

Wolfram Horstmann*

Forschungsdaten und Bibliotheken – Gedankenexperimente zur Covidpandemie und Einschätzungen

<https://doi.org/10.1515/bfp-2021-0029>

Zusammenfassung: Was wäre wenn eine erneute Covid-pandemie im Jahre 2070 zeigen würde, dass eine schnelle Eindämmung deshalb nicht möglich ist, weil die Forschungsdaten aus den Jahren um 2020 nicht nachhaltig gespeichert wurden. Ausgehend von dieser Frage werden Einschätzungen zur Rolle von Bibliotheken für Forschungsdaten entwickelt. Grundsätzlich wird geschlossen, dass Bibliotheken eine besondere Chance haben, einen wichtigen Anteil an der Bewältigung der Herausforderungen im Kontext von Forschungsdaten zu tragen und damit eine neue Identität der Bibliothek zu prägen.

Schlüsselwörter: Forschungsdaten; Bibliothek; Covid; Infrastruktur; Wissenschaftspraxis

Research Data and Libraries-A Thought Experiment on the Covid Pandemic and Considerations

Abstract: What if a new Covid pandemic in 2070 showed that rapid containment is not possible because today's research data from the years around 2020 were not sustainably preserved. Based on this question, considerations of the role of libraries for research data will be developed. It is concluded that libraries have a chance to play an important part in overcoming the challenges in the context of research data and thus to shape a new identity for the library.

Keywords: Research data; library; Covid; infrastructure; scientific conduct

Dieser Beitrag nähert sich dem Thema Forschungsdaten in Form eines Gedankenexperimentes und eigener Einschätzungen. Er ist damit nicht als ein wissenschaftlicher Artikel mit Bezug auf die umfassende Forschungsliteratur zu dem Thema zu verstehen, sondern soll vielmehr noch einmal die grundsätzlichen Herausforderungen in möglichst einfacher Form aufzeigen.

*Kontaktperson: Prof. Dr. Wolfram Horstmann,
horstmann@sub.uni-goettingen.de

Stellen wir uns vor, dass im Jahr 2070 eine neue Covid-pandemie entsteht. Die Pandemie, die sich im Jahr 2020 zeigte, konnte in den Folgejahren vollständig bewältigt werden. Das Thema beherrschte die Wissenschaft noch einige Jahre, wurde jedoch daraufhin vollständig von der wissenschaftlichen Behandlung der Klimakatastrophe abgelöst. Die Ausbreitung einer neuen Form von Coronaviren im Jahr 2070 stellt Gesellschaft und Wissenschaft vor ähnliche Herausforderungen wie im Jahr 2020. Es stellt sich die Frage, was man aus der Vergangenheit lernen kann, um einen besseren Umgang mit der erneuten Pandemie zu ermöglichen?

Überliefert ist der Korpus der Forschungsliteratur und Medienberichterstattung. Die Klimakatastrophe hat zu großen Umwälzungen in der Informationswelt geführt. Die großen Unternehmen, die Informationen aus Wissenschaft und Medien gespeichert hatten, existieren nicht mehr. Das öffentliche System zur Informationsversorgung über Bibliotheken, Archive und andere Einrichtungen wurde zugunsten lebensrettender Maßnahmen zur Bekämpfung der Klimakatastrophe weitgehend abgewickelt. Eine Überführung der Datenbestände von Unternehmen und öffentlichen Einrichtungen fand nur in reduzierter Form statt. Eine Rekonstruktion der Ereignisse und Untersuchungen während der Covid-19-Pandemie aus dem Jahr 2020 muss also auf Basis von Texten, die überliefert worden sind, erfolgen.

Dabei sind Analyseverfahren und Simulationstechniken auf Basis künstlicher Intelligenz im Jahr 2070 so weit vorangeschritten, dass eine Ausbreitung des neuen Coronavirus durch gezielte Maßnahmen schnell einzudämmen wäre. Was fehlt sind Daten. Das Wissenschaftssystem im Jahr 2020 publizierte Forschungsergebnisse in Form von Artikeln in einer fixierten Form, der sogenannten Portable Document Format, kurz PDF. Daten waren lediglich als Grafiken oder in stark verkürzter Form enthalten. Die Daten, die den Forschungsergebnissen zugrunde lagen, wurden separat verarbeitet und gespeichert. Die Reproduzierbarkeit der Ergebnisse sollte gewährleistet sein, indem Forschende, die Ergebnisse überprüfen wollten, sich die Daten separat besorgten. Die Daten lagen verteilt in verschiedenen Umgebungen, teils bei Unternehmen, teils in

Forschungseinrichtungen, teils in Informationsinstitutionen. Standards, die eine Überführung der Daten in nachhaltige Pflege ermöglichten, waren nicht ausreichend etabliert, um eine Rettung der Daten während der Klimakatastrophe zu ermöglichen. Die Gesellschaft im Jahr 2070 muss neue Daten erheben während die Covidpandemie erneut ihren Lauf nimmt.

Eine Debatte über die Schuldfrage entbrennt. Die Politik kritisiert die Wissenschaft, in 2020 keine ausreichend gute wissenschaftliche Praxis etabliert gehabt zu haben, die es erzwungen hätte, dass Forschungsdaten immer und in jedem Fall bei Veröffentlichungen mitzuführen sind, um Forschung reproduzierbar zu machen. Die Wissenschaft kritisiert die Politik, die Datenpflege nicht auskömmlich finanziert zu haben, und dass die Förderung der Forschung nur auf den schnellen Fortschritt unter immer mehr Wettbewerb bedacht war. Die Gesellschaft insgesamt fragt sich, warum Bibliotheken, Archive und andere wissenschaftliche Infrastrukturen zwar in der Lage waren, die Texte zu retten, aber Daten nicht? Alle kritisieren, dass die Industrie jahrelang an den Daten verdient, aber keine Vorsorge für die nachhaltige Bereitstellung betrieben hat – und dass Wissenschaftler*innen die Daten über die industriellen Angebote gespeichert und verarbeitet haben.

Das Gedankenexperiment soll Fragen aufwerfen. Warum gibt es keine Verpflichtung für die Reproduzierbarkeit von Forschungsergebnissen? Warum hat das Forschungsdatensystem keinen doppelten Boden, der verhindert, dass unkalkulierbare Risiken für die Wissenschaft entstehen?

1 Daten sind nicht gleich Forschungsdaten

Dabei werden Daten doch seit Jahren als Rohstoff des 21. Jahrhunderts gesehen, wie der interessierten Öffentlichkeit beispielsweise im Artikel von WIRED¹ erklärt wurde. Daten sind nicht immer Forschungsdaten. Zu den Daten, die die Industrie benötigt, gehören zum Beispiel Daten über Menschen zu Ihrem Konsumverhalten oder ihrem Alltag. Diese und auch Daten aus sozialen Medien, z. B. Twitter, sind nicht primär für die Forschung ‚gemacht‘, werden aber in der wissenschaftlichen Forschung verwendet. Forschungsdaten im engeren Sinne sind solche, die aus der Forschung, d. h. dem wissenschaftlichen Prozess, stam-

men. Der Diskurs in der Wissenschaft über diesen Typus von Forschungsdaten geht bis ins letzte Jahrhundert zurück – spätestens auf das Human Genome Project, das in den 1990er-Jahren ein weltweites System zum Teilen von Daten entwickelte, die an vielen Stellen der Welt produziert wurden.

Auch in Bibliotheken und Rechenzentren in Deutschland wurde das Thema adressiert.² Die Bedeutung des Themas Forschungsdaten in Bibliotheken ist hoch. Open Access, das andere beherrschende Thema in Bibliotheken, wurde und wird zum Teil vom Thema Forschungsdaten überspielt. Langfristig können Forschungsdaten das zentrale Thema für die Zukunft von Bibliotheken werden.

2 Warum Forschungsdaten in Bibliotheken?

Die Gründe, warum Forschungsdaten in Bibliotheken behandelt werden sollten, sind ebenso hintergründig wie offensichtlich. Es ist nicht ungewöhnlich, dass Bibliotheken erst einmal Unverständnis entgegengebracht wird, wenn das Thema angesprochen wird: „Bibliotheken machen doch Bücher – was haben die mit Forschungsdaten zu tun? Wichtiger ist doch, dass die Daten in den Rechenzentren gespeichert werden. Die Forschung übernimmt doch das Management von Forschungsdaten selbst!“ In der Tat ist die Herausforderung des Managements von Forschungsdaten nur kooperativ zu bewältigen. Forschungsdaten im engeren Sinne werden von der Forschung produziert und nur dort liegt auch die Expertise vor, die für das aktive Forschungsdatenmanagement benötigt wird. Bibliotheken und Rechenzentren sollten zudem arbeitsteilig vorgehen, wenn es um die Speicherung und Verarbeitungskapazitäten geht – es macht wenig Sinn, dass Doppelstrukturen entstehen und beide Einrichtungstypen sollten sich auf ihre Stärken konzentrieren. Die Stärken der Bibliotheken liegen in der nachhaltigen Bereitstellung. Es gibt dabei klar abgrenzbare Aufgaben: Sicherzustellen, dass Metadaten und nachhaltige und eindeutige Identifikationsnummern vorliegen, die in internationalen Netzwerken verwendet und verarbeitet werden können, Konzepte für die langfristige Verfügbarkeit umzusetzen und Beratung zu rechtlichen und organisatorischen Aspekten anzubieten. Unterscheidungskriterien für diejenigen Forschungsdaten, die in Bibliotheken gehalten werden und solche, die in der Forschung oder in Rechenzentren ver-

¹ <https://www.wired.com/insights/2013/02/is-big-data-the-new-black-gold>.

² <https://dini.de/publikationen/dini-publikationen/>.

bleiben, sind ebenfalls klar: Solche, die eine Bedeutung für die Wissenschaftskommunikation haben, sollten im globalen Bibliothekssystem nachgewiesen sein. Das betrifft also alle Forschungsdaten, die in der Forschungsliteratur verweisen werden und zur Überprüfbarkeit der veröffentlichten Forschungsergebnisse herangezogen werden können müssen und diejenigen Forschungsdaten, die direkt veröffentlicht werden. Hier wird es offensichtlich, warum Bibliotheken eine Rolle spielen. Sie sind verantwortlich für die nachhaltige und offene Versorgung der Forschung, Lehre und Öffentlichkeit mit Literatur – und die digitalen Möglichkeiten und die veränderten Forschungspraktiken erweitern das Genre der wissenschaftlichen Literatur um die Komponente „Forschungsdaten“. Es wäre also geradezu fahrlässig, wenn Bibliotheken diese Rolle nicht wahrnehmen würden. Denn dann wird die Nachhaltigkeit der Wissenschaft eingeschränkt und es besteht keine Möglichkeit, Forschungsergebnisse nachvollziehbar in wissenschaftliches Kulturgut zu überführen.

Somit ergibt sich ein einfacher Dreiklang: die Forschung produziert und bearbeitet Forschungsdaten aktiv, die Rechenzentren stellen Speicher und Verarbeitungskapazitäten zur Verfügung und die Bibliotheken beschreiben und pflegen diejenigen Forschungsdaten, die für die Wissenschaftskommunikation wichtig sind.

3 Bibliotheks-Forschungsdaten

Es gibt darüber hinaus die Perspektive der Bibliotheks-Forschungsdaten. Bibliotheken haben eigene Forschungsdaten. Zunächst halten alle Bibliotheken Texte. Texte sind Forschungsdaten. Das ist so offensichtlich, dass es häufig übersehen wird. Liegen Texte in digitalisierter Form vor, sind sie bereits als Bild Forschungsdaten. Digitalisate von Handschriften etwa werden regelmäßig von der Wissenschaft untersucht, ob rein inhaltlich oder mit Bilderkennungsverfahren. Wenn dann auch noch der Text über automatisierte Verfahren ausgelesen wurde, kann dieser bearbeitet werden, zum Beispiel um Digitale Editionen zu erstellen. Viele solche Texte können einen Korpus erbringen, den man mit Verfahren der Computerlinguistik untersuchen kann. Zusätzlich gibt es in Bibliotheken große Bestände von lizenzierter Literatur, etwa Zeitschriftenartikel, für die sogar häufig auch Rechte für sogenanntes Text-and-Data-Mining vorliegen. Eine systematische Erschließung dieser Möglichkeiten, Texte der Bibliotheken als Forschungsdaten zu behandeln, existiert am ehesten noch in den digitalen Geisteswissenschaften für die Erstellung von Digitalen Editionen. Die Nutzung dieses Datenschatzes mit Methoden der Künstlichen Intelligenz und

des maschinellen Lernens ist jedoch eher selten zu beobachten.

Die zweite Kategorie von Bibliotheks-Forschungsdaten sind Metadaten. Früher als bibliografische Daten für Bücher als Grundlage der Bibliothekskataloge entwickelt, sind Metadaten heute allgegenwärtig, wenn auch selten sichtbar. Sie liegen nahezu jedem Suchergebnis im Internet zugrunde, speichern Zeit und Ort des Fotos auf dem Smartphone und sind eben auch für Forschungsdaten wichtig. Bibliotheken haben bereits im 20. Jahrhundert weltweit mit Metadaten für Bücher ein einmaliges öffentlich gestütztes System zum Suchen und Finden von Wissensressourcen aufgebaut, das global vernetzt ist und sowohl kontinental, national, regional und am Ort digitale und physische Ressourcen zugänglich macht. Darauf auszubauen und dies für Forschungsdaten zu entwickeln, könnte die entscheidende Herausforderung für Bibliotheken im 21. Jahrhundert sein. Es geht dabei zum einen darum sicherzustellen, dass überhaupt Metadaten vorliegen, aber auch, in welchem Format und wie sie ausgetauscht werden können. Zum anderen geht es aber auch darum, die im Hintergrund etablierten Verfahren, wie die Benutzung von kontrollierten Vokabularen, Thesauri und Ontologien zur Vernetzung, Standardisierung, Normalisierung der Metadaten weiterzuentwickeln. Denn darüber lässt sich eines der größten Probleme für das Forschungsdaten-Management angehen: die Verknüpfung unterschiedlicher fachwissenschaftlicher Domänen. Forschungsdaten, die in der Wissenschaft produziert werden, werden auch dort fachspezifisch mit Metadaten ausgezeichnet. Dadurch existieren zahllose Verfahren und, wenn überhaupt vorhanden, Spezifikationen für Metadaten. Jeden Metadatensatz intellektuell zu prüfen ist unmöglich, so dass automatisierte oder semi-automatisierte Verfahren notwendig sind. Wie bei den Texten sind hier Verfahren der Künstlichen Intelligenz und des maschinellen Lernens notwendig. Hier haben Bibliotheken bisher wenig Ansätze entwickelt, die Chance zu nutzen und damit fachübergreifende, interdisziplinäre Wissenschaft stärker zu unterstützen. Die Forschung im Bereich semantischer Netzwerke ist weit vorangeschritten, aber eine Überführung in die Praxis eines weltweiten öffentlich gestützten Systems steht aus.

4 Forschungsdateninfrastrukturen und ein Appell an die Wissenschaftspolitik

Forschungsdateninfrastrukturen sind im Prinzip nichts anderes als eine Institutionalisierung von den verschiedenen Services, die oben genannt wurden. Bibliotheken sind zwar auf dem Radar bei deren Entwicklung, aber die Rollen sind noch nicht vollständig etabliert. Bibliotheken sind in den Geisteswissenschaften stark vertreten. Es ist quasi das „Heimspiel“, weil die in Bibliotheken vorhandenen Textmaterialien dort am stärksten genutzt werden. In den Naturwissenschaften ist der Ruf der Bibliothek als primärer Partner für wissenschaftliche Information eingeschränkt – wohl auch deswegen, weil die Bedeutung wissenschaftlicher, digitaler Zeitschriftenartikel im Vordergrund steht und die Rolle der Bibliotheken bei deren Bereitstellung nicht so offensichtlich ist wie bei gedruckter Literatur.

Es ist eine kritische Kommunikationsaufgabe für Bibliotheken, ein neues Rollenverständnis von Bibliotheken als Akteur in einem global verteilten Netzwerk wissenschaftlicher digitaler Information zu erklären. Dies hat interne und externe Kommunikationsaspekte. Zu naheliegend und gewohnt ist es, Bibliotheken auf Bücher und Lernraum zu reduzieren – besonders dann, wenn eine neue Aufgabe der Bibliotheken als Eingriff in das eigene wissenschaftliche Handeln verstanden werden kann und erst recht, wenn es zusätzliche Finanzierungen erfordert, die gegebenenfalls auch im eigenen Labor landen oder gar zu eigenen Mehrkosten oder Einsparungen führen könnten. Die Forschung kann jedoch auch als inhärent verantwortungsvoll, rational oder pragmatisch gesehen werden. Und die Botschaft, dass neue Kooperationsformen zwischen Forschung und Bibliothek entwickelt werden müssen, um den Herausforderungen der Forschungsdaten begegnen zu können, kann Gehör finden, wenn sie klar ausgesprochen wird.

Auch bibliotheksintern ist wichtig zu betonen, dass die Übernahme dieser neuen Aufgabe einen massiven Strukturwandelprozess erfordert. Aufgaben des Personals verändern sich. Das bedeutet Schulung und ein langfristiges Konzept zur Veränderung des Qualifikationsprofils. Neue Kooperationsformen müssen institutionalisiert werden. Das bedeutet neues Vertragswerk und die aktive Exploration neuer rechtlicher Rahmenbedingungen. Zum Teil kann dies aus der Transformation der heutigen Bibliotheksaufgaben gestemmt werden, zum Teil sind aber auch zusätzliche Mittel erforderlich, besonders als mittelfristige Programme für Personal und digitale Systeme. Die Wis-

senschaftspolitik ist hier gefordert, intelligente Angebote zu entwickeln, die es transformationsfähigen Bibliotheken erlauben, ihr Schicksal in den nächsten ein oder zwei Jahrzehnten selbst zu gestalten und dabei erfolgreiche Kulturen der Bibliothekswelt auf ein neues Problem anzuwenden. Es handelt sich hier um nicht weniger als die Aufgabe, wissenschaftliches Kulturgut des 21. Jahrhunderts nachhaltig zu sichern und die Identität der Bibliothek als gesamtgesellschaftliche Institution weiterzuentwickeln.

5 Die „institutionelle Forschungsdatenbibliothek“ als erfolgskritische Infrastruktur der Wissenschaft

Die Aufgaben beginnen vor der eigenen Haustür. Die Diversität von Forschungsdaten hat in der Wissenschaftspraxis vor Ort zwei gegenläufige aber verbundene Effekte: Je hochspezialisierter Forschende und damit die Forschungsdaten sind, desto unwahrscheinlicher wird es, dass eine einzelne Wissenschaftlerin oder ein einzelner Wissenschaftler Ansprechpartner*innen im benachbarten Labor oder Büro findet. Es bilden sich global verteilte Netzwerke, in denen Daten, Methoden und Ergebnisse ausgetauscht werden. Auf der anderen Seite werden Gemeinsamkeiten zwischen den Forschenden vor Ort verstärkt – und zwar bei den Anforderungen an institutionelle Services. Es ist weniger das gemeinsame Labor oder der gemeinsame Korridor, der entscheidend für die Zusammenarbeit ist, sondern Fragen, wie Daten und Methoden in der Kommunikation über das Internet und in den praktischen und rechtlichen Rahmenbedingungen von virtuellen Verbünden ausgetauscht werden können. Dies erfordert an Universitäten und Forschungseinrichtungen ein Umdenken, was gute Infrastruktur für Forschung überhaupt bedeutet? Viele Services verlagern sich weg von der Universität oder Forschungseinrichtung in Richtung virtueller Services industrieller oder öffentlich gestützter Anbieter. Was aber an der Universität verbleibt, sind Fragen der wissenschaftlichen Qualität, Reputation und rechtlicher Verbindlichkeiten der Wissenschaft, die an einer Universität oder Forschungseinrichtung praktiziert wird. Denn der Name der Universität oder Forschungseinrichtung steht letztendlich bei der Veröffentlichung für immer neben dem Namen der Autorin oder des Autors.

Diese Effekte führen zusammengefasst zu einer vielleicht überraschenden Sicht auf das, was man als

Infrastruktur bezeichnen kann. Eine Infrastruktur, die Fragen der wissenschaftlichen Qualität, Reputation und rechtlicher Verbindlichkeiten der Wissenschaft adressiert, wird ein kritischer Erfolgsfaktor für die langfristige Entwicklung von Universitäten und Forschungseinrichtungen. Denn Reputation ist ein zentrales Lebenselixier der Wissenschaft. Also muss Infrastruktur an Universitäten und Forschungseinrichtungen völlig neu gedacht werden. Es geht bei der Infrastruktur nicht mehr um Gebäude, Wandfarben oder Kaffeeküchen und es geht auch nicht primär um technische Geräte und Systeme. Es geht um die Bereitstellung eines virtuellen Raumes für die Forschenden vor Ort, der ihnen alle Möglichkeiten gibt, gute wissenschaftliche Praxis auszuüben.

Bezogen auf Forschungsdaten bedeutet dies, dass Universitäten und Forschungseinrichtungen Sorge dafür tragen müssen, dass diese gut dokumentiert sowie dauerhaft und rechtssicher gespeichert und für spezifizierte Nutzungsformen abrufbar bereitgestellt werden können. All dies sind klassische Anforderungen an Bibliotheken. Man könnte diese Infrastruktur eine „institutionelle Forschungsdatenbibliothek“ nennen.

6 Vor-Ort-Aktivitäten zu Forschungsdaten in Bibliotheken

Der Begriff der „institutionellen Forschungsdatenbibliothek“ suggeriert ein monistisches, in sich geschlossenes, wenn auch zugängliches Ding, das in gewisser Art und Weise unzeitgemäß und zweckfremd wirkt. Es ist anzunehmen, dass der Begriff seitens der Forschenden und des Managements nicht selten Stirnrunzeln hervorrufen wird. Eingedenk der oben beschriebenen Anforderungen muss sich dahinter jedoch ein ausgeklügeltes virtuelles System verbergen, das Forschenden alle Freiheitsgrade gibt, sich agil in global verteilten, hochspezialisierten Netzwerken zu bewegen. Solche Infrastrukturen existieren derzeit nur in Teilen und implizit in den komplexen Netzwerken der Wissenschaft, die aus Menschen, Information und Technologie bestehen.

Welchen Begriff man auch immer wählt, den Anforderungen können sich Forschende und Bibliothek vor Ort annähern. Dazu muss eine häufig ungewohnte Kooperation von Bibliothek und Forschung erlernt werden. Es bietet sich an, konsequent Forschungsprojekte in das Zentrum der Aktivitäten zu stellen – nicht etwa gleich mit institutionalisierten Services, die auf dem Reißbrett entworfen wurden, zu beginnen. Forschungsprojekte sind ein zentraler *modus operandi* der Wissenschaft, den sich Bibliothe-

ken auch selbst zu Eigen machen müssen, wenn Infrastrukturen gemeinsam entwickelt werden.

In der Göttinger eResearch Alliance³ etwa, die gemeinsam von Bibliothek und Rechenzentrum getragen wird, dreht sich alles um Projekte. Es beginnt bei der Planung des Forschungsprojektes mit der Beratung von Forschenden in der Antragstellung bei Forschungsförderern. Die Forschungsförderer berücksichtigen mehr und mehr Forschungsdatenmanagement als wesentliches Kriterium für die erfolgreiche Antragstellung in den Begutachtungskriterien. Für die Veröffentlichungen am Ende des Forschungsprojektes wird ein Daten-Repository angeboten, in dem Metadaten und persistente Identifikationsnummern für Datensätze gespeichert werden. In einer institutionellen Literaturdatenbank können Artikel und Forschungsdaten verknüpft werden. Im Kontext der Veröffentlichungen sind es die wissenschaftlichen Journale, die mehr und mehr die Anforderung haben, dass neben dem Artikel auch die Forschungsdaten veröffentlicht sein müssen.

Auch bei Durchführung des eigentlichen Forschungsprojektes kann die eResearch Alliance beteiligt werden. Sie entwickelt Software, stellt Systeme bereit oder kuratiert Daten. Darüber hinaus werden Beratungsangebote zu rechtlichen Fragen und Schulungen zu guter Forschungsdatenpraxis angeboten. Auch werden Angebote für Forschungsdaten, etwa Datenbanken, die rein aus der Forschung entstanden sind, in nachhaltige institutionelle Angebote überführt.

Die gemeinsamen Projekte und die Beteiligung an Forschungsprojekten sind wesentlich, um ein gemeinsames Verständnis zu entwickeln, wie Infrastruktur in Zukunft gestaltet werden muss. Es geht letztendlich um Menschen, die vor Ort zusammenarbeiten um vor Ort optimale Bedingungen für gute wissenschaftliche Praxis im virtuellen Raum zu schaffen.

7 Vernetzung von Forschungsdateninfrastruktur

Lokale Angebote der Forschungsdateninfrastruktur müssen vernetzt sein. Regional beginnt es gegebenenfalls gleich vor Ort, und zwar dann, wenn institutionelle Grenzen überschritten werden müssen. Das ist zum Beispiel der Fall, wenn verschiedene Forschungseinrichtungen an einem Ort ansässig sind, denn Forschende, die unter-

³ <https://www.eresearch.uni-goettingen.de/>.

schiedlichen Forschungseinrichtungen angehören, wenden nicht immer dieselben Praktiken an, haben unterschiedliche Standards und unterliegen unterschiedlichen rechtlichen Rahmenbedingungen. In Göttingen etwa arbeitet die eResearch Alliance daher im Rahmen des „Göttingen Campus“,⁴ der neben der Universität die Max-Planck-Gesellschaft, die Leibniz-Gemeinschaft, die Helmholtz Gesellschaft sowie die Akademie der Wissenschaften zu Göttingen betrifft. Derselbe Ansatz kann in geeigneter Form in regionalen Netzwerken fortgesetzt werden, indem bilaterale Kooperationen mit Forschungseinrichtungen gebildet werden – oder eben auf Ebene des Bundeslandes, das in Deutschland auch die politische Zuständigkeit für Bildung und Forschung ausübt. In der Tat entwickeln mehrere Bundesländer Forschungsdaten-Initiativen.⁵

Auf nationaler Ebene existiert in Deutschland eine Bund-Länder-Vereinbarung für eine Nationale Forschungsdateninfrastruktur,⁶ die im Wesentlichen den Empfehlungen⁷ des Rates für Informationsinfrastrukturen folgt und deren Förderung von der Deutschen Forschungsgemeinschaft betreut wird.⁸ Ein Verein auf nationaler Ebene koordiniert allgemeine, fachübergreifende Entwicklungen und stellt Anschlüsse an das Europäische System her.⁹ In Europa ist vor allem die European Open Science Cloud¹⁰ zu nennen, die allgemeine, fachübergreifende Entwicklungen koordiniert und ebenfalls als Rechtsform etabliert wurde. Wie in Deutschland mit der Bund-Länder-Vereinbarung soll auch in Europa politisch verbindlich zwischen europäischen Staaten im Rahmen einer sogenannten ‚Partnership‘ agiert werden.

Auf nationaler und internationaler Ebene erfolgt der Aufbau der Infrastrukturen fachbezogen. In der Nationalen Forschungsdateninfrastruktur werden zurzeit ausschließlich fachbezogene Konsortien gefördert und in Europa existiert ein fachlich organisierter Prozess, der sich an der „European Strategy Roadmap for Research Infrastructures“¹¹ orientiert.

Wissenschaft richtet sich letztendlich nur zum Teil an politischen Systemen aus. Wissenschaft muss Ländergrenzen und geopolitische Rahmenbedingungen transzendieren können. Global ist es vor allem die Research Data Alliance,¹² die die soziale Plattform für den Diskurs zu Forschungsdaten zwischen den komplexen wissenschaftlichen Netzwerken stützt und so Standardisierung antreibt.

8 Am Ende bleibt die letzte Meile

Angenommen, dass alle internationalen, kontinentalen, nationalen und regionalen Netzwerke funktionieren, bleibt immer noch die Aufgabe, diese Angebote an die einzelne Wissenschaftlerin oder den einzelnen Wissenschaftler vor Ort zu bringen. Die Wissenschaft tendiert dazu, sich ihre eigenen pragmatischen Lösungen zu suchen. Was aber bleibt, ist der Mensch vor Ort und seine Angehörigkeit zur Institution. Sollte diese Angehörigkeit auch noch so kurz sein, so ist sie doch eine nicht zu unterschätzende Konstante, weil es dabei immer um ein konkretes Einzelschicksal geht, das bei wissenschaftlichen Veröffentlichungen lebenslang und darüber hinaus mit dem Namen der Institution verbunden bleiben wird.

Es ist eben dieser Hintergrund, warum die lokale Vernetzung hier ausführlicher behandelt wird als die weiteren Vernetzungen. Der Anschluss des hinter einer Forschungsleistung stehenden Menschen an Forschungsdateninfrastruktur geht notwendigerweise immer von einem Ort aus. Auch wenn Forschende mobil sind und nicht zwingend lokale Infrastruktur verwenden müssen, verbleibt eine Angehörigkeit, die sich in einer alltäglichen Lebensumwelt samt ihrer digitalen Endgeräte niederschlägt. Bibliotheken befinden sich aufgrund ihrer etablierten Systeme, ihrer Kultur und ihrer Erfahrung in einer besonders geeigneten Position, diese Angehörigkeit eines Forschenden zur Institution mit Leben zu füllen und lokale Dateninfrastruktur gemeinsam mit den Forschenden zu modellieren.

In der Telekommunikation ist der Begriff der „letzten Meile“ geprägt worden. Einfach ausgedrückt ist die Überbrückung der letzten Meile das Problem, dass es einfacher ist, dicke Kabel für viele Anschlüsse zu verlegen als dünne Kabel für einzelne Anschlüsse. Bibliotheken könnten in der Forschungsdateninfrastruktur diese schwierige Herausforderung bewältigen und dabei eine neue Identität der Bibliothek als gesamtgesellschaftliche Institution er-

⁴ <https://goettingen-campus.de/de/>.

⁵ <https://www.forschungsdaten.info/fdm-im-deutschsprachigen-raum/>.

⁶ <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.

⁷ <https://www.rfii.de/download/rfii-empfehlungen-2016/>.

⁸ <https://www.dfg.de/foerderung/programme/nfdi/>.

⁹ <https://www.nfdi.de/verein>.

¹⁰ https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en.

¹¹ https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/esfri_en#:~:text=The%20European%20Strategy%20Forum%20on,also%20includes%20a%20Commission%20representative.

¹² <https://rd-alliance.org/>.

schließen. Es kann derzeit wohl nicht abschließend vorhergesagt werden, ob ein dystopisches oder eine utopisches Szenario wahrscheinlicher ist. Das dystopische Szenario ist im Rahmen des Gedankenexperiments eingangs ausführlich geschildert worden. Das utopische Szenario ist so kurz, dass es trivial erscheint: Stellen wir uns vor, dass im Jahr 2070 eine neue Covidpandemie entsteht. Die Pandemie, die sich im Jahr 2020 zeigte, konnte in den Folgejahren vollständig bewältigt werden. Das Thema beherrschte die Wissenschaft noch einige Jahre und wurde sowohl in der Forschungsliteratur als auch in den Forschungsdatenbibliotheken gut dokumentiert. Durch Simulationen der Virus-Ausbreitung und rasche Impfstoff-Entwicklung hatte die Politik im Jahr 2070 leichtes Spiel, Maßnahmen zu ergreifen, um die Pandemie binnen Wochen einzudämmen und zu beenden.



Prof. Dr. Wolfram Horstmann

Georg-August-Universität Göttingen

SUB Göttingen

Platz der Göttinger Sieben 1

D-37073 Göttingen

horstmann@sub.uni-goettingen.de