

# Anforderungen an das Forschungsdatenmanagement an einer mittelgroßen Universität und Konzeption einer prototypischen Lösung

Robin Jegan (robin.jegan@uni-bamberg.de), Otto-Friedrich-Universität Bamberg, Deutschland

Tobias Gradl (tobias.gradl@uni-bamberg.de), Otto-Friedrich-Universität Bamberg, Deutschland

Andreas Henrich (andreas.henrich@uni-bamberg.de), Otto-Friedrich-Universität Bamberg, Deutschland

Am Lehrstuhl für Medieninformatik der Otto-Friedrich-Universität Bamberg wird derzeit ein Prototyp zum Management von Forschungsdaten entwickelt (FDMS), der heterogene Daten, die momentan nicht digital katalogisiert werden, speichern, beschreiben, und veröffentlichen soll. Von besonderer Bedeutung sind hierbei die FAIR-Prinzipien, wonach wissenschaftliche Daten auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable), und wiederverwendbar (Re-usable) sein sollten (Wilkinson 2016). Es existiert bereits ein Projekt zur Einführung eines Forschungsinformationssystems (FIS) an der Universitätsbibliothek Bamberg (Franke 2019), jedoch liegt dort der Fokus auf Publikation, Forschenden, und Projekten. Forschungsdaten werden hier bisher nur am Rande als Anhang bei der Publikation von Dissertationen und anderen Arbeiten adressiert. Weitere Forschungsdaten, die mit keiner Publikation zusammenhängen, werden hingegen in diesem FIS nicht miteinbezogen. Aus diesem Grund soll ein FDMS entwickelt werden, welches genau diese Art von Forschungsdaten ohne zugehörige Publikation, sowie auch alle anderen Forschungsdaten abdeckt.

Eine wichtige Anforderung und Herausforderung für das FDMS stellt die Heterogenität der Daten dar. Diese Heterogenität manifestiert sich durch die unterschiedlichen Fachrichtungen der Lehrstühle an der Universität Bamberg und dementsprechend in einer Vielzahl an Metadatenschemata und Datenformaten. Die Erschließung und weitere Verwendung dieser Metadaten ist von großer Bedeutung, da die effiziente Suche sowie weitere Dienste, wie Filterung oder Visualisierung, von umfangreichen Metadaten profitieren (Neuroth 2017) und zudem Fördergeber wie die Deutsche Forschungsgemeinschaft Anforderungen an ein Forschungsdatenmanagement stellen, das den FAIR-Prinzipien folgt.

Für die erste Umsetzung wurde DSpace verwendet (Smith 2003, Donohue 2018), eine Software zur Verwaltung von digitalen Forschungs- und Lehrmaterialien. DSpace wird in Bamberg auch zur Realisierung des FIS eingesetzt, dort jedoch in der Erweiterung als DSpace-CRIS (Donohue 2019). Während der Installation und Anpassung von DSpace an die Testdaten wurde eine Einschränkung und zugleich ein Nachteil von DSpace deutlich, welche die Nutzung dieser Software für den Zweck eines FDMS ungeeignet erscheinen lassen. Der Import von unterschiedlichen Metadatenschemata für Forschungsdaten wird in DSpace zwar unterstützt, jedoch ist dieser Import in der DSpace Standard-Installation nur mit den originalen Dublin-Core Elementen eingerichtet. Außerdem können keine hierarchischen Datenelemente erzeugt werden, weswegen der Import von Daten sowie deren Metadaten, welche in hierarchischen XML-Dateien vorliegen, verhindert wird.

Aufgrund dieser Einschränkung von DSpace, welche sich bei einer Vielzahl von Forschungsdaten als restriktiv darstellen würde und somit einer der Anforderungen an ein FAIR-basiertes FDMS widerspricht, wurde eine alternative Software benötigt und in Dataverse gefunden (Dataverse 2019a). Dataverse, als Projekt vom Harvard Institute for Quantitative Social Science, der Harvard Library, und anderen Partnern initiiert (Dataverse 2019b), ermöglicht die Modellierung der Metadaten in mehreren Varianten. Zum einen können über die Administrationsoberfläche der Dataverse Installation händisch Elemente der bestehenden Metadatenschemata angepasst werden. Zum anderen, und für das Umsetzungsszenario der Universität Bamberg passender, kann ein sogenannter Metadatenblock angelegt werden, mit dem ein komplett neues Metadatenchema für Datenobjekte in der Dataverse Installation zur Verfügung gestellt wird (Dataverse 2019c). Dieser selbst erzeugte Metadatenblock, in Form einer TSV (Tab-Separated Values) Datei, kann mithilfe eines Kommandozeilen Befehls importiert werden und steht daraufhin neben den in Dataverse bereits vorgefertigten Standard-Metadaten schemata zur Verfügung.

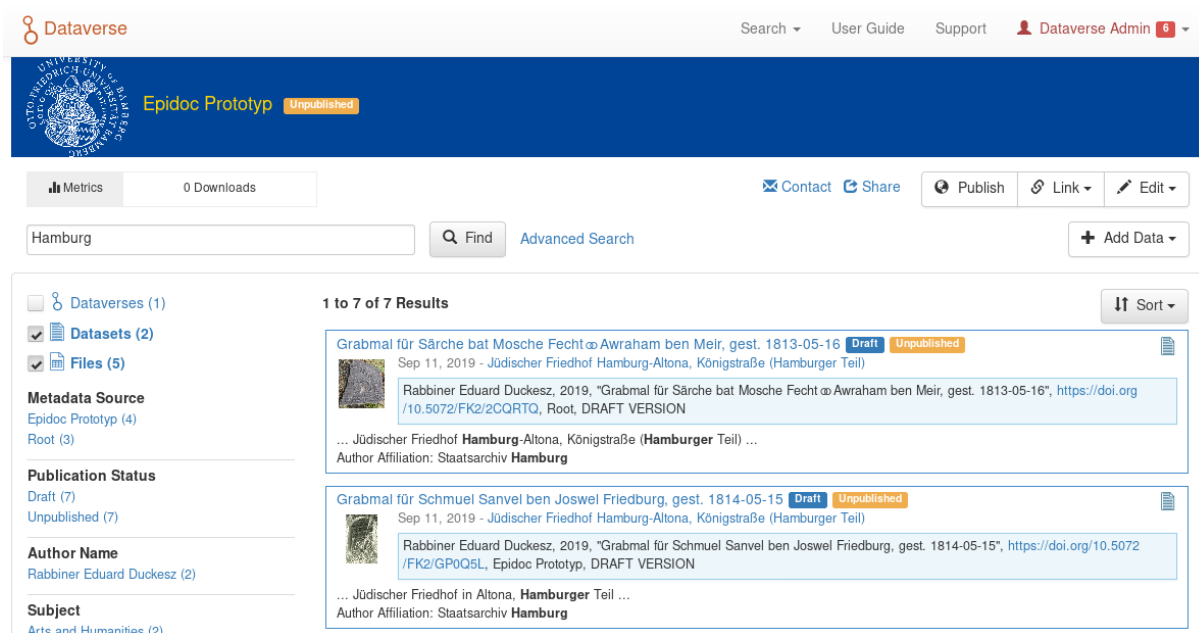


Abbildung 1. Oberfläche des FDMS-Prototyps mit einer Suchanfrage und der Ergebnisseite.

Als prototypische Anwendung wurden Bilder von Grabsteinen mit weiteren zugehörigen Forschungsprimärdaten verwendet, die mehrere Eigenschaften eines typischen Use Cases für das künftige FDMS abdecken. Diese Forschungsdaten liegen zum einen als Bilder vor, die ohne zugehörige Publikation bisher nicht in einem FIS abgespeichert werden können, zum anderen sind komplexe Metadaten vorhanden, welche in DSpace, wie oben beschrieben, nur eingeschränkt darzustellen, in Dataverse jedoch besser einzufügen sind. Die Testdaten für diesen Prototyp basieren auf dem epidat Projekt des Steinheim-Instituts (Steinheim-Institut 2019), die dem Datenbestand der Professur für Judaistik der Universität Bamberg ähneln, welcher ein konkretes künftiges Anwendungsszenario darstellt.

Auch für die Analyse der Metadatenelemente, die im FDMS verwendet werden, wurde das epidat Projekt als Testfall genutzt. Hierzu wurden die XML-Dateien, welche die Gräber einer Vielzahl von Friedhöfen in Deutschland mit Metadaten im Epidoc-TEI-Format (Elliott 2006-2017) beschreiben, analysiert und relevante Metadatenfelder identifiziert. Diese relevanten Metadatenelemente wurden als Basis für ein Dataverse im Kontext von Fotografien von Grabsteinen verwendet. Der Begriff Dataverse umfasst dabei nicht nur die Software Dataverse, sondern dient auch als Oberbegriff für eine Sammlung von Datensätzen (King 2007). Die Vorgehensweise zur Erstellung eines eigenen

Metadaten-Schemata wurden verwendet, da das im epidat-Projekt verwendete Epidoc-Format sehr umfangreich und dementsprechend für ein initiales Anwendungsszenario zu mächtig scheint. Die Anzahl der Elemente für den Prototyp wurde daher eingeschränkt.

Die identifizierten Metadaten-Elemente, welche für die forschungsorientierte Beschreibung der Grabsteine notwendig sind, wurden mithilfe eines der oben erwähnten Metadatenblöcke modelliert. Das hiermit erstellte Metadaten-Schema „epitaph“ – betitelt in Anlehnung an den Fachbegriff für Grabinschriften – enthält 25 Elemente, die unter anderem den zugehörigen Friedhof, den Zustand des Grabmals, die Inschrift inklusive Übersetzung, und weitere Datenfelder umfassen. Weiterhin wurden hierarchische Beziehungen zwischen den Elementen aufrechterhalten, beispielsweise für die verstorbene Person, deren Todestag, und die erwähnten Verwandten dieser Person über die Datenfelder „person“, „personDeathdate“, und „personRelationship“.

In Kooperation mit dem Rechenzentrum und der Universitätsbibliothek wird aufbauend auf internen Diskussionen die Weiterentwicklung des in diesem Poster vorgestellten Prototyps geplant, welche in einer größeren Testphase 2020 vorgenommen werden soll. Die Anbindung an das bereits im Betrieb befindliche FIS soll ebenso wie die Integration der lokalen Shibboleth-Authentifizierungsverfahren dort umgesetzt werden.

Das Poster verdeutlicht die Anforderungen und Anstrengungen, die für die Konzeption eines FDMS an einer mittelgroßen Universität notwendig sind, sowie erste positive Ergebnisse. Weiterhin werden durch die Betrachtung zweier Softwarelösungen die Probleme in der Praxis, also in der Umsetzung eines derartigen Forschungsdatenmanagements, näher beleuchtet. Die Erfahrungen verdeutlichen wie wichtig eine systematische Anforderungsanalyse bei der Auswahl eines Systems zum Forschungsdatenmanagement ist. Hierbei sollten insbesondere Aspekte wie die Unterstützung unterschiedlicher Metadatenformate und deren hierarchische Ausprägungen berücksichtigt werden. Diese und weitere Erfahrungen werden am Poster geteilt und ausgetauscht.

Dataverse (2019): The Dataverse Project: Open source research data repository software. Online: <https://dataverse.org/> [letzter Zugriff: 19.12.2019]

Dataverse (2019): The Dataverse Project: About the project. Online: <https://dataverse.org/about> [letzter Zugriff: 20.12.2019]

Dataverse (2019): The Dataverse Project Admin Guide: Metadata Customization. Online: <http://guides.dataverse.org/en/latest/admin/metadatatocustomization.html> [letzter Zugriff: 19.12.2019]

Donohue, Tim (2018): DSpace User FAQ. Online: <https://wiki.lyrasis.org/display/DSPACE/User+FAQ> [letzter Zugriff: 20.12.2019]

Donohue, Tim (2019): DSpace-CRIS Home. Online: <https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home> [letzter Zugriff: 19.12.2019]

Elliott, Tom / Bodard, Gabriel / Cayless, Hugh / et al. (2006-2017): EpiDoc: Epigraphic Documents in TEI XML. Online: <https://sourceforge.net/p/epidoc/wiki/Home/> [letzter Zugriff: 19.12.2019]

Franke, Fabian (2019): Laufende Projekte unter Beteiligung der Universitätsbibliothek Bamberg. In: Projekte – Otto-Friedrich-Universität Bamberg. Online: <https://www.uni-bamberg.de/ub/ueberuns/projekte/> [letzter Zugriff: 19.12.2019]

King, Gary (2007): "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing". In: Sociological Methods & Research 36(2): 173–199.

Neuroth, Heike (2017): "Bibliothek, Archiv, Museum". In: Digital Humanities: 213–222.

Smith, MacKenzie / et al. (2003): "An Open Source Dynamic Digital Repository". In: D-Lib Magazine 9.1.

Steinheim-Institut (2019): Datenbank: Jüdische Grabsteinepigraphik. Online: <http://www.steinheim-institut.de/cgi-bin/epidat> [letzter Zugriff: 20.12.2019]

Wilkinson, Mark D., et al. (2016): "The FAIR Guiding Principles for scientific data management and stewardship". In: Scientific data 3.