

Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum

Grundsatzpapier zur Sicherung der langfristigen
Verfügbarkeit von Forschungsdaten

Arbeitsgruppe Datenzentren im Verband DHd



digital humanities im
deutschsprachigen raum

Grundsatzpapier der DHd-AG „Datenzentren“

Der 2013 gegründete Verband DHd („Digital Humanities im deutschsprachigen Raum“) versteht sich als Forum und formelle Interessenvertretung für alle, die sich im deutschsprachigen Raum in Forschung und Lehre sowie in Dienstleistungen im Bereich digitale Geisteswissenschaften („Digital Humanities“) engagieren.

In der DHd-Arbeitsgruppe Datenzentren sind Einrichtungen und Infrastrukturen im deutschsprachigen Raum vertreten, die teils schon jetzt, teils perspektivisch konkrete Dienstleistungen zum nachhaltigen Umgang mit Forschungsdaten anbieten.

Im vorliegenden Papier werden Potenziale und Perspektiven der beteiligten Datenzentren sowie der AG Datenzentren beleuchtet und mit einigen Narrativen ergänzt, die einen Einblick in den Themenkomplex geben.

Impressum

Verabschiedet am 01.08.2017

Verband Digital Humanities im deutschsprachigen Raum (DHd) e.V.

Sitz des Vereins: Hamburg

Redaktionsverantwortlich:

Kai Wörner

Zentrum für nachhaltiges Forschungsmanagement

Universität Hamburg

Monetastr. 4

20146 Hamburg

kai.woerner@uni-hamburg.de

Zitiervorschlag:

DHd-AG Datenzentren: Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum.

Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten, Hamburg 2017

DOI: [10.5281/zenodo.1134760](https://doi.org/10.5281/zenodo.1134760)

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0

Internationale Lizenz: <https://creativecommons.org/licenses/by/4.0/deed.de>

Satz und Layout: Kai Wörner, Julia Sorouri

Der Druck wurde gefördert vom Verband DHd

Inhalt

1. Ausgangslage	4
2. Begriffsbestimmung	5
3. Leistungsspektrum eines Datenzentrums	12
4. Ziele und Aufgaben	16
5. Organisationsformen	19
6. Entwicklungsbereiche und Empfehlungen	25
7. Zukunftsperspektiven	27
Mitglieder der DHd AG Datenzentren	28
Bibliographie (Literatur und Webressourcen)	30

Narrative

<i>Erhalt von archäologischen Kulturdenkmälern durch digitale Techniken</i>	6
<i>Das Nachleben von Forschungsprojekten und Forschungsdaten</i>	9
<i>Datenverlust aufgrund fehlender Infrastrukturen</i>	11
<i>Eine Altstudie der Psychologie als Quelle für die Geisteswissenschaften</i>	15

1. Ausgangslage

Das gemeinsame Ziel der Sicherung der langfristigen Verfügbarkeit digitaler Daten und Ergebnisse der geisteswissenschaftlichen Forschung stellt die damit befassten Akteure vor große Herausforderungen, die nur gemeinsam gelöst werden können. WissenschaftlerInnen, Wissenschaftseinrichtungen, Forschungsförderer und zuletzt der Rat für Informationsinfrastrukturen sehen gemeinsam die positiven Auswirkungen dieser Zielstellungen auf Qualität, Transparenz (vergleiche „Sicherung Guter Wissenschaftlicher Praxis“¹), Sichtbarkeit und Nachnutzung von Forschungsergebnissen und -daten und somit auf den wissenschaftlichen Diskurs insgesamt sowie auf die Effizienz der Forschungsförderung.

Von den für den deutschsprachigen Raum ausschlaggebenden nationalen Forschungsförderern erwartet z.B. die Deutsche Forschungsgemeinschaft (DFG) bei Projektanträgen eine Stellungnahme zum geplanten Umgang mit bereits existierenden und neu entstehenden Daten, in der Maßnahmen skizziert werden, die die Nachhaltigkeit und Nachnutzbarkeit von Projektergebnissen auch nach der finanzierten Projektphase sicherstellen sollen.² Im Förderprogramm der Union

der deutschen Akademien der Wissenschaften können keine Anträge gestellt werden, die keine „digitale Komponente“ enthalten – wobei auch Aussagen zur nachhaltigen Datenbereitstellung erwartet werden. Der Schweizerische Nationalfonds (SNF) verlangt seit 2017 die Vorlage von Datenmanagementplänen für eine Projektförderung³ ebenso wie der österreichische Wissenschaftsfonds (FWF)⁴, der darüber hinaus 2016 auch ein eigenes Pilotprogramm zum Thema offene Forschungsdaten (Open Research Data – ORD)⁵ ausgeschrieben hat. Insbesondere das europäische Förderprogramm Horizon 2020 stellt weitgehende Anforderungen an das Datenmanagement, die sich an den FAIR-Data-Prinzipien orientieren.⁶

Auf dem Weg zur Sicherstellung der langfristigen Verfügbarkeit geisteswissenschaftlicher Forschungsergebnisse und -daten müssen noch grundlegende Fragen und Probleme gelöst werden. Bestehende und im Aufbau befindliche geisteswissenschaftliche Datenzentren unterscheiden sich teils stark darin, wie sie in die wissenschaftliche Gemeinschaft eingebunden sind, wie sie organisiert sind und wie sie

1 Siehe „Deutsche Forschungsgemeinschaft (DFG) (1998, 2013)“.

2 Für Grundsätze und Leitlinien siehe „DFG / Umgang mit Forschungsdaten“.

3 Siehe ‚SNF | Open Research Data‘.

4 Siehe ‚FWF | Open Access Policy‘.

5 Siehe ‚FWF | Open Research Data Pilot‘.

6 Fair steht für Findable, Accessible, Interoperable, Re-Usable, siehe ‚FORCE11 | The FAIR Data Principles‘. Siehe auch ‚European Commission (2016)‘.

ihre Angebote langfristig sicherstellen wollen.

Obwohl Konzepte z.B. zur Langzeitarchivierung seit den 1990er Jahren entwickelt werden und andere wissenschaftliche Fachgemeinschaften bereits erfolgreich globale Einrichtungen für das Datenmanagement oder sogar explizite Datenzentren aufgebaut haben und nachhaltig betreiben, erfüllen die meisten bestehenden Angebote die sich abzeichnende breite Anforderungspalette geisteswissenschaftlicher digitaler Forschung noch nicht vollständig. Ein geeignetes Instrument, um fachspezifische Lösungen zu entwickeln, sind dezidierte geisteswissenschaftliche Datenzentren. Dieses Positionspapier der AG Datenzentren⁷ des Verbandes „Digital Humanities im deutschsprachigen Raum“ (DHd)⁸ beschreibt ein gemeinsames Verständnis der speziellen Gegebenheiten in den Geisteswissenschaften und leitet daraus die Anforderungen an sowie die spezifischen Aufgaben von geisteswissenschaftlichen Datenzentren ab. Dabei dient dieses Dokument zur Sicherung des Erkenntnisstandes in einer anhaltenden Diskussion: Datenzentren müssen nicht nur aufgebaut und nachhaltig betrieben werden, sondern sich ständig weiterent-

wickeln, um mit den wissenschaftlichen, technischen und organisatorischen Neuerungen und Veränderungen Schritt halten zu können. Dies geschieht idealerweise im Rahmen einer Vernetzung der einzelnen Zentren miteinander sowie mit weiteren Infrastruktureinrichtungen.

2. Begriffsbestimmung

Ein geisteswissenschaftliches Datenzentrum sorgt dafür, dass für die Geisteswissenschaften relevante Ressourcen über die üblichen Lebenszyklen von Projekten, Datenformaten und Software hinaus verfügbar sowie über die Aktivität der ursprünglichen DatenerzeugerInnen hinaus nutzbar bleiben. Für eine weitere Diskussion der Aufgaben und Rollen eines Datenzentrums ist ein gemeinsames Verständnis darüber, wie der Forschungszyklus in den Geisteswissenschaften aussieht und was geisteswissenschaftliche Ressourcen sind, eine essenzielle Ausgangsbasis.

„Ressourcen“ können alle (digitalen) Artefakte sein, die während eines Forschungsprozesses verwendet werden, in ihm entstehen oder sein Ergebnis sind. Diese Ressourcen unterscheiden sich in den Geisteswissenschaften häufig (1.) in ihrer Art und (2.) in ihrem Bezug zum Forschungsprozess, also in ihrer Entstehung und Verwendung, von nicht-geisteswissenschaftlichen Disziplinen und werden daher von den gängigen Definitionen zu Forschungsdaten nur unzureichend erfasst:

⁷ Siehe ‚Verband Digital Humanities im deutschsprachigen Raum (DHd) | AG Datenzentren‘.

⁸ Siehe ‚Verband Digital Humanities im deutschsprachigen Raum (DHd) | Homepage‘.

Erhalt von archäologischen Kulturdenkmälern durch digitale Techniken

Felix Schäfer, IANUS – Forschungsdatenzentrum Archäologie & Altertumswissenschaften, Deutsches Archäologisches Institut, Berlin

Die kriegerischen Auseinandersetzungen der letzten Jahre in Nordafrika und im Nahen Osten veranschaulichen auf dramatische Weise, dass auch das in Stein gebaute archäologische und kulturelle Erbe von Zerstörung und Zerfall bedroht ist. Analoge und digitale Dokumentationen können diesen gewalttätigen physischen Prozess zwar nicht aufhalten, aber sie können einen wesentlichen Beitrag dazu leisten, den Umfang der Zerstörung festzuhalten sowie die wissenschaftliche, kulturhistorische und damit auch identitätsstiftende Bedeutung der Denkmäler und Objekte herauszustellen. Eine umfassende Datengrundlage bildet die Voraussetzung dafür, dass auch in Zukunft eine Forschung an nicht mehr erhaltenen Kulturgütern möglich ist und dass im Bedarfsfalle z.B. beschädigte Gebäude authentisch wiederhergestellt oder unter Schutz gestellt werden können.

Damit diese Informationsfülle auch tatsächlich genutzt werden kann, ist es einerseits notwendig, die in der Regel einmaligen analogen Bestände zu digitalisieren und international zur Verfügung zu stellen, wie dies beispielhaft in dem Syrian Heritage Archive Project des Deutschen Archäologischen Institutes zusammen mit dem Museum für Islamische Kunst in Berlin erfolgt (<https://www.dainst.org/projekt/-/project-display/199951>). Andererseits geht es aber auch darum, alle digitalen Daten – sowohl die retrodigitalisierten als auch die seit einigen Jahren erzeugten born-digital – langfristig nutzbar und verfügbar zu halten, eine Zielsetzung, die das Projekt IANUS – Forschungsdatenzentrum Archäologie & Altertumswissenschaften verfolgt (<https://www.ianus-fdz.de>).

So wie heute ArchäologInnen noch mit Grabungsunterlagen des 19. und 20. Jahrhunderts als Primärquellen arbeiten, werden auch künftige Generationen auf die digitalen Daten von heute zurückgreifen wollen und müssen. Das Problem der langfristigen Archivierung und Bereitstellung von Forschungsdaten betrifft somit zentrale Grundlagen des Fachs. Dies gilt insbesondere deshalb, da im Verlauf von archäologischen Ausgrabungen und Oberflächenbegehungen einmalige historische Befundsituationen irreversibel zerstört werden. Die vollständige, nachvollziehbare und zugängliche Dokumentation solcher Maßnahmen besitzen daher eine wissenschaftlich wie gesellschaftlich zeitlich unbegrenzte Relevanz und Gültigkeit.

(1.) Während in den Natur- und Sozialwissenschaften in der Regel Mess- oder Erhebungsdaten die Grundlage der Forschung bilden, bauen die Geisteswissenschaften zusätzlich häufig auf digitalen Repräsentationen von kulturellen Artefakten wie Manuskripten, Texten, Bildern, Audio- und Videoaufnahmen oder anderen materiellen Objekten auf. Die in den Natur- und (quantitativen) Sozialwissenschaften vorgängig definierten Modelle führen meist zur Erhebung klar strukturierter Daten. Dagegen betreiben die Geisteswissenschaften in der Regel eine nachgängige Modellierung in Form fortschreitender Repräsentation, Erschließung, Bearbeitung, Visualisierung, Interpretation und Annotation. Für ihre häufig semi- oder komplex strukturierten Daten sind darüber hinaus Historizität und Kontextualität wichtige Dimensionen. Geisteswissenschaftliche Ressourcen bilden verschiedene Perspektiven ab und können in unterschiedlichen Formaten und Aggregationsstufen vorliegen, sind also in vielen Fällen hinsichtlich Struktur und Inhalt sehr heterogen. Im Allgemeinen muss von einem offenen, nichtlinearen Lebenszyklus der Ressourcen⁹ ausgegangen werden, bei dem Ressourcen in immer neuen Kontexten genutzt, angereichert und mit anderen Daten verbunden werden; die inhärent dynamische Natur der Daten stellt eine wesentliche Herausforderung für die Datenrepräsentation dar.

⁹ In diesem Kontext greifen traditionelle Datenformate (Datenbank-, XML-basiert) und Datenmodelle (Baum, Prozessschichten, Tabellen) zu kurz. Als Denkmodell könnte ein stets wachsender heterogener globaler (Knowledge) Graph angemessener erscheinen.

Gleichzeitig besitzen viele digitale Repräsentationen von geisteswissenschaftlichen Quellen sowie born digital-Materialien eine anhaltende wissenschaftliche Relevanz, vor allem, wenn es sich um einmalige Dokumente handelt, z.B. um Texte aus früheren Jahrhunderten, Film-/Ton-Aufnahmen von historischen Ereignissen und Dokumentationen von vergangenen Zuständen (z.B. bei Bauwerken). Diese Ressourcen sind deshalb nicht für einen bestimmten Zeitraum (z.B. zehn Jahre) zur Wahrung der Transparenz und Reproduzierbarkeit von Analysen, sondern für einen unbestimmten Zeitraum (im Prinzip, wie bei den Archiven und Bibliotheken: für immer) für eine wiederkehrende Nutzung vorzuhalten.

(2.) Anders als in den Naturwissenschaften sind in den Geisteswissenschaften die Schritte im Forschungsprozess manchmal nicht deutlich voneinander getrennt. Daher haben auch „Forschungsdaten“ in den Geisteswissenschaften ihre eigene Spezifik. Eine klare Definition ist schwierig, da eine scharfe Trennung zwischen den einzelnen Bearbeitungsstufen der Daten oft nicht eindeutig möglich ist.¹⁰

Häufig werden alle Forschungsarbeiten auf einer Datenbasis durchgeführt, die kontinuierlich erweitert, verfeinert, annotiert, angereichert und verlinkt wird. Während

¹⁰ Prinzipiell gibt es verschiedene Daten und Stufen der Bearbeitung, etwa: Ausgangsdaten – „Quellen“; unmittelbare Repräsentanzen (Digitalisate), Born Digital-Materialien;

in den Natur- und Sozialwissenschaften oft vollständige Datensätze oder Datensammlungen zur Nachnutzung vorgesehen werden, zielt der geisteswissenschaftliche Forschungsprozess regelmäßig auf einzelne Objekte aus einer oder mehreren verteilten Sammlungen. Diese Daten sind deshalb auch nur in einer spezialisierten, mehrere Phasen des Forschungszyklus unterstützenden Softwareumgebung sinnvoll zu nutzen, die das Retrieval unterstützt, die Objekte präsentiert, ihre Verlinkung erlaubt und ihre weitere Bearbeitung oder Nachnutzung ermöglicht. Neben Daten können auch Funktionen auf den Daten bzw. komplette Anwendungen wie Arbeits- und Präsentationsumgebungen relevante Ressourcen sein. Auf einer abstrakten Ebene müssen deshalb neben den Daten auch „Behavior“ und „Appearance“ als „Significant Properties“ ihrer Benutzungsumgebungen bewahrt werden.¹¹ Daher sind die erhaltenswerten Forschungsergebnisse in den Geisteswissenschaften mehr als nur Forschungsdaten im engeren Sinne: Auch Präsentationsumgebungen und Applikationen, die Daten interpretieren, Möglich-

aufbereitete (intermediäre) Daten - angereicherte Repräsentanzen; Annotationen; Ergebnisdaten - Publikationen; Texte, Präsentationssysteme, Visualisierungen etc. Es lassen sich jedoch keine klaren Grenzen zwischen den verschiedenen Bearbeitungsstufen bestimmen: So können Inhalte in Datenbanken zum Beispiel als Ausgangsdaten, intermediäre Daten oder auch als Ergebnisdaten fungieren.

¹¹ Siehe ‚JISC | The significant Properties of Digital Objects‘ und ‚NARA 2009‘, insbes. Annex C.

keiten zur Recherche und Verknüpfung schaffen sowie Services zur Nutzung und Weiterverarbeitung anbieten, zählen dazu.

Die Komplexität geisteswissenschaftlicher Ressourcen ergibt sich so insgesamt aus den Anforderungen an Formate (in ihrer Vielfalt und Komplexität), aus den (nicht in klar abgrenzbare Prozessschritte aufzutrennenden) Datenschichten, aus der bedeutungstragenden Verlinkung (zwischen unterschiedlichen Datentypen und mit anderen Datenquellen) und den „lebenden Systemen“ zur kontinuierlichen projektspezifischen Bereitstellung und Präsentation. Eine weitere, nicht unerhebliche Komplexität liegt in den Rechtsfragen, die mit geisteswissenschaftlichen Forschungsdaten verbunden sind.¹² Die Daten können dem Urheberrecht, den Verwertungs-, Eigentums- und Persönlichkeitsrechten, teilweise auch mehrfach, unterliegen. Zum Beispiel können bei der Transkription eines Interviews, das zuvor in einer Zeitschrift veröffentlicht wurde, alle diese Rechtsbereiche betroffen sein. Auch wenn häufig stillschweigend davon ausgegangen wird, dass Forschende auch ohne eine vertragliche Regelung ihre Daten zumindest der wissenschaftlichen Welt zur Verfügung stellen können, ist dies ohne eine explizite Einwilligung aller beteiligten Akteure und Rechteinhaber rechtlich nicht zulässig. Datenzentren können zwar versuchen,

¹² Siehe dazu die jüngst erschienen Empfehlungen und Problematisierung des Rats für Informationsinfrastrukturen ‚RfII – Rat für Informationsinfrastrukturen (2017)‘.

sich in Verträgen mit den Datenbereitstellenden von Ansprüchen Dritter freistellen zu lassen, aber im Zweifel müssen sie zumindest nachweisen, dass die Datenbereitstellenden Verantwortung z.B. im Rahmen von rechtswirksamen Datenüberlassungsverträgen übernehmen. Selbst bei

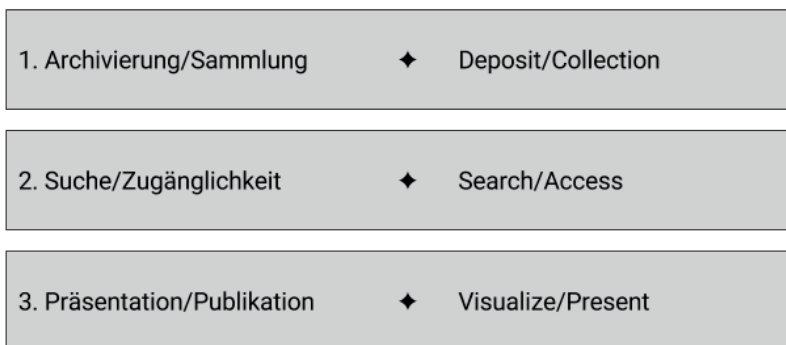
der Bereitstellung von retrodigitalisierten Quellen können unter Umständen Urheberrechte an Datenbanken oder Fragen des Rechtes am Lichtbild relevant sein.¹³

Datenverlust aufgrund fehlender Infrastrukturen

Jörg Wettlaufer, Humanities Data Centre, Göttingen

Nicht immer geht es gut aus, wenn digitale Forschungsdaten nach einem längeren Zeitraum erneut verwendet und unter anderen Fragestellungen bearbeitet werden sollen. 1990 erschien eine Untersuchung zu mittelalterlichen Rechtsquellen (sog. Urfehden) in Regensburg (Steffen Wernicke & Martin Hoernes (1990): „Umb die unzucht die ich handelt han ...“. Quellen zum Urfehdewesen, St. Katharinen [Halbgraue Reihe zur Historischen Fachinformatik. Serie A ; Bd. 9]). Höchst fortschrittlich wurden die Quellen damals digital erfasst und analysiert. Die Texte selber, also die Forschungsdaten, verblieben auf verschiedenen Sicherungsmedien zunächst erstmal bei den Forschern. Von den Autoren der Studie wurden die Daten auf Disketten gesichert und zudem befanden sich die Texte nach dem Ortswechsel des Herausgebers der Reihe in einem Rechenzentrum in einem größeren Konvolut von Sicherungskopien. Heute, 27 Jahre nach der Veröffentlichung, gestaltet sich die Nachnutzung der damals nachweislich schon digital vorliegenden Texte schwierig. Beide Autoren konnten zwar ausfindig gemacht werden, aber die Disketten waren schon vor langer langer Zeit entsorgt worden. Letzte Hoffnung ist also das Rechenzentrum, das vielleicht noch an irgendeiner Stelle die Sicherheitskopie gespeichert haben könnte. Aufgrund fehlender Metadaten und Standards ist die Suche danach jedoch ähnlich schwierig wie die nach der berühmten Stecknadel im Heuhaufen. Und dieses Beispiel ist kein Einzelfall, sondern steht stellvertretend für viele wertvolle digitale Forschungsdaten in den Geisteswissenschaften, die täglich unwiederbringlich verloren gehen, weil sich niemand rechtzeitig um ihre langfristige Verfügbarkeit gekümmert hat.

¹³ Vergleiche u.a. „Deutsche Forschungsgemeinschaft (DFG) (2015)“, bes. S. 17; Klimpel und Weitzmann (2015).



Paradigmatische Aufgabenbereiche eines geisteswissenschaftlichen Forschungsdatenzentrums

Dazu kommen Besonderheiten der verwendeten Standards und Regeln für Datenformate und Beschreibungen, die der Heterogenität der Daten sowie der Unschärfe der Trennung zwischen Ausgangs-, Zwischen- und Ergebnisdaten geschuldet sind. Diese Standards und Regeln sind durch eine große Offenheit und weitgehende Flexibilität gekennzeichnet, was beispielhaft an dem Variantenreichtum bei Auszeichnungen nach den Vorgaben der Text Encoding Initiative (TEI)¹⁴ für geisteswissenschaftliche Daten erkennbar ist.

Auf dieser Grundlage lässt sich ein geisteswissenschaftliches Datenzentrum begrifflich von anderen, rein auf die bitstream preservation orientierten Archivierungseinrichtungen oder auf gleichmäßige Datenformate ausgerichtete Repository-Systeme unterscheiden. Ein geisteswissenschaftliches Datenzentrum muss wie ein „Daten-

archiv“ institutionell auf Dauer angelegt und abgesichert sein und in gleicher Weise auf die nachhaltige Verfügbarkeit seiner Inhalte abzielen. Es sollte auch den permanenten Online-Zugang zu einzelnen Objekten und den Betrieb laufender generischer und teilweise projektspezifischer Softwaresysteme gewährleisten sowie aktives Forschungsdatenmanagement während und nach Ende der Laufzeit von Projekten betreiben. Geisteswissenschaftliche Datenzentren sind in diesem Sinne „lebende Archive“.

Wie Kompetenz- oder Forschungszentren müssen sie integraler Bestandteil des Forschungsprozesses sein, forschungsnah agieren und im Idealfall schon während der Projektentwicklung eingebunden sein, um die Grundlagen nachhaltiger Datenerzeugung zu gewährleisten. Den Unterschied zu Einrichtungen, die stärker auf zeitlich begrenzte Fachforschungsprojekte konzentriert sind, macht ihre Verantwor-

¹⁴ Siehe „TEI | Homepage“.

tung für die Pflege von Ressourcen über Projektlaufzeiten hinaus aus. Geisteswissenschaftliche Datenzentren definieren sich des Weiteren dadurch, dass sie fachspezifische und fachorientierte Kom-

petenzen aufbauen, die für eine Betreuung von spezialisierten Forschungsvorhaben und die langfristige Kuratierung der Daten erforderlich sind.

Das Nachleben von Forschungsprojekten und Forschungsdaten

Patrick Sahle, Data Center for the Humanities, Köln

Forschung findet heute überwiegend projektgetrieben statt. Für einen bestimmten Zeitraum kann Energie und können Ressourcen in Aktivitäten gesteckt werden, die zu Daten führen, die wiederum der Forschung zugänglich gemacht werden. In den Geisteswissenschaften sind die häufig sehr idiosynkratischen Modellen und Formaten folgenden "reinen Daten" außerhalb ihrer Präsentationssysteme praktisch unbrauchbar. Benötigt werden deshalb spätestens zum Projektende tiefgreifende Übergabeprozesse, bei denen die Projekte, ihre Daten und Anwendungen genau dokumentiert werden und der weitere Umgang mit ihnen geplant wird.

Praktische Erfahrungen mit diesen Prozessen gibt es zum Beispiel seit Jahren mit dem "altägyptischen Totenbuch" (<http://tottenbuch.awk.nrw.de/>), dessen Förderung durch das Programm der Akademien der Wissenschaften (hier: Nordrhein-Westfälische Akademie der Wissenschaften und der Künste) 2012 ausgelaufen ist. Das Ergebnis war hier eine sehr spezielle Datenbasis, ein digitales Bildarchiv und eine Präsentationsumgebung, die seitdem kontinuierlich von einem geisteswissenschaftlichen Datenzentrum (DCH) gepflegt wird. Dabei wird nicht nur das technische System gewartet und ggf. aktualisiert, sondern die Daten selbst werden in einem moderierten Prozess von der Fachgemeinschaft weiter aktualisiert und erweitert.¹

Diese Betreuung kann aufrechterhalten werden, solange die notwendigen Mittel zur Verfügung stehen und sich ihr Einsatz durch die Benutzung von System und Daten rechtfertigen lässt. Für die Zukunft wird von Zeit zu Zeit zu prüfen sein, ob (1.) das System weiter gepflegt werden soll, (2.) das System abgeschaltet und nur noch die reinen Daten (Datenbank und Bildarchiv) bereitgestellt werden sollen oder (3.) Daten und Anwendung in einem Folgeprojekt aufgehen und damit obsolet werden – wobei man auch hier den Stand der Daten archivieren würde.

¹ Ulrike Henny, Jonathan Blumtritt, Marcel Schaeben, Patrick Sahle: The life cycle of the Book of the Dead as a Digital Humanities resource. In: Digital Classics 3/2 (2017). DOI: 10.11588/dco.2017.0.35896

3. Leistungsspektrum eines Datenzentrums

Ein Datenzentrum trägt aktiv dazu bei, dass digitale Ressourcen kontinuierlich im Forschungszyklus eingebunden werden können. Dazu müssen die Ressourcen auffindbar, zugänglich, zitierbar, vernetzt und nutzbar sein.

Zu den Grundaufgaben eines Datenzentrums gehört die sichere und strukturierte Aufbewahrung von Ressourcen, die auch eine Referenzierung ermöglicht. Für die Anforderungen der meisten Förderer ist heute zwar eine Sicherung und Aufbewahrung von Primärdaten für einen gewissen Mindestzeitraum (üblicherweise zehn Jahre) ausreichend. Der Anspruch und die Aufgabe der Geisteswissenschaften, einen wesentlichen Beitrag zur langfristigen Erhaltung und Verstehbarkeit des kulturellen Erbes, seiner Repräsentations-, Erschließungs- und Bearbeitungsformen beizutragen, verlangt jedoch nachhaltigere, zeitlich theoretisch unbegrenzte Archivierungs- und Bereitstellungsstrategien.

Ein an der sich kontinuierlich weiterentwickelnden Wissenschaft ausgerichtetes Datenzentrum wird außerdem jenseits der Archivierung eine weitere Einbindung in den wissenschaftlichen Diskurs anstreben und eventuell zusätzliche Angebote entwickeln wollen, die die Sichtbarkeit und die Nutzbarkeit von digitalen Ressourcen weiter erhöhen. Man kann deshalb von zunächst drei Aufgabenbereichen für die praktischen

Aufgaben eines Datenzentrums sprechen: 1.) Archivierung, 2.) Bereitstellung, und 3.) Präsentation (siehe Abbildung).

1. Aufgabenbereich: Archivierung

Ein wesentliches Element nachhaltiger wissenschaftlicher Praxis ist die Archivierung aller in einem Projekt entstandenen relevanten Ressourcen, aus denen die Projektergebnisse abgeleitet wurden oder die selbst als Ergebnisse zu betrachten sind (deposit / collection). Neben der Erhaltung der Daten und möglicher Maßnahmen zur Langzeitarchivierung (z.B. regelmäßige Migrationsschritte) müssen zukünftige NutzerInnen, d.h. sowohl Menschen als auch Software-Systeme, die Ressourcen interpretieren und verstehen können. Inhaltliche Interpretierbarkeit setzt dabei voraus, dass die digitalen Ressourcen (a) selektiert und strukturiert, (b) mit fachspezifischen Kontextinformationen z.B. über Forschungsfrage und -methode beschrieben sowie (c) mit technischen und administrativen Metadaten versehen sind. In einem Datenzentrum archivierte Ressourcen sind im Rahmen dieses Aufgabenbereiches inhaltlich eingefroren; dies ist sowohl zur Bewahrung der Datenintegrität als auch zur Zitierbarkeit der Ressourcen notwendig, da Migrationsschritte zum Zweck der langfristigen Erhaltung der Daten auch inhaltliche und funktionale Auswirkungen haben können, die von einem Datenzentrum wissenschaftlich anhand der erhaltenswerten signifikanten Eigenschaften eines digitalen Objektes gegeneinander abgewogen werden müssen.

2. Aufgabenbereich: Zugänglichkeit

Damit archivierte Ressourcen im Forschungszyklus eingebunden bleiben können, muss ein Datenzentrum Ressourcen bereitstellen, auffindbar und zugänglich machen, Browse- und Suchfunktionen anbieten, Daten über standardisierte Schnittstellen bereitstellen und die Zitierbarkeit der Ressourcen sicherstellen (search / access). Eingeschränkt zugängliche Ressourcen, bei denen Lizenzfragen oder explizite Nutzungsvereinbarungen eine Rolle spielen, müssen dabei vor unerlaubtem Zugriff geschützt werden. Für eine minimale Nachnutzbarkeit können Ressourcen unverändert in der Form heruntergeladen werden, in der sie im Projektkontext genutzt wurden. Im Sinne einer einfacheren inhaltlichen Zugänglichkeit der Ressourcen werden Datenzentren zuweilen eine Form der Vereinheitlichung („Normalisierung“) anstreben. Im Rahmen nachträglicher Kuratierung kann es zu Homogenisierungen kommen, bei denen zum einen Detailinformationen oder idiosynkratische Strukturen verloren gehen, zum anderen aber auch Informationen angereichert werden.

Bei den Systemen sind spezialisierte Datenbanken (je nach Significant Properties) ohne eine Zugriffsschicht meist nicht unmittelbar nutzbar, und interaktive Visualisierungen müssen mit ihrer technischen Umgebung archiviert oder aus ihr heraus extrahiert werden. Normalisierte Ressourcen sind insofern u.U. nur dann nachnutzbar, wenn NutzerInnen sie selbst übernehmen, wiederherstellen und wieder für ihre

Forschungen einsetzbar machen können. Die Nutzbarkeit von Ressourcen bemisst sich für WissenschaftlerInnen an dem Aufwand, der nötig ist, um Ressourcen einzusehen und in die eigenen Forschungsarbeiten einzubetten. Um die Nutzung zu vereinfachen, bieten Datenzentren daher üblicherweise Dienste und Lösungen an, die über minimale Dienste zur Archivierung und Zitierbarkeit hinausgehen.

3. Aufgabenbereich: Präsentation

Gerade in den Geisteswissenschaften zielt das primäre Nutzungsszenario häufig nicht auf die systematische algorithmische Auswertung ganzer Datensammlungen, sondern auf die Auffindbarkeit und Nutzung einzelner Datenobjekte (visualize / present). Um solche digitalen Objekte kontinuierlich durch Online-Präsentation bzw. Publikation zugänglich zu machen, bietet sich der Aufbau von integrierten Repositorien an, die verschiedene normalisierte Datenbestände aufnehmen können. Zugleich kann das Datenzentrum dadurch übergreifende Werkzeuge zum Retrieval und zur Analyse bereitstellen.

Weil generische Repositories niemals alle individuellen Ressourcen ohne funktionale Verluste aufnehmen können, wird man grundsätzlich prüfen müssen, ob auch einzelne projektspezifische Präsentationsumgebungen in Betrieb zu halten sind. Diese werden im Regelfall stärker gegenstandsspezifische Nutzungsformen unterstützen. Es ist aber aus technischen, finanziellen und rechtlichen Gründen heute

noch unklar, wie weit diese funktionale Anwendungserhaltung im Einzelfall unterstützt werden kann. Statt einer generellen 1:1-Übernahme werden einige Datenzentren möglicherweise abgestufte Angebote aufbauen, bei denen die Mehrheit der Ressourcen entweder in ein allgemeines Repository eingehen oder als Datensammlungen archiviert und angeboten werden. Andere werden übergreifende Systeme anbieten, die das Nutzungsschema der betreffenden Ressourcen unterstützt; oder sie werden Präsentationssysteme auch aus Einzelprojekten über eine längere Zeit hinweg unterstützen und warten.

Idealerweise arbeiten Projekte und Datenzentren von vornherein zusammen, um projektspezifische technische Lösungen gleich bei ihrer Erstellung hinsichtlich Wartbarkeit zu optimieren. Datenzentren können hierbei modulare erweiterbare Basisapplikationen bieten, die auf die Archivierbarkeit abgestimmt sind. Dadurch ergeben sich Synergien und ein Mehrwert für beide Seiten.

Bei Forschungsprojekten, deren Daten (meist in Form einer Anwendung) bearbeitet und ergänzt werden, obwohl die Förderung des Projektes bereits ausgelaufen ist, stellt die zeitgleiche Archivierung und Konservierung von Forschungsdaten ein grundsätzliches Problem dar, weil eine Spannung zwischen der digitalen Repräsentation des fortlaufenden Forschungsprozesses und der Archivierung stabilisierter Daten besteht, die nicht

ohne weiteres aufgelöst werden kann. Die dauerhafte Pflege von Arbeitsumgebungen, in denen Ressourcen kontinuierlich weiter bearbeitet und angereichert werden können (process / analyze), kann durchaus zu den Aufgaben eines Datenzentrums gehören, wenn nach dem Ende eines Forschungsprojektes und seiner Förderung deren interaktive Plattform dauerhaft vorgehalten werden soll, um von der wissenschaftlichen Community weiter genutzt zu werden. Neben den technischen Schwierigkeiten ist für diese Fälle konzeptionell die Dualität der zur Zitation stabilisierten Daten und den im weitergeführten Forschungsprozess veränderlichen Daten zu lösen. Forschungsdatenzentren haben hier u.U. eine doppelte Aufgabe im Betrieb der Plattformen und in der Archivierung von zeitlich gebundenen Datenständen.

Darüber hinaus gibt es eine Reihe von Angebotstypen, die über die beschriebenen Anwendungsszenarien hinausgehen. Dazu zählen neben den im Aufbau befindlichen Ressourcen noch laufender Forschungsprojekte derzeit vor allem zwei Ressourcentypen. Zum einen sind das sogenannte „lebende Archive“, deren Konzept, die kontinuierliche Anreicherung und auch Weiterentwicklung der zugrundeliegenden Forschungsdaten und Präsentationsmechanismen vorsieht. Zum anderen sind dies Anwendungen wie Aggregatoren (dissemination services), die ihren Datenbestand teilweise oder sogar vollständig aus anderen Anwendungen akkumulieren. In beiden Fällen verändern sich mit

Altstudien der Psychologie als Quelle für die Geisteswissenschaften: Das Beispiel der Bonner Längsschnittstudie des Alterns (BOLSA) 1964-1984

**Katrin Moeller, Historisches Datenzentrum Sachsen-Anhalt (Hist-Data),
Universität Halle-Wittenberg**

In den Kellerräumen einer deutschen Universität fand die Historikerin Christina von Hohenberg im Jahr 2014 das, was sie nach dem Studium von Forschungsliteratur dringend gesucht hatte: Die Bonner Längsschnittstudie des Alterns. Die in London tätige Wissenschaftlerin erforscht die 1968er Jahre nicht aus der Sichtweise der jungen, sondern der alten Menschen und nimmt so einen originellen Perspektivwechsel vor. Die zwischen 1964 und 1984 entstandene teure Drittmittelstudie hatte in der psychologischen Forschung weitgehend an Aktualität verloren, die beteiligten WissenschaftlerInnen waren 2014 längst in den Ruhestand gegangen. Die Kellerräume sollten entrümpelt, die Studie entsorgt werden. Der Bestand wurde nach vertraglichen Regelungen in das wissenschaftsgetriebene Historische Datenzentrum Sachsen-Anhalt überführt und zur Digitalisierung aufbereitet (Verzeichnung, Metadaten, Kollationierung). Entscheidend für die Rettung der Studie war auch die Situierung des Forschungsdatenzentrums in der Mitte der geisteswissenschaftlichen Fakultät mit täglichen Kontakten, kurzen Wegen und bekannten Verantwortlichkeiten.

Die Finanzierung der Digitalisierung wird nach einigen Schwierigkeiten von der VW-Stiftung getragen. Geisteswissenschaftliche Datenkuratierung fällt mit ihren Anliegen weder in die DFG-Bereiche des Aufbaus von technischen Infrastrukturen noch der eigentlichen wissenschaftlichen Fachförderung. Zudem sollen Doppelstrukturen neben Archiven und Bibliotheken verhindert werden. Fachgetriebene Forschungsdatenzentren haben es daher schwer.

Besonders die Tonbandinterviews mit den zwischen 1880-1885 und 1900-1905 geborenen Frauen und Männern stellen allerdings einen besonderen Schatz für linguistische, soziologische und historische Forschung dar. Es handelt sich um die frühesten biografischen Interviews im deutschsprachigen Raum überhaupt, die darüber hinaus Frauen aus den mittleren und unteren Sozialschichten zu Wort kommen lassen. Nur wie lassen sich fast 3.000 Stunden Tonbandaufzeichnungen effektiv nutzen? Die Wissenschaft hat hierzu jeweils fachspezifische Wege der Erschließung, Auszeichnung, Transkription und Dokumentation entwickelt. Wissenschaftlicher Mehrwert entsteht nicht allein aus Metadaten, sondern durch Tiefenerschließung und wissenschaftliche Analyse. Darüber hinaus sind vertiefte Kenntnisse der automatisierten Worterkennung und Transkription, des Umgangs mit SPSS-Daten und des Programmierens sowie Fachkenntnisse alter psychologischer Testverfahren notwendig, um

den Bestand adäquat aufzubereiten. Die Arbeiten dazu können nur FachwissenschaftlerInnen leisten. Wir haben uns dazu fachkompetente Hilfe bei SprachwissenschaftlerInnen (Archiv für gesprochenes Deutsch am IDS Mannheim) und PsychologInnen (Universität Halle-Wittenberg, Heidelberg) geholt. Fachspezifisches Forschungsdatenmanagement und Datenkuratierung ist das Herzstück der Wissenschaft.

Für Fragen der Langzeitarchivierung (ITZ Halle-Wittenberg), des Webportals (Servicezentrum eSciences Trier) und der Metadaten-speicherung (Universitäts- und Landesbibliothek Halle) werden dagegen die Träger bisheriger und neuer Infrastruktureinrichtungen aktiv. Es geht nicht um den Aufbau von Doppelstrukturen, sondern um wissenschaftsrelevantes Forschungsdatenmanagement. Gemeinsam werden wir die BOLSA für alle Forschenden offen und langfristig anbieten.

wachsendem oder auch schrumpfendem Datenbestand Rechercheergebnisse und Zitierbarkeit der Ressource. Mit zukünftig wachsenden Möglichkeiten der Nachnutzbarkeit von digitalen Forschungsdaten werden diese Ressourcentypen zunehmende Bedeutung erlangen. Sie bewegen sich häufig in einer Grauzone zwischen Datensammlung und Daten-Service. Hier wird zukünftig zu entscheiden sein, ob und welche Aufgaben für den Betrieb eines Daten-Services in das Aufgabengebiet eines Datenzentrums fallen.

4. Ziele und Aufgaben

Generell ist zwischen einer Begleitung der Projekte und WissenschaftlerInnen von Anfang an (ab ovo), einer Betreuung zur Laufzeit (in vita) und einer Übernahme nach Projektende (post mortem) zu unterscheiden. Die Frage, was mit Forschungsdaten nach dem Ende eines Projektes geschieht, führt unmittelbar zu der Aufgabe, diese Projekte nach Möglichkeit von Anfang an (ab ovo) zu begleiten, um die Nachhaltigkeit ihrer Arbeiten sicherzustellen. Unter optimalen Bedingungen fängt die Aufgabe „Archivierung“ nicht erst mit der Erstellung des ersten digitalen Objektes, sondern mit der Planung des Projekts an. Datenzentren sind im Idealfall gut sichtbare Akteure im Forschungsprozess, die alle Schritte von der Beratung bei Projektentwicklungen und -anträgen, über die Datenmodellierung und Datenerzeugung bis hin zur Ergebnispro-

duktion und schließlich der Übergabe von Ressourcen am Projektende begleiten. Datenzentren müssen deshalb eng mit den FachwissenschaftlerInnen zusammenarbeiten, um schon im Vorfeld eingangsseitig die Qualität der Daten im Blick zu haben und ihre Verstetigung vorzubereiten sowie ausgangsseitig ihre Bewahrung, Adressierbarkeit und Nachnutzbarkeit sicherzustellen.

Zum allgemeinen Aufgabenspektrum geisteswissenschaftlicher Datenzentren zählen deshalb:

1. Forschung und Vermittlung

Dies beinhaltet den internen Kompetenzaufbau in der fachspezifischen Beratung, Projektentwicklung und -umsetzung sowie die Vermittlung von Kompetenzen an Externe in den Bereichen der Data Literacy, Medienkritik, Informationskritik, Metadaten, Forschungsdaten, Werkzeuge und Arbeitsumgebungen durch Fortbildung, Schulung, Lehre und Online-Angebote. Datenzentren übernehmen auch eine Vermittlerfunktion zur Vernetzung von Forschung und Dienstleistungen.

2. Öffentlichkeitsarbeit

Nur wenn die Datenzentren bekannt und sichtbar sind, werden auch die durch sie langfristig verfügbar gehaltenen Ressourcen findbar sein. Durch hohe Sichtbarkeit tragen Datenzentren dazu bei, dass Ressourcen und Daten als Publikationen und eigenständige wissenschaftliche Leistung anerkannt werden.

3. Datenkuratierung

Die Pflege und das Monitoring digitaler Ressourcen im Sinne des Grundziels sichtbarer, zugänglicher und nutzbarer Daten. Dazu gehören das Kuratieren von Daten, die von Zeit zu Zeit ggf. für neue Formate und Standards aufzubereiten sind, die inhaltliche Erschließung und der Einsatz von Anreicherungsdiensten für bestehende Daten sowie die Aggregation von Ressourcen und Metadaten in zentralen Wissensspeichern (z.B. übergreifenden Portalen oder Services). Datenzentren fungieren als allgemeine „Data Provider“.

4. Aufbau und Betrieb von Infrastruktur

Datenzentren müssen sich um Technik, Infrastruktur, Prozesse und Entwicklung kümmern. Hierzu gehören die Bereitstellung und der Betrieb von Speicherinfrastrukturen sowie weitere Elemente des Datenmanagements: die Entwicklung, die Standardisierung, der langfristige Betrieb, die Kuratierung und das Monitoring technischer Services und Schnittstellen der Datenbearbeitung und Datenbereitstellung einschließlich der Regelung und Durchsetzung von Zugriffsrechten. Dies beinhaltet die Entwicklung und Bereitstellung generischer Präsentations- und Disseminationssysteme, z.B. auf der Basis zentralisierter Repository-Systeme, in die möglichst viele Projekte eingehen können. Es kann aber auch die anhaltende Betreuung individueller, projektspezifischer Anwendungen betreffen.

5. Beratung zu rechtlichen und ethischen Fragen sowie Standards

Rechtliche und ethische Fragen und die Auswahl und Anwendung von Standards und Normen erfordern den Aufbau von Expertise in diesem Bereich, da hier nicht nur technische Fragestellungen zu beachten sind, sondern auch langfristige Erfahrungen sowie die Vermittlung von Kompetenzen und das Einhalten von Verfahren erforderlich sind. Will man konkrete Dienste für Projekte und WissenschaftlerInnen über die verschiedenen Szenarien (ab ovo, in vita und post mortem) hinweg aufspannen, dann lassen sich mindestens folgende Aufgaben identifizieren:

- Begleitung von Projekten bei Entwurf und Beantragung bzw. Unterstützung bei der Antragstellung
- Modellierung und Formalisierung von Forschungsfragen
- Datenmodellierung für die Projektdurchführung
- Auswahl und Anpassung geeigneter Formate und Standards
- Bereitstellung der technischen Infrastruktur
- Bei post mortem-Übernahme ggf. Rekonstruktion und / oder Re-Engineering der Projektssysteme
- Datenkuratierung mit Qualitätssicherung
- Ggf. Datenanreicherung
- Strukturierte Erschließung der Inhalte mittels Metadaten
- Vergabe von PIDs; Sicherung der Zitierfähigkeit
- Betrieb der technischen Schnittstellen
- Projekt- und Ressourcenbeschreibung,

Dokumentation

- Dauerbetrieb der Speicherung / Datenvorhaltung
- Dauerbetrieb der Präsentation bzw. Publikation
- Aktualisierung und ggf. Relaunch von Präsentations- und Publikationssystemen; disaster recovery bei Systemen, die durch veränderte Rahmenbedingungen (Updates etc.) nicht mehr funktionieren
- Rechtemanagement; Zugriffskontrolle
- Ggf. Migration von Daten oder Systemen in generische Strukturen; Homogenisierung
- Technische Aufgaben der Langzeitarchivierung (bitstream preservation, technology watch, preservation planning)
- Auswahlkriterien und Kassationsprinzipien vor, während und nach dem Ingestprozess
- Veränderungen im Servicelevel im Laufe der Zeit (z.B. Aufgabe des Hostings und Beschränkung auf Datenarchivierung)

Zu den Qualitätsmerkmalen von Datenzentren gehören die Standardisierung und Zertifizierung¹⁵ sowohl ihrer technischen Infrastrukturen als auch ihres Leistungsspektrums sowie die Zusammenarbeit mit anderen, generisch ausgerichteten Infrastruktureinrichtungen (z.B. Rechenzentren, Bibliotheken etc.). Zu prüfen ist, ob und

¹⁵ Relevante Zertifikate sind hier u.a. das Data Seal of Approval, siehe ‚DSA | Homepage‘, das nestor-Siegel für vertrauenswürdige digitale Langzeitarchive, siehe ‚nestor | nestor-Siegel für vertrauenswürdige Langzeitarchive‘, ISO 16363, siehe ‚ISO | ISO 16363 - Audit and certification of trustworthy digital repositories‘ sowie ‚Consultative Committee for Space Data Systems (CCSDS) (2011)‘.

welche Dienste und Aufgaben zwischen den Datenzentren arbeitsteilig bewältigt werden können und wo Forschung und Entwicklung abgestimmt voranzutreiben sind.

5. Organisationsformen

Zuschnitt und Zuständigkeit von Datenzentren können höchst unterschiedlich ausfallen. So können hinsichtlich der Fachspezifik entweder die Geisteswissenschaften insgesamt, Teilbereiche der Forschung oder einzelne Fächer adressiert werden. Neben nicht weiter spezifizierten „Humanities Data Centers“ existieren z.B. bereits Datenzentren für die Sprachwissenschaften oder für die Archäologie. Diese haben dabei einen lokalen, regionalen oder überregionalen Auftrag. Eine mögliche Ausprägung ist, dass in räumlicher oder institutioneller Hinsicht lokal organisierte Institutionen eine hohe Bandbreite an unterschiedlichen Disziplinen, Aufgaben und Datentypen repräsentieren müssen, während übergreifende, regionale oder nationale Institutionen sich stärker nach Fächern oder Technologien und Methoden spezialisieren können.

Mit einem „Datenzentrum“ kann aber auch vor allem ein konzeptioneller Begriff gemeint sein, der im Folgenden funktional definiert ist. Er beschreibt dann ein Set an Zielen und Aufgaben, die im Einzelfall sehr unterschiedlich umgesetzt werden können. Ein Datenzentrum kann deshalb sowohl von einer Einzelinstitution als auch

in Kooperation zwischen verschiedenen Beteiligten betrieben werden, wobei Begriffe wie „Forschungszentrum“, „Rechenzentrum“, „Bibliothek“ oder „Archiv“ zunächst Rollen bezeichnen, sehr wohl aber auch institutionelle Partner meinen können. Sowohl Fachspezifik als auch Regionalität sind Faktoren für die notwendige „Nähe zur Forschung“, weil diese sowohl die fachliche Kompetenz als auch die räumliche Nähe betreffen.

Außerdem ist zu beachten, dass der Begriff Geisteswissenschaften eine Vielzahl von Einzeldisziplinen (z.B. Geschichtswissenschaft, Musik- und Theaterwissenschaften, Kulturwissenschaften, Kunstgeschichte, Archäologie, Ethnologie, Sprach- und Literaturwissenschaften und viele weitere geistes- und kulturwissenschaftlich forschende Disziplinen) umfasst, die methodisch und hinsichtlich ihrer Fragestellungen und Erkenntnisinteressen divergent sind. Trotz vieler Gemeinsamkeiten weisen sie Unterschiede zueinander auf, die sich auf die Ausgestaltung der Datenzentren auswirken können. Der Begriff „fachspezifisch“, der im Kontext digitaler Infrastrukturen eine Rolle spielt, hat daher eine doppelte Bedeutung. Einerseits steht er für die Geisteswissenschaften insgesamt als Abgrenzungen zu anderen Disziplinen (Sozial-, Rechts-, Ingenieurs-, Wirtschafts- und Naturwissenschaften), andererseits werden dadurch auch die unterschiedlichen Anforderungen und Bedürfnisse einzelner Fächer innerhalb der Geisteswissenschaften beschrieben. Im

Folgenden werden die verschiedenen idealtypischen Formen geisteswissenschaftlicher Datenzentren grob charakterisiert.

1. Datenzentren innerhalb von Institutionen

Viele der derzeitigen geisteswissenschaftlichen Datenzentren entstehen innerhalb von großen Institutionen, vor allem an Universitäten, Hochschulen, Bibliotheken und Akademien, oder werden von diesen betrieben. Zum Teil sind diese institutionellen Datenzentren auch nationalen Verbänden angeschlossen oder arbeiten mit diesen zusammen (siehe den Unterpunkt 3). Die Datenzentren innerhalb von Institutionen verfolgen das Ziel, alle Herausforderungen und Lösungen zu den digitalen Geisteswissenschaften im Kontext einer Einrichtung zu bündeln, um im Idealfall umfassende IT-Unterstützung anzubieten. Deshalb wird dort ein Ansatz verfolgt, der sowohl eine Vielzahl von Einzeldisziplinen adressiert als auch alle Phasen des digitalen Forschungsdatenlebenszyklus. Dies beinhaltet in der Regel eine enge Begleitung von Forschungsprojekten und FachwissenschaftlerInnen von der ersten Antragsidee bis hin zur dauerhaften Übernahme der Daten nach Projektende. Diese Betreuung des gesamten Forschungszyklus „aus einer Hand“ lässt sich lokal vergleichsweise gut organisieren, insbesondere weil die räumliche Nähe oder die institutionelle Zusammengehörigkeit eine direkte Verbindung, kurze Wege und persönliche Kontakte ermöglichen. Auch können meistens rechtliche und finanzielle

Aspekte des Datenmanagements relativ einfach geklärt werden, weil alle Akteure zu einer einzigen Institution gehören. Mit dem projektspezifischen Wissen der ForscherInnen und den technischen Erfahrungen der DatenkuratorInnen lassen sich gemeinsam Datenmanagement-Strategien entwickeln, die auf die Möglichkeiten und Grenzen der konkreten Institution zugeschnitten sind und bei denen – je nach personeller und technischer Ausstattung eines Zentrums – auch individuelle Anforderungen berücksichtigt werden können.

Da in einem institutionellen Datenzentrum in der Regel eine Vielzahl geisteswissenschaftlicher Einzeldisziplinen betreut und beraten werden, kann es allerdings eine Herausforderung sein, in allen Bereichen die notwendige Expertise aufzubauen, um adäquat und langfristig mit den sehr heterogenen technischen wie fachlichen Spezifika der vielfältigen Daten umgehen zu können. In der Zusammenarbeit verschiedener Datenzentren kann hier eine entsprechende Expertise trotzdem angeboten werden. Ferner kann der Auftrag, eine digitale Unterstützung für alle geisteswissenschaftlichen Fächer anzubieten, dazu führen, dass Angebote und Lösungen eher generisch ausgerichtet sind und fachspezifische Anforderungen aus Einzeldisziplinen in geringerem Maße berücksichtigt werden können. Aber auch und gerade in der eventuellen Weiterverweisung an andere, z.B. fachspezifische Einrichtungen sind die lokalen Datenzentren grundsätzlich Teil einer übergreifenden Infrastruktur. Beispie-

le für bereits bestehende Datenzentren an Forschungseinrichtungen sind gwin (Hamburg) oder das Data Center for the Humanities (DCH, Köln)¹⁶. Zu den Einrichtungen, die über Abteilungen verfügen, die de facto die Aufgaben von Datenzentren übernehmen, gehören u.a. die Standorte des Austrian Centre for Digital Humanities in Graz und Wien, die Berlin-Brandenburgische Akademie der Wissenschaften oder die Herzog August Bibliothek in Wolfenbüttel.

2. Datenzentren mit regionaler oder fachwissenschaftlicher Ausrichtung

Neben Datenzentren, die eher institutionell ausgerichtet sind, existieren Datenzentren mit einer regionalen Ausrichtung, die dadurch charakterisiert sind, dass sich mehrere Akteure einer Stadt oder einer Region zusammenschließen, um deren Expertisen (z.B. aus dem Bereich der Rechenzentren, Bibliotheken, Softwareentwicklung, Archive und Forschungsinstitute) zu bündeln und gemeinsam Lösungen und Angebote für digital arbeitende und forschende Geisteswissenschaften (Digital Humanities) zu entwickeln. Zum Teil sind diese regionalen Datenzentren nationalen Verbänden angeschlossen oder arbeiten mit diesen zusammen (siehe den folgenden Unterpunkt 3).

Auch im Fall der regionalen Datenzentren können die ForscherInnen die Vorteile der räumlichen Nähe und einer ganzheitlichen Betreuung nutzen, auch wenn diese Verbände häufig nicht ausschließlich für die Region agieren, sondern ihre Dienste auch für ein größeres Umfeld (insbesondere das jeweilige Bundesland) anbieten. Die Vorteile eines solchen Modells liegen einerseits in den Synergieeffekten sowie in der größeren personellen Ausstattung, andererseits aber auch in der Tatsache, dass sich mehrere Institutionen an den Kosten eines Datenzentrums beteiligen und damit ein dauerhafter Betrieb einfacher zu gewährleisten ist. Oftmals ist zu beobachten, dass derartige regionale Datenzentren sich in thematischer oder technischer Hinsicht spezialisieren und dadurch weiterhin eine überregionale Bedeutung erlangen.

Identifizieren kann man etwa übergreifende fachwissenschaftliche Zentren, die gleichzeitig nationale Organisationsstrukturen für Fachcommunities aufbauen. Hierzu ist etwa das Forschungsdatenzentrum IANUS zu zählen, das als nationale fachspezifische Einrichtung Angebote für alle Archäologien und Altertumswissenschaften in Deutschland entwickelt und vom Deutschen Archäologischen Institut in Berlin für diese Fachcommunity koordiniert wird. In diesem Sinne wirken solche Einrichtungen wie die in Unterpunkt 3 aufgeführten nationalen Infrastrukturen.

Teilweise adressieren solche Datenzentren auch nur bestimmte Aspekte

¹⁶ Eine Aufstellung der genannten Einrichtungen findet sich unter „Mitglieder der DHD AG Datenzentren“ auf S. 28.

des Forschungsdatenlebenszyklus, wie beispielsweise die Unterstützung bei der Datengenerierung, bei der Datenkuration, der digitalen Langzeitarchivierung oder die Online-Verfügbarkeit von Daten. Eine Expertise in spezifisch fachwissenschaftlichen kontrollierten Vokabularien, nationalen Standards und Analysemethoden baut etwa das Historische Datenzentrum Sachsen-Anhalt (Halle) auf.

Fachwissenschaftliche Akteure bilden ein Bindeglied zwischen den Institutionen, übernehmen fachspezifische Aspekte des Datenmanagements, unterstützen institutionelle/regionale Datenzentren und arbeiten eng mit den nationalen Infrastrukturen zusammen.

3. Datenzentren als Teil einer übergreifenden Infrastruktur mit (inter-)nationaler Perspektive

Ein weiterer Typus von Datenzentren ist derjenige, der versucht, Angebote auf nationaler Ebene bzw. in einem internationalen Kontext für eine Teilmenge der Digital Humanities zu etablieren. In Deutschland und im europäischen Forschungsraum sind hier insbesondere zwei digitale Forschungsinfrastrukturen zu nennen, die neben weiteren Angeboten auch Aufgaben von Datenzentren übernehmen. Dies sind die nationalen Forschungsinfrastrukturen von CLARIN (Common Language Resources and Technology Infrastructure, z.B. CLARIN-D) mit einem Fokus auf Sprachressourcen aus den Geistes- und

Sozialwissenschaften, und DARIAH (Digital Research Infrastructure for the Arts and Humanities, z.B. DARIAH-DE) mit einem Schwerpunkt auf digitalen Ressourcen für Geistes- und Kulturwissenschaften. In der Schweiz hat das Data and Service Center for the Humanities (DaSCH) seinen Betrieb als disziplinär ausgerichtete, nationale Einrichtung aufgenommen, die neben Beratungsleistungen auch technische Infrastruktur für die langfristige Übernahme von Forschungsdaten anbietet und aktiv an der Standardisierung von Modellen und Methoden sowie der internationalen Vernetzung arbeitet.

Nationale Initiativen sind örtlich nicht immer in der Nähe der DatenproduzentInnen und DatenkuratorInnenen, was dazu führt, dass die Stärken der lokalen Zentren nicht immer auch für diese gelten. Dafür gibt es stärker ausgeprägte organisatorische und technische Spezialisierungen.

Durch ihr gebündeltes jeweiliges Know-How können sie die institutionellen, fachwissenschaftlichen und regionalen Datenzentren in ihrer Arbeit beraten und entweder generische oder (fach-)spezifische Anforderungen arbeitsteilig übernehmen. Für Akteure, die keinen Zugang zu institutionellen oder regionalen Infrastrukturen haben oder zu klein sind, um eigene dauerhafte IT-Infrastrukturen aufzubauen (auch im Bereich der außeruniversitären Forschung), können nationale Zentren eine wichtige Unterstützung darstellen. Außerdem können sie übergreifende zentrale

Aufgaben und Services wahrnehmen, etwa bei der Entwicklung von Standards und Best-Practices. Solche Standards können zu größerer Interoperabilität führen, etwa bei der Zusammenarbeit mit internationalen, vergleichbaren Einrichtungen oder bei der virtuellen Zusammenführung von Informationen und Ressourcen, die auf unterschiedliche Datenzentren verteilt sind, z.B. im Sinne eines zentralen Nachweiskataloges. Aus Sicht der WissenschaftlerInnen ergeben sich so neue Möglichkeiten zur Nachnutzung und Analyse quer über alle Daten. Allerdings stellen bei nationalen Zentren insbesondere rechtliche, finanzielle und organisatorische Fragen sowie die Beteiligung der Fachcommunities und die Verzahnung mit deren Anforderungen größere Herausforderungen dar.

Zusammenwirken und vernetzte Datenzentren im DHd-Verband

Die Organisationsmodelle der geisteswissenschaftlichen Datenzentren weisen zum jetzigen Zeitpunkt auf nationaler Ebene verschiedene Ansätze auf. Erstens betreiben einzelne Institutionen ein eigenes Datenzentrum, das meist viele Einzeldisziplinen und alle Phasen des Forschungsdatenlebenszyklus adressiert. Zweitens betreiben mehrere Institutionen gemeinsam ein lokal bzw. regional ausgerichtetes Datenzentrum. Diese spezialisieren sich oftmals in thematischer und technischer Hinsicht. Drittens haben sich mehrere

lokale Datenzentren in einem Verbund zusammengeschlossen, in dem nur bestimmte technische Aufgaben gemeinsam gelöst werden und bei dem eine inhaltliche Spezialisierung zur Arbeitsteilung möglich ist. Viertens werden von einzelnen Institutionen gemeinsam unterschiedliche Aspekte des Forschungsdatenlebenszyklus abgedeckt und in einem Verbund bzw. durch Kooperation entsprechende Expertisen eingebracht. Und fünftens betreiben mehrere relevante Institutionen gemeinsam Angebote für spezifische Fachcommunities.

Angesichts der vorgestellten Organisationsformen stellt sich die Frage, welches Organisationsmodell langfristig am sinnvollsten ist. In einigen naturwissenschaftlichen Bereichen, z.B. in der Astrophysik und den Erd- und Klimawissenschaften, werden globale Netze als „World Data Centres“ aufgebaut, die zumindest die Frage aufwerfen, ob dies auch für die Geisteswissenschaften möglich und wünschenswert sei. Mit anderen Worten: Sollte es ein einziges, zentrales geisteswissenschaftliches Datenzentrum mit einem Standort geben, an dem alle Kompetenzen gebündelt sind, oder ist ein föderales System und somit ein der Forschungslandschaft näher stehender verteilter Zentrenverbund, der nur übergreifende Aufgaben zentral bündelt, zu präferieren?

Datenzentren entstehen aus lokalen, institutionellen oder fachlichen Bedürfnissen und disziplinären Besonderheiten heraus.

Aus Sicht der AG Datenzentren hat daher jede der oben skizzierten Organisationsformen eines geisteswissenschaftlichen Datenzentrums eine Daseinsberechtigung, da die einzelnen Datenzentren fachspezifische Aufgaben übernehmen und unterschiedliche Zuständigkeitsbereiche bedienen. Unter den bestehenden Möglichkeiten und Bedingungen besitzen sie jeweils ihre eigenen Stärken und Schwächen, ergänzen sich arbeitsteilig und erarbeiten Kooperationsmodelle. Im Sinne der Empfehlungen des Rats für Informationsinfrastrukturen, der sich für eine Schwerpunktbildung und Arbeitsteilung im Rahmen eines föderalen Systems mit stärkerer Koordination und Anschlussfähigkeit zu europäischen Entwicklungen ausgesprochen hat, sieht die AG Datenzentren im Verband DHD für Deutschland in diesem verteilten Modell eine gute Ausgangsbasis für die Weiterentwicklung der nationalen geisteswissenschaftlichen Forschungsdatenzentren – eben eine echte „Leistung aus Vielfalt“.¹⁷

Zentral für die zukünftige Landschaft der geisteswissenschaftlichen Datenzentren sind aus Sicht der AG Datenzentren daher vor allem drei Strategien:

- Die Etablierung und Verstetigung Erfolg versprechender, aber befristeter Projekte und deren Überführung in langfristige Infrastrukturen.

- Die Verstärkung der bereits bestehenden Koordination zwischen den einzelnen Akteuren, um Doppel- und Mehrfachentwicklungen zu vermeiden und verteilte, aber aufeinander abgestimmte Kompetenzen und Dienstleistungen aufzubauen.
- Fokussierung auf Interoperabilität und eine flexibel gestaltbare Vernetzung. Hierdurch werden wissenschaftliche Anforderungen abgedeckt, die sich aus fachübergreifenden Fragestellungen, dem Aufbrechen disziplinärer Strukturen im Zuge der Digitalisierung (wobei die Methodik und die verfügbaren Daten als Bindeglied jenseits überkommener Fachstrukturen fungieren), sowie die Nutzung von unterschiedlichen Daten- und Metadatenformaten und weiteren fachspezifischen Anforderungen ergeben.

Eine noch bessere Vernetzung des Ökosystems der geisteswissenschaftlichen Datenzentren kann Synergien hervorbringen, Leistungsspektren schärfen, einen Mehrwert an Erfahrung und Expertise schaffen und in seiner Gesamtheit zu einer nationalen Forschungsdateninfrastruktur der Geisteswissenschaften beitragen. Durch nationale Forschungsinfrastrukturen wie CLARIN-D und DARIAH-DE, aber auch durch die vielfältigen Kooperationsbeziehungen auf der Ebene der Einzeldisziplinen und lokalen Institutionen ist gleichzeitig der weitere Rahmen der europäischen ESFRI-Strategie im Blickwinkel und eine internationale Vernetzung angelegt. Die AG Datenzentren des DHD-Verbandes, in der sowohl lokale, regionale als auch nationale

¹⁷ Vergleiche auch ‚RfII – Rat für Informationsinfrastrukturen (2016)‘, insbes. S. 40-41.

Datenzentren organisiert sind, sieht es als eine ihrer Aufgaben, eine solche nationale wie globale Vernetzung weiter voranzubringen.

6. Entwicklungsbereiche und Empfehlungen

Für die Langzeitarchivierung digitaler Daten im Sinne einer bitstream preservation gibt es seit einigen Jahren Forschungsergebnisse, tragfähige Konzepte und aktive Netzwerke (nestor¹⁸) – auch wenn die Lösungen noch nicht den Reifegrad einfach buchbarer Dienstleistungen erreicht haben. Ebenso ist der Bereich der Repositorien schon so weit entwickelt, dass hier Best Practices, etablierte Vorbilder und eine ganze Reihe von Zertifizierungen existieren, die auch für geisteswissenschaftliche Datenzentren leitend sind.¹⁹ Die Herausforderung besteht nunmehr darin, die vorhandenen Ansätze und Lösungen in den konkreten rechtlichen, finanziellen, technischen und inhaltlichen Rahmenbedingungen eines Datenzentrums umzusetzen.

18 Siehe ‚nestor | Homepage‘.

19 Um hier nur einige zu nennen: die Liste akkreditierter Datenzentren des Rates für Sozial- und Wirtschaftsdaten, siehe ‚Rat SWD | Akkreditierte Datenzentren‘, INSPIRE als Beispiel für ein Datenzentrum aus den Naturwissenschaften, siehe ‚INSPIRE | Homepage‘, DataCite als PID-Provider, siehe ‚DataCite | Homepage‘ und Zertifikats-Checklisten für Infrastrukturen, siehe CRL und OCLC (2007) für TRAC (Trusted Repository Audit Checklist) und TDR (Trustworthy Digital Repository Checklist, <http://public.ccsds.org/publications/archive/652x0m1.pdf>). Für relevante Zertifikate siehe Anmerkung 15.

Dagegen scheint der Umgang mit „lebendigen“ Präsentationssystemen und das Problem einer kontinuierlichen Datenkuratierung noch nicht in gleicher Weise konzeptionell durchdrungen oder exemplarisch umgesetzt zu sein. Insbesondere in diesem Bereich gibt es verschiedene Felder, die aus Sicht von Datenzentren insgesamt oder hinsichtlich ihrer spezifischen Herausforderungen in den Geisteswissenschaften weiter erforscht und entwickelt werden müssen. Zu diesen Entwicklungsbereichen gehören insbesondere:

- Entwicklung von Workflows zur Übernahme von Ressourcen sowie umfassende Project/Resource Descriptions. Diese müssen dabei sowohl die strukturierte Aufnahme von Daten nach Abwicklung von Datenzentren durch externe Partner samt Übernahme der Metadaten, Zugangsrechten und Identifikatoren-Portierung umfassen, als auch die Entgegennahme von Daten von einem Datenzentrum beschreiben, z. B. am Ende eines Projekts.²⁰ Bestehende Lösungen und Verfahren sollten dabei vereinheitlicht und transparent gemacht werden.

20 Das DCH Köln arbeitet hierzu an einem allgemeinen Übernahme-Workflow in dessen Zentrum ein Fragebogen bzw. eine Anleitung zu tief reichenden Interviews steht und der zu einer umfassenden Dokumentation und Maßnahmenplanung führt. Das Clarin-D Datenzentrum in Tübingen verfolgt z. B. eine differenzierte Praxis: 1.) Übernahmeworkflows von fachspezifischen Repositorien in Kerninfrastrukturen der Bibliothek (unter Verlust bestimmter fachwissenschaftlichen Anforderungen, Services und Beschreibungsmächtigkeit). 2.) Workflows für den Ingest von externen Daten in das Archivsystem. Noch nicht ausreichend bearbeitete Bereiche sind insbesondere:

- Abgestimmtes Ressourcen-Monitoring bzw. Services-Monitoring als Katalysator für die Standardisierung von Daten und Diensten und für bessere Sichtbarkeit von Inhalten und Funktionen.
- Standards für Speicherinfrastrukturen und grundlegende technische Infrastrukturen: Allgemeine Repository-Systeme zur Aufnahme von Forschungsdaten und -projekten; generische Disseminationssysteme und Schnittstellen; RDF/LOD-Repräsentation von Projekthinhalten.
- Langzeitarchivierung im Verbund mehrerer Datenzentren und als etablierter Dienst.
- Zwischen den Datenzentren abgestimmte Datenanreicherungsdienste wie generische Annotationsdienste, Werkzeuge für die Metadatengenerierung und Persistent-Identifier-Dienste (PID) für verschiedene Ressourcenarten und Granularitäten.
- Entwicklung und Anwendung von Standards: Metadatenformate auf Objekt- oder Sammlungsebene; Normdaten und kontrollierte Vokabularien für die Anreicherung von Forschungsdaten. Definition von Daten- und Objektklassen als Grundlage für weitere Standardisierung.
- Vernetzung der Angebote durch Datenzentren und übergreifende Such- und Nachweissysteme.
- Administration, Lizenzen und Rechte: Rechtemanagement und Zugangsorganisation für Forschungsdaten; Kostenmodelle, Geschäftsmodelle, Finanzierungsmodelle.²¹
- Technische und konzeptionelle Forschung zur Lösung des Problems des Widerspruchs zwischen zur Zitiation stabilisierten Daten und im Forschungsprozess veränderlichen Daten, das sich insbesondere für dynamische Systeme stellt (Beispiel: TextGrid²²).
- Ausgleich zwischen Standardisierungsnotwendigkeit (Interoperabilität, Effizienz) und wissenschaftlicher Innovationsleistung.
- Etablierung von Open Access und Open Data (Transparenz) in den Geisteswissenschaften mit Niederschlag in der geisteswissenschaftlichen Forschungs(daten)kultur.
- Identifikation von Lücken und entsprechende Weiterentwicklung bestehender Angebote bzw. Neuaufbau von Ressourcen (insbesondere bezüglich der teilweise betreuungsintensiven komplexen und lebenden Systeme, die bislang unzureichend adressiert werden).

Die Entwicklung dieser Themenbereiche geschieht im Idealfall in Abstimmung der geisteswissenschaftlichen Datenzentren untereinander bzw. auf (inter-)nationaler Ebene, auch unter Einbeziehung anderer Fachbereiche und Disziplinen.²³

a) Übernahme von Daten nach Abwicklung eines Zentrums durch externe Institutionen (wegen unklarer Rechtsnachfolge wären unter Umständen Datenüberlassungsverträge nicht gültig, selbst wenn die Übernahme technisch geklärt wäre);

b) Generische, standardisierte, dokumentierte Workflows, die nicht nur an einem Datenzentrum funktionieren.

²¹ Siehe u.A. ‚AC | Homepage‘.

²² Siehe ‚TextGrid | Homepage‘.

²³ Relevante Netzwerke sind z.B. DINI, siehe ‚DINI | Homepage‘, nestor, siehe ‚nestor | Homepage‘, die Research Data Alliance, siehe ‚RDA | Homepage‘, die Allianzinitiative ‚Digitale

7. Zukunftsperspektiven

Auch dank der schon erwähnten Publikation „Leistung durch Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland“ des Rats für Informationsinfrastrukturen (RfII 2016) ist inzwischen eine Agenda für die koordinierte Einrichtung einer nationalen Forschungsdateninfrastruktur verfügbar, an denen sich die konkrete Umsetzung der oben genannten Aufgaben und Perspektiven orientieren kann. Die AG Datenzentren stimmt in vielen Punkten den dort aufgestellten fünf dringlichsten Handlungsfeldern für Wissenschaft und Politik und sowie den umfangreichen, an verschiedene Adressatengruppen gerichtete, Einzelempfehlungen zu, und sieht sich hinsichtlich der Forderung nach Vernetzung von regionalen kooperierenden Datenzentren bestätigt.²⁴

Eine Stellungnahme zur NFDI und dem RfII-Papier hat die DHd-AG in Abstimmung mit dem Vorstand des DHd-Vereins am 03. September 2017 veröffentlicht.²⁵

Information“, siehe ‚Schwerpunktinitiative „Digitale Information“ | Homepage‘ und Forschungsdaten.org, siehe ‚Forschungsdaten.org | Homepage‘.

²⁴ Die fünf dringlichsten Handlungsfelder laut ‚RfII – Rat für Informationsinfrastrukturen (2016)‘, S. 66: „Fördermechanismen anpassen“; „Effizienz und Koordination durch eine (verteilte) nationale Infrastruktur sicherstellen“; „Forschungsdatenkultur befördern“; „Übergreifendes Monitoring und Qualitätssicherung einrichten“, sowie „Personalentwicklung auf allen Ebenen betreiben“. Für die Einzelempfehlungen siehe ‚RfII – Rat für Informationsinfrastrukturen (2016)‘, S. 67-69

²⁵ Siehe ‚DHd, Verband Digital Humanities im deutschsprachigen Raum (DHd) | NFDI-Stellungnahme‘

Mitglieder der DHd AG Datenzentren

[Austrian Centre for Digital Humanities \(ACDH-ÖAW\)](#), Wien

Matej Ďurčo

[Berlin-Brandenburgische Akademie der Wissenschaften \(BBAW\)](#), Berlin

Alexander Czmiel, Gerald Neumann

[CLARIN-D-Zentrum Tübingen](#), Seminar für Sprachwissenschaft (Clarín-D), Tübingen

Erhard Hinrichs, Thorsten Trippel

[DARIAH-DE – Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften](#)

Peter Gietz (Tübingen)

Lisa Klaffki (Wolfenbüttel)

[Data and Service Center for the Humanities \(DaSCH\)](#), Basel

Lukas Rosenthaler

[Data Center for the Humanities \(DCH\)](#), Köln

Patrick Sahle, Jonathan Blumtritt, Patrick Helling

[eScience Center](#), Tübingen

Matthias Lang

[Herzog August Bibliothek \(HAB\)](#), Wolfenbüttel

Lisa Klaffki

[Hessische Forschungsdateninfrastrukturen \(HeFDI\)](#), Marburg

Stefan Schulte

[Historisches Datenzentrum Sachsen-Anhalt \(Hist-Data\)](#)

Halle-Wittenberg: Katrin Moeller / Forschungsinfrastrukturen Universitäts- und

Landesbibliothek Sachsen-Anhalt: Roberto Cozatl

[Humanities Data Centre \(HDC\)](#), Göttingen

Sven Bingert, Claudia Engelhardt, Daniel Kurzawe, Jörg Wettlaufer

[IANUS am Deutschen Archäologischen Institut](#), Berlin

Maurice Heinrich, Felix Schäfer

[Institut für Deutsche Sprache – Archiv für Gesprochenes Deutsch](#), Mannheim

Thomas Schmidt

[Projekt PARTHENOS](#), FH Potsdam

Ulrike Wuttke (stellv. Convenor)

[Schweizerische Akademie der Geistes- und Sozialwissenschaften \(SAGW\)](#), Bern

Beat Immenhauser

[Servicezentrum eSciences](#), Universität Trier

Marina Lemaire

[Universitäts- und Landesbibliothek](#), Darmstadt

Thomas Stäcker

[Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities \(ZIM-ACDH\)](#), Graz

Johannes Stigler

[Zentrum für nachhaltiges Forschungsdatenmanagement/Geisteswissenschaftliche Infrastruktur für Nachhaltigkeit \(gwin\)](#), Hamburg

Kai Wörner (Convenor)

Bibliographie (Literatur und Webressourcen)

4C (Collaboration to Clarify the Costs of Curation) | Homepage,
Online: <http://www.4cproject.eu>
(letzter Zugriff 21.05.2017).

Allianz der deutschen Wissenschaftsorganisationen, Schwerpunktinitiative „Digitale Information“ | Homepage,
Online: <http://www.allianzinitiative.de/de>
(letzter Zugriff 21.05.2017).

BMBF (2013), Roadmap für Forschungsinfrastrukturen: Pilotprojekt des BMBF,
Online: https://www.bmbf.de/pub/Roadmap_Forschungsinfrastrukturen.pdf
(letzter Zugriff 21.05.2017).

CLARIN-D | Homepage,
Online: <https://www.clarin-d.de/de>
(letzter Zugriff 21.05.2017).

Consultative Committee for Space Data Systems (CCSDS) (2011). Recommendations for Space Data Systems Practices: Audit and certification of trustworthy digital repositories. Recommended Practice, CCSDS 652.0-M-1,
Online: <https://public.ccsds.org/pubs/652x0m1.pdf>
(letzter Zugriff 21.05.2017).

CRL und OCLC (2007), Trusted Repositories Audit & Certification: Criteria and Checklist [TRACK], Version 1.0,
Online: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
(letzter Zugriff 21.05.2017).

DataCite | Homepage,
Online: <https://www.datacite.org>
(letzter Zugriff 21.05.2017).

Data Seal of Approval (DSA) | Homepage,
Online: <http://datasealofapproval.org>
(letzter Zugriff 21.05.2017).

DARIAH-DE | Homepage,
Online: <https://de.dariah.eu>
(letzter Zugriff 21.05.2017).

Deutsche Forschungsgemeinschaft (DFG) (1998, 2013), Sicherung Guter Wissenschaftlicher Praxis (Safeguarding Good Scientific Practice), Denkschrift, Ergänzte Auflage,
Online: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf
(letzter Zugriff 21.05.2017).

Deutsche Forschungsgemeinschaft (DFG) (2015), Handreichung: Informationen zu rechtlichen Aspekten bei der Handhabung von Sprachkorpora,
Online: http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_recht.pdf
(letzter Zugriff 21.05.2017).

Deutsche Forschungsgemeinschaft (DFG), DFG | Umgang mit Forschungsdaten,
Online: http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/antragstellung/nachnutzung_forschungsdaten/index.html
(letzter Zugriff 21.01.2018).

Deutsche Initiative für Netzwerkinformation e.V. (DINI) | Homepage,
Online: <https://dini.de>
(letzter Zugriff 21.05.2017).

Deutsche Initiative für Netzwerkinformation e.V. (DINI) (2017), DINI-Stellungnahme zu den Empfehlungen des Rates für Informationsinfrastrukturen „Leistung aus Vielfalt“,
Online: <https://dini.de/fileadmin/docs/DINI-Stellungnahme-RfII-2017.pdf>
(letzter Zugriff 21.05.2017).

DHd, Verband Digital Humanities im deutschsprachigen Raum (DHd) | Homepage,
Online: <https://dig-hum.de>
(letzter Zugriff 21.05.2017).

DHd, Verband Digital Humanities im deutschsprachigen Raum (DHd) | AG Datenzentren,
Online: <https://dig-hum.de/ag-datenzentren>
(letzter Zugriff 21.05.2017).

DHd, Verband Digital Humanities im deutschsprachigen Raum (DHd) | NFDI-Stellungnahme,
Online: <https://dig-hum.de/aktuelles/dhd-stellungnahme-zur-nfdi>
(letzter Zugriff: 08.12.2017)

European Commission (2016), H2020 Programme: Guidelines on FAIR Data Management
in Horizon 2020, Version 3.0, 26 July 2016,
Online: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
(letzter Zugriff 21.05.2017).

Fonds zur Förderung der wissenschaftlichen Forschung (FWF) | Open Access Policy,
Online: <https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/>
(letzter Zugriff 21.05.2017).

Fonds zur Förderung der wissenschaftlichen Forschung (FWF) | Open Research Data Pilot,
Online: <https://www.fwf.ac.at/de/news-presse/news/nachricht/nid/20160118-2166/>
(letzter Zugriff 21.01.2018).

FORCE11 | The FAIR Data Principles,
Online: <https://www.force11.org/group/fairgroup/fairprinciples>
(letzter Zugriff 21.05.2017).

Forschungsdaten.org | Homepage,
Online: <http://www.forschungsdaten.org>
(letzter Zugriff 21.05.2017).

International Organization for Standardization (ISO) | ISO 16363 – Audit and certification of
trustworthy digital repositories,
Online: <https://www.iso.org/standard/56510.html>
(letzter Zugriff 21.05.2017).

IANUS – Forschungsdatenzentrum Archäologie & Altertumswissenschaften | Homepage,
Online: <https://www.ianus-fdz.de>
(letzter Zugriff 21.05.2017).

INSPIRE (High Energy Physics Information System), INSPIRE | Homepage,
Online: <https://inspirehep.net>
(letzter Zugriff 21.05.2017).

JISC | The significant Properties of Digital Objects, JISC Workshop 2008,
Online: <https://www.webarchive.org.uk/wayback/archive/20140615130716/http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops.aspx>
(letzter Zugriff 21.05.2017).

Klimpel und Weitzmann (2015), Paul Klimpel, John H. Weitzmann, Forschen in der digitalen Welt: Juristische Handreichung für die Geisteswissenschaften, DARIAH-DE Working Papers, 12,
Online: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-5-0>
(letzter Zugriff 21.05.2017).

NARA 2009, Significant Properties,
Online: <https://www.archives.gov/files/era/acera/pdf/significant-properties.pdf>
(letzter Zugriff 21.05.2017).

nestor | Homepage,
Online: <http://www.langzeitarchivierung.de/Subsites/nestor>
(letzter Zugriff 21.01.2018).

nestor | nestor-Siegel für vertrauenswürdige Langzeitarchive,
Online: http://www.langzeitarchivierung.de/Subsites/nestor/DE/Siegel/siegel_node.html
(letzter Zugriff 21.05.2017).

OAster | Homepage,
Online: <http://www.oclc.org/en/oaister.html>
(letzter Zugriff 21.05.2017).

Rat für Sozial- und Wirtschaftsdaten (RatSWD) | Akkreditierte Datenzentren,
Online: <https://www.ratswd.de/forschungsdaten/fdz>
(letzter Zugriff 21.05.2017).

Research Data Alliance (RDA) | Homepage,
Online: <https://www.rd-alliance.org>
(letzter Zugriff 21.05.2017).

Rfil – Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Online: <http://www.rfii.de/?wpdmdl=1998>
(letzter Zugriff 21.05.2017).

Rfil – Rat für Informationsinfrastrukturen (2017), Datenschutz und Forschungsdaten. Aktuelle Empfehlungen, Online: <http://www.rfii.de/?wpdmdl=2249>
(letzter Zugriff 22.05.2017).

Schweizer Nationalfonds (SNF), SNF | Open Research Data,
Online: http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx
(letzter Zugriff 21.05.2017).

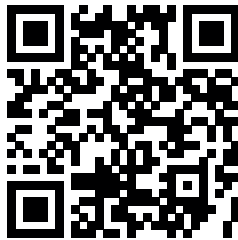
TEI | Homepage,
Online: <http://www.tei-c.org/index.xml>
(letzter Zugriff 21.05.2017).

TextGrid | Homepage,
Online: <https://textgrid.de>
(letzter Zugriff 21.05.2017).



digital humanities im
deutschsprachigen raum

<https://dig-hum.de/>



<http://dx.doi.org/10.5281/zenodo.1134760>



<https://creativecommons.org/licenses/by/4.0/deed.de>