

Umgang mit Forschungsdaten in den Naturwissenschaften

Eine Fallstudie aus der Mikrobiologie

Dr. Sibylle Kaspar

Einleitung

Forschungsdaten sind so vielfältig wie die verschiedenen Disziplinen, innerhalb derer wissenschaftlich gearbeitet wird. Sie stellen die Basis zur Beantwortung wissenschaftlicher Fragestellungen dar, entstehen zum Beispiel bei Messungen, Analysen, Beobachtungen, Erhebungen oder Befragungen und liegen inter- und interdisziplinär in sehr verschiedenen Formen und Formaten vor. Diese Daten werden mit unterschiedlichen Begriffen wie zum Beispiel Messdaten, Rohdaten, Quelldaten oder Primärdaten betitelt. Im deutschsprachigen Raum wird hauptsächlich von Forschungsdaten gesprochen.¹

Forschung dient der Fortentwicklung von Wissen und damit zur Weiterentwicklung der Gesellschaft durch Beantwortung von Fragestellungen. Die generierten Daten müssen aus unterschiedlichen Gründen langfristig zur Verfügung stehen. Dazu gehören zum Beispiel die Nachvollziehbarkeit, Verifizierbarkeit und Integrität im Sinne einer «guten wissenschaftlichen Praxis» (siehe S.355), Verhinderung oder Aufdeckung von Fälschungen, wissenschaftsgeschichtliche Beweggründe, Einmaligkeit beziehungsweise Unwiederholbarkeit und erneute disziplinäre wie auch interdisziplinäre Nutzung (Nachnutzung). Im Rahmen von Projekten, die durch staatliche Förderorganisationen finanziert wurden, besteht ein begründetes gesellschaftliches Interesse an den entstandenen Daten – nicht nur, weil die Projekte sehr geldintensiv sind, sondern weil die gewonnenen Daten und daraus resultierende Erkenntnisse auch einen berechtigten Anteil am öffentlichen Kulturgut ausmachen.²

Verlässlichkeit, Nachvollziehbarkeit und Integrität von Forschungsdaten sind von Behörden geforderte Gütekriterien, um einen gesicherten und strukturierten Umgang mit den gewonnenen Daten zu gewährleisten. Die entsprechenden Richtlinien und Forderungen entstanden nach Bekanntwerden von verschiedenen Fällen

1 Böttner, S.; Hobohm, H.-C.; Müller, L. (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef 2011, Kap. 1.1.

2 Osswald, A.; Scheffel, R.; Neuroth, H. et al.: Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen. In: H. Neuroth et al. (Hg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Göttingen 2012, Kap.1.

von Fälschungen in der Wissenschaft. Aussenstehende können Qualität und Aussagekraft wissenschaftlicher Daten nicht einschätzen und nur wenige Spezialisten sind fähig, neue Erkenntnisse zu beurteilen und im Kontext einzuordnen. Es ist daher eine Pflicht der Gesellschaft und der Wissenschaftsgemeinschaft gegenüber, einen angemessenen Umgang mit Forschungsdaten zu pflegen und zu garantieren. Doch auch wenn man redlich handeln möchte, können sich unwissentlich Fehler ergeben, die zum Beispiel durch Unterlassung oder Unwissenheit aufgetreten sind. Von Forschenden muss und kann ein sorgfältiger und gewissenhafter Umgang mit Daten verlangt werden. Das setzt jedoch auch eine Problemsensibilisierung und die Kenntnis möglicher Handlungsweisen voraus. Eine sorgfältige Aufzeichnung von wissenschaftlichen Daten während der Projektphase verhindert Fehler und Missverständnisse und gewährleistet die Nachvollziehbarkeit und spätere Verfügbarkeit derselben.

Aus diesen Gründen wurden in den letzten 15 Jahren verschiedene Empfehlungen und Grundsatzpapiere verfasst, Konzepte erstellt und Initiativen gegründet.³ Kernpunkte dieser Richtlinien sind unter anderem die Sicherstellung eines effizienten und professionellen Umgangs mit wissenschaftlichen Daten,⁴ ein offener und freier Zugang zu Daten, die aus öffentlich geförderter Forschung entstanden sind, die Berücksichtigung des gesetzlichen Rahmens, der Rechte der Wissenschaftler und fachspezifischer Aspekte,⁵ Standardisierung von Aufbewahrung, Sicherung, Metadatenvergabe und die Entwicklung von Methoden zu einer adäquaten Qualitätssicherung.⁶ Daran schliessen sich die aktuellen Gesamtkonzepte und Positionspapiere zur Schaffung einer geeigneten Informationsinfrastruktur an.⁷

Langzeitarchivierung von Forschungsdaten

Unter Langzeitarchivierung verstehen Neuroth et al.⁸ ganz pragmatisch und doch sehr treffend «die für den jeweiligen Kontext definierte Zeitspanne der Bewahrung digitaler Objekte auch über grundlegende technologische und soziokulturelle Wandlungsprozesse hinaus. Mit ihr wird ermöglicht, den Zugriff auf und die Nachnutzung von Forschungsdaten auch in der Zukunft zu sichern».

3 Z. B. 9, 5, 1, 2, 8.

4 Allianz-Initiative: Allianz Schwerpunktinitiative «Digitale Information» (2008).

5 Allianz der deutschen Wissenschaftsorganisationen, 2010. Grundsätze zum Umgang mit Forschungsdaten.

6 DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Unterausschuss für Informationsmanagement (2009).

7 WGL (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V./Leibniz Gemeinschaft). Kommission Zukunft der Informationsstruktur: Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder, April 2011. Übergreifende Empfehlungen zu Informationsinfrastrukturen, 2011. Positionspapier des Wissenschaftsrates. Drs. 10 466–11.

8 Neuroth, Langzeitarchivierung.

Doch zur Realisierung einer guten Langzeitarchivierung müssen noch einige Herausforderungen bewältigt werden.⁹ Es müssen zum Beispiel Kriterien definiert werden, die zur Auswahl der archivwürdigen Forschungsdaten verwendet werden können. Wie im klassischen Archivwesen muss auch hier eine Auseinandersetzung mit dem Thema Bewertung stattfinden: Was wird für zukünftige Forscher, die Wissenschaftsgeschichte und die Gesellschaft interessant sein? Die zukünftige Nutzbarkeit und Interpretierbarkeit der gespeicherten Daten müssen durch standardisierte Metadaten (Erschließung) und durch die Dokumentation der technischen und organisatorischen Rahmenbedingungen (Bestandserhaltung) gewährleistet sein. Kooperationen auf institutioneller, nationaler und internationaler Ebene sind für eine erfolgreiche Langzeitarchivierung von Forschungsdaten massgebend. Im Idealfall entstehen standardisierte Strukturen und Verfahren, die eine zunehmend globalisierte Forschungsgemeinschaft berücksichtigt.

Der Kulturwandel im Umgang mit Forschungsdaten ist ohne Konzept, spezifisches (technisches) Fachwissen und Informationsspezialisten nicht zu bewerkstelligen. In Deutschland ist mit dem Kompetenznetzwerk Nestor¹⁰ ein Kooperationsverbund zur digitalen Langzeitarchivierung entstanden, innerhalb dessen eine Zusammenarbeit von Bibliotheken, Archiven, Museen und führenden Experten zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen erfolgt. Die Deutsche Initiative für Netzwerkinformationen (DINI) hat sich zum Ziel gesetzt, die «Entwicklung der Informations-Infrastrukturen an den Hochschulen sowie regional und überregional zu fördern», um dadurch die Informations- und Kommunikationsdienstleistung zu verbessern.¹¹

Für eine Bedarfsanalyse, zur Ermittlung der Rahmenbedingungen und Nutzenziele zum Thema Forschungsdatenmanagement wurden in den letzten Jahren auf nationaler und internationaler Ebene verschiedene Umfragen und Konzeptstudien¹² durchgeführt. Dabei entstanden verschiedene Pläne und Bestandsaufnahmen, wie zum Beispiel das Gesamtkonzept für die Informationsinfrastruktur in Deutschland der Kommission «Zukunft der Informationsinfrastruktur» (KII),¹³ die Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz (ikeep, 2008)¹⁴ oder die

9 Neuroth, Langzeitarchivierung.

10 www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html (Zugriff am 26.7.2012).

11 Deutsche Initiative für Netzwerkinformationen e.V. (DINI). Positionspapier Forschungsdaten, 2009.

12 Z.B. SURFfoundation (Feijen, 2011), ikeep (2008), Konzeptstudie Primärdateninfrastruktur Chemie (TIB, FIZ Chemie, Uni Paderborn, 2012).

13 WGL, Gesamtkonzept.

14 Ikeep, Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz, Version 1.4, 2008.

Nestor-Bestandsaufnahme zur Langzeitarchivierung.¹⁵ Die Leibniz-Gemeinschaft (WGL) hat ganz aktuell ein Konzept zur Schaffung eines Leibniz-Forschungsverbundes Science 2.0 ausgearbeitet, mit dem Ziel «web 2.0-Konzepte in wissenschaftliche Forschungs- und Publikationsprozesse zu integrieren».¹⁶ Beispiele für internationale Berichte und Studien auf diesem Gebiet sind der *PARSE.Insight survey report* (2009) und die Studie der *SURFfoundation* (2011).¹⁷

Die technische Grundlage des Managements von Forschungsdaten bilden die Repositorien.¹⁸ Sie werden für die Langzeitarchivierung von Forschungsdaten, zur gemeinsamen Datenhaltung sowie zum Austausch und zur kooperativen Nutzung einer Forschungs-*Community* eingesetzt.¹⁹ Aschbrenner und Neuroth definieren «[...] Repository als eine Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit digitaler Objekte sowie für ihre Interpretierbarkeit zum Zwecke der Nutzung durch eine bestimmte Zielgruppe [...] übernommen hat».

Jedoch existieren national und international noch nicht für alle Fachbereiche solche Repositorien, und die, die schon existieren, sind unabhängig voneinander entstanden und befinden sich in unterschiedlichsten Kontexten. Allgemeingültige Definitionen und Standards gibt es derzeit noch nicht.²⁰

Forschungsdatenmanagement an Universitäten

Gemessen an der Anzahl Wissenschaftler stellen die Universitäten den grössten Anteil der Forschungs- und Entwicklungsgemeinschaft weltweit dar und produzieren eine signifikante Menge an neuen Daten.²¹ In den meisten Fällen wird diesen nach deren Publikation keine Beachtung mehr geschenkt. Stattdessen werden im Bedarfsfall eher ressourcenintensive Neumessungen in Kauf genommen.²² Missstände im Umgang mit Forschungsdaten sind leider eher noch die Regel als die Ausnahme. So berichtet die Technische Informationsbibliothek (TIB) (TIB Hannover, FIZ Chemie Berlin & Universität Paderborn, 2010, S. 5) bezüglich Forschungsdaten in der Chemie: «Der bisherige Umgang mit Forschungsdaten in der Chemie beinhaltet keine

15 Neuroth, Langzeitarchivierung.

16 Leibniz Gemeinschaft. Leibniz-Forschungsverbund Science 2.0 – Konzeptvorschlag.

17 PARSE.Insight, Survey report 2009. (D3.4); Feijen, M.: What Researchers want. Utrecht, SURF Foundation (2011).

18 Rümpel, S.: Data Librarianship – Anforderungen an Bibliothekare im Forschungsdatenmanagement. Diplomarbeit, Fachhochschule Potsdam 2012.

19 Aschbrenner, A.; Neuroth, H.: Forschungsdaten-Repositorien. In: Büttner/Hobohm/Müller, Forschungsdatenmanagement.

20 Aschbrenner/Neuroth, Forschungsdaten-Repositorien.

21 Thessen, A. E., Patterson, D. J.: Data issues in the life sciences. *ZooKeys* 150 (2011), 15–51.

22 Rümpel, S.: Der Lebenszyklus von Forschungsdaten. In: Büttner/Hobohm/Müller, Forschungsdatenmanagement, Kap. 1.1.

allgemein anerkannten Standards hinsichtlich einer Nachnutzbarkeit oder langfristigen Verfügbarkeit. Überwiegend existiert keine Qualitätssicherung, keine gesicherte Langzeitarchivierung, kein gesicherter Nachweis sowie keine Erschließung der Forschungsdaten und somit keine Datensicherheit.»²³

Die Universitäten übernehmen (mit einzelnen Ausnahmen)²⁴ nicht die Verantwortung, Massnahmen zur Verbesserung der Lage zu entwickeln. Die meisten Universitäten besitzen zwar eine Richtlinie zur Einhaltung der guten wissenschaftlichen Praxis, um sich gegen Anschuldigungen zur Fälschung abzusichern, haben aber keine übergreifende *Policy* zur Langzeitarchivierung von Daten. Die Forschenden verwenden dafür oft institutseigene Angebote und Lösungen, die jedoch nicht die Anforderungen von Repositorien an Nachhaltigkeit oder Überregionalität erfüllen.²⁵

Wie sieht die derzeitige Praxis im Umgang mit Forschungsdaten in einem Labor an einer Universität aus? Gibt es bereits eine Sensibilisierung seitens der Forschenden für diesen Themenbereich? Welche Anforderungen werden an die Beteiligten gestellt? Durch eine Fallstudie wird in dieser Arbeit exemplarisch Einsicht genommen und Erkenntnisse gewonnen, wie der gegenwärtige Umgang mit Daten innerhalb eines Universitätslabors ist und inwieweit dort ein Handlungsbewusstsein im Vergleich zum derzeitigen Forschungsstand eingetreten ist.

Ein Labor im Fachbereich Mikrobiologie – eine Fallstudie

Rahmen und Vorgehensweise

Als empirisches Beispiel zum Umgang mit Daten, die während des Forschungsprozesses entstehen, dient eine universitäre Arbeitsgruppe im Bereich Mikrobiologie. Der Lehrstuhl bearbeitet den Forschungsschwerpunkt mikrobielle Antibiotikaproduzenten. Es handelt sich dabei zum einen um Grundlagenforschung, zum anderen um angewandte Forschung in Kooperation mit externen Firmen. Geldgeber für die Forschungen sind neben der Universität die DFG, das BMBF, der SFB, die EU, Landesstipendien und externe Firmen im Rahmen einer Zusammenarbeit.

Mithilfe eines Interviews mit einer Privatdozentin des Lehrstuhls vor Ort wurde die Datenthematik eruiert und diskutiert. Zur Vorbereitung wurde ein Fragenkatalog zur konkreten Praxis von Entstehung und Umgang mit Forschungsdaten ausgearbeitet. Aspekte zum Forschungsprozess, der Art der Aufzeichnungen, der Nach-

23 TIB (Technische Informationsbibliothek) Hannover, FIZ Chemie Berlin & Universität Paderborn. Konzeptstudie «Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie». Paderborn 2010, 5.

24 Z.B. hat die Niedersächsische Staats- und Universitätsbibliothek in Göttingen die Initiative ergriffen (Neuroth, Langzeitarchivierung, 25).

25 Winkler-Nees, S.: Stand der Diskussionen (National). In: Neuroth, Langzeitarchivierung, Kap. 2.1.

vollziehbarkeit, deren Dokumentation und Aufbewahrung gaben dazu den Rahmen. Die Wahl fiel auf dieses Labor, da die Autorin dort ihre Diplomarbeit angefertigt hat. Durch den Einblick in die Fachbereiche Biologie und Informationswissenschaft war ihr die Möglichkeit gegeben, durch eine Schnittstellenfunktion einen wechselseitigen Eindruck der wissenschaftlichen Kulturen zu vermitteln und so zum gegenseitigen Verständnis beizutragen.

Die Auswertung des Gesprächs wurde mit archivischen Grundthemen verflochten und die Ergebnisse anschliessend mit verschiedenen Richtlinien zu wissenschaftlicher Arbeitsweise und Integrität, dem ISO-Standard zur Aktenführung (ISO 15489) und der Guten Laborpraxis (GLP) verglichen und in Bezug gesetzt.

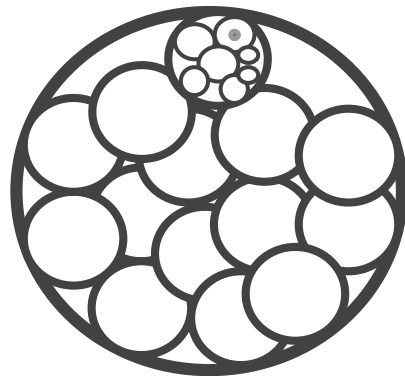
Arbeitsweise eines naturwissenschaftlichen Forschungsprojekts mit Fokus auf den Informationsfluss und Umgang mit Forschungsdaten

Ausgangslage eines naturwissenschaftlichen Forschungsprojekts ist das Vorhandensein eines grossen Gesamtthemas, das in verschieden grosse Unterprojekte unterteilt wird. Das kann modellhaft mit einer Himbeere verglichen werden, die aus verschiedenen Teilkügelchen besteht, die zusammen eine ganze Beere ergeben (Abb. 1). Die einzelnen Projekte stehen also nicht alleine, sondern immer im Zusammenhang mit Vorgänger-, Parallel- und Nachfolgeprojekten.

Zu Beginn der Bearbeitung eines Projekts (im Modell Grösse 2) bekommt der Forschende Informationen zum Vorgängerprojekt, im Idealfall vom Vorgänger selbst. Bei kleineren Projekten erfolgt dies durch den Betreuer. Informationen werden meist mündlich gegeben und durch niedergeschriebene wissenschaftliche Arbeiten ergänzt. Oft müssen Publikationen ausreichen, um an das Vorgängerprojekt anschliessen zu können.

Während des Forschungsprozesses entstehen Planungsdaten, Messdaten und Auswertungsdaten. Diese werden handschriftlich im Laborjournal festgehalten oder

Abb. 1: Modell eines Forschungsprojekts.
Äusserer Kreis (Grösse 1): Gesamtprojekt.
Dieses kann über Jahre/Jahrzehnte hinweg bearbeitet werden. Nächstkleinere Kreise (Grösse 2): Teilprojekte, die von Doktoranden oder Postdoktoranden bearbeitet werden.
Diese beinhalten kleinere Kreise und stellen Projekte verschiedenen Umfangs dar, die von Diplomanden, Masterstudenten, Bachelorstudenten oder Praktikanten durchgeführt und vom Doktoranden oder Postdoktoranden betreut werden. Der Grundbaustein eines Projekts ist das Experiment. Dieses ist als dicker, ausgefüllter Punkt dargestellt. (Quelle: Dr. Sibylle Kaspar.)



als Computerausdruck dort eingeklebt. Das Laborjournal enthält (wenn es sorgfältig geführt wird) alle Aufzeichnungen des Prozesses: Zielsetzung des Experiments, Durchführungsplanung, Ergebnisse und Kommentare zu den Ergebnissen.

Zur Konzeption eines Experiments sind verschiedene Informations- und Daten-Inputs nötig, die aus unterschiedlichen Quellen stammen, wie zum Beispiel Literatur, Daten von Vorgängern oder aus selbst durchgeführten Vorgängerexperimenten. Zur Durchführung der Versuche wird eine Vielzahl von Materialien und Geräten benötigt. Letztere werden durch diverse Eingaben wie zum Beispiel verwendete Menge an Substanzen, Volumenangaben, Geräteeinstellungen etc. (=Daten) gespeist. Während und nach dem durchgeführten Experiment werden durch Prozessieren wieder Daten generiert (Daten-Output), die als Input für ein Folgeexperiment dienen können oder eventuell aufzeigen, dass der eingeschlagene Weg in eine Sackgasse führt. Die gewonnenen Ergebnisse liegen in ganz unterschiedlichen Formen und Formaten, einem typischen mikrobiologischen/biotechnologischen Labor entsprechend, als neu generierte Daten vor.²⁶ Sie befinden sich auf verschiedenen Servern, in geschützten persönlichen Verzeichnissen oder im Laborjournal. Zur Bearbeitung dieser Dateien kommen bis zu 16 verschiedene Software-Programme zum Einsatz.²⁷ Damit unterscheidet sich die Arbeitsgruppe nicht von der Vorgehensweise der meisten anderen Universitäten. Eine Umfrage an drei australischen Hochschulen hat ergeben, dass die Situation dort ähnlich gestaltet ist.²⁸ Neben digitalen Daten existiert eine grosse Menge an nicht digitalen Dokumenten, so Handnotizen, Umfrage- und Evaluationsformulare, Bilder, ausgedruckte Diagramme und anderes mehr. Die von den Wissenschaftlern produzierten Anmerkungen liegen oft in einer unstrukturierten, nicht standardisierten Form vor.²⁹ Auch die digitalen Daten kommen in vielen verschiedenen Datenformaten vor und sind stark disziplinspezifisch geprägt.³⁰ Nicht kompatible Datentypen, veraltete Dateiformate, Software und Hardware stellen das Management von Forschungsdaten vor grosse Herausforderungen.³¹

Von den aufgezeichneten Daten liegen in der untersuchten Forschergruppe ungefähr 25 Prozent in elektronischer Form und 75 Prozent als handschriftliche Aufzeichnungen vor. Für Letztere dient das gebundene Laborjournal als Hauptaufzeichnungsort. Eine Richtlinie zur Erfassung der Daten existiert nicht. Die einzige Vorgabe,

26 Aufzeichnungs- und Datentypen sind z.B. Planungsdaten, handschriftliche Aufzeichnungen und Anmerkungen, Berichte und Präsentationen, Messdaten und prozessierte Daten, Plasmidkarten, DNA-Sequenzen oder Informationen zu Bakterienstämmen.

27 Z.B. MS Office, SciEd Clone Manager, ARTEMIS, Agilent (Bruker) Data Analysis (HPLC-MS), ChemSketch, NCBI-Blast etc.

28 Henty, M. et al.: Investigating Data Management Practices in Australian Universities. APSR, July 2008.

29 Verhaar, P. et al.: Data Curation in Arts and Media Research, Utrecht, 2010.

30 Henty, Investigating. PARSE. Insight, Survey Report 2009. (D3.4)

31 Feijen, What Researchers want.

die eingehalten werden muss, ist die, dass die handschriftlichen Aufzeichnungen in einem vom Departement zur Verfügung gestellten, gebundenen Heft (DIN A4) zu erfolgen haben. Wie und zu welcher Zeit die Einträge erfolgen, ist jedem Forschenden selbst überlassen. Die Archivierung der handschriftlichen Aufzeichnungen ist über das Laborjournal (10 Jahre Aufbewahrungspflicht, siehe unten) eindeutig geregelt. Die Speicherung der digitalen Daten erfolgt auf lokalen Rechnern, die von der Universität zur Verfügung gestellt werden, oder dem Departements-Server. Es steht ein Verzeichnissystem zur Verfügung, das eine grobe Struktur aufweist und durch unterschiedliche Zugänglichkeiten geregelt ist. Jeder Mitarbeitende hat Zugang zu einem allgemeinen Laufwerk und zusätzlich ein persönliches Verzeichnis, auf das nur er Zugriff hat. Letzteres ist eigenverantwortlich zu führen und unterliegt keiner Vorschrift. Oft werden für die gesamte Gruppe wichtige Daten dort gespeichert, obwohl sie auf das allgemein zugängliche Verzeichnis gestellt werden müssten. Geschieht beim Weggang der betreffenden Person kein Transfer, sind diese meist nicht mehr greifbar. Daten zu Arbeitsmaterial sollten zentral in einem Verzeichnis abgelegt werden, das allen Mitarbeitenden des Departements zugänglich ist. Leider werden diese nicht ausreichend gepflegt und daher fehlt es an Feinstruktur und Übersichtlichkeit. Die Information, die als Arbeitsgrundlage für das jeweilige Projekt notwendig ist, wird daher von jedem selbst beschafft und in einem eigenen Verzeichnis gespeichert. Als Quelle dienen meist die langjährigen Mitarbeitenden (meist Gruppenleiter) oder elektronische Versionen von Dissertationen. Daten, die an Messgeräten entstehen, werden auch dort ausgewertet. Die erforderliche Software steht aus Kostengründen nur lokal an den geräteeigenen Computern zur Verfügung. Jeder Wissenschaftler legt sich dort (meist nur mit dem Vornamen) selbst ein Verzeichnis an, in das er seine Daten speichert. Diese sind so hauptsächlich für die Bearbeitenden auffindbar und nachvollziehbar. Wenn diese den Lehrstuhl verlassen, sind sie für Nachfolgende meist unbrauchbar.

Aus verschiedenen Umfragen ist ersichtlich, dass dieses Szenario an Universitäten meist üblich ist.³² Kroll und Forsman sprechen im OCLC-Bericht³³ in Bezug auf die Datenspeicherung in diesen Institutionen von einem «hodge-podge of approaches». Mit diesem Sammelsurium an Speichermethoden kommen die Wissenschaftler während der laufenden Projekte gut zurecht. Es gibt jedoch einen grossen Unterschied zwischen der Projektphase und der Phase danach:³⁴ Während der Laufzeit eines Projekts legen sie sehr viel Wert auf den Schutz ihrer wertvollen Daten und verwalten sie eigenverantwortlich (auch wenn sie unstrukturiert und verstreut vor-

32 Henty, Investigating. PARSE.Insight; Kroll, S.; Forsman, R.: A Slice of Research Life: Information Support for Research in the United States. Report commissioned by OCLC Research in support with the RLG partnership. Dublin, Ohio, 2010.

33 Kroll/Forsman, A slice.

34 Feijen, What Researchers want.

liegen, die Forschenden kennen deren Relevanz und die Zusammenhänge). Ist diese beendet, geht der Fokus auf das nächste Projekt über und der Forschende möchte dieses durch eine mögliche Förderung finanziert wissen. Die Daten des Vorgängerprojekts verbleiben dort, wo sie als Letztes abgespeichert wurden: kein sicherer und vertrauenswürdiger Ort (im Sinne von Datenmanagement). Dabei handelt es sich oft um unzuverlässige und kurzlebige Speichermedien und es ist wahrscheinlich, dass sie über kurz oder lang nicht mehr zu gebrauchen sind. Obwohl sich alle dessen bewusst sind, finden keine Gegenmassnahmen statt.³⁵ Gründe für diese Vorgehensweise sind Zweckmässigkeit, Benutzerfreundlichkeit, Komfort, Backup-Möglichkeit und ein leichter Zugang.

Relevanz – Bewertung der Datensätze

Die Bewertung der Relevanz der Daten wird in der Literatur nicht eigenständig diskutiert. Sie wird implizit vorausgesetzt. Experimenten geht eine Fragestellung voraus. Wenn ein Experiment diese bestätigt oder untermauert, werden die Daten als relevant erachtet, gesichert und weiterverwendet. Relevant sind auch Daten, die eindeutig zeigen, dass die zuvor aufgestellte These nicht bestätigt wird. Die Studierenden (Praktikanten, Diplomanden, Master- und Bachelorstudenten) bearbeiten je einen Teilbereich des Themas eines Doktoranden oder eines Postdoktoranden. Sie verwalten ihre Aufzeichnungen selbstständig und geben sie nach Beendigung ihrer Arbeit dem Betreuer ab. Dieser wählt aufgrund der Relevanz (Nutzbarkeit) für seine Doktorarbeit aus, welche Daten er weiterverwenden kann.

Nachvollziehbarkeit

(Langzeit-)Verfügbarkeit im Sinne von Nachvollziehbarkeit bedeutet mehr als eine reine *Bitstream-Preservation* (Bewahrung der Bits und Bytes), sie bedeutet auch über soziokulturelle Veränderungsprozesse hinaus eine zukünftige Interpretierbarkeit und Nutzbarkeit der wissenschaftlichen Daten. Daher müssen die Forschungsdaten und auch der technische, organisatorische und inhaltliche Kontext durch (standardisierte) Metadaten beschrieben werden. Doch auch schon für kürzere Zeitspannen werden deskriptive, technische und administrative Metadaten benötigt.³⁶ Denn Ergebnisse und Daten, die nicht ausreichend beschrieben vorliegen, haben bereits einige Jahre nach deren Entstehung,³⁷ spätestens jedoch nachdem der Forscher die Arbeitsgruppe verlassen hat, nicht genügend Kontextinformation, um sie sinnvoll wieder einsetzen zu können. So müssen die deskriptiven Metadaten systematisch gegliederte Informa-

35 Feijen, What Researchers want.

36 Osswald/Scheffel/Neuroth, Langzeitarchivierung, Kap. 1.

37 Osswald/Scheffel/Neuroth, Langzeitarchivierung, Kap. 1.

tionen über Auswahlkriterien des Untersuchungsgegenstands, die Untersuchungs-, Erhebungs- oder Messmethode und deren Anwendung und die Ergebnisse selbst enthalten. Die Nachvollziehbarkeit von Daten wird in den Richtlinien zur «Guten wissenschaftlichen Praxis» (z. B. DFG) gefordert und sollte im Rahmen einer Metadatenvergabe gewährleistet werden können.

Die Situation der untersuchten Arbeitsgruppe treffend beschrieben, heisst es in einer Präsentation über den Umgang mit Forschungsdaten in den Wirtschaftswissenschaften auf dem Bibliothekartag im Mai 2012³⁸ zum Thema Nachvollziehbarkeit sinngemäss: Daten sind nicht selbsterklärend, es bedarf des Aufwands, um diese nachvollziehbar und nutzbar zu machen – wie ausführlich das geschieht, ist der Entscheidung jedes Einzelnen überlassen und erfolgt unsystematisch – die Schilderungen sind meist noch von Kollegen zu verstehen, aber nicht geeignet für eine allgemeine Nachnutzung. Eine Metadatenvergabe erfolgt selten, und dann meist auch nur kryptisch.

Publizierte Daten und Experimente sind im Allgemeinen nachvollziehbar aufgeschrieben und werden nach für diese Community gültigen Regeln erstellt.

Datenintegrität – Datensicherheit

Datenintegrität wird im analysierten Labor dadurch gewährleistet, dass jedes Experiment zwei Mal wiederholt werden muss, um das Ergebnis zu verifizieren. Nur nachvollziehbare, verifizierte und logisch begründete Ergebnisse werden von den Betreuern akzeptiert. Eine weitere Kontrollinstanz zur Datenintegrität stellen die Seminare dar. Die Forschenden präsentieren regelmässig ihre Ergebnisse vor der gesamten Arbeitsgruppe. Beim Verfassen der Abschlussarbeit werden die Ergebnisse und Daten durch den Departement-Leiter ein weiteres Mal kontrolliert.

Da das Laborjournal als Hauptaufzeichnungsort in Papierform vorliegt, ist es zunächst vor Datenzerstörung geschützt. Die von der DFG geforderten zehn Jahre Aufbewahrungsfrist überdauert auch ein Heft, das aus säurehaltigem Papier besteht. Nachträgliche Änderungen (oder kriminelle Fälschungsversuche) würden schnell und deutlich auffallen. Ausserdem werden sie für die geforderte Aufbewahrungszeit in einem verschlossenen Schrank gelagert und sind nur über eine gezielte Nachfrage erhältlich. Die Sicherung der elektronischen Daten ist nicht auf eine Langzeitarchivierung ausgelegt, aber für den Zeitraum von zehn Jahren in der Regel ausreichend. Die Daten können meist noch nach ein bis zwei Software-Updates ihres korrespondierenden Programms gelesen werden. Zur Absicherung gegen physikalische Schäden wird jede Nacht vom Departement-Gesamtlaufwerk ein Backup erstellt.

38 Bahls, D.: Leibniz-Informationszentrum Wirtschaft. Präsentation auf dem 101. Deutschen Bibliothekartag. Hamburg 2012.

Während der Zeit des wissenschaftlichen Arbeitens am Lehrstuhl hat jeder Forschende neben dem (für das Departement) öffentlichen Netzwerk einen nur ihm zugänglichen Speicherplatz. Nach Beendigung der wissenschaftlichen Arbeit und Verlassen der Arbeitsgemeinschaft werden die Daten der Wissenschaftler dem Systemadministrator des Departements zur Speicherung und Verwaltung übergeben. Sie werden in diesem Zustand bei jeder Migration mitgezogen, jedoch ohne Kontrolle, ob sie danach noch lesbar sind. Sie sind zugangssicher abgelegt und werden nur nach Einverständnis des Produzenten herausgegeben. Falls dieser nicht erreichbar ist, entscheidet der Gruppen- oder der Departement-Leiter. Die Abschlussarbeiten werden auf dem Netzwerk als PDF-Datei zur Konsultation bereitgestellt.

Vergleich der Situation mit Anforderungen und Empfehlungen von Standards

Der im Interview eruierte Umgang mit Forschungsdaten wird im Folgenden mit offiziellen Richtlinien und Normen der Deutschen Forschungsgemeinschaft (DFG) zur Sicherung der guten wissenschaftlichen Praxis,³⁹ den Grundsätzen und Verfahrensregeln der Akademien der Wissenschaften Schweiz⁴⁰ und fünf institutsinternen Richtlinien zur wissenschaftlichen Integrität verglichen.⁴¹ Des Weiteren erfolgt eine Gegenüberstellung mit den Standards ISO 15489:2002 (Teil 1 und 2)⁴² zur Schriftgutverwaltung und vier Dokumenten der OECD-Schriftenreihe über Grundsätze der Guten Laborpraxis (GLP)⁴³

Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis Grundprinzipien für wissenschaftliche Tätigkeit stellen Ehrlichkeit, Vertrauen und Verantwortung dar. Alle von der DFG oder anderen nationalen Wissenschaftsgemeinschaften, auch den Schweizer Akademien a+⁴⁴ abgeleiteten Richtlinien enthalten diese ethischen Grundwerte als Basis wissenschaftlichen Arbeitens. Kernpunkte dieses

-
- 39 DFG (Deutsche Forschungsgemeinschaft) (Hg): Sicherung guter wissenschaftlicher Praxis, Denkschrift. Empfehlungen der Kommission «Selbstkontrolle der Wissenschaft». Weinheim, Wiley-VCH 1998.
- 40 a+ Akademien der Wissenschaften Schweiz (Hg): Wissenschaftliche Integrität. Grundsätze und Verfahrensregeln, 2008.
- 41 Hier: Richtlinien der ETH Zürich und der Universitäten Genf, Tübingen, Ulm und Bern (zufällig ausgewählt).
- 42 Information und Dokumentation: Schriftgutverwaltung (DIN ISO 15489), Teil 1 und 2. Berlin 2002.
- 43 OECD-Schriftenreihe über Grundsätze der Guten Laborpraxis und Überwachung ihrer Einhaltung. Nr. 1: OECD-Grundsätze der Guten Laborpraxis, 1999. Nr. 4: Qualitätssicherung und GLP, 2001. Nr. 10: Die Anwendung der GLP-Grundsätze auf computergestützte Systeme, 1995. Nr. 15: Einrichtung und Betrieb von Archiven in Übereinstimmung mit den Grundsätzen der Guten Laborpraxis, 2007.
- 44 a+ Akademien der Wissenschaften Schweiz (Hg): Wissenschaftliche Integrität. Grundsätze und Verfahrensregeln, 2008.

Rahmens sind unter anderem, «lege artis» zu arbeiten, Ergebnisse aufzuzeichnen, diese selbstkritisch zu hinterfragen, strikte Ehrlichkeit, Zusammenarbeit innerhalb der Arbeitsgruppe, die Betreuung des wissenschaftlichen Nachwuchses (auch hinsichtlich der Sicherung der guten wissenschaftlichen Praxis), Umgang mit wissenschaftlichem Fehlverhalten und die Sicherung und Aufbewahrung von Primärdaten. In den DFG-Regularien wird empfohlen, dass «Primärdaten als Grundlage für Veröffentlichungen [...] auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für 10 Jahre aufbewahrt werden» sollen. Die Schweizer Akademien formulieren, dass «Daten und Materialien nach Abschluss des Projektes während einer für das Fachgebiet adäquaten Dauer aufbewahrt bleiben [sollen]». Diese allgemeinen Formulierungen ermöglichen einen grösseren Handlungsspielraum und eine grössere Flexibilität, schaffen jedoch andererseits durch die vage Formulierung Unsicherheiten und Interpretationsspielraum. In den Regularien wird an Ehrlichkeit, Verantwortungsbewusstsein, Redlichkeit, Selbstkritik und Vertrauen appelliert. Dies sind moralische Werte, sie können daher auch nur auf moralischer Ebene sanktioniert werden. Eine Bekanntmachung der Zuwiderhandlung und ein Ausschluss aus der wissenschaftlichen Gemeinschaft sind gegebenenfalls die Folge.

Viele Institutionen und Universitäten stellen im Internet ihre Richtlinien zur Verfügung und zeigen damit auf, dass sie die Empfehlungen zur Sicherung der guten wissenschaftlichen Praxis kennen, sie bekannt machen und ihren Bedürfnissen angepasst haben. Dies stellt zwar nicht sicher, dass sie auch von den Wissenschaftlern gelesen und beachtet werden, gibt aber der Universität bei einem Verstoss eine Möglichkeit, diesen zu sanktionieren, ohne selbst Schaden zu nehmen. Der Grundsatz zur wissenschaftlichen Redlichkeit und Ehrlichkeit und der Appell zu verantwortlichem Handeln ist allen gemeinsam. Sie legen Rahmenbedingungen fest, die eine integre und qualitativ hochwertige Forschung erlauben. Die Universität, zu der die hier untersuchte Arbeitsgruppe gehört, verlangt, dass die Richtlinie bei der Einstellung überdies unterschrieben wird. Die Arbeitsgruppe kennt die DFG-Richtlinie zur guten wissenschaftlichen Praxis⁴⁵ und hat diese in ihre Arbeitsweise integriert. Die praktische Umsetzung der Empfehlungen wird durch die Forschenden vorgelebt und während der Einarbeitungszeit den Nachwuchswissenschaftlern vermittelt. Ehrlichkeit und Redlichkeit werden zunächst einmal vorausgesetzt und würden erst dann angezweifelt werden, wenn ein (begründeter) Verdacht aufkäme.

45 DFG (Deutsche Forschungsgemeinschaft) (Hg): Sicherung guter wissenschaftlicher Praxis, Denkschrift. Empfehlungen der Kommission «Selbstkontrolle der Wissenschaft». Weinheim, Wiley-VCH 1998.

Vergleich des Standards ISO 15489:2002 zur Schriftgutverwaltung mit der Arbeitsweise in der untersuchten Forschergruppe

Bei ISO 15489:2002 handelt es sich um eine Norm für Grundsätze und Verfahren der Schriftgutverwaltung. Die Schriftgutverwaltung in einer Institution dient der Regulierung der Nutzung von Unterlagen und enthält unter anderem eine Aufstellung von Grundsätzen und Zielvorgaben, Festlegungen zu Verantwortlichkeiten, Befugnissen, Richtlinien und der Bereitstellung eines speziellen Schriftgutverwaltungssystems. Schriftgut wird als wertvolle Ressource verstanden, mit der sorgfältig umgegangen werden muss. Dieses «Betriebskapital» erlaubt es, die Geschäfte effizient und rechenschaftsfähig zu führen, eine gleichmässige Qualität, die Unterstützung von Führungsentscheidungen, gesetzliche Vorschriften zu erfüllen, Nachweise bei Rechtsstreitigkeiten vorweisen zu können und «das organisationsspezifische, persönliche und kollektive Gedächtnis zu bewahren».⁴⁶ Institutionen und Organisationen, die das internationale ISO-15489-Zertifikat zur Schriftgutverwaltung erhalten wollen, müssen alle geforderten Punkte erfüllen. Auch wenn die Aufzeichnung von Forschungsdaten keiner Norm unterliegt, so gibt es doch gemeinsame Aspekte, die in beiden Fällen zum Tragen kommen: Es muss eine adäquate Dokumentation von Daten erfolgen. Vergleicht man damit die Praxis in der Arbeitsgruppe, so fällt auf, dass Forschungsdaten hier ebenfalls das «Betriebskapital» darstellen, den Kern der Forschung ausmachen und eine sorgfältige Aufzeichnung notwendig ist zum Verfassen von Publikationen, zum Erbringen der Nachweispflicht und Ablegen von Rechenschaft. Sie erfüllen im Prinzip den groben Rahmen, den die ISO-Norm vorschreibt, jedoch ohne eine explizite Aufstellung von Grundsätzen und Richtlinien bezüglich der Datenaufzeichnung. Der Motor liegt mehr in der Forschernatur selbst und dem Ehrgeiz, gute, publizierbare Ergebnisse zu liefern. Forschende fühlen sich in erster Linie sich selbst und der Community verpflichtet und fokussieren auf ihr Ziel, nämlich brauchbare Ergebnisse zu erzielen, diese zu publizieren und die Anerkennung ihrer Leistungen innerhalb der Forschungsgemeinschaft zu erhalten. Dabei stehen die inhaltlichen Themen weit im Vordergrund, und die administrativen Fragen bezüglich Datenaufzeichnung, Speicherung, Langzeitverfügbarkeit, Sicherheit et cetera stellen sich erst dann, wann Probleme auftreten.

Anforderungen

In der ISO-Norm 15489-1/Punkt 7 werden die Anforderungen an die Schriftgutverwaltung wie zum Beispiel Festlegung, Dokumentation und Bekanntgabe von Grundsätzen, Zielvorgaben, Verfahren und Vorgehensweisen beschrieben, um «authentische, aussagefähige, zuverlässige und benutzbare Unterlagen [zu] erstellen und

46 Information und Dokumentation: Schriftgutverwaltung (DIN ISO 15489), Teil 1 und 2. Berlin 2002.

[zu] führen». Dabei wird ausführlich definiert, welche Merkmale das Schriftgut aufzuweisen hat. Ausserdem muss von den Organisationen bestimmt werden, welche Akten in welcher Form und Struktur erstellt werden müssen und wie die Metadatenvergabe zu erfolgen hat und verwaltet werden kann. Ausserdem müssen Wieder auffindbarkeit, Aufbewahrungsdauer und -bedingungen, Bearbeitungsprozesse und anderes festgelegt werden.

Die Daten, die vor, während und nach einem Forschungsprozess anfallen, müssen ebenfalls sauber dokumentiert werden und authentisch, zuverlässig, integer und benutzbar sein. Es wird jedoch nicht schriftlich festgehalten, wie dies zu erfolgen hat. Daher bleibt es jedem Wissenschaftler selbst überlassen, wie und in welchem Ausmass die Aufzeichnung erfolgt. Selbstverantwortung bekommt hier, systembedingt, einen höheren Stellenwert. Folgen einer schlechten Aufzeichnung sind Daten, die nicht verwendet werden können. Das hat in erster Linie jeder Einzelne selbst zu tragen, indem er entweder die Zeit aufbringen muss, Experimente wiederholt durchzuführen, oder im Falle einer Doktor-, Master- oder Bachelorarbeit eine schlechtere Note oder einen nicht verliehenen Titel in Kauf nehmen muss. In zweiter Linie sind Team und Geldgeber geschädigt, da mit einer adäquaten Aufzeichnungsmethode Ressourcenverluste vermieden und die Forschung schneller hätte vorangetrieben werden können. Sanktionen diesbezüglich sind in den seltensten Fällen schwerwiegend, daher ist hauptsächlich das Ernten der Früchte einer gelungenen Forschung, zum Beispiel eine sehr gute Dissertation oder eine Publikation in einem angesehenen Journal, Antriebsmotor für sorgfältiges Arbeiten.

Datenverwaltungssystem und Sicherheit

In ISO 15489:2002 werden die Merkmale von Schriftgutverwaltungssystemen beschrieben und detailliert festgehalten, welche Massnahmen für ein zuverlässiges, vollständiges, integriertes und kontrollierbares System, das im Einklang mit den Anforderungen des Regelungsumfeldes steht, getroffen werden müssen. Es sollte «den Zugang zu und das Wiederauffinden von Schriftgut, das für die laufende Geschäfts erledigung oder für einschlägige Nachweisanforderungen benötigt wird, zeitnah und effizient gewährleisten. [Es sollte] Zugangskontrollen beinhalten und anwenden, damit die Integrität der Unterlagen nicht beeinträchtigt wird.»

Wie bereits erwähnt, erfolgen 75 Prozent der Aufzeichnungen der Arbeitsgruppe handschriftlich in den Laborjournalen. Sie liegen chronologisch geordnet vor. Da es sich um ein Instrument zur täglichen Arbeit handelt, liegen diese meist auf dem Schreibarbeitsplatz der Forschenden oder in einer unabgeschlossenen Schublade und sind daher nicht explizit gegen unbefugte Einsicht geschützt. Nach Abschluss des Projekts kommen die Laborjournale, die meist nummeriert werden, in einen abgeschlossenen Schrank im Büro des Departement-Leiters.

Für die elektronischen Dokumente besteht für das gesamte mikrobiologische/biotechnologische Departement ein Dateiablagensystem. Dieses enthält unterschiedliche Verzeichnisse, deren Zugänglichkeit geregelt ist. Die Ablage ist nicht strukturiert und es wurde kein Verantwortlicher zur Einhaltung und Überwachung des Ablagesystems ernannt. Jeder Mitarbeitende hat ausserdem ein persönliches Verzeichnis, in dem er Daten speichert, die er für die Bearbeitung seines Projekts benötigt. Auch Informationen, die für alle wichtig sein könnten, werden dort erfasst. Das geschieht nicht aus Böswilligkeit, sondern ist eine Folge der Selbstverwaltung mit fehlenden Strukturvorgaben.

Bei Verlassen der Arbeitsgruppe werden ohne Triage alle Aufzeichnungen des entsprechenden Mitarbeiterverzeichnisses dem Computerbeauftragten des Departements übergeben und in einem «Archivierungsverzeichnis» auf dem Departement-Server abgelegt, zu dem nur der Computerbeauftragte Zugang hat. Es gibt folglich keine Regelung und Entscheidungshilfe zur Aussonderung von Daten, wie sie in der ISO-Norm unter 8.3.7 gefordert wird. Bei Migrationen werden diese Daten so lange mitgezogen, bis sie nicht mehr lesbar sind. Es wird angenommen, dass sie die Zehnjahresfrist überdauern. Zur Datensicherung wird vom Gesamtverzeichnis jede Nacht ein Back-up erstellt. Ein Plan zur Sicherung der Datenintegrität ist nicht vorhanden. Da bis anhin noch keine Probleme aufgetreten sind, hat sich auch keine Notwendigkeit gezeigt, dies zu ändern.

Bilanz

Da es in der untersuchten Forschergruppe kein schriftlich festgelegtes (Forschungsdaten-)Verwaltungssystem gibt, kann es auch nicht 1 : 1 mit den in der ISO-Norm vorgegebenen Prozessen zur Steuerung der Schriftgutverwaltung verglichen werden. Inhaltlich jedoch sind viele dieser Prozesse vergleichbar, wenn auch nicht so ausführlich geregelt. So erfolgt die Aufbewahrung der Forschungsdaten nach den durch die DFG herausgegebenen Empfehlungen. Damit werden schon zwei Punkte abgedeckt: Das Regelungsumfeld ist mit einbezogen und die Aufbewahrungsdauer der Aufzeichnungen festgelegt. Für elektronische Daten existiert eine Zugangsregelung und mit einem Back-up-System eine Absicherung gegen Verlust, Zerstörung oder einen Katastrophenfall. In ISO 15489:2002 werden verschiedene Techniken für eine Erfassung von Schriftgut vorgeschlagen (z. B. Klassifikation, Indexierung, Registrierung). Die handschriftlichen Aufzeichnungen werden in der Forschergruppe nach der logischen Struktur «Entstehungsdatum» aufgezeichnet. Die Erfassung der elektronischen Daten unterliegt keiner Struktur, was die Auffindbarkeit bestehender Daten deutlich erschwert, wenn nicht gänzlich unmöglich macht. Hier besteht Optimierungspotenzial. Anders sieht es aus, wenn es sich um Publikationen oder Abschlussarbeiten handelt. Dort ist eine Nachvollziehbarkeit der entsprechenden Experimente, Ergebnisse und

der anschließenden Diskussion eine Grundvoraussetzung zur Akzeptanz der Forschungsarbeit. Betreuer und Peergroup sind für eine entsprechende Ausbildung, Überwachung und Prüfung verantwortlich und bilden dabei das Regulativ.

Fazit

In der vorliegenden Arbeit wurde anhand einer Fallstudie die gängige Praxis zum Umgang mit Forschungsdaten beleuchtet und unter Einbeziehung aktueller Literatur diskutiert. Es konnte ein tiefer Einblick in die (Daten-)Welt eines naturwissenschaftlichen Labors gewonnen werden.

Information ist personengebunden

Während die Wissenschaftler für die Dauer der Bearbeitungsphase ihre Forschungsdaten genau kennen und einordnen können, liegen die Aufzeichnungen für Nachfolger oft ungenügend aufbereitet vor. Bei einer Suche nach Informationen wird meist auf bestehende Dissertationen, Diplomarbeiten oder Masterarbeiten zurückgegriffen. Die nicht verwendeten Daten und die Gründe dafür stehen meist nicht mehr zur Verfügung. Experimente, die nicht funktioniert haben, werden im schlechtesten Fall unwissentlich noch einmal durchgeführt, Ressourcen unnötig verbraucht, um (wiederum) festzustellen, dass dies nicht der richtige Weg ist. Eine Konsultation des Laborjournals ist aus praktischen Gründen nahezu unmöglich, denn man müsste Seite für Seite den Testverlauf rekonstruieren. Die elektronischen Daten sind auf dem Netzwerk-Server abgelegt und liegen dort relativ unstrukturiert vor. Stabile Informationsinseln bilden die Gruppenleiter, die in der Regel sehr lange am Institut verweilen und ein einzigartiges (notiertes und nicht notiertes) Wissen angesammelt haben. Viel Information ist somit personengebunden und geht mit deren Weggang unwiderruflich verloren. Der Einsatz eines elektronischen Laborjournals⁴⁷ könnte durch eine zentrale Datenhaltung und eine einheitliche Pflege der Daten und Dokumente zumindest teilweise zu einer Ablösung personengebundener Information führen. Dazu notwendig sind jedoch eine (zeitaufwändige und ressourcenverbrauchende) Erschließung mittels standardisierter Metadaten und eine aussagekräftige Struktur.

Bleibt die Frage, wie schwerwiegend die Ressourcenverschwendung zu bewerten ist. Im Gegensatz zu Wetterdaten oder Umfrageergebnissen sind Laborexperimente wiederholbar. Es muss eine Kosten-Nutzen-Rechnung aufgestellt werden, was teurer ist: ein Datenmanagement mit Metadatenvergabe oder eine Wiederholung

47 Potthoff, J. et al.: Elektronisches Laborbuch: Beweiserhaltung und Langzeitarchivierung in der Forschung. In: Schomburg, S.; Leggewie, C.; Lobin, H.; Puschmann, C. (Hg.): Konferenz: Digitale Wissenschaft. Stand der Entwicklung digital vernetzter Forschung in Deutschland. Köln 20.–21. September 2010. Köln ²2011, 149–156.

eines Experiments beziehungsweise eines ganzen Forschungsprojekts (womöglich mit neuerer Technologie).⁴⁸

Eigenverantwortlichkeit und Integrität

Neben Kontrollinstanzen wie Betreuer, Seminarvorträge, Abschlussarbeiten, Publikationen und Peergroup sind es auch sogenannte «weiche Faktoren» wie Sorgfalt, Genauigkeit und Verlässlichkeit, die die Güte der Daten prägen. Wenn Experimente und Daten nicht gut notiert sind, können sie nicht für Publikationen verwendet werden. Es handelt sich dabei «nur» um Ressourcenverschwendung. Die Integrität der Forschung und ihrer Daten erleidet damit in der Regel keinen Schaden. Dazu bräuchte es die bewusste Absicht, durch die Manipulation der Daten falsche Publikationen zu generieren. Das Interview zeigte auf, dass das Thema Forschungsdatenmanagement noch nicht im alltäglichen Gebrauch angekommen ist und langfristig ein Anstoß zum Umdenken und Handeln notwendig ist. Die Problemsensibilisierung ist in der Arbeitsgruppe aufgrund der geringen und noch leicht handhabbaren Datenmenge weniger ausgeprägt.

In der Gesamtbetrachtung kann festgestellt werden, dass in der untersuchten Arbeitsgruppe aus informationswissenschaftlicher Sicht Handlungsbedarf bezüglich Strukturierung, Erschließung und Archivierung von vorhandenen Daten besteht. Aus fachwissenschaftlichem Blickwinkel jedoch ist dieses System funktionstüchtig. Die Arbeitsweise und der Umgang mit den Forschungsdaten unterscheiden sich nicht von denen anderer, vergleichbarer Laboratorien weltweit. Es handelt sich bei der untersuchten Forschungsgruppe um ein funktionierendes, eigenverantwortliches System, das noch nicht in der *data-driven science*⁴⁹ angekommen ist, aber innerhalb «seiner» Community *lege artis* handelt und nach den Regeln der «Guten wissenschaftlichen Praxis» ehrliche Forschung betreibt.

Richtlinien und Regularien

Ein Vergleich mit dem Standard ISO 15489:2002⁵⁰ und den Grundsätzen zur «Guten Laborpraxis»⁵¹ zeigt, dass die dort verlangte Arbeitsweise starken Reglementierungen unterliegt und für ein universitäres Labor nicht geeignet wäre. Forschung ist von

48 Ludwig, J.: Zusammenfassung und Interpretation. In: Neuroth, Langzeitarchivierung, Kap. 15.

49 Hey, T.; Tansley, T.; Tolle, K.: Jim Gray on eScience: A Transformed Scientific Method. In: dies (Hg.): The Fourth Paradigm Data-Intensive Scientific Discovery. Redmond 2009, xix–xxxiii.

50 Information und Dokumentation: Schriftgutverwaltung (DIN ISO 15489), Teil 1 und 2, 2002. Berlin 2002.

51 OECD-Schriftenreihe über Grundsätze der Guten Laborpraxis und Überwachung ihrer Einhaltung. Nr. 1: OECD-Grundsätze der Guten Laborpraxis, 1999. Nr. 4: Qualitätssicherung und GLP, 2001. Nr. 10: Die Anwendung der GLP-Grundsätze auf computergestützte Systeme, 1995. Nr. 15: Einrichtung und Betrieb von Archiven in Übereinstimmung mit den Grundsätzen der Guten Laborpraxis, 2007.

sorgfältigem Arbeiten geprägt, aber auch von einer Art künstlerischer Freiheit, die neue Ideen kreiert. Schliesslich ist hier die Schmiede neuer Erkenntnisse, die nur so reglementiert wie nötig ist, um vertrauenswürdige und integre Daten zu liefern. Sie sollte so frei wie möglich sein, um neue Wege, neue Ideen, neue Gedanken nicht bereits im Keim zu ersticken, indem sie durch eine überregulierte Kontrollwalze unterdrückt werden. Forscher müssen einen Spagat vollziehen zwischen einer kontrollierbaren, strukturierten, nachvollziehbaren Welt und einer kreativen, intuitiven, ideenreichen schöpferischen Welt. Das ist die Kunst des Forschens!