

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Forschungsdatenmanagement und -infrastruktur im SFB 1187 “Medien der Kooperation”

Annette Strauchⁱ Michael Dahnkeⁱⁱ

2019

Zitiervorschlag

Strauch, Annette und Michael Dahnke. 2019. Forschungsdatenmanagement und
-infrastruktur im SFB 1187 “Medien der Kooperation”. *Bausteine
Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern* Nr. 2/2019: S. 54-60. DOI:
[10.17192/bfdm.2019.2.8071](https://doi.org/10.17192/bfdm.2019.2.8071).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱStiftung Universität Hildesheim. ORCID: [0000-0003-2355-325X](https://orcid.org/0000-0003-2355-325X)

ⁱⁱUniversität Siegen.

1 Einführung

Seit Mai 2016 leisten Forscherinnen und Forscher aus den Geistes-, Kultur-, Sozial- und Ingenieurwissenschaften einen Beitrag zum Verständnis der digitalen Gegenwarts-kultur im Sonderforschungsbereich *Medien der Kooperation* (SFB 1187)¹ der DFG. Das INF-Modul *Infrastrukturkonzepte für die Beforschung kooperativer Medien*² hilft beim Aufbau technischer Infra- und Beratungsstrukturen zum Forschungsdatenmanagement und wird vom Zentrum für Informations- und Medientechnologie (ZIMT)³ und dem Bereich Wirtschaftsinformatik der Universität Siegen⁴ getragen. Man orientiert sich an der universitären Forschungsdaten-Policy⁵ der Universität Siegen. In diesem Rahmen entwickelt ZIMT eine forschungsbegleitende Infrastruktur für die einzelnen Forscherinnen und Forscher beim Umgang mit ihren Forschungsdaten, und zwar von der Entstehung bis zur Langzeitarchivierung. Das ZIMT unterstützt seit Jahren insbesondere die medienwissenschaftliche Forschung an der Universität Siegen.

In einem ersten Schritt hatte das ZIMT die Bedürfnisse der Wissenschaftler und Wissenschaftlerinnen aus den Geistes-, Kultur-, Sozial- und Ingenieurwissenschaften sowie der Siegener Sozioinformatik erhoben. Dabei wurde zum einen auf Erfahrungen im Forschungsdatenmanagement zurückgegriffen (DINI⁶, DARIAH⁷, RADAR⁸, RDA⁹ und nestor¹⁰), zum anderen sollten fachspezifische Anforderungen in Kooperation von Universitätsbibliothek und ZIMT Siegen in zentrale Serviceleistungen umgewandelt werden, sodass diese auf die Forschergruppen im SFB 1187 – und darüber hinaus – entsprechend zugeschnitten wären. Die Entwicklung von Verfahren, Standards, technischen Infrastrukturen und Werkzeugen sollte also projektabhängig und interdisziplinär erfolgen, um die bestmöglichen Voraussetzungen sowohl für ein sinnvolles Forschungsdatenmanagement eines einzelnen Produzenten oder einer Produzentin im SFB als auch für kollaborative Forschungsk Kooperationen bereitzustellen. Generell ist der Charakter der Forschungsdaten qualitativer Art. Es handelt sich a) um textuelle Primärdaten - i. e. Protokolle, Briefe, Feldtagebücher, Feldnotizen und E-Mails -, b) textuelle Sekundärquellen wie Handbücher und Archivalien, c) visuelle Quellen wie Fotografien, Malereien, Zeichnungen und Bilder auf Flickr¹¹, d) Umfragen und Interviews und e) Social Media-Daten.

¹<https://www.mediacoop.uni-siegen.de>. Letzter Zugriff 10.06.2019.

²<https://www.mediacoop.uni-siegen.de/de/projekte/inf/>. Letzter Zugriff 10.06.2019.

³<https://www.zimt.uni-siegen.de/>. Letzter Zugriff 10.06.2019.

⁴<http://www.cscw.uni-siegen.de/team/pipek/>. Letzter Zugriff 10.06.2019.

⁵https://www.universi.uni-siegen.de/usi_forschungsdaten-policy-1.pdf. Letzter Zugriff 10.06.2019.

⁶<https://dini.de/>. Letzter Zugriff 10.06.2019.

⁷<https://www.dariah.eu/>. Letzter Zugriff 10.06.2019.

⁸<https://www.radar-service.eu/de>. Letzter Zugriff 10.06.2019.

⁹<https://rd-alliance.org/>. Letzter Zugriff 10.06.2019.

¹⁰https://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html. Letzter Zugriff 10.06.2019.

¹¹<https://www.flickr.com/>. Letzter Zugriff 10.06.2019.

2 Vorgehen

Der methodischen Vielfalt versuchte man von Seiten des ZIMT mit einer Unterscheidung in vier Forschergruppen Rechnung zu tragen, um die Unterstützung projektspezifischer gestalten zu können. Diese Gruppen sind a) einige Forscherinnen und Forscher, die davon ausgehen, mit der Erstellung von Videos und deren anschließendem Schnitt einen forschungsrelevanten Beitrag zum Verständnis der Verwendung von Mobiltelefonen im privaten Umfeld von Menschen zu liefern, und sich selbst als Medienethnologen bezeichnen, b) hauptsächlich mit gedruckten Archivalien Arbeitende, diese sprechen sich selbst eine hermeneutische Herangehensweise zu, c) sich selbst als Medien Gestaltende beziehungsweise in etwas Intervenierende Bezeichnende, und d) digitale Editionsphilologinnen und Editionsphilologen.

Es fand zunächst ein Dialog mit allen Akteurinnen und Akteuren der insgesamt 13 Teilprojekte statt. Die Klärung der Bedürfnisse für forschungsunterstützende Dienste orientierte sich bei diesem Austausch eng an den Forscherinnen und Forschern. Sie erfolgte zweiteilig erstens von Juni 2017 bis August 2018 als geschlossene Umfrage (LimeSurvey) und zweitens um die Jahreswende 2018/2019 in teilstandardisierten Interviews mit Vertreterinnen und Vertretern der Teilprojekte. Hieraus resultierten einige generalisierbare sowie weitere projektspezifische Anforderungen und Lösungen. Die Anforderungen werden immer aus der Perspektive der Arbeitsweise der Forscherinnen und Forscher heraus betrachtet, wobei das *Curation Domain Model* zur Beschreibung des Lebenszyklus von Forschungsdaten nach Treloar u.a.¹² für einzelne, möglichst exemplarische Anwendungsfälle als Orientierungsrahmen im Forschungsdatenmanagement herangezogen wird.

3 Anforderungen und Lösungen

Aus zwei Bedarfserhebungen 2017 und 2018 bei den einzelnen Teilprojekten des SFBs resultierten neben anderem erstens Anfragen zur Möglichkeit der kollaborativen Bearbeitung von Texten und der ebenfalls kollaborativen Verschlagwortung von Bild-, Ton- und Audiomaterial. Zweitens wurde immer wieder das Bedürfnis nach Speichermöglichkeiten für große Datenmengen genannt und drittens mit wachsendem Bewusstsein der Vertreterinnen und Vertreter der Teilprojekte für die wissenschaftlichen Vorteile einer Langzeitarchivierung auch diese zunehmend häufiger thematisiert. Viertens werden Möglichkeiten der Kodierung der verschiedenen Forschungsdaten nachgefragt und fünftens realisiert das INF-Teilprojekt Publikationen einzelner Teilprojekte mit dem OJS (Open Journal Systems).

¹²Treloar, Andrew, Groenewegen, David und Cathrine Harboe-Ree (2007). *The Data Curation Continuum*. D-Lib Magazine, 13 (9/10), <http://www.dlib.org/dlib/september07/treloar/09treloar.html> . Letzter Zugriff 10.06.2019.

Für die kollaborative Bearbeitung von Texten werden die Forscherinnen und Forscher aktuell mit dem File-hosting-Dienst zum Austausch von Dateien *Sciebo*¹³ vertraut gemacht, der eine Freigabeoption zum Teilen von Dokumenten mit Hochschulangehörigen und externen Projektpartnern und einen Dokument-Editor beinhaltet, der das simultane Bearbeiten von Text-, Tabellen- und Präsentationsdokumenten ermöglicht. Möglichkeiten der kollaborativen Verschlagwortung von Bild-, Ton- und Audiomaterial werden gegenwärtig geprüft. Neben den genannten Diensten bietet *Sciebo*¹⁴ 30 GB kostenlosen Speicher für jeden angemeldeten Nutzer.

Mit *DSpace*¹⁵ befindet sich derzeit ein Dienst für die Langzeitarchivierung von Forschungsdaten aller Formate in der Erprobung. Bei diesem wie auch dem beschriebenen *Sciebo* sind die besonderen Ansprüche der sozialwissenschaftlichen Forschung an die Sicherung der Privatsphäre der in die Forschung einbezogenen Probandinnen und Probanden zu berücksichtigen. Kurz gesagt geht es darum, jederzeit den kontrollierten Zugang auf die Daten zu gewährleisten und jeden unbefugten Zugriff zu verhindern. Dazu schreiben die Anbieter und Anbieterinnen von *Sciebo*: »Mit Standorten ausschließlich in NRW unterliegt *Sciebo* den strengen deutschen Richtlinien zu Datenschutz und Datensicherheit. Universitäre Betreiber garantieren, dass von Anbieterseite keinerlei kommerzielle Interessen an den Daten bestehen.«¹⁶ Eine gleichermaßen sichere Lösung erwarten die Nutzer auch bei der Verwendung von *DSpace*. Gegenwärtig wird hier in intensiven Diskussionen das Spannungsfeld zwischen dem Datenschutz einerseits, dem beispielsweise mit der Verschlüsselung der Daten vor dem Hochladen genügt werden könnte, und der grundlegenden Funktion eines Archivs andererseits ausgelotet. Ein Archiv unterscheidet sich gerade dank der Durchsuchbarkeit der Bestände, vulgo Daten, von einer bloßen Sammlung von Artefakten.

Für die Kodierung qualitativer Daten werden von einzelnen Teilprojekten bereits Datenbanken auf Basis von MAXQDA benutzt und weitere Teilprojekte haben mit dem Wachstum ihrer Datenbestände und der Erkenntnis der Notwendigkeit einer effizienten Verwaltung derselben mittlerweile ebenfalls deutliches Interesse an einer Einführung in diese Software/Technologie und der Installation auf ihrer Hardware zum Ausdruck gebracht.

Zwei Teilprojekte sind ebenfalls ein gutes Beispiel dafür, wie ein Bewusstsein für die Notwendigkeit bestimmter Kenntnisse und Fertigkeiten zum Umgang mit bestimmten Daten manchmal bereits von Beginn an vorhanden ist, in anderen Fällen aber auch erst im Verlauf der Forschung entsteht. So entschieden sich die Vertreter der Rekonstruktion des Briefwechsels der drei Herausgeber des Nachlasses Ludwig Wittgensteins (P01) bereits zu Beginn ihrer Arbeit für eine Einführung in den Quasistandard der geisteswissenschaftlichen Textkodierung TEI¹⁷. Einem Wissenschaftler und Wissen-

¹³<https://www.sciebo.de/features/index.html>. Letzter Zugriff 10.06.2019.

¹⁴<https://www.sciebo.de/features/index.html>. Letzter Zugriff 10.06.2019.

¹⁵<https://wiki.duraspace.org/display/DSPACE>. Letzter Zugriff 10.06.2019.

¹⁶<https://www.sciebo.de/features/> Letzter Zugriff 10.06.2019.

¹⁷<http://www.tei-c.org>. Letzter Zugriff 10.06.2019.

schaftlerinnen von B01, der zu Ludwig Heinrich Nicolays *Episteln* forsch, konnten die Vorzüge der Verwendung einer entsprechenden Textkodierung hingegen erst jüngst vermittelt werden. Mit ihm wird jetzt eine Transformation der bereits vorhandenen Daten vom proprietären XLS- in das quelloffene XML-Format TEI vorgenommen. Um hier zukünftig noch rascher so bedarfsgerecht wie möglich agieren zu können, werden als weiteres Kommunikationsmittel regelmäßige Rundschreiben vom INF-Projekt zu neuesten technischen Entwicklungen und Lösungen verschickt. Diese können bei weiteren Mitgliedern von Teilprojekten möglicherweise auch ein Bewusstsein von ihnen zukünftig noch zu lösender Aufgaben schaffen. Die Öffentlichkeitsarbeit des SFB unterstützt das INF-Projekt schließlich mit der Bereitstellung der Software OJS (Open Journal Systems)¹⁸, mit der »Media in Action. Interdisciplinary Journal on Cooperative Media«¹⁹ publiziert wird.

Die alldem zugrunde liegende Aufgabe der Schaffung einer nachhaltigen Forschungsdateninfrastruktur und der Entwicklung eines fachspezifischen Forschungsdatenmanagements wäre aber unvollständig ohne die beiden Eckpfeiler eines ausgeprägten Bewusstseins der Relevanz des Forschungsdatenmanagements bei den Vertreterinnen und Vertretern der einzelnen Teilprojekte und die Erstellung von Forschungsdatenmanagementplänen durch dieselben gelöst. Hier ist in einer Reihe von Einzeltreffen in den letzten Monaten mit den Mitgliederinnen und Mitgliedern der Teilprojekte erstens nochmals das Bewusstsein für die Notwendigkeit des Forschungsdatenmanagements per se geweckt beziehungsweise erhöht worden. Zweitens sind mit ihnen gemeinsam unter Verwendung der zuvor für die Bedürfnisse des SFB implementierten Software *RDMO*–*Research Data Management Organiser*²⁰ jeweils projektspezifische Forschungsdatenmanagementpläne erstellt worden.

4 Beispiele aus der Praxis

Der wissenschaftliche Kontext im SFB 1187 reicht von Projekten wie A01 zur Geschichte digital-verbundener Medien zwischen Spezialisierung und Universalisierung²¹ über B04 zu digitalen Öffentlichkeiten und gesellschaftlichen Transformationen im Maghreb²² bis zur Rekonstruktion des Briefwechsels der drei Herausgeber des Nachlasses Ludwig Wittgensteins (P01)²³. Geplant ist hierbei die Sammlung, Erschließung, Analyse und Präsentation des Korrespondenznetzwerks der Herausgeber der Werke Ludwig

¹⁸<https://pkp.sfu.ca/ojs/>. Letzter Zugriff 10.06.2019.

¹⁹<https://www001.zimt.uni-siegen.de/ojs/index.php/mia/about/editorialTeam>. Letzter Zugriff 10.06.2019.

²⁰<https://github.com/rdmorganiser/rdmo> und <https://rdmorganiser.github.io/>. Beide letzter Zugriff 10.03.2019.

²¹<https://www.mediacoop.uni-siegen.de/de/projekte/a01/>. Letzter Zugriff 10.06.2019.

²²<https://www.mediacoop.uni-siegen.de/de/projekte/b04/>. Letzter Zugriff 10.06.2019.

²³<https://www.mediacoop.uni-siegen.de/de/projekte/p01/>. Letzter Zugriff 10.06.2019.

Wittgensteins. Dieser Briefwechsel ist das dritte wesentliche Element der Rezeptionsgeschichte Wittgensteins neben erstens der Edition Wittgensteins eigenen Briefwechsels und zweitens seines eigenen Nachlasses. Die Forschungsdaten, mit denen in P01 gearbeitet wird, sind erstens die Briefe aus den Bibliotheken als eingescannte PDF und deren Transkriptionen, zweitens XML-erschlossene Textdateien von *Wittgenstein-Source*²⁴ und drittens Archivalien der *Bergen Text Edition* (BTE). Die Briefe werden mit ihren Metadaten erfasst, ausgewählte Quellen mit Abstracts, Inhaltsschlagworten (Personen, Orte, Werke, Ereignisse, Sachschlagworte) angereichert, ediert und kommentiert. Die Quellen sollen mit den Daten in *Wittgenstein-Source*²⁵ verknüpft werden. *Wittgenstein-Source* bietet freien Zugriff auf Wittgensteins Werke, inkl. der *Bergen Faksimile Edition* (BFE) und *Bergen Text Edition* (BTE), die 5000 Seiten von Wittgensteins Nachlass enthalten. Dafür wurden Gespräche mit *net*²⁶ bezüglich zwei Kostenmodelle geführt, die bislang nicht zu einer Lösung geführt haben. Die Verwaltung der dabei entstehenden Forschungsdaten im Rahmen eines nachhaltigen Forschungsdatenmanagements während der Laufzeit des SFBs und danach ist eine der zentralen Aufgaben des INF-Teilprojektes. Dem dienen die Schaffung einer nachhaltigen Forschungsdateninfrastruktur für die einzelnen Forscherinnen und Forscher im SFB 1187 und die Entwicklung von Lösungen für das Forschungsdatenmanagement. Die methodische Vielfalt ist neben der Unterschiedlichkeit der Forschungsdaten der im Sonderforschungsbereich versammelten Teilprojekte als eine große Herausforderung für das Datenmanagement qualitativ ausgerichteteter geistes-, kultur- und sozialwissenschaftlicher Forschung hervorzuheben.

5 Ausblick

Das INF-Projekt setzt sich auch weiterhin mit den fachspezifischen Anforderungen der verschiedenen Teilprojekte des SFB 1187 auseinander, um für diese forschungsunterstützende Dienstleistungen zu konzipieren. Dabei sollen die zuvor generischen und nicht disziplinspezifischen Basisinfrastrukturen, Prozesse und Services aus der ersten Förderphase noch stärker an den fachspezifischen Belangen der vier eingangs genannten Arten von Forscherinnen und Forschern ausgerichtet werden. Dies geschieht immer in Kooperation mit internen Partnern an der Universität Siegen oder externen Dienstleistern, wie das Beispiel P01 gezeigt hat. Die größte Herausforderung des INF-Projektes beim Umgang mit den Daten wird auch in der zweiten Förderphase die Diversität der Ansprüche und benutzten Software in den einzelnen Teilprojekten sein. Trotzdem soll den Ansprüchen eines vernetzten, globalen und vor allem nachhaltigen Forschungsdatenmanagements genügt werden, indem man sich noch intensiver mit anderen Institutionen, die ebenfalls mit dem Forschungsdatenmanagement

²⁴www.wittgensteinsource.org. Letzter Zugriff 10.06.2019.

²⁵www.wittgensteinsource.org. Letzter Zugriff 10.06.2019.

²⁶<https://www.netseven.it>. Letzter Zugriff 10.06.2019.

befasst sind, vernetzt. Zu nennen sind hier neben anderen die *Consortium of European Social Science Data Archives*²⁷, das *GESIS – Leibniz-Institut für Sozialwissenschaften in Mannheim*²⁸ und der *Rat für Sozial- und WirtschaftsDaten*²⁹. Innerhalb des INF-Projektes, das dem nachhaltigen Umgang mit Daten dient, fördern die genannten Teilprojekte die Entwicklung und Umsetzung eines Datenmanagementkonzeptes innerhalb des SFBs. Neben einem nachhaltigen Forschungsdatenmanagement ist die Weiterentwicklung und der Betrieb einer technischen Infrastruktur wichtig, die unter anderem aus den Forschungspraktiken und Bedürfnissen der Forscherinnen und Forscher hervorgehen. Zusammenfassend betreut das INF-Projekt alle Teilprojekte beim Umgang mit Daten, bei der Datenanalyse, dem Umgang mit den Tools sowie zu Fragen der Speicherung und Nachhaltigkeit. Manchmal kann es sinnvoll sein, mit externen Dienstleistern wie *net7* oder *FuD* zu kooperieren, wobei die entstehenden Kosten zu berücksichtigen sind.

²⁷<https://www.cessda.eu/>. Letzter Zugriff 10.06.2019.

²⁸<https://www.gesis.org/institut/impressum/>. Letzter Zugriff 10.06.2019.

²⁹<https://www.ratswd.de/ratswd/auf-einen-blick>. Letzter Zugriff 10.06.2019.