

# Viele Fronten

## Forschungsdatenmanagement als Ermöglichungspolitik

| PETRA GEHRING | **Der digitale Wandel verändert den Forschungsalltag und die methodischen Ansätze – und damit auch die Forschung selbst. Im Umgang mit Forschungsdaten geht es um die Frage, wie sich maximale Qualität von Daten, breite Nachnutzbarkeit und Datensouveränität sicherstellen lassen.**

**F**orschungsdatenmanagement? Das klingt nach trockenem Brot. Ebenso „digitale Infrastrukturen“ oder „Informationsinfrastrukturen“. Sollten das nicht Basisangelegenheiten sein, um die sich Fachleute im Hintergrund sorgen? Dienstleistungen, die (geräuschlos) funktionieren?

Man mag das wünschen. Der digitale Wandel jedoch reißt die Grenze zwischen Forscherhandeln und Infrastruk-

»Wer mit digitalen Forschungsdaten umgeht, kämpft um die Stabilisierung guter Routinen.«

turen, zwischen Forschungsinhalten und einem bloßen Verarbeiten oder Aufbewahren von Daten ein. Unter der Bedingung von Digitalität fordern alle Prozesse IT-Entscheidungen von methodischem Gewicht, diese geben auch dem Forschungsergebnis seine Gestalt und seine Materialität. Denn was sind digitale Daten? Man kann sie von Programmen, die sie prozessieren, nicht trennen. Man kann sie aufbewahren

nur um den Preis permanenter Übersetzung (von Information und Programm) in sich verändernde Maschinensprachen hinein. Und schon die bloße Sicherstellung der Wiederholbarkeit eines Forschungsprozesses – also Wissenschaftlichkeit überhaupt: die Möglichkeit des Prüfens von Ergebnissen – wird in der digitalen Forschung auf neue Weise zum Problem. Denn Datenbestände, Programme, Rechner und sogar Datenträger sind niemals ganz gleich. Von komplexen Datenprozessierungsketten (sogenannten „Pipelines“), die in der Forschung heute die Regel sind, ganz zu

schweigen.

Herausfordernd für die Wissenschaft ist nicht nur „Big Data“ (also schiere Menge), sondern es geht um Nachhaltigkeit und Methodenstrenge unter der Bedingung sich steigernder Heterogenität: um verlustreiche Übersetzungsprozesse, um ein Wimmelbild der Modelle und Formate, um Skalierungsprobleme, intransparente Software, Fragilität von Hardware, Schutz vor Angriffen und Dauersorgen hinsichtlich der Verlässlichkeit und Qualität von Datendiensten. Neue Arbeitsteilungen, eine zunehmende Abhängigkeit der Wissenschaft von Firmen bei unklarer Rechte- und Haftungslage sowie veränderte Kommunikationswege kommen hinzu. Wer mit digitalen Forschungsdaten umgeht, kämpft um die Stabilisierung guter Routinen. Digitales Arbeiten ist somit vielversprechend, aber tückisch. Und



Foto: mauritius-images

man kann nicht mal eben hineinschauen. Alles steckt im Prozess.

### Daten digital: Notwendige Weichenstellungen?

Der digitale Wandel wird immer noch unterschätzt. Vielen der neuen Möglichkeiten (mehr Daten, dramatisch beschleunigte Datenprozessierung, innovative Analyseverfahren, „Data at your fingertips“) stehen tiefgreifende Veränderungen des Forschungsalltags und der Methodiken gegenüber – und zwar in allen Fachkulturen. Mit der Qualität

### AUTORIN



Foto: Kathrin Böttner

**Petra Gehring** ist Professorin für Philosophie an der TU Darmstadt und derzeit Vorsitzende des Rats für Informationsinfrastrukturen.

der Methoden geht es aber um die Forschung selbst. So wird Forschungsdatenmanagement zur entscheidenden Ermöglichungsbedingung für Forschung überhaupt. Anders gesagt: Management – das ist der Wandel! Oder noch einmal anders gesagt: Wenn Daten das „neue Öl“ sind, dann kann es nicht sein, dass in der Wissenschaft lediglich mit Rohöl gearbeitet wird (Stichwort „Datengewinnung“), dann bedarf es neuer Arten von Raffinerien und Versorgungsnetzen (Stichworte „Datenintegration“, „Interoperabilität“), und es wird auch niemand Handelskar-

liefernden Dienste und infrastruktureller Leistungen wie Datenpflege und -verknüpfung ab. Stimmt die Datengrundlage nicht, liefern auch schlaue Pro-

»Neben neuen IT-Professuren wird ein neuer Gestaltungswille in Sachen Verknüpfbarkeit von Forschungsdaten gebraucht.«

gramme nur Wirrwarr. Und schon ist da wieder das Wort: Forschungsdatenmanagement. Ohne die Voraussetzun-

gleich auch mutiges, weil bottom up angelegtes Projekt. Gerade das föderal und selbstverantwortlich organisierte deutsche Wissenschaftssystem muss jedoch eigene Wege und Handlungsformen finden. Es ist zwar leistungsfähig und qualitätsbewusst, aber auch verzweigt, verinselt, projektabhängig und als Datenlandschaft betrachtet extrem heterogen. In vielem ist Vielfalt Stärke. Der

Umgang mit digitalen Forschungsdaten muss aber dringend durchlässiger und nachhaltiger ausgestaltet werden. Gebraucht werden volle Aufmerksamkeit für das Thema, dazu Geld, Know how, Rechte- und Prozessklärung. Vor allem aber – denn statt um Infrastruktur im herkömmlichen Wortsinn geht es um die Forschungsprozesse als solche – ist es die aktive Mitwirkung der Forschenden selbst, die benötigt wird.

#### **Konkreter: Herausforderungen, Gefahren**

Welche Probleme gilt es anzugehen, auf welche Mehrwerte ist zu hoffen? Die Qualitätssicherung, die durchgängige Nutzbarkeit und die nachhaltige Aufbewahrung (also: Pflege, „Kuration“) der Daten mitsamt Programmen, ohne die sie ihre Prozessierbarkeit verlieren, das sind Herausforderungen, die unmittelbar sinnfällig sind. Gefordert sind hier technische Entscheidungen: Formate müssen optimiert, Schnittstellen geschaffen, Prozessketten verbunden werden. Gebraucht werden aber auch umfangreiche Verständigungsprozesse, nämlich Verabredungen in Fachgemeinschaften: Welche gemeinsamen Metadaten und Beschreibungen werden genutzt, um digitale Ressourcen findbar zu machen? Wie werden Datenqualitäten gekennzeichnet? Für welche Forschungsformen sind welche digitalen Artefakte state of the art nutzbar? Wie steht es um Teilen, Nachnutzen und Fusion bzw. die Übersetzbarkeit von Forschungsdatenbeständen ineinander – und wie schafft Digitalität dann auch methodisch und im Ergebnis einen Gewinn?

Daten, Wissen und Methoden durch ein öffentlich finanziertes Forschungsdatenmanagement in der Wissenschaft selbst zu halten, ist auch wissenschaftspolitisch wichtig. Denn die Privatisierung von Daten bzw. Datendiensten gehört zu den großen Gefahren für die öffentliche Wissenschaft im Digitalzeit-

telle, eine weltweite OPEC, wollen. Hier lautet das Stichwort „Datensouveränität für die Forschung“. Denn Wissenschaft benötigt technisch und rechtlich die Grundlagen dafür, Daten möglichst ungehindert zu Forschungszwecken zu teilen.

Künstliche Intelligenz zum Beispiel: Derzeit redet man über (und investiert in) leistungsfähige Software. Aber auf welchen Daten werden die intelligenten Algorithmen laufen? Die Qualität des Ergebnisses hängt von der Qualität der Ausgangs- und Metadaten, der daten-

gen für gute Daten – Gewinnung, Qualitätssicherung, nachhaltiger Zugang, Standardisierung und Verknüpfung – wird „KI“ nichts leisten. Neben neuen IT-Professuren wird also ein neuer Gestaltungswille in Sachen Verknüpfbarkeit von Forschungsdaten gebraucht.

Dass Digitalpolitik nicht aus Abwarten bestehen kann, spricht sich herum. So haben sich Bund und Länder auf die Schaffung einer Nationalen Forschungsdateninfrastruktur – „NFDI“ – verständigt. Es handelt sich um ein neuartiges und im internationalen Ver-



alter. Hier drohen Intransparenz, Zugangshürden, unkontrollierbare Kosten – und auch: fehlende Nachprüfbarkeit, damit aber der Verlust wissenschaftlicher Standards von Forschung. Tatsächlich findet eine schleichende Enteignung bzw. Aneignung öffentlicher Wissenschaft bereits in dramatischem Umfang statt: Globale Internetgiganten (Softwarekonzerne, Großverlage) sammeln mittels kostenloser Publikationsdienste, Repositorien und sozialer Netze das Wissen europäischer Wissenschaftlerinnen und Wissenschaftler ein, um es für eigene, kommerzielle Zwecke zu nutzen. Dropbox, Google Drive, Academia.edu, Figshare heißt die Spitze des Eisbergs. Unlängst hat Microsoft mitgeteilt, für die sagenhafte Summe von 6,4 Milliarden EUR die zuvor altruistisch auftretende Software-Entwicklungsplattform GitHub zu kaufen, auf welcher die Wissenschaft bisher weltweit Computerprogramme teilte. Das Drama der teuren Lizenzpakete von Elsevier, SpringerNature und Wiley (wie überhaupt das Problem der Monopolbildung einer Datenwirtschaft, die Produzenten und Nutzer gleichermaßen an sich zu binden vermag) ist bekannt. Weniger beachtet werden die digitalisierten Alltags- und Laborwerkzeuge. Dass die ubiquitäre Nutzung von PDF-Dateien im Bibliotheksbereich Wissenschaft von der Preisgestaltung der Firma Acrobat abhängig macht, fiel erst auf, als diese ein Mietmodell einführte. Aber auch Messgeräte können zur Black-Box werden und von kommerziellen Mitspielern abhängig machen, wenn diese Geräte Daten mittels proprietärer (also intransparenter) Software erzeugen und Prozessdaten sowie Messergebnisse nicht bei der öffentlichen Forschungseinrichtung, sondern in der Cloud des Herstellers speichern. Für Spezialsoftware, deren Details nur eine Firma kennt, gilt Ähnliches. Zunehmend bieten private Akteure auch Datenauswertungen an. So lassen sich mehr oder weniger solide Unternehmen in Deutschland für die Aufbereitung von Standard-Erhebungen buchen; zuweilen werden aber auch online gedungene Internet-Klickarbeiter aus Übersee zur Auswertung von Fragebögen eingesetzt. Und auch die aktuelle Diskussion über windige Tagungs- und Publikationsmärkte (Stichwort „Raub-Verlage“) hat gezeigt, dass, was innerwissenschaftlich

klappen mag (etwa „Open Access“), ungewollte Ökonomisierungs-Effekte zeitigt. Drastischer ausgedrückt: Digitalität schafft eine nach unten offene Richter-Skala nicht nur was Verwertungsinteressen angeht, sondern auch Richtung Fake. Gegenüber beidem gibt ein bundesweit vernetztes Datenmanagement dem Wissenschaftssystem Handlungsfähigkeit. Seriöse Industriepartnerschaften können davon profitieren.

Datenschutz, gleichzeitig aber auch

»Ein ganzes soziales Subsystem – das, was wir als Wissenschaft kennen – verwandelt nicht nur seine Werkzeuge, sondern auch seine Basis.«

Rahmenbedingungen für eine verantwortliche Nutzung auch kritischer Daten (vor allem letzteres fehlt im Big Data-Bereich), und dazuhin Datensicherheit (also Schutz von Spionage, Datendiebstahl und Daten-Sabotage) sind weitere große Fragen, auf welche die Antwort nur heißen kann: aktives Forschungsdatenmanagement – und zwar in koordinierter, integrierter Form und in den Händen der Wissenschaft selbst. Durch die Presse gehen digitale Attacken auf öffentliche Versorgungsnetze und auf Unternehmen, aber auch deutsche Forschungseinrichtungen sind Angriffen ausgesetzt. Digitalität sorgt für Verletzlichkeiten, die man im Papier- und vor-digitalen Laborzeitalter nicht kannte.

Dass auch das Forscherverhalten sich hierauf einstellen muss (Stichworte „Fehlverhalten“ und „Ethik“), ist ebenso klar, wie dass über die Folgen von Digitalität als Faktor fast aller Formen und Prozesse der Forschung seinerseits geforscht werden muss. Ein ganzes soziales Subsystem – das, was wir als Wissenschaft kennen, – verwandelt nicht nur seine Werkzeuge, sondern eben auch seine Basis. Ausgang offen. So sind es insgesamt 13 grundlegende Empfehlungen, die der Rat für Informationsinfrastrukturen als beratende Stimme für Politik und Wissenschaft im Jahr 2016 zur Reorganisation des Umgangs mit Forschungsdaten in Deutschland wie auch im europäischen und globalen Kontext gegeben hat.

#### **Infrastrukturlösungen als Weg zu mehr Datensouveränität**

Forschungsdatenmanagement – ein Hebel also für die ganz schweren Fragen. Denn „von allein“ lösen diese sich

nicht. Laufenlassen ist für den digitalen Forschungsstandort Deutschland wie auch für Europa keine Option. Dabei war von den explodierenden Kosten für den enormen Mehraufwand, den digitale Prozesse mit sich bringen, noch gar nicht die Rede. Gerade die Datenhaltung, die Langzeitsicherung, die Tatsache, dass sich auch vor Ort nichts (dank stabiler Materialität) von selbst „aufbewahrt“, macht Digitalität unkalkulierbar teuer. Da die Budgets von Bund und Ländern als öffentliche Träger der Wissenschaft sich nicht vervielfachen werden, zwingt der digitale Wandel also auch zur Priorisierung. Hierzu sollten Entscheidungen im System am besten selbstbestimmt und möglichst nah an den Forschungsbelangen getroffen werden, und also im fachnahen Austausch über Forschungsdatenmanagement.

Für Deutschland befindet sich die Nationale Forschungsdateninfrastruktur (NFDI) auf der Schwelle zur Realisierung. Geplant ist eine nationale Integrations- und Vernetzungsaktivität, die bottom up nicht primär „Blech“, sondern Prozessklärung und Experteneinsatz fördert – also menschliche Arbeit an forschungseigenen digitalen Diensten, an dem unendlich vielen Kleinklein, das für qualitativ hochwertige Daten und wissenschaftlich wichtige Synergien in der Datennutzung dringend gebraucht wird. Fachgemeinschaften sollen Konsortien bilden können, die auf nachhaltige Kooperation, Perspektivplanung und dauerhafte Dienste-Portfolios für ganze fachlich-thematische Domänen angelegt sind. Langfristiges Engagement der Wissenschaftlerinnen und Wissenschaftler selber wird dabei, so ist zu hoffen, mobilisiert.

Datensouveränität, breite Nutzbarkeit von Daten und maximale Qualität der Daten(prozesse) und damit der Forschung sind gemeinsames Ziel. Ein Schlusstrich unter alle Probleme wird die NFDI natürlich nicht sein. Eher der sprichwörtliche Anfang. Und auf jeden Fall ein Stück Digitalitäts-Ermöglichungspolitik. Und dies mitten im Wandel, was Steuerung (auch Selbststeuerung) bekanntlich erschwert. Im Falle der Energiewende hatte man vor lauter Energiegewinnung die Transport- und Speichermedien, also die Infrastruktur, vergessen. Diesen Fehler dürfen wir im Bereich der Informationsinfrastrukturen nicht wiederholen.