

AR 3265

Fabian Cremer, Lisa Klaffki und Timo Steyer

Redaktionssache Forschungsdaten.

Ein Servicekonzept zur Forschungsdatenpublikation in den Geisteswissenschaften

Zusammenfassung: Der Beitrag fokussiert die Publikation von Forschungsdaten in den Geisteswissenschaften und vertritt die These, dass die Forschungsdatenpublikation im Wissenschaftsbetrieb als Dienstleistung realisierbar ist. Ausgehend von hinderlichen Barrieren im System und förderlichen Modellen in Verbundprojekten wird das Konzept der Forschungsdatenredaktion diskutiert. Dieses Konzept basiert auf einer redaktionellen Bearbeitung der Forschungsdaten nach dem Forschungsprozess und der Veröffentlichung als eigenständiger Publikationsform.

Schlüsselwörter: Forschungsdatenmanagement, Datenpublikation, Digitale Geisteswissenschaften

Research Data Publication – A Matter for an Editorial Board?

A service concept for research data publication in the humanities

Abstract: The article focuses on the publication of research data in the humanities and argues that data publication in the scientific ecosystem is feasible for a service model. The service concept of an editorial team preparing research data for publication is discussed on the basis of barriers in the system and success stories in collaborative projects. This concept is based on an editorial processing of the research data following the research phase and understands data publication as an autonomous publication form.

Keywords: Research Data Management, Data Publication, Digital Humanities

1 Produktionsbedingungen geisteswissenschaftlicher Forschungsdaten

Das Kompositum Forschungsdaten bezeichnet ein aktuelles Debattenthema in den Geistes- und Kulturwissenschaften – dies wirkt sich auch unmittelbar auf die Arbeit der Forschenden und die Angebote von Hochschulen sowie Gedächtniseinrichtungen aus.¹ Gleichfalls nehmen Forschungsdaten eine immer wichtigere Rolle im (digitalen) Forschungsprozess ein. Trotz dieser Relevanz sind die mit dem Phänomen Forschungsdaten verbundenen Begriffe und Prozesse in den Geisteswissenschaften weder abschließend definiert noch standardisiert:

¹ Vgl. dazu den vorliegenden Band sowie die Themenhefte der o-bib. Das offene Bibliotheksjournal, Bd. 5. Nr. 2 und 3 (2018), hierzu: Stäcker und Steenweg (2018) IV–V.

„In dieser noch nicht abgeschlossenen Begriffsgenese wird der Begriff Forschungsdaten auf unterschiedliche Qualitäts- und Aggregatsstufen von Daten, angefangen bei Grunddaten wie Digitalisaten oder maschinell erstellten Volltexten bis zu nach komplexen Richtlinien erschlossenen Datensammlungen oder Datenrepräsentation bzw. Datenbanken, angewendet und ist nicht frei von Widersprüchen und Akzeptanzproblemen.“²

Nähert man sich den geisteswissenschaftlichen Forschungsdaten von ihrer Erzeugung her, stellen sie sich als ein durch die Forschenden kreierte Produkt dar, das von manuellen und kreativen Prozessen geprägt ist und nur selten einer ausschließlichen maschinellen Generierung entstammt. Johanna Drucker schlägt daher den Begriff „Capta“ vor, welcher die aktive und bewusste Auswahl der Information unterstreicht.³ Christoph Schöch charakterisiert die geisteswissenschaftlichen Daten als „Smart Data“, die mit Auszeichnungen, Klassifikationen und Relationen angereichert sind.⁴ Bisherige Publikationen, die als geisteswissenschaftliche Forschungsdaten gewertet werden können, entsprechen diesen Charakteristika („Smart Capta“): es sind häufig sogenannte „Sammlungen“ aufwändig aufbereiteter Daten – Texteditionen, Datenbanken, Objektsammlungen.⁵

Diese Überlegungen, die Anforderungen der datengestützten Methoden in den Digital Humanities sowie die etablierten Praktiken in den Geisteswissenschaften führen daher zu der Schlussfolgerung, dass Forschungsdaten aufbereitet werden müssen, um Gegenstand von komplexen Analysen geisteswissenschaftlicher Fragestellungen zu werden. Damit können sie grundsätzlich auch als eine eigenständige und rezeptionsfähige wissenschaftliche Leistung gelten, die schon aufgrund ihrer wissenschaftlichen Relevanz und der Ermöglichung der Nachnutzung publiziert werden sollten.

Die Grundlagen für eine Publikation von Forschungsdaten bestehen zum einen in der Einhaltung technischer Mindestanforderungen, der wissenschaftlichen Qualität sowie im Vorhandensein einer entsprechenden digitalen Forschungsinfrastruktur, bestehend aus Datenzentren, Repositorien und Archivsystemen sowie entsprechenden Rechercheinstrumenten und Nachweissystemen. Diese Infrastruktur ist zu einem großen Teil bereits vorhanden, die einzelnen Lösungen jedoch noch nicht kohärent verbunden und nicht allen Forschenden bekannt. Das Konzept der nationalen Forschungsdateninfrastruktur (NFDI)⁶ kündigt an, sowohl die verteilten Lösungen als auch die

² Cremer et al. (2018) 143.

³ Drucker (2011).

⁴ Schöch (2013). Schöch argumentiert, dass Big Data für die Humanities gar nicht erstrebenswert ist, sondern Big Smart Data notwendig wäre.

⁵ Zu Forschungsdatensammlungen vgl. Schmunk und Oltersdorf (2016) sowie Palmer (2004).

⁶ RfII (2016).

Begrenzung auf einzelne Disziplinen oder Institutionen überwinden zu wollen. Nicht nur die Infrastruktur, auch viele Werkzeuge und neue Methoden wurden in den letzten Jahren vor allem im Feld der Digital Humanities entwickelt. Doch es fehlt nicht selten an den Forschungsdaten, um die Methoden anzuwenden, die Werkzeuge einzusetzen und die Infrastrukturen nutzen zu können. In den Geisteswissenschaften ist dies zwar teilweise auf einen Rückstand im Bereich der Digitalisierung von Kulturgut zurückzuführen,⁷ aber auch von Seiten der Forschenden lassen sich aktuell wenig Anfragen oder Praktiken für Datenpublikationen beobachten. Das aktuelle Dilemma besteht darin, dass auf der einen Seite die Relevanz von Forschungsdaten zum Aufbau von technischen und personellen Unterstützungsangeboten führt, auf der anderen Seite aber die Akzeptanz und Nutzung dieser Angebote zumindest vorerst gering bleibt.

2 Barrieren der Forschungsdatenpublikation

Neben den skizzierten gesellschaftlichen und forschungspolitischen Argumenten für die Veröffentlichung und offene Bereitstellung von Forschungsdaten gibt es für die Forschenden jedoch auch Gründe, die eigenen Daten nicht zu publizieren. Weiter sorgen eine Reihe an strukturellen und kulturellen Hindernissen dafür, dass gerade in den Geisteswissenschaften die Publikation von Forschungsdaten die Ausnahme darstellt.⁸

Das Reputationssystem der Wissenschaft bildet auch im Bereich der Forschungsdaten die handlungsleitende Motivation. So verknüpft Borgman die zentralen Anreize gegen das Teilen von Forschungsdaten mit der Reputation, die grundsätzlich auf der Publikation von Ergebnissen basiert.⁹ Forschende werden nach ihrer Bibliografie beurteilt, nicht nach ihren Leistungen im Datenmanagement. Auch fehlt es an etablierten Verfahren der Qualitätsbemessung – Peer Review oder andere Begutachtungsverfahren sind für Forschungsdaten nicht etabliert. Bei Berufungen und Entfristungen spielen Forschungsdatenpublikationen (und andere Aspekte offener Wissenschaft) bestenfalls eine nachrangige Rolle. Dies gilt ebenso für die Bewertung von Antragstellenden bei Drittmittelanträgen. Um den Aufwand bei der Aufbereitung und Dokumentation zu rechtfertigen, müsste diese Leistung jedoch entsprechend anerkannt werden. Weiter werden Forschungsdaten aber auch unveröffentlicht gehalten, um als Wertobjekt für Tausch oder weitere Forschungszuwendungen zu dienen.¹⁰ Das ökonomische System der Wissenschaft bietet für die Datenproduzierenden damit derzeit keine Motivation für eine Datenpublikation.

⁷ Vgl. Klaffki et al. (2018) 5f.

⁸ Siehe hierzu die Übersicht von Kaden (2018).

⁹ Borgman (2010) 7.

¹⁰ Vgl. Klump (2017).

Eines der hinreichenden Argumente für die Publikation von Forschungsdaten ist die erforderliche Transparenz, Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsergebnissen, die mitunter nur durch den Zugang zu den zugrundeliegenden Daten gewährleistet werden kann. In den Geisteswissenschaften ist dies häufig jedoch gar nicht der Fall, denn „während eine wissenschaftliche Erkenntnis selbst publiziert werden muss, um gelten zu können, ist dies für die ihr zugrundeliegenden Forschungsschritte nicht erforderlich. Für den Weg zur Erkenntnis reicht meist eine kurze Schilderung als Beleg des wissenschaftlichen Vorgehens.“¹¹ Die fehlende Notwendigkeit zur Offenlegung der Daten für die Nachvollziehbarkeit einer hermeneutisch orientierten Analyse konstatiert bereits Andorfer 2015, denn „die Nachvollziehbarkeit der Ergebnisse sollte durch die (formalisierte) Offenlegung der verwendeten Quellen und der konsultierten Literatur (Fußnotenapparat, Quellen- und Literaturverzeichnis) samt einer verschriftlichten Begründung der Auswahl der Quellen und Literatur gegeben sein.“¹² Für eine Publikationspraxis geisteswissenschaftlicher Forschungsdaten müsste sich daher das Verständnis von Forschungsdaten als Element in der Rezeption und Nutzbarkeit grundlegend wandeln.

Die Aufbereitung von Forschungsdaten erfordert nicht nur einen erheblichen Aufwand, sondern auch ein umfangreiches Portfolio an Kompetenzen. Grundlagen des Forschungsdatenmanagements spielen in den klassischen Curricula der Geisteswissenschaften keine Rolle und sind auch in den Digital Humanities nur implizit oder untergeordnet vorhanden.¹³ Die Trainings zum Forschungsdatenmanagement, etwa in den Doktorandenkolloquien an den Hochschulen, können dies nur bedingt ausgleichen.

Normdatenauszeichnung, Formatvalidierung und andere Standardisierungsmaßnahmen erfordern eine informationswissenschaftliche Fachkompetenz, die selbst eigene Ausbildungs- und Studienformate füllt, vor allem im bibliothekarischen Bereich („Data Librarian“). Auch die Metadatenvergabe und Klassifikation wird im klassischen Publikationswesen von den Bibliotheken übernommen, weil hier entsprechende Kompetenzen traditionell verortet sind. Jenseits der technischen und informatischen Kompetenzen werden von den Forschenden die fehlende Kenntnis der juristischen Rahmenbedingungen als größere Herausforderung beschrieben.¹⁴ Urheberrecht, Datenbankrecht und Datenschutzrecht führen zu aufwendigen Einzelfallprüfungen und in den Geisteswissenschaften verkompliziert sich die Lage durch die heterogenen Datensammlungen mit Objekten, Repräsentationen und Reproduktionen noch mehr. Selbst wenn sich in Zukunft eine

¹¹ Kaden (2017).

¹² Andorfer (2015) 19.

¹³ Vgl. bspw. Sahle (2013) 10.

¹⁴ Vgl. die Daten aus dem Projekt bwFDM-Communities: <http://bwfdm.scc.kit.edu/daten>, Bericht: Tristram et al. (2015).

Datenkompetenz („Data Literacy“) als gemeinsame Basis ausgebildet, bleiben informatische, informationswissenschaftliche und juristische Kompetenzen nicht ausreichend abdeckt. Auch könnte die Akzeptanz von Forschungsdatenpublikationen erhöht werden, wenn diese nicht nur in Repositorien, sondern auch in den etablierten Nachweissystemen verzeichnet werden. Insgesamt kann daher konstatiert werden, dass Forschungsdaten von den Forschenden alleine *nicht* publikationsreif und nachnutzbar erzeugt werden können, weil es an Motivation, Auftrag, Notwendigkeit und Kompetenz fehlt.

3 Institutionalisation des projektgebundenen Forschungsdatenmanagements

In der Genese der Entwicklung von Infrastrukturen für das Forschungsdatenmanagement spielen große Verbundforschungsprojekte wie die DFG-geförderten Sonderforschungsbereiche (SFB) eine einflussreiche Rolle. Bei den SFBs werden seit 2007 Teilprojekte zur Informationsinfrastruktur (INF-Projekte) gefördert, die sich unter anderem um Zugang, Erschließung, Pflege und Nachnutzung von Forschungsdaten kümmern.¹⁵ INF-Projekte sind kein Phänomen der naturwissenschaftlich orientierten Vorhaben, sondern sogar in den Geisteswissenschaften am prominentesten vertreten: In den rund 20–25 Sonderforschungsbereichen im Bereich der Geisteswissenschaften sind in der Hälfte der Vorhaben INF-Projekte verortet (SFB/INF, 2013: 23/13; 2018: 24/11).¹⁶

Die bisher konzipierten INF-Projekte sind dabei sehr unterschiedlich ausgestaltet, was auch in der Aufgabe begründet ist, die spezifischen Anforderungen des jeweiligen SFBs zu erfüllen.¹⁷ Wenngleich der primäre Fokus der INF-Projekte auf dem Management und der Nutzung der Daten innerhalb der Verbünde liegt, wird auch die Forschungsdatenpublikation zunehmend häufiger an den Aufgabenbereich der INF-Projekte adressiert.¹⁸ Neben der Bereitstellung notwendiger Infrastrukturkomponenten wie Webportale und Repositorien reichen die Aktivitäten von der „Entwicklung von Redaktions- und Publikationsworkflows sowie entsprechenden Werkzeugen und Diensten über die Unterstützung der Forscher bei der Veröffentlichung bis hin zur Unterstützung der Bibliothek bei der Verknüpfung von Literatur und Daten oder der Verlinkung und Referenzierung von Daten für das Retrieval.“¹⁹

¹⁵ Siehe DFG (2018) 10f.

¹⁶ Die Daten von 2013 stammen aus der Untersuchung von Engelhardt (2013) 115.

¹⁷ Eine kontinuierliche systematische Untersuchung oder koordinierte Begleitung der zentralen Projekte in Sonderforschungsbereichen erfolgt derzeit übrigens nicht. Die punktuelle Ausnahme bildet ein bundesweites Treffen 2013, siehe Engelhardt (2013) 120ff.

¹⁸ 2013 haben ein Drittel der INF-Projekte Datenpublikation als Aufgabe angegeben, siehe Umfrage von Engelhardt (2013) 117. 2018 geben von den elf SFBs mit INF-Projekten in den Geisteswissenschaften sechs INF-Projekte explizit die Aufgabe Datenpublikation in der Teilprojektbeschreibung oder Selbstdarstellung an, bei vier weiteren bleibt es unklar. Datensatz: Cremer (2018), als HTML: <https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-D464-5/data>.

¹⁹ Engelhardt (2013) 117.

In großen Verbundvorhaben können Forschungsdatenpublikationen über zentrale Einheiten innerhalb der Projektstrukturen realisiert werden. Bei einer Übertragung dieser Unterstützungsmodelle auf institutionelle Ebenen wie Hochschulen und außeruniversitäre Forschungseinrichtungen werden jedoch strukturelle Hindernisse und konzeptionelle Grenzen dieser projektorientierten Ansätze deutlich:

Ein gemeinsames Charakteristikum der INF- oder DH-Projekte ist die Integration der Datenmanager in den Projektkontext und eine enge Anbindung an den Forschungsprozess. Die in diesen Projekten tätigen Personen sind „eingebettet“: Erst durch den regelmäßigen, direkten und vertrauensvollen Kontakt ist es möglich, eine kontinuierliche Begleitung über den Lebenszyklus von Forschungsdaten anzubieten, die Grundlage für nachhaltiges Forschungsdatenmanagement.²⁰ Das Verhältnis der unterstützenden Personen zu den Forschenden in diesen Verbänden skaliert jedoch nicht mit den zur Verfügung stehenden Ressourcen für die Breite der Forschung an Hochschulen oder außeruniversitären Einrichtungen. Die institutionelle Zuordnung der Datenmanager ändert zudem deren Status – vom Teammitglied zum Dienstleistenden. Die für das Management der Daten, eine gemeinsame Arbeit auf Augenhöhe, notwendige Vertrauensbasis und Anerkennung ist damit erheblich reduziert.²¹

In Verbänden wirken sich bestimmte strukturelle Rahmenbedingungen positiv auf die Bereitschaft aus, Forschungsdaten zu teilen und aufzubereiten: andere Forschende sind auf die eigenen Daten angewiesen; projekteigene Regularien zwingen zur Freigabe; die interdisziplinäre Zusammenarbeit verspricht innovative Forschungsergebnisse.²² Wenn Datenproduktion und verschiedene Nutzungsformen innerhalb eines Projektkontexts angesiedelt sind, können die Anforderungen an Qualität und Form der Forschungsdaten gemeinsam evaluiert und umgesetzt werden. Ist diese unmittelbare Kooperationsmöglichkeit nicht gegeben, werden die Anforderungen der Nachnutzung nicht berücksichtigt, auch weil in vielen Bereichen keine Richtlinien und Standards gelten oder eine interdisziplinäre Nachnutzung nicht intendiert ist. Das Bemühen der Digital Humanities um Standards in der Erfassung und Digitalisierung von Kulturgütern zeugt von dieser Diskrepanz.²³

Die Grenzen der Übertragbarkeit des eingebetteten und projektorientierten Datenmanagements zeigen sich auch an den Elementen, die erfolgreich skalieren: Infrastrukturkomponenten wie Repositorien oder Softwarewerkzeuge können auch andere Zielgruppen bedienen und skalieren problemlos mit hoher Nutzungszahl. Schulungs- und Trainingsmaterialien im Bereich

²⁰ Vgl. Cremer et al. (2015) 23f.

²¹ Zur Bedeutung von Vertrauen beim Teilen von Daten Peer-to-Peer vgl. Wallis et al. (2013) 12f.

²² Vgl. Van den Eynden (2016) 24f.

²³ Vgl. Klaffki et al. (2018) 18–23.

Forschungsdatenmanagement, die auch für den Eigenbedarf in Forschungsprojekten entwickelt wurden, sind in großer Zahl und Vielfalt verfügbar.²⁴ Das eigentliche Erfolgsrezept der Verbundprojekte, die personelle Unterstützung, persönliche Beratung und kontinuierliche Begleitung, wird jedoch nur in Ausnahmefällen in einen institutionellen Service überführt.²⁵

4 Die Forschungsdatenredaktion

4.1 Portfolio einer Forschungsdatenredaktion

Unter den Voraussetzungen, dass 1) Forschungsdaten aufbereitet werden müssen, 2) die Aufbereitung nicht allein von den Forschenden übernommen kann, 3) ein den Forschungsprozess begleitendes Modell auf institutioneller Ebene nicht mehr skaliert, stellt sich die Frage nach einem tragfähigen Modell für die Unterstützung von Forschungsdatenpublikationen in Forschungseinrichtungen. Im Folgenden wird das Modell einer „Forschungsdatenredaktion“ diskutiert. Eine Forschungsdatenredaktion unterstützt Forschende bei der Aufbereitung der Daten zur Publikationsreife und übernimmt zentrale Aufgaben im Publikationsprozess. Das Modell orientiert sich hierbei an den klassischen Publikationen und den redaktionellen Tätigkeiten im Publikationswesen. Eine Redaktion übernimmt die zu publizierende Daten und führt in Zusammenarbeit mit den Forschenden die notwendigen Überarbeitungen durch. Mögliche Aufgaben einer Forschungsdatenredaktion sind:

- Anamnese der Daten (z.B. durch ein Dateninterview)²⁶
- Appraisal des Datensatzes (z.B. Auswahl der Dateien und Informationspakete)²⁷
- Explikation (z.B. Variablenbezeichnung, Abkürzungen)²⁸
- Konversion in standardisierte, nicht-proprietäre Formate (z.B. xls in csv)²⁹
- Wahl des Dateiformats und Durchführung der Formatvalidierung (z.B. TIFF)³⁰
- Transformation und Strukturierung (z.B. Texte in TEI-P5)³¹
- Automatisierte Anreicherung (z.B. Named Entity Recognition)³²
- Validierung gegen Standards (z.B. FAIR Data Prinzipien)³³
- Wahl des Datenrepositoriums (z.B. via re3data.org)³⁴

²⁴ Bange et al. (2018).

²⁵ Siehe dazu auch die Beiträge von Helbig et al. (2018) sowie Gnad und Steilen (2018).

²⁶ Siehe Witt und Carlson (2007).

²⁷ Siehe Whyte und Wilson (2010).

²⁸ Siehe Borer et al. (2009).

²⁹ Siehe "Comparison of File Formats" in Hoe (2006).

³⁰ Siehe Funk und Neubauer (2010).

³¹ Die Text Encoding Initiative (TEI) betreut und entwickelt den XML-basierten Standard für Textauszeichnung, inzwischen in Version 5: <http://www.tei-c.org>.

³² Eine Übersicht bietet Schuhmacher (2018).

³³ Siehe Wilkinson et al. (2016).

³⁴ re3data.org ist ein weltweites Verzeichnis für Forschungsdatenrepositorien: <https://www.re3data.org>.

- Ingest (z.B. Account, Paketierung, Uploadprozess)³⁵
- Lizenzvergabe (z.B. kompatibel mit der Offen-Definition³⁶ und den bereits erwähnten FAIR-Prinzipien)
- Metadatenvergabe (z.B. DataCite Metadata Schema)³⁷
- Dokumentation (z.B. natursprachliche Sammlungsbeschreibung)³⁸

Die Forschungsdatenredaktion grenzt sich damit dezidiert von den Konzepten von “Data Curator”, “Data Stewardship” und “Data Preservation” ab. Das Aufgabenfeld bezieht sich auf ein temporäres, aktives Eingreifen im Lebenszyklus von Forschungsdaten mit dem primären Ziel der Datenpublikation. Damit werden jedoch die oben als erfolgreich skizzierte kontinuierliche Begleitung während des Forschungsprozesses, die für eine Nachhaltigkeit notwendige langfristige Betreuung nach der Publikation sowie die für eine Kulturveränderung notwendige Informations-, Schulungs- und Beratungsstellen nicht adressiert.

4.2 Beispiel: Die Redaktion *perspectivia.net*

In der Max Weber Stiftung, die eine Vielzahl geisteswissenschaftlicher Forschungsinstitute im Ausland betreibt, wurde die „Redaktion *perspectivia.net*“ ursprünglich (2008) mit klassischen Aufgaben des Publizierens, wie Lektorat, und neuen Aufgaben des elektronischen Publizierens wie Lizenzen und HTML betraut.³⁹ Seit einigen Jahren wird diese konzeptionell und personell zu einer “Digitalen Redaktion” weiterentwickelt, zu deren Aufgaben auch Aspekte des Forschungsdatenmanagements gehören. Das Ziel ist, eine unterstützende Instanz für den wissenschaftlichen Publikationsprozess zu bilden, die in Form einer zentralen Redaktion mit einem ganzheitlichen Konzept den Herausforderungen der digitalen Transformation der Publikationspraxis und der Wissenschaft gerecht wird. Das anvisierte Aufgabenfeld reicht von Beratung und Vermittlung, Projektmanagement und Konzeptionierung, Auszeichnung und Anreicherung, Standardisierung und Normierung, Datenmanagement und Lizenzierung, Transformation und Archivierung, Katalogisierung und Vernetzung, Layout und Lektorat, Verteilung und Kommunikation bis zu Schulung und Betreuung der Forschenden in den Instituten der MWS.

Das Team setzt sich aus einer Leitung, zwei festen und mehreren projektbasierten Stellen sowie wissenschaftlichen Hilfskräften zusammen, die die Umsetzung der Publikationsvorhaben begleiten und umsetzen. Nicht abgedeckte Kompetenzen und Ressourcen werden über Kooperationen realisiert, bei denen Partnerschaften mit Infrastruktureinrichtungen sowie mit Forschungs- und

³⁵ Siehe Ludwig und Enke (2013) 28–33.

³⁶ Die “Offen-Definition” (Open Definition) präzisiert das Konzept Offenheit in Bezug auf Daten und klassifiziert bestehende Lizenzen hinsichtlich ihrer Offenheit: <http://opendefinition.org/od/2.1/en/>.

³⁷ Die Organisation Datacite betreut den Metadatenstandard für Forschungsdaten, die eine DOI von Datacite erhalten: <https://www.datacite.org/>.

³⁸ Bspw. in der DARIAH-DE Collection Registry: <https://colreg.de/dariah.eu>.

³⁹ Siehe <https://perspectivia.net>. Einen Einblick in die Anfangsphase von *perspectivia.net* liefert Landes (2009).

Infrastrukturprojekten eine Schlüsselrolle spielen. Die Kompetenzen im Team, in den Forschungsinstituten der Stiftung und dem Partnernetzwerk ermöglichen ein Serviceangebot zur Publikation von Forschungsdaten, das eine Aufbereitung der Daten durch die Redaktion inkludiert. Die Grundlage für das Serviceportfolio bildet zunächst der klassische Publikationsbereich: Kenntnisse und Komponenten für Identifikation, Referenzierung, Registrierung, Katalogisierung, Indexierung und Lizenzierung lassen sich für Forschungsdaten adaptieren. Das zweite Standbein des Portfolios kommt aus dem Bereich der Digital Humanities, für den in der Max Weber Stiftung eine Koordinationsstelle die Editions- und Datenbankprojekte betreut. Die hier verfügbare Expertise in Datenmodellierung, Softwarewerkzeugen und Standards erlaubt auch ein allgemeines Unterstützungsangebot für die Aufbereitung von Forschungsdaten. Auch die inzwischen auf XML-Transformation basierenden Publikationsworkflows der klassischen Publikationsformate bieten für textbasierte Forschungsdaten viele Anknüpfungspunkte.

Gleichwohl gelten für dieses Serviceangebot ebenso die eingangs geschilderten Voraussetzungen: Die strukturellen und kulturellen Hürden sorgen dafür, dass Forschungsdatenpublikationen bei *perspectiva.net* vorerst die Ausnahme darstellen. Dennoch erscheint das Angebot als notwendiger Schritt, um mit den vorhandenen Infrastrukturen überhaupt Forschungsdatenpublikationen realisieren zu können.

4.3 Verortung und Einbindung

Aus der oben skizzierten und exemplifizierten Forschungsdatenredaktion lassen sich institutionelle und strukturelle Rahmenbedingungen ableiten, die dem Aufbau eines solchen Angebots zugrunde liegen:

1. Abgesehen von dem grundsätzlichen Beratungsangebot, das als optionale Leistung nicht immer genutzt wird, setzt die Unterstützung am Ende der Forschung, nach den Verarbeitungsprozessen der Daten, an. Während dies für ein ganzheitliches, an dem gesamten Lebenszyklus orientierten Datenmanagement misslich ist, ergeben sich im Workflow auch Vorteile. Im Vorfeld oder nach einer mit den Daten zusammenhängenden Publikation bietet sich Gelegenheit, die Forschungsdatenpublikation "mitzumachen" und beides in Relation zu setzen und so jeweils zu stärken.
2. Das Aufgabenfeld der redaktionellen Aufbereitung der Forschungsdaten zeigt Überschneidungen und Synergiepotential mit den Bereichen des digitalen Publizierens und der digitalen Bibliothek sowie den Digital Humanities. Einheiten, Abteilungen oder Einrichtungen, die in diesen Feldern mit einem Serviceangebot aktiv sind, können daher bereits einen Teil der notwendigen Kompetenzen abdecken, so dass fehlende Bereiche gezielt ergänzt werden könnten.
3. Solange Forschungsdatenpublikationen eine Ausnahme darstellen und die Arbeitsprozesse nicht standardisiert sind, entsteht ein unregelmäßiger, zeitlich begrenzter, aber hoher Aufwand der Aufbereitung. Die Tätigkeiten können daher nur als Nebenaufgabe einer

mehrköpfigen Einheit übernommen werden, um Legitimation und Skalierbarkeit zu gewährleisten.

Unter diesen Voraussetzungen ist die Verortung einer Forschungsdatenredaktion in mehreren institutionellen Kontexten denkbar: Universitätsverlage, Datenzentren, Publikationsabteilungen oder andere zentrale Einrichtungen mit Publikationsaufgaben bieten sich an. Das Kompetenzprofil würden auch Digital Humanities-Abteilungen oder e-Research-Zentren erfüllen. Nicht zuletzt könnten sich hier auch die Universitätsbibliotheken einbringen, wenn es gelingt, die Kompetenzen entsprechend zu bündeln und zu ergänzen.

5 Diskussion

Der Artikel macht den mit der Generierung und der standardkonformen Publikation von Forschungsdaten verbundenen Aufwand deutlich. Solange Forschende nur geringe Vorteile für ihre wissenschaftliche Reputation aus der Publikation ihrer eigenen Forschungsdaten beziehen, verbleiben der Aufbau von Infrastrukturen und Informationsangeboten außerhalb des eigentlichen wissenschaftlichen Ökosystems. Das Reputationsproblem kann nur wissenschaftskulturell gelöst werden, aber Kompetenzdefizite und Rahmenbedingungen müssen infrastrukturell verbessert werden – Infrastrukturen und Serviceangebote können dabei als Katalysatoren wirken. Eine Forschungsdatenredaktion, die den Forschenden Aufgaben abnimmt, vermag als zusätzliche Ressource fehlende Zeit und Kompetenzen zu kompensieren. Für die wissenschaftliche Notwendigkeit der Publikation müsste jedoch die Datenveröffentlichung Renommee versprechen (oder das Fehlen ebensolcher negativ gewertet werden) und die Nachnutzung von Forschungsdaten Teil der geisteswissenschaftlichen Forschungspraxis werden.

Die klassische Redaktionsarbeit im Publikationswesen hat den Charakter einer Dienstleistung und unterscheidet sich somit signifikant von den stärker teamorientierten und partnerschaftlichen Konzepten in Verbundprojekten. Der Vorgang der Datenaufbereitung greift direkt in das Forschungsergebnis ein und bedarf umfangreicher Kontextinformationen. Das Forschungsdatenmanagement erfordert damit jedoch eine enge Zusammenarbeit und erzeugt einen hohen Kommunikationsbedarf. In einem Rahmen, in dem die Forschungsdatenredaktion als Service konzipiert ist, birgt dies große Schwierigkeiten, auf Augenhöhe zusammenzuarbeiten. Denn in dem vorgestellten Modell sind weder die Entscheidungsgewalt etwa über die formale Gestalt der Daten, noch der Umgang mit Konflikten geklärt. Auch wäre zu fragen, ob die erst am Ende des Forschungsprozesses eingreifende Forschungsdatenredaktion nicht zu spät kommt? Denn bereits bei der Erzeugung und Analyse sind viele Aspekte des Forschungsdatenmanagements zu berücksichtigen, wenn publikationsreife Forschungsdaten entstehen sollen. Gleichzeitig existieren in

den geisteswissenschaftlichen Disziplinen erst wenige Praktiken und keine ausreichenden Standards, die die Erzeugung nachnutzbarer Forschungsdaten begünstigen.

Eine Forschungsdatenredaktion verursacht in der Ökonomie des Wissenschaftssystems keine Veränderung zugunsten einer offeneren Wissenschaft. Sie dient lediglich als ein Instrument, um in bestimmten Fällen die Publikation von Forschungsdaten zu begünstigen, die sonst verschlossen geblieben wären. Die Forschungsdatenredaktion bietet ein Angebot für Forschende, denen Zugang und Ressourcen zum Forschungsdatenmanagement fehlen.

Literaturverzeichnis

- Andorfer, Peter (2015): Forschungsdaten in den (digitalen) Geisteswissenschaften. Versuch einer Konkretisierung. Göttingen (DARIAH-DE Working Papers, 14). URN:[urn:nbn:de:gbv:7-dariah-2015-7-2](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2015-7-2).
- Bange, Dörte; Biernacka, Katarzyna; Buchholz, Petra; Dolzycka, Dominika; Helbig, Kerstin; Langhanke, Gerald; Rex, Jessica; Vlaeminck, Sven (2018): Materialkatalog zum Forschungsdatenmanagement. Datensatz. DOI:[10.5281/zenodo.1209284](https://doi.org/10.5281/zenodo.1209284).
- Borer, Elizabeth T.; Seabloom, Eric W.; Jones, Matthew B.; Schildhauer, Mark (2009): Some Simple Guidelines for Effective Data Management. In: *Bulletin of the Ecological Society of America*, 90 (2), 205–14. DOI:[10.1890/0012-9623-90.2.205](https://doi.org/10.1890/0012-9623-90.2.205).
- Borgman, Christine L (2010): Research Data: Who will share what, with whom, when, and why? Berlin (RatSWD Working Paper Series, 161).
- Cremer, Fabian (2018): INF-Projekte in Sonderforschungsbereichen (SFB/DFG) der Geisteswissenschaften 2018. Datensatz. DARIAH-DE Repository. DOI:[10.20375/0000-000b-d462-7](https://doi.org/10.20375/0000-000b-d462-7).
- Cremer, Fabian; Engelhardt, Claudia; Neuroth, Heike (2015): Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale. In: *BIBLIOTHEK – Forschung und Praxis*, 39 (1), 13–31. DOI:[10.1515/bfp-2015-0006](https://doi.org/10.1515/bfp-2015-0006).
- Cremer, Fabian; Klaffki, Lisa; Steyer, Timo (2018): Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: *o-bib. Das offene Bibliotheksjournal*, 5 (2), 142–62. DOI:[10.5282/o-bib/2018H2S142-162](https://doi.org/10.5282/o-bib/2018H2S142-162).
- Deutsche Forschungsgemeinschaft (2018): Merkblatt Sonderforschungsbereiche. In: http://www.dfg.de/formulare/50_06/50_06_de.pdf.
- Drucker, Johanna (2011): Humanities Approaches to Graphical Display. In: *Digital Humanities Quarterly*, 005 (1). <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Engelhardt, Claudia (2013): Forschungsdatenmanagement in DFG-Sonderforschungsbereichen: Teilprojekte Informationsinfrastruktur (INF-Projekte). In: *LIBREAS. Library Ideas*, (23). URN:[urn:nbn:de:kobv:11-100212741](https://nbn-resolving.org/urn:nbn:de:kobv:11-100212741).
- Funk, Stefan E.; Neubauer, Matthias (2010): Formatcharakterisierung (Version 2.3). In: Nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen: Niedersächsische Staats- und Universitätsbibliothek. URN:[urn:nbn:de:0008-20100617150](https://nbn-resolving.org/urn:nbn:de:0008-20100617150).
- Gnadt, Timo; Steilen, Lena (2018): Beratungsangebote für Forschende. In: *Bausteine Forschungsdatenmanagement*, (1), 30–38. DOI:[10.17192/bfdm.2018.1.7816](https://doi.org/10.17192/bfdm.2018.1.7816).
- Helbig, Kerstin; Dierkes, Jens; Neumann, Janna (2018): Aufbau und Bekanntmachung von Informationsangeboten über Forschungsdatenmanagement für Forschende. In: *Bausteine Forschungsdatenmanagement*, (1), 1–6. DOI:[10.17192/bfdm.2018.1.7821](https://doi.org/10.17192/bfdm.2018.1.7821).

- Hoe, Nah Soo (2006): Free/Open Source Software. Open Standards. Verfügbar unter https://en.wikibooks.org/wiki/FOSS_Open_Standards.
- Kaden, Ben (2017): Warum die Publikation von Forschungsdaten nach wie vor ein begrenztes Phänomen bleibt. LIBREAS.Library Ideas Blog. Verfügbar unter <https://libreas.wordpress.com/2017/04/05/forschungsdaten>.
- Kaden, Ben (2018): Warum Forschungsdaten nicht publiziert werden. In: *LIBREAS. Library Ideas*, (33). URN:[urn:nbn:de:kobv:11-110-18452/20046-8](https://nbn-resolving.org/urn:nbn:de:kobv:11-110-18452/20046-8).
- Klaffki, Lisa; Schmunk, Stefan; Stäcker, Thomas (2018): Stand der Kulturgutdigitalisierung in Deutschland: Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“. Göttingen (DARIAH-DE Working Papers 26). URN:[urn:nbn:de:gbv:7-dariah-2018-1-3](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2018-1-3).
- Klump, Jens (2017): Data as Social Capital and the Gift Culture in Research. In: *Data Science Journal*, 16 (0). DOI:[10.5334/dsj-2017-014](https://doi.org/10.5334/dsj-2017-014).
- Landes, Lilian (2009): Open Access und Geschichtswissenschaften. In: *LIBREAS. Library Ideas*, (14). DOI:[10.18452/8918](https://doi.org/10.18452/8918).
- Ludwig, Jens; Enke, Harry, Hrsg. (2013): Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt. Glückstadt: Hülsbusch.
- Oltersdorf, Jenny; Schmunk, Stefan (2016): Von Forschungsdaten und wissenschaftlichen Sammlungen. In: *BIBLIOTHEK – Forschung und Praxis*, 40 (2), 179–85. DOI:[10.1515/bfp-2016-0036](https://doi.org/10.1515/bfp-2016-0036).
- Palmer, Carole L. (2007): Thematic Research Collections. In: *A Companion to Digital Humanities*, hg. v. Susan Schreibman, Ray Siemens und John Unsworth, 348–65. DOI:[10.1002/9780470999875.ch24](https://doi.org/10.1002/9780470999875.ch24).
- RfII – Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. RfII Empfehlungen. Verfügbar unter <http://www.rfii.de/?wpdmdl=1998>.
- Sahle, Patrick (2013): DH studieren! Auf dem Weg zu einem Kern- und Referenzcurriculum der Digital Humanities. Göttingen (DARIAH-DE Working Papers 1). Verfügbar unter <http://resolver.sub.uni-goettingen.de/purl/?dariah-2013-1>.
- Schöch, Christof (2013): Big? Smart? Clean? Messy? Data in the Humanities. In: *Journal of Digital Humanities*. Verfügbar unter <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities>.
- Schumacher, Mareike (2018): Named Entity Recognition (NER). Blog. forTEXT. Literatur digital erforschen. Verfügbar unter <https://fortext.net/routinen/methoden/named-entity-recognition-ner>.
- Stäcker, Thomas; Steenweg, Helge (2018): Forschungsdaten – Aufgabe und Herausforderung für Bibliotheken. In: *o-bib. Das offene Bibliotheksjournal*, 5 (2), IV–V. DOI:[10.5282/o-bib/2018H2SIV-V](https://doi.org/10.5282/o-bib/2018H2SIV-V).
- Tristram, Frank; Bamberger, Peter; Çayoğlu, Uğur; Hertzner, Jörg; Knopp, Johannes; Kratzke, Jonas; Rex, Jessica; u.a. (2015): Öffentlicher Abschlussbericht von bwFDM-Communities – Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs. DOI:[10.5445/IR/1000083272](https://doi.org/10.5445/IR/1000083272).
- Van den Eynden, Veerle; Bishop, Libby (2014): Incentives and motivations for sharing research data, a researcher’s perspective. Knowledge Exchange Report. Verfügbar unter https://data-archive.ac.uk/media/492924/ke_report-incentives-for-sharing-research-data.pdf.
- Wallis, Jillian C.; Rolando, Elizabeth; Borgman, Christine L. (2013): If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. In: *PLoS ONE*, 8 (7), e67332. DOI:[10.1371/journal.pone.0067332](https://doi.org/10.1371/journal.pone.0067332).

Whyte, Angus; Wilson, Andrew (2010): How to Appraise & Select Research Data for Curation. DCC How-to Guides. Verfügbar unter <http://www.dcc.ac.uk/resources/how-guides>.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; u.a. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data*, 3, 160018. DOI:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

Witt, Michael; Carlson, Jake R (2007): Conducting a data interview. Poster. 3rd International Digital Curation Conference, 2007 Washington D.C., Verfügbar unter http://docs.lib.purdue.edu/lib_research/81.



Fabian Cremer
Max Weber Stiftung
Rheinallee 6
D-53173 Bonn

cremer@MaxWeberStiftung.de



Lisa Klaffki
DARIAH-DE
Herzog August Bibliothek
Postfach 1364
D-38299 Wolfenbüttel
klaffki@hab.de



Timo Steyer
Herzog August Bibliothek
Postfach 1364
D-38299 Wolfenbüttel
steyer@hab.de