

Nachträglich ist nicht gleich nachnutzbar: Ansätze für integrierte Prozessdokumentation im Forschungsalltag

Sibylle Hermann, Universitätsbibliothek Stuttgart

Uli Hahn, kiz Universität Ulm

Markus Gärtner, Institut für maschinelle Sprachverarbeitung Universität Stuttgart

Florian Fritze, Universitätsbibliothek Stuttgart

Zusammenfassung:

Um Forschungsdaten auffindbar zu machen, müssen diese mit ausreichend Metadaten beschrieben werden. Damit die durch die Metadaten beschriebenen Forschungsdaten für andere Wissenschaftlerinnen und Wissenschaftler reproduzierbar sind, ist es notwendig, den Kontext ihrer Entstehung mit abzubilden. Gerade die Dokumentation dieses Entstehungsprozesses wird aber oft durch mangelnde Zeit im Forschungsalltag vernachlässigt. Auch fehlt es hier noch an niederschwelliger Unterstützung im Arbeitsprozess. Einige Methoden sind gerade dabei sich zu etablieren oder befinden sich in der Entwicklung. Im Folgenden werden Softwareanwendungen, die die Dokumentation erleichtern sollen, vorgestellt und mit der aktuell im Projekt RePlay-DH entwickelten Lösung verglichen. Der Ansatz der Virtuellen Forschungsumgebung setzt auf die Zusammenarbeit über eine gemeinsame Plattform. Das Elektronische Laborbuch unterstützt die Dokumentation im Labor. Das Workflow-Management definiert, im Gegensatz zum Workflow-Tracking, einen Workflow vor der Ausführung der einzelnen Arbeitsschritte. Dabei steht die prozessbegleitende Dokumentation im Mittelpunkt. Der Lösungsansatz, der im Projekt RePlay-DH verfolgt wird, besteht in der unterstützenden Dokumentation des Forschungsprozesses mit Metadaten durch ein vereinfachtes Workflow-Tracking. Die Integration in bestehende Arbeitsabläufe von Wissenschaftlerinnen und Wissenschaftlern und die einfache Bedienbarkeit stehen dabei im Vordergrund.

Summary:

In order to make research data retrievable, it must be described with sufficient metadata. It is also necessary to include the context in which the research data has come into being, so that other scientists can reproduce the research data described by the metadata. However, the documentation of this origination process is often neglected due to a lack of time in everyday research. In addition, low-threshold support in day-to-day research is often missing. At present, some methods are being developed or about to be established. In this paper, software applications designed to facilitate documentation are presented and compared with the solution currently developed in the RePlay-DH project. The approach of the virtual research environment relies on cooperation via a shared platform. The electronic laboratory notebook supports documentation in the laboratory. In contrast to workflow tracking, workflow management defines a workflow before the individual work steps are carried out. Here, the focus is set on a process-accompanying documentation. The approach of the RePlay-DH project for a solution is to support documentation of the research process with metadata through simplified workflow tracking. Special emphasis is laid on the integration in the existing work processes of scientists and scholars and easy handling.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H3S32-45>

Autorenidentifikation: Hermann, Sibylle: GND 1073989070, ORCID: <http://orcid.org/0000-0001-9239-8789>; Hahn, Uli: GND 1130888878, ORCID: <https://orcid.org/0000-0002-1316-881X>;

Gärtner, Markus: ORCID: <https://orcid.org/0000-0002-2687-4350>; Fritze, Florian: ORCID: <https://orcid.org/0000-0002-9949-3815>

Schlagwörter: Forschungsdatenmanagement

1. Einleitung

Ziel dieses Artikels ist, darzulegen, wie komplexe Forschungsprozesse so dokumentiert werden, dass diese reproduzierbar gemacht werden können.¹ Neben den Datenobjekten, die als Forschungsdaten verarbeitet werden, sind dabei die bei der Verarbeitung anfallenden Arbeitsschritte von Interesse. Aus diesem Grund ist für die Beschreibung eines Forschungsprozesses eine prinzipielle Zweiteilung in eine Objekt- und eine Prozessdomäne wichtig. Die Objektmetadaten dienen der Beschreibung des veröffentlichten Ergebnisses, um dieses auffindbar zu machen. Objektmetadaten werden durch Standards wie das DataCite-Metadatenschema² disziplinübergreifend diskutiert. Neben der Beschreibung der Objekte geht es bei der Reproduzierbarkeit wissenschaftlicher Ergebnisse aber vor allem um eine ausführliche Prozessdokumentation.³ Die hier betrachtete Prozessdokumentation konzentriert sich auf den disziplinspezifischen Arbeitsprozess und beschreibt, wie das Ergebnis zustande kam. Eine Dokumentation des Arbeitsprozesses ist zwar wünschenswert, im Arbeitsalltag von Forschenden fehlen dazu aber oft die Zeit und geeignete Werkzeuge.⁴ Zwar wird eine Archivierung der Daten im Sinne der guten wissenschaftlichen Praxis von der Deutschen Forschungsgemeinschaft gefordert,⁵ allerdings bleibt es dabei den Forschenden überlassen, wie diese zu erfolgen hat. So werden Dateien oft unstrukturiert in Ordnern abgelegt, was ein Nachvollziehen der eigenen Arbeit zu einem späteren Zeitpunkt erschwert. Wenn die Dokumentation dazu noch einer eigenen Logik folgt, ist es unter Umständen für andere nicht oder nur mit einem hohen Arbeitsaufwand möglich, das Ergebnis nachzuvollziehen.

1 Nicht nur Daten, sondern auch der Code und der Workflow sollen verfügbar und zitierfähig sein. Vgl. Stodden, Victoria, McNutt, Marcia; Bailey, David H. u.a.: Enhancing reproducibility for computational methods, in: *Science* 354 (6317), 2016, S. 1240–1241. Online: <<https://doi.org/10.1126/science.aah6168>>, Stand: 01.08.2018.

2 DataCite, <https://www.datacite.org/>, Stand: 01.08.2018.

3 Die Replizierbarkeit bezieht sich auf die Wiederholung, um gleiche Ergebnisse in der gleichen Umgebung zu erlangen. Reproduzierbarkeit bezieht sich auf eine unabhängige Wiederholung in einer anderen Umgebung. Nachnutzbarkeit bezieht sich darauf die Methode für andere Zwecke zu nutzen. Vgl. Fehr, Jörg; Heiland, Jan; Himpe, Christian u.a.: Best practices for replicability, reproducibility and reusability of computer-based experiments exemplified by model reduction software, in: *AIMS Mathematics* 1 (3), 2016, S. 261–281. Online: <<https://doi.org/10.3934/Math.2016.3.261>>, Stand: 01.08.2018.

4 Vgl. den Abschlussbericht des Projekt bwFDM-Communities. Tristram, Frank: Öffentlicher Abschlussbericht von bwFDM-Communities. Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs. Online: <<http://bwfdm.scc.kit.edu/downloads/Abschlussbericht.pdf>>, Stand: 01.08.2018. Die Umfrage ergibt, dass ein Grund für fehlendes Datenmanagement der hohe Zeitaufwand für die Datenverwaltung ist. (S. 77 f.)

5 Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 30.09.2015. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 01.08.2018.

Die Dokumentation der eigenen Arbeit kann im Falle komplexer Prozesse eine erhebliche Zeiterparnis bedeuten, da für das Suchen und Verstehen der eigenen Gedankengänge weniger Zeit aufgewendet werden muss. Eine normierte, detaillierte Beschreibung des Arbeitsprozesses würde es Wissenschaftlerinnen und Wissenschaftlern, die mit dem Forschungsprozess selbst nicht vertraut sind, ermöglichen, schnell ein Verständnis für den Prozess und dessen Ergebnis zu erhalten. Darüber hinaus sollte für die Veröffentlichung von Forschungsdaten angestrebt werden, diese einfach nachnutz- und nachvollziehbar zur Verfügung zu stellen.

Um dieses Problem zu lösen, existieren bereits Ansätze zur Prozessdokumentation. Vier dieser Lösungen werden im Folgenden vorgestellt und in Beziehung zueinander gesetzt. Zunächst werden dazu notwendige Schritte für eine lückenlose Prozessdokumentation definiert. In einem weiteren Abschnitt werden technische Verfahren beschrieben, wie eine Dokumentation von Arbeitsschritten erfolgen kann. Abschließend wird der Lösungsansatz des Projektes RePlay-DH in Bezug zu den vorgestellten Methoden der Forschungsdokumentation vorgestellt.

2. Voraussetzungen zur Dokumentation von Forschungsprozessen

Der Forschungsprozess wird nachvollziehbar, wenn zusätzlich zu den im Prozess verwendeten Input-Daten und Werkzeugen eine genaue Beschreibung der durchgeführten Schritte und vorgenommenen Änderungen erfolgt. Eine manuelle Anreicherung der Arbeitsschritte mit Metadaten hilft dabei, Inhalte auffindbar zu machen. Auch für die Veröffentlichung der Forschungsdaten sollten interne Strukturen und fachspezifische Beschreibungen durch allgemein verständliche Beschreibungen in den Metadaten ergänzt werden, um die interdisziplinäre Nachnutzung nach den FAIR-Prinzipien⁶ zu ermöglichen. Idealerweise werden dabei Inhalte über persistente Links referenziert und erfahren vor der Veröffentlichung eine Qualitätssicherung durch disziplinspezifische Standards. Die Nachvollziehbarkeit fremder Inhalte dient auch dazu, die Datenveröffentlichung bei Wissenschaftlerinnen und Wissenschaftlern als wichtiges und zitierfähiges Verfahren zu etablieren. Der Aufwand, den Forschende leisten müssen, um eine Dokumentation vorzunehmen, darf dabei allerdings nicht außer Acht gelassen werden. Nur wenn der Nutzen höher als der Aufwand ist, wird in der Regel ein zusätzlicher Aufwand für die Dokumentation in Kauf genommen.

Es gibt also drei wesentliche Voraussetzungen für eine nachnutzbare Dokumentation:

- Beschreibung des durchgeführten Arbeitsschritts mit Metadaten
- Kontrollierte bzw. persistente Beschreibung der Zwischenergebnisse
- Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Aktuell deutet sich im Wissenschaftsbereich noch kein etabliertes Verfahren an, das händische Annotation und technische Lösungen so miteinander verbindet, dass eine einfache Prozessdokumentation

6 Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3 (160018), 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 01.08.2018.

für Wissenschaftlerinnen und Wissenschaftler möglich wird. Dennoch gibt es verschiedene Lösungsansätze, die den Forschungsprozess unterstützen können. Im Folgenden werden diese Lösungen in vier prinzipielle Methoden eingeteilt und deren Herangehensweise in Bezug auf die vorher aufgeführten Thesen hin diskutiert.

3. Methoden zur Dokumentation von Forschungsprozessen

Die hier vorgestellten Lösungen für die Forschungsdokumentation lassen sich grob dadurch klassifizieren, wo und wie sie im Workflow eingreifen oder verwendet werden. Virtuelle Forschungsumgebungen (VFU) umfassen den gesamten Forschungsdatenlebenszyklus und beinhalten als Teilaspekt die Workflowdokumentation. Elektronische Laborbücher ersetzen und ergänzen die traditionelle Dokumentation (auf Papier) von Laborversuchen. Bei Workflow-Management-Systemen wird ein kontrollierter Ablauf von einzelnen Arbeitsschritten vor deren Durchführung definiert. Im Gegensatz dazu stehen Workflow-Tracking-Systeme, welche die einzelnen Arbeitsschritte während oder nach deren Entstehungsprozess dokumentieren.

3.1. Virtuelle Forschungsumgebungen (VFU)

Auch wenn in Deutschland der Trend der VFU abzunehmen scheint, gibt es im nationalen⁷ und internationalen⁸ Kontext weiterhin Projekte, die den Aufbau einer solchen Plattform zum Ziel haben. Im Allgemeinen stimmen die verschiedenen Definitionen von VFU in den wesentlichen Punkten überein.⁹ Der Zugriff auf Ressourcen und die Veröffentlichung von Forschungsergebnissen stehen dabei im Vordergrund. Aus Sicht der Wissenschaftlerinnen und Wissenschaftler definieren Neuroth et al. die VFU als Dienstleistung, bei der Forschende im Arbeitsprozess bei allen Schritten „effizient und wirksam“ unterstützt werden sollen und dabei über die VFU auf alle für ihre Arbeit relevanten Inhalte und Forschungsdaten zugreifen können.¹⁰ Dabei gehen Neuroth et al. davon aus, dass Wissenschaftlerinnen und Wissenschaftler ihren Arbeitsprozess in die VFU verlegen und Zwischenergebnisse für sie selbst und andere bereitgestellt werden.¹¹ Aus struktureller Sicht führt Lossau die drei Grundbausteine Informationsobjekte, Werkzeuge und Systeme an, aus denen eine VFU besteht,¹² Candela et al. dagegen nennen fünf Eigenschaften einer virtuellen Forschungsumgebung.¹³ Sie definieren virtuelle

7 Was sind virtuelle Forschungsumgebungen, [forschungsdaten.info, https://www.forschungsdaten.info/praxis-kompakt/aktuelle-infrastrukturprojekte/virtuelle-forschungsumgebungen/](https://www.forschungsdaten.info/praxis-kompakt/aktuelle-infrastrukturprojekte/virtuelle-forschungsumgebungen/), Stand: 01.08.2018.

8 Wie z.B. Hydroshare, <https://www.hydroshare.org/>, Stand: 01.08.2018.

9 Beispielhaft sei hier die Definition der Allianz-AG Virtuelle Forschungsumgebungen genannt, <http://doi.org/10.2312/ALLIANZOA.028>: „Eine Virtuelle Forschungsumgebung ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse – während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert.“

10 Neuroth, Heike, Jannidis, Fotis; Rapp, Andrea u.a.: Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften, in: BIBLIOTHEK Forschung und Praxis 33 (2), 2009, S. 161-169. Online: <<https://doi.org/10.1515/bfup.2009.017>>, Stand: 01.08.2018.

11 Ebd., S.163.

12 Lossau, Norbert: Virtuelle Forschungsumgebungen und die Rolle von Bibliotheken, in: Zfbb 58, (3-4), 2011, S. 154-163. Online: <<http://resolver.sub.uni-goettingen.de/purl?gs-1/7057>>, Stand: 01.08.2018.

13 Candela, Leonardo; Castelli, Donatella; Pagano, Pasquale: Virtual Research Environments: An Overview and a Research Agenda, in: Data Science Journal 12, 2013, GRDI75-GRDI81. Online: <<https://doi.org/10.2481/dsj.GRDI-013>>, Stand: 27.07.2018.

Forschungsumgebungen zunächst als webbasierte Arbeitsumgebungen (i), die auf die Bedürfnisse einer „community of practice“ angepasst sind (ii) und über alle Werkzeuge verfügen, um die Ziele der Community realisieren zu können (iii), des Weiteren sollen VFU ein offenes und flexibles Serviceangebot anbieten (iv) und ein kontrolliertes Teilen von (Zwischen-)Ergebnissen ermöglichen (v).¹⁴ Die VFU zielen also darauf ab, einer abgeschlossenen fachlichen Community die Zusammenarbeit und den Austausch ihrer Ergebnisse zu erleichtern. Die Beschreibung dieser Ergebnisse ist dabei mit einem gemeinsamen Vokabular einfacher zu gewährleisten.

3.1.1. Beschreibung des durchgeführten Arbeitsschritts mit Metadaten

Metadaten werden allerdings nur bei Lossau explizit als wesentlicher Bestandteil einer VFU beschrieben.¹⁵ Sie werden als Basis für die Funktionen der Benutzeroberfläche der VFU hervorgehoben, um in dieser arbeiten zu können sowie um die Ressourcen zu finden, sie zu verknüpfen und deren digitale Langzeitverfügbarkeit zu gewährleisten.¹⁶ Darüber hinaus verweist Lossau auf die Rolle der Metadaten für „die Rekonstruktion der Produktionsbedingungen von Forschungsdaten inkl. ihres Entstehungsworkflows.“¹⁷ Dabei betont er, dass die Erfassung der Metadaten möglichst automatisiert erfolgen sollte.¹⁸

Inwieweit diese Beschreibung mit Metadaten in der VFU geleistet werden kann, ist jedoch stark von der Fachcommunity und deren Bestreben nach Vereinheitlichung geprägt.

3.1.2. Kontrollierte bzw. persistente Beschreibung des Zwischenergebnisses

Die Publikation der (Zwischen-)ergebnisse wird in den oben genannten Definitionen als wesentlicher Bestandteil einer VFU angesehen. So stellen Candela et al. fest: „Finally, the VRE’s characteristic of supporting fine-grained controlled sharing of both intermediate and final research results while guaranteeing ownership, provenance, and attribution is somehow a consequence of the scenarios VREs are going to serve.“¹⁹ Dabei müssen allerdings die Bedenken der Forschenden, ihre Daten zu teilen, berücksichtigt werden. Eine Forschungsinfrastruktur bestimmt den Handlungsspielraum, in dem sich Nutzerinnen und Nutzer bewegen können, und den „Rahmen der möglichen Handlungen der Forschenden.“²⁰

3.1.3. Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Ein weiteres Ziel einer VFU ist die Zusammenarbeit innerhalb einer Community. Das heißt aber auch, dass das Arbeiten in der VFU nur dann Sinn macht, wenn die komplette Community darin arbeitet. Dies kann auf einen Forschungsverbund oder ein gemeinsames Thema, das über ein Projekt bearbeitet wird, zutreffen. Existiert keine VFU oder ist die Community sich in deren Auswahl nicht einig, kann

14 Ebd.

15 Lossau: Virtuelle Forschungsumgebungen, 2011, S. 157.

16 Ebd.

17 Ebd.

18 Ebd.

19 Candela, Castelli, Pagano: Virtual Research Environments, 2013, S. GRD176.

20 Kaden, Ben; Rieger, Simone: Usability in Forschungsinfrastrukturen für die Geisteswissenschaften, in: Neuroth, Heike; Rapp, Andrea; Söring, Sibylle (Hg.): TextGrid: Von der Community – für die Community, Glückstadt 2015, S. 63-75. Online: <<http://doi.org/10.3249/webdoc-3947>>, Stand: 01.08.2018., S. 66.

die Unterstützung im eigenen Arbeitsprozess sinnvoller sein, um schnell teilbare Zwischenstände zu erhalten. Ein Ergebnis einer begleitenden Studie zu Textgrid, einer VFU in den Digitalen Geisteswissenschaften, zeigt, „dass die Darstellung der Anforderungen der Nutzer/innen in allen [in der Studie untersuchten] Kategorien stark durch die Perspektive auf ihre (digitalen) Forschungsgegenstände und die damit verbundenen spezifischen Methoden und Ziele geprägt ist, nicht durch den Wunsch nach einem universellen digitalen Werkzeugkasten.“²¹ Lösungswege, die Wissenschaftlerinnen und Wissenschaftler außerhalb der VFU suchen, bleiben somit für andere zunächst unsichtbar.

3.2. Elektronische Laborbücher

Elektronische Laborbücher (Electronic Laboratory Notebook - ELN) werden vor allem in der Medizin und den Naturwissenschaften eingesetzt. Sie ersetzen oder ergänzen traditionelle Papier-Laborbücher. In den Laborbüchern werden Rahmenbedingungen zur Durchführung eines Experiments und dessen Ergebnisse dokumentiert. Nussbeck et al.²² führen drei Gründe an, warum sich ELN in Zukunft durchsetzen werden: Erstens liegen immer mehr Daten digital vor, welche dokumentiert und archiviert werden müssen. Zweitens können Datenformate und Protokolle leichter standardisiert werden und zum Dritten können Daten nur langzeitverfügbar gemacht werden, wenn sie ausreichend mit Metadaten beschrieben werden.

3.2.1. Beschreibung des durchgeführten Arbeitsschritts mit Metadaten

Wie Nussbeck et al. erklären, stellen die in einem Laborbuch dokumentierten Parameter und Versuchsprotokolle implizit Metadaten für die Reproduzierbarkeit und Nachnutzung der Forschungsdaten dar.²³ Darüber hinaus fordern sie, das Beschreiben der Daten mit Metadaten zu erzwingen, sowie ein standardisiertes Metadatenschema zu etablieren.²⁴ Da Laborbücher schon immer den Zweck hatten, einen Versuch reproduzierbar zu machen, besteht in der elektronischen Variante die Möglichkeit, dies mit einem vereinheitlichten Vokabular zu bewerkstelligen. Wie Menzel et al. ausführen, bedarf es eines höheren Aufwands, um sich auf die einheitliche Beschreibung eines Versuches zu einigen, aber die spätere Dokumentation und das Finden, sowie das Vergleichen mit anderen Ergebnissen, wird dadurch deutlich vereinfacht.²⁵

3.2.2. Kontrollierte bzw. persistente Beschreibung des Zwischenergebnisses

Die Universität Edinburgh, die Universität Göttingen und die Universität Wisconsin haben zusammen mit der Firma Research Space das ELN RSpace²⁶ entwickelt, dass sich in die universitäre Infrastruktur

21 Bender, Michael: *Forschungsumgebungen in den Digital Humanities. Nutzerbedarf, Wissenstransfer, Textualität*, Berlin; Boston 2016.

22 Nussbeck, Sara Y., Weil, Philipp; Menzel, Julia u.a.: *The laboratory notebook in the 21st century: The electronic laboratory notebook would enhance good scientific practice and increase research productivity*, in: *EMBO reports* 15 (6), 2014, S. 631-634. Online: <<http://doi.org/10.15252/embr.201338358>>, Stand: 01.08.2018.

23 Ebd., 632.

24 Ebd., 633.

25 Menzel, Julia; Weil, Philipp; Nussbeck, Sara Y.: *Minimierung des Dokumentationsaufwandes für die medizinische Grundlagenforschung mithilfe eines elektronischen Laborbuches: Aufgezeigt am Beispiel eines Western Blot Metadatenschemata*, 59. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, Göttingen 2014. Online: <<https://dx.doi.org/10.3205/14gmds120>>, Stand: 01.08.2018.

26 RSpace, <<https://www.researchspace.com/>>, Stand: 01.08.2018.

einbinden lässt.²⁷ Daneben besteht die Möglichkeit, Laborbücher über Plattformen wie Open Notebook Science Network²⁸ oder OpenWetWare²⁹ frei zugänglich zu führen. Auch das von der DFG geförderte Projekt Chemotion verbindet ein eigenentwickeltes ELN mit einem Repositorium.³⁰ Da konventionelle Laborbücher quasi schon immer eingesetzt wurden, kann der Wechsel zur elektronischen Variante mit einem kontrollierten Vokabular, auf das sich die Community geeinigt hat, die Interoperabilität und Nachvollziehbarkeit deutlich steigern. Wie die Standardisierungsbemühungen z.B. in der Chemie zeigen, wird das Potential einer persistenten einheitlichen Beschreibung der Daten erkannt und umgesetzt.³¹

3.2.3. Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Soweit sich das beurteilen lässt, hat sich momentan noch kein Softwareprodukt an Universitäten durchsetzen können. Da auch hier die Bedürfnisse variieren, ist die Benutzerfreundlichkeit vom Anwendungsfall abhängig. Die Harvard-Universität hat zu diesem Zweck eine Matrix erstellt, bei der viele gängige ELN-Angebote anhand von verschiedenen Kriterien evaluiert wurden.³² Umso besser die Anbindung an bestehende Repositorien und die Universitätsinfrastruktur gelingen wird, desto höher wird die Akzeptanz bei den Anwenderinnen und Anwendern ausfallen. Wie die Untersuchung von Krause zeigt, gehen die bereits an Universitäten eingesetzten Lösungen meist mit einer Anbindung an ein Repositorium einher.³³

3.3. Workflow-Management

Workflow-Management-Systeme (WMS) sind in der Software-Entwicklung bereits seit langem etabliert. Es existiert eine Vielzahl von kommerziellen und freien Softwarelösungen für unterschiedliche Einsatzgebiete. Systeme für das Management von Workflows sind Ansätze zur Unterstützung von vordefinierten Prozessen. „Das Workflow-Management-System ist ein System, welches die Phasen des Prozess-Lifecycles, also das Workflow-Management, durch IT-Werkzeuge unterstützt.“³⁴ Das vorherrschende Konzept zur Modellierung solcher Workflows ist eine Abfolge voneinander abhängiger Arbeitsschritte.³⁵

27 Macdonald, Stuart; Macneil; Rory: Service Integration to Enhance Research Data Management: RSpace Electronic Laboratory Notebook Case Study, in: International Journal of Digital Curation 10 (1), 2015, S. 163–172. Online: <<https://doi.org/10.2218/ijdc.v10i1.354>>, Stand: 01.08.2018.

28 Open Notebook Science Network, <<http://onsnetwork.org/>>, Stand: 01.08.2018.

29 OpenWetWare, <<https://openwetware.org/>>, Stand: 01.08.2018.

30 Chemotion, <<https://chemotion.net>>, Stand: 01.08.2018.

31 McEwen, Leah R.: IUPAC Facilitating Chemistry Data Exchange in the Digital Era, in: Chemistry International 39 (2), 2017. Online: <<https://doi.org/10.1515/ci-2017-0205>>., Stand: 01.08.2018.

32 Harvard Biomedical Data Management, Harvard University, <<https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>>, Stand: 01.08.2018.

33 Krause, Evamaria: Elektronische Laborbücher im Forschungsdatenmanagement – Eine neue Aufgabe für Bibliotheken?, in: ABI Technik 36 (2), 2016, S. 78–87. Online: <<https://doi.org/10.1515/abitech-2016-0013>>, Stand: 01.08.2018.

34 Müller, Joachim: Workflow-based Integration. Grundlagen, Technologien, Management, Berlin 2005, S. 11.

35 Deelman, Ewa; Gannon, Dennis; Shields, Matthew u.a.: Workflows and e-Science: An overview of workflow system features and capabilities, in: Future Generation Computer Systems 25 (5), 2009, S. 528–540. Online: <<https://doi.org/10.1016/j.future.2008.06.012>>, Stand: 01.08.2018.

3.3.1. Beschreibung des durchgeführten Arbeitsschritts mit Metadaten

WMS werden in proaktiver Weise dazu eingesetzt, um Workflows und die diesen zugrunde liegenden individuellen Schritte oder Prozesse im Vorfeld der Workflow-Ausführung formell zu beschreiben. Je nach Funktionsumfang des jeweiligen WMS hat diese Beschreibung entweder rein dokumentarischen Charakter oder kann unter Umständen auch direkt zur automatischen Ausführung einzelner Schritte oder auch des kompletten Workflows genutzt werden. Lassen sich diese Beschreibungen aus dem WMS heraus exportieren, so können diese direkt als erster Schritt einer Workflow-Dokumentation genutzt werden.

3.3.2. Persistente Beschreibung des Zwischenergebnisses

Diverse Forschungsdisziplinen haben WMS-Lösungen bisher in unterschiedlichem Umfang adaptiert. Insbesondere in Bereichen mit traditionell rechenlastigen Workflows, wie z.B. in der biomedizinischen Forschung oder Genetik, existieren etablierte Systeme wie Galaxy³⁶ oder GenPattern³⁷. Diese ermöglichen es Forschenden, aus umfangreichen Katalogen bereitgestellter Werkzeuge individuelle Sammlungen zusammenhängender Prozessierungsschritte zu definieren und diese auszuführen. Durch die Notwendigkeit einer formellen Definition des gesamten (auszuführenden) Workflows liefern WMS-Instanzen bereits nachnutzbare und/oder zum Zwecke der Dokumentation archivierbare Beschreibungen der geplanten oder abgeschlossenen Abläufe.³⁸

3.3.3. Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Die Anwendbarkeit von WMS ist häufig beschränkt auf eher homogene Workflows in Bezug auf die Ausführungsmodalität einzelner Workflowschritte. Mit zunehmender Heterogenität der zugrunde liegenden Systeme steigt die Problematik, ganze Workflows weiterhin automatisch ausführen zu lassen. Finden manuelle Arbeitsschritte als Teil eines Workflows statt, so gelangen viele WMS an ihre Grenzen und alternative Ansätze gewinnen an Bedeutung.

3.4. Workflow-Tracking

Im Gegensatz zum Konzept des Workflow-Managements stellt Workflow-Tracking eine reaktive Lösung für die Problematik nachhaltiger Dokumentation dar. Hierbei werden vordefinierte Ressourcen überwacht und bei relevanten Änderungen oder Ereignissen die Dokumentation der damit verbundenen Prozesse ermöglicht oder verlangt. Die wohl am weitesten verbreiteten Vertreter dieses Lösungsansatzes sind Systeme zur Versionsverwaltung oder Versionskontrolle. Ursprünglich aus dem Bereich der Software-Entwicklung stammend dienen diese dazu, Änderungen in Sammlungen von Dateien (dem Repositorium), typischerweise Text-Dokumente, die Quellcode enthalten, zu verfolgen und Änderungen zu dokumentieren, d.h. zu versionieren. Der Fokus liegt auf der Möglichkeit, dass

36 Afgan, Enis; Baker, Dannon; van den Beek, Marius u.a.: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update, in: *Nucleic Acids Research* 44 (W1), 2016, W3-W10. Online: <<https://dx.doi.org/10.1093%2Fnar%2Fgkw343>>, Stand: 27.07.2018.

37 Reich, Michael; Liefeld, Ted; Gould, Joshua u.a.: GenePattern 2.0, in: *Nature Genetics* 38 (5), 2006, S. 500-501. Online: <<https://doi.org/10.1038/ng0506-500>>, Stand: 01.08.2018.

38 Zur Archivierung und Reproduzierbarkeit dieser Inhalte wird aktuell in dem Baden-Württemberg Landesprojekt „CITAR“ an Lösungen gearbeitet, da hier hardwarebedingte Fragestellungen bei der Archivierung im Mittelpunkt stehen, die einen Einfluss auf die errechneten Forschungsergebnisse haben. CITAR, <<https://www.forschungsdaten.info/praxis-kompakt/aktuelle-infrastrukturprojekte/fdm/citar/>>, Stand: 01.08.2018.

mehrere Benutzerinnen und Benutzer gleichzeitig kollaborativ und verteilt an denselben Ressourcen arbeiten und konkurrierende Änderungen über ein öffentliches oder zugriffsbeschränktes System zusammenführen können.

3.4.1. Beschreibung des durchgeführten Arbeitsschritts mit Metadaten

Die eigentliche Dokumentation erfolgt hauptsächlich durch von den Anwenderinnen und Anwendern in so genannte Commit-Nachrichten eingetragenen Freitext bei der Erfassung von neuen Änderungen. Bekannte Versionierungssysteme sind unter anderem Git³⁹ als verteilte Lösung, bei der ein lokales unabhängiges Repositorium unter Versionskontrolle gestellt wird und dies zur Kollaboration an ein zentrales Repositorium angebunden werden kann. Daneben findet Apache Subversion (SVN)⁴⁰ als von Grund auf zentralisierter Ansatz Verwendung.

3.4.2. Persistente Beschreibung des Zwischenergebnisses

Die persistente Beschreibung eines Zwischenergebnisses ist in den Versionierungssystemen so nicht vorgesehen. Bei öffentlichen Repositorien wie GitLab⁴¹ ist der gesamte Prozess mit allen Änderungen sichtbar, allerdings kann nicht gesagt werden, ob ein Betreiber nicht irgendwann seinen Dienst wieder einstellt. Die Daten sind zwar nach wie vor existent, da alle, die an einem Projekt arbeiten, ein komplettes Abbild auf ihrer eigenen Rechenumgebung besitzen. Allerdings ist keine nachhaltige URL gewährleistet. GitHub⁴² bietet die Möglichkeit an, Zwischenstände direkt über Zenodo⁴³ zu veröffentlichen. Dabei wird das gesamte Projekt archiviert. Einzelne Zwischenergebnisse lassen sich damit bisher nicht veröffentlichen.⁴⁴ Wie eine Veröffentlichung von bestimmten Teilen aus einem Git-Repositorium ermöglicht werden und in universitäre Veröffentlichungsinfrastrukturen integriert werden kann, wird momentan im Projekt SARA⁴⁵ erforscht.

3.4.3. Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Neben oben erwähnten räumlichen Aspekten kann im Falle von Systemen zur Versionsverwaltung auch die Art der zu versionierenden Daten Probleme aufwerfen. Zur detaillierten Erkennung von Änderungen müssen die Daten mindestens einmal komplett dupliziert werden. Da diese Systeme ursprünglich für klar strukturierte Dateien (konkret für Textdokumente, in denen Informationen zeilenweise hinterlegt sind) konzipiert wurden, können insbesondere sehr große oder unstrukturierte Daten, wie beispielsweise die meisten binären Formate, einen unverhältnismäßig großen Speicherverbrauch mit sich bringen und diese Art des Workflow-Trackings unpraktikabel machen. Auch birgt der Zeitpunkt, zu dem bei dieser Art des Workflow-Trackings die Dokumentation erstellt wird, den Nachteil, scheinbar kleine und unwichtige, für die nachfolgende Verarbeitung oder Auswertung aber

39 Git, <<https://git-scm.com/>>, Stand: 01.08.2018.

40 SVN, <<https://subversion.apache.org/>>, Stand: 01.08.2018.

41 GitLab, <<https://gitlab.com/>>, Stand: 01.08.2018.

42 GitHub, <<https://github.com/>>, Stand: 01.08.2018.

43 Zenodo, <<https://zenodo.org/>>, Stand: 01.08.2018.

44 Potter, Megan; Smith, Tim: Making Code citeable with Zenodo and GitHub, Software Sustainability Institute Blog, 28.07.2015, <<https://www.software.ac.uk/blog/2016-09-26-making-code-citable-zenodo-and-github>>, Stand: 01.08.2018.

45 SARA, <<https://www.forschungsdaten.info/praxis-kompakt/aktuelle-infrastrukturprojekte/fdm/sara/>>, Stand: 01.08.2018.

bedeutsame Informationen auszulassen. Auch die Eingabe von Freitext sollte durch unterstützende Hinweise gesteuert werden, da sonst eine mangelnde Aussagekraft der Dokumentation zu befürchten ist. Alternativen zum rein reaktiven Verfolgen von Änderungen schließen unter anderem die direkte Integration von Werkzeugen zur Dokumentation in die Ausführung bestehender Abläufe ein. Beispiele hierfür sind Systeme wie YesWorkflow⁴⁶, die eine Annotation von auszuführenden Skripten mit Instruktionen erlauben, auf Basis derer dann eine Dokumentation des damit zusammenhängenden Workflows generiert werden kann. Analog zu übergeneralisierenden WMS sind derartige Nischenlösungen aber auch nur eingeschränkt für viele Workflows im Forschungsalltag anwendbar.

4. Workflow-Tracking mit RePlay-DH

Ziel des vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg geförderten Projekts RePlay-DH ist es, die Arbeitsprozesse auf der individuellen Arbeitsebene der jeweiligen Forschenden zu erfassen und diese mit möglichst geringem Mehraufwand zu dokumentieren. Die vorgestellten WMS bieten nicht die gewünschte Flexibilität, da sie darauf abzielen, vordefinierte Workflows zu optimieren. Daher wurde ein Workflow-Tracking-Ansatz gewählt und versucht, die erkannten Defizite, wie die hohe Komplexität und die große Einarbeitungshürde, zu reduzieren. Das Projekt RePlay-DH nutzt die Vorteile der Versionierung mit Git und reichert die Möglichkeit des kommentierten Zwischenspeicherns mit kontrollierten Prozessmetadaten an. Der Arbeitsprozess findet in der Arbeitsumgebung der Wissenschaftlerinnen und Wissenschaftler statt und kann auf die individuellen Bedürfnisse angepasst werden. Dazu wird ein Client entwickelt, der auf allen gängigen Rechnerplattformen ausführbar ist und nur eine Java-Laufzeitumgebung zum Starten benötigt.⁴⁷ Der Quellcode und eine ausführbare Programmdatei stehen frei im Netz und können ausprobiert werden.⁴⁸

4.1. Beschreibung des durchgeführten Arbeitsschritts mit Metadaten

Der Client bietet, neben der Versionskontrolle, die (semi-)automatische Anreicherung der generierten und verarbeiteten Forschungsdaten mit Metadaten. Ziel ist es, heterogene und intellektuelle bzw. händisch geprägte Arbeitsprozesse zu dokumentieren, um eine Reproduzierbarkeit wissenschaftlicher Ergebnisse zu ermöglichen. Das hinterlegte Metadatenschema zur Abbildung eines fachlich geprägten Forschungsprozesses kann an verschiedene Bedürfnisse angepasst werden. Weiterhin besteht die Möglichkeit, Metadaten von nachgenutzten Materialien einzubinden. Die Dokumentationsarbeit kann damit auf ein Minimum reduziert werden. Die Verknüpfung der Forschungsdaten mit Metadaten kann schon während des Forschungsprozesses erfolgen. Damit kann eine große Hürde,

46 McPhillips, Timothy M.; Song, Tianhong; Kolisnik, Tyler u.a.: YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts, in: *International Journal of Digital Curation* 10 (1), 2015, S. 298-313. Online: <<https://doi.org/10.2218/ijdc.v10i1.370>>, Stand: 01.08.2018.

47 Eine genauere Erläuterung von Architektur, Schnittstellen und der Anbindung an Informationsinfrastruktur findet sich im Artikel von Gärtner, Markus; Hahn Uli; Hermann, Sibylle: Supporting Sustainable Process Documentation, in: Rehm, Georg; Declerck, Thierry (Hg.): *Language Technologies for the Challenges of the Digital Age*, Cham, 2018 (Lecture Notes in Computer Science 10713), S. 284-291. Online: <https://doi.org/10.1007/978-3-319-73706-5_24>, Stand: 01.08.2018. Über die Homepage des Projektes werden aktuelle Entwicklungen bekannt gegeben. RePlay, <<http://www.ub.uni-stuttgart.de/replay>>, Stand: 01.08.2018.

48 RePlay-DH-Client Sourcecode, GitHub, <<https://github.com/RePlay-DH>>, Stand: 01.08.2018.

Daten zu veröffentlichen, überwunden werden. Darüber hinaus hilft die Dokumentation auch, die eigenen Ergebnisse wiederzufinden.

4.2. Persistente Beschreibung des Zwischenergebnisses

In diesem Kontext ist geplant, dass Nutzerinnen und Nutzer des RePlay-DH-Clients dabei unterstützt werden, Dateien inklusive Metadaten zu exportieren. Diese für eine Veröffentlichung vorgesehen Datensätze bzw. Forschungsdaten können entweder automatisiert über eine Schnittstelle in einem entsprechenden Repositorium abgelegt werden oder in einem passenden Format exportiert werden. Möchte der Forschende ein Backup der Daten erstellen oder mit anderen Kolleginnen und Kollegen kollaborieren, ist eine Anbindung an eine zentrale Instanz eines Git-Servers am jeweiligen Forschungsinstitut des Nutzers möglich. Die Anbindung an weitere Komponenten einer Forschungsdateninfrastruktur ist optional, Wissenschaftlerinnen und Wissenschaftler können den RePlay-Client auch nur lokal als Desktop-Programm auf ihren Rechnern starten und die Vorzüge (Metadatenanreicherung und Versionskontrolle) nutzen. Der Client ist so autonom, dass er nicht zwangsläufig mit entfernten Infrastrukturkomponenten interagieren muss, auch wenn diese Möglichkeit besteht.

4.3. Benutzerfreundlichkeit des Dokumentations- und Veröffentlichungswerkzeugs

Um die Benutzerfreundlichkeit zu erhöhen, wird die zugrunde liegende Versionierungssoftware Git durch eine grafische Benutzeroberfläche abgebildet, die einmal aufgerufen auch im Hintergrund laufen kann. Durch die Dokumentation des Forschungsprozesses mit Git kann der Prozess in Form eines Workflowgraphen visuell dargestellt werden. Die Darstellung folgt dabei logischen Mustern, die in Git über sogenannte „Branches“ und „Commits“ realisiert werden und den Forschenden damit die Möglichkeit bieten, sämtliche Einzelschritte innerhalb eines Prozesses zu überblicken und diese somit besser zu überwachen und bei Bedarf steuern zu können. In der Regel wird bei jeder größeren Veränderung der Daten der Status quo der Forschung festgehalten und kann bei Bedarf wiederhergestellt werden. Somit wird der Prozess nicht nur nach den Prinzipien von Git dokumentiert und visualisiert, sondern von Nutzerinnen und Nutzern auch semantisch beschrieben, um eine intellektuelle Zuordnung der getätigten Arbeitsschritte zu gewährleisten. Dabei soll die Interaktion zwischen den Forschenden und dem RePlay-DH-Client auf ein Mindestmaß reduziert werden, indem schon vergebene Metadaten erkannt und nachgenutzt werden können. Der Fokus bei RePlay liegt hierbei auf der Unterstützung des Arbeitsprozesses und nicht auf der Zusammenarbeit. Darüber hinaus wird eine höhere Flexibilität angestrebt, da das zugrunde liegende Metadatenschema an verschiedene Fachdisziplinen angepasst werden kann. Momentan finden dazu erste Gespräche mit der ingenieurwissenschaftlichen Fachcommunity statt.

5. Fazit

Die Archivierung von Forschungsergebnissen steht schon lange im Interesse der Forschungsförderer. Mit neuen Initiativen und neuen technischen Möglichkeiten rückt nun die nachhaltige Dokumentation und – wo möglich – die Veröffentlichung der Forschungsdaten in den Fokus. VFU und WMS verfolgen den Ansatz einer maßgeschneiderten forschungsgerechten Umgebung. Sie bieten Archivierungs- und Nachnutzungslösungen auf übergreifender Ebene an. Allerdings sind diese Lösungen

stark an Einzelfällen ausgerichtet und müssen sich erst etablieren. Laborbücher, die traditionell schon lange zur Dokumentation eingesetzt werden, tragen in ihrer elektronischen Form dazu bei, Prozesse zu standardisieren und damit für Maschinen und Menschen besser nachvollziehbar zu machen. Dies bezieht sich allerdings nur auf Daten, die in einem Labor anfallen und für die Dokumentation bereits etabliert sind. Für viele Fächer gibt es keine spezifische Lösung. Mit dem Thema Workflow-Tracking existiert ein Lösungsansatz für Wissenschaftsdisziplinen, in denen bisher wenig dokumentiert wurde. Die zu etablierende technische Dokumentationshilfe soll dazu dienen, den Forschenden im Arbeitsalltag eine Arbeitserleichterung zu bieten.

Trotz des Vorhandenseins oben beschriebener Ansätze erfolgt die Dokumentation von Forschungsprozessen vor oder während deren Ausführung in vollständiger und nachnutzbarer Form nur selten und wenn überhaupt, dann meist nur nachträglich. Neben motivationsbezogenen Gründen, wie dem Fehlen von Anerkennung und/oder der Einforderung einer Dokumentation, treten technische Inkompatibilitäten und mangelnde Standardisierung von Austauschformaten in der Fachgemeinschaft als Barrieren auf. Der RePlay-DH-Client versucht, die Hürden zur Dokumentation, wie die Einarbeitungszeit und den zusätzlichen Aufwand, die neben der eigentlichen Forschungsarbeit geleistet werden müssen, zu reduzieren. Ein zu starkes Abweichen von bisherigen gewohnten Abläufen wird damit vermieden. Für die Standardisierung der Prozessbeschreibung kann damit aber nur ein Impuls gesetzt werden.

Literaturverzeichnis

- Afgan, Enis; Baker, Dannon; van den Beek, Marius u.a.: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update, in: *Nucleic Acids Research* 44 (W1), 2016, W3-W10. Online: <<https://dx.doi.org/10.1093%2Fnar%2Fgkw343>>, Stand: 27.07.2018.
- Arbeitsgruppe Virtuelle Forschungsumgebungen in der Allianz der deutschen Wissenschaftsorganisationen: Definition Virtuelle Forschungsumgebung, 2011, Online: <<http://doi.org/10.2312/ALLIANZOA.028>>, Stand: 01.08.2018.
- Bender, Michael: Forschungsumgebungen in den Digital Humanities. Nutzerbedarf, Wissenstransfer, Textualität, Berlin; Boston 2016.
- Candela, Leonardo; Castelli, Donatella; Pagano, Pasquale: Virtual Research Environments: An Overview and a Research Agenda, in: *Data Science Journal* 12, 2013, GRDI75-GRDI81. Online: <<https://doi.org/10.2481/dsj.GRDI-013>>, Stand: 27.07.2018.
- Deelman, Ewa; Gannon, Dennis; Shields, Matthew u.a.: Workflows and e-Science: An overview of workflow system features and capabilities, in: *Future Generation Computer Systems* 25 (5), 2009, S. 528–540. Online: <<https://doi.org/10.1016/j.future.2008.06.012>>, Stand: 01.08.2018.

- Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 30.09.2015. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 01.08.2018.
- Fehr, Jörg; Heiland, Jan; Himpe, Christian u.a.: Best practices for replicability, reproducibility and reusability of computer-based experiments exemplified by model reduction software, in: AIMS Mathematics 1 (3), 2016, S. 261-281. Online: <<https://doi.org/10.3934/Math.2016.3.261>>, Stand: 01.08.2018.
- Gärtner, Markus; Hahn Uli; Hermann, Sibylle: Supporting Sustainable Process Documentation, in: Rehm, Georg; Declerck, Thierry (Hg.): Language Technologies for the Challenges of the Digital Age, Cham, 2018 (Lecture Notes in Computer Science 10713), S. 284-291. Online: <https://doi.org/10.1007/978-3-319-73706-5_24>, Stand: 01.08.2018.
- Harvard Biomedical Data Management, Harvard University, <<https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>>, Stand: 01.08.2018.
- Kaden, Ben; Rieger, Simone: Usability in Forschungsinfrastrukturen für die Geisteswissenschaften, in: Neuroth, Heike; Rapp, Andrea; Söring, Sibylle (Hg.): TextGrid: Von der Community – für die Community, Glückstadt 2015, S. 63-75. Online: <<http://doi.org/10.3249/webdoc-3947>>, Stand: 01.08.2018.
- Krause, Evamaria: Elektronische Laborbücher im Forschungsdatenmanagement – Eine neue Aufgabe für Bibliotheken?, in: ABI Technik 36 (2), 2016, S. 78-87. Online: <<https://doi.org/10.1515/abitech-2016-0013>>, Stand: 01.08.2018.
- Lossau, Norbert: Virtuelle Forschungsumgebungen und die Rolle von Bibliotheken, in: ZfBB 58, (3-4), 2011, S. 154-163. Online: <<http://resolver.sub.uni-goettingen.de/purl?gs-1/7057>>, Stand: 01.08.2018.
- Macdonald, Stuart; Macneil; Rory: Service Integration to Enhance Research Data Management: RSpace Electronic Laboratory Notebook Case Study, in: International Journal of Digital Curation 10 (1), 2015, S. 163-172. Online: <<https://doi.org/10.2218/ijdc.v10i1.354>>, Stand: 01.08.2018.
- McEwen, Leah R.: IUPAC Facilitating Chemistry Data Exchange in the Digital Era, in: Chemistry International 39 (2), 2017. Online: <<https://doi.org/10.1515/ci-2017-0205>>., Stand: 01.08.2018.
- McPhillips, Timothy M.; Song, Tianhong; Kolisnik, Tyler u.a.: YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts, in: International Journal of Digital Curation 10 (1), 2015, S. 298-313. Online: <<https://doi.org/10.2218/ijdc.v10i1.370>>, Stand: 01.08.2018.

- Menzel, Julia; Weil, Philipp; Nussbeck, Sara Y.: Minimierung des Dokumentationsaufwandes für die medizinische Grundlagenforschung mithilfe eines elektronischen Laborbuches: Aufgezeigt am Beispiel eines Western Blot Metadatenschemata, 59. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, Göttingen 2014. Online: <<https://dx.doi.org/10.3205/14gmids120>>, Stand: 01.08.2018.
- Müller, Joachim: Workflow-based Integration. Grundlagen, Technologien, Management, Berlin 2005.
- Neuroth, Heike, Jannidis, Fotis; Rapp, Andrea u.a.: Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften, in: BIBLIOTHEK Forschung und Praxis 33 (2), 2009, S. 161-169. Online: <<https://doi.org/10.1515/bfup.2009.017>>, Stand: 01.08.2018.
- Nussbeck, Sara Y., Weil, Philipp; Menzel, Julia u.a.: The laboratory notebook in the 21st century: The electronic laboratory notebook would enhance good scientific practice and increase research productivity, in: EMBO reports 15 (6), 2014, S. 631-634. Online: <<http://doi.org/10.15252/embr.201338358>>, Stand: 01.08.2018.
- Potter, Megan; Smith, Tim: Making Code citeable with Zenodo and GitHub, Software Sustainability Institute Blog, 28.07.2015, <<https://www.software.ac.uk/blog/2016-09-26-making-code-citable-zenodo-and-github>>, Stand: 01.08.2018.
- Reich, Michael; Liefeld, Ted; Gould, Joshua u.a.: GenePattern 2.0, in: Nature Genetics 38 (5), 2006, S. 500–501. Online: <<https://doi.org/10.1038/ng0506-500>>, Stand: 01.08.2018.
- Stodden, Victoria, McNutt, Marcia; Bailey, David H. u.a.: Enhancing reproducibility for computational methods, in: Science 354 (6317), 2016, S. 1240–1241. Online: <<https://doi.org/10.1126/science.aah6168>>, Stand: 01.08.2018.
- Tristram, Frank: Öffentlicher Abschlussbericht von bwFDM-Communities. Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs. Online: <<http://bwfdm.scc.kit.edu/downloads/Abschlussbericht.pdf>>, Stand: 01.08.2018.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific data 3 (160018), 2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>, Stand: 01.08.2018.