

Maxi Kindling, Berlin

# Qualitätssicherung im Umgang mit digitalen Forschungsdaten

Der Beitrag behandelt das Thema Qualitätssicherung im Umgang mit digitalen Forschungsdaten. Es werden drei Ebenen herausgearbeitet, auf die sich die Qualitätssicherung beziehen kann: die der Forschungsdaten selbst, der Metadaten zur ihrer Erschließung sowie der Forschungsdaten-Infrastrukturen. Es erfolgt eine Annäherung an den Qualitätsbegriff im Kontext digitaler Forschungsdaten sowie an die qualitätssichernden Verfahren. Exemplarische nationale und internationale Initiativen werden vorgestellt, die Empfehlungen und Kriterien zu disziplinübergreifenden Aspekten der Qualitätssicherung formulieren. Zudem wird ein disziplinäres Beispiel aus der historischen Korpuslinguistik vorgestellt.

**Deskriptoren:** Qualität, Forschung, Datensicherung, Metadaten, Management, Infrastruktur, Langzeitarchivierung, Definition

## Quality assurance of digital research data

The article deals with quality assurance of digital research data. Three categories of quality assurance are identified: digital research data itself, its metadata and research data infrastructures. An approach to define quality of research data and quality assurance is provided. The recommendations of national and international initiatives on interdisciplinary aspects of quality assurance are presented. Moreover a disciplinary example from historical corpus linguistics is introduced.

**Keywords:** Quality assurance, research data, metadata, information management, definition

## La garantie de la qualité des données numériques de recherche

L'auteur aborde la question de la garantie de la qualité dans le domaine des données numériques de recherche. La garantie de la qualité peut agir à trois niveaux différents: les données de la recherche elles-mêmes, leurs métadonnées ainsi que leurs infrastructures. L'auteur propose une définition de la notion de qualité dans le contexte des données de recherche numériques ainsi que

des procédures d'assurance de qualité. Ensuite, il énumère des exemples d'initiatives nationales et internationales pour formuler des recommandations et des critères pour les aspects interdisciplinaires de la garantie qualité. Finalement, il présente un exemple de la linguistique de corpus historique.

**Mots-clés:** qualité, recherche, sauvegarde des données, métadonnées, gestion, infrastructure, préservation à long terme, définition

## 1 Einleitung

Digitale Forschungsdaten werden zunehmend planvoll archiviert und zugänglich gemacht. Diese Entwicklung wird durch Infrastrukturen wie Forschungsdaten-Repositoryn unterstützt, die zuvor so nicht bestanden. Dies steht in enger Wechselwirkung damit, dass Forschungsförderorganisationen einerseits die Nachnutzung verfügbarer digitaler Forschungsdaten aus ökonomischen Gründen fordern, zum anderen Forschungsergebnisse reproduziert und damit nachvollziehbar gemacht werden. Um dies nicht nur in der Theorie, sondern auch in der Praxis möglich zu machen, müssen die Forschungsdaten adäquat aufbereitet sein und zuverlässig archiviert werden. Um dies auf einem bestimmten Niveau zu erreichen und damit Qualität zu sichern, werden Richtlinien und Kriterien gebraucht. Es liegt nahe, dass es so viele Kriterien gibt, wie beteiligte Akteure in diesem Umfeld und Ebenen, auf die sich Qualität beziehen kann. Darüber hinaus erfolgt die Bestimmung von Qualität nach objektiven und auch subjektiven Kriterien. Unter diesen Voraussetzungen ist die Qualitätssicherung im Umgang mit digitalen Forschungsdaten ein komplexes Themenfeld. Es scheint nahezu unmöglich, die Anforderungen aller Beteiligten zu ihrer Zufriedenheit zu berücksichtigen. Dennoch ist es wichtig, die Diskussion zu Qualitätskriterien, Maßnahmen zu ihrer Sicherung sowie Verfahren zu ihrer Überprüfung zu führen, um sich einigen Richtlinien und Standards anzunähern, die die Nachvollziehbarkeit von Forschung und die Nachnutzung digitaler Forschungsdaten möglich machen.

Digitale Forschungsdaten und die Forschungsdaten-Infrastrukturen sind Forschungsgebiete am Lehr- und Forschungsbereich Informationsmanagement (vgl. Kindling & Schirnbacher, 2013, in diesem Heft S. 127–136). Die Bibliotheks- und Informationswissenschaft als reflektierende Wissenschaft über Prozesse der Wissenschaftskommunikation und der sie unterstützenden Informationsinfrastrukturen kann durch die Analyse und Bewertung der vorhandenen Initiativen zur Qualitätssicherung sowie der integrativen Anforderungsanalyse an Qualität und ihre Sicherung einen wichtigen Beitrag leisten, die Qualitätssicherung im Umgang mit digitalen Forschungsdaten zu unterstützen. Dieser Beitrag versteht sich als Problematisierung und bietet eine disziplinübergreifende Annäherung an das Thema. Die Vielschichtigkeit der notwendigen Auseinandersetzung zur Qualitätssicherung im Umgang mit digitalen Forschungsdaten wird aufgezeigt.

## 2 Digitale Forschungsdaten zur Qualitätssicherung und Qualitätssicherung digitaler Forschungsdaten

Transparente Forschungsprozesse und damit die Gewährleistung der Nachvollziehbarkeit von Forschung sind grundlegend für die Qualitätssicherung der Wissenschaftskommunikation. Zu dieser Transparenz zählt die Publikation der der Forschung zugrunde liegenden Forschungsdaten<sup>1</sup>: Erst wenn diese zugänglich bzw. veröffentlicht sind, wird es möglich, die tatsächliche (bspw. empirisch fundierte) Aussagekraft der auf dieser Grundlage ermittelten Erkenntnisse parallel zur methodischen Korrektheit und Logik in der Schlussfolgerung zu prüfen.

Dass die Qualität des Ausgangsmaterials einer Forschung wesentlich zur Gesamtqualität beiträgt und daher jede Qualitätssicherung in der Wissenschaftskommunikation auch eine entsprechende Prüfung zulassen muss, liegt eigentlich nahe. Allerdings gibt es unterschiedliche Hindernisse, weshalb die planvolle Archivierung und Bereitstellung von digitalen Forschungsdaten bisher in vielen Wissenschaftsdisziplinen eine bestenfalls nachgeordnete Rolle spielten. Teilweise war und ist dies rein praktisch bedingt: Mitunter fehlten technische Infrastrukturen zur Speicherung und Zugänglichkeit des Materials oder Forschungsdaten lagen nur analog und nicht

in digitaler Form vor, wie etwa in kultur- und geisteswissenschaftlichen Disziplinen. Die Entwicklung vernetzter Forschungsinfrastrukturen in den vergangenen zehn Jahren lässt darüber hinaus heute neue und veränderte Möglichkeiten der Generierung, Bearbeitung und Analyse digitaler Forschungsdaten zu. (vgl. Kindling, 2012, 148 ff.)

## 3 Relevanz und Akteure

Aufmerksamkeit seitens der Wissenschaftspolitik erlangte die Archivierung und Zugänglichmachung durch Fälle wissenschaftlichen Fehlverhaltens (vgl. DFG, 1998, 5). Daraufhin veröffentlichte in Deutschland beispielsweise die Deutsche Forschungsgemeinschaft 1998 Empfehlungen zur Sicherung guter wissenschaftlicher Praxis (vgl. DFG, 1998). Auch international wird der Zugang zu digitalen Forschungsdaten thematisiert und gefordert (vgl. bspw. The Royal Society, 2012; Office of Science and Technology Policy, 2013). Eng damit verbunden sind die Bemühungen um geeignete Forschungsdaten-Infrastrukturen – sie werden zum einen durch internationale Fördermaßnahmen der Forschungsförderorganisationen unterstützt, zum anderen wird die Benutzung bestehender Forschungsdaten-Infrastrukturen wie Repositorien empfohlen (vgl. DFG, 2012, 6).

Forschungsdaten-Repositorien sind Infrastrukturen, die die dauerhafte Archivierung und den Zugang zu digitalen Forschungsdaten über das World Wide Web (WWW) ermöglichen. Im Projekt [re3data.org](http://www.re3data.org)<sup>2</sup> und begleitenden Forschungsprojekten wurde festgestellt, dass die Landschaft der Forschungsdaten-Repositorien bislang als heterogen beschrieben werden muss (vgl. Pampel, 2013). Vor diesem Hintergrund ist die Identifizierung und Umsetzung von Anforderungen an Forschungsdaten-Repositorien, die letztlich als Qualitätskriterien festgehalten werden können, eine Herausforderung (vgl. Kindling & Schirnbacher, 2013, in diesem Heft, S. 134).

Die Verantwortung dafür, den Forschenden Infrastrukturen für die Archivierung und Bereitstellung digitaler Forschungsdaten und entsprechende Services anzubieten, liegt bei unterschiedlichen Akteuren. Eindeutige Zuständigkeiten durch bestimmte Akteure lassen sich kaum identifizieren. Sind es die IT-Verantwortlichen an ihren Forschungsinstituten? Sind es zentrale IT-Serviceeinrichtungen einer Hochschule oder Forschungseinrichtung? Sind es die Bibliotheken, die Archive, die Museen? Sind es die Wissenschaftsverlage, die Daten speichern und aufbereiten oder wird die Verantwortung zur For-

<sup>1</sup> Zur Begriffsbestimmung digitaler Forschungsdaten vgl. Kindling & Schirnbacher, 2013, in diesem Heft S. 130.

<sup>2</sup> <http://www.re3data.org>

schungsdatensicherung und -publikation an die Autorinnen und Autoren einer Publikation bzw. an ihre Heimat-einrichtung zurückverwiesen?

Neben der unklaren Zuständigkeit besteht eine weitere Herausforderung: Damit die Qualitätssicherung in der Wissenschaftskommunikation gelingt, gilt es für die benannten Akteure, also die Forschenden und Infrastruktureinrichtungen, gemeinsame und allgemein anerkannte Kommunikations- und Verarbeitungsrichtlinien zu entwickeln. Ein gutes Beispiel sind Richtlinien für das Zitieren von Forschungsdaten oder die Erhebung von Nutzungsdaten für Forschungsdaten und die Etablierung standardisierter Verfahren zu deren Auswertung. Dies gilt umso mehr, wenn Forschungsdaten über ihre reine Bereitstellung hinaus nachgenutzt werden sollen.

Ein naheliegendes Nachnutzungsszenario ist die Anwendung neuer Forschungsfragen auf bereits bestehenden Forschungsdaten (*Research Data Re-use*). Diese Form der Nachnutzung von vorhandenen Forschungsdaten und damit die Vermeidung von Doppelerhebungen ist nicht zuletzt ökonomisch sinnvoll. Zudem gibt es in einigen Disziplinen einzigartige Erhebungen (z. B. Klimadaten), die nicht ein weiteres Mal erhoben werden können oder es liegen in textorientierten Disziplinen umfängliche Textkorpora vor, die die Auswertung durch viele Forschende und nach unterschiedlichen disziplinären Gesichtspunkten zulassen. Zudem lassen sich Forschungsarbeiten, die auf der gleichen Datenbasis beruhen, im Idealfall besser zueinander in Beziehung setzen oder vergleichen. Abgesehen von den Vorbehalten gegen und Anreizen für die Bereitstellung der Forschungsdaten<sup>3</sup> (vgl. The ODE Project, 2011) hängt die Nachnutzung sehr stark davon ab, inwieweit ihre Erhebung, Auswertung, Aufbereitung und Archivierung bestimmten, eindeutigen Richtlinien oder Standards folgt, die eine Nachnutzung überhaupt erst möglich machen. Die Kompetenz, Forschungsdaten als Vorlage für Anschlussforschung auf ihre Relevanz, Solidität und Konsistenz hin einschätzen zu können, gehört mittlerweile zur wissenschaftlichen Grundausstattung. Die Funktion standardisierter Qualitätssicherungsverfahren ist in dieser Hinsicht eindeutig die Unterstützung der Forschenden.

Somit lässt sich festhalten, dass die Qualitätssicherung in der Wissenschaftskommunikation und im Umgang mit digitalen Forschungsdaten die Ziele

- (a) der Gewährleistung der Nachvollziehbarkeit von Forschung(sergebnissen) und
- (b) der Nachnutzbarkeit der Forschungsdaten verfolgt.

<sup>3</sup> Zu diesem Thema entstand am IBI die Dissertation von Sünje Dallmeier-Tiessen. Erscheint.

## 4 Forschungsförderung und Qualitätssicherung

Die qualitätsgesicherte Archivierung und der Zugang zu qualitätsgesicherten Forschungsdaten werden durch Wissenschaftspolitik und Forschungsförderorganisationen gefordert. Die *European Science Foundation* (ESF) und die *European Heads of Research Councils* (EUROHORCS) stellten 2009 den ökonomischen Nutzen des *Research Data Re-use*-Ansatzes und die Bedeutung von Qualitätssicherung und Interoperabilität von Forschungsdaten heraus:

„The collection of research data is a huge investment. Permanent access to such data, if quality controlled and in interoperable formats, will allow better use to be made of this investment because it allows other researchers to (re)use them. Furthermore it allows re-analysis and could play a role in ensuring research integrity.“ (ESF & EUROHORCS, 2009, 17)

Die Deutsche Forschungsgemeinschaft (DFG) griff den Aspekt der Qualitätssicherung in ihren 2009 veröffentlichten Empfehlungen zum Umgang mit digitalen Forschungsdaten auf und forderte:

„Es sind in den Fachcommunities Kriterien und Verfahren zu entwickeln, die die Qualitätssicherung der Daten gewährleisten [...].“ (DFG, 2009, 3)

Wenn Forschung nachvollzogen wird und Forschungsdaten nachgenutzt werden, dann möglichst mit einem Material, das den wissenschaftlichen bzw. innerhalb der Fachcommunities beschlossenen Kriterien entspricht. Die DFG betont die Heterogenität der Fachcommunities in der Wissenschaft, die mit ihrer Forschung auf unterschiedliche Datentypen und -formate referieren und sie mit unterschiedlichen Methoden auswerten.

Mit ihrer Position folgt die DFG einer Linie der *Organisation for Economic Co-Operation and Development* (OECD), die bereits 2007 für öffentlich geförderte Forschung formulierte:

„The value and utility of research data depends, to a large extent, on the quality of the data itself. Data managers, and data collection organisations, should pay particular attention to ensuring compliance with explicit quality standards. Where such standards do not yet exist, institutions and research associations should engage with their research community on their development. Although all areas of research can benefit from improved data quality, some require much more stringent standards than others. For this reason alone, universal data quality standards are not practical. Although all areas of research can benefit from improved data quality, some require much more stringent standards than others. For this reason alone, universal data quality standards are not practical. [...]“ (OECD, 2007, 19)

Allgemeingültige Kriterien für die Qualität von digitalen Forschungsdaten zu definieren, ist allein aufgrund der hohen fachlichen Spezifität und der daraus resultierenden Anforderungen an Methodik, Forschungsdesign und Interpretation der Daten kompliziert. Aber auch innerhalb von Fachcommunities kann man es nicht als gegeben voraussetzen, dass Einigkeit über alle Kriterien herrscht. Das Dilemma ist: Für die Forschungsdatenarchivierung und den Zugang im Kontext interdisziplinärer Projekte oder einer multidisziplinär ausgerichteten Einrichtung ist wenigstens eine Orientierung an allgemein anerkannten Richtlinien zur Qualitätssicherung notwendig. Letztere sind darüber hinaus von Bedeutung, wenn berücksichtigt wird, dass die einzelnen Einrichtungen die Nachvollziehbarkeit und Nachnutzung der Forschungsdaten mit effizient eingesetzten Ressourcen erreichen sollten.

## 5 Kriterien der Qualität(ssicherung)

Das britische Research Information Network (RIN) findet eine für die Qualitätssicherung im wissenschaftlichen Kontext anwendbare Definition:

“The relatively modern concept of quality assurance may be defined as the set of procedures designed to ensure a desired level of quality in a service or product.” (RIN, 2010, 7).

Damit sie gesichert werden kann, müssen zunächst Qualitätskriterien identifiziert werden, die zur Bestimmung von Qualität herangezogen werden können. Für den Qualitätsbegriff lässt sich zunächst abstrakt festhalten, dass Qualität eine in einer Community oder Domäne herrschende Übereinkunft über eine bestimmte Sammlung von Kriterien ist, die erfüllt sein müssen, damit Forschungsdaten als valide Ausgangsbasis für Forschung anerkannt werden können. Einige der Kriterien, die zur Bestimmung der Qualität digitaler Forschungsdaten dienen, sind operationalisierbar und bspw. durch formale oder technische Prüfung „messbar“, andere nicht.

Entscheidend ist der Aspekt der Übereinkunft: Die Fachcommunities müssen selbst, besser explizit als implizit, aushandeln, wann Kriterien erfüllt sind, die einen Qualität ausmachen. Über einige Kriterien herrscht darüber hinaus in der Wissenschaft Fachcommunity-übergreifende Einigkeit, wie folgende Ausführung zeigt:

“In the academic environment, there is widespread agreement that quality involves adherence to key principles such as intellectual rigour, accurate recording and honest reporting of results, and integrity in recognising the work of other researchers.” (RIN, 2010, 7)

Spätestens an dem Punkt, wo die Interoperabilität von Forschungsdaten in Forschungsdaten-Infrastrukturen erforderlich wird, ist eine Formalisierung notwendig, die nicht mehr disziplinär beliebig bestimmt werden kann. Der Spagat muss zwischen den Ansprüchen der Einzelcommunities und der Notwendigkeit einer standardisierten Verarbeitung gemeistert werden. Aus Sicht der Anbieter einer Forschungsdaten-Infrastruktur ist also eine spezifische, idealerweise formalisierbare Qualität relevant. Diese muss so gestaltet sein, dass sie gegebenenfalls besondere Qualitätsvorstellungen der Fachcommunities mit einschließt.

Die Zahl der Untersuchungen zur Bestimmung von übergreifend gültigen Qualitätskriterien und zur Qualitätssicherung digitaler Forschungsdaten ist bislang überschaubar (vgl. RIN, 2008; RIN, 2010; Pampel, 2012; APARSEN Report 2012; Waaijers & van der Graaf, 2011; Hense & Quadt, 2011, Peters-Kottig & Hasler, 2013). Dem Thema wird aber zunehmend mehr Aufmerksamkeit geschenkt. So wurde kürzlich die international agierende *Research Data Alliance*<sup>4</sup> gegründet, die sich in ihrer Auftaktveranstaltung dem Thema Data Peer Review widmete.

Bestehende Kriterien für die Qualität von *Information*, wie sie in der Informationswissenschaft und -praxis verwendet werden, lassen sich auf digitale Forschungsdaten anwenden. Dabei sollte die jeweilige Aggregationsstufe (auch: Prozessierungsstatus) digitaler Forschungsdaten in Anlehnung an den Lebenszyklus digitaler Forschungsdaten (vgl. Rümpel, 2011) der Referenzpunkt sein. So gibt es Kriterien, die bereits vor der Generierung oder Aggregation von Forschungsdaten relevant sind, so wie es auch Qualitätskriterien für ihre Aufbereitung und Anreicherung durch Metadaten, ihre Archivierung und damit die entsprechende technische Infrastruktur und den rechtlichen und organisatorischen Rahmen gibt.

Basierend auf Untersuchungen von Wang & Strong (1996) beschreiben Lee et al. (2002) beispielsweise die folgenden Qualitätskategorien für Information mit den entsprechenden Kriterien (in Klammern):

- „Intrinsic Information Quality“ (Exaktheit, Konsistenz, Korrektheit, Glaubwürdigkeit, Verlässlichkeit, Objektivität)
- „Contextual Information Quality“ (zeitlicher Kontext; Mehrwert, Relevanz, Vollständigkeit, Aktualität, Angemessenheit)
- „Representational Information Quality“ (Verständlichkeit, Interpretierbarkeit, Prägnanz der Darstellung und konsistente Darstellung)

<sup>4</sup> <http://rd-alliance.org/>

- „Accessibility Information Quality“ (Zugänglichkeit, Verfügbarkeit des Systems, die angebotenen Funktionen und Operationen, Bedienbarkeit und die System- und Datensicherheit.)

Die ersten beiden Kategorien *Intrinsic (Information) Quality* und *Contextual (Information)* basieren im Wesentlichen auf Kriterien, die durch die Fachcommunities bereits bei den Prozessen der Datengenerierung (z. B. durch automatisierte Verifizierung der datenerzeugenden Werkzeuge wie Messgeräte in den Naturwissenschaften; vgl. RIN, 2008, 48) und der Datenaufbereitung berücksichtigt werden sollten. Kriterien wie die Erzeugung anerkannter bzw. standardisierter Formate und die Weiterverarbeitung der Forschungsdaten in anerkannten bzw. standardisierten Formaten kommen ergänzend dazu. Die erfüllten Kriterien sollten in den beschreibenden Metadaten erwähnt werden.

*Representational (Information) Quality* bezieht sich darüber hinaus konkret auf die Anreicherung mit Metadaten. Hierbei ist die Semantik verwendeter Metadaten-schemata von Bedeutung, die ggf. durch entsprechendes (standardisiertes) Vokabular für die Beschreibung digitaler Forschungsdaten ergänzt wird. Die *Accessibility (Information) Quality* legt den Bezug zur Forschungsdaten-Infrastruktur nah. In einigen naturwissenschaftlichen Fächern ist es bereits etabliert, dass Datenzentren und Datenbanken Qualitätskriterien etwa hinsichtlich der erforderlichen Datenformate, ihrer Struktur sowie der Metadaten formulieren (vgl. RIN, 2008, 49; Waaijers & van der Graaf, 2011).<sup>5</sup>

Die Reichweite der Kriterien lässt sich erweitern. Waaijers & van der Graaf (2011) nennen beispielsweise folgende Kategorien, die für die Datenproduktion bzw. -gewinnung relevant sind:

- die Qualität der Ausstattung, die zur Datenproduktion verwendet wird („accuracy of the equipment and the refinement of the applied algorithms“)
- sowie die Qualität der Methodologie zur Datengewinnung („appropriate methodology, selection of data, integrity aspects, ethical requirements“)

Hier stellt sich schließlich die Frage, wie viel Komplexität in der Praxis der Qualitätssicherung noch handhabbar ist.

Wie gezeigt wurde ist es für den Gesamtprozess der Qualitätssicherung von Forschungsdaten notwendig, die

fachdisziplinär-definierten Ansprüche an die Forschungsdatenqualität mit den Ansprüchen an die Forschungsdaten-Infrastrukturen (FDI) zu integrieren. Daraus werden zwei untrennbar verbundene Ebenen von Qualität abgeleitet: die erste bezieht sich auf die Forschungsdaten selbst, die zweite auf die FDI und damit die unterstützenden Prozesse zu ihrer Archivierung und Bereitstellung.

Für diese Form der Harmonisierung und Standardisierung von Forschungsdaten und FDI in ihrer Wechselbeziehung gibt es bereits Ansätze. So sind beispielsweise die europäische Forschungsinfrastruktur für den Bereich der Sprachwissenschaften und -ressourcen *Common Languages and Technology Infrastructure* (CLARIN)<sup>6</sup> oder die *Digital Research Infrastructure for the Arts and Humanities* (DARIAH)<sup>7</sup> für den geisteswissenschaftlichen Bereich zu nennen. In den Naturwissenschaften wird bereits seit den 50er Jahren ein System von Weltdatenzentren entwickelt, das heute unter der Bezeichnung ICSU World Data System (WDS) geführt wird.<sup>8</sup> Für das Gebiet der Sozialwissenschaften lässt sich die Dachorganisation europäischer Datenzentren und -initiativen *European Council of European Social Science Data* (CESSDA) nennen.<sup>9</sup>

## 6 Verfahren zur Qualitätskontrolle

Qualitätssicherung bezieht sich nicht nur auf die Bestimmung von Qualität, sondern auch darauf, ein qualitatives Grundniveau zu stabilisieren und Überprüfungsroutinen zu entwickeln.

Die Sicherung und Überprüfung von Qualität bzw. ihrer Kriterien kann an diversen Stufen des Lebenszyklus von Forschungsdaten bzw. Prozessierungsstadien der Forschungsdaten ansetzen – von der Planung eines Forschungsprojektes über die Datenerhebung bis zur Publikation (vgl. RIN, 2008, 48).

Waaijers & van der Graaf (2011) identifizieren dahingehend drei Varianten der *Qualitätskontrolle*:

- “quality control in the production of data;
- data-management: ensuring permanence and accessibility (including metadata, documentation, formats, rights);
- actual content quality: “the scholarly merit of the dataset””

<sup>5</sup> Auch für disziplinübergreifende Datenarchive lassen sich Beispiele finden, siehe etwa „UK Data Archive“: <http://data-archive.ac.uk/curate/archive-quality>

<sup>6</sup> <http://www.clarin.eu/external/>

<sup>7</sup> <http://www.dariah.eu/>

<sup>8</sup> <http://www.icsu-wds.org/>

<sup>9</sup> <http://www.cessda.org/>

Sind die Stadien im Lebenszyklus der Forschungsdaten als Bezugspunkte der Qualitätssicherung und -überprüfung identifiziert, sind die eigentlichen Qualitätssicherungsverfahren darauf anzuwenden: In den meisten Arbeiten wird der Einsatz von Peer-Review-Verfahren in verschiedenen Stadien des Forschungsprozesses neben dem sog. *Author Approval* (vgl. Hense & Quadt, 2011) beschrieben.

Die Verlage haben an der Qualitätssicherung der über ihre Kanäle verbreiteten Ergebnisse ein erwartungsgemäß großes Interesse. Werden digitale Forschungsdaten publiziert, sind sie selbstverständlich Teil dieses Interesses, wie das prominente Beispiel des Journals *Nature* zeigt.<sup>10</sup> Traditionell hat sich Peer Review für die Qualitätssicherung wissenschaftlicher Artikel in Fachzeitschriften etabliert, mitunter werden die dem Artikel zugrundeliegenden Forschungsdaten begutachtet (vgl. Pampel et al., 2012; Waaijers & van der Graaf, 2011).<sup>11</sup> Allerdings gibt es angesichts des enormen inhaltlichen und zeitlichen Aufwands gegenüber Peer-Review-Verfahren als qualitätssichernder Maßnahmen für Forschungsdaten auch Vorbehalte (vgl. RIN, 2008; Pampel et al., 2012). Im Interesse eines angemessenen Aufwandes ist auch für Qualitätssicherungsverfahren erforderlich, Richtlinien und Kriterien beispielsweise für das Peer-Review-Verfahren zu entwickeln, die für Gutachterinnen und Gutachter als Orientierung für die Überprüfung dienen (vgl. Pampel et al., 2012).

Forschungsförderorganisationen fordern die Forschenden dazu auf, ihre Forschungsdaten in vertrauenswürdigen Archiven langfristig zugänglich zu machen (vgl. DFG, 2009). Häufig geben Verlage diese Verantwortung an ihre Autorinnen und Autoren weiter. Wissenschaftlerinnen und Wissenschaftler in interdisziplinären Forschungsprojekten oder kleinen Fachdisziplinen, die ihre Forschungsdaten nicht in disziplinären Forschungsdaten-Repositoryn oder Datenzentren ablegen, sind auf die Forschungsdaten-Infrastrukturen ihrer Heimateinrichtungen und ausreichende Unterstützung bei der Bereitstellung und Archivierung von digitalen Forschungsdaten angewiesen. Die Qualitätssicherung von Forschungsdaten-Infrastrukturen ist somit sowohl für disziplinäre Forschungsdaten-Repositoryn wie auch für institutionelle relevant. Sie stellen aber somit zugleich hohe Anforderungen an die Service- und Infrastruktureinrichtungen ihrer Hochschule oder Forschungseinrichtungen.

<sup>10</sup> <http://www.nature.com/authors/policies/availability.html>

<sup>11</sup> Darüber hinaus werden in einigen Zeitschriften originäre Arbeiten zu Forschungsdaten publiziert, siehe das Beispiel „Earth System Science Data“: <http://www.earth-system-science-data.net/home.html>

## 7 Zertifizierung von Forschungsdaten-Infrastrukturen

Das *Data Seal of Approval* (DSA)<sup>12</sup>, eine europäische Initiative von Datenarchivaren und Datenmanagern für die Qualitätssicherung und -zertifizierung von Forschungsdaten-Repositoryn, geht auf das dänische Datenarchiv *Data Archiving & Networked Services* (DANS) zurück. Es ist das erste Zertifikat mit dem expliziten Ziel der Qualitätssicherung, das einen internationalen Bekanntheitsgrad erreicht hat. Die Richtlinien umfassen 13 Kriterien in Form von Empfehlungen, die für eine Zertifizierung nachgewiesen werden müssen. Die Kriterien sprechen die Zielgruppen der Datenproduzenten, der Daten-Repositoryn und der Daten-Konsumenten an. Das DSA unterscheidet, wie dieser Beitrag, zwischen der Qualität der Forschungsdaten selbst sowie der Qualität der Dateninfrastruktur. Die Qualität der Forschungsdaten wird bestimmt durch

- „1. Their intrinsic scientific and scholarly quality,
2. The format in which the research data and supporting information are stored,
3. The documentation (metadata or contextual information) regarding the research data.“

Die dritte Empfehlung lautet beispielsweise: “The data producer provides the research data together with the metadata requested by the data repository”. Die Empfehlung verdeutlicht, dass zwar die Datenproduzenten, also die Wissenschaftlerinnen und Wissenschaftler selbst, angesprochen werden. Welche Information geliefert werden soll, wird indes seitens der Informationsinfrastruktur vorgegeben – was in Absprache mit den jeweiligen Fachdisziplinen geschehen muss. Die Notwendigkeit dafür wird in weiteren Untersuchungen konstatiert, wie etwa folgendes Zitat deutlich macht:

“Researchers have varying degrees of understanding about metadata, but often do not have a sense of what metadata should be applied to their data set to enable it to be discovered, understood, administered or used by others.” (Carlson, 2012, 17)

Neben der Qualität der Forschungsdaten bestimmt das DSA die Qualität von Forschungsdaten-Repositoryn durch zwei Faktoren:

- “The quality of the organizational framework in which the data repository is incorporated (organization and processes)”

<sup>12</sup> <http://datasealofapproval.org/>

- “The quality of the technical infrastructure of the data repository”.

Diese Faktoren betreffend lassen sich im Umfeld der digitalen Langzeitarchivierung bzw. des nachhaltigen Betriebs von FDR weitere relevante Projekte finden. Exemplarisch wird das DRAMBORA-Projekt (am britischen *Digital Curation Center*) angeführt.<sup>13</sup> DRAMBORA ist ein Akronym für „*Digital Repository Audit Method Based on Risk Assessment*“. In diesem Projekt wird betont, in welcher engen Beziehung die Qualität der Information eines FDR mit dem FDR selbst steht:

“Digital preservation is nowadays often defined as a risk management exercise where the aim is to convert the uncertainty about maintaining usability of authentic digital objects into quantifiable risks. The purpose of a digital repository is to do everything it can to mitigate the risks that impede its ability to provide access to authentic digital information. The measure of success of a repository’s work is the ‘quality’ of information it releases to its users.” (Mc Hugh et al., 2007, 4)

Zur „Information“ für die Nutzerinnen und Nutzer zählen zweifelsohne die die Forschungsdaten beschreibenden Metadaten. Neben dem DRAMBORA-Projekt sind die folgenden Standards und Zertifikate zu nennen, die den Anwendungsbereich der Sicherung und der Überprüfung vertrauenswürdiger Langzeitarchivierung gemeinsam haben:

- ISCU WDC – World Data System Zertifizierung<sup>14</sup>,
- ISO 16363 Audit and Certification of Trustworthy Digital Repositories<sup>15</sup>,
- Nestor – Vertrauenswürdige Archive // DIN 31644<sup>16</sup>
- Trusted Repository Audit and Certification (TRAC).

Die Zertifikate und Standards Data Seal of Approval, DIN 31644 und ISO 16363 werden in einem *European Framework for Audit and Certification of Digital Repositories* zusammenggeführt (vgl. Mauer & Schumann, 2013).

Darüber hinaus sind in verschiedenen nationalen und internationalen Projekten Hilfsmittel zur Einordnung und Bewertung von Forschungsdaten-Infrastrukturen entstanden bzw. werden vorbereitet. Darunter ist das deutsche WissGrid-Projekt, in dem eine Checkliste für das Forschungsdatenmanagement erstellt wurde (Ludwig & Enke, 2013, 83 ff.) und das britische Projekt *Data Asset*

*Framework*<sup>17</sup> als Unterstützung für sog. „Data Audits“. Im Projekt re3data.org werden perspektivisch Empfehlungen für den Aufbau und Betrieb von Forschungsdaten-Repositoryn in Anlehnung an das DINI-Zertifikat für Dokumenten- und Publikationsservices entwickelt (vgl. Kindling & Schirnbacher, 2013, in diesem Heft S. 133–134).

## 8 Die Bedeutung von Metadaten bei der Qualitätssicherung

Gestützt auf die obigen Ausführungen lassen sich drei Akteursgruppen identifizieren, die die Qualität digitaler Forschungsdaten und der Forschungsdaten-Infrastrukturen bestimmen:

- die Fachcommunities (zugleich Nutzerinnen und Nutzer)
- den Anbietern der Informationsinfrastruktur (darunter auch Verlage)
- der Wissenschaftspolitik (Forschungsförderer) für die Wissenschaftsgemeinschaft

Dabei ist zu berücksichtigen, dass diejenigen, die Forschungsdaten produzieren, nicht notwendigerweise diejenigen sind, die Forschungsdaten nutzen. Darüber hinaus ist eine partielle Ungewissheit hinsichtlich des konkreten Zwecks festzuhalten, auf den die Kriterien im Einzelnen gerichtet werden. Deutlich wird dies im Zusammenhang der Nachnutzung digitaler Forschungsdaten. Diese werden gerade mit dem Ziel aufbereitet und vorgehalten, für zukünftige, möglicherweise heute noch gar nicht absehbare Zwecke zur Verfügung zu stehen. So verwenden wir heute Wetterdaten und meteorologische Aufzeichnungen aus den vergangenen Jahrhunderten für die Untersuchung der globalen Erderwärmung – ein Problem, was für diese Forschungsdaten sammelnden Wetterforscher mutmaßlich nicht spezifizierbar war. Dies bedeutet, dass digitale Forschungsdaten ihren Wert aus potentiellen (zukünftigen) Forschungszwecken erhalten, die wir nur teilweise abschätzen können. Sie müssen in einer Form vorliegen, die möglichst aufwandsarm ihre Nutzung und Weiterverarbeitung ermöglicht. Darüber hinaus sollten sie möglichst ohne spezifischen Kontext verständlich sein oder zumindest Hinweise auf die Mittel bereithalten, mit deren Hilfe das Gelingen kann. Die Grundbedingung dafür ist eine adäquate und eindeutige Beschreibung durch Metadaten und ggf. weiteren Ausführungen in Form einer Datendokumentation.

<sup>13</sup> <http://www.repositoryaudit.eu/>

<sup>14</sup> <http://www.wdc.rl.ac.uk/wdc/guide/wdcguide.html>

<sup>15</sup> [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510)

<sup>16</sup> <http://nestor.cms.hu-berlin.de/moinwiki/>

AG\_Vertrauensw%C3%BCrdige\_Archive\_-\_Zertifizierung

<sup>17</sup> <http://www.data-audit.eu/>

Dieser Umstand führt uns zu einem weiteren wichtigen Aspekt der Qualitätssicherung: Eine zentrale Schnittmenge digitaler Forschungsdaten und Forschungsdaten-Infrastrukturen bilden die Metadaten. Ihre Bedeutung muss aus zwei Blickwinkeln spezifiziert und typisiert werden. Ohne Beschreibung ist die Benutzung digitaler Forschungsdaten gerade in interdisziplinären Kontexten unmöglich, ohne eine adäquate Forschungsdaten-Infrastruktur ist die (langfristige) Verfügbarkeit und Auffindbarkeit von Metadaten, Datendokumentation und Forschungsdaten selbst nicht möglich. Die Forschungsdaten, ihre Beschreibung durch Metadaten sowie die Forschungsdaten-Infrastrukturen sind demnach eng miteinander verwoben.

Die Herausforderungen für den skizzierten Bereich der Qualität von Metadaten und Datendokumentation (auch von Forschungsdaten und Forschungsdaten-Infrastrukturen) bestehen darin,

- zwischen den verschiedenen angedeuteten Domänen (Informationsinfrastruktur, Informationswissenschaft, Fachdisziplinen/Communities) und innerhalb einer einzelnen Fachdisziplin/Community allgemein anerkannte Kriterien und daraus abgeleitete Kriterien zu finden und zu etablieren;
- die Kriterien operationalisierbar zu machen;
- und zu einer Erfassung der Kriterien zu gelangen, die auf einem möglichst hohen Maß an Objektivität beruht.

Darüber hinaus besteht der Anspruch, zum einen eine möglichst einfache und generische Beschreibung zu haben, zum anderen aber auch möglichst ausreichend fachspezifische oder datenspezifische Beschreibungsanforderungen zu erfüllen. Die besondere Bedeutung von Metadaten für Forschungsdaten wird durch die Arbeitsgruppe KIM (Kompetenzzentrum Interoperable Metadaten) treffend formuliert:

„Metadaten haben eine besondere Relevanz bei Forschungsdaten, da diese ausschließlich anhand dieser Daten beschrieben, verwaltet und klassifiziert werden können. Digitale Objekte (s. a. Digitalisierungsmetadaten) verfügen von sich aus über keine ausreichenden Beschreibungen, welche dem Nutzer schnelle Rückschlüsse, z. B. auf Herkunft der Daten, ihren Autor/-in, das Entstehungsdatum oder die Sprache des Dokumentes, ermöglichen.“<sup>18</sup>

Metadaten für Forschungsdaten enthalten (idealerweise) Informationen über

- den Entstehungskontext der Forschungsdaten (Projekt-, Erhebungs-, Digitalisierungsinformationen),
- die Forschungsfrage und Methode ihrer Generierung und Analyse,
- die Werkzeuge, mit denen sie erstellt, bearbeitet, analysiert oder gespeichert werden,
- ihre (inhärenten) formalen Eigenschaften wie Format, Typ, Größe usw.
- die Rahmenbedingungen ihrer Speicherung und Benutzung (Organisation, Recht, Langzeitarchivierung).

Mit ihrer Hilfe sollten

- 1 Forschungsdaten nachgewiesen und aufgefunden werden können,
- 2 bei Bedarf die Erkenntnisse der auf diesen Daten beruhenden Forschung geprüft werden können,
- 3 Nachnutzung und Weiterverarbeitung der Forschungsdaten ermöglicht werden.

Angesichts der oben genannten Ziele stehen Forschungsdaten, Metadaten und Forschungsdaten-Infrastruktur zwangsläufig in enger Verbindung. Das bedeutet, dass sich die Qualität der Metadaten hinsichtlich der Ziele direkt auf die Qualität der Forschungsdaten auswirkt, so wie die Qualität der Forschungsdaten-Infrastruktur auf die Metadaten zurückwirkt (z. B. wenn es um die Schnittstellen zur Metadatenauslieferung und die Verfügbarkeit geht). Daher sollten alle drei Komponenten der Diskussion als Gesamtgefüge betrachtet werden, auch wenn die Zuständigkeiten, wie gezeigt, in der Regel bei unterschiedlichen Akteuren liegen dürften.

Die DFG hat für die Beschreibung von Forschungsdaten durch Metadaten folgende Empfehlung formuliert:

„Die Daten werden durch Metadaten beschrieben. Mit den Metadaten (mindestens nach Dublin Core) werden zum einen die bibliographischen Fakten festgehalten. Es sind dies der Name des Forschers, der die Daten erhoben hat, die Benennung des Datensatzes, Ort und Jahr der Veröffentlichung sowie technische Daten (Format etc). In den inhaltsbezogenen Metadaten werden die Primärdaten umfassend beschrieben. Hier finden sich die Angaben zu den Rahmenbedingungen, unter denen sie erhoben bzw. gemessen wurden. Hier beschreibt der Autor auch die Fragestellung, unter der die Daten entstanden. Es sollen hier alle Informationen vorliegen, die für eine wiederholte Nutzung der Daten in anderen Fragestellungen erforderlich sind. Die Kriterien des Information Life Cycle Management sollen dabei berücksichtigt werden.“ (DFG, 2009, 3)

Die DFG nennt in dieser Ausführung bereits die formale und inhaltliche sowie technische Erschließung; zu ergänzen wäre die strukturelle Erschließung (bspw. in welcher Beziehung die Elemente eines Forschungsdatensatzes zu-

<sup>18</sup> <https://wiki.d-nb.de/display/DINIAGKIM/Forschungsdaten>



einander stehen oder eine Aggregation von Forschungsdaten(sätzen) beschrieben werden kann). Ein disziplinübergreifendes Minimal-Set an Metadaten fordert Data Cite (vgl. Data Cite, 2011). Darüber hinaus bieten die im Projekt WissGrid formulierten Empfehlungen Metadaten einen Ansatz (vgl. Ludwig & Enke, 2013). Jensen et al. (2011) führen ebenfalls aus, welche inhaltliche Reichweite die Beschreibung digitaler Forschungsdaten haben kann. Die Empfehlungen „Recommended Practices for Online Supplemental Journal Article Materials“ der US-amerikanischen *National Information Standards Organization* (NISO) und *National Federation of Advanced Information Services* (vgl. NISO & NFAIS, 2013) dienen als Guideline zur Beschreibung von Forschungsdaten als Supplemental Material zu Artikeln in Fachzeitschriften.

Eine intensive Auseinandersetzung zur Durchdringung des Themas „Metadaten“ für Forschungsdaten erfolgt am britischen Digital Curation Center.<sup>19</sup> In den Fachcommunities selbst haben sich Metadatenstandards etabliert. Einen Überblick zu den Gebräuchlichkeiten in einzelnen Fachdisziplinen bietet beispielsweise das *nestor-Handbuch zur Langzeitarchivierung digitaler Forschungsdaten* (vgl. Neuroth et al., 2012).

## 9 Exemplarische Fallstudie historische Korpuslinguistik

Wie spezifisch die Anforderungen an Metadaten und Datendokumentation sind und was überhaupt darunter zu verstehen ist, zeigt eine exemplarische Fallstudie für die historische Korpuslinguistik im Projekt LAUDATIO (Long-term Access and Usage of Deeply Annotated Information). In diesem DFG-geförderten Projekt wird an der Humboldt-Universität zu Berlin unter Beteiligung der Fachdisziplinen (Korpuslinguistik und Historische Linguistik), der Informationsinfrastruktur (Computer- und Medienservice) sowie der Informationswissenschaft (Lehr- und Forschungsbereich Informationsmanagement am IBI) ein Forschungsdaten-Repository für die historische Korpuslinguistik aufgebaut. Die historischen Korpora im LAUDATIO-Repository werden dauerhaft und in standardisierten Formaten gespeichert, zur Nachnutzung aufbereitet und nach den Bedingungen von Open Access zur Verfügung stehen. Das LAUDATIO-Repository ist bereits in einer Demo-Version für die Fachwissenschaftlerinnen und Fach-

wissenschaftler verfügbar, die Freigabe ist für den frühen Sommer 2013 geplant.<sup>20</sup>

Das historische Sprachmaterial besteht aus Digitalisaten und deren Transkriptionen, die linguistisch annotiert werden. Meist liegen für die Korpora tiefe linguistische Annotationen vor, sie sind komplex strukturiert und umfangreich, aber u. U. uneinheitlich oder nicht ausreichend dokumentiert. Aus diesen Eigenschaften resultieren spezifische Anforderungen an das Forschungsdaten-Repository und insbesondere an die Erschließung. Sie wird durch ein komplexes Metadatenframework umgesetzt. Das Metadatenframework wird durch die Fachdisziplinen erarbeitet und weiterentwickelt (vgl. Odebrecht & Zipser, 2013). Darüber hinaus konnten Probleme identifiziert werden, die bereits in der theoretischen informationswissenschaftlichen Auseinandersetzung zu digitalen Forschungsdaten auftraten: Ein besonderer Aspekt war es beispielsweise, sich gegenseitig über Begriffsbestimmungen zu verständigen. Ein Beispiel dafür sind die Aggregationsstufen von digitalen Forschungsdaten. Im Fall der Korpuslinguistik lassen sich sowohl Rohdaten (Texte in Form von Digitalisaten), Primärdaten (Transkriptionen) und Forschungsdaten (Annotationen) bestimmen. Darüber hinaus müssen aber alle Verarbeitungsschritte in Bezug auf die Aggregationsstufen sowie die Annotationschritte und -formate vollständig beschrieben werden. Auf allen Aggregationsstufen sind somit bibliografische Beschreibungen wie beispielsweise Autorinnen und Autoren (etwa für die Texte oder Versionen davon oder die Annotationen) erforderlich. Anhand dieser Prozesse wird die Relevanz des Lebenszyklus von Forschungsdaten deutlich, der auch in den Metadaten dokumentiert wird.

Weitere bedeutende Aspekte in der Diskussion sind die strukturelle Erschließung der Korpora und ihrer Annotationen sowie ihre Versionierung. Dies ist insoweit bedeutend für das Forschungsdaten-Repository, als dass es möglich sein soll, auf Korpusausschnitte z. B. in einer Publikation Bezug zu nehmen und dafür möglichst einen persistenten Identifier anzubieten.

Die Dokumentation der Forschungsdaten in LAUDATIO beschreibt darüber hinaus das Korpusdesign sowie Angaben zu den konkreten Annotationen, darunter auch die Annotationsformate und Methoden der disziplinspezifischen Qualitätsprüfung durch z. B. das Inter-Coder Agreement. „Diese Angaben sind für jede (Nach-)Nutzung digitaler Forschungsdaten in der Historischen Korpuslinguistik und für das Hinzufügen von Annotationen

<sup>19</sup> <http://www.dcc.ac.uk/resources/metadata-standards>

<sup>20</sup> <http://www.laudatio-repository.org/>

oder Dokumenten zu einem Korpus bzw. für das Erstellen neuer Korpora auf Grundlage gewonnener Erkenntnisse ('best practice standard') essentiell.<sup>21</sup>

Abgesehen von den fachspezifischen Anforderungen, die aus historischer Linguistik und Korpuslinguistik resultieren, sind die meisten der genannten Anforderungen auch aus Untersuchungen der Informationswissenschaft bzw. der Projekten der Informationsinfrastruktur zumindest abzuleiten, um ein zufriedenstellendes Qualitätsniveau zu erreichen (Anforderungen übergreifender Fachcommunities: CLARIN, DARIAH; Anforderungen seitens der Informationsinfrastruktur: vertrauenswürdige Archiv nach nestor, Data Seal of Approval; Anforderungen seitens der Forschungsförderer), erfolgt in LAUDATIO eine enge Zusammenarbeit zwischen Fachwissenschaft, Infrastruktur und Informationswissenschaft. Vorstellbar ist darüber hinaus, dass die in LAUDATIO erstellten Forschungsdaten der Wissenschaftlerinnen und Wissenschaftler der HU über einen zentralen Metadaten-Pool für Forschungsdaten nachgewiesen werden. Auch für diesen Zweck ist es erforderlich, Kriterien für die Qualität der Erschließung festzulegen, die disziplinübergreifend angewendet werden können.

## 10 Fazit und Ausblick

Empfehlungen und Richtlinien für die Generierung und Verarbeitung digitaler Forschungsdaten, ihre Archivierung und Bereitstellung in Forschungsdaten-Infrastrukturen sowie zu ihrer Erschließung durch Metadaten liegen vor, sie werden aber selten gemeinsam betrachtet. Aufgrund der hohen Komplexität des Themas ist es schwierig, sich ein (interdisziplinäres) Gesamtbild zur Qualitätssicherung von Forschungsdaten zu verschaffen. In einigen Fällen werden die Empfehlungen und Richtlinien ausdrücklich mit dem Ziel formuliert, die Qualität der Forschungsdaten und der Forschungsdaten-Infrastrukturen zu sichern. Häufig sind diese aber allgemein gehalten und können in dieser Form nur ansatzweise als konkretes Kriterium für die Bestimmung der Qualität gelten. Als Desiderate müssen daraus folgend die fehlende Operationalisierbarkeit der Kriterien und fehlende Mittel zu ihrer Überprüfung benannt werden.

Für eine adäquate Qualitätssicherung muss das gesamte Forschungsfeld digitaler Forschungsdaten noch viel weiter durchdrungen und differenziert werden. Am Lehr- und Forschungsbereich Informationsmanagement

am IBI wird dies durch Dissertationsvorhaben und Forschungsprojekte erreicht. Ziel ist dabei, Hilfsmittel zur Bestimmung und Sicherung der Qualität digitaler Forschungsdaten, ihrer Metadaten und der Forschungsdaten-Infrastrukturen für einen möglichst breiten Anwenderkreis zu entwickeln, um etwa das Forschungsdatenmanagement an einer multidisziplinär ausgerichteten Universität wie der HU oder interdisziplinäre Forschungsprojekte zu unterstützen. Auf diesem Weg werden anhand ausgewählter Beispiele die Reichweite und Grenzen möglicher Qualitätskriterien unter Berücksichtigung disziplinärer und disziplinübergreifender Anforderungen überprüft.

Der Umgang mit digitalen Forschungsdaten sowie die Analyse und Evaluation von Informationsinfrastrukturen für die Wissenschaft sind Forschungsgebiete der Informationswissenschaft. (vgl. Kindling & Schirmbacher, 2013, in diesem Heft S. 127–136). Sie reflektiert die Prozesse und Entwicklungen in den Fachwissenschaften, der Informationsinfrastruktur und der Forschungsförderer. Eine strikte Trennung der Fachgebiete wie auch der zugehörigen Akteursgruppen ist weder erforderlich noch sinnvoll. Indes ist wichtig, welche Kernkompetenzen an welcher Stelle am besten eingebracht werden können.

## Literatur

- APARSEN Report. Report on Peer Review of Digital Repositories. 2012. [http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33\\_1B-01-1\\_0.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33_1B-01-1_0.pdf) [18.03.2012]
- Data Cite: Data Cite Metadata Schema. 2011. [http://schema.datacite.org/meta/kernel-2.1/doc/DataCite-MetadataKernel\\_v2.1.pdf](http://schema.datacite.org/meta/kernel-2.1/doc/DataCite-MetadataKernel_v2.1.pdf) [18.03.2013].
- DFG: Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. 1998. [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf) [18.03.2013].
- DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. 2009. [http://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) [18.03.2013].
- DFG: Leitfaden für die Antragstellung : Projektanträge. 2012. [http://www.dfg.de/formulare/54\\_01/54\\_01\\_de.pdf](http://www.dfg.de/formulare/54_01/54_01_de.pdf) [18.03.2013].
- ESF & EUROHORCS: EUROHORCS and ESF Vision on a Globally Competitive ERA and their Road Map for actions. 2009. [http://www.era.gv.at/attach/EUROHORCS-ESF\\_Vision\\_and\\_RoadMap.pdf](http://www.era.gv.at/attach/EUROHORCS-ESF_Vision_and_RoadMap.pdf) [18.03.2013].
- Hense, Andreas; Quadt, Florian: Acquiring high quality research data. 2011. D-Lib Magazine 17. doi:10.1045/january2011-hense [18.03.2013].

<sup>21</sup> <http://www.laudatio-repository.org/metadaten-und-dokumentation-von-forschungsdaten/>

- Jensen, Uwe; Katsanidou, Alexia; Zenk-Möltgen, Wolfgang: Metadaten und Standards. 2011. In: Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. 2011. urn:nbn:de:kobv:525-opus-2318 [18.03.2013].
- Kindling, Maxi: e-Research und Bibliotheken. 2012. In: Umlauf, Konrad; Gradmann, Stefan: Handbuch Bibliothek: Geschichte, Aufgaben, Perspektiven. Stuttgart: Metzler-Verlag – ISBN: 978-3-476-02376-6. S. 146–152.
- Kindling, Maxi; Schirnbacher, Peter: „Die digitale Forschungswelt“ als Gegenstand der Forschung. Lehrstuhl Informationsmanagement. In: Information. Wissenschaft und Praxis 64 (2013) 2–3, S. 127–136
- Lee, Yang W.; Strong, Diane M.; Kahn, Beverly K.; Wang, Richard Y.: AIMQ: a methodology for information quality assessment. 2002. Information & Management 40. S. 133–146.
- Ludwig, Jens & Enke, Harry (Hrsg.): Leitfaden zum Forschungsdaten-Management. 2013. [http://webdoc.sub.gwdg.de/univerlag/2013/leitfaden\\_DGRID.pdf](http://webdoc.sub.gwdg.de/univerlag/2013/leitfaden_DGRID.pdf) [18.03.2013].
- Mauer, Rainer; Schumann, Natascha: Langfristige Sicherung von sozialwissenschaftlichen Forschungsdaten. Das GESIS Datenarchiv für Sozialwissenschaften. 102. Deutscher Bibliothekartag 2013. urn:nbn:de:0290-opus-14028 [21.03.2013].
- McHugh, Andrew; Innocenti, Perla; Seamus, Ross; Ruusalepp, Raivo: Risk management foundations for digital libraries: DRAMBORA (Digital Repository Audit Method Based on Risk Assessment). 2007. ECDL Conference, Budapest, Hungary, 20. September 2007. <http://146.48.87.21/OLP/UI/1.0/Disseminate/1362805510tcoi0Soj57/a221362805510Q8NpM9DL> [18.03.2013].
- Neuroth, Heike; Strathmann, Stefan; Oßwald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (Hrsg.): Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. nestor-Handbuch. 2012. [http://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor\\_lza\\_forschungsdaten\\_bestandsaufnahme.pdf](http://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf) [18.03.2013].
- NISO & NFAIS: Recommended Practices for Online Supplemental Journal Article Materialis. 2013. [http://www.niso.org/apps/group\\_public/download.php/10055/RP-15-2013\\_Supplemental\\_Materials.pdf](http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf) [18.03.2013].
- Odebrecht, Carolin & Zipser, Florian: LAUDATIO – Eine Infrastruktur zur linguistischen Analyse historischer Korpora. DTA-/CLARIN-D Konferenz und -Workshops: Historische Textkorpora für die Geistes- und Sozialwissenschaften. Fragestellungen und Nutzungsperspektiven, Berlin 2013.
- Office of Science and Technology Policy: Increasing Access to the Results of Federally Funded Scientific Research. 2013. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf) [18.03.2013]
- Pampel, Heinz: Forschungsdaten-Repositoryn. Stand und Perspektiven. 102. Deutscher Bibliothekartag 2013. 2013. urn:nbn:de:0290-opus-13804 [21.03.2013].
- Pampel, H.; Pfeiffenberger, H.; Schäfer, A.; Smit, E.; Pröll, S.; Bruch, C. Report on Peer Review of Research Data in Scholarly Communication. 2012. [hdl:10013/epic.39289](https://doi.org/10.1001/epic.39289) [18.03.2013].
- Peters-Kottig, Wolfgang & Hasler, Tim: Ist auch drin was draufsteht?: Qualitätssicherung bei der Langzeitarchivierung von Forschungsdaten. 2013. 102. Deutscher Bibliothekartag 2013. urn:nbn:de:0290-opus-14241 [21.03.2013].
- RIN: Quality assurance and assessment of scholarly research: A guide for researchers, academic administrators and librarians. 2010. [http://www.rin.ac.uk/system/files/attachments/Quality\\_Assurance\\_screen\\_0.pdf](http://www.rin.ac.uk/system/files/attachments/Quality_Assurance_screen_0.pdf) [18.03.2013].
- RIN: To Share or not to Share: Publication and Quality Assurance of Research Data Outputs. 2008. <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf> [18.03.2013].
- The Royal Society: Science as an open enterprise: The Royal Society Science Policy Centre report 02/12. 2012. [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf) [18.03.2013].
- The ODE Project: Ten Tales and Barriers of Drivers and Barriers in Data Sharing. 2011. [http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836\\_ODE\\_brochure\\_final.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836_ODE_brochure_final.pdf) [18.03.2013]
- OECD: OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007. <http://www.oecd.org/sti/sci-tech/38500813.pdf> [18.03.2013]
- Rümpel, Stefanie: Der Lebenszyklus von Forschungsdaten. In: Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. 2011. urn:nbn:de:kobv:525-opus-2268
- Wang, Richard Y.; Strong, Diane M.: Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems 12 (1996) 4, S. 5–34.



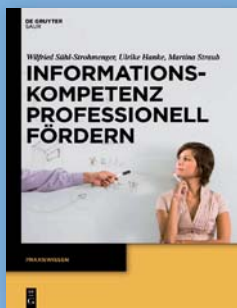
**Maxi Kindling, M.A.:**

Humboldt-Universität zu Berlin  
 Institut für Bibliotheks- und Informationswissenschaft  
 Lehrstuhl Informationsmanagement  
 Unter den Linden 6  
 10099 Berlin  
[maxi.kindling@hu-berlin.de](mailto:maxi.kindling@hu-berlin.de)  
<http://www.ibi.hu-berlin.de/institut/personen/kindling>

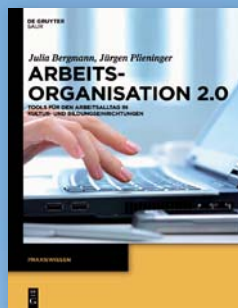
**Maxi Kindling** studierte Bibliothekswissenschaft und Germanistische Linguistik an der Humboldt-Universität zu Berlin und promoviert derzeit am IBI zum Thema Qualitätssicherung im Umgang mit digitalen Forschungsdaten. Sie ist wissenschaftliche Mitarbeiterin am Lehr- und Forschungsbereich Informationsmanagement des IBI und betreut Drittmittelprojekte in den Bereichen Open Access und digitale Forschungsdaten des IBI und der Arbeitsgruppe Elektronisches Publizieren (AGEP).

# FIT FÜR DEN BIBLIOTHEKSALLTAG

## DE GRUYTER SAUR PRAXISWISSEN



2012. 128 Seiten  
**Br.** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-027371-7  
**eBook** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-027438-7  
**Print + eBook** € 79,95 / RRP US\$ 112,00  
 ISBN 978-3-11-027439-4



2012. 149 Seiten  
**Br.** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-026950-5  
**eBook** Open Access



2012. 163 Seiten  
**Br.** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-026897-3  
**eBook** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-026955-0  
**Print + eBook** € 79,95 / RRP US\$ 112,00  
 ISBN 978-3-11-026956-7

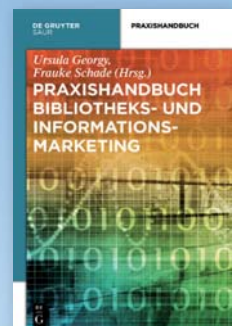


2013. Ca. 200 Seiten  
**Br.** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-030694-1  
**eBook** € 49,95 / RRP US\$ 70,00  
 ISBN 978-3-11-030704-7  
**Print + eBook** € 79,95 / RRP US\$ 112,00  
 ISBN 978-3-11-030705-4

### Praxisrelevantes Wissen auch in unseren Handbüchern:



2012. 592 Seiten  
**Gebunden** € 128,95 / RRP US\$ 181,00  
 ISBN 978-3-11-025473-0  
**eBook** € 128,95 / RRP US\$ 181,00  
 ISBN 978-3-11-025518-8  
**Print + eBook** € 199,95 / RRP US\$ 280,00  
 ISBN 978-3-11-916726-0



2012. 670 Seiten  
**Gebunden** € 129,95 / RRP US\$ 182,00  
 ISBN 978-3-11-026042-7  
**eBook** € 129,95 / RRP US\$ 182,00  
 ISBN 978-3-11-026043-4  
**Print + eBook** € 199,95 / RRP US\$ 280,00  
 ISBN 978-3-11-916351-4



2013. Ca. 360 Seiten  
**Gebunden** € 99,95 / RRP US\$ 140,00  
 ISBN 978-3-11-025553-9  
**eBook** € 99,95 / RRP US\$ 140,00  
 ISBN 978-3-11-025554-6  
**Print + eBook** € 149,95 / RRP US\$ 210,00  
 ISBN 978-3-11-220457-3



2013. Ca. 300 Seiten  
**Gebunden** € 99,95 / RRP US\$ 140,00  
 ISBN 978-3-11-025933-9  
**eBook** € 99,95 / RRP US\$ 140,00  
 ISBN 978-3-11-025947-6  
**Print + eBook** € 149,95 / RRP US\$ 210,00  
 ISBN 978-3-11-220384-2

Weitere Bände sowie Leseproben finden Sie unter [www.degruyter.com/praxiswissen](http://www.degruyter.com/praxiswissen)

