

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Infrastruktur-Dienste für das aktive und passive Forschungsdatenmanagement eines neurowissenschaftlichen Sonderforschungsbereichs

Thorsten Arendtⁱ Alexander C. Schützⁱⁱ

2019

Zitiervorschlag

Arendt, Thorsten und Alexander C. Schütz. 2019. Infrastruktur-Dienste für das aktive und passive Forschungsdatenmanagement eines neurowissenschaftlichen Sonderforschungsbereichs. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2019: S. 50-53. DOI: [10.17192/bfdm.2019.2.8070](https://doi.org/10.17192/bfdm.2019.2.8070).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱPhilipps-Universität Marburg. ORCID: [0000-0002-4866-6405](https://orcid.org/0000-0002-4866-6405)

ⁱⁱPhilipps-Universität Marburg. ORCID: [0000-0002-7742-8123](https://orcid.org/0000-0002-7742-8123)

1 Einleitung

Das Informationsinfrastrukturprojekt NOWA (Neuroscientific Workflow Assistance)¹ des DFG-Sonderforschungsbereichs SFB/TRR 135 *Kardinale Mechanismen der Wahrnehmung: Prädiktion, Bewertung, Kategorisierung* entwickelt und kombiniert Werkzeuge für die gemeinsame Nutzung von Forschungsdaten. Dabei zielt NOWA darauf ab, einen organisatorischen und technologischen Rahmen zu schaffen, der die Arbeitsabläufe entlang des gesamten Forschungsdatenlebenszyklus unterstützt. Dies beinhaltet zum einen die Verwaltung von Forschungsdaten während Planung und Durchführung einer Studie (aktives Forschungsdatenmanagement) sowie die Bereitstellung der finalen Forschungsergebnisse bei Abschluss der Studie (passives Forschungsdatenmanagement). Dieser Erfahrungsbericht befasst sich mit der infrastrukturellen Seite von NOWA. Ausgehend von einer Zusammenfassung der Anforderungen der Wissenschaftlerinnen und Wissenschaftler an der Justus-Liebig-Universität Gießen und der Philipps-Universität Marburg stellen wir diese zunächst den von den jeweiligen Rechenzentren aktuell und zukünftig bereitgestellten Infrastruktur-Diensten gegenüber. Anschließend skizzieren wir unseren grundlegenden Ansatz für das aktive und passive Forschungsdatenmanagement im Sonderforschungsbereich 135.

2 Anforderungen an das Forschungsdatenmanagement im SFB 135

Wie sich in den ersten Monaten des Projektes herausstellte, können die von den Wissenschaftlern und Wissenschaftlerinnen an NOWA gestellten Anforderungen grundsätzlich in die Bereiche **Organisation** und **Optimierung** eingeteilt werden. Aus organisatorischer Sicht ist in erster Linie der Bedarf an Unterstützung für ein *kollaboratives Arbeiten* zu nennen, was in der Natur eines Sonderforschungsbereichs mit interdisziplinären Teilprojekten begründet liegt. Auch der Bedarf an einer strukturierten Planung, Umsetzung und Verwaltung des Teilprojekt-internen *Forschungsdatenmanagements* fällt in diese Kategorie. Optimiert werden sollen zum einem der Umgang mit *großen Dateien* (z.B. fMRT-Daten) sowie zum anderen diejenigen Aktivitäten, die bei der Veröffentlichung von Forschungsdaten im Kontext von Open Science bzw. Open Data anfallen. Dazu gehören beispielsweise die Auswahl der zu der Publikation passenden *Version* sowie die Sicherstellung einer für die Reproduzierbarkeit der Forschungsergebnisse ausreichende *Qualität* der Forschungsdaten.

¹Arendt, Thorsten, Ortrun Brand, Christian Krippes, Andreas Gabriel, Matteo Valsecchi, Clemens Helf, Karl R. Gegenfurtner und Alexander C. Schütz. *Neuroscientific Workflow Assistance (NOWA)*, DINI Jahrestagung 2018. DOI: <https://doi.org/10.17192/es2019.0002>.

3 Infrastruktur-Dienste an den Rechenzentren

Im Zuge der Digitalisierung unterliegen die Hochschulrechenzentren einer Transformation weg von einem reinen Hardware- hin zu einem kombinierten IT-Dienstleister. Infolgedessen stellen diese Einrichtungen in zunehmenden Maß verschiedene, im Kontext von NOWA nützliche, Dienste und Werkzeuge bereit. Mit **Planio**² stellt das Rechenzentrum der Philipps-Universität Marburg eine integrierte Plattform zur Unterstützung von Projektarbeit und Projektmanagement zur Verfügung. Im Rahmen eines vom Hessischen Ministerium für Wissenschaft und Kunst geförderten Gemeinschaftsprojekts der hessischen Hochschulen zur Bereitstellung einer Datenschutz-konformen Sync&Share-Lösung (**HessenBox**) befinden sich an beiden Universitäten Installationen der Anwendung *PowerFolder*³ zurzeit in der Test- bzw. Pilotphase. In Zusammenarbeit mit dem Verbundprojekt *Hessische Forschungsdateninfrastrukturen* (HeFDI)⁴ befindet sich mit **HeRDMO** eine prototypische Installation der von der DFG geförderten Lösung zum Forschungsdatenmanagement (*Research Data Management Organiser* (RDMO)⁵) ebenfalls in der Testphase. Derzeit befindet sich eine auf **DSpace**⁶ basierende Lösung zur Langzeitarchivierung von Forschungsdaten in Form eines Systems zur sicheren Speicherung digitaler Objekte und der beschreibenden strukturierten Metadaten (Repositorium) im Aufbau. Schließlich stellen die Bibliotheken der beiden beteiligten Universitäten im Rahmen von *Open Access* entsprechende **Publikationsserver**⁷⁸ zur Veröffentlichung von Forschungsergebnissen zur Verfügung, ggf. in Kombination mit den durch die DFG geförderten Open-Access Publikationsfonds⁹¹⁰.

4 NOWA-Ansatz zum Forschungsdatenmanagement im SFB 135

Die folgende Tabelle stellt den Anforderungen der Wissenschaftlerinnen und Wissenschaftler die etablierten und zukünftigen Dienste an den beiden Rechenzentren gegenüber und fasst unseren Ansatz für das aktive und passive Forschungsdatenmanagement im SFB 135 zusammen.

²Planio: <https://plan.io/de/> (Letzter Aufruf am 12.07.2019).

³PowerFolder: <https://www.powerfolder.com/de/> (Letzter Aufruf am 12.07.2019).

⁴HeFDI: <https://www.uni-marburg.de/hefdi> (Letzter Aufruf am 12.07.2019).

⁵RDMO: <https://rdmorganiser.github.io/> (Letzter Aufruf am 12.07.2019).

⁶DSpace: <https://duraspace.org/dspace/> (Letzter Aufruf am 12.07.2019).

⁷Giessener Elektronische Bibliothek: <http://geb.uni-giessen.de/geb/> (Letzter Aufruf am 12.07.2019).

⁸Publikationsserver der Philipps-Universität Marburg: <https://archiv.ub.unimarburg.de/ubfind/> (Letzter Aufruf am 12.07.2019).

⁹9OA Publikationsfonds der Justus-Liebig-Universität Gießen: <https://www.unigiessen.de/ub/digitales-publizieren/openaccess/oafonds> (Letzter Aufruf am 12.07.2019).

¹⁰Open-Access-Publikationsfonds der Philipps-Universität Marburg: <https://www.unimarburg.de/de/ub/forschen/open-access/publikationsfonds> (Letzter Aufruf am 12.07.2019).

	Anforderungen	Dienste
<i>Organisation</i>	Kollaboratives Arbeiten Planung und Umsetzung des FDM Versionierung von FD	GitLab, plan.io, HessenBox RDMO git
<i>Optimierung</i>	Verwaltung großer Dateien Veröffentlichung von Ergebnissen und FD Qualitätssicherung von FD	Git LFS Publikationsserver, DSpace CI-Server in GitLab

Den Kern unserer Workflow-Assistenz bildet eine auf dem verteilten Versionskontrollsystem git¹¹ basierende lokale Instanz der Webanwendung **GitLab**¹². Das zentrale Feature einer kontinuierlichen und auszeichnenden Versionierung wird von GitLab um weitere nützliche Funktionen ergänzt. Zum einen beinhaltet GitLab ein System zur kontinuierlichen Integration (CI-Server), mit dessen Hilfe Maßnahmen zur Sicherung der Qualität der Forschungsdaten durchgeführt werden können. Zum anderen bietet GitLab mit Git LFS (Large File Storage)¹³ eine Erweiterung für die Behandlung großer Dateien im Repository. Ebenso ermöglicht GitLab den interdisziplinären Arbeitsgruppen ein kollaboratives Arbeiten, was durch die Projektmanagement-Plattform Planio sowie die Sync&Share-Lösung HessenBox optimal ergänzt werden kann. Abgerundet wird das aktive Forschungsdatenmanagement durch die Verwendung von RDMO, mit dessen Hilfe alle relevanten Planungsinformationen in Datenmanagementplänen erfasst und alle Datenmanagementaufgaben über den gesamten Datenlebenszyklus hinaus verwaltet werden können. Im passiven Forschungsdatenmanagement kann die Veröffentlichung von Forschungsergebnissen und zugehörigen Forschungsdaten dann schließlich über die entsprechenden Publikationsserver in Verbindung mit dem Daten-Repositorium DSpace erfolgen.

Danksagung Die vorliegende Arbeit wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG), Projekt Nr. 222641018, SFB/TRR 135 TP INF.

¹¹Git: <https://git-scm.com/> (Letzter Aufruf am 12.07.2019).

¹²GitLab: <https://about.gitlab.com/> (Letzter Aufruf am 12.07.2019).

¹³Git LFS: <https://git-lfs.github.com/> (Letzter Aufruf am 12.07.2019).