

"Das ist nicht in unserer Verantwortung" - Strategien zur nachhaltigen Bereitstellung lebender Systeme

Helling, Patrick

patrick.helling@uni-koeln.de

Data Center for the Humanities (DCH), Universität zu Köln, Deutschland

ORCID: 0000-0003-4043-165X

Einleitung

Forschungsdaten als zentrale Ergebnisse (geisteswissenschaftlicher) Forschungsvorhaben sollten im Sinne der FAIR-Prinzipien (Wilkinson et al. 2016) möglichst auffindbar, zugänglich, interoperabel und nachnutzbar gehalten werden. In diesem Zusammenhang unterstützen insbesondere Beratungseinrichtungen und Infrastrukturanbietende Forschende dabei sowohl auf einer generischen als auch auf einer fachspezifischen Ebene Maßnahmen des Forschungsdatenmanagements (FDM) zu planen und umzusetzen.¹

Neben Forschungsdaten können auch lebende Systeme² bzw. Ressourcen Ergebnisse von Forschungsvorhaben sein. Sie fungieren häufig als Repräsentations-, Visualisierungs- oder Zugangsschichten zu Forschungsdaten und -ergebnissen und stellen diese in einem spezifischen (Nutzungs-)Kontext bereit. Sofern Forschungsdaten und -ergebnisse nicht zusätzlich als vom jeweiligen lebenden System losgelöste Datenpublikation zur Nachnutzung veröffentlicht wurden, bspw. weil die Nutzung der Daten ausschließlich im Kontext eines lebenden Systems sinnvoll möglich ist, stellen diese Ressourcen die einzige Zugangsschicht zu Ergebnissen dar und müssen entsprechend in gleichem Maße wie Forschungsdaten auffindbar, zugänglich, interoperabel und nachnutzbar gehalten werden. Darüber hinaus können auch lebende Systeme i.S.v. Tools oder Analysesoftware zur Datenverarbeitung und -auswertung Resultate von Forschungsvorhaben sein. Auch diese gilt es nach den FAIR-Prinzipien zu sichern und nachnutzbar zu halten.

Während es bereits einige etablierte Maßnahmen, Infrastrukturangebote und Best Practices im Umgang mit Forschungsdaten über den gesamten Forschungsdatenlebenszyklus hinweg (siehe u.a. Minn und Lemaire 2017; Witt et al. 2018), bezüglich spezifischer Bedarfe wie bspw. die Archivierung (Helling und Rau 2024), oder im Kontext bestimmter Fachbereiche (siehe u.a. Jung, Helling und

Pielström 2023) gibt, scheint, trotz einiger Ansätze (von Suchodoletz et al. 2023), der standardisierte und nachhaltige Umgang mit lebenden Systemen hingegen noch weitestgehend ungeklärt zu sein (Helling et al. 2024). Vielmehr noch, sie gehen regelmäßig verloren (s.u.) und mit ihnen essenzielle Forschungsergebnisse (bzw. der Zugang zu ihnen).

In diesem Beitrag werden die Herausforderungen im nachhaltigen Umgang mit lebenden Systemen und Ressourcen diskutiert und mit bereits bestehenden Strategien sowie ihrer Schwachstellen kritisch in Bezug gesetzt. Dabei wird ein neuer Ansatz vor dem Hintergrund verschiedener Verantwortungen unterschiedlicher Stakeholder, die im Prozess der Planung, Entwicklung und Bereitstellung von lebenden Systemen eine Rolle spielen – Forschende, Drittmittelgeber, Datenzentren und Bibliotheken –, sowie Möglichkeiten der Orchestrierung dieser Stakeholder, vorgestellt.

Realitätscheck – die nachhaltige Bereitstellung von lebenden Systemen

Der *Katalog Digitaler Editionen*,³ der von Patrick Sahle und seinem Team gepflegt wird, umfasst insgesamt 846 eingetragene digitalen Editionen (Stand Juli 2024). Die Datenbank lässt sich nach unterschiedlichen Kriterien wie bspw. Edenda, den jeweiligen Fachbereich, aus dem eine digitale Edition stammt oder der verwendeten Sprache in einer digitalen Edition durchsuchen. Die Einträge des Katalogs sind als XML-Datei zwecks Nachnutzung verfügbar.⁴



Abbildung 1: Gesamtzahl der gelisteten digitalen Editionen und ihre Erreichbarkeit.

Eine Auswertung der Erreichbarkeit, der in diesem umfangreichen Katalog von Sahle gelisteten digitalen Editionen ergibt, dass mit lediglich 383 von 789 ausgewerteten digitalen Editionen weniger als die Hälfte der Ressourcen über ihre dort gelisteten URLs erreichbar sind (siehe Abb. 1). Die 406 Einträge, die nicht mehr erreichbar sind, verfügen dabei aus unterschiedlichen Gründen (siehe Abb. 2) über fehlerhafte URLs. Über die Hälfte der nicht mehr erreichbaren Ressourcen (210 Fälle) geben einen Status Code 301 – moved permanently zurück, 83 Ressourcen wur-

den zwar gefunden, sind aber temporär umgezogen (Status Code 302 – found (moved temporarily)) und 63 Ressourcen gaben den Status Code 404 – not found zurück. Darüber hinaus wurden viele verschiedene weitere Fehlermeldungen registriert (siehe Abb. 2).

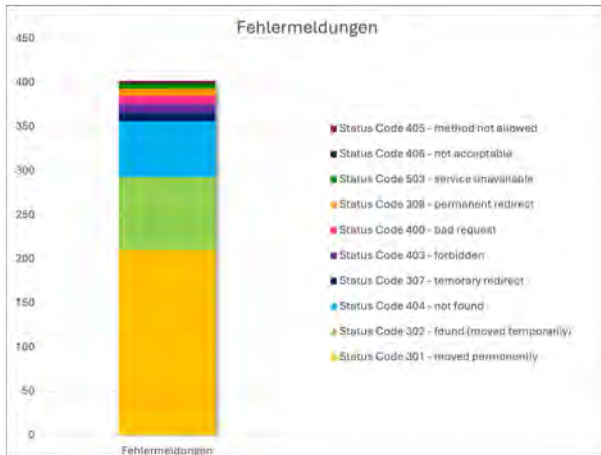


Abbildung 2: Registrierte Fehlermeldungen der 406 nicht erreichbaren URLs.

Vor dem Hintergrund, dass mir nicht bekannt ist, in welchem Umfang die Einträge in Sahles Katalog gepflegt und aktualisiert werden, könnte die hohe Anzahl an nicht erreichbaren Ressourcen darauf zurückzuführen sein, dass der Katalog schlichtweg nicht aktuell ist. Allerdings sollten Ressourcen wie digitale Editionen (genauso wie Forschungsdaten und -ergebnisse), allein für ihre Zitierbarkeit, aber vor allem auch für die Sicherstellung einer Nachnutzbarkeit, möglichst nachhaltig und permanent auf einer beständigen Infrastruktur abgelegt und zur Verfügung gestellt werden, was offensichtlich nicht der Fall ist. Entsprechend deutet die Analyse der Erreichbarkeit von digitalen Ressourcen, die im Katalog Digitaler Editionen von Patrick Sahle gelistet sind, darauf hin, dass es noch keine umfassenden Ansätze gibt, um ebenjene Forschungsergebnisse langfristig zur Verfügung zu stellen (Helling et al. 2024) und, dass Ressourcen wie lebende Systeme nicht selten einfach verschwinden und mit ihnen wichtige Forschungserkenntnisse.

Mit anderen Worten: Mit jeder nicht erreichten Ressource sind potenziell Forschungsergebnisse, die in durch Drittmittel finanzierten Forschungsvorhaben entstanden sind und potenziell von Dritten zitiert und nachgenutzt wurden, um eigene Forschungsfragen zu beantworten, verloren gegangen. Daraus resultiert nicht nur eine sinkende Qualität und Verlässlichkeit von Forschung, bspw. wenn zitierte und verwendete Quellen im Kontext von Journal-Publikationen nicht mehr erreichbar sind, sondern auch, insbesondere im Kontext geisteswissenschaftlicher Forschung, häufig der Verlust eines Teils des kulturellen Erbes, den es eigentlich zu bewahren gilt.

Bestehende Strategien für die nachhaltige Bereitstellung von lebenden Systemen

Es existieren bereits einige Strategien für den nachhaltigen Betrieb lebender Systeme und Ressourcen: **(1)** Eine Möglichkeit, um lebende Systeme langfristig zur Verfügung zu stellen, ist die Investition ausreichender finanzieller und personeller Ressourcen (Smithies et al. 2019), bspw. für Infrastrukturen und Datenzentren sowie Bibliotheken, die lebende Systeme über die Projektlaufzeit hinweg übernehmen und betreiben könnten, da es i.d.R. ihre Aufgabe ist, das Management von Forschungsdaten in irgendeiner Weise zu unterstützen bzw. zu übernehmen und Ressourcen bereitzustellen. Auch **(2)** die Einschränkung zu nutzender Technologie-Stacks (Arneil et al. 2019) im Rahmen der Entwicklung lebender Systeme, idealerweise auf Basis von community-getriebenen de-facto Standards (siehe bspw. Cugliana und van Zundert 2022: S. 49-50), aber auch vor dem Hintergrund von Kompetenzprofilen und technischer Ausstattungen von Infrastrukturen, Datenzentren und Bibliotheken, die lebende Systeme potenziell übernehmen und hosten, kann eine valide Vorgehensweise für die Sicherstellung der langfristigen Verfügbarkeit von lebenden Systemen sein. Mit der zunehmenden Nutzung von Software wie bspw. Docker⁵ stellt auch **(3)** die Virtualisierung oder Kapselung lebender Systeme eine mögliche Strategie dar (Smithies et al. 2019), um ebenjene Ressourcen langfristig verfügbar zu halten. Zuletzt kann durch die **(4)** Statistisierung, bspw. als HTML-Instanz, der Kurations- und Betreuungsaufwand bezüglich lebender Systeme im Hosting für entsprechende Infrastrukturen, Datenzentrum und Bibliotheken, oder sogar für die Forschenden selbst, minimiert werden und somit den Betrieb mehrerer lebender Systeme skalierbar gehalten werden (Arneil et al. 2019).

Allerdings sind all diese Strategien in unterschiedlicher Ausprägung restriktiv: **(1)** Der umfassende Betrieb von lebenden Systemen und Ressourcen skaliert bei wachsender Anzahl an Anwendungen i.d.R. nicht, da häufig verschiedene Technologie-Stacks bei der Entwicklung von Systemen genutzt werden, die eine individuelle Betreuung verlangen. Eine Investition von ausreichend finanzieller und personeller Ressourcen würde entsprechend unkalkulierbar wachsen. **(2)** Obwohl es eine vergleichsweise vielversprechende Strategie wäre, existieren noch deutlich zu wenige community-getriebene Standards und Best Practices für die Entwicklung lebender Systeme, die für eine gewisse Einschränkung von genutzten Technologie-Stacks sorgen könnten. Eine Einschränkung von zu verwendenden Technologie-Stacks durch Infrastruktureinrichtungen steht hingegen im Widerspruch zur Freiheit der Wissenschaft. **(3)** Für die Nutzung von Technologien wie bspw. Docker bedarf es nicht nur spezifischer technischer Ausstattung, sondern vor allem auch spezifisch geschultes, technisches Personal. Außerdem kann selbst die Betreuung von Container-Infrastrukturen einen gewissen Aufwand mit sich brin-

gen. (4) Das statische Hosting lebender Systeme und Ressourcen bedeutet darüber hinaus häufig auch den Verlust von (notwendigen) Funktionalitäten, die im Rahmen von Forschungsvorhaben aufwendig entwickelt wurden und die für die Nutzung von lebenden Systemen oder den Zugang zu in ihnen enthaltenen Forschungsdaten und -ergebnissen entscheidend sind, allerdings in statischer Form eben verloren gehen.

Verantwortlichkeiten verschiedener Stakeholder

Bestehende Strategien im Umgang mit lebenden Systemen stehen aktuell im Spannungsfeld zwischen einer personellen und finanziellen Skalierbarkeit, der Freiheit von Wissenschaft, dem Verlust von Funktionalitäten und der effektiven Nachnutzbarkeit der Ressourcen. Es fehlt weiterhin ein Ansatz, der sowohl finanzielle und personelle Anforderungen niedrig hält, gleichzeitig das Bestehen substantieller Funktionalitäten pauschal gewährleisten kann und die möglichst freie Gestaltung und Entwicklung von lebenden Ressourcen ermöglicht.

Grundsätzlich gilt es in diesem Zusammenhang zu berücksichtigen, dass der (langfristige) Betrieb lebender Systeme und Ressourcen in der Verantwortung aller Stakeholder, nämlich der **Forschenden, Drittmittelgeber und Infrastrukturen/Datenzentren/Bibliotheken**, liegt.

Forschende sind i.d.R. Expert:innen in ihrem jeweiligen Fachbereich. Vor dem Hintergrund, dass sie im Rahmen von durch Drittmittel geförderten Forschungsvorhaben meist öffentliche Gelder erhalten, liegt es mit in ihrer Verantwortung, dass Forschungsdaten und -ergebnisse genauso wie lebende Systeme, die im Rahmen ihrer Forschung entstehen, unter der Berücksichtigung möglicher, rechtlicher Schranken nachhaltig öffentlich zur Verfügung stehen. Schlussendlich ist es in diesem Zusammenhang elementar, dass Forschende sich insbesondere ihrer Verantwortung gegenüber durch sie entwickelten lebenden Systemen und Ressourcen bewusst sind, v.a. darüber, dass diese Verantwortung nicht mit dem Ende der Finanzierung, bspw. im Rahmen eines Drittmittelprojektes, endet.

Um dieser Verantwortung gerecht zu werden, benötigen Forschende allerdings nicht nur das Bewusstsein für ebenjene dauerhafte Verantwortung ihrerseits sondern auch Expertise und Unterstützung in / bei der Planung und Entwicklung lebender Systeme und Ressourcen sowie (im Rahmen von durch Drittmittel geförderten Forschungsvorhaben) eine ausreichende finanzielle Ausstattung für die nachhaltige Entwicklung dieser Systeme. Gleichzeitig müssen Forschende in der Lage sein, Konzepte für den nachhaltigen Betrieb zu planen und über eine Projektlaufzeit hinaus umzusetzen, was mit einschließt, dass auch diese Aspekte im Rahmen von Drittmittelanträgen mit eingepreist werden können sollten, woraus allerdings keine Kürzung der eigentlichen Forschungsgelder resultieren darf.

Drittmittelgeber wiederum fördern Forschungsvorhaben i.d.R. mit öffentlichen Geldern. Sie koordinieren Reviewprozesse für Forschungsprojektanträge und haben in diesem Rahmen die Möglichkeit insbesondere in solche Vorhaben zu investieren, in denen nachhaltige Systeme und Ressourcen sowie plausible Lösungen für die langfristige Bereitstellung ebenjener Systeme und Ressourcen entwickelt bzw. angewendet werden sollen. Dafür müssen allerdings Nachhaltigkeitsaspekte sowohl in Bezug auf die Entwicklung als auch in Bezug auf Bereitstellungsstrategien in Reviewprozesse von Forschungsprojektanträgen integriert werden und Reviewer:innen zur Beurteilung jener Aspekte Kompetenzen aufbauen. Außerdem sollten Drittmittelgeber die Möglichkeit geben, Gelder für das langfristige Hosting und den Betrieb sowie die Archivierung von lebenden Systemen und Ressourcen zu beantragen, ohne dass dies einen Einfluss auf die eigentlichen Forschungsgelder hat.

Infrastrukturen, Datenzentren und Bibliotheken sowie die Institutionen, an denen sie angesiedelt sind, sind i.d.R. für die Unterstützung von Forschenden im Bereich Forschungsdaten- und Softwaremanagement zuständig. Zu ihren Aufgaben gehört es nicht selten spezifische Angebote wie Beratungs- und Infrastrukturservices für Forschende bereitzustellen. Sie sollten in der Lage sein einerseits Forschende darin zu schulen nachhaltiges Daten- und Softwaremanagement umzusetzen und andererseits Infrastrukturkomponenten und Services anbieten zu können, mit denen insbesondere auch lebende Systeme und Ressourcen langfristig gehostet und betrieben werden können. Hierfür bedarf es allerdings, neben entsprechenden Kompetenzen, einer ausreichenden, vor allem permanenten, Finanzierung dieser Einrichtungen.

Vor dem Hintergrund, dass nicht jede infrastrukturen-bietende / FDM-Service Einrichtung alle Bedarfe im Forschungsdaten- und Softwaremanagement adressieren und abdecken kann, bedarf es zusätzlich geeigneter Geschäftsmodelle, die es ermöglichen, dass vor allem Infrastrukturen, Datenzentren und Bibliotheken, die einen klar definierten Adressat:innenkreis haben wie bspw. eine Universität, eigene Services auch über Institutions- und Ländergrenzen hinweg anbieten können, sodass Forschende bezüglich der dauerhaften Bereitstellung ihrer Ressourcen nicht durch möglicherweise defizitäre Angebote der eigenen Einrichtung eingeschränkt werden. Dabei gilt es in diesem Rahmen v.a. auch auf Basis der eigenen Kapazitäten und Ausstattung klare Kostenmodelle für entsprechende Services zu definieren, sodass entsprechende Kosten durch Forschende im Rahmen von Drittmittelanträgen mit eingepreist werden können. Diese Modelle schießen dabei explizit Konzepte zur erweiterten Zuständigkeit von Einrichtungen, zur Finanzierung und zum Ressourcentransfer mit ein.

Eine Möglichkeit den Bedarf entsprechender Geschäftsmodelle zu adressieren, könnte möglicherweise die Nationale Forschungsdateninfrastruktur (NFDI)⁶ mit ihren dezentralen, im Verbund organisierten Konsortien darstellen. Inwieweit dies allerdings einer als Projekt und zeitlich begrenzt finanzierten Infrastrukturinitiative gelingen könnte, bleibt bisweilen abzuwarten.

Zusammenfassung

Zentrale Voraussetzungen in der Sicherstellung einer nachhaltigen Verfügbarkeit von lebenden Systemen und Ressourcen sind einerseits die Erkenntnis, dass verschiedene Stakeholder innerhalb des gesamten Forschungsprozesses unterschiedliche Verantwortungen und Aufgaben haben und andererseits, dass nur durch eine enge Zusammenarbeit zwischen all diesen Stakeholdern eine realistische Chance besteht, nachhaltige Konzepte für den Dauerbetrieb von lebenden Systemen und Ressourcen zu entwickeln und strukturell nachhaltig zu implementieren.

Drittmittelgeber sollten nachhaltige, lebende Systeme und Ressourcen finanzieren und Konzepte der nachhaltigen Entwicklung Teil des Begutachtungsprozesses machen. **Forschende** sollten das Management von lebenden Systemen vom Beginn ihrer Forschungsvorhaben an mitberücksichtigen und sich über ihre Verantwortung über die Grenzen eines drittmittelfinanzierten Projekts hinweg bewusst sein. **Infrastrukturen, Datenzentren und Bibliotheken** sowie die **Institutionen**, an denen sie angesiedelt sind, sollten spezifische Unterstützung für Forschende auf Dauer und über ihre Institutions- sowie über Ländergrenzen hinweg anbieten (können), bspw. durch das Angebot von Software Management Plänen (Alves et al. 2021), aber vor allem durch die Entwicklung von Strategien und dem Betrieb von technischer Infrastruktur zur nachhaltigen Übernahme lebender Systeme und Ressourcen.

In diesem Vortrag werde ich die verschiedenen Verantwortlichkeiten der unterschiedlichen Stakeholder im Detail vorstellen und für eine Orchestrierung ebenjener argumentieren, um auf diese Weise Restriktionen in bestehenden Strategien für den nachhaltigen Umgang mit lebenden Systemen und Ressourcen zu umgehen.

Fußnoten

1. <https://www.forschungsdaten.org/index.php/FDM-Kontakte>, (letzter Zugriff: 21. Juni 2024).
2. Im Vergleich zu Forschungsdaten sind lebende Systeme dynamische bzw. regelmäßig zu kuratierende Umgebungen, Zugangs- oder Repräsentationsschichten zu Forschungsdaten. Ihre Funktion ist i.d.R. die Zugänglichkeit und Nachnutzung von Forschungsdaten in einer spezifischen Form zu ermöglichen.
3. www.digitale-edition.de/, letzter Zugriff: 21. Juni 2024.
4. <https://git.uni-wuppertal.de/dhsfu/sde-catalog>, letzter Zugriff: 21. Juni 2024.
5. <https://www.docker.com/>, letzter Zugriff: 21. Juni 2024.
6. <https://www.nfdi.de/>, letzter Zugriff: 09. Juli 2024.

Bibliographie

Alves, Renato, Dimitrios Bampalikis, Leyla Jael Castro, José María Fernández, Jennifer Harrow,

Mateusz Kuzak, Eva Martin, Fotis E. Psomopoulos, und Allegra Via. 2021. „ELIXIR Software Management Plan for Life Sciences“. Preprint. BioHackrXiv. <https://doi.org/10.37044/osf.io/k8znb>.

Arneil, Stewart, Martin Holmes, und Greg Newton. 2019. „Project Endings: Early Impressions From Our Recent Survey On Project Longevity In DH“. DataverseNL. <https://doi.org/10.34894/SIKOBN>.

Cugliana, Elisa und Joris van Zundert. 2022. „A Computational Turn in Digital Philology.“ In: La filologia germanica e il paradigma digitale: modelli, metodi e strumenti per i testi germanici medievali = Germanic philology and the digital paradigm: models, methods and tools for medieval germanics texts. o. J. Milano: Prometheus. S. 40/50.

Helling, Patrick und Felix Rau. 2024. "Ein blinder Fleck in der FDM Versorgungslandschaft? Dark und Cold Archiving Services am Data Center for the Humanities". b.i.t. online 27 (2024) Nr. 1, S. 27–34. <https://www.b-i-t-online.de/heft/2024-01-fachbeitrag-helling.pdf>.

Helling, Patrick, Felix Rau, Philip Schildkamp, Lisa Dieckmann, Johanna Puhl, und Ulrike Henny-Krahmer. 2024. „Still alive?! - Vom Umgang mit lebenden Systemen in den Digital Humanities“, Februar. <https://doi.org/10.5281/ZENODO.10698430>.

Jung, Kerstin, Patrick Helling, und Steffen Pielström. 2023. „Forschungsdatenmanagement in den Computational Literary Studies - Bericht über die FDM-Landschaftsvermessung im DFG Schwerpunktprogramm 2207 „Computational Literary Studies““. Zenodo. <https://doi.org/10.5281/ZENODO.10419199>.

Minn, Gisela und Marina Lemaire. 2017. „Forschungsdatenmanagement in den Geisteswissenschaften. Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans“, Universität Trier eSciences Working Papers, Nr. 03. Online: https://ubt.opus.hbz-nrw.de/opus45ubtr/frontdoor/deliver/index/docId/799/file/WP_03_DMP_final_Juli_2017.pdf (letzter Zugriff: 21. Juni 2024).

Sahle, Patrick. 2016. „What is a scholarly digital edition (SDE)?“. In: Driscoll, Matthew James, und Elena Pierazzo, Hrsg. 2016. *Digital Scholarly Editing: Theories and Practices*. Open Book Publishers. <https://doi.org/10.11647/OBP.0095>.

Smithies, James, Carina Westling, Anna-Maria Sichani, Pam Mellen und Arianna Ciula. 2019. „Mapping 100 Digital Humanities Projects: Digital Scholarship & Archiving in King’s Digital Lab.“ digital humanities quarterly. Online: <http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html> (letzter Zugriff: 17. Juli 2024).

Suchodoletz, Dirk von, Peter Brettschneider, Alexandra Axtmann, Maximilian Heber, Lars Oberländer, Jan Leendertse, Irene Schumm, u. a. 2023. „Bausteine Forschungsdatenmanagement:

2023, 5Sicherstellung der Reproduzierbarkeit von Forschungsergebnissen durch Bewahrung des Zugriffs auf Forschungssoftware“. Application/pdf, Dezember. <https://doi.org/10.17192/BFDM.2023.5.8555>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, u. a. 2016. „The FAIR Guiding Principles for Scientific Data Management and Stewardship“. *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Witt, Andreas, Jonathan Blumtritt, Patrick Helling, Brigitte Mathiak, und Felix Rau. 2018. „Forschungsdatenmanagement in den Geisteswissenschaften an der Universität zu Köln“. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, September, 104-117 Seiten. <https://doi.org/10.5282/O-BIB/2018H3S104-117>.