

Tagungsbericht

DHd2017

13.-18. Februar 2017 in Bern

Melanie Janßen
Laura Rothfritz
Dorothea Wenz

31. März 2017

Exkursionsleiterin: Prof. Dr. Heike Neuroth

1 Einleitung

Die 4. Jahrestagung der DHd (Digital Humanities im deutschsprachigen Raum) fand vom 13. - 18. Februar an der Universität Bern statt. Der DHd ist ein 2013 gegründeter Verband mit bereits über 300 Mitgliedern die sich im Bereich Digital Humanities in der Forschung und Lehre engagieren¹.

Dabei umfasst der Begriff *Humanities* im Deutschen die Geistes- und Kulturwissenschaften. Innerhalb der digital humanities werden Methoden entwickelt, mit denen computergestützt geisteswissenschaftliche Quellengrundlagen analysiert und aufgearbeitet werden können. Dies führt dazu, dass sich a) neue Arbeitsweisen innerhalb der Geisteswissenschaften entwickeln (kollaboratives Arbeiten), b) große Mengen an Quellen schneller bearbeitet werden können (Quantifizierung/Empirisierung der geisteswissenschaftlichen Forschung, *distant reading*) und c) darauf aufbauend neue Forschungsfragen entwickelt werden können.

Beispiele für Projekte aus den digitalen Geisteswissenschaften sind digitale Sammlungen, Editionen, Annotationen, sowie Visualisierungen oder Netzwerkanalysen von Korpora.

Die Konferenz stand unter dem Motto *Digitale Nachhaltigkeit*². In Workshops, Panels und Vorträgen ging es darum, Möglichkeiten zu diskutieren, wie „digitale Wissensgüter auf eine langfristig zugängliche und Ressourcenschonende Weise eingesetzt und weiterentwickelt werden“ können³. Nachhaltigkeit betrifft nicht nur die Daten als „Güter“ selbst, sondern auch ihren Kontext und Interpretierbarkeit zur Wissensgenerierung. Der Begriff *digitale Nachhaltigkeit* wird sowohl im technischen Sinn von Langzeitverfügbarkeit, als auch (beispielsweise in den Bewegungen open source und open data) im Sinn von gesicherter freier Zugänglichkeit von Ressourcen verwendet. (Stu14) Peter SEELE ordnete in seiner Eröffnungkeynote digitale Nachhaltigkeit in das Spektrum von Mittel und Zweck sowie Methodik und Theorie ein.

Er warf zum Auftakt der Konferenz die Fragen auf, was Theorien der Digitalisierung sein könnten, in wie weit sich die digital Humanities normativ positionieren können und wie die Digitalisierung sowohl *methodisch* als Mittel die Geisteswissenschaften fördern als auch *theoretisch* ihren Zweck fördern könnte. Im Zuge der Keynote wurde die Frage diskutiert, in wie fern eine Konzentration auf Nachhaltigkeit weg von einer Wissensproduktion hin zu einer reinen Wissensverwaltung führen könnte. Seele betonte, dass auch die digitalen Geisteswissenschaften intellektuell dialektisch blieben, sie aber in ihrer Umsetzung unbedingt nachhaltig sein müssten. Nur so könnten sie sich selbst verstetigen.

¹Siehe: <https://dig-hum.de/ueber-dhd>

²Eine Sammlung von Vortragsfolien befindet sich auf der Webseite der DHd: <http://www.dhd2017.ch/presentationen>

³Siehe Call for Papers: <http://www.dhd2017.ch/calls>

2 Nachhaltige Forschung

Forschungsdaten in den Geisteswissenschaften

Geisteswissenschaftliche Methoden begründen sich traditionell auf textuellen Grundlagen. Im Unterschied zu den Naturwissenschaften steht bei der ursprünglichen Analyse von (Quellen-)Literatur allerdings nicht darum, Hypothesen mit Hilfe von empirischen Studien möglichst ausreichend zu verifizieren. Vielmehr begründen sich die Geisteswissenschaften auf hermeneutischer Auswertung von Textmaterial.

Der Begriff „Daten“ als Forschungsgrundlage wurde lange direkt den Natur- bzw. Sozialwissenschaften zugeordnet. Jedoch lässt sich beobachten, dass durch die Digitalisierung der Geisteswissenschaften ebenfalls Daten im Sinne der Definition Buzettis („*Data is the representation of information in a form that can be processed by a machine.*“)(Buz09) Forschungsgrundlage und Forschungsergebnisse sind.

Peter Andorfer definiert geisteswissenschaftliche Forschungsdaten folgendermaßen: Unter Forschungsdaten werden „*Daten bzw. Dateien verstanden, die im Zuge des Forschungsprozesses gesammelt und erstellt wurden und die nicht in weitgehend unveränderter Form Teil der wissenschaftlichen Publikation sind.*“ (And15) ⁴

Niklaus STETTLER⁵ sprach in seinem Vortrag über Möglichkeiten, Forschungsdaten nachhaltig zu erhalten. Er diskutierte zwei Ansätze: Zum Einen soll das Bewusstsein für das Management von Forschungsdaten durch die Forschenden gestärkt werden. Wissensvermittlung, wie das Schweizer Projekt „Train2Dacar“⁶ kann hier Abhilfe schaffen, aber auch noch mehr Verunsicherung schüren. Die Einforderung von Datenmanagementplänen für öffentlich finanzierte Projekte führt zwar zu erhöhter Aufmerksamkeit auf den Umgang mit Daten, diese ist jedoch meist eher auf Qualitätsmanagement als auf wirkliche Nachhaltigkeit von Forschungsdaten fokussiert. Andererseits fehlt es auf Seiten der Infrastruktur an Disziplin übergreifenden Repositories. Ziel müsse es sein, eine zentrale Institution aufzubauen, die das Forschungsdatenmanagement für die Wissenschaftler niederschwelliger im Workflow zu verankert und die nachhaltige Nutzung der Daten ermöglicht.

⁴In der Interviewreihe handelte es sich hierbei um Exzerpte von Quellen und Sekundärliteratur, Transkripte von Quellen, Notizen und Anmerkungen zu konkreten Texten/Bildern/Objekten oder zu Ideen, Thesen, Konzepten, (kommentierte und/oder strukturierte) Bibliographien sowie (mehr oder weniger legale) digitale Reproduktionen von Fachliteratur und/oder Quellen, ggf. mit Anmerkungen versehen. (Ebd.)

⁵Schweizerisches Institut für Informationswissenschaft (SII) Chur

⁶http://campus.hesge.ch/id_bilingue/projekte/train2dacar/

Nachhaltigkeit durch Nachnutzbarkeit

Nachhaltig werden Forschungsdaten jedoch erst dadurch, dass sie sich in einen Kontext einbetten lassen und somit verständlich werden. Dies ist gerade für geisteswissenschaftliche (qualitative) Datensätze wichtig. Im Sinne der Nachnutzbarkeit bedeutet dies, dass die Daten auch für Menschen verständlich und interpretierbar bleiben. Peggy GROßE und Sarah WAGNER⁷ stellten eine Möglichkeit vor, wie mit Hilfe von CIDOC-CRM Forschungsdaten semantisch angereichert werden können und somit nicht nur langfristig nutzbar sondern auch interpretierbar bleiben. CIDOC-CRM ist eine ISO-zertifizierte Ontologie, mit der Informationen formal kategorisiert, definiert und ausgetauscht werden können. Relationen der Informationsobjekte können mit Hilfe von CIDOC-CRM so dargestellt werden, dass sie gleichzeitig maschinenlesbar und in natürlicher Sprache verständlich bleiben. Anwendung fand die Ontologie bei dem Projekt MUSICES bei dem in Zusammenarbeit mit dem Germanisches Nationalmuseum Nürnberg (GNM) und dem Entwicklungszentrum Röntgentechnik (EZRT) des Fraunhofer-Instituts für Integrierte Schaltungen (IIS) ein Standard für die dreidimensionale Röntgen-Computertomographie (3D-CT) von Musikinstrumenten entwickelt werden sollte. Die Forschungsdaten wurden prozessorientiert erfasst, mit CIDOC-CRM semantisch strukturiert und vernetzt und schließlich über die Projektwebsite veröffentlicht.

Im Workshop Nachhaltiges Metadatenmanagement mit XMP, exiftool und FotoStation gab Oliver POHL⁸ ein Beispiel für nachhaltig nachnutzbare Bildmetadaten mit Hilfe von XMP. XMP ist eine RDF/XML-basierte Auszeichnungssprache für Bildmetadaten, das von Adobe geschaffen wurde und zunächst für PDF Dateien benutzt wurde. Diese werden an die Bilddatei angeheftet, die Interpretierbarkeit der Originaldatei wird dabei jedoch nicht gefährdet. Metadaten und Bilddateien bilden dabei eine Einheit, jedoch werden die XMP-Daten von den meisten Programmen nicht interpretiert. Da XMP auf RDF/XML gründet, ist das Metadatenschema leicht erweiterbar und interoperabel. Im Workshop wurde den TeilnehmerInnen das Schreiben von Bildmetadaten in XMP mit Hilfe der Software exiftool und FotoStation näher gebracht und praktisch erprobt.

Nachhaltigkeit des Forschungsprozesses

Christof SCHÖCH⁹ referierte in seinem Vortrag über die Möglichkeiten die digitalen Geisteswissenschaften als wiederholende und wiederholbare Forschung zu definieren. Im Sinne der wiederholenden Forschung können in den digital Humanities Studien, die noch mit intellektuellen Mitteln durch-

⁷Beide: Germanisches Nationalmuseum Nürnberg

⁸Berlin-Brandenburgische Akademie der Wissenschaften

⁹Universität Würzburg

geführt wurden mit Hilfe digitaler Techniken reproduziert werden. Als Beispiel stellte er François Richeaudeaus Studie zur Satzlänge in den Werken des belgischen Autors Georges Simenon vor, die 1982 in einem Aufsatz veröffentlicht wurde. Richeaudeaus analysierte 25 Texte Simeons, in der Wiederholungsstudie wurden sowohl diese Textgrundlage verwendet in einem zweiten Schritt jedoch 102 zusätzliche Texte des Autors sowie 252 zeitgenössische Romane analysiert. Mit Hilfe computerlinguistischer Verfahren konnten weit mehr Daten ausgewertet werden als es Richeaudeau möglich gewesen wäre. Die Ergebnisse sprechen letztendlich eher gegen seine Thesen.

Wird Wiederholende Forschung in den digital Humanities betrieben, so schafft dies eine Möglichkeit, den Diskurs innerhalb der einzelnen Disziplinen, in denen geforscht wird, voranzutreiben. Doch auch für die digitale Geisteswissenschaft als eigene Disziplin schafft wiederholende Forschung eine Auseinandersetzung mit den eignen Methoden. Dabei ist besonders das Verhältnis zur Ursprungsstudie bezüglich der zugrundeliegenden Daten, Fragestellungen und Methoden wichtig. (Analoge) Originalstudien sind in den Geisteswissenschaften häufig nicht auf Wiederholbarkeit ausgelegt. In wiederholenden Studien können im Forschungsprozess neue Eigenheiten der Originalstudie, aber auch reflektierend Kriterien der digitalen Forschungsprozessmodellierung herausgearbeitet werden.

Nachhaltigkeit von Forschung in den digital Humanities lässt sich durch wiederholbare Forschung erwirken. Philipp STEINKRÜGER¹⁰ stellte Kriterien einer Evaluation von Nachhaltigkeit im *Review Journal für digitale Editionen RIDE* vor. Punkte aus dem RIDE Kriterienkatalog, die die Nachhaltigkeit von Editionsprojekten evaluieren sind neben der fertigen Modellierung der Edition und der dauerhaften Nutzbarkeit sowohl des Ergebnisses als auch der Daten, die Dokumentation der Ziele und Methoden um den Forschungsprozess nachvollziehbar und transparent zu machen. Bei 28% der evaluierten Editionen wurden die Methoden nicht expliziert dokumentiert. Eine Dokumentation des benutzten Datenmodells sowie eine Begründung für die Auswahl fehlte bei 40% der Editionen. Die größten Probleme ergab die Bewertung des Datenverfügbarkeit, wobei gerade Skripte für benutzte Software nicht zur Verfügung gestellt wurden. In der anschließenden Diskussion wurde die Frage aufgeworfen, ob die Nachhaltigkeit eines Forschungsprojektes sich immer immer auf das Endprodukt ausrichten sollte oder aber, ob nicht eine genaue Dokumentation und Offenlegung des Forschungsprozesses und der dabei benutzten Instrumente (in diesem Fall Software) einen beträchtlichen Anteil an nachhaltiger Forschung in den digitalen Geisteswissenschaften hat. Ein Kulturwandel in der geisteswissenschaftlichen Forschung, die auf transparente Prozesse und wiederholbare, nicht bloß nachnutzbare Ergebnisse ausgerichtet ist, könnte einen Kern der digital humanities bilden. Intrinsisches Interesse der Forschenden, Inhalte ausdauernd zu sichern

¹⁰Institut für Dokumentologie und Editorik

in Kombination mit Governance für langfristig nutzbare Forschung sowie mehr Fördermittel für Nachhaltigkeit in der Wissenschaft wurden in der Diskussion als Lösungsansätze genannt.

Zusammenfassung

Nachhaltigkeit als Motto der DHd 2017 spiegelte sich in den Vorträgen zu großen Teilen im Bezug auf Forschungsdaten, deren Management und eine Nachnutzungsmöglichkeit. Zudem wurden Möglichkeiten, den Forschungsprozess in den Geisteswissenschaften transparent und wiederholbar zu gestalten diskutiert. Auffällig war, dass nach der Vorstellung von Forschungsprojekten fast immer nach Nachnutzungsmöglichkeiten der Forschungsergebnisse, zum Beispiel durch Bereitstellung von API-Schnittstellen, gefragt wurde. Wie weit der Begriff „Forschungsdaten“ in den Geisteswissenschaften gefasst wird und dass auch Software und eine Dokumentation des Forschungsprozesses für die Nachvollziehbarkeit der Ergebnisse zur Verfügung gestellt werden sollte, tauchte in vielen Diskussionen auf. Informationsspezialisten können nachhaltige Forschung u.A. durch drei wesentliche Praktiken unterstützen: 1. Unterstützung der Forschenden bei dem Management von Forschungsdaten, 2. Langfristig interpretierbare und standardisierte Metadaten Vergabe durch technische Lösungen und 3. Mitwirkung bei Forschungsdokumentation, deren Verfügbarmachung sowie die langfristige Bereitstellung von Forschungssoftware.

3 Informationswissenschaften und DH

Kernkompetenzen der Informationswissenschaft

Im Rahmen von Workshop 2a „Daten sammeln, modellieren und durchsuchen mit DARIAH-DE“ bot Lisa KLAFKI¹¹ diverse Beispiele der Kernkompetenzen von Informationswissenschaftlern die bereits im Kontext der Digital Humanities angewandt werden. Darunter der Best Practice Vorschlag, Nutzungsrechte frühzeitig zu klären, eine möglichst offene Lizenz zu wählen und diese auch in den Metadaten transparent festzuhalten. Der drohende Kontextverlust der Daten soll durch ausführliche Dokumentation verhindert werden. Die Vielfalt veralteter Formate soll von offenen Archivformaten abgelöst werden.

Danach zählte Anna ASCHAUER¹² Probleme in den digitalen Geisteswissenschaften auf, die vielfach von Methoden aus den Informationswissenschaften profitieren können: Eine Übersicht über diverse verfügbare Quellen schaffen, heterogene Quellen in Verbunddatenbanken oder Metasuchmaschinen

¹¹Herzog August Bibliothek Wolfenbüttel

¹²Institut für Europäische Geschichte

zusammenführen, die Qualität von Digitalisaten überwachen und verbessern, isolierte Forschungsfelder durch einen Raum der Begegnung zusammenführen, sowie allgemein Kommunikation und Kooperation anregen. Im Hands-On Teil des DARIAH-DE Workshops fielen diverse Stichworte mit denen der moderne Informationswissenschaftler bereits umgeht: Dublin Core, Datenmapping, XML, Normdateien, Disambiguierung, facettierte Suche, MEDS/MODS, Precision und Recall. Lediglich die Programmiersprache Java fehlt im Potsdamer Curriculum. Ein Nachteil für solche Informationswissenschaftler die an der Schnittstelle von Informatik und Geisteswissenschaft wirken wollen.

In Panel 1 „Virtuelle Forschungsumgebung für objekt- und raumbezogene Forschung“ stellte Mark FICHTNER¹³ „WissKI“, die Wissenschaftliche Kommunikations-Infrastruktur vor. Hervorgegangen aus einem DFG Projekt aus dem Jahr 2009 ist es eine virtuelle Forschungsumgebung für den musealen Kontext. Das Suchinterface ist an OPACs und Wikis orientiert, ein Vorteil für Fachwissenschaftler und Informationswissenschaftler gleichzeitig, da die Gewohnheiten der Nutzer unterstützt werden.

Herr FICHTNER merkte auch an, dass semantische Datenmodellierung besonders im musealen Kontext wichtig ist, denn sie erlaubt das Knüpfen von Verbindungen, und die Darstellung von Netzwerkanalysen mit geringem Aufwand. Informationswissenschaftler finden hier also diverse Anknüpfungspunkte um mit Projekten dieser Art zu kooperieren.

Danach präsentierte WERNER KÖHLER¹⁴ mit dem „Corpus der barocken Deckenmalerei in Deutschland“ ein DFG Langzeit-Projekt, welches „WissKI“ als Arbeitsplattform nutzt. Besonders interessant war, dass die Ontologieentwicklung zusammen mit Konfiguration, Programmierung und Linked Data Schnittstellen dem Bereich „[Computer] Engineering“ zugeordnet war und nicht etwa den Informationswissenschaften. Herr KÖHLER erwähnte in einem Nebensatz, dass die Entwicklung der Ontologie für das Team aus Informatikern doch überraschend arbeits- und zeitaufwändig war.

Normen und Standards

Ebenfalls Teil von Panel 1 war das von Piotr KUROCZYŃSKI¹⁵ vorgestellte Leibniz-Projekt „Virtuelle Rekonstruktionen in transnationalen Forschungsumgebungen“, welches digitale 3D Rekonstruktionen von Schlössern und Denkmälern im ehemaligen Ostpreußen anfertigt. Es nutzt den W3C Standard RDF und Linked Data, sowie CIDOC CRM als „semantischen Klebstoff“. Die zugrundeliegenden Standards sollen die digitale Nachhaltigkeit sichern: Open Source Software, Creative Commons Inhalte, standardisiertes Datenmodell, Linked Data per Semantic Web und sind allesamt Deckungs-

¹³Germanisches Nationalmuseum

¹⁴Deutsches Dokumentationszentrum für Kunstgeschichte

¹⁵Herder-Institut für historische Ostmitteleuropaforschung

gleich mit den Kerninhalten der Informationswissenschaftlichen Studien. Im Resümee von Panel 1 wurde bedauert, dass in Forschungsprojekten meist keine Mittel für Akzeptanzerhöhung (Spaßfaktor, Communitybildung) oder Untersuchungen zum Informationsverhalten vorgesehen sind. Informationswissenschaftler könnten hier mit den bereits vorliegenden Fachkenntnissen eine Lücke schließen.

In den Vorträgen 3b „Bilder“ stellten Lisa DIECKMANN und Claes NEUEFEIND¹⁶ die automatische inhaltliche Erschließung und Annotation kunsthistorischer Daten aus der Prometheus Datenbank vor. In diesem digitalen Bildarchiv wurden Daten aus neunzig verschiedenen Datenbanken mit verschiedenen fachlichen Schwerpunkten aus Museen, Forschung und Lehre zusammengeführt. In der dreijährigen Laufzeit wurde das Projekt durch das BMBF¹⁷ gefördert und finanziert sich seitdem mit einem Lizenzmodell¹⁸.

Eine besondere Herausforderung war das Dublettenproblem und die Frage ob ein Objekt dasselbe oder nur das gleiche ist. Die Objekte sollte aus den vorliegenden Bildbeschreibungen (Fließtext) identifiziert werden. Dafür wurden Namen, Orte, etc. per Name Entity Recognition, und Datierungen per Temporal Expression Detection extrahiert, diese wurden in strukturierter Form den Metadaten beigegeben. Der automatischen Extraktion folgt eine manuellen Korrektur durch Crowdsourcing. Die Ergebnisse der maschinellen Verfahren werden hier als Vorschläge betrachtet und das Feedback der Prometheus Nutzer als Bewertung verstanden.

Diese Herausforderungen, die aus fehlender Normierung und Standardisierung auf Seiten der Geisteswissenschaft entstanden, könnten in Kooperation mit Informationswissenschaftlern für zukünftige Projekte, Datenbanken und Sammlungen vermieden werden.

Interoperabilität informationswissenschaftlicher Infrastruktur

Stefan SCHMUNK¹⁹ und Tobias GRADL²⁰ sprachen in Workshop 2a zum Thema Modellierung und Interoperabilität. Hier wurden Bibliotheken klar kritisiert: In einer Literaturwissenschaftlichen Rezeptionsanalyse sollten Volltexte analysiert werden, dafür wurden 180 Bibliotheksbestände durchsucht. Keine davon hat eine offene API²¹ um eine automatisierte Abfrage in einem realistischen Rahmen zu ermöglichen. Hier fehlt in den Informationswissenschaften scheinbar großflächig die Unterstützung moderner Standards um allen Nutzern adäquaten Zugriff zu gewähren.

In Workshop 8 stellte Canan HASTIK²² die virtuelle Forschungsumgebung

¹⁶Beide: Universität zu Köln

¹⁷Bundesministerium für Bildung und Forschung

¹⁸Prometheus Lizenzmodell: <http://prometheus-bildarchiv.de/de/license/index>

¹⁹Staats- und Universitätsbibliothek Göttingen

²⁰Universität Bamberg

²¹Application Programming Interface: <https://techterms.com/definition/api>

²²Technische Universität Darmstadt

TextGrid vor, die seit Ende der Projektförderung in 2015 ein Baustein der Forschungsinfrastruktur DARIAH-DE ist. Die Erstellung und Verwaltung von Annotationen kann besonders vom professionellen Umgang mit Standardisierung und Ontologien für die Qualitätssicherung profitieren. Zusätzlich erwähnte Frau HASTIK auch, dass es ihres Wissens nach bisher keine Linked Open Data Sammlungen aus den Geisteswissenschaften in Produktion befindlich sind, ein Umstand der von der Zusammenarbeit mit Informationswissenschaftlern und deren Kenntnissen zum Semantic Web profitieren kann.

Informationsvisualisierung

Evelyn GIUS²³ stellte in den Vorträgen 1a „Visualisierung“ das Projekt 3DH vor.²⁴

Ein Zwischenergebnis des Projektes sind grafische Aktivatoren für bestimmte Konzepte, wie spezifische Formen, Größen, Farbwerte oder Texturen für qualitative oder quantitative Unterschiede. Das nächste Ziel ist die Entwicklung einer Visualisierungs-Grammatik, welches ein interessanter Schnittpunkt mit Informationswissenschaftlern sein kann. Dazu passend bietet Marian DÖRK im neu akkreditierten Studiengang Informationswissenschaften an der FH Potsdam ein Wahlpflichtfach „Informationsvisualisierung“ an.

In den Vorträgen 2a „Visualisierung“ stellten Ingrid MATSCHINEGG und Isabelle NICKA²⁵ den Relaunch der 2002 konzipierten Bilddatenbank „REA-Online“ vor. Im Neubau fanden sich für Informationswissenschaftler relevante Aufgaben, wie der Aufbau eines Thesaurus im Backend, sowie die Semantic Web Vernetzung und Konzepte zur Darstellung der Metadaten im Frontend. Die idealerweise Informationswissenschaftlichen Arbeiten wurden von Geisteswissenschaftlern und Informatikern erledigt. Der überarbeitete Thesaurus ist Institutionseigen und nicht standardisiert, Iconclass und GND Normdaten sollen aber noch eingefügt werden. Ebenso auf der To Do Liste stehen: erweiterte Taggingfunktionalitäten, Informationsvisualisierung, ein Ausbau der Bezugsinformationen und bessere Schnittstellen zu anderen Datenbanken und dem Semantic Web. Ebenso soll die automatisierte Erkennung der Bildinhalte ermöglicht werden.

Florian WINDHAGER²⁶ referierte folgend zur „polykubistischen Informationsvisualisierung von Biographiedaten“, deren Grundgedanke es ist kulturelle Sammlungen und ihre Metadaten in einer kartenbasierten Umgebung zur visuellen Analyse aufzubereiten. Mit Hilfe von Chronogeographie können Bewegungen von Personen und Gruppen über einen Zeitraum dynamisch

²³Universität Hamburg

²⁴Three-Dimensional Dynamic Data Visualisation and Exploration for Digital Humanities Research, <http://threedh.net/3dh/>

²⁵Beide: Universität Salzburg

²⁶Donau-Universität Krems

dargestellt werden.

Neben der Informationsvisualisierung an sich ist hier fachlich auch das Textmining interessant, welches die Datengrundlage liefert. Hier werden Entitäten im Text (Personen, Orte, Zeitstempel) verknüpft und in Korpora gesammelt.

Auch hier zeigte sich wie sehr Informationsvisualisierung zum Verständnis von fachwissenschaftlichen Daten beitragen kann und wie viel Mehrwert dies für die Digitalen Geisteswissenschaftler bedeutet.

Kollaboration und Community

Stefan SCHMUNK sprach in Workshop 8 zur Langzeitaufbewahrung von digitalen Forschungsdaten und merkte an, dass Digital Humanities Wissenschaftler ihre Daten weiterhin oft auf Privatrechnern aufbewahren, statt ihre zugehörigen Rechenzentren und Bibliotheken damit zu betrauen. Dem soll die Webapplikation DARIAH Publikator entgegenwirken. Die Geisteswissenschaftler sollen ihr Repository kontinuierlich mit Daten füttern und zum Projektende dann mit einem Knopfdruck live schalten.

Auf die Frage nach der Kompatibilität von Werkzeugen im Beta Status und Nachhaltigkeit antwortete Dr. SCHMUNK mit Nachdruck, dass im DH Bereich keine Produkte generiert werden, sondern Projektergebnisse und Know How. Die Verstetigung kann nur durch eine Community und deren Nutzung der Tools geschehen. Nach Projektende leben Werkzeuge in der Regel von Elan, Interesse und Eigenressourcen.

Nachhaltigkeit ist Communityabhängig und Feedback der Nutzer ist zentral um Geldgebern die Intention der Masse beweisen, nicht anders als in Bibliotheken und Archiven also. Er fügte hinzu, dass besonders Langzeitprojekte Zeit und Etat für die Entwicklung von Tools schon im Projektantrag beachten sollten. Dem folgend sollten auch wissenschaftliche Bibliotheken und Institutszugehörige Archive weiterhin für Forschungsdatenmanagement werben und die Infrastruktur dahingehend ausbauen.

Andrea SCHILZ²⁷ referierte im Rahmen der Vorträge 4c „Kollaboration“ zum Crowdsourcing.

Vorgestellt wurden diverse Projekte die unter den Begriff der Bürgerforschung oder Citizen Science fallen und auf dem partizipativen Bildungsgedanken basieren, unter anderem das DFG-geförderte Projekt „Artigo“²⁸, welches Kunstwerke über ein „Tabu“-artiges Spiel verschlagworten lässt und dadurch akkurate Schlagworte generiert, die aber nicht offensichtlich sind. Oder „DigiTalkoot“, ein Projekt der finnischen Nationalbibliothek, dessen OCR-Korrekturen signifikant zur Erstellung der „Historical Newspaper Library“beitrugen²⁹.

²⁷Universität Passau

²⁸<http://www.artigo.org/about.html>

²⁹<http://www.digitalkoot.fi>

In der Regel werden durch die Crowd Annotationen, OCR Korrekturen oder sogar kontrollierte Vokabulare generiert, der Anreiz wird oft über „Gamification“, also Elemente aus der Videospielebranche geboten.

Crowdsourcing wird bereits in Bibliotheken und Archiven durchgeführt und nutzt diese “Games with a Purpose” als didaktisches Medium und um ihre Angebote zu verbessern.

Friederike FLESS³⁰ beschrieb in ihrer Keynote die Archäologie als Erforschung nachhaltiger Dinge, die aber in den Methoden nicht nachhaltig und nicht reproduzierbar ist, denn Ausgraben ist ein irreversibler Vorgang. Die Dokumentation (heutzutage meist 3D-Umgebungsscans) ist zentral, denn ohne diese Daten ist der Vorgang der Grabung verloren.

Das DFG-Projekt IANUS³¹ soll ein Forschungsdatenzentrum für die Archäologie werden und Daten nachhaltig sichern. Auf Nachfrage nach der Einordnung von IANUS in die Problematik der Nachhaltigkeit beschrieb sie das Forschungsdatenzentrum als erweiterungsbedürftige Teillösung für das spezifische Problem unstrukturierter Datenbestände in dauerhaften Strukturen. Darum herum muss eine Community aus Forschern, Museen und Lehre entstehen, vor allem mit starker globaler Vernetzung. Nachhaltige und nachgefragte Strukturen sind nötig, die für Nutzer konzipiert und beworben werden. Denn jede Disziplin der Geisteswissenschaften hat spezifische Probleme mit der Nachhaltigkeit, aber gleichzeitig gibt es große Überschneidungen die gemeinsam gelöst werden können.

Rollen und Ziele des Forschungsdatenmanagements sind in den Geisteswissenschaftlichen Disziplinen nahezu identisch und spielen eine zentrale Rolle in der Forschungspolitik, werden aber von einer Tendenz zu Einzellösungen und Umsetzungsproblemen von Archivierungsstandards wie OAIS zurückgehalten.

Die Problemstellen sind laut Frau FLESS vor allem Arbeitspotential („Manpower“), IT Infrastruktur und nachhaltige Systeme. Neben enger Kooperation mit Rechenzentren können Informationswissenschaftler sowohl als Datenmanager oder Datenkurateure, sowie auch informationswissenschaftliche Einrichtungen als Teil der Infrastruktur zur Lösung beitragen.

Zusammenfassung

Zusammenfassend ist zu sagen, dass die Kernkompetenzen der Informationswissenschaft nahezu deckungsgleich mit spezifischen Bedarfen der Digital Humanities sind. Der Potsdamer Fokus auf Informationstechnik bereitet hier den Weg für eine fruchtbare Zusammenarbeit.

Von Hilfe bei der Entwicklung von Normen und Standards würden die digitalen Geisteswissenschaften ebenfalls profitieren und können hier auf den bereits gesammelten Erfahrungen und getesteten Modellen aufbauen.

³⁰Freie Universität Berlin, Deutsches Archäologisches Institut

³¹<http://www.ianus-fdz.de/>

Auf dem Bereich der Interoperabilität finden sich noch einige Baustellen für die Informationswissenschaft. Hier gilt es vor allem, die Infrastrukturaufgabe zu priorisieren und auf dem Stand der Technik zu bleiben, um mit dem Daten- und Informationsbedarf Schritt zu halten.

Die Visualisierung von Informationen wird nicht nur in den Geisteswissenschaften zentraler und in den Zeiten von Big Data vermutlich unumgänglich werden. Mit den vorhandenen Ressourcen besteht vor allem in Potsdam die Chance eine noch relativ junge Marktnische zu erobern.

Im Bereich der Öffentlichen Bibliotheken ist das Konzept des Dritten Raumes und die Wichtigkeit der Nutzercommunity lang bekannt, wenn auch wissenschaftliche Bibliotheken und Archive noch etwas zögerlicher damit umgehen. An dieser Stelle können vor allem Studien und Expertise zu Informationsverhalten und -management besonders wertvoll für die noch vergleichsweise junge digitale Geisteswissenschaft sein.

4 Informationsinfrastrukturen

Im Verlauf der Konferenz wurde mehrfach die Bedeutung von Infrastrukturen für die digitale Nachhaltigkeit betont. Der Rat für Informationsinfrastrukturen (RfII) definiert Informationsinfrastrukturen als „[...] technisch und organisatorisch vernetzte Dienste und Angebote für den Zugang zu und die Erhaltung von Daten-, Informations- und Wissensbeständen.“ (RfII16) Informationsinfrastrukturen sind demnach keine isolierten Datenspeicher. Sie orientieren sich an den Bedürfnissen der Nutzer, sorgen aktiv für die Verbreitung von Inhalten und verfügen über Schnittstellen zu anderen Diensten.

Einzellösungen

Die Vorträge 3c zu digitalen Editionen machten deutlich, dass projektbasierte Einzellösungen in den Digital Humanities weit verbreitet sind.

Dass diese Zersplitterung in Kleinprojekte häufig der Beschaffenheit des Editionsmaterials geschuldet ist, zeigte der Vortrag von Peter DAENGELI³² und Simon ZUMSTEG³³. Hermann Burgers *Lokalbericht* kann als Mosaik aus Textfragmenten angesehen werden. Die digitale Edition gibt darum nicht nur den Text wieder, sondern dokumentiert auch den komplexen Entstehungszusammenhang der Textbestandteile. Das Projekt griff zwar auf bestehende Standards (TEI) und Normdaten (GND, VIAF) zurück; für das textgenetische Dossier musste aber eine projektspezifische Lösung gefunden werden.

Vera FAßHAUER³⁴ stellte ein Editionsprojekt vor, welches Teile der hand-

³²Universität Köln

³³Schweizerisches Literaturarchiv

³⁴Goethe-Universität Frankfurt am Main

schriftlichen Tagebücher des Frankfurter Arztes Johann Christian Senckenberg digitalisiert. Von insgesamt 53 Bänden wurden 6 ausgewählt und digitalisiert. Mit halbautomatischen Transkriptionsverfahren entstand so eine TEI-basierte Auswahledition. Durch die Beschränkung des Digitalisierungsvolumens wurde mit den Projektmitteln eine inhaltlich tiefere Erschließung (beispielsweise durch Annotationen der Personen oder verwendeter alchemischer Symbole) möglich. Diese tiefe Erschließung sichert durch Kontextanreicherung nachhaltig die Interpretierbarkeit der Tagebücher. Der Ansatz des Projektes lässt sich mit „Nachhaltigkeit durch Auswahl“ zusammenfassen. Einen umfangreicheren Bericht des Diskurses zur Nachhaltigkeit digitaler Editionen während der DHd 2017 hat Anna Busch auf dem DHd-Blog veröffentlicht.³⁵

Partnerschaften

In Panel 4 diskutierten Vertreter aus dem Museumsbereich, wie Museen sich durch die Digitalisierung verändern. Mögliche Ebenen einer Partnerschaft der Gedächtniseinrichtung Museum und den Digital Humanities stellte Malte REHBEIN³⁶ vor. Da Museen nicht nur Datenlieferanten für Geisteswissenschaftler sind, sondern auch selbst Forschung betreiben, könnten sich zwischen Museen und den Digital Humanities zukünftig verstärkt Synergien herausbilden. Dies gilt insbesondere für die Bereiche Erschließung, Standardisierung und Vermittlung.

Georg HOHMANN³⁷ betonte, dass die Digitalisierung in Museen bestehende Organisationsstrukturen und Rollenverständnisse in Frage stellt. Dieser Strukturwandel muss durch Wissenstransfer nachhaltig gestaltet werden. Insbesondere ist das Outsourcen von Digitalisierungsaufgaben kritisch zu sehen: nach Projektende bleibt die Expertise nicht in der Einrichtung und gewonnenes Wissen geht dadurch verloren.

Regina DOPPELBAUER³⁸ sprach über den Konflikt zwischen best practices und der tatsächlich zur Verfügung stehenden Ressourcen. Oft ist nicht genug Personal vorhanden, um Projekte nach aktuellem Kenntnisstand bestmöglich zu realisieren. Die Albertina erwägt daher Crowdsourcing als Lösungsoption.

Die Erfahrungen mit Digitalisierungsprojekten unterscheiden sich je nach Gedächtniseinrichtung. So haben Bibliotheken früher mit der Digitalisierung der Bestände begonnen als Archive. Museen haben damit erst in den letzten Jahren begonnen. Kooperationen und ein verstärkter Erfahrungsaustausch der Gedächtniseinrichtungen würde die Digitalisierung des kulturellen Erbes

³⁵http://dhd-blog.org/?p=7772&utm_content=buffer8a83a&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

³⁶Universität Passau

³⁷Deutsches Museum München

³⁸Albertina Wien

voranbringen.

Offene Lizenzierung

Im Verlauf der Konferenz wurde mehrfach die Bedeutung lizenzrechtlicher Fragen für die Nachhaltigkeit betont. In Workshop 3 stellten Vertreter von Informationsinfrastrukturen (Vanessa HANNESSCHLÄGER³⁹; Walter SCHOLGER⁴⁰) und ein Jurist (Pawel KAMOCKI⁴¹) verschiedene Lizenzmodelle für Daten und Software vor.

Da Forschungsergebnisse erst durch freie Lizenzen nachnutzbar werden, ist die Wahl einer geeigneten Lizenz für nachhaltige Forschung maßgebend. Eine Orientierung an den Konventionen der Fachcommunity stellt neben der rechtlichen Interoperabilität verschiedener Quellen auch eine hohe Rechtssicherheit bei den Nachnutzern sicher, da Lizenzbedingungen und deren praktische Implikationen bekannt sind. Als Faustregel bei der Lizenzwahl kann daher gelten: „Use what your community uses!“

Infrastruktureinrichtungen wie CLARIN unterstützen Daten- und Softwareproduzenten durch Informationsangebote bei der Lizenzwahl und helfen so dabei, die nachhaltige Zugänglichkeit von Forschungsergebnissen und -werkzeugen sicherzustellen. Die Komplexität der rechtlichen Rahmenbedingungen wurde im Vortrag an einem Modell der Datenpublikation verdeutlicht (Abbildung 1). An jeder Schnittstelle zwischen Beteiligten, beispielsweise zwischen dem Urheber und seiner Forschungseinrichtung, sind rechtliche Vereinbarungen erforderlich. Rechtliche Fragestellungen sollten bereits bei der Projektplanung berücksichtigt werden, um die Nachnutzbarkeit zu gewährleisten.

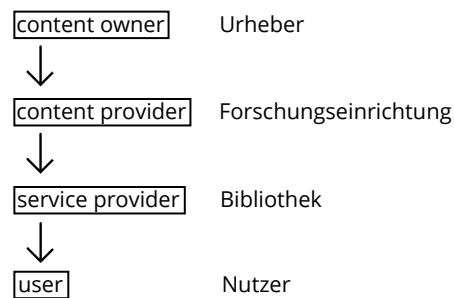


Abbildung 1: Rechteübertragung zwischen Beteiligten

Erfolgsmessung

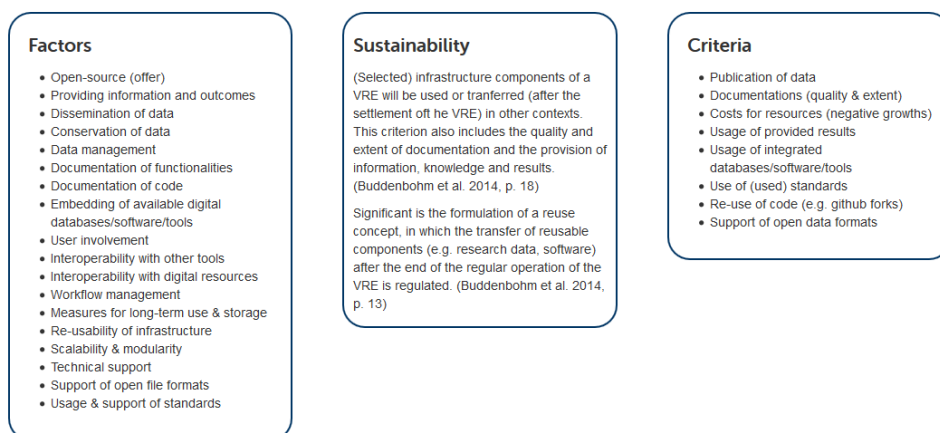
Mit der Frage der Messbarkeit von Erfolg und Impact beschäftigte sich Klaus THODEN⁴² im Vortrag 2b. Das Projekt, das DARIAH-DE als Begleitforschung unterstützt, befasst sich mit der Identifikation von Erfolgskriterien und Impactfaktoren für Tools und Infrastrukturkomponenten in den digita-

³⁹ ACDH

⁴⁰ ZIM-ACDH

⁴¹ IDS Mannheim

⁴² Max-Planck-Institut für Wissenschaftsgeschichte

Abbildung 2: Faktoren und Kriterien für den Impactbereich *Nachhaltigkeit*

len Geisteswissenschaften. Impact und Erfolg sind in den digitalen Geisteswissenschaften bisher nicht allgemein definiert. Erfolgsmessung könnte ein Mittel für mehr Transparenz in Forschungsprojekten und eine verbesserte Kommunikation der Projektergebnisse (beispielsweise an Förderer) sein.

In der Impactmatrix wurden zunächst Bereiche, die für die Erfolgsmessung relevant sind, ermittelt. Für alle Bereiche wurden Faktoren, die den Erfolg beeinflussen, und Kriterien, anhand derer sich der Erfolg bestimmen lässt, dokumentiert. Die Matrix soll unter anderem die Projektplanung unterstützen, indem Maßnahmen stärker am gewünschten Impact orientiert werden. Geeignete Kriterien können dann in der Evaluation der Projektevaluation angewendet werden, um den Erfolg zu messen.

Nachhaltige Strukturen

In der Abschlusskeynote zeichnete Friederike FLESS⁴³ die Geschichte der Informationsinfrastrukturen in Deutschland nach. Nachdem der Start des sowjetischen Satelliten Sputnik 1957 den Westen in eine Krise stürzte, versuchten staatlich gesteuerte Initiativen die Informationsversorgung in Form von Fachinformationsdiensten zentral zu organisieren. In den 80er Jahren wurde dieses Vorhaben zugunsten der Privatisierung von Informationsdiensten aufgegeben. Inzwischen stellt sich erneut die Frage, wie Informationsinfrastrukturen nachhaltig und zukunftsfähig organisiert werden können.

Ob die Ausgestaltung zukunftsfähiger Infrastrukturen ein Prozess *von oben* (durch politische Vorgaben) oder *von unten* (durch Selbstorganisation der Wissenschaftler) sein soll, wird immer wieder diskutiert. Hierfür wurde der Rat für Informationsinfrastrukturen (RfII) 2014 von der Gemeinsamen Wissenschaftskonferenz von Bund und Ländern (GWK) eingesetzt. In ihrer

⁴³Deutsches Archäologisches Institut

Keynote berichtete Frau Fless als Mitglied über die bisherige Arbeit und Positionen des RfII.

Der Rat vertritt die Interessen von Politik und Wissenschaft und soll „[...] als Koordinationsgremium Synergiepotenziale aufzeigen und die Verzahnung der bereits vielfältigen Aktivitäten in diesem Bereich verbessern.“⁴⁴ 2016 legte der Rat erste Empfehlungen vor. Er sprach sich für die Einrichtung einer nationalen Forschungsdateninfrastruktur (NFDI) aus. Die NFDI soll bestehende Dienste in Form eines Kompetenznetzwerks organisieren und dadurch für Nutzer bündeln. Dies soll in einem gesteuerten Prozess geschehen, der an Nutzerbedürfnissen ausgerichtet ist. Hierfür sollen sich wissenschaftliche Communities in Konsortien organisieren. Ein hauptamtliches Management und Aufsichtsgremien koordinieren die Prozesse. Die Finanzierung der NFDI soll dauerhaft gesichert sein.

Das Vorhaben fasst der RfII in seinen Empfehlungen zusammen: „Ziel sollte sein, ein System nachhaltiger Informationsinfrastrukturen zu schaffen, die sowohl verlässliche Arbeitsstrukturen bieten als auch eine dynamische Weiterentwicklung ermöglichen und eine qualitätsgesicherte Grundversorgung in Deutschland sichern.“(RfII16)

Zusammenfassung

Bisher sind die digitalen Geisteswissenschaften von vereinzelten Inselprojekten bestimmt, die häufig zwischen Qualität und Umfang von Digitalisierungsprojekten abwägen müssen. Infrastruktureinrichtungen und die Digital Humanities könnten zukünftig stärker kooperieren und voneinander profitieren. Als Projektpartner müssen Informationsinfrastrukturen bestenfalls schon in der Projektplanung mit einbezogen werden. Der Messbarkeit von Erfolg in Forschungsprojekten sollte mehr Aufmerksamkeit geschenkt werden, da sie schon bei der Zielsetzung unterstützen kann. Die Evolution der Informationsinfrastrukturen ist eine gemeinsame Aufgabe von wissenschaftlichen Communities und bestehenden Informationseinrichtungen. Eine gelungene Zusammenarbeit ist für beide Seiten von Vorteil.

5 Schluss

Auf der Konferenz zeigte sich, dass es fruchtbare Kooperationen zwischen den Digital Humanities und den Informationswissenschaften auf Augenhöhe geben kann. Für Informationswissenschaftler entstehen aus dieser Zusammenarbeit spannende neue Aufgabenfelder und Fragestellungen. Im Gegenzug profitieren die Digital Humanities von informationswissenschaftlichen Methoden und Konzepten (wie Informationsverhaltensforschung und Datenmodellierung). Informationswissenschaftler haben als Fachfremde in den

⁴⁴<http://www.rfii.de/de/category/fakten/>

Digital Humanities den Vorteil, dass sie auf der Metaebene agieren und dadurch eine übergreifende Perspektive einbringen.

Außerdem wurde klar, dass Nachhaltigkeit eine große Herausforderung für die digitalen Geisteswissenschaften darstellt. Die bisher vorhandenen Speziallösungen wie CLARIN, IANUS und DARIAH adressieren zwar einzelne Fachcommunities, eine koordinierende Instanz (zum Beispiel in Form einer disziplinübergreifenden Plattform) fehlt bisher allerdings. Die Informationsinfrastrukturen sind noch nicht ausreichend auf den Paradigmenwechsel eingestellt.

Innerhalb der geisteswissenschaftlichen Projekte besteht Bedarf an informationswissenschaftlicher Arbeit. Dass es in diesem Bereich ausgebildete Fachkräfte gibt, ist in vielen Projekten nicht bekannt. Forschungsvorhaben in den Digital Humanities profitieren von Informationswissenschaftlern unter anderem in Hinblick auf Wissens- und Forschungsdatenmanagement. Die von vielen Forschungsförderern verlangten Datenmanagementpläne werden häufig als Last wahrgenommen. Wenn Datenkuratoren diese Aufgabe übernehmen, können sie die Nachnutzbarkeit der Forschungsdaten sicherstellen. Hierfür muss der Bedarf nach Datenkuratoren in den jeweiligen Fachkreisen explizit kommuniziert werden.

Die langfristige Finanzierung von Projekten bleibt ein Problem. Hierfür muss sich auf lange Sicht die Förderstruktur ändern.

Interessant für folgende DHd Tagungen wäre es den Fokus auf verstärkten Austausch zwischen Projekten und Reflexion der Projekterfahrungen zu legen, da die reine Vorstellung bisher etwas isoliert wirkt. Das Thema „Kritik der digitalen Vernunft“ der DHd 2018 in Köln verspricht hier einen interessanten Ansatz.

Für Informationswissenschaftler hat sich der Besuch der DHd 2017 gelohnt. In den digitalen Geisteswissenschaften gibt es diverse Anknüpfungspunkte und die Voraussetzungen für einen ertragreichen Austausch mit den Informationswissenschaften sind gegeben. Zudem schafft die Anwesenheit von Informationswissenschaftlern bei fachwissenschaftlichen Tagungen das Bewusstsein für mögliche Kooperationen.

Anhang: Besuchte Veranstaltungen

Montag	Workshop 1: Einführung in das PANDORA Linked Open Data Framework
	Workshop 2: Daten sammeln, modellieren und durchsuchen mit DARIAH-DE
	Workshop 3: open your data, open your code: Offene Lizenzierung für geisteswissenschaftliche Projekte
Dienstag	Workshop 4: CUTE: CRETA Unshared Task zur Entitätenreferenzen
	Workshop 5: Dokumente segmentieren und Handschriften erkennen: Arbeiten mit der Plattform Transkribus
	Workshop 6: How to build your own „Digital Edition Web-App“
	Workshop 7: Nachhaltiges Management von Bildmetadaten mit XMP, exiftool und Fotostation
	Workshop 8: Annotieren und Publizieren mit DARIAH-DE und TextGrid
	Eröffnungsk keynote: Digitale Nachhaltigkeit: Mittel und/oder Zweck?
Mittwoch	Vorträge 1/2a: Visualisierung
	Vorträge 1/2b: Forschungsprozess
	Panel 1: Virtuelle Forschungsumgebung für objekt- und raumbezogene Forschung
	Panel 3: Zugänglichkeit und dauerhafte Nutzbarkeit historischer Bildrepositorien für Forschung und Vermittlung
Donnerstag	Mittelkeynote: Ein Bild sagt mehr als tausend Worte, aber sagen tausend Pixel mehr als ein Wort?
	Vorträge 3a: Modellierung
	Vorträge 3b: Bilder
	Panel 6: Virtuelle Forschungsplattformen im Vergleich: MONK, Textgrid, Transcribo und Transkribus
	Panel 9: eValuation - Kriterien zur Evaluation digitaler Angebote und Forschungsinfrastrukturen
	Posterslam
Freitag	Vorträge 4b: Forschungsdaten
	Vorträge 4c: Kollaboration
	Abschlusskeynote: Wenn Forschen ein nicht reproduzierbarer Prozess ist – Nachhaltigkeit als Herausforderung in der Archäologie

Literatur

- [And15] ANDORFER, Peter: *DARIAH-DE Working Papers*. Bd. 10: *Forschen und Forschungsdaten in den Geisteswissenschaften: Zwischenbericht einer Interviewreihe*. Göttingen : DARIAH-DE, 2015
- [Buz09] BUZETTI, Dino: Digital Editions and Text Processing. In: DEEGAN, Marilyn (Hrsg.) ; SUTHERLAND, Kathryn (Hrsg.): *Text editing, print and the digital world*. Farnham, Surrey : Ashgate, 2009 (Digital research in the arts and humanities), S. 45–61
- [Rfi16] RFI – RAT FÜR INFORMATIONENINFRASTRUKTUREN (Hrsg.): *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen, 2016
- [Stu14] STUERMER, Matthias: *Characteristics of Digital Sustainability*. Bern : University of Bern, Research Center for Digital Sustainability, 2014

Abbildungsverzeichnis

Abbildung 1

nach einer Vortragsfolie von Workshop 3 am 13.02.2017

Abbildung 2

<https://dariah-de.github.io/Impactomatrix/sustainability.html>
(abgerufen am 29.03.2017)