-based tools. The history of science has shown how the nature and value of diversity have been co-produced by the material-semiotic devices used to record it—taxonomic tables, indices, double-entry book-keeping, museum collection designs, drawers and the like.[1] How we document diversity is

thus co-constitutive of how we cognitively, politically and experientially apprehend it.[2] Consequently, the data practices that have emerged in the past 10 years are critical for understanding the relationships between the biological and the social that implicate 'diversity'.

What difference do digital data make when it comes to the nature of bio/diversity? In posing that question, we seek to go beyond the standard issues of access, speed and preservation and attend instead to the specific relations—between people, objects, politics, environments and institutions—that are compelled, maintained or, at times, severed in data practices. To do so, we examine a particular case of citizen science that exemplifies the evolution of museum digitisation efforts. *Forschungsfall Nachtigall*, a project devoted to the study of nightingales, was initiated by the Museum für Naturkunde Berlin to encourage public participation in science. The project is an example of lay people and scientists working together in the production of data and scientific knowledge. It also shows how biodiversity—in this case, that of nightingales—can be a vehicle for social diversity through the inclusion of non-scientists ('citizens') in the research process. In this chapter we speak of 'data creatures' to denote both nightingales and citizens made in and through the data practices designed to capture biodiversity and social diversity. We borrow the term from Jennifer Gabrys's work on citizen sensing. According to Gabrys, 'citizen-sensing' projects are processes of '*creaturing data*, where actual environmental entities that come together are creations that materialize through distinct ways of perceiving and participating in environments' (2017: 13). In our chapter, the term 'data creature' serves a two-fold purpose: it foregrounds the making of data in the constitution of bio/diversity and focuses on the coeval becoming of entities and distinct ways of assembling (in) data.

While histories of science have shown how lay knowledges have always been a core part of scientific knowledge-making, the term citizen science has gained spectacular traction in recent years (Kimura and Kinchy 2016; Strasser et al. 2019). Knowledge and memory institutions as well as various sciences have responded to calls for more participation and transparency by enrolling lay people through specifically designed research projects, from mapping galaxies to deciphering historical shipping log books. But these developments have also sparked critical responses, triggered not least by the reduction of participation to data gathering and the lack of critical reflection of citizen science projects on the exclusions, politics and economics involved. Questions about the particular form of 'citizen' or 'science' in use, the struggles over what constitutes legitimate knowledge and expertise and the underlying institutional interests are usually not part of the project design. Our chapter addresses some of these problems by following data practices of the *Forschungsfall Nachtigall* project, from nightingales in Berlin public parks to digital spectrograms in data workshops and the Museum's Animal Sound Archive. We use the notion of 'data creatures' to propose a particular constellation of citizen and natural history object.

The natural history museum and bio/diversity

Forschungsfall Nachtigall began in 2018 as a two-year research project funded by the Federal Ministry of Education and Research (BMBF) as part of a programme to increase cooperation between science and 'citizens'. The nightingale project aimed to study nightingales' 'dialects' and habitat requirements in Berlin (2018) and Germany (2019) by generating a map of nightingale locations and collecting recordings for bioacoustic analyses. Both the map and the recordings were to be created through the collective efforts of volunteers who would venture into Berlin late at night and in the early morning to capture nightingale songs on their smartphones.

The drive to include mainstream citizens in scientific research has been a central concern of European research policy for the past several years. Horizon 2020, a framework programme governing Europe's research priorities, set aside explicit funding for efforts to involve civil society in research and innovation. In their guidelines for the programme, the German government states that science and innovation need to be 'closer to the people' to ensure 'public transparency and acceptance of scientific processes' (Federal Government Position Paper 2017: 14). It further stresses that 'digital transformation…offers new possibilities for exchange and communication', which can ultimately contribute to 'reawakening people's interest in Europe' (Devictor and Bensaude-Vincent 2016: 8). Citizen science is a tool to address three distinct contemporary issues: the apparent 'disconnect' between Brussels and the 500 million people that live in the European Union; the increasing scepticism towards scientific methods and results; and the yet-to-be realised potenial of digital technologies.

The government guidelines require that citizen science 'be distinguished from civic participation, which refers to the possibility of (political) participation in political decisions and processes, for example to further develop our model of society' (Federal Government Position Paper 2017: 14 fn. 6). The disclaimer is curious because the very name 'citizen science' makes an explicit appeal to the political. Moreover, it indirectly references the untenable division between science and society. It is this tension that we want to address when situating bio/diversity in museum data, not least because museums have become a key site for doing and promoting citizen science. The Museum für Naturkunde Berlin was one of the founding members of the *Bürger schaffen Wissen—Wissen schafft Bürger (GEWISS)* platform, a portal and advocacy programme for German-based citizen science projects. It also hosts ECSA, the European Citizen Science Association, and runs its own citizen science projects such as the nightingale project.

This chapter is based on ethnographic fieldwork carried out by Chiara Garbellotto, who accompanied parts of the nightingale project as a participant observer, and on research carried out by Tahani Nadim on 'data creatures' (Nadim 2022). In what follows, we present this citizen science project before focusing on nightingales and citizens as forms of data creatures produced by data practices. Following the work of feminist science & technology studies at the intersection of sociology and cultural anthropology, the chapter contributes to debates concerned with tracking the interpolations of nature and culture (Franklin, Lury, and Stacey 2000; Haraway 2018) and with the infrastructural powers of datafied classification systems (Bowker and Star 2000).

The nightingale project

The biologists, PhD students and other museum team members with whom the first author spoke described the nightingale project as a 'modular network' comprising different activities, research practices and foci. The project had three main components. The first was an education programme designed for seventh- and eighth-graders in Berlin schools to foster awareness of conservation, encourage direct contact with nature and assess students' knowledge of nightingales. This assessment consisted of questionnaires before, immediately after and eight weeks after scientist-led excursions into Berlin's urban nature. In addition, a second set of questionnaires targeted the correlation between individual learning and motivation and the provision of information about the project's research and context.

The second component consisted of an artistic exploration of the nightingale's symbolism in different cultures. It was accompanied by the collection of biographical anecdotes and memories related to the birds from a second cohort of project participants, which the project team variously referred to as 'refugees', 'new citizens' and 'new Berliners'. (On the use of essentializing categories, see Gram and Tinius in this volume). In the first year, the collection was expanded to include other participant contributions such as written stories and recorded audios transcribed by the project team. The work revolved around activities such as crafting and sewing bird puppets, live music, cross-stich and poetry performances. A map of Berlin designed and sewn by the members of the *Nachtigall Projekt* group on a black textile localised the stories gathered. With the unfolding of the project, this became the place where stories and images were narrated and stitched together.

The third and final component of the project focused on the collection of nightingales' songs by citizen scientists and their visualization through a digital, interactive map on the project's website. Data collection was enabled by the popular and free-of-charge app Naturblick developed by the museum in 2015. Funded by the Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety, the app is meant to enable 'conscious access to nature in the city' for valuing and supporting this nature'. This app allows people to 'discover city-nature' by recording and identifying animals and plants through image and sound. It integrates different taxonomic keys that let users answer a series of questions about the anatomy or morphology of the yet-to-be identified species, and optical pattern recognition that pulls up similar-looking flowers or trees. Participants in the nightingale project used the app during guided tours, though any app user could feed data to the project. Aside from the collection of song data, this component included two free workshops on bioacoustics analysis, which was used to study the songs that had been recorded and collected via the app.

10.1 People assemble to look for nightingales during an early morning guided tour in the Tiergarten Berlin. Photograph by Chiara Garbellotto.

This brief outline of the project shows that it brought together realms conventionally thought of as scientific (bioacoustics, natural history) with realms conventionally thought of as social or cultural (museum pedagogy and informal learning, story-telling, cross-stich). Jenny Reardon, in her work on diversity in modern genetics, has observed how these realms 'are inextricably interconnected and come into being together' (Reardon 2001: 381). And indeed, non-scientists have been constitutive of modern natural history that began with colonial expansions in the 15th century. The history of natural history has been made by amateurs, from colonial officers collecting specimens to hobby breeders recording the character traits of peas and finches and sharing them with Charles Darwin and other scientists. In what might be the first museum-led citizen science project, Hugo Gonwetz, the director of the natural history museum in Gdansk (then a part of the German Empire), sent out questionnaires in the early 1880s to local groups asking for data on fauna and flora (Nyhart 2009: 244). In this capacity, the museum assumed an informational and, we would argue, governmental function as a producer and arbiter of bio/diversity. The museum not only enrolled lay people in its official activities, it catalogued the species diversity of a territory, a territory *contested* across the political upheavals that marked German-Polish relations from the partitions in the 18th century to the end of Second World War.

The biological perspective

In her history of the 'biological perspective' in Germany, Lynn Nyhart examines the changes in structural and social relations that occurred in Museum für Naturkunde Berlin in the late 19th century. According to Nyhart, the 'biological perspective' replaced classical natural history and its focus on system classification with a concern for evolution, ecology and biogeography. The 'biological perspective' introduced what was to become the dominant view of nature as an interdependent and functionalist entanglement of organisms and their environments. The rise of modern biology compelled the new naturalists to venture outside and study living organisms in situ (see also Kohler 2006). Moreover, it prompted museum curators and researchers to use specimens as evidence for complex ecological dynamics outside the museum. In moving from stable form to function, and from collection to environment, the value of diversity and the meaning and measurement of difference changed, too. Diversity became a scalar phenomenon that still encompassed morphological differences at the individual level but could now also include many other levels as well, from microbial communities to ecoregions. Importantly, diversity was rendered into a tractable object for experimental systems, something that could be manipulated and scientifically verified (Kohler 2002). This shift is indicative of much broader changes that looked to science for delivering and shoring-up social facts by appealing to 'nature' or innate 'biology'. Not coincidentally, the late 19th century also saw the invention of a scientific theory of society in the work of Émile Durkheim, who published sociology's foundational text, *The Rules of Sociological Method*, in 1895. Social and cultural differences could now be reified as natural differences and their function related to the progress and well-being of society as a social whole. Difference, in the form of variation, thus attained a new valence. Patterns of variation were studied and evaluated by an increasing repertoire of differentiating techniques with regard to phenotype, behaviour, ecology and genetics (Subramaniam 2014). Natural history museums were central actors in the circulation and popularisation of these ideas. According to Nyhart, the spread and institutionalisation of the 'biological perspective' manifested itself in the invention of new display formats such as dioramas and coincided with two further transformations: the 'governmentalization' (Nyhart 2009: 201) of museums and the mutation of natural history museums into 'centres for gathering, dispensing and exchanging biological information' (Nyhart 2009: 240).

Nyhart's account situates the Museum für Naturkunde Berlin at the intersection of aesthetic, scientific and political formations that have gained novel traction amid its recent digitisation and data efforts. Specifically, the enmeshments of the museum's governmental and informational functions have intensified. The unprecedented investment of €660 million as well as large-scale projects to integrate natural history collections across Germany and the European Union attest to the political valence of data held in museums. Similarly, with exhibitions such as *Artefacts* (2018–19), co-produced with the European Commission's Joint Research Centre, the Berlin museum has positioned itself as a distributor of independent, fact-based evidence, seeking to produce well-informed subjects. The political valence is thus manifold, covering both economic rationales (bioeconomy, automation, application labs) and the production of data-literate and scientifically informed citizens. For instance, the nightingale project combines 'scien-

tific' and 'socio-cultural' dimensions that connect the birds across complex interactions of people, stories, histories, technologies and infrastructures. But just as the textual practices and infrastructures of natural history introduced with Linnaean nomenclature and taxonomy contributed to the making of particular versions of diversity, so do the practices of data-based biodiversity. Indeed, we argue that the production and process of digital data construct a different notion diversity. Like Margareta von Oswald (see her essay in this volume), we contend that (digital) data practices and their infrastructures perform specific versions of bio/diversity.

Nightingales in data

On a summer's night in 2018, a group of 17 people gathered at the southern entrance of Berlin's Volkspark Friedrichshain, one of Berlin's largest parks, to participate in a 'nightingale tour'. The tour leader, a biologist and staff member at the Natural History Museum, welcomed the group and introduced the project. Participants learned that Berlin is the nightingale capital of the world, and that it offers an ideal environment to carry out research into the birds' different songs, or 'dialects'. She instructed them on how to use the Naturblick app to record bird songs before providing additional information about *Luscinia megarhynchos* on laminated A4 sheets. The sheets bore images of a small, brown bird, chicks in a nest, diagrams, numbers and maps. People clustered around the tour leader as she started moving through the park. The group congealed into a sensing organism, mobile phones ready, and headed into the silent dark.

> A nightingale starts to sing. The tour guide smiles and points her finger up in the air. The nightingale's vocalizations travel through the leaves and the chill night air down to our ears; the group momentarily becomes silent. The father and his child move closer to the tree from which the song just reached us. Torchlights are dancing here and there trying to spot the bird among the green leaves. Screens are touched to activate the digital recording.
> (Tour guide—Fieldnotes)

The bird songs were picked up by the mobile phone's microphone and converted into a stream of discrete numbers by an analogue-to-digital converter on the phone's logic board before being saved as an mp4 file. The Naturblick app then generated an image of audio signal frequencies and amplitude change over time, which served as the species' representative signature. The file was also sent directly to the server of the Naturblick team at the Museum für Naturkunde Berlin, where members of the nightingale project could access them.

Aside from recording bird songs, the participants were invited to join two (open) workshops at the museum to learn about bioacoustics analysis. They were shown how to use software (Avisoft-SASLab Pro) for analysing the recorded bird songs. Participants learned how scientists visualise, describe and use the nightingale songs.[3] Avisoft runs on Windows, its graphical user interface is sparse and technical: grey background, a mixture of familiar icons and obscure buttons for generating different statistical reports and manipulating variables (e.g. threshold, group time, hysteresis).

As Chiara Garbellotto sat in one of the newly built and fairly nondescript education rooms at the museum, the standard screens and grey and white office-style tables provided a stark contrast to the park. A sense of place vanished as nightingale songs turned into spectrograms (also called sonograms). The spectrograms in Avisoft are set within a diagrammatic space, showing black audio waves on a white background running along an x-axis (time) and a y-axis (frequency). Using spectrogram samples, a team member explained the categories for quantitative analysis: size of repertoire, category of song, song rate, duration, stereotypy (repeated utterance), frequency, number of elements. These variables are used to compare the song's elements and correlate them with additional data, such as the location and date of the recording. One team member, a PhD student, then compiled a list of all the different song elements identified in the analysed recordings into a catalogue. The instructor urged participants to select 'good data' for the group exercise, which in this case meant choosing spectrograms that were distinct and easy to label, such as trills, which were recognisable by bold, black, sharp traces on the spectrograms.

Participants were informed that Avisoft was used to 'clean' the recordings prior to analysis, that is, it removed any 'ambient' sound in order to isolate the single nightingale's song. While seemingly sensible, this type of 'data cleaning' reveals the tension that inheres in classification systems. On the one hand, the Avisoft interface assumes the abstract ideal of stable and unique classificatory principles and mutually exclusive categories (Bowker and Star 2000: 11). On the other, as with any classification system, objects have to be standardised in order to be processed across media and languages. And there are always moments when birds and their songs don't quite fit.

Recording and analysing digital bird songs showed the constructedness of our notions of data and species. The data were both abstract—the bits and bytes travelling from mobile phone microphones to servers, softwares and databases—and material (phones, laptops, the print-outs of sonograms). It took a good deal of work to create a functionable data set—standardisation, cleaning, quantification, visualisation—work that occurred beyond the sight of the participants. And it required a sequence of decisions, big and small, setting the type of protocols, formats and research questions. The same was true for the nightingales as they moved from living beings to a set of diagrammatic representations on a computer screen, classified by Linnaean nomenclature, data points on the map of the project homepage. According to Staffan Müller-Wille, Linnaean taxonomy, introduced in the late 18th century, 'enhanced flows of data' and established 'a system of relations and equivalence, rather than difference' (2017: 126). Taxonomy and nomenclature turn nightingale into *Liscinia megarhyncho* and into a member of the Old World flycatcher family. For Müller-Wille, name and classification function as 'containers' that can be added to or divided and that have no value in and of themselves aside from facilitating human communication, exchange and ordering. Rising concurrently with the establishment of natural history collections—the Berlin museum's collection dates back to 1805—the Linnaean system turned animals and plants turned into units that could be managed, reorganised and exchanged within and between institutions (Müller-Wille 2017). In this sense, the Linnaean taxonomy had long ago made animals into data creatures.

10.2 A nightingale sings on a tree branch while the camera lens focuses to compose the portrait. Photograph by Daniela Friebel.

Modern digital tools preserved the Linnaean system —the Naturblick app provides the Latin binomial for each taxa—but they also brought novel paradigms. One such novel paradigm is 'interoperability', a property allowing data to flow across different systems, products and infrastructures. The paradigm is crucial because communication and exchange occur between humans, machines and computers. It requires shared terminologies, institutional and disciplinary domains and semantic metadata. This entails a host of new data, standards, infrastructures, etc. Interoperability, next to time and space, is a vector for ordering. As such, it is associated with another paradigmatic change, which Geoffrey Bowker (2000) has identified in biodiversity sciences: the transformation of data and databases into ends in themselves. Vincent Devictor and Bernadette Bensaude-Vincent (2016) argue that 'global biodiversity' is an artefact of data infrastructures and that the concept 'is, above all, a policy-making platform' (2016: 13). The collection of occurrence data together with the massive digitisation of natural history collections—a potential area for citizen science—works to legitimise a particular research style (Devictor and Bensaude-Vincent 2016: 13). Nyhart's 'biological perspective' has become datafied with biodiversity science and ecology, and rendered into technoscience.

As a data creature, the nightingale is an aggregate of different data points that circulate from park to mobile phone to museum to global database to policy reports and biodiversity assessments. According to Müller-Wille, those data points can be expanded indefinitely. There will always be another occurrence, another song to record. But what his diagnosis of natural history misses is that the form of data compels some logics and not others. The discovery and management of biodiversity require standardisation to produce generalizable knowledge. 'No science', Lorraine Daston and Peter Galison write,

'can do without such standardised working objects, for unrefined natural objects are too quirkily particular to cooperate in generalizations and comparisons' (1992: 85). In other words, the nightingale as data creature is an object of science composed of standard protocols and international classification systems. Through their ability to standardise, nomenclature and taxonomy carry prescriptive power. But who decides what counts as a standard and whom or what does this decision serve?

Citizens in data

The 'citizen' in citizen science is difficult to pin down. In the considerable literature on citizen science, the citizen is imagined as a non-descript person positioned outside of science proper who is distinguished, primarily, by his or her activity. Citizens are described as 'people who actively observe, report, collect, analyse and disseminate information via text, audio, or video messages' (Sheth 2009). In a similar vein, the EU report 'Using New Data Sources for Policymaking' (Schade et al. 2017) states that citizen scientists engage with policy, interact with scientists and politicians and contribute to scientific information and knowledge. They are variously presented as distinct from 'stakeholders' and as representing 'one particular stakeholder group'. Once enrolled in citizen science projects, they become 'participants'. On the nightingale project website, citizen science is described as scientific projects undertaken by 'interested lay people and amateurs'. So while the 'citizen' appears prominently in project descriptions and citizen science discourse, the term remains a 'floating signifier' (Lévi-Strauss 1987), a signifier with a vague or unspecifiable signified. These diverse designations all assume that citizens—whatever they are—come into being performatively through practices of interacting or engaging with science. Rather than focusing on what citizenship is, a semiotic understanding of diversity turns our attention to how 'citizenship' is done.

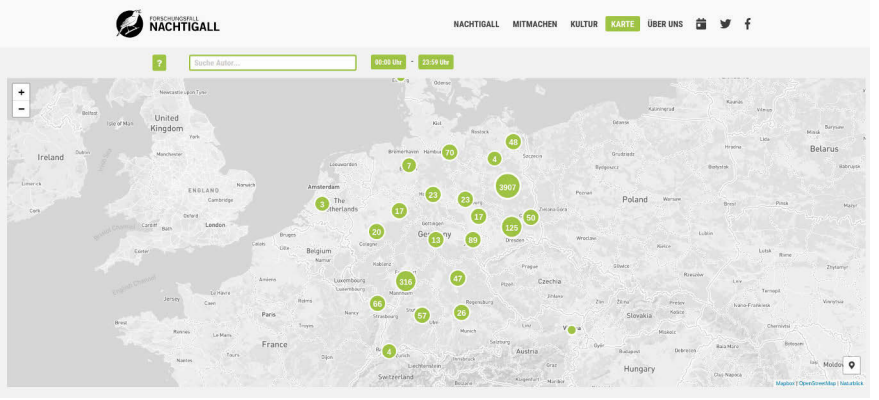
A central concern of the nightingale project, and a key practice through which citizenship is enacted, is 'language'. The study of bird songs and learning behaviour gained popularity in the 1950s with the spread of sonogram technology. This turned the practices of amateur naturalists in the birding community into a scientific method (Marler and Slabbekoorn 2004: 11). The study of single-species song repertoires navigates a complex terrain between variability (due to 'dialects' and improvisation) and universalism at the species level (Devictor and Bensaude-Vincent 2016: 13). Moreover, it brings together concepts of phonotypical and genetic evolution, sexual selection and social learning (Devictor and Bensaude-Vincent 2016: 25) to address intraspecies behaviours within the (acoustic) environment individual birds occupy. One of the main applications of research on vocal learning is determining the biological substrate of human language evolution. Language is not only a 'mystery' for evolutionary theory, it is a key biological and sociological unit for establishing bio/diversity. Reardon notes that the lead scientist at the Human Genome Diversity Project, a controversial effort begun in the 1990s at Stanford University to gather genetic material from interesting 'populations', proposed collecting samples 'from 50 individuals from 200 isolated "aboriginal tribes" defined according to language' (2001: 363). Here, language is identified as a stable classifier by which to assemble a particular social group—'aboriginal tribes'—and distinguish them from other groups. At the

same time, the choice of the term 'aboriginal tribe' makes evident the social and political values that persist in the doing of biological sciences and that stand in stark contrast to the relational and process-based approach to culture in the social and cultural sciences. A third conception of language appears through the data-based infrastructures that subtend the project in the form of 'semantic metadata', i.e. the 'metadata that describes the "meaning" of data' (Melton and Buxton 2006: 75). Given the 'fundamental relationship between the speaking subject and the citizen subject' (Isin and Ruppert 2015: 52), the data-based engagements prompted by the app suggest two important problematisations of this relationship. First, the introduction of apps to facilitate or, at times, enable civic duties add a novel range of utterances to the speech of the citizen subject, from clicking consent boxes to allowing the extraction of data streams from devices. Second, this type of speech is no longer entirely readable by humans and as such is incongruent with conventional practices of accountability.

Biological sciences have long incorporated sociological categories and, as the passage from Reardon shows, have used them to explicate orders in the natural world. Here, human language, a complex and continuously changing socio-cultural phenomenon, becomes representative of and for a particular genetic make-up. Much of feminist science studies is dedicated to unpacking the problems that arise when contested terminology is elided in favour of a narrowly defined scientificity. Messy and complex histories inhere in the term 'citizen'. There is of course no requirement to prove national citizenship status in order to participate in citizen science, although the nightingale's project's second component did specifically enrol 'new citizens'. For the third component, the collection of bird song data, participants needed to use the Naturblick app, which is available on iOS and Android accounts. The smart phone, its operating system and the app thus become material expressions or correlates of 'citizenship'. For some users, then, citizenship depends on being in possession of a Google account or being registered at the Google App Store. Given the well-documented malpractices of Google and its parent company Alphabet, such prerequisites are not trivial.[4] The app download and sign-in procedure show that the project not only collects data on nightingales but, through its infrastructures, also collects data on its participants. The invisible data traffic suggests that even while users might be able to read some of the data-based utterances, they might not have the literacy required for understanding their consequences. To put it bluntly, while many app-based citizen science projects intend to create scientifically literate citizens, they might also entrench illiteracy of data privacy issues.

Here emerges another prominent figuration of the citizen, the 'user'. Upon opening the Naturblick app on a smartphone, the catchy logo of a stylised eye with a quill for a pupil (a visual nod to the app's name, which translates as 'nature's gaze') hovers above an extravagantly coloured bird wing. Throughout the app, eye-catching buttons bearing simple logos or photographs of animals let users easily navigate and interact with the app. When first opening the app, a pop-up window informs the user about the new data protection guidelines. It spells out the terms regarding the authorship of image, sound recordings and their automatically issued geo-coordinates. Users have to agree that all recordings will fall under the CC BY SA 4.0 licence. The agreement form notes that where names are made known, they will be identified as authors. If no name is provided, the default author will be the museum's Naturblick app. It also informs users that they can share

their recordings with project partners such as the nightingale project and other publicly funded citizen science projects. The section on the collection of usage data states that all data will be collected for purely scientific purposes. Should they no longer be needed, they will be deleted. The app collects device ID and 'user interaction data'. The audio files recorded are saved on the Museum's Naturblick server and available on the user's device in a folder called Naturblick, which the app automatically creates. From there, any user is able to access or reproduce them on other devices. They are also available online as part of the map on the nightingale project's website, from which they are downloadable.



10.3 An online map that locates the verified recordings shared by the App users at the end of the second season of the nightingale project. Reproduced courtesy of the Forschungsfall Nachtigall team at the Museum für Naturkunde Berlin.

The scientific institutions of Western modernity have also been dedicated to registering humans in addition to non-human animals. Such registering and counting always require classifications and categorisations: who to count is a central question for state-sponsored efforts targeting the arithmetrics of population. Writing about the technology of census-taking, Theodore Porter argues that 'any actual count depends on the specification and political acceptance of a whole array of conventions' (1994: 391). The first hurdles are 'definitional problems', that is, the question of who to count and the constitution of the 'who'. In other words, who to count is tied to the question of who counts prior to the study design. In critical citizenship studies, attention has focused on how conceptions of citizenship, which traditionally focus on rights, obligations, expectations and nation-state membership, have changed with the rise of supranational (and subnational) institutions such as the European Union, introducing a notion of citizenship that is 'performed in relation to sociotechnical arrangements' (Isin and Ruppert 2015: 4) and based on specific norms, values and technologies. Critical data studies expand on these sociotechnical arrangements by examining the ways in which governments and state agencies define citizenship in contradistinction to 'foreignness' through the logic of data-based surveillance regimes (e.g. Browne 2015).

On the nightingale project's website, participants are addressed simultaneously as 'laymen and amateurs' and invited to become 'citizen researchers'. Participation is described as a 'dialogue between science and society', with the museum as mediator. This enrolment of the project's participants in the museum's mission returns us to Nyhart's argument about the governmental function of museums. A crucial component of this function is the creation of narratives of participation to legitimise the institution. The coexistence of 'citizen science' in the museum with other professional terminology and related practices such as 'knowledge transfer', 'participation' and 'open science' is part of a professionalization and standardization of participatory technologies (Chilvers and Kearnes 2016: 44). These take the form of data-gathering efforts at various scales, and are part of a data-driven 'biological perspective' that relies on the binary opposition of science/society. One technology that appears to have been particularly generative in the case of the nightingale project was the creation of imagined participant categories for funding schemes and accounting procedures. The need to identify target publics for both the project's initial proposal and the final evaluation required defining a list of subjects—'citizens', 'clubs', 'new citizens', 'school and kindergarten students', 'scientific team', and 'cooperation partners, collaborators, input providers'. Moreover, under the umbrella of 'citizen science', the explicit separation of the scientific activities from educational and cultural activities re-produce divisions at the epistemological and ontological levels. Citizens emerge as data-literate 'researchers' who record nightingales and produce data; students, who lack conservationist knowledge, learn how to identify nightingales; new Berliners as representatives of their country of origin provide memories of their 'previous life' where nightingales figure as cultural symbols. The clear delineation of these identities is a response to the formal requirements of projects within European and state funding schemes. At the same time, such categorisations might render invisible the partial or undisciplined connections between participants and between research objects and subjects (see also the contributions to this volume by Gerbich, Puzon and Tinius).

In sociology, anthropology and political theory, 'citizen' and 'citizenship' are highly contested concepts in both theory and practice. Nowadays, citizenship depends on being informed. We see this in test requirements for citizenship candidates as well as in the enormous amounts of information we must process if we are to make use of our rights as citizens, consumers, students, employees etc. In the nightingale project, data and participants are co-produced. As participants gather and collect data via their smartphones, they generate actionable data for science, the museum, the German government, European funders and a host of other entities not readily identifiable, or, not yet assembled given the rapid developments in the current data economy. At the same time, data production turns citizens into participants and scientists. But the exact remit of these designations, the political consequences and efficacies and historical meanings of these subject positions remain obscure. The history of statistics has demonstrated how populations and individuals have been rendered visible and governable through data. A future history of museums might demonstrate how data have been rendered visible and valent through participation.

Conclusion

> CS [citizen science] is definitively needed, but also not all the scientific projects can be turned into CS, though some can. For biodiversity monitoring, CS is the future. It helps raise awareness of conservation issues: if people know how to look, they can also care. Species conservation is an issue.
> (Tour guide, private conversation)

In this chapter we focused on the ambiguous nature of bio/diversity on the basis of a citizen science project conceived and run by the Naturkunde Museum für Berlin and dedicated to recording and mapping nightingale songs. In particular, we attended to the data practices of the project—from recording and collecting to analysing—in order to capture how different versions of citizens and species are made in and with data. On the basis of this, we put forward the notion of the 'data creature' to understand the ways in which nightingales and 'citizens' are brought into being.

Our key argument was that doing diversity moves between biological and social registers and that data are novel mediators for terms and practices to criss-cross these domains. Data practices have also extended the 'biological perspective' that is associated with the move from a taxonomy-oriented natural history to an ecology-oriented biology. Natural history museums have been instrumental in construing and disseminating this perspective and, by doing so, have assumed both informational and political functions. These functions have only intensified as data-based efforts to capture, collect and communicate biodiversity have increased. And we can see a reintroduction of taxonomy-oriented logics and formalisms in the ways in which standards and protocols are structuring the capture of biodiversity. One of the problems this poses is the intransparency of data and their digital infrastructures. If the aim of citizen science is the production of scientifically literate people and their participation in scientific practices, then it is crucial to help participants develop a 'critical data consciousness' (Nadim 2018) and to explore the possibility of making scientific epistemology—its historicity and its social and political constructedness—an object of collaboration among scientists and non-scientists as well. This would also help integrate citizen science and civic participation. In fact, for citizen science to work, it requires civic participation.

The form of problems posed by diversity in natural history and citizenship discourse share a similar tension between universalism and particularism. Like the citizen, the species is a 'historical and geographic figure' (Isin and Ruppert 2015: 19) whose form and substance is tied to ideas, practices and institutions such as natural history museums. Questions as to the unit of analysis (how to measure/diagnose diversity?), scale (is diversity a trait of individuals or groups?) and mode of ordering (population, culture, flock, participants, etc.) connect practices and categories across biological and socio-political discourses. Importantly, the 'liberal subject form' (Brown 2006), which shapes the ontological foundation of Western concepts of citizenship, is deeply connected to the history and scientific understanding of (human) bodies. The history of this understanding is also a history of data practices.

An important application of the research serves the monitoring of species within the urban environment, especially through the production of occurrence maps. This data is

directly relevant to the development of conservation and protection policies in connection to urban planning. Nightingales are particularly interesting in this regard because of their preferred habitat: 'bushes with branches close to the ground (as food and reproduction areas), tree/shrub over-holders (as singing and warning stations) and high perennial vegetation (as food areas)' (Witt 1996: 13). Berlin is nowadays considered 'the capital of nightingales' because of its particular urban conditions, which themselves are the result of its multiple reconstructions and divisions. For Witt, bush pruning and the 'cleaning' of public spaces were to blame for the decrease in nightingale species populations as part of clean-up works conducted in the eastern neighbourhoods of the city in the years after German reunification. Human activity shapes nature just as nature and environmental changes shape human activities. Losing social diversity—in the form of public spaces and mixed-used housing—leads to the loss of biodiversity. The data creatures produced in the project will live on, in databases, project reports, friendships, scientific papers, mobile phones and other unlikely places. They are ready to inform urban planning and make the city more liveable for humans and nightingales alike.

Acknowledgements

Chiara Garbellotto's doctoral research has been funded by the Alexander von Humboldt Foundation as part of the research award for Sharon Macdonald's Alexander von Humboldt Professorship (2017–2020). She thanks the colleagues at the Centre for Anthropological Research on Museums and Heritage (CARMAH) for the inspiring discussions and the support received, especially by her cohort companion doctoral researcher Nazlı Cabadağ. We would like to give special thanks to the *Forschungsfall Nachtigall* team at the Museum für Naturkunde and all the project's participants for the curiosity and availability. We are particularly grateful to Silke Voigt-Heucke and Sarah Darwin for their comments on an earlier draft of the paper.

Notes

1 On the role of lists in the production of natural orders, see Müller-Wille and Charmantier (2012) and Pugliano (2012); on note-taking and double-entry bookkeeping, see te Heesen (2005a; (2005b); on the role of description as a type of technology in early natural history, see Ogilvie (2008).

2 With reference to the performativity of technologies and language, we suggest that how we account for biodiversity also shapes how we relate to it. See, for example Asdal (2008) and Höhler and Ziegler (2010). This accounting entails a range of objects and narratives through which diversity becomes enacted differently. By way of, say, remote-sensing data and global environmental monitoring networks, biodiversity is turned into a set of variables that allow the calculation of global biodiversity trends. These are subsequently ameniable to and necessitate specific neoliberal forms of governance, what Turnhout et al. (2014) have termed 'measurementality', that reify diversity as a marketable value for cost-benefit analyses.

3 Avisoft is made by a German software developer and since its release in the mid-1990s has become a central tool for 'investigating acoustic communication in various animal species including birds, mammals, rodents, frogs, fish and insects' (Avisoft website). It is used by researchers worldwide and currently costs 2,400 euros (1,800 euros for additional licences).

4 Most recent accusations focus on racism (Barmann 2022), sexism (Glaser and Adams 2021), labour relations (Robitzski 2021), algorithmic bias (Wong 2020), censorship (Roth 2021), data privacy and the facilitation of misinformation and 'fake news' (Ohlheiser 2017).

References

Asdal, K. 2008. 'Enacting things through numbers: Taking nature into account/ing', *Geoforum*, 39(1): 123–132.

Barmann, J. 2022. 'Alphabet hit with racial discrimination. Lawsuit by former diversity recruiter for Google', 22 March 2022, *SFist* https://sfist.com/2022/03/22/alphabet-hit-with-racial-discrimination-lawsuit-by-former-diversity-recruiter-for-google/ (accessed 4 April 2022).

Bowker, G. C. 2000. 'Biodiversity Datadiversity', *Social Studies of Science*, 30(5): 643–83.

Bowker, G. C., and S. L. Star. 2000. *Sorting Things Out*. Cambridge, MA: MIT Press.

Brown, W. 2006. 'American nightmare: Neoliberalism, neoconservatism, and de-democratization', *Political Theory*, 34(6): 690–714.

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

Chilvers, J., and M. Kearnes (eds). 2016. *Remaking Participation: Science, Environment and Emergent Publics*. London: Routledge.

Daston, L., and P. Galison. 1992 'The image of objectivity', *Representations*, (40): 81–128.

Devictor, V., and B. Bensaude-Vincent. 2016. 'From ecological records to Big Data: The invention of global biodiversity', *History and Philosophy of the Life Sciences*, 38(4).

Federal Government Position Paper. 2017. *Guidelines for the New EU Framework Programme for Research and Innovation*. https://www.bmbf.de/files/Federal_government_FP9_guidelines_September_2017.pdf (accessed 4 April 2022).

Franklin, S., C. Lury, and J. Stacey. 2000. *Global Nature, Global Culture*. London: SAGE Publications.

Gabrys, J. 2017. 'The becoming environmental of computation from citizen sensing to planetary computerization', *Italian Journal of Science & Technology Studies*, 8(1): 5–21.

Glaser, A., and C. Adams. 2021. 'Google advised mental health care when workers complained about racism and sexism', 7 March 2021, *NBCNews*. https://www.nbcnews.com/tech/tech-news/google-advised-mental-health-care-when-workers-complained-about-racism-n1259728 (accessed 4 April 2022).

Haraway, D. 2018. *Modest_Witness@Second_Millennium. FemaleMan_Meets_OncoMouse: Feminism and Technoscience*. New York: Routledge.

Heesen, A. te. 2005a. 'The notebook: A paper technology', in B. Latour, and P. Weibel (eds), *Making Things Public. Atmospheres of Democracy*, 582–89. Cambridge, MA: MIT Press.

Heesen, A. te. 2005b. 'Accounting for the natural world: Double-entry bookkeeping in the field', in L. Schiebinger, and C. Swan (eds), *Colonial Botany: Science, Commerce, and Politics in the Early Modern World*, 237–251. Philadelphia: University Pennsylvania Press.

Höhler, S. and R. Ziegler 2010. 'Nature's accountability: Stocks and stories', Science as Culture 19(4): 417–430.

Isin, E. F., and E. Ruppert 2015. *Being Digital Citizens*. London: Rowman & Littlefield.

Kimura, A. H., and A. Kinchy. 2016. 'Citizen science: Probing the virtues and contexts of participatory research', *Engaging Science, Technology, and Society*, 2: 331–361.

Kohler, R. 2002. *Landscapes & Labscapes: Exploring the Lab-Field Border in Biology*. Chicago: University of Chicago Press.

Kohler, R. 2006. *All Creatures: Naturalists, Collectors, and Biodiversity, 1850–1950*. Princeton University Press.

Lévi-Strauss, C. 1987. *Introduction to the Work of Marcel Mauss*. London: Routledge.

Marler, P., and H. W. Slabbekoorn (eds). 2004. *Nature's Music: The Science of Birdsong*. Amsterdam: Elsevier Academic.

McGhie, H. 2019. 'Museum collections and biodiversity conservation', *Curating Tomorrow*. https://curatingtomorrow236646048.files.wordpress.com/2019/08/museum-collections-and-biodiversity-conservation-2019.pdf (accessed 4 April 2022).

Melton, J., and S. Buxton. 2006. 'Metadata—An overview', in J. Melton, and S. Buxton (eds), *Querying XML*, 67–84. Burlington: Morgan Kaufmann.

Müller-Wille, S. 2017. 'Names and numbers: "Data" in classical natural history, 1758–1859', *Osiris*, 32(1): 109–128.

Müller-Wille, S., and I. Charmantier. 2012. 'Lists as research technologies', *Isis*, 103(4): 743–752.

Nadim, T. 2018. 'c u soon humans need to sleep now so many conversations today thx', in T. Nadim, and N. Wagner (eds), *The Influencing Machine*, 72–81. Berlin: nGbK.

Nadim, T. 2022. 'All the data creatures', in E. J. Gonzalez-Polledo, and S. Posocco (eds), *Bioinformation Worlds and Futures*, 20–35. London: Routledge.

Nyhart, L. N. 2009. *Modern Nature: The Rise of the Biological Perspective in Germany*. Chicago: University of Chicago Press.

Ohlheiser, A. 2017. 'How far-right trolls named the wrong man as the Las Vegas shooter, 2 October 2017, *The Washington Post.* https://www.washingtonpost.com/news/the-intersect/wp/2017/10/02/how-far-right-trolls-named-the-wrong-man-as-the-las-vegas-shooter/ (accessed 4 April 2022).

Ogilvie, B. W. 2008. *The Science of Describing: Natural History in Renaissance Europe*. Chicago: University of Chicago Press.

Porter, T. M. 1994. 'Making things quantitative', *Science in Context*, 7(3): 389–407.

Pugliano, V. 2012. 'Specimen lists: Artisanal writing or natural historical paperwork?', *Isis*, 103(4): 716–26.

Reardon, J. 2001. 'The human genome diversity project: A case study in coproduction', *Social Studies of Science*, 31(3): 357–388.

Robitzski, D. 2021 April 27. 'Google investors condemn company for punishing whistleblowers', 27 April 2021, *Futurism*.

https://futurism.com/the-byte/google-investors-condemn-punishing-whistleblowers (accessed 4 April 2022).

Roth, A. 2021. 'This article is more than 6 months old Apple and Google accused of 'political censorship' over Alexei Navalny app', 17 September 2021, *The Guardian*. https://www.theguardian.com/world/2021/sep/17/apple-and-google-accused-of-political-censorship-over-alexei-navalny-app (accessed 4 April 2022).

Schade, S., M. Manzoni-Brusati, C. Tsinaraki, et al. 2017. *Using New Data Sources for Policymaking, JRC Technical Reports*. Luxembourg: Publications Office of the European Union.

Sheth, A. 2009. 'Citizen sensing, social signals, and enriching human experience', *IEEE Internet Computing*, 13(4): 87–92.

Strasser, B. J., L. Baudry, D. Mahr, G. Sanchez, and E. Tancoigne. 2019. '"Citizen Science"? Rethinking science and public participation', *Science & Technology Studies*: 52–76.

Subramaniam, B. 2014. *Ghost Stories for Darwin: The Science of Variation and the Politics of Diversity*. Urbana: University of Illinois.

Turnhout, E., K. Neves, and E. de Lijster 2014. '"Measurementality" in biodiversity governance: Knowledge, transparency, and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Ipbes)', *Environment and Planning A: Economy and Space*, 46(3): 581–597.

Witt, K. 1996. *Bestand der Nachtigall (Luscinia Megarhynchos) in Berlin in den Jahren 1993–95*. Berl. Ornithol. Ber. 6: 3 22.

Wong, J. C. 2020. 'More than 1,200 Google workers condemn firing of AI scientist Timnit Gebru', 4 December 2020, *The Guardian*. https://www.theguardian.com/technology/2020/dec/04/timnit-gebru-google-ai-fired-diversity-ethics (accessed 4 April 2022).