
Datenmodellierung für Forschungsdatenmanagementpläne

Masterarbeit an der Fachhochschule Potsdam am Fachbereich 5 Informationswissenschaften

Eingereicht von:

Martin Heger (Matr.-Nr.: 11151)

Syrische Straße 10

13349 Berlin

martin.heger@fh-potsdam.de

Erstgutachterin: Prof. Dr. Heike Neuroth, Fachhochschule Potsdam (FHP)

Zweitgutachter: Dr. Jochen Klar, Leibniz-Institut für Astrophysik Potsdam (AIP)

15.08.2016, Potsdam

Inhalt

1	Einführung.....	1
1.1	Forschungsdaten/Forschungsdatenmanagement.....	3
2	Forschungsdatenmanagementpläne.....	6
2.1	Horizon 2020.....	8
2.2	Open Research Data Pilot.....	9
3	Übersicht DMP-Tools Deutschland/weltweit.....	11
3.1	Bewertungsmatrix.....	13
4	Bestandsanalyse.....	24
4.1	Zielsetzung/Scope des DFG-Projekts.....	26
5	Einbettung der Arbeit in den Kontext des DFG-Projekts RDMO.....	28
6	Vorgehen/Methodik.....	29
6.1	Verwendete Werkzeuge und Bibliotheken.....	30
6.2	Ausgangslage.....	31
7	Ergebnisse.....	34
7.1	Attribute und Entitäten.....	34
7.2	Metadaten.....	37
7.2.1	RDMO-Mapping.....	39
7.3	Technische Implementierung.....	58
8	Diskussion.....	66
9	Ausblick.....	69
	Quellenverzeichnis.....	72
	Abbildungsverzeichnis.....	78
	Danksagung.....	79
	Anlagen.....	80

1 Einführung

Wissenschaft und Forschung sind in Deutschland traditionell stark differenzierte Bereiche. In Deutschland gilt das Prinzip der wissenschaftlichen Selbstverwaltung und -organisation unter der Schirmherrschaft der zentralen Selbstverwaltungsorganisation der deutschen Wissenschaft in Form der Deutschen Forschungsgemeinschaft e. V. (DFG)¹. Der Förderkatalog der Bundesregierung² offenbart eine Vielzahl von gegenwärtig laufenden Forschungsprojekten und eine vermehrte Investition von Fördergeldern in die Forschungsförderung. Doch mangelt es oft noch an einer ausreichenden digitalen Infrastruktur, die es u. a. ermöglicht die Forschungsdaten und Ergebnisse auszutauschen sowie langfristig zu sichern.³ Die wissenschaftspolitische Forderung für Forschungsdaten spricht von einer Zugänglichmachung der angefallenen Daten, ihrer sachgemäßen Verwaltung sowie einer angemessenen Dokumentation. Dies soll zu einer besseren Nachprüfbarkeit der Ergebnisse führen und sicherstellen, dass die Forschungsdaten auch langfristig aufbewahrt, operational zugänglich gemacht werden und somit für nachfolgende Projekte nutzbar sind. Die Umsetzung dieser Forderungen kann nur erfolgen, wenn von den Forschenden, den beteiligten Institutionen, dem Gesetzgeber sowie den Forschungsförderern gemeinsame Ziele und diesen Zielen entsprechende Strategien, Werkzeuge und Standards entwickelt werden.

Diskussionen und Bestrebungen der letzten Jahre hinsichtlich dieser Fragestellungen und Ziele mündeten schlussendlich in der Entwicklung und zunehmend größeren Verbreitung von Datenmanagementplänen (DMP). Datenmanagementpläne sind ein wichtiges Element der gesamten Entwicklung und der Forschungsdateninfrastruktur insgesamt. Datenmanagementpläne können ein Forschungsprojekt schon ab einem frühen Stadium begleiten und zur systematischen Beschreibung von Strategien und Maßnahmen zum Umgang mit Forschungsdaten während und auch nach Beendigung des Projekts dienen. Inzwischen verlangen EU-Forschungsförderprogramme wie Horizon 2020⁴ den Einsatz und die Benutzung von Datenmanagementplänen, wodurch sich eine deutlich steigende Nachfrage nach entsprechenden Werkzeugen herausgebildet hat. Im angloamerikanischen Bereich und insbesondere in Australien wird der Einsatz von Datenmanagementplänen bereits jetzt schon stark fokussiert.

Die vorliegende Masterarbeit ist im Kontext des DFG-Projektes „Entwicklung und Implementierung eines Werkzeugs für die Planung, Umsetzung und Kontrolle des Forschungsdatenmanagements“ entstanden. Dieses Projekt läuft nach derzeitiger Planung (Stand: 15.08.2016) bis April

¹ Deutsche Forschungsgemeinschaft e. V. URL: <http://www.dfg.de/>.

² BMBF: Förderkatalog. URL: <http://foerderportal.bund.de/foekat/jsp/StartAction.do>.

³ Herbold: Europas digitales Gedächtnis ist löchrig. In: Der Tagesspiegel, online verfügbar unter URL: <http://www.tagesspiegel.de/wissen/forschungsdaten-von-akademien-europas-digitales-gedaechtnisist-loechrig/12906856.html>.

⁴ European Commission: Horizon 2020. URL: <https://ec.europa.eu/programmes/horizon2020/>.

2017 und entwickelt ein Datenmanagementplanwerkzeug: Research Data Management Organiser (RDMO)⁵. Das fertige Datenmanagementplanwerkzeug soll Forschende, Antragstellende, Forschungsfördernde und weitere wichtige Akteure im Forschungsdatenmanagement bei der Verwaltung von Forschungsdaten unterstützen, begleiten und darüber hinaus auch die Möglichkeit zur textuellen Ausgabe eines Forschungsdatenmanagementplans als Beilage zur Einreichung eines Forschungsantrages bieten.

Der Fokus dieser Masterarbeit liegt dabei auf der Schaffung von Grundlagen zur Interoperabilität von Forschungsdatenmanagementplänen, die mit dem Datenmanagementplanwerkzeug des genannten DFG-Projektes erstellt werden. Dies wird einerseits durch eine durchdachte Datenmodellierung erreicht, die sinnvolle Attribute und ein passendes Vokabular beim Fragenkatalog des Werkzeuges zum Forschungsdatenmanagement benutzt. Andererseits werden damit die Weichen zur Erstellung eines Mappings gestellt, welches dem Austausch zwischen verschiedenen Datenmanagementplanwerkzeugen dient und auch Schnittstellen zu anderen Programmen ermöglicht. Weiterhin wird für das Werkzeug eine Export-Funktion programmiert, die die eingetragenen Daten im XML-Format ausgibt. Dabei wird insgesamt auch auf die Zusammenführung und Implementation dieser inhaltlichen und technischen Funktionalitäten eingegangen.

Zunächst wird eine Einführung in die Begrifflichkeiten Forschungsdaten und Forschungsdatenmanagementpläne gegeben. Danach erfolgt ein Blick auf den Einsatz von Datenmanagementplänen im EU-Forschungsförderprogramm Horizon 2020 und eine kritische Beobachtung und Auseinandersetzung mit den am Markt derzeit vorhandenen Datenmanagementplanwerkzeugen. Beschrieben wird – vor allem im Kontext der Anforderungen der verschiedenen Forschungsdatenmanagementakteure –, welche Funktionalitäten von welchem Werkzeug abgedeckt werden, welche Funktionalitäten noch ausbaufähig sind und welche Features noch nicht umgesetzt wurden.

An dieses Kapitel schließt die Vorstellung des entwickelten Datenmanagementplanwerkzeuges des DFG-Projektes an. Im Kontext der Hausarbeit zur Marktsichtung von Datenmanagementplänen mit dem Titel „Datenmanagementpläne – Eine Bestandsübersicht“, die im Rahmen des Moduls „Konzeptionelle Entwicklung eines Werkzeugs für die Planung des Forschungsdatenmanagements“ unter Leitung von Prof. Dr. Neuroth im Wintersemester 2015 an der Fachhochschule Potsdam von Heger und Heinrich erstellt und eingereicht wurde, wird deren Ergebnis reflektiert und eine Problemstellung wird formuliert. Den Erläuterungen zur Zielsetzung und des vom Projekts angestrebten Wirkungsbereichs folgt darauf aufbauend die Formulierung der Methodik und des Vorgehens beim Mitwirken des Autors an diesem DFG-Projekt und der daraus resultierenden Masterarbeit.

⁵ AIP: RDMO - Research Data Management Organiser. URL: <http://rdmorganiser.github.io/>.

1.1 Forschungsdaten/Forschungsdatenmanagement

Es gibt keine einheitliche Definition von Forschungsdaten. Die Deutsche Forschungsgemeinschaft e. V. (DFG) definiert den Begriff in ihren Leitlinien⁶ zum Umgang mit Forschungsdaten wie folgt:

„Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“⁷

Weiterhin heißt es dort:

„Die langfristige Sicherung und Bereitstellung der Forschungsdaten leistet einen Beitrag zur Nachvollziehbarkeit und Qualität der wissenschaftlichen Arbeit und eröffnet wichtige Anschlussmöglichkeiten für die weitere Forschung.“⁸

Das Management dieser oftmals stark heterogenen Forschungsdaten ist dabei sehr von der Wissenschaftsdisziplin abhängig. Einheitliche Modelle bzw. Verfahrensweisen sind nicht vorhanden.⁹ Eine Grundlage bieten die eingangs genannten Leitlinien zum Umgang mit Forschungsdaten der DFG. Weiter heißt es dort, dass idealerweise bereits bei der Antragsstellung Angaben zu den möglichen Forschungsdaten und deren Qualitätssicherung gemacht werden sollen. Auch die Allianzinitiative (Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisation)¹⁰ rät in der Präambel¹¹ zu ihren Grundsätzen zum Umgang mit Forschungsdaten:

„Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und können unabhängig von ihrem ursprünglichen Erhebungszweck vielfach Grundlage weiterer Forschung sein. [...] Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient daher nicht nur der Prüfung früherer Ergebnisse, sondern in hohem

⁶ DFG: Umgang mit Forschungsdaten. URL: http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung_nachnutzung_forschungsdaten/.

⁷ Ebd.

⁸ Ebd.

⁹ Büttner, Hobohm, Müller (Hrsg.): Handbuch Forschungsdatenmanagement. S. 7.

¹⁰ Schwerpunktinitiative „Digitale Information“. URL: <http://www.allianzinitiative.de/>.

¹¹ Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V.: Grundsätze zum Umgang mit Forschungsdaten. URL: <http://www.allianzinitiative.de/handlungsfelder/forschungsdaten/grundsätze/>.

Maße auch der Erzielung künftiger Ergebnisse. Sie bildet eine strategische Aufgabe, zu der Wissenschaft, Politik und andere Teile der Gesellschaft gemeinsam beitragen müssen.“¹²

Die Europäische Kommission gibt ebenfalls entsprechende Richtlinien zum Forschungsdatenmanagement in Form des EU-Forschungsrahmenprogrammes Horizon 2020 vor¹³. Basierend auf diesem EU-Förderprogramm werden in der dazugehörigen Initiative des Open Data Research Pilots (ODRP)¹⁴ detaillierte Datenmanagementpläne gefordert, durch den Antragsteller bereits zu Beginn der Antragsphase erstellt und in der Folge kontinuierlich aktualisiert und erweitert. Bezogen auf den Lebenszyklus von Forschungsdaten¹⁵ befindet sich diese Aktivität am Anfang und somit noch vor dem Entstehen der eigentlichen Daten. Trotzdem sind alle Aspekte des gesamten Kreislaufs betroffen, da diese Überlegungen sämtliche Aspekte der Datenerstellung, -verarbeitung, -analyse, -archivierung, -zugänglichmachung und -nachnutzung beschreiben (vgl. Abb. 1).



Abbildung 1: Lebenszyklus von Forschungsdaten

¹² Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V.: Grundsätze zum Umgang mit Forschungsdaten. URL: <http://www.allianzinitiative.de/handlungsfelder/forschungsdaten/grundsaeetze/>.

¹³ European Commission: Guidelines on Data Management in Horizon 2020. URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-datamgt_en.pdf.

¹⁴ OpenAIRE: What is the Open Research Data Pilot? URL: <https://www.openaire.eu/opendatapilot>.

¹⁵ DAI: Der Lebenszyklus von Forschungsdaten. URL: <http://www.ianus-fdz.de/it-empfehlungen/lebenszyklus>.

Die Unterstützung der Forschenden beim Forschungsdatenmanagement kann auf verschiedene Arten erfolgen: Leitfäden und andere Informationsmaterialien können einen ersten Einblick in die Notwendigkeit und die Anwendung von Forschungsdatenmanagement geben. Diese Kenntnisse können durch Beratung, z. B. von Datenmanagern, noch weiter vertieft werden. Damit die tägliche Forschungsarbeit nicht unter dem zusätzlichen Aufwand des Managements der Daten leidet, ist der Schritt zum Embedded Data Management¹⁶ eine sinnvolle Option. Die Einbindung von recherche- und informationstechnisch erfahrenen Datenmanagern in die Forschungs- und Arbeitsabläufe direkt vor Ort, erlaubt die Entwicklung von individuellen Lösungsmöglichkeiten für das Forschungsdatenmanagement.

Ein weiteres Hilfsmittel, welches beim Forschungsdatenmanagement zunehmend zum Einsatz kommt, sind Werkzeuge bzw. Tools, die bei der Erstellung eines Datenmanagementplans assistieren. Bekannte Tools sind z. B. DMPonline¹⁷ von der DCC (Digital Curation Center)¹⁸ in Großbritannien und das DMPTool¹⁹ der California Digital Library (CDL)²⁰.

¹⁶ Cremer, Engelhardt, Neuroth: Embedded Data Manager. DOI: 10.1515/bfp-2015-0006.

¹⁷ Digital Curation Centre: DMPonline. URL: <https://dmponline.dcc.ac.uk/>.

¹⁸ DCC: The Digital Curation Centre. URL: <http://www.dcc.ac.uk/>.

¹⁹ California Digital Library: Data Management Planning Tool. URL: <https://dmptool.org/>.

²⁰ California Digital Library. URL: <http://www.cdlib.org/>.

2 Forschungsdatenmanagementpläne

Forschungsdatenmanagementpläne können zur Unterstützung von Fachwissenschaftlern bei der Dokumentation und den digitalen Forschungsprozessen dienen. Sie sind mehr als nur eine Anforderung des Forschungsförderers, sondern essentiell für die Optimierung des Datenmanagements von Projektbeginn bis zum -ende und auch darüber hinaus. Sie helfen bei der systematischen Beschreibung von Strategien und Maßnahmen und zeigen als Leitfaden auf, wie mit Forschungsdaten allgemein oder auch speziell in einem Projekt umgegangen wird. Forschungsdatenmanagementpläne stellen zudem übersichtlich dar, wie und mit welchen Mitteln Forschungsdaten während eines Projekts gesichert, verzeichnet, gepflegt, verarbeitet und zugänglich gemacht werden. Zusätzlich werden sämtliche Prozesse und Technologien beschrieben, die während des gesamten Lebenszyklus der Forschungsdaten zum Einsatz kommen. Ziel ist es, diese Prozesse und Technologien sichtbar und nachvollziehbar zu machen. Auch Richtlinien zu Forschungsdaten für verschiedene Interessengruppen, wie z. B. die Planung und Vorbereitungen von Publikationen und der Langzeitaufbewahrung der Daten, werden darin beschrieben und festgelegt. Erst durch einen umfassenden Datenmanagementplan, werden die Forschungsergebnisse für Dritte auch interpretierbar, verifizierbar und letztendlich nachnutzbar, denn der Datenmanagementplan gibt die Richtlinien dafür vor und hilft bei der Organisation einer effizienten Arbeit in diesem Sinne. Somit kommen detaillierte Forschungsdatenmanagementpläne sowohl den Forschenden, der Institution, als auch dem Forschungsförderer zu Gute. Sie dienen langfristig zur Verbesserung der Qualität und Effizienz der wissenschaftlichen Arbeit. Wichtig bei der Entwicklung von Forschungsdatenmanagementplänen ist eine durchdachte und flexible Datenmodellierung. Damit soll einerseits eine Zusammenführung von mehreren Forschungsdatenmanagementplänen ermöglicht werden, andererseits die reibungslose Einbindung und Anpassung von Forschungsdatenmanagementplänen ans eigene institutionelle Umfeld und die jeweilige Forschungsdomäne sichergestellt werden. Hinweise und Empfehlungen zu den entsprechenden nationalen und internationalen Metadatengepflogenheiten und damit einen wichtigen Beitrag zur Anbindung von Community-Metadaten in Deutschland und der Sicherstellung von Interoperabilität von digitalen Informationsbeständen gibt das KIM Kompetenzzentrum Interoperable Metadaten²¹.

Essentielle Fragen zur Erstellung eines DMP finden sich z. B. im Leitfaden²² bzw. der Checkliste²³ von WissGrid²⁴, Horizon 2020, dem DFG-Leitfaden zum Umgang mit Forschungsdaten sowie den

²¹ KIM. URL: <http://www.kim-forum.org>.

²² Ludwig, Enke (Hrsg.): Leitfaden zum Forschungsdaten-Management. URL: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf.

²³ WissGrid: Checkliste zum Forschungsdaten-Management. URL: <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Checkliste-Forschungsdaten-Management.pdf>.

²⁴ Georg-August-Universität Göttingen: WissGrid. URL: <http://www.wissgrid.de/>.

Checklisten zur DMP-Erstellung des Digital Curation Centre (DCC)²⁵ und der National Science Foundation²⁶ (NSF)²⁷. Nachfolgend sollen nur exemplarisch einige aufgeführt werden:

- Um was für ein Projekt handelt es sich?
- Welche Daten werden erzeugt und genutzt?
- Um welche Art(en) von Daten handelt es sich?
- Welche Daten sollen oder müssen aufbewahrt werden, und warum?
- Sind Zusatzinformationen für das Verstehen der Daten notwendig?
- Wann erfolgt die Datenauswahl?
- Wie lange sollen die Daten aufbewahrt werden (Archivierung)?
- Wann werden die Daten übergeben (Datenaustausch, Datenpublikation)?
- Wer darf die Daten nutzen?
- Welche Kosten entstehen?
- Welche Ressourcen werden benötigt?

Der Einsatz von Forschungsdatenmanagementplänen und den damit verbundenen Tools wird inzwischen von EU-Forschungsförderprogrammen wie Horizon 2020 gefordert, welches im folgenden Kapitel näher erläutert wird.

²⁵DCC: Checklist for a Data Management Plan. URL: <http://www.dcc.ac.uk/resources/data-management-plans/checklist>.

²⁶ NSF: National Science Foundation. URL: <https://www.nsf.gov/>.

²⁷ Princeton University Library: NSF Data Management Plan. URL: http://libguides.princeton.edu/ld.php?content_id=2940897.

2.1 Horizon 2020

Bei Horizon 2020 (dt.: Horizont 2020) handelt es sich um ein EU-Rahmenprogramm für Forschung und Innovation. Er setzt damit das 7. EU-Forschungsrahmenprogramm (7. FRP)²⁸ fort und integriert zugleich das Europäische Innovations- und Technologieinstitut²⁹ (EIT) und die Innovationselemente des Rahmenprogramms für Wettbewerbsfähigkeit und Innovation³⁰ (CIP).³¹ Horizon 2020 ist ein Teil der vom Europäische Rat im Jahr 2010 verabschiedeten Europa 2020 Strategie³². Zudem ist Horizon 2020 Teil der Umsetzung des Europäischen Forschungsraums (EFR)³³. Das Budget beträgt rund 80 Milliarden Euro über einen Zeitraum von sieben Jahren (2014-2020) und stellt damit das finanzstärkste Forschungsförderprogramm in Europa dar.

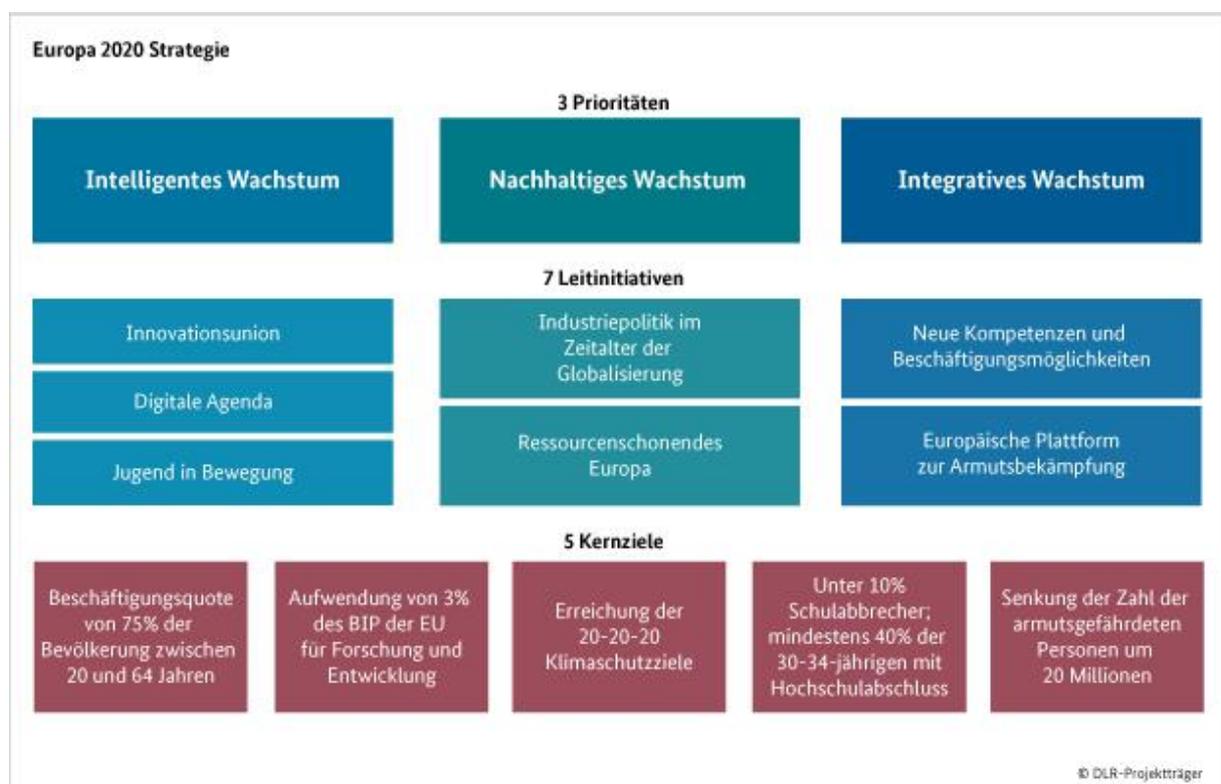


Abbildung 2: Europa 2020 Strategie

Horizon 2020 besitzt, noch deutlicher als die vorherigen Forschungsrahmenprogramme, eine große Ausrichtung auf Innovation und führt damit den Prozess der bisherigen FRP fort. Neben der Grundlagenforschung und den daraus resultierenden Ergebnissen, wird daher zudem sehr

²⁸ BMBF: 7. EU-Forschungsrahmenprogramm im Überblick. URL: <http://www.forschungsrahmenprogramm.de/frp-ueberblick.htm>.

²⁹ European Institute of Innovation & Technology: EIT. URL: <https://eit.europa.eu/>.

³⁰ Europäische Kommission: Rahmenprogramm für Wettbewerbsfähigkeit und Innovation (CIP). URL: http://ec.europa.eu/cip/index_de.htm.

³¹ BMBF: Programmaufbau von Horizon 2020. URL: <http://www.horizont2020.de/einstieg-programmaufbau.htm>.

³² European Commission: Europa 2020. URL: http://ec.europa.eu/europe2020/index_de.htm.

³³ BMBF: Der Europäische Forschungsraum. URL: <http://www.horizont2020.de/einstieg-era.htm>.

darauf geachtet, wie die aus Projekten entstandenen Daten und Ergebnisse aufbewahrt und weiterverwendet werden können. Zusätzlich zu den Teilen der gesellschaftlichen Herausforderungen und der führenden Rolle der Industrie, spielt die Wirtschaftsexzellenz eine große Rolle. Diese soll vor allem Forschungsstrukturen aufbauen und verstärken, in dem z. B. eine bessere Vernetzung der Forschenden untereinander und Forschungsaufenthalte im Ausland stattfinden. Den Forschenden sollen zudem ausgezeichnete Forschungsinfrastrukturen zugänglich gemacht werden. Darüber hinaus sollen durch den Ausbau von forschungs- und innovationsschwächeren Regionen Wissenschaft, Forschung und Innovation noch stärker in der Gesellschaft verankert werden.³⁴

2.2 Open Research Data Pilot

Eine der innovativen Komponenten von Horizon 2020 ist die Initiative des Pilotprogramms Open Research Data Pilot, der zusätzlich zur allgemeinen Open-Access-Verpflichtung in Horizon 2020 existiert. Ziel des Open Research Data Pilots ist es, die von einem Projekt generierten und zur Validierung der Ergebnisse notwendigen Forschungsdaten besser zugänglich und nachnutzbar zu machen. Die rechtliche Grundlage dafür wurde mit dem Artikel 29.3 im Model Grant Agreement geschaffen. Dort wird die Erstellung eines DMP als zwingend notwendig vorgeschrieben. Außerdem müssen sämtliche angefallenen Daten inklusive Metadaten in einem Datenrepositorium kostenlos für Dritte verfügbar, nachnutzbar, reproduzierbar und verbreitbar (z. B. durch Veröffentlichung unter einer Creative Commons License wie CC-BY³⁵ oder CC0³⁶) abgelegt werden. Die Teilnahme am Open Research Data Pilot und somit an den verpflichtenden Vorgaben, erfolgt für einige Forschungsfelder automatisch, u. a. neue und zukünftige Technologien, Forschungsinfrastrukturen bzw. e-Infrastrukturen und gesellschaftliche Herausforderungen. Einreichungen in teilnehmenden Forschungsbereichen müssen bereits beim Projektantrag auf Planungen und Überlegungen zum anvisierten Datenmanagement eingehen. Auch eine freiwillige Teilnahme am Open Research Data Pilot ist möglich.³⁷

Eine Nicht-Teilnahme (Opt-out) am Open Research Data Pilot ist ebenfalls möglich. Gründe hierfür können u. a. der Schutz der erzielten Ergebnisse sein, falls davon ausgegangen werden kann, dass sie kommerziell oder industriell ausgenutzt werden könnten oder generelle Datenschutz-

³⁴ BMBF: Horizont 2020 - das europäische Forschungsrahmenprogramm. URL:

<https://www.bmbf.de/de/horizont-2020-das-europaeische-forschungsrahmenprogramm-281.html>.

³⁵ CC: Namensnennung 2.0 Deutschland. URL: <https://creativecommons.org/licenses/by/2.0/de/>.

³⁶ CC: Public Domain Dedication. URL: <https://creativecommons.org/publicdomain/zero/1.0/deed.de>.

³⁷ FFG: Open Data in Horizon 2020. URL: https://www.ffg.at/europa/recht-finanzen/h2020-open_data.

bedenken bestehen. Die Teilnahme am Open Research Data Pilot bei der Einreichung des Projektantrags ist ausdrücklich nicht Teil der Evaluierung und ein Opt-out hat keinen negativen Einfluss auf die Bewertung und Bewilligung des Projektantrags.³⁸

³⁸ EC: Guidelines on FAIR Data Management in Horizon 2020. URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, S. 3.

3 Übersicht DMP-Tools Deutschland/weltweit

Vorreiter bei DMP-Tools sind das Digital Curation Center (DCC) in Großbritannien, die California Digital Library (CDL) sowie die Universität Bielefeld. Andere Einrichtungen bieten häufig nur reine Checklisten und keine eigenen Tools für die DMP-Erstellung an.

Im Rahmen der Hausarbeit mit dem Titel „Datenmanagementpläne – Eine Bestandsübersicht“, die im Rahmen des Moduls „Konzeptionelle Entwicklung eines Werkzeugs für die Planung des Forschungsdatenmanagements“ unter Leitung von Prof. Dr. Neuroth im Wintersemester 2015 an der Fachhochschule Potsdam von Heger und Heinrich eingereicht wurde, wurden verschiedene DMP-Tools untersucht. Es handelt sich dabei nicht um eine quantitative Analyse. Stattdessen wurden bekannte und häufig verwendete DMP-Tools mittels Internetrecherche ermittelt und überprüft, ob eine detailliertere Analyse sinnvoll erscheint. Diese Entscheidung wurde primär durch die Zugänglichkeit beeinflusst, denn einige Tools (das DMPonline der Universität Queensland in Australien und Research Data Footprints der Deakin Universität, ebenfalls in Australien) lassen sich nur von Universitätsangehörigen benutzen und konnten daher bei der Analyse von vornherein nicht berücksichtigt werden.

Die folgende Auflistung zeigt alle gefundenen DMP-Tools:

- Clarin-D (DataWizard)
- Data Management Planning Tool (Queensland Universität für Technologie Brisbane, Australien)
- DMP Assistant (Universität Alberta, Kanada)
- DMPonline (DCC)
- DMPonline (Universität Queensland, Australien)
- DMPTool (Universität Kalifornien)
- DMP Tool (Universität Bielefeld)
- IEDA (Interdisciplinary Earth Data Alliance) DMP-Tool
- LabArchives
- Research Data Footprints (Deakin Universität, Australien)

Die abschließende Auswahl zur Analyse erfolgte anhand des jeweiligen Entstehungskontextes bzw. der Zielsetzung und den damit verbundenen unterschiedlichen Schwerpunkten der gefundenen DMP-Tools. Dadurch sollte eine möglichst große Bandbreite an DMP-Tools abgedeckt werden. Nachfolgende DMP-Tools wurden letztendlich für die nähere Untersuchung ausgewählt:

- DMPTool (Universität Bielefeld)
- DMPonline (DCC)
- Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool
- LabArchives

Das DMP-Tool der Universität Bielefeld ist im institutionellen Kontext entstanden und orientiert sich vom Aufbau und der Funktionsweise her am DMPTool der Universität Kalifornien. Das DMP-Tool des Digital Curation Centres besitzt hingegen einen nationalen Entstehungshintergrund. Die beiden genannten DMP-Tools sind domänenübergreifend, daher wurde mit dem DMP-Tool der Interdisciplinary Earth Data Alliance (IEDA) zusätzlich ein Tool mit eher fachspezifischem Fokus untersucht. Bei LabArchives handelt es sich schließlich um eine Art virtuelles Laborbuch und nicht um ein DMP-Tool im eigentlichen Sinne. Dennoch sollte überprüft werden, ob dieser Dienst einen potentiellen Mehrwert für die Entwicklung neuer DMP-Tools bieten kann bzw. generell für die Verwaltung von Daten und Dateien sowie zum kollaborativen Arbeiten geeignet sein könnte. Diese vier ausgewählten DMP-Tools wurden anschließend in einer tabellarischen Bewertungsmatrix hinsichtlich verschiedener Kriterien untersucht. Dabei wurden zunächst allgemeine Gesichtspunkte untersucht, wie z. B. der Entstehungskontext des jeweiligen DMP-Tools und die primäre Zielgruppe. Bei den Kriterien zur technischen Infrastruktur wurde – soweit ermittelbar – u. a. geprüft, ob der Dienst lokal oder zentral betrieben wird und ob eine Art der Gewährleistung für die Datensicherheit übernommen wird. Der Schwerpunkt der Untersuchung der DMP-Tools lag auf den Funktionalitäten und dem Inhalt. Dort wurde einerseits die grundlegende Bedienung untersucht und in welchen Sprachen das jeweilige DMP-Tool vorliegt. Andererseits wurde u. a. verglichen, ob ein kollaboratives Arbeiten möglich ist, mehrere DMP verwaltet werden können und ob z. B. Templates und Metadatenstandards zur Verfügung stehen. Auch, ob für die Benutzung der DMP-Tools Kosten entstehen, wurde überprüft. Damit folgte die Auswahl der Kriterien den aus der Sicht der Autoren wichtigsten und auch möglichst überprüfbaren Blickwinkeln auf die Belange von Forschenden bei der Auswahl eines DMP-Tools.

3.1 Bewertungsmatrix

Die folgende Bewertungsmatrix ist ein Teil der Hausarbeit mit dem Titel „Datenmanagementpläne – Eine Bestandsübersicht“, die im Rahmen des Moduls „Konzeptionelle Entwicklung eines Werkzeugs für die Planung des Forschungsdatenmanagements“ im Wintersemester 2015 an der Fachhochschule Potsdam von Heger und Heinrich erstellt und eingereicht wurde. Die vollständige Hausarbeit ist als Anlage 1 dieser Masterarbeit angefügt. Stand der Analyse ist der 08.02.2016.

DMP Tool (Universität Bielefeld)

Allgemein	
Entstehung/Herkunft	Entstehung im institutionellen Kontext
Domäne	Berücksichtigung der Profilschwerpunkte der Forschungseinrichtungen sowie interdisziplinäre Forschungsvernetzung
Zielgruppe/Zielsetzung	alle potenziellen Antragssteller für Forschungsvorhaben
Größe des Anwenderkreises	gesamte Universität Bielefeld
Policies	Einrichtung eines DMP wird in der vorhandenen Policy explizit genannt ¹⁸
Sicherheit und Authentifizierung	Zugang ist frei, aber Freischaltung durch Administrator nötig
Technische Infrastruktur	
zentral oder lokal betrieben	zentral betrieben durch die Universität Bielefeld
Installation	entfällt, da zentral betrieben
Aktualisierung/Update	nur Datum, keine Versionierung
Gewährleistung/Datensicherheit	gemäß den Grundsätzen zum Umgang mit Forschungsdaten ist der Schutz von sensiblen Daten verpflichtend

Funktionalitäten und Inhalt

Usability	DMP wird kapitelweise ausgefüllt, eine Grafik zeigt das aktuelle Kapitel bzw. den Bearbeitungsstatus an. Kapitel können nur in vorgegebener Reihenfolge bearbeitet werden.
Sprachen	deutsch, englisch
Hilfefunktionen	umfangreiche Hinweise direkt im DMP unterstützen bei der Erstellung
Kollaboration	keine Kollaboration möglich
Exportfunktionen	DMP kann im PDF-Format heruntergeladen werden
Community/Erweiterbarkeit	keine Erweiterung möglich; Änderungen werden zentral durch die Universität Bielefeld durchgeführt; Quelltext nicht frei verfügbar bzw. es existiert keine Entwickler-Community
Versionsmanagement	kein Versionsmanagement vorhanden, lediglich Datum der Erstellung sichtbar
Verwaltung mehrerer DMP	einfache Auflistung der erstellten DMP und alphabetische Sortierung möglich
Templates von Fördereinrichtungen vorhanden/Bindung an Policy des Förderers	Vorlage der Universität Bielefeld mit einem hohen Detaillierungsgrad für datenintensive Forschungsvorhaben, Cluster of Excellence Cognitive Interaction Technology (CITEC), Horizon 2020
Kosten/Ressourcen	Personalkosten und -aufwand, Kosten vor und nach der Projektlaufzeit sowie das Gesamtbudget für das Datenmanagement werden erfasst

Metadatenstandards sinnvolle (automatische) Integration von Metadaten wird nahegelegt; Voraussetzungen für Hard- u. Software sowie benötigte Fachkenntnisse zum Umgang mit Metadaten werden erfragt

DMPonline (DCC)

Allgemein

Entstehung/Herkunft Entwicklung durch Digital Curation Centre (UK).

DCC hat eine Vielzahl an strategischen Partnerschaften (darunter regionale Universitäten sowie überregionale staatlich geförderte Organisation weltweit (Bsp: Australian National Data Service))

Domäne aufgrund der Entstehung domänenübergreifend verwendbar

Zielgruppe/Zielsetzung berücksichtigt werden alle potenziellen Antragsteller mit Schwerpunkt in UK; aber auch Nutzer außerhalb UK

Größe des Anwenderkreises sehr großer Anwenderkreis in UK, aber auch international

Es gibt internationale Anwender, die DMPonline als Grundlage für eigene Entwicklungen nutzen.

Zudem sehr aktive Community auf GitHub¹⁹.

Policies aufgrund der Herkunft keine Policies; Es sind jedoch Policies der Förderer zentral gesammelt und es gibt Anleitungen zur Erstellung von institutionellen Policies

Sicherheit und Authentifizierung	Zugang ist frei. Nutzer aus UK können sich mit ihrem Instituts-Login anmelden (Weiterleitung auf Anmeldeseite des Instituts). Andere Nutzer können Accounts selbstständig anlegen.
----------------------------------	--

Technische Infrastruktur

zentral oder lokal betrieben	zentral betrieben durch die DCC an der Universität Edinburgh. Es gibt jedoch eigene Entwicklungen die folglich lokal betrieben werden.
------------------------------	--

Installation	entfällt, da zentral betrieben
--------------	--------------------------------

Aktualisierung/Update	Datum und Versionsnummer (aktuell Version 4.1)
-----------------------	--

Gewährleistung/Datensicherheit	Die E-Mail wird gespeichert und ggf. unter DCC-Partner ausgetauscht. Der Schutz der persönlichen Daten wird jedoch zugesichert. Administratoren der Universität Edinburgh dürfen Zugang erhalten, sofern es der Wartung dient. Die Eigentumsrechte der eingegebenen Daten verbleiben bei dem Urheber.
--------------------------------	---

Funktionalitäten und Inhalt

Usability	<p>Planerstellung unterscheidet sich je nach Förderer: in drei Stufen bei Horizon 2020 (Initial DMP, Mid-Term Review DMP und Final Review DMP), einfache Abfrage der Fragen bei dem Großteil der Förderer.</p> <p>Eine Grafik zeigt den Status an.</p> <p>Beiträge bzw. Änderungen werden durch Zeit und Bearbeiter gekennzeichnet.</p> <p>ResearcherID kann angegeben werden (wichtig für mögliche Verknüpfungen)</p>
-----------	--

Sprachen	englisch
----------	----------

Hilfefunktionen	umfangreiche Hinweise direkt im DMP unterstützen bei der Erstellung. Weiterhin gibt es bei der Wahl einiger Einrichtungen aus UK institutionelle Hilfen. Für Einsteiger gibt es ein E-Tutorial, dass alle Basisfunktionen erklärt.
Kollaboration	stark ausgeprägte Kollaborationsmöglichkeiten. Zu jeder Frage können Notizen erstellt werden. Zudem können weitere Nutzer zum DMP hinzugefügt werden und gemeinsam arbeiten (bzw. nur lesen -> Rechtemanagement).
Exportfunktionen	DMP kann in allen Phasen exportiert werden. Diverse Formate vorhanden (csv, html, json, pdf, xml, text, docx). Schriftart, Dateiname und Inhalte können vor dem Export festgelegt werden.
Community/Erweiterbarkeit	Quelltext kann frei heruntergeladen (GNU Affero General Public License) und installiert werden. Es gibt eine Entwickler-Community und regelmäßige Updates. Die Diskussionen können auf Github nachvollzogen werden. Auf der Webseite von DMPonline gibt es ebenso entsprechende Informationen. Weiterhin gibt es eine E-Mail-Liste für Entwickler (DMPONLINE-USER-GROUP@JISCMail.AC.UK)
Versionsmanagement	Datum und Bearbeitet der letzten Änderung kann eingesehen werden.
Verwaltung mehrerer DMP	Freie Suche nach mehreren DMP möglich. Viele Filtermöglichkeiten sowie Sortierung beliebigen Kriterien möglich.
Templates von Fördereinrichtungen vorhanden/Bindung an Policy des Förderers	Vorlagen schwerpunktmäßig aus UK, aber auch Horizon 2020, National Science Foundation (USA) und ZonMw (Niederlande)

Kosten/Ressourcen	Angaben, ob Fachexperten bzw. Training und spezielle Hardware notwendig sind. Personalkosten und -aufwand werden nicht abgefragt.
-------------------	---

Metadatenstandards	sinnvolle (automatische) Integration von Metadaten wird nahegelegt; Verantwortliche Personen für Metadaten müssen benannt werden.
--------------------	---

Interdisciplinary Earth Data Alliance (IEDA) DMP Tool

Allgemein

Entstehung/Herkunft	2011 offiziell gestartet, um auf einfache Art und Weise DMP für die Einbindung in NSF-Vorschläge (National Science Foundation) zu erstellen, aus Kollaboration zwischen EarthChem20 und Marine Geoscience Data System21 (MGDS) entstanden
---------------------	---

Domäne	Fokus liegt auf maritimen, geologischen und polaren Daten, ist aber dennoch so grundlegend aufgebaut, dass auch Anträge von anderen NSF-Abteilungen damit erstellt und bearbeitet werden können
--------	---

Zielgruppe/Zielsetzung	Primäre Zielgruppe sind alle wissenschaftlichen NSF-Divisionen, bei denen die o.g. Daten anfallen (EarthChem und Marine Geoscience Data System) sowie ähnliche NSF-Abteilungen
------------------------	--

Größe des Anwenderkreises	Hauptnutzerkreis in den USA, aber auch Mitglied bei ICSU World Data System22 (WDS) und Federation of Earth Science Information Partners23 (ESIP), weiterer Outreach durch Einbindung der Community, Konferenzen, Meetings, Webinars, Workshops und Tool-Schulungen, Mail-Support.
---------------------------	---

Policies	Data Policy je nach NSF-Division unterschiedlich, wird beim Auswählen automatisch angezeigt
----------	---

Sicherheit und Authentifizierung Zugang ist frei. Kostenlose GeoPass ID wird benötigt

Technische Infrastruktur

zentral oder lokal betrieben zentral gehostet am Lamont-Doherty Earth Observatory²⁴ der Columbia University²⁵

Installation entfällt, da zentral betrieben

Aktualisierung/Update Datum und Versionsnummer (aktuell Version 2 vom 17.04.2012)

Gewährleistung/Datensicherheit Anmeldung erfolgt mit GeoPass ID und Passwort, Daten werden mit einem DOI unter der Creative Commons License (CC BY-NC-SA 3.026) veröffentlicht

Funktionalitäten und Inhalt

Usability Einfache Formatvorlage. Fokus liegt auf der Art der vorliegenden bzw. entstehenden Daten (Type of Data), die detailliert angegeben werden können

Sprachen englisch

Hilfefunktionen Einbindung der Community, Konferenzen, Meetings, Webinars, Workshops, Tool-Schulungen, Mail-Support

Kollaboration Co-Autoren können eingetragen werden, keine gemeinsame Bearbeitung bzw. Notizen o.ä. möglich

Exportfunktionen Art der Einreichung sowie ein automatisches Einreichungsdatum bei der NSF mit Deadline können festgelegt werden, Export als PDF-Datei mit vorheriger Syntax-Prüfung möglich

Community/Erweiterbarkeit Daten werden nach einer Prüfung durch die IEDA unter der Creative Commons Lizenz veröffentlicht und mit einem DOI verknüpft und nach einer durch den Autor festgelegten Frist der Öffentlichkeit zur Verfügung gestellt. Nach der Veröffentlichung können keine

weiteren Änderungen vorgenommen werden und bei Änderungsbedarf muss eine neue Version angelegt, registriert, geprüft und veröffentlicht werden. Sollten veröffentlichte Daten zurückgezogen werden, verbleiben die Katalogmetadaten im System (als "zurückgezogen" markiert)

Versionsmanagement	Datum der Erstellung und Einreichung werden angegeben
Verwaltung mehrerer DMP	Verwaltung mehrerer DMP möglich
Templates von Fördereinrichtungen vorhanden/Bindung an Policy des Förderers	Verschiedene Templates je nach Art der Daten und Fördereinrichtung auswählbar, Policies richten sich nach jeweiligen Einrichtungen bzw. NSF-Divisionen und werden nach Auswahl automatisch angezeigt und berücksichtigt
Kosten/Ressourcen	Kosten werden nicht abgefragt
Metadatenstandards	Mischung aus Pflichtfeldern und optionalen Metadatenfeldern wird benutzt, die die Beziehungen zu anderen Publikations-Datensammlungen weiter beschreiben sollen; verwendet wird das DataCite ²⁷ -Metadatenchema

LabArchives

Allgemein

Entstehung/Herkunft	2009 gegründet und als einfach zu benutzende und kostengünstige Lösung (Software-as-a-Service) zur Labororganisation und kollaborativem Arbeiten entwickelt. Soll traditionelle Notizbücher aus Papier durch elektronische ersetzen
---------------------	---

Domäne	Fokus liegt auf der Datenorganisation von Laboren, kann allerdings auch allgemein zum Speichern, Organisieren, Teilen und Publizieren von Daten genutzt werden
Zielgruppe/Zielsetzung	ursprünglich als elektronisches Notizbuch für Wissenschaftler von Forschungs- und Institutslaboren entwickelt, kann jedoch auch in einem allgemeineren Datenkontext genutzt werden
Größe des Anwenderkreises	vorwiegend große Institute und Universitäten, darunter das California Institute of Technology, Cornell University, Tufts University, UT Southwestern, University of Sidney, Monash University, The Garvan Medical Research Institute, University of Wisconsin, Yale University u.a. sowie Kooperation mit dem Internet228-Projekt
Policies	Terms of Service seitens LabArchives
Sicherheit und Authentifizierung	kostenlose Version nach Registrierung (Anmeldung erfolgt mit E-Mail-Adresse und Passwort) verfügbar, außerdem eine Classroom Edition (hauptsächlich an Studenten/individuelle Benutzer gerichtet), Professional Edition (für PIs, Laborleiter etc.) sowie eine campusweite Enterprise-Lizenz kostenpflichtig verfügbar
Technische Infrastruktur	
zentral oder lokal betrieben	zentral betrieben als Software-as-a-Service, cloudbasiertes Electronic Lab Notebook (ELN)
Installation	keine Installation nötig
Aktualisierung/Update	nicht ersichtlich/bekannt
Gewährleistung/Datensicherheit	LabArchives übernimmt keinerlei Haftung für die Datensicherheit

Funktionalitäten und Inhalt

Usability	übersichtlich und einfach zu benutzen, Benutzung erinnert an eine Mischung aus Windows Explorer und CMS-Systemen wie WordPress
Sprachen	englisch
Hilfefunktionen	extrem umfangreiche Hilfefunktionen: Quick Start Guides, PowerPoint- u. Video-Tutorials, Support-Anfragen per Mail und Ticket direkt im Tool möglich, Kundenhotline, Anmeldung zu verschiedenen Webinars über Website
Kollaboration	Einladefunktion von Personen per Mail, individuelle Vergabe von Rechten (Schreiben, Editieren, Sehen, kein Zugriff etc.) bis hinunter auf Datei-Ebene. Jeder Mitarbeiter sieht nur genau das, was er sehen darf. Activity Feed, das aktuelle Änderungen anzeigt
Exportfunktionen	automatischer Export per Mail möglich, an bestimmte Gruppenmitglieder, Benutzergruppen, Export u.a. als URL, komfortable DOI-Vergabe, Share-Funktion
Community/Erweiterbarkeit	Community-Blog vorhanden, ansonsten eher über Internet2
Versionsmanagement	Versionierung mit Datum, Uhrzeit (+Zeitzone), Name, IP-Adresse sowie Aktualisierung dieser Daten bei Änderungen, Versionsgeschichte ähnlich wie bei Wikipedia einsehbar und beliebig wiederherstellbar
Verwaltung mehrerer DMP	komfortabel möglich, erstellte DMP nach Name, der eigenen Nutzerrolle, Aktivitäten, letzten Aktivitäten, Navigation und Verfügbarkeit (offen, geschlossen) möglich, DMP können gelöscht, geklont (inkl. Benutzerrechten und Inhalten), kopiert und sogar vollständig oder partiell zusammengeführt werden

Templates von Fördereinrichtungen vorhanden/ nur eigene Templates bzw. exportrelevante Templates (z.B. für Google Docs/Calendar) verfügbar

Bindung an Policy des Förderers

Kosten/Ressourcen Aufgrund der Herkunft bzw. der Zielsetzung existiert standardmäßig keine Kostenabfrage

Metadatenstandards Erstellung von verschiedenen Textformaten (PDF, Word, Plain/Rich Text etc.) möglich. Aufgrund der Herkunft bzw. der Zielsetzung der Software erfolgt keine Abfrage von Metadatenstandards

4 Bestandsanalyse

Die Bestandsanalyse in der o. g. Hausarbeit zu DMP-Tools hat vor allem gezeigt, dass es speziell bei den Punkten der Versionierung und der Standardisierung noch großen Verbesserungs- und Anpassungsbedarf gibt. Idealerweise wird bereits die Eingabeoberfläche so entworfen und modelliert, dass unterschiedliche Schreibweisen bei der Dateneingabe verhindert werden, um von vornherein eine höhere Datenqualität zu ermöglichen. Insbesondere die Verwendung von einzelnen Metadaten-Elementen, z. B. für das Eingeben von Schlagwörtern, von standardisierten Listen und kontrollierten Vokabularen, z. B. zur Auswahl von Forschungsfeldern, tragen hierzu bei.

Auch ein hohes Maß an Flexibilität wird von DMP-Tools immer mehr erwartet, bevor eine Benutzung überhaupt in Erwägung gezogen wird, wie User Tests und das Feedback zu RDMO auf Konferenzen gezeigt haben. Dies bedeutet u. a., dass die Nutzer auch lokal auf ihrem Rechner damit arbeiten können, ohne Zugriff auf einen zentralen Web-Service haben zu müssen. Nach dem Zeitraum des Offline-Arbeitens muss es natürlich auch problemlos möglich sein, die in der Zwischenzeit eingetragenen Daten einfach wieder in den bestehenden Datenmanagementplan einzupflegen und die Daten zu synchronisieren. Auch ein Institutionswechsel wäre für diese Funktionalitäten ein häufiger Anwendungsfall, damit die Forschungsdaten nicht an die Software der Institution gekoppelt sind und damit für die Nachnutzung an der neuen Institution verloren gehen, sondern sozusagen „mit Umziehen“. Anknüpfend daran sollte die Möglichkeit zur technischen und inhaltlichen Konfiguration und Anpassbarkeit des DMP-Werkzeugs an das Umfeld der eigenen Institution bzw. der Forschungsdomäne bestehen.

Die Nutzung von Schnittstellen zu anderen Software-Umgebungen, anderen Datenbanken und Repositorien – sowohl institutionelle, als auch welche innerhalb der Community – ist ebenfalls ein entscheidender Faktor, der bei den Vorüberlegungen zur Benutzungen eines DMP-Tools eine große Rolle spielt. Die Möglichkeit des Zugriffs auf Linked-Open-Data-Services, z. B. der Gemeinsamen Normdatei (GND), der Klassifikation des Journal of Economic Literature (JEL), des Standard-Thesaurus Wirtschaft (STW) oder auch „nur“ eine Mitarbeiterdatenbank des Instituts und der daraus resultierenden Einpflege von bereits existierenden Personendaten, hilft bei der Vermeidung von Redundanzen und garantiert eine einheitliche sowie eindeutige Dateneintragung und damit eine höhere Datenqualität.

Die bisher genannten Punkte wurden bei den untersuchten Tools leider nicht weiter betrachtet. Häufig wurde nur ein Online-Zugriff auf Tools angeboten, welche sich kaum an die eigene Forschungsdomäne anpassen ließen und nur wenige Schnittstellen zu Linked-Open-Data-Services ermöglichten. Zudem liegt der Fokus der möglichen Eingaben und Features meist eher auf den

Anforderungen der Forschungsförderer, nicht so sehr auf der aktiven Abdeckung des Datenmanagements über die gesamte Projektlaufzeit. Dadurch wird eine umfassende Einbindung der am Datenmanagement Beteiligten erschwert.

Letztendlich wird die Verbreitung und Akzeptanz innerhalb der Community eines DMP-Tools auch davon bestimmt, welche Institution bereits wie erfolgreich ein Tool eingesetzt hat und wie groß die daraus resultierende Akzeptanz innerhalb der Forschungsgemeinschaft erwächst. Ein nicht vorhandener oder nur begrenzter Zugriff auf ein DMP-Tool für Nicht-Institutionsmitglieder ist in dieser Hinsicht kontraproduktiv; ebenso die Eigenentwicklung von DMP-Tools, die sich oft nur minimal von bestehenden Tools unterscheiden und die Interoperabilität nicht erhöhen.

Weitere Problemfelder hinsichtlich des Forschungsdatenmanagements sind oft grundlegender Natur. Oft existiert noch keine universitäre bzw. institutionelle Richtlinie zu Forschungsdaten, und es fehlt insgesamt ein Fachkonzept an Forschungseinrichtungen. Forschungsdatenmanagement hat bisher kaum ein fachliches Profil, was dazu führt, dass es keine einheitlichen Standards zum Umgang mit Forschungsdaten gibt und diese Aufgabe von der Projektleitung und/oder einer studentischen Hilfskraft „nebenher“ erledigt wird. Auch die Abgabe dieser Aufgabe an die IT-Abteilung an der jeweiligen Institution hinsichtlich des Forschungsdatenmanagements, ist keine langfristige Lösung.

Forschungsdaten sind außerhalb von Forschungsdatenrepositorien bisher nur schwierig zitierbar und der Vorgang ist noch nicht etabliert genug, so dass es bisher an Anreizen fehlt, die Metadaten für sie bereitzustellen. Die Dokumentation von Metadaten sollte noch weiter ausgebaut und gefördert werden, denn nur durch eine umfassende Metadatenbereitstellung und einer einheitlichen Plattform (z. B. an der jeweiligen Institution), wo diese mit Metadaten aufbereiteten Forschungsdaten abgelegt werden, sind die Forschungsdaten auch für die eigene Nutzung verfügbar und sinnvoll für Dritte nachnutzbar. Auch eine Integration von Metadaten in disziplinspezifische Portale (z. B. Google Scholar) mit einem persistenten Identifikator (PID) wäre hierbei ein sinnvoller Ansatzpunkt.

Diese Punkte wiegen umso schwerer, da nachhaltiges Forschungsdatenmanagement eine wachsende Bedeutung bei den Projektträgern bzw. -förderern zugemessen wird und auch, wie o. g., die DFG bei einem Projektantrag dargelegt haben möchte, wie das Forschungsdatenmanagement im jeweiligen Projekt voraussichtlich aussehen wird. Hier wären verbindliche Vorgaben zu gemeinsamen Datenmanagementrichtlinien seitens der Forschungsförderer, statt Empfehlungen, wünschenswert.

4.1 Zielsetzung/Scope des DFG-Projekts

Das Kooperationsprojekt mit dem Titel „Entwicklung und Implementierung eines Werkzeugs für die Planung, Umsetzung und Kontrolle des Forschungsdatenmanagements“ ist ein gemeinsames Projekt des Leibniz-Instituts für Astrophysik (AIP) und der Fachhochschule Potsdam. Es wird von der Deutschen Forschungsgemeinschaft e. V. seit November 2015 mit ca. 165.000 Euro über eine Projektlaufzeit von zunächst 1,5 Jahren finanziert und läuft noch bis April 2017. Ein nutzbarer Prototyp soll als Beta-Version Ende 2016/Anfang 2017 fertig sein. Der erste voll funktionsfähige Prototyp wird im Mai 2017 erwartet. Davor finden jedoch bereits Tests mit assoziierten Partnern von Mitte bis Ende 2016 statt.

Mit dem Datenmanagement-Tool „Research Data Management Organiser“ (RDMO) soll das Datenmanagement über die gesamte Laufzeit eines Projekts vollständig abgedeckt und unterstützt werden. Damit soll eine verbesserte Organisation, Durchführung und Überprüfung des Forschungsdatenmanagements gewährleistet werden. Anders, als bei vielen anderen Tools für das Forschungsdatenmanagement, steht nicht die Generierung eines textuellen Datenmanagementplans für den Forschungsförderer im Vordergrund der Funktionalität. Es soll eher ein Tool bereitgestellt werden, das zur Unterstützung des Forschungsdatenmanagements über den gesamten Projektzeitraum und unter der Einbeziehung aller Akteure (Forschende, Institutionsleitung, IT-Abteilung, Forschungsförderer, Gesetzgeber) im Datenmanagement dient. Verschiedene Stakeholder sollen zudem unterschiedliche Perspektiven des Datenmanagementplans sehen können, so dass die jeweils relevanten Informationen in den Vordergrund rücken. Generell sollen spezifische Rollen für die Akteure im Datenmanagementplan vergeben werden, die über unterschiedliche Rechte und Ansichten verfügen. Auch eine kollaborative Bearbeitung der eingegebenen Daten soll möglich sein. Dies geschieht vor allem durch ein strukturiertes Interview zur Eingabe aller Daten für ein nachhaltiges Datenmanagement. Damit werden redundante und irrelevante Fragen vermieden und gleichzeitig konfigurierbare Antwortoptionen erlaubt.

Das Tool soll inhaltlich und technisch sowohl an die jeweilige Forschungsdomäne, das Forschungsfeld sowie an den institutionellen Kontext anpassbar sein. Ersteres ist durch die Erstellung bzw. Einbindung eines eigenen, disziplinspezifischen Fragenkatalogs möglich. Letzteres geschieht durch eine White-Label-Lösung, um ein individuelles Branding (z. B. Corporate Identity) zu ermöglichen. Außerdem soll das Tool sowohl lokal (eigener PC) genutzt als auch in eine bestehende IT-Infrastruktur vor Ort eingebettet werden können, so dass es von jedweder Institution möglichst einfach eingesetzt und angepasst werden kann. Eine Interoperabilität mit anderen Instanzen von RDMO und anderer Software gehört ebenfalls dazu, wie die Bereitstellung von Schnittstellen zum Zugriff auf die gespeicherten Informationen und die Nutzung externer Schnittstellen (z. B. Repositorien und Forschungsinformationssysteme). Ein einfacher Einsatz

und die Verwaltung in verschiedenen Kontexten, z. B. innerhalb einer Universität, eines Instituts und Verbundprojekts, ist somit denkbar.

Das Projekt baut inhaltlich vor allem auf den Vorarbeiten aus dem WissGrid-Leitfaden zum Forschungsdatenmanagement auf, in dem bereits viele Anforderungen an ein modernes Forschungsdatenmanagement formuliert wurden. Während der Projektlaufzeit wird ein enger Kontakt mit Forschenden, ihren Institutionen sowie der Forschungscommunity allgemein gepflegt, um bereits in der Entwicklungsphase wertvolles Feedback, Hinweise und Wünsche zu sammeln und ggf. umsetzen zu können und den disziplinspezifischen Anpassungsbedarf auszuloten. Dies geschieht u. a. durch gezielte Expertengespräche in verschiedenen Forschungsdisziplinen (u. a. Astrophysik und Sozialwissenschaften) und der Vorstellung und Evaluierung von Prototypen des Tools im Rahmen von Workshops, Fachkonferenzen und Nutzertests (u. a. in der Gruppe des GREGOR Fabry-Pérot-Interferometer im Bereich der Sonnenphysik und der sozialwissenschaftlichen Zwillingsstudie TwinLife³⁹). Der dadurch gewonnene Input wird zur Erstellung von User Stories und des Fragen-Aufgaben-Katalogs verwendet. Die Entwicklung eines eigenen RDMO-Vokabulars sowie eines RDMO-Mappings sollen den Datenexport und die Interoperabilität sowohl innerhalb des RDMO als auch mit anderen Forschungsumgebungen sicherstellen. Um von Beginn an eine hohe Verbreitung zu ermöglichen, wird das Tool in seiner ersten Iteration in Deutsch und Englisch veröffentlicht.

³⁹ Universität des Saarlandes: TwinLife. URL: <http://www.twin-life.de/>.

5 Einbettung der Arbeit in den Kontext des DFG-Projekts RDMO

Im Kontext dieses Projekts ist die vorliegende Masterarbeit entstanden und wurde in einer frühen Phase nach Einladung durch die Deutsche Gesellschaft für Information und Wissen e. V. (DGI) bereits auf der DGI-Konferenz 2016⁴⁰ von Martin Heger in einer Präsentation⁴¹ vorgestellt.

Prof. Dr. Heike Neuroth, Professorin für Bibliothekswissenschaften an der Fachhochschule Potsdam und Erstgutachterin dieser Masterarbeit sowie Claudia Engelhardt, wissenschaftliche Mitarbeiterin an der Fachhochschule Potsdam und der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB), wirken als Vertreter der Fachhochschule Potsdam an dem Projekt mit. Die weiteren Projektmitglieder sind Dr. Jochen Klar, wissenschaftlicher Mitarbeiter am Leibniz-Institut für Astrophysik (AIP) und Zweitgutachter dieser Masterarbeit, Dr. Harry Enke, Leiter der E-Science-Gruppe am AIP und Sprecher des Arbeitskreis Forschungsdaten der Leibniz-Gemeinschaft sowie Jens Ludwig, Sprecher der gemeinsamen Arbeitsgruppe "Forschungsdaten" der Deutschen Initiative für Netzwerkinformation e. V. (DINI) und nestor - Deutsches Kompetenznetzwerk zur digitalen Langzeitarchivierung und an der Staatsbibliothek zu Berlin - Stiftung Preußischer Kulturbesitz tätig.

Die Projektwebsite befindet sich unter der Webadresse <http://rdmorganiser.github.io/>. Der gesamte, in der Entwicklung befindliche Quellcode des Programms kann bei GitHub⁴² eingesehen werden, wo er unter einer offenen Lizenz veröffentlicht wird.

⁴⁰ DGI: DGI-Konferenz 2016: Erfahrung reloaded. URL: <http://dgi-info.de/events/dgi-konferenz-erfahrung-reloaded-vom-mundaneum-zum-web-of-everything/>.

⁴¹ Heger: Datenmodellierung für Forschungsdatenmanagementpläne. URL: http://dgi-info.de/wp-content/uploads/2015/11/DGI-Pr%C3%A4sentation_Martin-Heger.pdf.

⁴² GitHub: RDMO. URL: <https://github.com/rdmorganiser/>.

6 Vorgehen/Methodik

Das Ziel dieser Masterarbeit ist die Schaffung von Interoperabilität zwischen verschiedenen DMP. Die in einer Instanz von RDMO vorhandenen Daten sollen mittels Datenexport im XML-Format in eine andere RDMO-Instanz übertragbar und importierbar sein. Ein wichtiges Anwendungsbeispiel hierfür ist der Umzug an eine neue Institution: Es soll möglich sein, die bisher eingetragenen Daten mitzunehmen und dort weiter bearbeiten zu können. Die Daten werden in einer Datenbank strukturiert abgelegt. Der Inhalt wird serialisiert, wenn er transportiert werden soll, entweder zum Import in eine andere Datenbank etc., oder zur Verwendung in einem Renderer, der z. B. textuelle oder graphische etc. Ausgaben ermöglicht. Dies geschieht durch einen Serializer, der die Daten in einen Stream zerlegt, eine Bearbeitung möglich macht und sie an den Renderer weiterleitet, der sie dann als XML-Daten ausgibt. Die Möglichkeit zur Offline-Arbeit ist ebenfalls wichtig: Falls die eingetragenen Daten mangels einer Internetverbindung nicht zeitnah zum RDMO der Institution hochgeladen werden können, muss es zu einem späteren Zeitpunkt möglich sein, die in dieser Zeit angefallenen Daten wieder in das Institutions-RDMO zu integrieren.

Zu diesem Zweck besitzt jede Frage im RDMO ein eigenes Attribut und einen Wert, die sowohl exportiert als auch wieder importiert werden können. Zudem erfolgt ein Mapping der RDMO-Attribute auf Terme aus anderen Vokabularen, um eine Schnittstelle zu weit verbreiteten Vokabularen wie z. B. DCMI Terms und CERIF bzw. SEMCERIF zu ermöglichen und Interoperabilität zwischen verschiedenen Vokabularen zu schaffen (siehe Kapitel 7.2). Hierfür muss ein RDMO-Mapping erstellt werden. Dies ist ein Mapping-Schema, welches Datenelemente (Terme) aus mehreren Namensräumen (engl.: Namespaces) kombiniert und auf die Anwendung des RDMO zugeschnitten ist. Zu den im RDMO vorhandenen Attributen werden Terme aus etablierten Vokabularen gemappt, welche die gleiche semantische Bedeutung besitzen und somit einen reibungslosen Austausch der Daten erlauben. In dieser Masterarbeit wird ein Mapping der Attribute zu den Termen aus bestehenden Vokabularen vorgenommen, soweit sinnvolle semantische Äquivalente vorhanden sind. In den Fällen, wo ein eindeutiges Mapping nicht möglich ist oder die vorhandenen Vokabulare lückenhaft sind, muss ein eigenes RDMO-Vokabular definiert und erstellt werden.

Als nächster logischer Schritt erfolgt die technische Implementierung dieser inhaltlichen Funktionalitäten durch die Erstellung einer XML-Output-Funktion mittels eines Renderers in der Programmiersprache Python⁴³. Dieser Renderer läuft über die in diesem Projekt verwendete Schnittstelle des Django REST Frameworks⁴⁴ und wandelt die Daten aus der RDMO-Datenbank

⁴³ Python Software Foundation: Python. URL: <https://www.python.org/>.

⁴⁴ Christie: Django REST framework. URL: <http://www.django-rest-framework.org/>.

in das XML-Datenformat um. Für diesen Prozess wurden verschiedene Werkzeuge und Bibliotheken verwendet, welche in dem nachfolgenden Unterkapitel genauer betrachtet werden.

6.1 Verwendete Werkzeuge und Bibliotheken

Server: Django (Web Application Framework in Python)

Serverseitig kommt das Django Webframework⁴⁵ zum Einsatz, welches zur Entwicklung von dynamischen Websites und Web-Applikationen in der Programmiersprache Python genutzt wird. Ein Framework ist eine Art Strukturgerüst, innerhalb dessen das eigentliche Programm entwickelt wird. Dazu werden integrierte Klassen und Funktionen sowie die entsprechenden Schnittstellen des Frameworks benutzt.⁴⁶ Mittels eines Object-Relational-Mappers (ORM) wird vom Django Framework eine Schnittstelle zu Datenbanken geschaffen. Django folgt der DRY-Devise (Don't Repeat Yourself)⁴⁷, welche auf dem MVC-Prinzip (Model View Controller) beruht und das Vermeiden von Redundanzen zum Ziel hat. Zudem wird die Entwicklung von datenbankbasierten Webseiten durch die Wiederverwendung einzelner Komponenten beschleunigt.⁴⁸ Das Django Webframework wird von der Django Software Foundation (DSF) als Open-Source-Software unter einer BSD-Lizenz vertrieben, so dass der Quellcode eingesehen und erweitert werden kann.⁴⁹

SQLite (SQL-Datenbank)

Als Grundlage für die Datenbank wird SQLite⁵⁰ verwendet. Dabei handelt es sich um eine Programm-bibliothek, die ein relationales Datenbanksystem enthält und auch Schnittstellen zu verschiedenen Programmiersprachen und Framework (u. a. Django) mit sich bringt. Durch die direkte Integration in eine Anwendung wird keine zusätzliche Server-Software benötigt.

Django REST Framework (API-Toolkit)

Das Django REST Framework ist ein Toolkit, um Web-APIs (Application Programming Interface) zu programmieren und ist die Grundlage für die Serialisierung. REST (REpresentational State Transfer) ist ein Client-Server-Protokoll (benutzt zumeist das HTTP-Protokoll) und ein architektonisches Software-Konzept für Netzwerkapplikationen. Statt komplexen bzw. bandbreitenintensiven XML-Mechanismen zur Kommunikation (z. B. Simple Open Access Protocol (SOAP))

⁴⁵ Django Software Foundation: Django. URL: <https://www.djangoproject.com/>.

⁴⁶ Django Girls: What is Django?. URL: <http://tutorial.djangogirls.org/en/django/#what-is-django>.

⁴⁷ Django Software Foundation: Design philosophies. URL: <https://docs.djangoproject.com/en/1.10/misc/design-philosophies/#don-t-repeat-yourself-dry>.

⁴⁸ Django Software Foundation: FAQ. URL: <https://docs.djangoproject.com/en/1.10/faq/general/#django-appears-to-be-a-mvc-framework-but-you-call-the-controller-the-view-and-the-view-the-template-how-come-you-don-t-use-the-standard-names>.

⁴⁹ Deutscher Django-Verein e. V.: Über Django. URL: <http://www.django-de.org/ueber-django/>.

⁵⁰ The SQLite Consortium: SQLite. URL: <https://www.sqlite.org/>.

wird für sämtliche Aktionen das HTTP-Protokoll verwendet, z. B. um Daten zu erstellen, aktualisieren, lesen und zu löschen.⁵¹ Damit stellt REST im Prinzip den softwaretechnischen Stil des World Wide Web dar und nutzt die bereits vorhandene REST-kompatible Infrastruktur.⁵²

Client: AngularJS (JavaScript-Webframework)

Zur clientseitigen Kommunikation mit dem Server wird AngularJS⁵³ benutzt. Es handelt sich dabei um ein reines JavaScript-Framework für die Erstellung von Web-Anwendungen und wird von Google und weiteren Firmen und Mitwirkenden gepflegt und weiterentwickelt. AngularJS kann sowohl für das Model-View-ViewModel-Prinzip (MVVM) als auch das Model-View-Controller-Prinzip (MVC) verwendet werden. Das MVC-Prinzip, das bei diesem Projekt Verwendung findet, basiert auf der Trennung der Programm-Komponenten zur Repräsentation der Daten (Model), Darstellung (View) und Logik (Controller).

SAX Serializer (Renderer)

SAX⁵⁴ steht für Simple API for XML und dient als breit genutzte Public-Domain-Schnittstelle zum Parsen, d. h. der Zerlegung und Umwandlung von Datenbestandteilen, von serialisierten XML-Daten. Bei der Serialisierung werden die vorliegenden XML-Daten als sequentieller Datenstrom eingelesen. Bei dieser ereignisorientierten Zugriffsmethode werden die XML-Daten wie in einer XML-Pipeline verarbeitet. Die vorliegenden XML-Daten werden sequentiell gelesen, und beim Identifizieren bestimmter Syntaxstrukturen können vorher festgelegte Ereignisse ausgeführt werden, in der Software wird das als Callback-Funktion (Rückruffunktion) realisiert.⁵⁵ In der vorliegenden Masterarbeit wird der SAX-Parser dazu genutzt, die im RDMO eingetragenen Daten auszulesen und als XML-Daten auszugeben.

6.2 Ausgangslage

Ausgangslage für die Erarbeitung der Ergebnisse war der bereits vorliegende generische Fragenkatalog für RDMO (siehe Anlage 2). Die Entstehung des Fragenkatalogs basiert hauptsächlich auf den Vorarbeiten aus dem WissGrid-Projekt und dem daraus entstandenen WissGrid-Leitfaden zum Umgang mit Forschungsdaten. Neben dem WissGrid-Leitfaden wurde sich auch an Checklisten des Max-Planck-Instituts zur Erforschung multireligiöser und multiethnischer Gesellschaften (MPI MMG) orientiert. Außerdem wurden Fragen in anderen Tools angesehen, z. B. dem DMP Tool der Universität Bielefeld und dem DataWizard von Clarin-D⁵⁶.

⁵¹ OIO: REST Web Services. URL: <http://www.oio.de/public/xml/rest-webservices.htm>.

⁵² TechTarget: REST. URL: <http://www.searchenterprisesoftware.de/definition/Representational-State-Transfer-REST>.

⁵³ Google Inc.: AngularJS. URL: <https://docs.angularjs.org/misc/faq>.

⁵⁴ Megginson: SAX. URL: <http://www.saxproject.org/>.

⁵⁵ Oracle: Simple API for XML. URL: <https://docs.oracle.com/javase/tutorial/jaxp/sax/index.html>.

⁵⁶ Clarin-D: Data Management Plan. URL: <http://www.clarin-d.de/en/preparation/data-management-plan>.

Ziel war es, einen möglichst generischen Fragenkatalog zu entwickeln, der auf möglichst viele Institutionen und Forschungsfelder anwendbar ist. Mit wenig Aufwand sollen zudem Anpassungen vorzunehmen sein, um den Fragenkatalog ans eigene institutionelle Umfeld und die eigene Forschungsdomäne anpassen zu können. Auch die Entwicklung eines komplett eigenen Fragenkatalogs soll einfach und ohne technische Vorkenntnisse erfolgen können.

Die Beantwortung des Fragenkatalogs soll wie ein geführtes Interview ablaufen. Die Nutzer werden nach und nach durch die einzelnen thematischen Fragenblöcke geführt und beantworten darin zunehmend speziellere Fragen. Dabei sollen neben allgemeinen Fragen, wie z. B. zum Forschungsförderer des Projekts, auch Punkte ins Bewusstsein der Verantwortlichen gerufen werden, die bei Forschenden noch unbekannt sind bzw. momentan eher vernachlässigt werden, beispielsweise die Anforderungen der Forschungsförderer an das Datenmanagement. Zudem sollen auch Aspekte gefragt werden, die später mit Aufgaben und Bedingungen verknüpft werden können. So soll es u. a. möglich sein, dass die verantwortliche Person für das Datenbackup eine automatische Erinnerung erhält, die ihr einen bevorstehenden Sicherungszeitpunkt ins Gedächtnis ruft. Generell sollen die verantwortlichen Personen benachrichtigt werden, sobald etwas in ihrem Zuständigkeitsbereich geschieht bzw. zeitlich fällig ist oder wird.

Das zugrundeliegende Datenmodell von RDMO ist in drei Teile gegliedert: Fragen, Domäne und Projekte. Die Fragen sind grundsätzlich zunächst in einem übergeordneten Fragenkatalog abgelegt. Dieser Fragenkatalog ist in verschiedene, thematische Abschnitte (u. a. Allgemein, Datennutzung, rechtliche und ethische Fragen etc.) unterteilt, die wiederum Unterabschnitte (u. a. Projektpartner, Qualitätssicherung, sensible Daten etc.) beinhalten. In diesen Unterabschnitten können sowohl einzelne Fragen, als auch Fragen-Sets enthalten sein, die aus mehreren, thematisch und logisch zusammenhängenden Fragen bestehen. Das Zusammenfassen von einzelnen Fragen zu Fragen-Sets hat den Vorteil, dass manche Fragen-Sets nur bestimmten Personen angezeigt werden können oder auch gleich beim Ausfüllen übersprungen werden können, falls sie sich als nicht relevant herausstellen (z. B. Fragen-Sets zu sensiblen Daten, falls keine sensiblen Daten anfallen bzw. erhoben werden). Unter dem Punkt Domäne befinden sich die Entitäten (Äquivalent zu Klassen in Metadatenschemata), an die Bedingungen geknüpft werden können, sowie die Attribute, welche den Properties in Metadatenschemata entsprechen, d. h. den Eigenschaften von Objekten. An die Attribute geknüpft sind die Optionen (z. B. Auswahl der Disziplinen aus der DFG-Fachsystematik, die dem Projekt zuzuordnen sind), der Bereich (z. B. Anzahl der Monate, die ein Projekt maximal dauern kann) und die Bedingungen (wurde z. B. eine Frage verneint, dann wird das nächste Fragen-Set dazu nicht angezeigt). Im dritten und letzten Teil des Datenmodells, Projekte, sind die Werte abgelegt, die ein Attribut annehmen kann.

Datenmodell

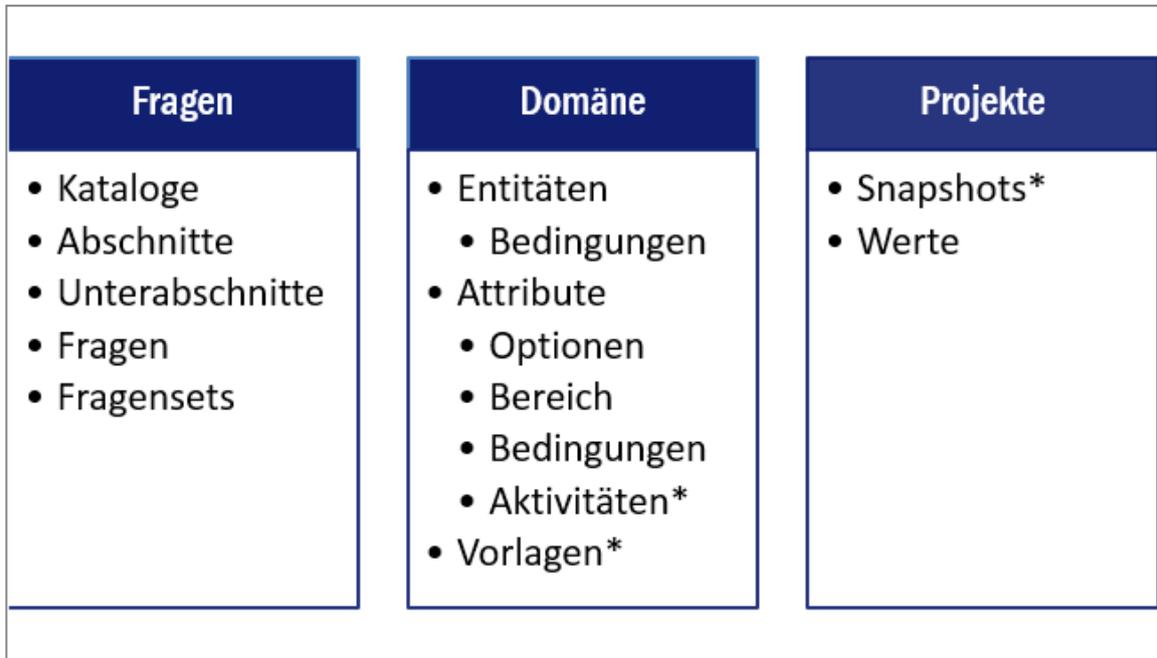


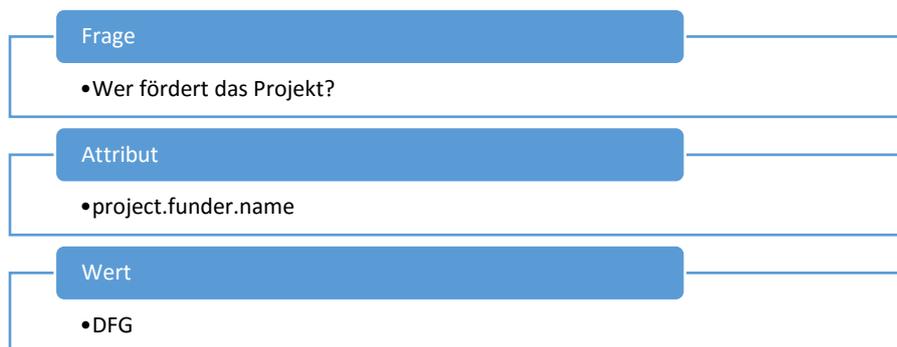
Abbildung 3: RDMO-Datenmodell

7 Ergebnisse

7.1 Attribute und Entitäten

Der ausgearbeitete Fragenkatalog wird zunächst in RDMO eingepflegt und im Anschluss daran ins Englische übersetzt. Erst dann kann die Vergabe der Attribute erfolgen. Jeder Frage und jedem Fragen-Set des Fragenkatalogs werden ein eindeutiges Attribut bzw. eine Entität zugewiesen, welche nach der Dateneingabe einen bestimmten Wert beinhalten.

Beispiel:



Die vollständige Liste der Attribute und Entitäten befindet sich in Anlage 3.

Zusätzlich wird aus möglichst etablierten und häufig verwendeten Vokabularen ein Term ausgewählt, sofern er exakt zum Attribut der Frage passt und natürlich auch in einem bereits bestehenden Vokabular vorhanden ist, und als Metadaten-Term verknüpft. Die Kombination aus Attribut, Wert und Metadaten-Vokabular erlaubt sowohl die Interoperabilität zwischen verschiedenen Instanzen von RDMO als auch zwischen verschiedenen Vokabularen. Sollte kein sinnvolles Mapping zwischen den RDMO-Daten und einem äquivalenten Term aus einem Metadaten-Vokabular möglich sein, werden eigene RDMO-Terme erstellt und mit den kompatiblen Termen zusammen ins RDMO-Metadatenschema integriert. Weiterhin wird jedem Attribut genau ein Wert-Typ zugewiesen, den dieses annehmen kann. Folgende Wert-Typen wurden für Attribute vergeben:

- Text
- Ganzzahl
- Kommazahl
- Boolesche Variable
- Datum und Zeit
- Optionen

Beispiele für die Verwendung dieser Wert-Typen im Kontext des Fragenkatalogs sind:

- Beschreibung des Projektvorhabens (Text)
- Projektlaufzeit in Monaten (Ganzzahl)
- Kosten pro Monat für das Datenbackup (Kommazahl)
- Gibt es eine Versionierungsstrategie? (Ja/Nein) (Boolesche Variable)
- Projektbeginn (Datum und Zeit)
- Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen? (Optionen)

Als letzter Schritt werden Widget-Typen für die Benutzeroberfläche vergeben. Ein Widget ist programm-technisch ein (abstraktes) Element der Benutzeroberfläche kombiniert mit Content und Eigenschaften für einen bestimmten Zweck. Jede Frage erhält einen Widget-Typ, der mit dem Wert-Typ korrespondieren muss. Folgende Widget-Typen gibt es in der RDMO-Benutzeroberfläche:

- Text (eine Zeile zur Texteingabe)
- Textarea (ein großes Feld zur Eingabe von längeren Texten)
- Yes/No (Feld zum Ankreuzen von Ja oder Nein)
- Checkboxes (von mehreren Antwortmöglichkeiten können mehrere ausgewählt werden)
- Radio buttons (von mehreren Antwortmöglichkeiten kann nur eine ausgewählt werden)
- Select drop-down (Drop-Down-Liste mit mehreren Antwortmöglichkeiten, von denen nur eine ausgewählt werden kann)
- Range slider (Schieberegler)
- Date picker (Öffnet Kalender zur Auswahl eines Datums)

Beim o. g. Beispiel sind demnach folgende Kombinationen von Wert-Typen und Widget-Typen sinnvoll:

- Beschreibung des Projektvorhabens (Wert-Typ: Text, Widget-Typ: Textarea)
- Projektlaufzeit in Monaten (Wert-Typ: Ganzzahl, Widget-Typ: Range slider)
- Kosten pro Monat für das Datenbackup (Wert-Typ: Kommazahl, Widget-Typ: Text)
- Gibt es eine Versionierungsstrategie? (Ja/Nein) (Wert-Typ: Boolesche Variable, Widget-Typ: Yes/No)
- Projektbeginn (Wert-Typ: Datum und Zeit, Widget-Typ: Date picker)
- Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen? (Wert-Typ: Optionen, Widget-Typ: Select drop-down)

- Unter welchen Nutzungsbedingungen sollen die Daten veröffentlicht bzw. geteilt werden? (Wert -Typ: Text, Widget-Typ: Checkboxes)
- Welche Lizenz soll hierfür verwendet werden? (Wert -Typ: Text, Widget-Typ: Radio buttons (wenn man Lizenztypen vorgibt))

Nachfolgend ein Beispiel für die Benutzeroberfläche des RDMO-Fragenkatalogs mit den dazugehörigen Widgets anhand von Fragen zur Datennutzung/Kosten:

Datennutzung / Kosten

Welcher Personalaufwand für das Datenmanagement entsteht im Rahmen der Erhebung, Erstellung oder Akquise der Daten im Projekt?

Bitte schätzen sie den Aufwand in Personenmonaten.

Welche Sachkosten für das Datenmanagement entstehen im Rahmen der Erhebung, Erstellung oder Akquise der Daten im Projekt?

Bitte schätzen sie die Kosten in Euro.

Welcher Personalaufwand für das Datenmanagement entsteht im Zusammenhang mit der Nutzung der Daten im Projekt?

Bitte schätzen sie den Aufwand in Personenmonaten.

Welche Sachkosten für das Datenmanagement entstehen im Zusammenhang mit der Nutzung der Daten im Projekt?

Bitte schätzen sie die Kosten in Euro.

Welcher Personalaufwand entsteht im Zusammenhang mit der Speicherung der Daten während des Projekts?

Bitte schätzen sie den Aufwand in Personenmonaten.

Welche Sachkosten entstehen im Zusammenhang mit der Speicherung der Datensätze während des Projekts?

Bitte schätzen sie die Kosten in Euro.

Abbildung 4: Ausschnitt des RDMO-Fragenkatalogs für Datennutzung/Kosten

7.2 Metadaten

Ein Mapping von Metadaten kann z. B. über ein Metadata Application Profile erfolgen. Dies ist ein Metadaten-Schema, welches vorhandene Datenelemente aus mehreren Namespaces bestehender Vokabulare (z. B. von DCMI Terms, CERIF, KDSF) aufgreift, kombiniert und benutzt, um größtmögliche Interoperabilität durch Linked Data zu ermöglichen. Es wird an die jeweilige Anwendung individuell angepasst.

Bei der Auswahl der Vokabulare gilt es möglichst wenige Vokabulare zu benutzen, die dafür viele sinnvolle und passende Äquivalente in Form von Termen zu den Attributen beinhalten. Außerdem sollten strategisch wichtige Vokabulare eingesetzt werden, die auch eine höhere Verbreitung innerhalb der Forschungsgemeinschaft aufweisen (z. B. eher das etablierte CERIF, als CASRAI). CERIF ist zudem kompatibel zum Kerndatensatz Forschung. Bei der Erstellung eines Metadata Application Profiles müssen beim Mapping auch die unterschiedlichen Maßstäbe für verschiedene Disziplinen berücksichtigt werden sowie die Vorgaben im Vocabulary-Encoding-Scheme, welches die mögliche Bandbreite der Vokabular-Einträge festlegt. Zur Erstellung eines Metadata Application Profiles gibt es eine Vorlage von DCMI Terms und des KIM-Forums⁵⁷.

Beim Datenaustausch von RDMO mit anderen (Verwaltungs-)Systemen muss ein eigener Standard für die zu den Attributen äquivalenten Terme geschaffen werden. Viele Attribute können jedoch mit einem Mapping auf verbreitete Vokabulare übertragen werden. In dieser Masterarbeit werden – soweit möglich – RDMO-Attribute auf bestehende Vokabulare gemappt. Ziel ist es, dass in RDMO eingegebene Daten als maschinenlesbarer Output zu anderen Forschungsinformationssystemen für eine Datenaufnahme bereitgestellt werden können. Auch soll es möglich sein, dass in diesen Systemen bereits eingetragene Daten, z. B. über Projekte und Personen, genutzt und in RDMO voreingetragen werden können, so dass ein erneutes Eintragen in RDMO entfällt und neue Projekte schnell und komfortabel erstellt werden können. Eine Verbindung von RDMO-Entitäten und Attributen mit Vokabularen wie CERIF ist eine wichtige Voraussetzung für die Akzeptanz des RDMO insgesamt.

Folgende Vokabulare und Schemata wurden für das Mapping von RDMO-Termen in die nähere Betrachtung einbezogen:

Friend of a Friend (FOAF)

FOAF⁵⁸ ist ein Vokabular, welches Personen, ihre Aktivitäten und ihre Beziehungen zu anderen Personen beschreibt. Grundlage ist ein RDF-Schema, dessen Klassen und Eigenschaften in einem

⁵⁷ KIM: Metadatenprofile. URL: http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadatenprofile.pdf?__blob=publicationFile.

⁵⁸ FOAF. URL: <http://www.foaf-project.org/>.

FOAF-Profil verwendet werden können. Dazu gehören z. B. Name, Alter, Geschlecht sowie Verbindungen zu anderen Personen. Durch diese Daten können Personen eindeutig beschrieben und identifiziert werden.⁵⁹ Da FOAF sehr ausgereift ist und sich auf die Beschreibung von Personen spezialisiert hat, ist es für das Mapping auf entsprechende RDMO-Terme gut geeignet.

Dublin Core Metadata Initiative Metadata Terms (DCMI Terms)

Bei DCMI Terms handelt es sich um ein Kern-Set von Metadaten-Elementen, die auf dem bibliographischen Datenformat von Dublin Core⁶⁰ beruhen und von der Dublin Core Metadata Initiative gepflegt werden. Es gibt 15 Kernfelder (engl.: Core Elements) und sowie weitere Felder, speziell zur möglichst granularen Beschreibung von Objekten. Sie beinhalten u. a. Klassen und Properties, Informationen zu Syntax- und Vokabular-Schemata und liefern normierte Grundregeln zur Deskription von Objekten und Webressourcen mit Metadaten. Diese werden dadurch leichter auffindbar, insbesondere durch stichwortbasierte Suchmaschinen. Die Darstellung der Metadaten kann u. a. im XML-/RDF-Format erfolgen. DCMI Terms wird vor allem von Bibliotheken und Museen verwendet.⁶¹

Common European Research Information Format (CERIF)

CERIF⁶² wurde ursprünglich mit Unterstützung der Europäischen Kommission entwickelt und den Mitgliedsstaaten zur Benutzung empfohlen; entsprechend groß ist die Verbreitung. Es ist ein Model zum Organisieren und dem Austausch von Forschungsdaten, der Forschungsdomäne und ihrer Beziehungen zueinander, sowohl auf einer konzeptuellen, logischen und physischen Ebene. Das Datenmodell umfasst u. a. Organisationen, deren Projekte, Finanzierungen und generell sämtliches, was beim Forschungsprozess entsteht bzw. mit ihm verbunden ist. Dem Linked-Data-Muster folgend, dient es der Interoperabilität zwischen verschiedenen Forschungsdatensystemen. CERIF soll als Model für den homogenen Zugriff auf heterogene Datensysteme sowie als Definition eines Datenaustauschformats dienen. Das ultimative Ziel von CERIF ist es, als Interoperabilitätsebene zwischen der elektronischen Infrastruktur und den Forschungsdaten zu dienen, und durch Standardisierung die Integration und den Austausch zu fördern.⁶³ Bei SEMCERIF handelt es sich um ein semantisches Vokabular, welches auf CERIF 1.5 basiert.

Consortia Advancing Standards in Research Administration Information (CASRAI)

CASRAI ist eine Initiative von Forschungsinstitutionen, Forschenden und Förderern. Ziel dieser Initiative ist es, dass Format von Forschungsinformationen zu standardisieren. Dies soll über eine zentrale Forschungsdatendatei laufen, die dann bei jedem Einreichen von Projektberichten

⁵⁹ FOAF: Vocabulary Specification. URL: <http://xmlns.com/foaf/spec/>.

⁶⁰ DCMI. URL: <http://dublincore.org/>.

⁶¹ DCMI: Metadata Terms. URL: <http://dublincore.org/documents/dcmi-terms/>.

⁶² euroCRIS: Main features of CERIF. URL: <http://www.eurocris.org/cerif/main-features-cerif>.

⁶³ ERCIM: CERIF. URL: <http://ercim-news.ercim.eu/en68/european-scene-qsupport-of-the-research-process/cerif-the-common-european-research-information-format>.

benutzt wird. Dort sind alle Informationen zu Fördermitteln, Datenmanagement, etc. verzeichnet und können von allen Akteuren im Forschungsprozess abgerufen werden.⁶⁴ Die Voraussetzung dafür ist allerdings, dass die entsprechenden Akteure auch CASRAI in ihre jeweilige Software und Prozesse mit eingebunden haben. Zudem entspricht CASRAI eher einem Wörterbuch bzw. Glossar, in dem Standardterme aus der Forschung gesammelt werden, als einem eigenen Vokabular bzw. Metadatenchema, so dass es für ein Mapping ungeeignet ist.⁶⁵

Kerndatensatz Forschung (KDSF)

Der Kerndatensatz Forschung⁶⁶ beruht auf Empfehlungen des Wissenschaftsrates und beschreibt in standardisierter Form Angaben zu Forschungsaktivitäten. Dadurch sollen qualitätsgesicherte Forschungsaktivitäten für Forschungsberichte mit wenig Arbeitsaufwand verglichen und mehrfach verwendet werden können. Das Ziel ist die Bereitstellung eines Standards für Deutschland; die Zielgruppen dafür sind u. a. Hochschulen und Forschungseinrichtungen.⁶⁷

Teile des Kerndatensatz Forschung sind mit dem technischen Datenmodell von CERIF kompatibel. Der KDSF ist in einigen Bereichen, z. B. den verschiedenen Publikationstypen, deutlich spezieller als CERIF, so dass nicht alle Daten abgebildet werden können. Der zu starke Fokus auf diesen Publikationsdaten und auch den Personalkategorien war auch dafür ausschlaggebend, den KDSF bei der Erstellung des RDMO-Mappings nicht zu benutzen und die Daten gleich auf CERIF zu mappen.

7.2.1 RDMO-Mapping

Bei der Auswahl des Vokabulars für das Mapping auf die Entitäten und Attribute von RDMO wird zunächst CERIF bzw. SEMCERIF ausgewählt. Die dort enthaltenen Klassen (Objekte) können auf die Entitäten des RDMO-Datenmodells gemappt werden. Attribute in RDMO entsprechen Properties (Eigenschaften von Objekten) und werden aus DCMI Terms und FOAF ausgewählt und gemappt. Jedes verwendete Schema deckt dabei andere Bereiche von RDMO ab, so kommt FOAF bei der Beschreibung von Personen zum Einsatz. Bei der näheren Betrachtung der vorab ausgewählten Vokabulare stellte sich heraus, dass ein Mapping auf SEMCERIF besonders sinnvoll ist, da einerseits viele Klassen auf die Entitäten von RDMO gemappt werden können und andererseits CERIF den Metadatenstandard für viele Forschungsinformationssysteme darstellt. Damit deckt der Export von RDMO-Daten die institutionelle Ebene ab. Von einer Verwendung von CASRAI wird abgesehen, da es sich eher um ein Glossar bzw. Lexikon handelt und nicht um ein voll ausgearbeitetes Metadatenchema. Das Vokabular des Kerndatensatz Forschung stellte sich

⁶⁴ CASRAI. URL: <http://casrai.org/about>.

⁶⁵ CASRAI: The CASRAI Dictionary. URL: http://dictionary.casrai.org/Main_Page.

⁶⁶ DZHW: Kerndatensatz Forschung. URL: <http://www.kerndatensatz-forschung.de/>.

⁶⁷ WR: Empfehlungen zur Spezifikation des KDSF. URL: <http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>, S. 5-7.

als nicht speziell genug für das RDMO-Datenmodell heraus und legt den Fokus hauptsächlich eher auf Metadaten für Publikationen.

Betont werden soll an dieser Stelle noch einmal, dass im Rahmen dieser Masterarbeit kein eigenes Exportformat für RDMO erstellt wird. Es erfolgt eine Transformation, d. h. ein 1:1-Mapping zu anderen Vokabularen. In den letztendlich ausgewählten Metadatenschemata werden semantische, statt syntaktischer Äquivalente für RDMO-Attribute und -Entitäten ausgesucht.

project.dataset (Entität)

Term: dc:dataset

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/dcmitype/Dataset>

Description:

Data encoded in a defined structure. Examples include lists, tables, and databases. A dataset may be useful for direct machine processing.

Anmerkung:

Die Entität „project.dataset“ kann für die Metadatenbeschreibung einer oder mehrerer Ressourcen auf den DCMI-Term „dc:dataset“ gemappt werden.

project.dataset.creator (Entität)

Term: semcerif:Creator

Metadatenschema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Creator>

Description:

This term belongs to the 'Person Output Contributions' classification scheme.

Anmerkung:

Die Entität „project.dataset.creator“ entspricht dem SEMCERIF-Term „semcerif:Creator“.

project.dataset.creator.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.dataset.creator.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wenn nachgenutzt, wer hat den Datensatz erzeugt?

project.dataset.description (Attribut)

Term: dc:description

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/description>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.dataset.description“ kann auf den DCMI-Term „dc:description“ gemappt werden.

Frage:

Um was für einen Datensatz handelt es sich?

project.dataset.format (Attribut)

Term: dc:format

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/format>

Description:

The file format, physical medium, or dimensions of the resource. Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types (MIME).

Anmerkung:

Das Attribut „project.dataset.format“ kann auf den DCMI-Term „dc:format“ gemappt werden. Die Benutzung eines kontrollierten Vokabulars wie MIME wird empfohlen.⁶⁸

Frage:

In welchen Formaten liegen die Daten vor?

project.dataset.id (Attribut)

Term: dc:identifier

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/identifier>

Description:

An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.

Anmerkung:

Das Attribut „project.dataset.id“ kann auf den DCMI-Term „dc:identifier“ gemappt werden.

project.dataset.ipr.owner (Entität)

Term: semcerif:Owner

Metadatenschema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Owner>

Description:

This term belongs to the 'Organisation Research Infrastructure Roles; Organisation Research Infrastructure Roles; Organisation Research Infrastructure Roles' classification scheme.

⁶⁸ [MIME]: Media Types. URL: <http://www.iana.org/assignments/media-types/>.

Anmerkung:

Die Entität „project.dataset.ipr.owner“ entspricht dem SEMCERIF-Term „semcerif:Owner“ und kann auf ihn gemappt werden.

project.dataset.ipr.owner.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.dataset.ipr.owner.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wer hält diese Rechte? Welche Nutzungsrechte werden eingeräumt bzw. wurden / werden eingeholt?

project.dataset.metadata.description (Attribut)

Term: dc:description

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/description>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.dataset.metadata.description“ kann auf den DCMI-Term „dc:description“ gemappt werden.

Frage:

Welche Standards, Ontologien, Klassifikationen etc. werden zur Beschreibung der Daten und Kontextinformation genutzt?

project.dataset.metadata.responsible_person.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.dataset.metadata.responsible_person.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wer ist verantwortlich für die Dokumentation und Prüfung der Metadaten und Kontextinformationen auf Richtigkeit und Vollständigkeit?

project.dataset.sharing_license (Attribut)

Term: dc:license

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/license>

Description:

A legal document giving official permission to do something with the resource.

Anmerkung:

Das Attribut „project.dataset.sharing_license“ kann auf den DCMI-Term „dc:license“ gemappt werden.

Frage:

Welche Lizenz soll hierfür verwendet werden?

project.dataset.size (Attribut)

Term: dc:extent

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/extent>

Description:

The size or duration of the resource.

Anmerkung:

Das Attribut „project.dataset.size“ kann auf den DCMI-Term „dc:extent“ gemappt werden. Für die Maßeinheiten bzw. Größenordnungen sollte ein kontrolliertes Vokabular verwendet werden.

Frage:

Was ist die erwartete Größe des Datensatzes?

project.dataset.storage.responsible_person.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.dataset.storage.responsible_person.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wer ist verantwortlich für die Erstellung der Backups?

project.dataset.storage.uri (Attribut)

Term: dc:uri

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/URI>

Description:

The set of identifiers constructed according to the generic syntax for Uniform Resource Identifiers as specified by the Internet Engineering Task Force.

Anmerkung:

Das Attribut „project.dataset.storage.uri“ entspricht dem DCMI-Term „dc:uri“.

Frage:

Unter welcher URL kann der Datensatz während des Projekts abgerufen werden?

project.dataset.usage_description (Attribut)

Term: dc:description

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/description>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.dataset.usage_description“ kann auf den DCMI-Term „dc:description“ gemappt werden.

Frage:

Wozu / wie wird dieser Datensatz während des Projektes genutzt?

project.funder (Entität)

Term: semcerif:Funder

Metadatenschema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Funder>

Description:

This term belongs to the 'Inter-Organisational Structure; Organisation Project Engagements; Organisation Output Roles; Organisation Output Roles; Organisation Output Roles; Organisation Output Roles; Output Funding Roles; Output Funding Roles; Output Funding Roles; Output Funding Roles; Research Infrastructure Funding Roles; Research Infrastructure Funding Roles; Research Infrastructure Funding Roles' classification scheme.

Anmerkung:

Die Entität „project.funder“ entspricht dem SEMCERIF-Term „semcerif:Funder“ und kann daher auf ihn gemappt werden.

project.funder.id (Attribut)

Term: dc:identifier

Metadatenchema: DCMI Terms

URI: <http://purl.org/dc/terms/identifier>

Description:

An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.

Anmerkung:

Das Attribut „project.funder.id“ kann auf den DCMI-Term „dc:identifier“ gemappt werden.

project.funder.name (Attribut)

Term: foaf:name

Metadatenchema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.funder.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wer fördert das Projekt?

project.funder.program (Entität)

Term: semcerif:FundingProgramme

Metadatenchema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#FundingProgramme>

Description:

This term belongs to the 'Funding Source Types' classification scheme.

Anmerkung:

Die Entität „project.funder.program“ entspricht dem SEMCERIF-Term „semcerif:Funder“ und kann daher auf ihn gemappt werden.

Frage:

In welchem Förderprogramm wird das Projekt gefördert?

project.funder.program.title (Attribut)

Term: dc:title

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/title>

Description:

A name given to the resource.

Anmerkung:

Das Attribut „project.funder.program.title“ kann auf den DCMI-Term „dc:title“ gemappt werden.

project.legal_aspects.other_sensitive_data.description (Attribut)

Term: dc:description

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/description>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.legal_aspects.other_sensitive_data.description“ kann auf den DCMI-Term „dc:description“ gemappt werden.

Frage:

Um welche nicht personenbezogenen sensiblen Daten handelt es sich?

project.partner (Entität)

Term: semcerif:Organisation

Metadatenchema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Organisation>

Description:

In CERIF, the concept of organisation is physically (cfOrgUnit) and logically (cfOrganisationUnit) defined as an entity in an ERM, described through basic attributes and through semantically neutral relationships with e.g. person (cfPers_OrgUnit), events (cfOrgUnit_Event), classification systems (cfOrgUnit_Class), projects (cfProj_OrgUnit). In the CERIF vocabulary, we introduce the concept of organisation. In the physical CERIF data model, the classification of types for organisations happens with the cfOrgUnit_Class entity, where a type is represented by a classification term (cfTerm) identified by its cfClassificationIdentifier (a uuid), and reused at instance level for assigning organisation types to organisation records. This term belongs to the 'CERIF Entities; Activity Subtypes; CERIF Entities' classification scheme.

Anmerkung:

Die Entität „project.partner“ kann auf den SEMCERIF-Term „semcerif:Organisation“ gemappt werden. Ein alternatives Mapping kann auf den CERIF-Term cfOrgUnit erfolgen.

project.partner.id (Attribut)

Term: dc:identifier

Metadatenchema: DCMI Terms

URI: <http://purl.org/dc/terms/identifier>

Description:

An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.

Anmerkung:

Das Attribut „project.partner.id“ kann auf den DCMI-Term „dc:identifier“ gemappt werden.

project.partner.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.partner.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Projektpartner

project.partner.contact_person (Entität)

Term: foaf:person

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_Person

Description:

A person. The Person class represents people. Something is a Person if it is a person. We don't nitpic about whether they're alive, dead, real, or imaginary. The Person class is a sub-class of the Agent class, since all people are considered 'agents' in FOAF.

Anmerkung:

Die Entität „project.partner.contact_person“ kann auf den FOAF-Term „foaf:Person“ gemappt werden. Ein alternatives Mapping kann auf den CERIF-Term cfPers erfolgen.

Frage:

Wer ist bei diesem Partner der/die Ansprechpartner/in für das Datenmanagement?

project.partner.contact_person.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.partner.contact_person.name“ entspricht dem FOAF-Term „foaf:name“.

project.preservation.responsible_person.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.preservation_responsible_person.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Durch wen erfolgt diese Auswahl?

project.coordination (Entität)

Term: semcerif:Coordinator

Metadatenschema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Coordinator>

Description:

In CERIF terms, coordinator is a role in e.g. the person-project relationship; that is, the classification term (cfTerm) "Coordinator" identified by its cfClassificationIdentifier (a uuid) and reused in link entities. This term belongs to the 'Person Project Engagements; Organisation Project Engagements' classification scheme.

Anmerkung:

Die Entität „project_coordination“ entspricht dem CERIF-Term „semcerif:Coordinator“.

project.coordination.name (Attribut)

Term: foaf:name

Metadatenschema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Die Entität „project_coordination“ enthält das Attribut „project_coordination.name“, welches dadurch zur gleichen Frage wie von der Entität „project_coordination“ gehört. Da bei RDMO nur das Attribut „project_coordination.name“ vorhanden ist und nicht in Vor- und Nachname unterteilt ist, macht ein Mapping auf den FOAF-Property „name“ Sinn, um grundlegende Interoperabilität zu ermöglichen. Eine noch feinere Unterteilung wäre u. a. mit den Properties „foaf:firstName“ und „foaf:lastName“ der FOAF-Klasse „Person“ möglich, allerdings müsste dafür ein weiteres Attribut in RDMO erstellt werden.

Frage:

Welche Personen oder Institutionen sind verantwortlich für die Projektkoordination?

project.research_field.title (Attribut)

Term: dc:title

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/title>

Description:

A name given to the resource.

Anmerkung:

Das Attribut „project.research_field.title“ kann auf den DCMI-Term „dc:title“ gemappt werden.

Frage:

Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen?

project.research_question.keywords (Attribut)

Term: dc:subject

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/subject>

Description:

The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.

Anmerkung:

Das Attribut „project.research_question.keywords“ kann auf das Property „dc:subject“ aus DCMI-Terms gemappt werden, so dass im Idealfall Keywords aus einem kontrollierten Vokabular benutzt werden können, um die Forschungsfrage näher zu beschreiben. Auch ein Mapping auf das Property „cerif:keyword“ wäre denkbar.

Frage:

Bitte geben sie einige Schlagworte zur Forschungsfrage an.

project.research_question.title (Attribut)

Term: dc:title

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/title>

Description:

A name given to the resource.

Anmerkung:

Das Attribut „project.research_question.title“ kann auf das DCMI-Term „dc:title“ gemappt werden.

Frage:

Wie lautet die primäre Forschungsfrage des Projektes?

project.schedule.end_date (Attribut)

Term: dc:date

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/date>

Description:

A point or period of time associated with an event in the lifecycle of the resource. Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601.

Anmerkung:

Das Attribut „project.schedule.end_date“ kann auf das DCMI-Term „dc:date“ gemappt werden. Die Benutzung eines Kodierungsschemas wird empfohlen. Auch ein Mapping auf das Property „cerif:endDate“ wäre denkbar.

Frage:

Wann endet die Projektlaufzeit?

project.schedule.start_date (Attribut)

Term: dc:date

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/date>

Description:

A point or period of time associated with an event in the lifecycle of the resource. Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601.

Anmerkung:

Das Attribut „project.schedule.start_date“ kann auf das DCMI-Term „dc:date“ gemappt werden. Die Benutzung eines Kodierungsschemas wird empfohlen. Auch ein Mapping auf das Property „cerif:startDate“ wäre denkbar.

Frage:

Wann beginnt die Projektlaufzeit?

project.software (Entität)

Term: dc:software

Metadatenchema: DCMI Terms

URI: <http://purl.org/dc/dcmitype/Software>

Description:

A computer program in source or compiled form. Examples include a C source file, MS-Windows .exe executable, or Perl script.

Anmerkung:

Die Entität „project.software“ kann auf den DCMI-Term „dc:software“.

project.software.creator (Entität)

Term: semcerif:Creator

Metadatenchema: SEMCERIF

URI: <http://eurocris.org/ontologies/semcerif/1.3/#Creator>

Description:

This term belongs to the 'Person Output Contributions' classification scheme.

Anmerkung:

Die Entität „project.software.creator“ entspricht dem SEMCERIF-Term „semcerif:Creator“.

project.software.creator.name (Attribut)

Term: foaf:name

Metadatenchema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.software.creator.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wenn nachgenutzt, wer hat den Code entwickelt?

project.software.description (Attribut)

Term: dc:description

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/description>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.software.description“ kann auf den DCMI-Term „dc:description“ gemappt werden.

Frage:

Um was für eine Art Code handelt es sich?

project.software.id (Attribut)

Term: dc:identifier

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/identifier>

Description:

An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Anmerkung:

Das Attribut „project.software.id“ kann auf den DCMI-Term „dc:identifier“ gemappt werden.

project.software.ipr.owner.name (Attribut)

Term: foaf:name

Metadatenchema: FOAF

URI: http://xmlns.com/foaf/spec/#term_name

Description:

A name for some thing. The name of something is a simple textual string.

Anmerkung:

Das Attribut „project.software.ipr.owner.name“ entspricht dem FOAF-Term „foaf:name“.

Frage:

Wer hält diese Rechte? Welche Nutzungsrechte werden eingeräumt bzw. wurden / werden eingeholt?

project.software.sharing_license (Attribut)

Term: dc:license

Metadatenchema: DCMI Terms

URI: <http://purl.org/dc/terms/license>

Description:

A legal document giving official permission to do something with the resource.

Anmerkung:

Das Attribut „project.software.sharing_license“ kann auf den DCMI-Term „dc:license“ gemappt werden.

Frage:

Welche Lizenz soll hierfür verwendet werden?

project.software.title (Attribut)

Term: dc:title

Metadatenchema: DCMI Terms

URI: <http://purl.org/dc/terms/title>

Description:

A name given to the resource.

Anmerkung:

Das Attribut „project.software.title“ kann auf den DCMI-Term „dc:title“ gemappt werden.

project.software.uri (Attribut)

Term: dc:uri

Metadatenschema: DCMI Terms

URI: <http://purl.org/dc/terms/URI>

Description:

The set of identifiers constructed according to the generic syntax for Uniform Resource Identifiers as specified by the Internet Engineering Task Force.

Anmerkung:

Das Attribut „project.software.uri“ entspricht dem DCMI-Term „dc:uri“.

Frage:

Wenn nachgenutzt, unter welcher Adresse oder URL ist der Code verfügbar?

7.3 Technische Implementierung

Grundlage für die Erstellung des Renderers für den XML-Output ist der Django REST Framework XML-Renderer⁶⁹. Darin wird bereits eine grundlegende XML-Ausgabe-Funktionalität geboten, die an das RDMO angepasst und erweitert werden muss.

Ausgangspunkt ist hierbei die von RDMO verwendete Datenbank (SQLite). Die Datenbankabstraktion von Django übersetzt die Datenbankabfrage auf Python-Objekte, sog. QuerySets⁷⁰, und übergibt es an den Serializer des Django REST Frameworks. Dort wird aus den Daten ein OrderedDict-Objekt⁷¹ erstellt. Auf dieses wiederum kann als das Attribut „data“ des Serializer-Objekts zugegriffen werden. Anschließend werden die Daten bzw. „data“ an den Renderer des Django REST Frameworks übergeben. Dieser Renderer ist für den Output der Daten verantwortlich und legt fest, in welchem Format diese ausgegeben werden. Genau hier setzt die vorliegende

⁶⁹ Padilla: REST Framework XML. URL: <https://github.com/jpadilla/django-rest-framework-xml>.

⁷⁰ Django Software Foundation: QuerySet API reference. URL: <https://docs.djangoproject.com/ja/1.9/ref/models/querysets/>.

⁷¹ Python Software Foundation: collections - Container datatypes. URL: <https://docs.python.org/3/library/collections.html#collections.OrderedDict>.

Masterarbeit an und formuliert den entsprechenden Python-Programmcode für den Renderer, um Daten im XML-Format auszugeben. Die Daten werden anschließend per HTTP an den Client gesendet.

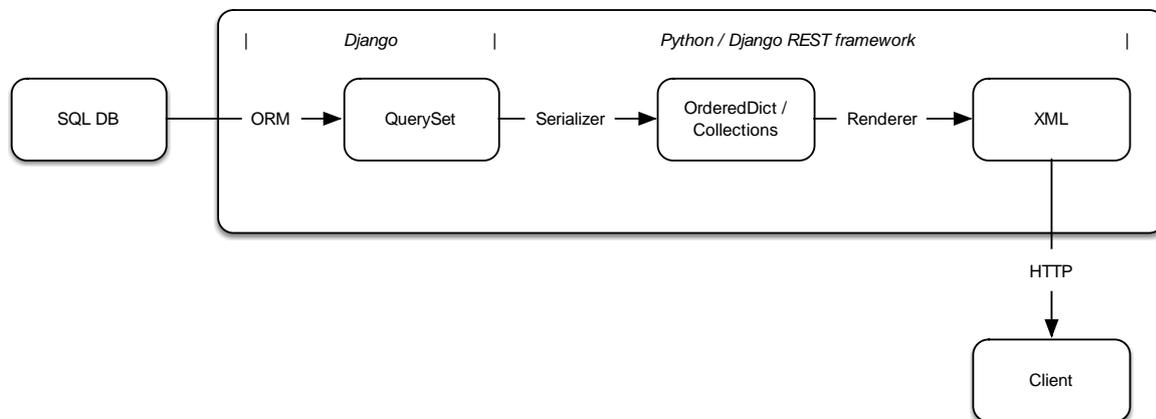


Abbildung 5: Ablaufschema Datenumwandlung

Der Aufbau und die Struktur der XML-Modellierung basieren auf den Richtlinien der DCXML-Terms⁷². Diese geben Hinweise und Ratschläge dazu, welche Daten als Attribut zu einem Element hinzugefügt werden sollten und welche Daten eher ein eigenes Element sein sollten. Im vorliegenden Python-Code für den XML-Output werden die ausgegebenen Entitäten und Attribute jeweils als eigene XML-Elemente bzw. XML-Unterelemente erstellt und auf die Vergabe von XML-Attributen verzichtet, um eine größtmögliche Flexibilität und Interoperabilität zu ermöglichen. Lediglich das Wurzelement „Domain“ erhält als XML-Attribut den Namespace von Dublin Core, damit die RDMO-Attribute „title“, „description“ und „uri“ ihren syntaktischen Äquivalenten aus DCMI Terms zugeordnet werden können.

```

from __future__ import unicode_literals

from django.utils import six
from django.utils.xmlutils import SimplifiedXMLGenerator
from django.utils.six.moves import StringIO
from django.utils.encoding import smart_text
from rest_framework.renderers import BaseRenderer

class XMLRenderer(BaseRenderer):
    media_type = 'application/xml'
  
```

⁷² Dublin Core Metadata Initiative: Expressing Dublin Core metadata using the Resource Description Framework. URL: <http://dublincore.org/documents/dc-rdf/>.

```
format = 'xml'
```

Die Klasse XMLRenderer aus dem REST-Framework erbt von der Klasse BaseRenderer, u. a. die Output-Funktionalitäten zur Datenausgabe. Damit wird die Funktionalität vom XMLRenderer konkretisiert.

```
def render(self, data):  
    if data is None:  
        return ''
```

„render“ ist die Methode des XMLRenderers, die den XML-String ausgibt. An diese wird „data“ übergeben und als OrderedDict ausgegeben.

```
stream = StringIO()
```

„StringIO“ aus dem XMLRenderer wird übernommen und das „stream“-Objekt wird instanziiert.

```
xml = SimplerXMLGenerator(stream, "utf-8")  
xml.startDocument()  
xml.startElement('Domain', {  
    'xmlns:dc': "http://purl.org/dc/elements/1.1/"  
})
```

Der XMLGenerator wird instanziiert und das „stream“-Objekt, in dem später die XML-Daten gespeichert werden, wird übergeben. Der XMLGenerator erlaubt die prozedurale Erzeugung eines XML-Dokuments. Anschließend wird das XML-Dokument mit dem Root-Element „Domain“ gestartet.

```
for attribute_entity in data:  
    if attribute_entity['is_attribute']:  
        self._attribute(xml, attribute_entity)  
    else:  
        self._attribute_entity(xml, attribute_entity)
```

Es wird geprüft, ob es sich um ein Attribut oder um eine AttributeEntity handelt. Ist das Element „is_attribute“ vorhanden und „true“, ist es ein Attribut, ansonsten ist es eine AttributeEntity. Die beiden Elemente Attribute und AttributeEntity sowie ihre Unterelemente werden in eigenen ausgelagerten Funktionen erstellt, welche weiter unten im Code implementiert werden.

```
xml.endElement('Domain')
xml.endDocument()
return stream.getvalue()
```

Das Root-Element „Domain“ und das XML-Dokument werden geschlossen. Die Werte von „stream“ werden ausgelesen.

```
def _attribute(self, xml, attribute):
    xml.startElement('Attribute', {})
    self._text_element(xml, 'dc:title', {}, attribute["title"])
    self._text_element(xml, 'dc:description', {}, attribute["description"])
    self._text_element(xml, 'dc:uri', {}, attribute["uri"])
    self._text_element(xml, 'is_collection', {}, attribute["is_collection"])
    self._text_element(xml, 'value_type', {}, attribute["value_type"])
    self._text_element(xml, 'unit', {}, attribute["unit"])

    if 'options' in attribute and attribute['options']:
        xml.startElement('Options', {})

        for option in attribute['options']:
            self._option(xml, option)

        xml.endElement('Options')

    if 'range' in attribute and attribute['range']:
        self._range(xml, attribute['range'])

    if 'conditions' in attribute and attribute['conditions']:
        xml.startElement('Conditions', {})

        for conditions in attribute['conditions']:
            self._conditions(xml, conditions)

        xml.endElement('Conditions')

    if 'verbose_name' in attribute and attribute['verbose_name']:
        self._verbose_name(xml, attribute['verbose_name'])
```

```
xml.endElement('Attribute')
```

Die Attribute-Elemente (title, description, uri, is_collection, value_type, unit, options, range, conditions und verbose_name) und ihre entsprechenden Unterelemente werden - falls vorhanden - erstellt. Anschließend wird das Attribute-Element geschlossen. Lagerst aus in andere Funktionen, bessere Lesbarkeit

```
def _attribute_entity(self, xml, attribute_entity):
    xml.startElement('AttributeEntity', {})
    self._text_element(xml, 'dc:title', {}, attribute_entity["title"])
    self._text_element(xml, 'dc:description', {}, attribute_entity["description"])
    self._text_element(xml, 'dc:uri', {}, attribute_entity["uri"])
    self._text_element(xml, 'is_collection', {}, attribute_entity["is_collection"])

    if 'children' in attribute_entity:
        xml.startElement('Children', {})

        for child in attribute_entity['children']:
            if child['is_attribute']:
                self._attribute(xml, child)
            else:
                self._attribute_entity(xml, child)

        xml.endElement('Children')
```

Es wird anhand des Elements „is_attribute“ geprüft, ob im jeweiligen Attribut oder AttributeEntity Kinderelemente vorhanden sind. Sollte dies der Fall sein, wird das „Children“-Element und für jedes Kinderelement ein „child“-Element erstellt.

```
if 'conditions' in attribute_entity and attribute_entity['conditions']:
    xml.startElement('Conditions', {})

    for conditions in attribute_entity['conditions']:
        self._conditions(xml, conditions)

    xml.endElement('Conditions')

if 'verbose_name' in attribute_entity and attribute_entity['verbose_name']:
```

```
self._verbose_name(xml, attribute_entity['verbose_name'])
```

```
xml.endElement('AttributeEntity')
```

Die AttributeEntity-Elemente (title, description, uri, is_collection, conditions und verbose_name) und ihre entsprechenden Unterelemente werden - falls vorhanden - erstellt. Zur besseren Lesbarkeit wird der dafür zuständige Code in anderen Funktionen weiter unten im Programm ausgelagert. Anschließend wird das AttributeEntity-Element geschlossen.

```
def _option(self, xml, option):  
    xml.startElement('Option', {})  
    self._text_element(xml, 'order', {}, option["order"])  
    self._text_element(xml, 'text_de', {}, option["text_de"])  
    self._text_element(xml, 'text_en', {}, option["text_en"])  
    self._text_element(xml, 'additional_input', {}, option["additional_input"])  
    xml.endElement('Option')
```

Falls bei der Prüfung weiter oben im Code festgestellt wird, dass das Element „options“ in einem Attribute-Element vorhanden ist, werden mit der Funktion „_option“ die Unterelemente (order, text_de, text_en und additional_input) von „options“ erstellt.

```
def _range(self, xml, range):  
    xml.startElement('range', {})  
    self._text_element(xml, 'minimum', {}, range["minimum"])  
    self._text_element(xml, 'maximum', {}, range["maximum"])  
    self._text_element(xml, 'step', {}, range["step"])  
    xml.endElement('range')
```

Falls bei der Prüfung weiter oben im Code festgestellt wird, dass das Element „range“ in einem Attribute-Element vorhanden ist, werden mit der Funktion „_range“ die Unterelemente (minimum, maximum und step) von „range“ erstellt.

```
def _conditions(self, xml, conditions):  
    xml.startElement('condition', {})  
    self._text_element(xml, 'source_attribute', {}, conditions["source_attribute"])  
    self._text_element(xml, 'relation', {}, conditions["relation"])  
    self._text_element(xml, 'target_text', {}, conditions["target_text"])  
    self._text_element(xml, 'target_option', {}, conditions["target_option"])
```

```
xml.endElement('condition')
```

Falls bei der Prüfung weiter oben im Code festgestellt wird, dass das Element „conditions“ in einem Attribute-Element oder einem AttributeEntity-Element vorhanden ist, werden mit der Funktion „_condition“ die Unterelemente (source_attribute, relation, target_text und target_option) von „conditions“ erstellt.

```
def _verbosename(self, xml, verbosename):
    xml.startElement('verbosename', {})
    self._text_element(xml, 'name_en', {}, verbosename["name_en"])
    self._text_element(xml, 'name_de', {}, verbosename["name_de"])
    self._text_element(xml, 'name_plural_en', {}, verbosename["name_plural_en"])
    self._text_element(xml, 'name_plural_de', {}, verbosename["name_plural_de"])
    xml.endElement('verbosename')
```

Falls bei der Prüfung weiter oben im Code festgestellt wird, dass das Element „verbosename“ in einem Attribute-Element oder einem AttributeEntity-Element vorhanden ist, werden mit der Funktion „_verbosename“ die Unterelemente (name_en, name_de, name_plural_en und name_plural_de) von „verbosename“ erstellt.

```
def _text_element(self, xml, tag, attributes, text):
    xml.startElement(tag, attributes)
    if text is not None:
        xml.characters(smart_text(text))
    xml.endElement(tag)
```

Die Funktion „text_element“ wird definiert, um die häufige Erstellung von XML-Elementen zu vereinfachen und den geschriebenen Code im restlichen Teil möglichst übersichtlich und gering zu halten.

Aus der in der Datenbank gespeicherten Entität „project.research_question“ zur Forschungsfrage des Projekts (Wie lautet die primäre Forschungsfrage des Projektes?) und dem dazugehörigen Attributen „project.research_question.keywords“ (Bitte geben sie einige Schlagworte zur Forschungsfrage an.) und „project.research_question.title“ rendert die Implementation folgenden XML-Code:

```

- <AttributeEntity>
  <dc:title>research_question</dc:title>
  <dc:description/>
  <dc:uri/>
  <is_collection>False</is_collection>
- <Children>
  - <Attribute>
    <dc:title>keywords</dc:title>
    <dc:description/>
    <dc:uri/>
    <is_collection>True</is_collection>
    <value_type>text</value_type>
    <unit/>
  - <verbosename>
    <name_en>keyword</name_en>
    <name_de>Schlagwort</name_de>
    <name_plural_en>keywords</name_plural_en>
    <name_plural_de>Schlagwörter</name_plural_de>
  </verbosename>
</Attribute>
- <Attribute>
  <dc:title>title</dc:title>
  <dc:description/>
  <dc:uri/>
  <is_collection>False</is_collection>
  <value_type>text</value_type>
  <unit/>
</Attribute>
</Children>
</AttributeEntity>

```

Abbildung 6: XML-Ausgabe

Der zusammenhängende Programmcode für den XML-Export ist als Anlage 4 beigefügt.

8 Diskussion

Die in dieser Masterarbeit beschrieben und erzielten Ergebnisse sind die ersten Schritte in der Richtung zur Ermöglichung und Erzielung von Interoperabilität, sowohl zwischen verschiedenen Instanzen von RDMO, als auch zu anderen Schnittstellen, wie z. B. Forschungsinformationssystemen.

Probleme gibt es noch beim Mapping von Attributen und Entitäten zu anderen Vokabularen. Das baumartige RDMO-Datenmodell lässt sich häufig nur unzureichend auf umfangreiche und sehr detaillierte Datenmodelle wie den europäischen CERIF-Standard mappen. Oftmals handelt es sich bei passenden semantischen Äquivalenten im Zielvokabular um Klassen, beim entsprechenden Äquivalent in RDMO jedoch um ein Attribut und nicht um eine Entität, so dass ein Mapping nicht erfolgen kann. Das gleiche gilt für Properties im Zielvokabular und Entitäten in RDMO. Eine Anpassung des RDMO-Datenmodells und somit die Umwandlung von einigen Entitäten in Attribute und andersherum wäre an einigen Stellen denkbar, jedoch aus Gründen der internen Programmlogik nicht immer möglich. Bei der Verwendung von SEMCERIF sollte auch bedacht werden, dass dies bisher nur in einer Version verfügbar ist, die auf dem älteren CERIF 1.5 basiert und nicht auf der aktuellen Version 1.6. Vokabulare wie der Kerndatensatz Forschung legen den Fokus des Metadatenschemas auf wesentlich spezifischere Bereiche, wie z. B. Publikationen, oder sind, wie im Fall von CASRAI, eher standardisierte Wörterbücher bzw. Glossare. Generell sollten in der verbleibenden Projektlaufzeit die bisher ausgewählten Vokabulare noch einmal genau untersucht und das RDMO ggf. angepasst und erweitert werden. Langfristig gesehen sollte die Erstellung eines eigenen Metadata Application Profiles für RDMO entstehen und ein dazugehöriges Klassifikationsschema geschaffen werden. Auch die Funktion zum Import und Einlesen von exportierten (XML-)Daten muss in der weiteren Projektlaufzeit noch programmiert werden, damit eine Interoperabilität zwischen verschiedenen Instanzen von RDMO hergestellt wird.

Die Entwicklung eines Werkzeuges wie RDMO stellt auf jeden Fall einen Schritt in die richtige Richtung dar, wie auch die Resonanz von der Präsentation von RDMO auf Konferenzen und das Nutzerfeedback des ersten Workshops gezeigt hat. Es füllt eine Lücke im Forschungsdatenmanagement aus, die vor allem in Deutschland und Europa spürbar ist. Bisherige Ansätze haben entweder das Problem, dass sie sich zu stark an etablierten DMP Tools, wie z. B. DMPonline und DMPTool, orientieren und dadurch nicht ausreichend an die deutsche bzw. europäische Situation des Forschungsdatenmanagements angepasst sind (z. B. bei rechtlichen Aspekten, Vorgaben des Förderers etc.). Auch die Zugänglichkeit ist oftmals ein Problem, denn einige Tools (z. B. DMP Tool der Universität Bielefeld) sind nur für Universitätsangehörige nutzbar und können somit per se nicht von anderen Institutionen für ein Forschungsprojekt eingesetzt werden.

Dadurch wird der potentielle Nutzerkreis von vornherein stark limitiert, was auch zur Folge hat, dass es für diese Tools schwieriger ist, eine aktive Community mit regem Austausch aufzubauen,

die in der Zukunft vielleicht auch Ideen und Anregungen einbringt, um die weitere Entwicklung des Tools aktiv mitzugestalten.

Eine weitere Einschränkung vieler DMP-Tools ergibt sich durch den häufig sehr disziplinspezifischen Fokus. Ein generischer Ansatz ist zwar häufig erkennbar (DMPonline, DMPTool), jedoch nicht konsequent genug umgesetzt. Durch die fehlende oder oft nur zeit- und kostenintensive Anpassbarkeit der Tools an die eigene Wissenschaftsdisziplin ist ein Einsatz der Tools nur begrenzt möglich oder mit Einschränkungen verbunden. Da bei RDMO die Anpassung eines vorhandenen Fragenkatalogs und sogar die komplette Neuerstellung eines eigenen Fragenkatalogs auch ohne Programmierkenntnisse durchgeführt werden kann, liegt die Hürde zum Einsatz deutlich niedriger als bei bisherigen DMP-Tools.

Idealerweise gibt ein möglichst generischer Fragenkatalog auch Denkanstöße zu Dingen, die beachtet und in der Projektlaufzeit (und oft auch darüber hinaus) berücksichtigt werden müssen, z. B. die Nutzungsbedingungen und Lizenzen zur Veröffentlichung und Weitergabe von Projektdatensätzen und Überlegungen zur Langzeitarchivierung. Ein generischer Fragenkatalog spricht möglichst viele verschiedene Wissenschaftsdisziplinen an und zeigt für alle Gebiete relevante Überlegungen auf, die in einem Projekt generell eine wichtige Rolle spielen. Ein Fragenkatalog kann jedoch niemals hundertprozentig generisch sein und eine völlige Abdeckung sämtlicher Belange, Fallstricke und Eventualitäten ist schlichtweg nicht möglich und auch gar nicht das Ziel. Dennoch kann, insbesondere, wenn dieser Aspekt bei der Entwicklung eines DMP-Tools von Beginn an angestrebt und stets berücksichtigt wird, ein recht hoher Grad an Allgemeingültigkeit erreicht werden. Das frühzeitige Einholen von Nutzerfeedback, z. B. durch Workshops und Erfahrungen von Early Adopters, ist dabei sehr hilfreich und sollte weiterverfolgt werden. Ein hoher Grad an Diversität der Fachrichtungen bei der Auswahl der Teilnehmer ist ebenfalls ein wichtiger Faktor, der weiterhin berücksichtigt werden sollte, denn jede Fachdisziplin stellt andere Anforderungen an einen DMP-Fragenkatalog. Im Fall von RDMO wurde frühzeitig das Feedback Sonnenphysiker des GREGOR Fabry-Pérot-Interferometer und Sozialwissenschaftlicher der Zwillingsstudie TwinLife eingeholt. Die dadurch gewonnenen Erkenntnisse flossen nicht nur in die allgemeine Gestaltung der Benutzeroberfläche ein, sondern sollen auch dafür verwendet werden, zwei auf diese beiden Fachdisziplinen angepasste Fragenkataloge für RDMO zu erstellen.

Angesichts des zur Verfügung stehenden Budgets und der restlichen Laufzeit des Projekts können viele Aspekte nur angedacht werden, aber nicht sämtliche Punkte bis zum Ende durchdekliniert bzw. umgesetzt werden. Deshalb ist es auch besonders wichtig, dass frühzeitig ein Netzwerk mit fachspezifischen Partnern aufgebaut wird. Diese können schon frühzeitig Meinungen zu fachspezifischen Ansichten und Standpunkten an die Projektverantwortlichen weitergeben, was von diesen sonst nicht geleistet werden kann. Aufgepasst werden sollte dabei jedoch auch,

dass wiederum nicht zu viele institutionelle Sichten eingeholt werden. Zu viele verschiedene Blickwinkel von Hochschulverwaltungen und Forschungsinformationssystemen können sonst schnell dazu führen, dass der ursprüngliche Fokus von RDMO unscharf wird und/oder zu sehr in den Hintergrund rückt. Eine Lösung für dieses potentielle Problem können ausgewählte Early Adopter sein, also frühzeitige Anwender und Tester, die sich RDMO schon während der Entwicklung anschauen, einsetzen und versuchen, das Tool dementsprechend zu erweitern und an ihre institutionellen und disziplinspezifischen Anforderungen anzupassen. Generell muss eine Identifikation relevanter Partner (Fachwissenschaft, Wissenschaftseinrichtungen, EU-Partner) erfolgen. Dabei ist es auch unabdingbar, dass die verschiedenen institutionellen Tester regelmäßig mit den Projektverantwortlichen zusammenkommen, um sich abzustimmen. Zudem erfordert so ein Vorgehen auch, dass das Prinzip dieser verteilten Programmierung funktioniert und am Projektende ein ausgereiftes und durchdachtes Tool zur Verfügung steht. Speziell bei diesen Schritten, aber auch insgesamt, ist es ausgesprochen wichtig seitens der Projektverantwortlichen deutlich aufzuzeigen, wo die Grenzen von RDMO liegen und die Grenzen des Projekts klar definiert werden. Es muss während der Projektlaufzeit fortwährend darauf geachtet werden, welche Aspekte RDMO nicht abdecken kann und vor allem, was es letztendlich auch gar nicht leisten will. Besonders groß ist hierbei die Gefahr, dass RDMO mit Anforderungen und Funktionalitäten des Projektmanagements überfrachtet wird und der Aspekt der Forschungsdatenverwaltung in den Hintergrund tritt. Hier stellt sich die grundsätzliche Frage, wie das Verhältnis zwischen einer Strukturvorgabe und der Bereitstellung eines konfigurierbaren Systems aussehen soll. Ziel sollte es nach wie vor sein, den involvierten Akteuren des Forschungsdatenmanagements eines Forschungsprojekts ein Tool zur Unterstützung über den gesamten Projektzeitraum an die Hand zu geben. RDMO sollte dabei eher als Baukasten gesehen werden, der bei Bedarf vielfältige Anpassungen an Institutionen und Disziplinen ermöglicht, jedoch nicht immer alle individuellen Bedürfnisse zwangsläufig befriedigt. Dennoch sollte eine Balance gewahrt werden, bei der Fehler, Features und Erweiterungswünsche gemeldet werden können und alle Interessierten und potentiellen Nutzer dazu eingeladen sind Early Adopter und Backend-Tester zu werden. Für diese Pilotanwender ist eine entsprechende Organisationsstruktur nötig, die dabei nicht zu viele Ressourcen des Projekts beanspruchen darf.

9 Ausblick

Offensichtlich noch ausstehende Erfordernisse sind, wie eingangs im vorherigen Kapitel bereits angesprochen, sowohl die Programmierung einer (XML-)Import-Funktionalität, damit exportierte (XML-)Daten wiedereingelesen werden können, als auch die weitere Verfeinerung und langfristig sicherlich auch die Erweiterung und Überarbeitung des Mappings. Für fehlende semantische Äquivalente muss eine eigene RDMO-Definition des Vokabulars erfolgen und ein konkretes Metadata Application Profile für RDMO entwickelt werden, um langfristig eine solide Interoperabilität zu leisten.

Für die potentielle Langlebigkeit von RDMO und die Entwicklung von zukünftigen Features wäre es natürlich ideal, wenn sich auch nach Ablauf des Projektes eine Community (z. B. bei GitHub) um RDMO bilden würde. Auch ein Forum, bei dem sich die Nutzer austauschen können, wäre denkbar und wünschenswert und würde langfristig dazu beitragen, dass RDMO mit weiteren Funktionen ergänzt wird und eine größere Verbreitung erfährt. Um den Grundstein für diese vorstellbare Art der langfristigen Community und Vernetzung zu legen, ist eine frühe Einbeziehung der Forschungsdatencommunity und denkbarer Partner und Benutzer unabdingbar und sinnvoll. Die Organisation dieses verteilten Programmierens (Community-Building-Code) sollte ebenfalls möglichst frühzeitig durch das RDMO-Projekt erfolgen. Ende Juni 2016 wurde bereits ein erster RDMO-Workshop am Leibniz-Institut für Astrophysik durchgeführt. Dort wurde eine frühe, aber dennoch funktionsfähige Vorabversion von RDMO ausgewählten Interessenten vorgestellt und auch ein eigenständiges Ausprobieren ermöglicht. Auf dem Workshop wurde nicht nur ein generelles Interesse an einem generischen Werkzeug zum Forschungsdatenmanagement festgestellt, sondern auch Wünsche und Anregungen eingeholt, die in der restlichen Projektlaufzeit noch angedacht und eventuell auch umgesetzt werden können. Dabei stellten sich, je nach Fachdisziplin- und Institutionszugehörigkeit, teilweise sehr unterschiedliche Sichten auf RDMO und der Nachfrage nach bestimmten Funktionen heraus. Der Aspekt des Community Buildings sollte in der verbleibenden Projektlaufzeit auf jeden Fall weiter beachtet und ausgebaut werden, um langfristig die Relevanz von RDMO zu gewährleisten.

Generell wird von den Nutzern auf eine möglichst visuell ansprechende Übersicht Wert gelegt, bei welcher sämtliche Fragen überschaubar und thematisch geordnet sind. Hinweise, Vorschläge und auch Warnungen bei den einzelnen Fragen(-themen), wie z. B. bei Datenrichtlinien und Richtlinien der eigenen Institution, sind genauso gewünscht, wie auch eine grundlegende Hilfefunktion und ganz konkrete Anwendungsbeispiele und die Auskunft zu Best-Practice-Verfahren. Generische Hilfetexte sind jedoch nur bis zu einem bestimmten Grad möglich, so dass hier auch das Engagement der verschiedenen Fachdisziplinen zur eigenen Anpassung gefordert ist. Auch die Implementierung der verschiedenen Ansichten des Fragenkatalogs für die unterschiedlichen

Stakeholder bzw. Akteure im Forschungsdatenmanagement, z. B. ein Gutachter-View, ist eine oft geforderte Funktionalität, die noch umgesetzt werden muss. Eigene Zugänge zum kollaborativen Arbeiten mit verschiedenen Sicht- und Schreibrechten gehen damit ebenfalls einher, genauso wie das Herstellen eines zeitlichen Bezugs auf Workflows zur Unterstützung und bestimmte Ansichten für bestimmte Zeitpunkte der Projektlaufzeit. Grundsätzlich sollte es auch eine Möglichkeit geben, Dokumente, Bilder etc. hochzuladen, anzuhängen und einzubetten, wie es bei vielen anderen kollaborativen Tools möglich ist. Können mehrere Nutzer an den gleichen Daten arbeiten, muss auch eine Dokumentation der Fehlerbehebungen bzw. eine Liste der erfolgten Änderungen einsehbar sein und vielleicht auch alle verschiedenen Release-Versionen eines Datensatzes als Snapshot verfügbar sein, die während der Projektlaufzeit angefallen sind. All dies sind, entsprechend des Workshop-Feedbacks populäre Features, die noch umgesetzt werden sollten, um ein modernes und konkurrenzfähiges Forschungsdatenwerkzeug in der Community zu etablieren.

Aus institutioneller Perspektive ist vor allem die Anbindung und Integration von RDMO in bestehende Forschungsinformationssysteme interessant, der insbesondere für Universitäten und Fachhochschulen eine wichtige Überlegung für den Einsatz eines DMP-Tools darstellt. In Publikationsdatenbanken sollte es zudem möglich sein, eine Verknüpfung zwischen Publikationen und den dazugehörigen Forschungsdaten zu schaffen. Zum effektiven kollaborativen Arbeiten ist außerdem das Einfügen einer Funktion zum Gruppenmanagement essentiell, bei dem verschiedenen Nutzern unterschiedliche Zugriffs- und Schreibrechte zugewiesen werden können. Auch eine Weiterleitungsfunktion zu bestimmten Kontaktpersonen an der jeweiligen Institution bei Fragen, Hinweisen etc. ist erforderlich. Weiterhin sollten die Ethikrichtlinien der jeweiligen Fachdisziplin bzw. Institution Berücksichtigung finden. Aus fachspezifischer Sicht muss neben einer Angleichung der Fragen auch eine Anpassung der Hinweis- und Hilfetexte erfolgen. Hierbei wären auch eigene Plug-Ins für disziplinspezifische Fragen, Texte und relevanter Fachgesellschaften eine denkbare Lösung.

Für die restliche Projektlaufzeit ist generell geplant, die Mehrsprachigkeit von RDMO mit weiteren EU-Sprachen und die Skalierbarkeit zur Anpassung von RDMO auszubauen. Des Weiteren ist u. a. angedacht, schon vorliegende Fragenkataloge von Fachdisziplinen einzulesen und zu prüfen, ob dies ohne größere Probleme möglich ist. Ein nächster denkbarer und sinnvoller Schritt ist die Einpflege bzw. Erstellung von Datenmanagementplänen, die bereits auf die Anforderungen bestimmter Förderer zugeschnitten sind. Dann wäre auch eine Kollektion von Fragenkatalogen für verschiedene Fachdisziplinen umsetzbar, die in RDMO standardmäßig integriert sind. Der dafür nötige inhaltliche Input muss in Kooperation mit den Fachdisziplinen gesammelt werden. Anfang 2017 könnte vor Ende der bisherigen Projektlaufzeit noch ein weiterer RDMO-Workshop stattfinden, bei dem diese Themen erarbeitet werden können. Langfristige Ziele, die

in einem Nachfolgeantrag berücksichtigt werden könnten, sind u. a. die Umsetzungen der o. g. fachspezifischen und organisationsspezifischen Komponenten. Dazu gehört auch die Schaffung von weiteren Infrastrukturschnittstellen, sowohl zu lokalen Forschungsinformationssystemen, als auch überregionalen Forschungsdatenrepositorien wie re3data⁷³. Auch die Formulierung von Metadaten in normierter Weise für die Entwicklung eines vollständigen Metadata Application Profiles und weiteren Metadaten-schnittstellen (wie z. B. Kerndatensatz Forschung) sind für einen größeren Austausch erstrebenswerte Ziele.

Für die ursprünglich geplante Konzeption ist der aktuelle Stand von RDMO bereits erfreulich und wird in der restlichen Projektlaufzeit noch weiter verbessert werden. Die in dieser Masterarbeit erarbeiteten Ergebnisse sind dabei als Beginn eines Prozesses zu verstehen und sollen dabei helfen, das komplexe Feld der Interoperabilität im Kontext von RDMO erfolgreich zu komplettieren.

⁷³ re3data.org Project Consortium: Registry of Research Data Repositories. URL: <http://www.re3data.org/>.

Quellenverzeichnis

- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock + Herchen, 2011.
- Bundesministerium für Bildung und Forschung (Hrsg.): 7. EU-Forschungsrahmenprogramm im Überblick. URL: <http://www.forschungsrahmenprogramm.de/frp-ueberblick.htm>. [letzter Zugriff: 14.08.2016].
- Bundesministerium für Bildung und Forschung (Hrsg.): Der Europäische Forschungsraum. URL: <http://www.horizont2020.de/einstieg-era.htm>. [letzter Zugriff: 14.08.2016].
- Bundesministerium für Bildung und Forschung (Hrsg.): Förderkatalog. URL: <http://foerderportal.bund.de/foekat/jsp/StartAction.do>. [letzter Zugriff: 14.08.2016].
- Bundesministerium für Bildung und Forschung (Hrsg.): Horizont 2020 - das europäische Forschungsrahmenprogramm. URL: <https://www.bmbf.de/de/horizont-2020-das-europaeische-forschungsrahmenprogramm-281.html>. [letzter Zugriff: 14.08.2016].
- Bundesministerium für Bildung und Forschung (Hrsg.): Programmaufbau von Horizon 2020. URL: <http://www.horizont2020.de/einstieg-programmaufbau.htm>. [letzter Zugriff: 14.08.2016].
- Bundesministerium für Bildung und Forschung (Hrsg.): Europa 2020 Strategie. URL: http://www.horizont2020.de/img/content/Europa2020_Strategie_Web.jpg. [letzter Zugriff: 14.08.2016].
- California Digital Library. URL: <http://www.cdlib.org/>. [letzter Zugriff: 14.08.2016].
- California Digital Library (Hrsg.): Data Management Planning Tool. URL: <https://dmptool.org/>. [letzter Zugriff: 14.08.2016].
- CASRAI. URL: <http://casrai.org/about>. [letzter Zugriff: 14.08.2016].
- CASRAI (Hrsg.): The CASRAI Dictionary. URL: http://dictionary.casrai.org/Main_Page. [letzter Zugriff: 14.08.2016].
- Christie, Tom: Django REST framework. URL: <http://www.django-rest-framework.org/>. [letzter Zugriff: 14.08.2016].
- Clarín-D: Data Management Plan. URL: <http://www.clarin-d.de/en/preparation/data-management-plan>. [letzter Zugriff: 14.08.2016].
- Creative Commons (Hrsg.): Namensnennung 2.0 Deutschland. URL: <https://creativecommons.org/licenses/by/2.0/de/>. [letzter Zugriff: 14.08.2016].

- Creative Commons (Hrsg.): Public Domain Dedication. URL: <https://creativecommons.org/publicdomain/zero/1.0/deed.de>. [letzter Zugriff: 14.08.2016].
- Cremer, Fabian; Engelhardt, Claudia; Neuroth, Heike: Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale. In: Bonte, Achim; Degkwitz, Andreas; Horstmann, Wolfram u. a. (Hrsg.): Bibliothek Forschung und Praxis. Band 39, Heft 1, 2015, S. 13-31, online verfügbar unter DOI: 10.1515/bfp-2015-0006.
- DCC: The Digital Curation Centre. URL: <http://www.dcc.ac.uk/>. [letzter Zugriff: 14.08.2016].
- DCC (Hrsg.): Checklist for a Data Management Plan. URL: <http://www.dcc.ac.uk/resources/data-management-plans/checklist>. [letzter Zugriff: 14.08.2016].
- Deutsche Forschungsgemeinschaft e. V. URL: <http://www.dfg.de/>. [letzter Zugriff: 14.08.2016].
- Deutsche Forschungsgemeinschaft e. V. (Hrsg.): Umgang mit Forschungsdaten. URL: http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/. [letzter Zugriff: 14.08.2016].
- Deutsche Gesellschaft für Information und Wissen e. V. (Hrsg.): DGI-Konferenz 2016: Erfahrung reloaded – Vom Mundaneum zum Web of Everything. URL: <http://dgi-info.de/events/dgi-konferenz-erfahrung-reloaded-vom-mundaneum-zum-web-of-everything/>. [letzter Zugriff: 14.08.2016].
- Deutscher Django-Verein e. V. (Hrsg.): Über Django. URL: <http://www.django-de.org/ueber-django/>. [letzter Zugriff: 14.08.2016].
- Deutsches Archäologisches Institut (Hrsg.): Datenlebenszyklus. URL: <http://www.ianus-fdz.de/it-empfehlungen/sites/default/files/medialibrary/Datenlebenszyklus.png>. [letzter Zugriff: 14.08.2016].
- Deutsches Archäologisches Institut (Hrsg.): Der Lebenszyklus von Forschungsdaten. URL: <http://www.ianus-fdz.de/it-empfehlungen/lebenszyklus>. [letzter Zugriff: 14.08.2016].
- Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) (Hrsg.): Kerndatensatz Forschung. URL: <http://www.kerndatensatz-forschung.de/>. [letzter Zugriff: 14.08.2016].
- Digital Curation Centre: DMPonline. URL: <https://dmponline.dcc.ac.uk/>. [letzter Zugriff: 14.08.2016].
- Django Girls (Hrsg.): What is Django? URL: <http://tutorial.djangogirls.org/en/django/#what-is-django>. [letzter Zugriff: 14.08.2016].

Django Software Foundation: Django. URL: <https://www.djangoproject.com/>. [letzter Zugriff: 14.08.2016].

Django Software Foundation (Hrsg.): Design philosophies. URL: <https://docs.djangoproject.com/en/1.10/misc/design-philosophies/#don-t-repeat-yourself-dry>. [letzter Zugriff: 14.08.2016].

Django Software Foundation (Hrsg.): FAQ. URL: <https://docs.djangoproject.com/en/1.10/faq/general/#django-appears-to-be-a-mvc-framework-but-you-call-the-controller-the-view-and-the-view-the-template-how-come-you-don-t-use-the-standard-names>. [letzter Zugriff: 14.08.2016].

Django Software Foundation (Hrsg.): QuerySet API reference. URL: <https://docs.djangoproject.com/ja/1.9/ref/models/querysets/>. [letzter Zugriff: 14.08.2016].

Dublin Core Metadata Initiative. URL: <http://dublincore.org/>. [letzter Zugriff: 14.08.2016].

Dublin Core Metadata Initiative (Hrsg.): Expressing Dublin Core metadata using the Resource Description Framework. URL: <http://dublincore.org/documents/dc-rdf/>. [letzter Zugriff: 14.08.2016].

Dublin Core Metadata Initiative (Hrsg.): Guidelines for implementing Dublin Core in XML. URL: <http://dublincore.org/documents/dc-xml-guidelines/>. [letzter Zugriff: 14.08.2016].

Dublin Core Metadata Initiative (Hrsg.): Metadata Terms. URL: <http://dublincore.org/documents/dcmi-terms/>. [letzter Zugriff: 14.08.2016].

ERCIM: CERIF. URL: <http://ercim-news.ercim.eu/en68/european-scene-qsupport-of-the-research-processq/cerif-the-common-european-research-information-format>. [letzter Zugriff: 14.08.2016].

euroCRIS (Hrsg.): Main features of CERIF. URL: <http://www.eurocris.org/cerif/main-features-cerif>. [letzter Zugriff: 14.08.2016].

Europäische Kommission (Hrsg.): Rahmenprogramm für Wettbewerbsfähigkeit und Innovation (CIP). URL: http://ec.europa.eu/cip/index_de.htm. [letzter Zugriff: 14.08.2016].

European Commission (Hrsg.): Europa 2020. URL: http://ec.europa.eu/europe2020/index_de.htm. [letzter Zugriff: 14.08.2016].

European Commission (Hrsg.): Guidelines on Data Management in Horizon 2020. URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-datamgt_en.pdf. [letzter Zugriff: 14.08.2016].

European Commission (Hrsg.): H2020 Programme - Guidelines on FAIR Data Management in Horizon 2020. URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. [letzter Zugriff: 14.08.2016].

European Institute of Innovation & Technology: EIT. URL: <https://eit.europa.eu/>. [letzter Zugriff: 14.08.2016].

Friend of a Friend: URL: <http://www.foaf-project.org/>. [letzter Zugriff: 14.08.2016].

Friend of a Friend (Hrsg.): Vocabulary Specification. URL: <http://xmlns.com/foaf/spec/>. [letzter Zugriff: 14.08.2016].

Georg-August-Universität Göttingen (Hrsg.): WissGrid – Grid für die Wissenschaft. URL: <http://www.wissgrid.de/>. [letzter Zugriff: 14.08.2016].

GitHub: RDMO. URL: <https://github.com/rdmorganiser/>. [letzter Zugriff: 14.08.2016].

Google Inc.: AngularJS. URL: <https://docs.angularjs.org/misc/faq>. [letzter Zugriff: 14.08.2016].

Heger, Martin: Datenmodellierung für Forschungsdatenmanagementpläne. Präsentation im Rahmen der DGI-Konferenz 2016 in Frankfurt am Main. URL: http://dgi-info.de/wp-content/uploads/2015/11/DGI-Pr%C3%A4sentation_Martin-Heger.pdf. [letzter Zugriff: 14.08.2016].

Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V. (Hrsg.): Schwerpunktinitiative „Digitale Information“ der Allianz der Deutschen Wissenschaftsorganisationen. URL: <http://www.allianzinitiative.de/>. [letzter Zugriff: 14.08.2016].

Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V. (Hrsg.): Grundsätze zum Umgang mit Forschungsdaten. URL: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze.html>. [letzter Zugriff: 14.08.2016].

Herbold, Astrid: Europas digitales Gedächtnis ist löchrig. In: Der Tagesspiegel, online verfügbar unter URL: <http://www.tagesspiegel.de/wissen/forschungsdaten-von-akademien-europas-digitales-gedaechtnisist-loechrig/12906856.html>. [letzter Zugriff: 14.08.2016].

KIM Kompetenzzentrum Interoperable Metadaten. URL: <http://www.kim-forum.org>. [letzter Zugriff: 14.08.2016].

KIM Kompetenzzentrum Interoperable Metadaten (Hrsg.): Kleines Handbuch Metadaten - Metadatenprofile. URL: http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadatenprofile.pdf?_blob=publicationFile. [letzter Zugriff: 14.08.2016].

Leibniz-Institut für Astrophysik Potsdam (Hrsg.): RDMO - Research Data Management Organiser. URL: <http://rdmorganiser.github.io/>. [letzter Zugriff: 14.08.2016].

Ludwig, Jens; Enke, Harry (Hrsg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt, 2013. URL: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf. [letzter Zugriff: 14.08.2016].

Megginson, David: SAX. URL: <http://www.saxproject.org/>. [letzter Zugriff: 14.08.2016].

NSF: National Science Foundation. URL: <https://www.nsf.gov/>. [letzter Zugriff: 14.08.2016].

Österreichische Forschungsförderungsgesellschaft mbH (Hrsg.): Open Data in Horizon 2020. URL: https://www.ffg.at/europa/recht-finanzen/h2020-open_data. [letzter Zugriff: 14.08.2016].

OpenAIRE: What is the Open Research Data Pilot? URL: <https://www.openaire.eu/opendatapi-lot>. [letzter Zugriff: 14.08.2016].

Oracle: Simple API for XML. URL: <https://docs.oracle.com/javase/tutorial/jaxp/sax/index.html>. [letzter Zugriff: 14.08.2016].

Orientation in Objects GmbH: REST Web Services. URL: <http://www.oio.de/public/xml/rest-webservices.htm>. [letzter Zugriff: 14.08.2016].

Padilla, José: REST Framework XML. XML support for Django REST Framework. URL: <https://github.com/jpadilla/django-rest-framework-xml>. [letzter Zugriff: 14.08.2016].

Princeton University Library (Hrsg.): NSF Data Management Plan. URL: http://libguides.princeton.edu/ld.php?content_id=2940897. [letzter Zugriff: 14.08.2016].

Python Software Foundation: Python. URL: <https://www.python.org/>. [letzter Zugriff: 14.08.2016].

Python Software Foundation (Hrsg.): collections - Container datatypes. URL: <https://docs.python.org/3/library/collections.html#collections.OrderedDict>. [letzter Zugriff: 14.08.2016].

re3data.org Project Consortium: Registry of Research Data Repositories. URL: <http://www.re3data.org/>. [letzter Zugriff: 14.08.2016].

The SQLite Consortium: SQLite. URL: <https://www.sqlite.org/>. [letzter Zugriff: 14.08.2016].

TechTarget: Representational State Transfer (REST). URL: <http://www.searchenterprisesoftware.de/definition/Representational-State-Transfer-REST>. [letzter Zugriff: 14.08.2016].

Universität des Saarlandes (Hrsg.): TwinLife. URL: <http://www.twin-life.de/>. [letzter Zugriff: 14.08.2016].

Wissenschaftsrat (Hrsg.): Kerndatensatz Forschung. URL: http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/kerndatensatz_forschung.html. [letzter Zugriff: 14.08.2016].

Wissenschaftsrat (Hrsg.): Empfehlungen zur Spezifikation des KDSF. URL: <http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>.

WissGrid – Grid für die Wissenschaft (Hrsg.): Checkliste zum Forschungsdaten-Management. URL: <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Checkliste-Forschungsdaten-Management.pdf>. [letzter Zugriff: 14.08.2016].

Abbildungsverzeichnis

Abbildung 1: Lebenszyklus von Forschungsdaten.....	4
Abbildung 2: Europa 2020 Strategie.....	8
Abbildung 3: RDMO-Datenmodell.....	33
Abbildung 4: Ausschnitt des RDMO-Fragenkatalogs für Datennutzung/Kosten	36
Abbildung 5: Ablaufschema Datenumwandlung.....	59
Abbildung 6: XML-Ausgabe.....	65

Danksagung

Danken möchte ich zunächst Prof. Dr. Heike Neuroth für die kompetente und unkomplizierte Betreuung und die dadurch erhaltenen wertvollen Hinweise zur Erstellung dieser Masterarbeit.

Insbesondere danken möchte ich vor allem Dr. Jochen Klar und Dr. Harry Enke für die umfangreiche Betreuung vor Ort im Leibniz-Institut für Astrophysik in Potsdam. Ich konnte mich bei allen Problemen an sie wenden und sie haben für meine Fragen immer Zeit aufbringen können.

Ein großer Dank geht auch an die weiteren Projektmitglieder Jens Ludwig und Claudia Engelhardt für die freundschaftliche Einbindung in das Team.

Danke auch an die Personen, die mich beim Korrekturlesen und dem Layout unterstützt haben.

Anlagen

Anlage 1: Hausarbeit.....	1
Anlage 2: RDMO-Fragenkatalog.....	30
Anlage 3: RDMO-Attribute/-Entitäten.....	44
Anlage 4: XML-Export-Programmcode.....	51

Anlage 1: Hausarbeit

Marcus Heinrich, Martin Heger

Datenmanagementpläne

Eine Bestandsübersicht

Projektarbeit im Rahmen des Moduls M4.2 / WS15 im Studiengang M.A. Informationswissenschaften der FH Potsdam:

Konzeptionelle Entwicklung eines Werkzeugs
für die Planung des Forschungsdatenmanagements

Prof. Dr. Heike Neuroth

in Kooperation mit dem DFG-Projekt



<https://dmpwerkzeug.github.io/>

08.02.2016

Inhaltsverzeichnis

1. Abbildungsverzeichnis
2. Einleitung
3. Begriffsklärung
 - 3.1 Forschungsdaten und Forschungsdatenmanagement
 - 3.2 Datenmanagementplan
4. Marktsichtung
 - 4.1 Auswahl der Datenmanagementpläne
 - 4.2 Bewertungskriterien
 - 4.2.1 Allgemein
 - 4.2.2 Technische Infrastruktur
 - 4.2.3 Funktionalitäten und Inhalt
 - 4.3 Bewertung ausgewählter DMP-Tools
5. Zusammenfassung
6. Empfehlungen
7. Literaturverzeichnis
8. Anhang

2. Einleitung

Die Forschung ist in Deutschland traditionell stark verankert. Eine Recherche in dem Förderkatalog der Bundesregierung zeigt, dass es eine Vielzahl an laufenden Projekten gibt und verstärkt Gelder in die Forschungsförderung investiert werden. Doch es mangelt an einer ausreichenden digitalen Vernetzung, die es ermöglicht die Forschungsdaten und die Ergebnisse auszutauschen sowie langfristig zu sichern.

Diskussionen und Bestrebungen der letzten Jahre zu diesen Fragestellungen mündeten schlussendlich in der Entwicklung von Datenmanagementplänen (DMP). Inzwischen verlangen Forschungsfördereinrichtungen wie die Deutsche Forschungsgemeinschaft (DFG) den Einsatz von Datenmanagementplänen, wodurch sich eine deutlich steigende Relevanz zeigt.

Ziel dieser Arbeit ist es, eine Liste der international verfügbaren Tools zur Erstellung von Datenmanagementplänen zu erarbeiten und in der Folge ausgewählte Programme anhand von Bewertungskriterien zu evaluieren. Abschließend werden die besten Funktionen der einzelnen Tools zusammenfassend dargestellt und es wird der Versuch unternommen darüber hinausgehende Empfehlungen für potenzielle Entwicklungen zu formulieren. Die Arbeit erhebt jedoch nicht den Anspruch auf Vollständigkeit, da die Möglichkeit durchaus besteht, dass Tools aufgrund von Zugangs- bzw. Sprachbarrieren nicht gefunden werden konnten.

Die Arbeit gliedert sich in einen theoretischen und praktischen Teil. Zu Beginn wird ein kurzer Einstieg in die Thematik vermittelt und somit Gründe für den Einsatz von Datenmanagementplänen genannt. Der Hauptteil befasst sich mit der Bewertung ausgewählter Tools. Auf Grundlage dieser Bewertungen werden im Anschluss Anregungen für die Entwicklung zukünftiger Tools entwickelt.

3. Begriffsklärung

Es werden die Begriffe Forschungsdaten und Forschungsdatenmanagement erläutert. Der Datenmanagementplan wird im Lebenszyklus von Forschungsdaten verortet. Schließlich folgt eine Beschreibung der Inhalte von Datenmanagementplänen.

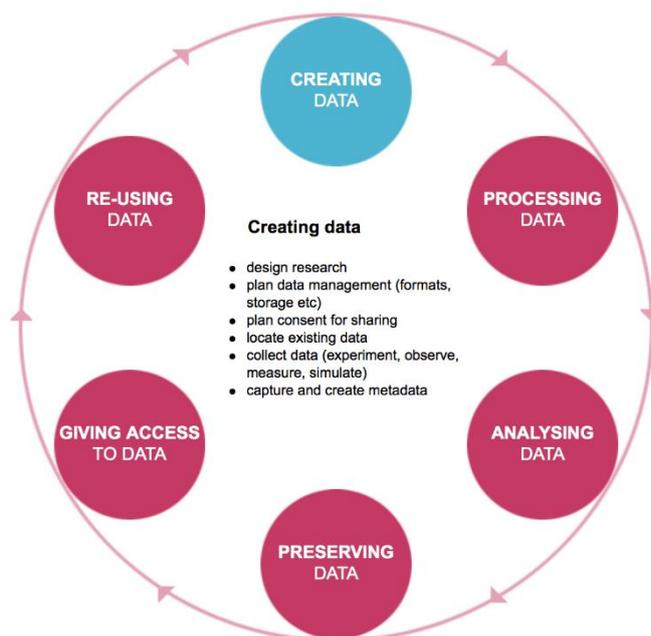
3.1 Forschungsdaten und Forschungsdatenmanagement

Die Deutsche Forschungsgemeinschaft (DFG) definiert Forschungsdaten wie folgt:

„Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“

Das Management dieser heterogenen Forschungsdaten ist dabei stark von der Wissenschaftsdisziplin abhängig und es gibt keine einheitlichen Modelle bzw. Verfahrensweisen. Eine gewisse Grundlage bieten jedoch die Leitlinien zum Umgang mit Forschungsdaten der DFG. Dort heißt es, dass bereits bei der Antragsstellung Angaben zu den möglichen Forschungsdaten und deren Qualitätssicherung gemacht werden sollen. Die Europäische Kommission gibt ebenfalls entsprechende Richtlinien vor.

Demzufolge werden Datenmanagementpläne durch den Antragsteller bereits in der Antragsphase erstellt und in der Folge stets aktualisiert. Bezogen auf den Lebenszyklus von Forschungsdaten befindet sich diese Aktivität am Anfang und somit noch vor der Erstellung der eigentlichen Daten. Trotzdem sind alle Aspekte des gesamten Kreislaufs betroffen, da diese Überlegungen Aspekte der Datenerhebung, -auswertung, -weiterverarbeitung, -speicherung, -zugänglichkeit und -nachnutzbarkeit beschreiben (vgl. Abb. 1).



<http://www.data-archive.ac.uk/create-manage/life-cycle>

3.2 Datenmanagementplan

Datenmanagementpläne dienen zur systematischen Beschreibung von Strategien und Maßnahmen, wie mit Forschungsdaten allgemein oder auch speziell in einem Projekt verfahren wird.

Erfasst werden nicht nur die Art der Speicherung, Verzeichnung, Pflege und Verarbeitung der Daten, sondern sämtliche Prozesse und Technologien, die während des gesamten Lebenszyklus der Forschungsdaten (während der Projektlaufzeit und auch nach Projektabschluss) zum Ein-

satz kommen. Ziel ist es, diese Prozesse und Technologien sichtbar und nachvollziehbar zu machen sowie zusätzlich die technischen, organisatorischen und rechtlichen Rahmenbedingungen zu definieren. Erst durch einen umfassenden Datenmanagementplan werden die Daten für Dritte interpretierbar und nachnutzbar.

Wichtige Fragen, die bei der Erstellung eines DMP beachtet werden sollten, finden sich z. B. im Leitfaden bzw. der Checkliste von WissGrid sowie den aktuellen Checklisten zur DMP-Erstellung der DCC (The Digital Curation Centre) und NSF (National Science Foundation):

- Um was für ein Projekt handelt es sich?
- Welche Daten werden erzeugt und genutzt?
- Um welche Art(en) von Daten handelt es sich?
- Welche Daten sollen oder müssen warum aufbewahrt werden?
- Sind Zusatzinformationen für das Verstehen der Daten notwendig?
- Wann erfolgt die Datenauswahl?
- Wie lange sollen die Daten aufbewahrt werden (Archivierung)?
- Wann werden die Daten übergeben (Datenaustausch, Datenpublikation)?
- Wer darf die Daten nutzen?
- Welche Kosten entstehen?
- Welche Ressourcen werden benötigt?

4. Marktsichtung

Grundlage für die Marktsichtung war eine Internetrecherche nach Tools für das Datenmanagement. Auffällig dabei war, dass viele Einrichtungen und Förderer Checklisten für die Planerstellung anbieten, jedoch keine eigenen Tools. Es sei zudem angemerkt, dass verstärkt Aktivitäten zur Entwicklung von derartigen Tools zu beobachten sind und neue Tools in Planung sind.

Die folgende alphabetische Auflistung zeigt die gefundenen DMP-Tools:

- Clarin-D (DataWizard)
- Data Management Planning Tool (Queensland Universität für Technologie Brisbane, Australien)
- DMP Assistant (Universität Alberta, Kanada)
- DMPonline (DCC)
- DMPonline (Universität Queensland, Australien)
- DMPTool (Universität California)
- DMP Tool (Universität Bielefeld)
- IEDA (Interdisciplinary Earth Data Alliance) DMP-Tool
- LabArchives
- Research Data Footprints (Deakin Universität, Australien)

4.1 Auswahl der Datenmanagementpläne

Bei der Auswahl der Tools spielte die Zugänglichkeit eine entscheidende Rolle. So können einige Tools nur von Vertretern der zugehörigen Institution verwendet werden. Aus diesem Grund wurden folgende DMP-Tools nicht weiter berücksichtigt: DMPonline (Universität Queensland, Australien) und Research Data Footprints (Deakin Universität, Australien).

Die finale Auswahl erfolgte schließlich nach dem Entstehungskontext bzw. der Zielsetzung und den damit verbundenen unterschiedlichen Schwerpunkten der Datenmanagementpläne. Auf diese Weise sollte ein möglichst breites Spektrum abgedeckt werden. Folgende DMP-Tools wurden für die weitere Bewertung ausgewählt:

- DMP Tool (Universität Bielefeld)
- DMPonline (DCC)
- Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool
- LabArchives

Der ausgewählte Datenmanagementplan der Universität Bielefeld ist im institutionellen Kontext entstanden. Im Gegensatz dazu hat das Tool des Digital Curation Centre einen nationalen Hintergrund und muss somit eine wesentlich größere Zielgruppe abdecken. Waren die erstgenannten Tools domänenübergreifend, so soll mit dem DMP-Tool der IEDA zusätzlich ein Tool mit fachspezifischem Fokus untersucht werden. Bei dem zuletzt genannten LabArchives handelt es sich um ein virtuelles Laborbuch und nicht um ein DMP-Tool im eigentlichen Sinne. Dennoch soll überprüft werden, ob dieser Dienst einen Mehrwert für die potenzielle Entwicklung neuer DMP-Tools bietet bzw. generell für die Verwaltung von Daten und Dateien sowie zum kollaborativen Arbeiten geeignet sein könnte.

4.2 Bewertungskriterien

4.2.1 Allgemein

- Entstehung/Herkunft
- Domäne
- Zielgruppe/Zielsetzung
- Größe des Anwenderkreises
- Policies
- Sicherheit und Authentifizierung

Die Motivation zur Erstellung von Datenmanagementplänen kann je nach Auftraggeber variieren. Typisch ist dabei die Entwicklung durch einen einzelnen Fachbereich an einer Hochschule oder als Gesamtlösung für alle Fachbereiche. Ebenfalls ist es möglich, dass nationale Verbände Tools entwickeln und diese national bzw. international anbieten. Darüber hinaus kann ein Tool mit kommerziellem Hintergrund durch Firmen entwickelt werden. Welche Domäne und Zielgruppe für das DMP-Tool in Frage kommen, ist im engen Zusammenhang mit dem Entstehungskontext zu sehen. Folglich kann der Schwerpunkt der Tools fachspezifisch und/oder interdisziplinärer sein.

Die Größe des Anwenderkreises und der Community gibt einen ersten Hinweis auf das Entwicklungspotenzial sowie den Support. Weiterhin wird untersucht, ob die Anbieter der Tools bekannte Policies zum Thema Forschungsdaten anbieten und inwiefern der Einsatz von DMP-Tools dort empfohlen bzw. vorgeschrieben wird. Der Zugang zu den Tools gestaltet sich nach der

Einrichtung eines Benutzerkontos in den meisten Fällen problemlos. In einigen Fällen wird jedoch eine Freischaltung durch den Administrator nötig oder es gibt keine Zugriffsmöglichkeit. Bei kommerziellen Angeboten ist unbedingt auf die entstehenden Kosten zu achten.

4.2.2 Technische Infrastruktur

- zentral oder lokal betrieben
- Installation
- Aktualisierung/Update
- Gewährleistung (Datensicherheit)

Die Begutachtung der technischen Kriterien umfasst den laufenden Betrieb der DMP-Tools. Es wird geprüft, ob der Dienst zentral oder lokal betrieben wird und ob das Tool auch offline genutzt werden kann. Des Weiteren werden Angaben zur Aktualisierungsstrategie bei neueren Softwareversionen und zum Datenschutz überprüft.

4.2.3 Funktionalitäten und Inhalt

- Usability
- Sprachen
- Hilfsfunktionen (E-Tutorials, FAQs, Hilfestellung direkt bei DMP-Erstellung)
- Kollaboration
- Exportfunktionen
- Community/Erweiterbarkeit
- Versionsmanagement
- Verwaltung mehrerer DMP
- Templates von Fördereinrichtungen vorhanden/Bindung an Policy des Förderers
- Kosten/Ressourcen
- Metadatenstandards

Bei den Funktionalitäten und dem Inhalt wird zuerst die Usability untersucht, d. h. wie intuitiv und flexibel das Tool zu bedienen ist. Auch wenn dieser Aspekt durchaus subjektiv ist, so trägt er dennoch mit einer Mischung aus Nice-to-have- und Must-have-Features zur möglichst ungehinderten und zügigen Erstellung eines DMP bei. Ebenso wird hierbei auf die Anpassbarkeit an das eigene institutionelle Umfeld und disziplinspezifische Aspekte eingegangen. Da sich DMP auch nach dem Projektstart ändern können, ist die Möglichkeit zu einer schnellen Anpassung des DMP besonders hervorzuheben (siehe Forschungsdatenkreislauf). Zusätzlich werden Verknüpfungsmöglichkeiten wie ResearcherID und ORCID aufgelistet.

Die zur Verfügung stehenden Sprachen des Tools werden erwähnt und etwaige Hilfefunktionen erörtert. Außerdem werden die Möglichkeiten zum kollaborativen Arbeiten und die vorhandenen Exportfunktionen näher beleuchtet. Gerade eine umfangreiche Share-Funktion ist wichtig, wenn Forscher international aktiv sind und über verschiedene Standorte verteilt vernetzt miteinander arbeiten möchten.

Ein Blick auf die Community und deren Engagement bezüglich Erweiterungen der vorhandenen Funktionalitäten (z.B. Aktivitäten bei GitHub) kann wertvolle Hinweise dazu liefern, ob das Tool Anerkennung und Einfluss in der Forscher-Community erhalten hat und ob spezielle Versionen bzw. Feature-Ergänzungen des Tools für bestimmte Fachdisziplinen zu erwarten sind. Aufgrund der Dynamik bei der Planerstellung ändert sich der Bearbeitungsstand stetig ("Living DMP") und daher ist eine Versionierung essenziell.

Die komfortable und übersichtliche Verwaltung mehrerer DMP ist für ein reibungsloses Arbeiten besonders wichtig. Genauso können bereits vorhandene Templates/Policies die Erstellung eines DMP beschleunigen und deutlich effizienter gestalten.

Zuletzt wird darauf eingegangen, ob eine Abfrage der Kosten und Ressourcen, die für ein Projekt entstehen bzw. benötigt werden, erfolgt und inwiefern auf die Einhaltung von Metadatenstandards geachtet wird und ob diese empfohlen werden.

4.3 Bewertung ausgewählter DMP-Tools

Folgende DMP-Tools werden anhand der o.g. genannten Kriterien bewertet: DMP Tool (Universität Bielefeld), DMPonline (DCC), Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool und LabArchives.

DMP Tool (Universität Bielefeld)

Kriterien/Tools	DMP Tool (Universität Bielefeld)
Allgemein	
Entstehung/Herkunft	Entstehung im institutionellen Kontext
Domäne	Berücksichtigung der Profilschwerpunkte der Forschungseinrichtungen sowie interdisziplinäre Forschungsvernetzung
Zielgruppe/Zielsetzung	alle potenziellen Antragssteller für Forschungsvorhaben
Größe des Anwenderkreises	gesamte Universität Bielefeld
Policies	Einrichtung eines DMP wird in der vorhandenen Policy explizit genannt
Sicherheit und Authentifizierung	Zugang ist frei, aber Freischaltung durch Administrator nötig
Technische Infrastruktur	
zentral oder lokal betrieben	zentral betrieben durch die Universität Bielefeld
Installation	entfällt, da zentral betrieben
Aktualisierung Update	nur Datum, keine Versionierung
Gewährleistung/ Datensicherheit	gemäß den Grundsätzen zum Umgang mit Forschungsdaten ist der Schutz von sensiblen Daten verpflichtend
Funktionalitäten und Inhalt	

Usability	DMP wird kapitelweise ausgefüllt, eine Grafik zeigt das aktuelle Kapitel bzw. den Bearbeitungsstatus an. Kapitel können nur in vorgegebener Reihenfolge bearbeitet werden.
Sprachen	deutsch, englisch
Hilfefunktionen	umfangreiche Hinweise direkt im DMP unterstützen bei der Erstellung
Kollaboration	keine Kollaboration möglich
Exportfunktionen	DMP kann im PDF-Format heruntergeladen werden
Community/Erweiterbarkeit	keine Erweiterung möglich; Änderungen werden zentral durch die Universität Bielefeld durchgeführt; Quelltext nicht frei verfügbar bzw. es existiert keine Entwickler-Community
Versionsmanagement	kein Versionsmanagement vorhanden, lediglich Datum der Erstellung sichtbar
Verwaltung mehrerer DMP	einfache Auflistung der erstellten DMP und alphabetische Sortierung möglich
Templates von Fördereinrichtungen vorhanden/Bindung an Policy des Förderers	Vorlage der Universität Bielefeld mit einem hohen Detaillierungsgrad für datenintensive Forschungsvorhaben, Cluster of Excellence Cognitive Interaction Technology (CITEC), Horizon 2020
Kosten/Ressourcen	Personalkosten und -aufwand, Kosten vor und nach der Projektlaufzeit sowie das Gesamtbudget für das Datenmanagement werden erfasst
Metadatenstandards	sinnvolle (automatische) Integration von Metadaten wird nahegelegt; Voraussetzungen für Hard- u. Software sowie

	benötigte Fachkenntnisse zum Umgang mit Metadaten werden erfragt
--	--

DMPonline (DCC)

Kriterien/Tools	DMPonline (DCC)
Allgemein	
Entstehung/Herkunft	Entwicklung durch Digital Curation Centre (UK). DCC hat eine Vielzahl an strategischen Partnerschaften (darunter regionale Universitäten sowie überregionale staatlich geförderte Organisation weltweit (Bsp: Australian National Data Service))
Domäne	aufgrund der Entstehung domänenübergreifend verwendbar
Zielgruppe/Zielsetzung	berücksichtigt werden alle potenziellen Antragsteller mit Schwerpunkt in UK; aber auch Nutzer außerhalb UK
Größe des Anwenderkreises	sehr großer Anwenderkreis in UK, aber auch international Es gibt internationale Anwender, die DMPonline als Grundlage für eigene Entwicklungen nutzen. Zudem sehr aktive Community auf GitHub.
Policies	aufgrund der Herkunft keine Policies; Es sind jedoch Policies der Förderer zentral gesammelt und es gibt Anleitungen zur Erstellung von institutionellen Policies
Sicherheit und Authentifizierung	Zugang ist frei. Nutzer aus UK können sich mit ihren Instituts-Login anmelden (Weiterleitung auf Anmeldeseite des Instituts). Andere Nutzer können Accounts selbstständig anlegen.
Technische Infrastruktur	

zentral oder lokal betrieben	zentral betrieben durch die DCC an der Universität Edinburgh. Es gibt jedoch eigene Entwicklungen die folglich lokal betrieben werden.
Installation	entfällt, da zentral betrieben
Aktualisierung/Update	Datum und Versionsnummer (aktuell Version 4.1)
Gewährleistung/ Datensicherheit	Die E-Mail wird gespeichert und ggf. unter DCC-Partner ausgetauscht. Der Schutz der persönlichen Daten wird jedoch zugesichert. Administratoren der Universität Edinburgh dürfen Zugang erhalten, sofern es der Wartung dient. Die Eigentumsrechte der eingegebenen Daten verbleiben bei dem Urheber.
Funktionalitäten und Inhalt	
Usability	<p>Planerstellung unterscheidet sich je nach Förderer: in drei Stufen bei Horizon 2020 (Initial DMP, Mid-Term Review DMP und Final Review DMP), einfache Abfrage der Fragen bei dem Großteil der Förderer.</p> <p>Eine Grafik zeigt den Status an.</p> <p>Beiträge bzw. Änderungen werden durch Zeit und Bearbeiter gekennzeichnet.</p> <p>ResearcherID kann angegeben werden (wichtig für mögliche Verknüpfungen)</p>
Sprachen	englisch
Hilfefunktionen	umfangreiche Hinweise direkt im DMP unterstützen bei der Erstellung. Weiterhin gibt es bei der Wahl einiger Einrichtungen aus UK institutionelle Hilfen. Für Einsteiger gibt es ein E-Tutorial, das alle Basisfunktionen erklärt.
Kollaboration	stark ausgeprägte Kollaborationsmöglichkeiten. Zu jeder Frage können Notizen erstellt werden. Zudem können weitere Nutzer

	zum DMP hinzugefügt werden und gemeinsam arbeiten (bzw. nur lesen -> Rechtemanagement).
Exportfunktionen	DMP kann in allen Phasen exportiert werden. Diverse Formate vorhanden (csv, html, json, pdf, xml, text, docx). Schriftart, Dateiname und Inhalte können vor dem Export festgelegt werden.
Community/Erweiterbarkeit	Quelltext kann frei heruntergeladen (GNU Affero General Public License) und installiert werden. Es gibt eine Entwickler-Community und regelmäßige Updates. Die Diskussionen können auf Github nachvollzogen werden. Auf der Webseite von DMPonline gibt es ebenso entsprechende Informationen. Weiterhin gibt es eine E-Mail-Liste für Entwickler (DMPONLINE-USER-GROUP@JISC-MAIL.AC.UK)
Versionsmanagement	Datum und Bearbeitet der letzten Änderung kann eingesehen werden.
Verwaltung mehrerer DMP	Freie Suche nach mehreren DMP möglich. Viele Filtermöglichkeiten sowie Sortierung beliebigen Kriterien möglich.
Templates von Förderinstitutionen vorhanden/Bindung an Policy des Förderers	Vorlagen schwerpunktmäßig aus UK, aber auch Horizon 2020, National Science Foundation (USA) und ZonMw (Niederlande)
Kosten/Ressourcen	Angaben, ob Fachexperten bzw. Training und spezielle Hardware notwendig sind. Personalkosten und -aufwand werden nicht abgefragt.
Metadatenstandards	sinnvolle (automatische) Integration von Metadaten wird nahegelegt; Verantwortliche Personen für Metadaten müssen benannt werden.

Interdisciplinary Earth Data Alliance (IEDA) DMP Tool

Kriterien/Tools	Interdisciplinary Earth Data Alliance (IEDA) DMP Tool
Allgemein	
Entstehung/Herkunft	2011 offiziell gestartet, um auf einfache Art und Weise DMP für die Einbindung in NSF-Vorschläge (National Science Foundation) zu erstellen, aus Kollaboration zwischen EarthChem und Marine Geoscience Data System (MGDS) entstanden
Domäne	Fokus liegt auf maritimen, geologischen und polaren Daten, ist aber dennoch so grundlegend aufgebaut, dass auch Anträge von anderen NSF-Abteilungen damit erstellt und bearbeitet werden können
Zielgruppe/Zielsetzung	Primäre Zielgruppe sind alle wissenschaftlichen NSF-Divisionen, bei denen die o.g. Daten anfallen (EarthChem und Marine Geoscience Data System) sowie ähnliche NSF-Abteilungen
Größe des Anwenderkreises	Hauptnutzerkreis in den USA, aber auch Mitglied bei ICSU World Data System (WDS) und Federation of Earth Science Information Partners (ESIP), weiterer Outreach durch Einbindung der Community, Konferenzen, Meetings, Webinars, Workshops und Tool-Schulungen, Mail-Support
Policies	Data Policy je nach NSF-Division unterschiedlich, wird beim Auswählen automatisch angezeigt
Sicherheit und Authentifizierung	Zugang ist frei. Kostenlose GeoPass ID wird benötigt
Technische Infrastruktur	
zentral oder lokal betrieben	zentral gehostet am Lamont-Doherty Earth Observatory der Columbia University

Installation	entfällt, da zentral betrieben
Aktualisierung/Update	Datum und Versionsnummer (aktuell Version 2 vom 17.04.2012)
Gewährleistung/ Datensicherheit	Anmeldung erfolgt mit GeoPass ID und Passwort, Daten werden mit einem DOI unter der Creative Commons License (CC BY-NC-SA 3.0) veröffentlicht
Funktionalitäten und Inhalt	
Usability	einfache Formatvorlage, Fokus liegt auf der Art der vorliegenden bzw. entstehenden Daten (Type of Data), die detailliert angegeben werden können
Sprachen	englisch
Hilfefunktionen	Einbindung der Community, Konferenzen, Meetings, Webinars, Workshops, Tool-Schulungen, Mail-Support
Kollaboration	Co-PI(s) können eingetragen werden, keine gemeinsame Bearbeitung bzw. Notizen o.ä. möglich
Exportfunktionen	Art der Einreichung sowie ein automatisches Einreichungsdatum bei der NSF mit Deadline können festgelegt werden, Export als PDF-Datei mit vorheriger Syntax-Prüfung möglich
Community/Erweiterbarkeit	Daten werden nach einer Prüfung durch die IEDA unter der Creative Commons Lizenz veröffentlicht und mit einem DOI verknüpft und nach einer durch den Autoren festgelegten Frist der Öffentlichkeit zur Verfügung gestellt. Nach der Veröffentlichung können keine weiteren Änderungen vorgenommen werden und bei Änderungsbedarf muss eine neue Version angelegt, registriert, geprüft und veröffentlicht werde. Sollten veröffentlichte Daten zurückgezogen werden, verbleiben die Katalogmetadaten im System (als "zurückgezogen" markiert)

Versionsmanagement	Datum der Erstellung und Einreichung werden angegeben
Verwaltung mehrerer DMP	Verwaltung mehrerer DMP möglich
Templates von Förder-einrichtungen vorhanden/Bindung an Policy des Förderers	Verschiedene Templates je nach Art der Daten und Fördereinrichtung auswählbar, Policies richten sich nach jeweiligen Einrichtungen bzw. NSF-Divisionen und werden nach Auswahl automatisch angezeigt und berücksichtigt
Kosten/Ressourcen	Kosten werden nicht abgefragt
Metadatenstandards	Mischung aus Pflichtfeldern und optionalen Metadatenfeldern wird benutzt, die die Beziehungen zu anderen Publikations-Datensamm-lungen weiter beschreiben sollen; verwendet wird das DataCite-Metadaten-schema

LabArchives

Kriterien/Tools	LabArchives
Allgemein	
Entstehung/Herkunft	2009 gegründet und als einfach zu benutzende und kostengünstige Lösung (Software-as-a-Service) zur Labororganisation und kollabo-rativem Arbeiten entwickelt. Soll traditionelle Notizbücher aus Pa-pier durch elektronische ersetzen
Domäne	Fokus liegt auf der Datenorganisation von Laboren, kann allerdings auch allgemein zum Speichern, Organisieren, Teilen und Publizie-ren von Daten genutzt werden
Zielgruppe/Zielsetzung	ursprünglich als elektronisches Notizbuch für Wissenschaftler von Forschungs- und Institutslaboren entwickelt, kann jedoch auch in einem allgemeineren Datenkontext genutzt werden

Größe des Anwenderkreises	vorwiegend große Institute und Universitäten, darunter das California Institute of Technology, Cornell University, Tufts University, UT Southwestern, University of Sidney, Monash University, The Garvan Medical Research Institute, University of Wisconsin, Yale University u.a. sowie Kooperation mit dem Internet2-Projekt
Policies	Terms of Service seitens LabArchives
Sicherheit und Authentifizierung	kostenlose Version nach Registrierung (Anmeldung erfolgt mit E-Mail-Adresse und Passwort) verfügbar, außerdem eine Classroom Edition (hauptsächlich an Studenten/individuelle Benutzer gerichtet), Professional Edition (für PIs, Laborleiter etc.) sowie eine campusweite Enterprise-Lizenz kostenpflichtig verfügbar
Technische Infrastruktur	
zentral oder lokal betrieben	zentral betrieben als Software-as-a-Service, cloud-basiertes Electronic Lab Notebook (ELN)
Installation	keine Installation nötig
Aktualisierung/Update	nicht ersichtlich/bekannt
Gewährleistung/ Datensicherheit	LabArchives übernimmt keinerlei Haftung für die Datensicherheit
Funktionalitäten und Inhalt	
Usability	übersichtlich und einfach zu benutzen, Benutzung erinnert an eine Mischung aus Windows Explorer und CMS-Systemen wie WordPress
Sprachen	englisch

Hilfefunktionen	extrem umfangreiche Hilfefunktionen: Quick Start Guides, Power-Point- u. Video-Tutorials, Support-Anfragen per Mail und Ticket direkt im Tool möglich, Kundenhotline, Anmeldung zu verschiedenen Webinars über Website
Kollaboration	Einladefunktion von Personen per Mail, individuelle Vergabe von Rechten (Schreiben, Editieren, Sehen, kein Zugriff etc.) bis hinunter auf Datei-Ebene. Jeder Mitarbeiter sieht nur genau das, was er sehen darf. Activity Feed, welches aktuelle Änderungen anzeigt
Exportfunktionen	automatischer Export per Mail möglich, an bestimmte Gruppenmitglieder, Benutzergruppen, Export u.a. als URL, komfortable DOI-Vergabe, Share-Funktion
Community/Erweiterbarkeit	Community-Blog vorhanden, ansonsten eher über Internet2
Versionsmanagement	Versionierung mit Datum, Uhrzeit (+Zeitzone), Name, IP-Adresse sowie Aktualisierung dieser Daten bei Änderungen, Versionsgeschichte ähnlich wie bei Wikipedia einsehbar und beliebig wiederherstellbar
Verwaltung mehrerer DMP	komfortabel möglich, erstellte DMP nach Name, der eigenen Nutzerrolle, Aktivitäten, letzten Aktivitäten, Navigation und Verfügbarkeit (offen, geschlossen) möglich, DMP können gelöscht, geklont (inkl. Benutzerrechten und Inhalten), kopiert und sogar vollständig oder partiell zusammengeführt werden
Templates von Förderer-einrichtungen vorhanden/Bindung an Policy des Förderers	nur eigene Templates bzw. exportrelevante Templates (z.B. für Google Docs/Calendar) verfügbar
Kosten/Ressourcen	Aufgrund der Herkunft bzw. der Zielsetzung existiert standardmäßig keine Kostenabfrage

Metadatenstandards	Erstellung von verschiedenen Textformaten (PDF, Word, Plain/Rich Text etc.) möglich. Aufgrund der Herkunft bzw. der Zielsetzung der Software erfolgt keine Abfrage von Metadatenstandards
--------------------	---

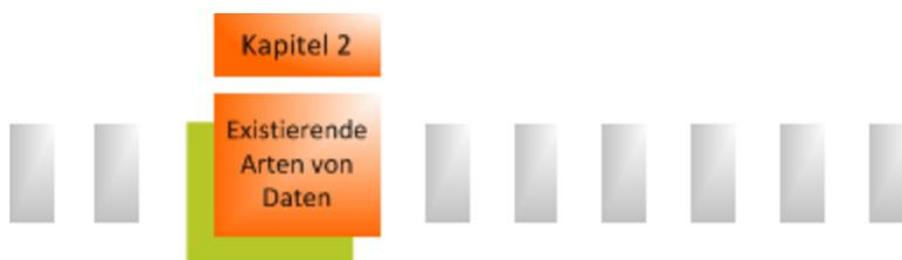
5. Zusammenfassung

Die vier begutachteten Datenmanagementplan-Tools haben im Kern die gleiche Intention, unterscheiden sich jedoch in ihrer Schwerpunktsetzung und haben folglich verschiedene Stärken und Schwächen.

Basierend auf den vorangegangenen Analysen sollen hier beispielhaft einige der besonders großen Pluspunkte der jeweiligen DMP-Tools nochmals ausdrücklich hervorgehoben werden, die für besonders nützlich und sinnvoll erachtet werden.

DMP Tool (Universität Bielefeld)

Beim Ausfüllen der Felder informiert eine jederzeit sichtbare Statusleiste über den gegenwärtigen Fortschritt, indem sie das stets aktuell bearbeitete Kapitel anzeigt (vgl. Abb. 2). Somit fällt die Orientierung und die Einschätzung des Gesamtfortschritts deutlich einfacher. Weiterhin ist eine Sortierung der DMP nach dem verwendeten Template möglich. BESCHRIFTUNG (Statusleiste DMP Tool der Universität Bielefeld (Screenshot))

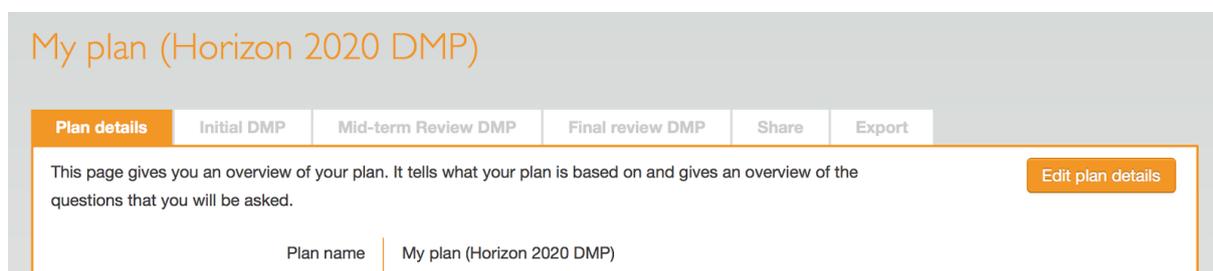


Auffällig ist, dass die Frage nach den Kostenaspekten in Bielefeld deutlich ausgeprägter ist (Personalaufwand muss angegeben werden bzw. Angaben zu Kosten vor und nach der Projektlaufzeit), als in den anderen untersuchten Tools.

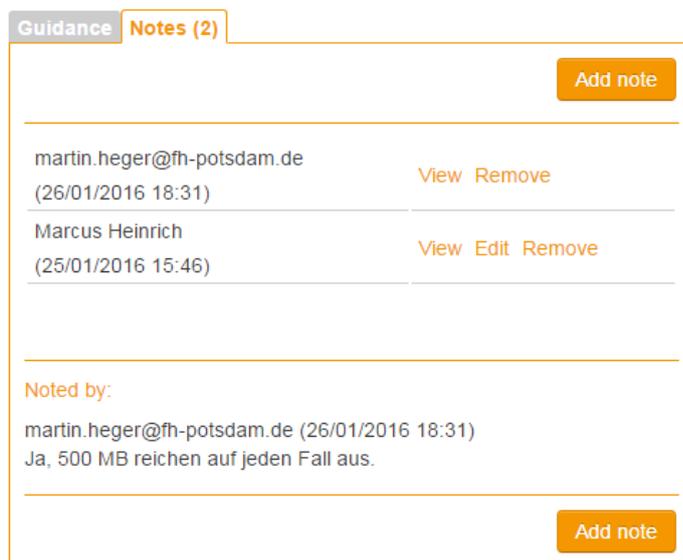
Eine Frage, die sich auch nur bei der Universität Bielefeld wiederfand, war, ob auch alle Beteiligten hinter den Plänen zum Datenmanagement stehen. Falls nein, wäre eine Begründung erforderlich.

DMPonline (DCC)

Es fallen zunächst die umfangreichen Hilfefunktionen auf. So gibt es ein E-Tutorial für Einsteiger und prägnante Hilfetexte zu jeder Frage bei der Planerstellung. Auch hier informiert eine Statusleiste über den aktuellen Fortschritt. Weiterhin ist es sogar möglich (im Gegensatz zur Universität Bielefeld), in ein beliebiges Kapitel zur Bearbeitung zu gehen und direkt dort weiterzuarbeiten, statt sich erst durch die vorangegangenen Kapitel klicken zu müssen. Hinzu kommen die sehr guten Suchmöglichkeiten. Die zahlreichen Filter- und Sortiermöglichkeiten ermöglichen eine optimale Verwaltung einer großen Anzahl an Datenmanagementplänen.



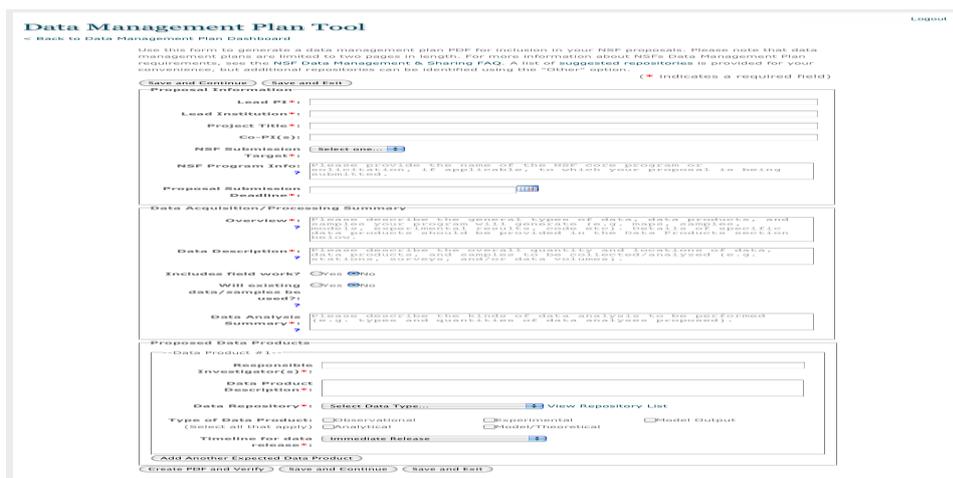
Dieses Tool bietet je nach Fördereinrichtung verschiedene Templates und der Funktionsumfang ändert sich dementsprechend. Hervorzuheben ist das Template von Horizon 2020, was im besonderen Maße die Dynamik von Datenmanagementplänen berücksichtigt. Dort unterteilt sich die Planerstellung in drei Phasen: Initial DMP, Mid-term Review DMP und Final review DMP (vgl. Abb. 3). Zudem unterscheiden sich je nach Phase die Fragen. Weiterhin kennzeichnet sich das Tool durch seine Kollaborationsmöglichkeiten aus. Der Ersteller kann weitere Personen einladen und ihnen unterschiedliche Rechte geben. Danach kann der Plan gemeinsam bearbeitet und zu jeder Frage Notizen gemacht werden. Dabei ist auch der letzte Bearbeitungszeitpunkt sowie der Bearbeiter sichtbar (vgl. Abb. 4).



Die Exportmöglichkeiten sind umfangreich. Es können sieben verschiedene Formate (pdf, csv, html, json, xml, text, docx) ausgewählt werden. Außerdem gibt es Formatierungsoptionen für die Schrift und der Export nur von ausgewählten Bereichen ist möglich.

Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool

Das IEDA DMP-Tool ist von den untersuchten Tools das einzige, welches es erlaubt die Art der vorliegenden Daten (Data Repository, Type of Data Product) anzugeben (vgl. Abb. 5).



Auch die Weise, wie Templates und Policies gehandhabt werden, ist - zumindest für US-Forscher einer der vielen NSF-Divisionen - sehr praktisch: Je nach Art der Daten und der jeweiligen Förderinstitution sind verschiedene DMP-Templates auswählbar, die angezeigten Policies richten sich nach den jeweiligen Einrichtungen bzw. NSF-Divisionen und werden nach der Auswahl automatisch angezeigt und berücksichtigt. Somit werden stets nur die jeweils relevanten Policies angezeigt und sind durch die Verlinkung sofort aufrufbar und können gelesen werden.

LabArchives

Bei dem LabArchives-Tool fallen besonders die extrem umfangreichen Hilfsfunktionen auf: Für Einsteiger stehen zunächst Quick Start Guides sowie PowerPoint- u. Video-Tutorials zur Verfügung, um sich einen ersten Überblick über die Bedienung und den Funktionsumfang zu verschaffen. Außerdem sind Support-Anfragen per Mail und per Ticket direkt im Tool möglich. Zusätzlich existiert eine Kundenhotline und die Benutzer können sich zu verschiedenen Webinars über die Website anmelden, die sogar nach diversen Themengebieten und Regionen (Europa, Nordamerika etc.) unterteilt sind. Eine weitere Anlaufstelle zum Informationsaustausch ist der eigene Community-Blog.

Auch das Versionsmanagement ist umfangreich: Es erfolgt eine Versionierung mit Datum, Uhrzeit (+Zeitzone), Name, IP-Adresse sowie einer Aktualisierung dieser Daten bei vorgenommenen Änderungen. Ähnlich wie bei Wikipedia ist eine Versionsgeschichte einsehbar und jeder Bearbeitungszeitpunkt kann beliebig wiederhergestellt werden. Sämtliche erfolgten Änderungen sind bequem über ein Activity Feed sichtbar (vgl. Abb. 6). Aktuelle Änderungen von Benutzern können beliebig nach Kategorien gefiltert werden und erleichtern so das kollaborative Arbeiten, da jeder Benutzer Änderungen sofort erkennen und mitverfolgen kann.

Feed Types		Filters	# of Events	
Notifications	FH Potsdam	<input checked="" type="checkbox"/> Comments <input checked="" type="checkbox"/> Notebook Modifications <input checked="" type="checkbox"/> Assignments <input checked="" type="checkbox"/> Course Announcements	1 event	Refresh Feed
<input type="checkbox"/> Show unread only				

Marcus Heinrich commented in [FH Potsdam/Protocols/Seminar](#) 28.01.16 Jan 27, 2016 @06:00 PM CET

Datum: 28.01.16
Ort: FH Potsdam

Du hast Meier und Schulze vergessen.

Ebenso sind die Funktionalitäten und der Inhalt insgesamt komfortabel gestaltet, sodass die allgemeine Usability hervorragend und die Verwaltung mehrerer DMP übersichtlich und leicht zu bewerkstelligen ist.

6. Empfehlungen

Die Erstellung von Datenmanagementplänen ist ein dynamischer Prozess. Daher sollte die Art der Versionierung stärker bei der Gestaltung neuer Tools berücksichtigt werden. Dahingehend bietet das Template Horizon 2020 im DMPonline Tool (DCC) eine gute Orientierungsmöglichkeit.

Weiterhin wird empfohlen, die Standardisierungsmöglichkeiten besser auszuschöpfen. Üblicherweise gibt es nur Auswahlfenster für Datumsangaben. Sinnvoll erscheint dieses Verfahren ebenfalls bei der Auswahl von Dateiformaten. So ließen sich unterschiedliche Schreibweisen verhindern und die Erstellung des DMP wird beschleunigt.

Schlussendlich ließe sich die Akzeptanz der Tools durch Empfehlungen von Forschungsförderinstitutionen sowie die Eintragung in Registries steigern. Beispielhaft sei hier die Projektförderungsseite der Australian National Data Service genannt. Dort kann unter anderem eingesehen werden, welche Institution welches Tool nutzt.

Tools wie LabArchives bieten mobile Applikationen zum Arbeiten für unterwegs an. Die Entwicklung von eigenen Anwendungen würde ein Bearbeiten von Datenmanagementplänen im Außeneinsatz auch ohne Internetverbindung ermöglichen. Wenn wieder eine Verbindung zum Internet hergestellt wird, könnten die eingegebenen Daten synchronisiert werden.

Auch mehr Möglichkeiten zur Anpassung der DMP-Tools an das eigene institutionelle Umfeld bzw. an domänenspezifische Aspekte wären wünschenswert. Von den untersuchten Tools bietet nur das Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool zumindest die Möglichkeit, die Art der vorliegenden Daten (Data Repository, Type of Data Product) anzugeben.

Gerade im universitären Bereich erscheint außerdem die Vernetzung der Tools mit Forschungsdaten-Repositorien sinnvoll. Derartige Bestrebungen sind an der Technischen Universität Berlin zu beobachten. Durch die Verknüpfung von Forschungsdaten und den finalen Publikationen, ließe sich ein einheitlicher Workflow etablieren und die Akzeptanz von Datenmanagementplan-Tools und Forschungsdaten könnte davon profitieren.

Die Untersuchung hat gezeigt, dass jedes Tool seine ausgeprägten Stärken besitzt und für seinen jeweiligen Anwenderkreis optimal angepasst ist. Es lässt sich also durchaus für jeden Anwendungsfall ein passendes Tool finden. Die Entwicklung eines Datenmanagementplan-Tools für sämtliche denkbaren Anwendungsfälle scheint aufgrund der großen Unterschiede in den einzelnen Disziplinen illusorisch. Es konnte durch die vorliegende Arbeit jedoch gezeigt werden, dass viele Funktionen auch in anderen Tools nutzbringend eingesetzt werden könnten. Ideal erscheint daher eine internationale Arbeitsgruppe oder eine Community, die einen gemeinsamen Austausch tätigt und somit alle Entwicklungen voranbringt. Denn es mangelt augenscheinlich nicht nur an der digitalen Vernetzung, wenn es um Forschungsdaten geht, sondern auch, wenn es um die Entwicklung entsprechender Tools geht.

7. Literaturverzeichnis

Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, 2011.

DCC Checkliste, URL: <http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data> [07.02.2016].

DFG: Leitlinien zum Umgang mit Forschungsdaten, URL: http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf [07.02.2016].

DMPonline (DCC), URL: <https://dmponline.dcc.ac.uk/> [07.02.2016].

European Commission: Guidelines on Data Management in Horizon 2020, URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [07.02.2016].

Fürste, Fabian: TUB-DMP: Der Datenmanagementplan als Bindeglied zwischen Informationssystem und Repository, URN: <urn:nbn:de:0290-opus-17540> [07.02.2016].

Herbold, Astrid: Europas digitales Gedächtnis ist löchrig, 2016, URL: <http://www.tagesspiegel.de/wissen/forschungsdaten-von-akademien-europas-digitales-gedaechtnis-ist-loechrig/12906856.html> [07.02.2016].

Interdisciplinary Earth Data Alliance (IEDA) DMP-Tool, URL: <http://www.iedadata.org/compliance/plan/> [07.02.2016].

LabArchives, URL: <http://www.labarchives.com/> [07.02.2016].

Princeton University Library: NSF Data Management Plan, URL: http://libguides.princeton.edu/ld.php?content_id=2940897 [07.02.2016].

Universität Bielefeld: DMP Tool, URL: <https://data.uni-bielefeld.de/de/data-management-plan> [07.02.2016].

Universität Jena: Datenmanagementplan, URL: https://www.uni-jena.de/FDM_DataManagementPlan.html [07.02.2016].

Universität Marburg: Datenmanagementplan, URL: <https://www.uni-marburg.de/projekte/forschungsdaten/service> [07.02.2016].

WissGrid: Leitfaden zum Forschungsdaten-Management, URL: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf [07.02.2016].

WissGrid: Checkliste zum Forschungsdaten-Management, URL: <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Checkliste-Forschungsdaten-Management.pdf> [07.02.2016].

8. Anhang

Auflistung der gefundenen Datenmanagementpläne (alphabetische Reihenfolge):

Clarín-D (DataWizard)

URL: <http://www.clarin-d.de/de/aufbereiten/datenmanagementplan-entwickeln>

- Forschungsinfrastrukturprojekt
- stellt forschungsbegleitende Infrastruktur bereit
- an verschiedenen Forschungseinrichtungen in Deutschland vertreten
- durch Bundesministerium für Bildung und Forschung (BMBF) gefördert
- rechtliche Vertretung durch Universität Tübingen

DMP Assistant (University of Alberta)

URL: <https://portagenetwork.ca/>

- Projekt der Canadian Association of Research Libraries (CARL), einem Verbund großer kanadischer Universitätsbibliotheken mit anderen großen kanadischen Forschungsbibliotheken
- Kooperation mit der University of Alberta Libraries und der Universität Montréal sowie Adaption von DMPonline (DCC)

DMPonline (DCC)

URL: <https://dmponline.dcc.ac.uk/>

- Projekt des Digital Curation Centre (DCC)
- speziell auf die Bedürfnisse der britischen Forschungscommunity zugeschnitten

DMPTool (University California)

URL: <https://dmp.cdlib.org/>

- Kostenfreies Tool der University of California
- Gemeinschaftsentwicklung mit acht anderen Partnerinstituten, wie z.B. der National Science Foundation (NSF) und den National Institutes of Health (NIH)

DMP Tool (Universität Bielefeld)

URL: <https://data.uni-bielefeld.de/de/data-management-plan>

- Eigenentwicklung durch Universität Bielefeld

figshare

URL: <https://figshare.com/>

- Repository mit der Möglichkeit zum kollaborativen Arbeiten, dem Erstellen von DOIs und Collections und einem großen Fokus auf Benutzerfreundlichkeit und Privatsphäre

IEDA (Interdisciplinary Earth Data Alliance) DMP Tool

URL: <http://www.iedadata.org/compliance/plan/>

- Community-Projekt für geowissenschaftliche Daten,
- Hauptpartner sind EarthChem und das Marine Geoscience Data System sowie die weiteren Divisionen der National Science Foundation (NSF)

LabArchives

URL: <http://www.labarchives.com/>

- Cloud-basiertes, elektronisches Laborbuch zum Ablegen und Managen aller Arten von Labordaten
- Kooperation mit vielen namhaften Universitäten,
- kostenlos testbar, Vollversion hingegen kostenpflichtig

labfolder

URL: <https://www.labfolder.com/>

- siehe LabArchives

Projects

URL: <https://projects.ac/>

- siehe LabArchives

Research Data Footprints (Deakin University)

URL: <http://www.deakin.edu.au/research/eresearch/manage-data/plan>

- Projekt der Deakin University Library in Victoria, Australien

TUB-DMP (TU Berlin)

URL: https://www.szf.tu-berlin.de/menue/dienste_tools/

- im Aufbau

- Entwicklung durch Servicezentrum Forschungsdaten und -publikationen der TU Berlin

University of Queensland (Australien)

URL: <https://dmponline.app.uq.edu.au/>

- Eigenentwicklung, basierend auf DMPonline (DCC)

Queensland University of Technology Brisbane (Australien)

URL: <https://dmp.qut.edu.au/>

- Tool der University of Queensland, Australien

Zenodo

URL: <http://zenodo.org/>

- hauptsächlich für EU-Projekte
- benutzt das CERN Data Center zur Datenverwaltung
- Anbindung an GitHub, OpenAIRE, ORCID, CrossRef, DropBox etc.
- Anlegen von ganzen Communities zum Datenaustausch bzw. -speichern möglich, DOIs

Anlage 2: RDMO-Fragenkatalog

Allgemein / Thema

- Wie lautet die primäre Forschungsfrage des Projektes?
- Bitte geben sie einige Schlagworte zur Forschungsfrage an.
- Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen?

Optionen:

- Geistes- und Sozialwissenschaften / Alte Kulturen
- Geistes- und Sozialwissenschaften / Geschichtswissenschaften
- Geistes- und Sozialwissenschaften / Kunst-, Musik-, Theater- und Medienwissenschaften
- Geistes- und Sozialwissenschaften / Sprachwissenschaften
- Geistes- und Sozialwissenschaften / Literaturwissenschaft

- Geistes- und Sozialwissenschaften / Außereuropäische Sprachen und Kulturen, Sozial- und Kulturanthropologie, Judaistik und Religionswissenschaft
- Geistes- und Sozialwissenschaften / Theologie
- Geistes- und Sozialwissenschaften / Philosophie
- Geistes- und Sozialwissenschaften / Erziehungswissenschaft
- Geistes- und Sozialwissenschaften / Psychologie
- Geistes- und Sozialwissenschaften / Sozialwissenschaften
- Geistes- und Sozialwissenschaften / Wirtschaftswissenschaften
- Geistes- und Sozialwissenschaften / Rechtswissenschaften
- Lebenswissenschaften / Grundlagen der Biologie und Medizin
- Lebenswissenschaften / Pflanzenwissenschaften
- Lebenswissenschaften / Zoologie
- Lebenswissenschaften / Mikrobiologie, Virologie und Immunologie
- Lebenswissenschaften / Medizin
- Lebenswissenschaften / Neurowissenschaft
- Lebenswissenschaften / Agrar-, Forstwissenschaften, Gartenbau und Tiermedizin
- Naturwissenschaften / Molekülchemie
- Naturwissenschaften / Chemische Festkörper- und Oberflächenforschung
- Naturwissenschaften / Physikalische und Theoretische Chemie
- Naturwissenschaften / Analytik, Methodenentwicklung (Chemie)
- Naturwissenschaften / Biologische Chemie und Lebensmittelchemie
- Naturwissenschaften / Polymerforschung
- Naturwissenschaften / Physik der kondensierten Materie
- Naturwissenschaften / Optik, Quantenoptik und Physik der Atome, Moleküle und Plasmen
- Naturwissenschaften / Teilchen, Kerne und Felder
- Naturwissenschaften / Statistische Physik, Weiche Materie, Biologische Physik, Nichtlineare Dynamik
- Naturwissenschaften / Astrophysik und Astronomie
- Naturwissenschaften / Mathematik
- Naturwissenschaften / Atmosphären- und Meeresforschung
- Naturwissenschaften / Geologie und Paläontologie
- Naturwissenschaften / Geophysik und Geodäsie
- Naturwissenschaften / Geochemie, Mineralogie und Kristallographie
- Naturwissenschaften / Geographie
- Naturwissenschaften / Wasserforschung
- Ingenieurwissenschaften / Produktionstechnik

- Ingenieurwissenschaften / Mechanik und Konstruktiver Maschinenbau
- Ingenieurwissenschaften / Verfahrenstechnik, Technische Chemie
- Ingenieurwissenschaften / Wärmeenergie-technik, Thermische Maschinen, Strömungsmechanik
- Ingenieurwissenschaften / Werkstofftechnik
- Ingenieurwissenschaften / Materialwissenschaft
- Ingenieurwissenschaften / Systemtechnik
- Ingenieurwissenschaften / Elektrotechnik
- Ingenieurwissenschaften / Informatik
- Ingenieurwissenschaften / Bauwesen und Architektur

Allgemein / Projektablauf

- Wann beginnt die Projektlaufzeit?
- Wann endet die Projektlaufzeit?

Allgemein / Projektpartner

- Welche Personen oder Institutionen sind verantwortlich für die Projektkoordination?
- Projektpartner
- Gibt es an der Einrichtung Regeln oder Richtlinien zum Umgang mit Forschungsdaten? Wenn ja, welches sind diese und wie ist der Grad an Verbindlichkeit?
- Wer ist bei diesem Partner der/die Ansprechpartner/in für das Datenmanagement?

Allgemein / Förderung

- Wer fördert das Projekt?
- In welchem Förderprogramm wird das Projekt gefördert?
- Gibt es von Seiten des Forschungsförderer Vorgaben oder Richtlinien bezüglich des Umgangs mit den im Projekt erhobenen Forschungsdaten? Wenn ja, welches sind diese und wie ist der Grad an Verbindlichkeit?

Allgemein / Weitere Anforderungen

- Gibt es von weiteren Seiten (z.B. von der Fachcommunity) Anforderungen an das Datenmanagement, die beachtet werden müssen?

Optionen:

- Ja
- Nein
- Noch zu klären

- Welches sind diese Anforderungen an das Datenmanagement?

Inhaltliche Einordnung / Daten

- Um was für einen Datensatz handelt es sich?
- Wird der Datensatz selbst erzeugt oder nachgenutzt?

Optionen:

- Erzeugt
- Nachgenutzt
- Wenn nachgenutzt, wer hat den Datensatz erzeugt?
- Wenn nachgenutzt, unter welcher Adresse oder URL ist der Datensatz verfügbar?
- Für welche Personen, Gruppen oder Institutionen könnte dieser Datensatz (für die Nachnutzung) von Interesse sein? Für welche Szenarien ist dies denkbar?
- Ist der Datensatz reproduzierbar, d.h. ließe sich er sich, wenn er verlorengehe, erneut erstellen oder erheben?

Optionen:

- ja, mit geringem Aufwand
- ja, mit mäßigem, aber vertretbarem Aufwand
- nein bzw. nur mit unverhältnismäßig großem Aufwand
- nein, die Daten sind per se nicht reproduzierbar

Inhaltliche Einordnung / Software

- Um was für eine Art Code handelt es sich?
- Wird der Code selbst erzeugt oder nachgenutzt?

Optionen:

- Erzeugt
- Nachgenutzt
- Wenn nachgenutzt, wer hat den Code entwickelt?
- Wenn nachgenutzt, unter welcher Adresse oder URL ist der Code verfügbar?
- Für welche Personen, Gruppen oder Institutionen könnte diese Software (für die Nachnutzung) von Interesse sein? Für welche Szenarien ist dies denkbar?
- Wie und durch wen könnte der Code nach Projektende gepflegt und/oder weiterentwickelt werden?

Technische Einordnung / Daten

- Wie groß ist die erwartete Größe des Datensatzes?

Optionen:

- weniger als 1 GB
- einige GB bis ein TB
- einige TB bis 100 TB
- mehr als 100 TB
- genau: _____
- noch nicht bestimmt
- Gegebenenfalls, wie hoch ist die erwartete Erzeugungsrate der Daten pro Monat?
- In welchen Formaten liegen die Daten vor?
- Welche Instrumente, Software, Technologien oder Verfahren werden zur Erzeugung oder Erfassung der Daten genutzt?
- Welche Software, Verfahren oder Technologien sind notwendig, um die Daten zu nutzen?
- Werden verschiedene Versionen des Datensatzes erzeugt?
- Welche Technologie bzw. welches Tool wird zur Versionierung verwendet?

Optionen:

- Einfaches Kopieren
- Versionskontrollsystem: _____
- Sonstiges: _____
- Noch nicht entschieden
- Welche Versionierungsstrategie wird für diesen Datensatz angewandt?

Technische Einordnung / Software

- Welche Technologie bzw. welches Tool wird zur Versionierung dieser Software verwendet?

Optionen:

- Einfaches Kopieren
- Versionskontrollsystem: _____
- Sonstiges: _____
- Noch nicht entschieden
- Welche Versionierungsstrategie wird für diese Software angewandt?
 - Datennutzung / Datennutzung
- Wozu / wie wird dieser Datensatz während des Projektes genutzt?
- Wie häufig wird dieser Datensatz genutzt?

- In welchem Umfang werden Infrastrukturressourcen benötigt (CPU-Stunden, Bandbreite etc.)?
- Gibt es beabsichtigte (ggf. auch potentielle) Nutzungsszenarien, für die Unterstützung durch Datenmanagement- oder IT-ExpertInnen sinnvoll oder notwendig ist?
 - Datennutzung / Datenspeicherung und Backup
- Wo wird der Datensatz während des Projekts gespeichert?
- Unter welcher URL kann der Datensatz während des Projekts abgerufen werden?
- Gibt es projektinterne Richtlinien zur einheitlichen Organisation der Daten? Wenn ja, wo sind diese festgehalten?
- Gibt es eine projektinterne Richtlinie zur Benennung der Daten? Wenn ja, wo ist diese festgehalten?
- Wer darf auf den Datensatz zugreifen?
- Wie und wie oft werden Backups der Daten erstellt?
- Wer ist verantwortlich für die Erstellung der Backups?

Datennutzung / Kollaboratives Arbeiten

- Werden die Daten kollaborativ genutzt / bearbeitet?

Optionen:

 - Ja, von mehreren Personen an verschiedenen Institutionen
 - Ja, von mehreren Personen derselben Arbeitsgruppe an derselben Institution
 - Nein
- Welche Plattform, welche Werkzeuge werden zum kollaborativen Arbeiten an Daten und Publikationen genutzt?
- Wie ist das kollaborative Arbeiten an denselben Dateien geregelt?

Datennutzung / Weitergabe und Veröffentlichung

- Soll dieser Datensatz veröffentlicht oder geteilt werden?

Optionen:

 - Ja, intern mit allen, solange sie die Daten nicht veröffentlichen oder nach außen weitergeben
 - Ja, extern in begrenztem Umfang mit individueller Freigabe
 - Ja, extern für alle
 - Nein
- Unter welchen Nutzungsbedingungen sollen die Daten veröffentlicht bzw. geteilt werden?

Optionen:

- Namensnennung
- keine kommerzielle Nutzung
- keine Bearbeitung
- Weitergabe unter gleichen Bedingungen
- Andere: _____
- Welche Lizenz soll hierfür verwendet werden?

Optionen:

- Der Datensatz wird unter folgender Lizenz veröffentlicht: _____
- noch zu klären
- Soll diese Software veröffentlicht oder geteilt werden?

Optionen:

- Ja, intern mit allen, solange sie die Daten nicht veröffentlichen oder nach außen weitergeben
- Ja, extern in begrenztem Umfang mit individueller Freigabe
- Ja, extern für alle
- Nein
- Unter welchen Nutzungsbedingungen sollen die Software veröffentlicht bzw. geteilt werden?

Optionen:

- Namensnennung
- keine kommerzielle Nutzung
- keine Bearbeitung
- Weitergabe unter gleichen Bedingungen
- Andere: _____
- Welche Lizenz soll hierfür verwendet werden?

Optionen:

- Der Code wird unter folgender Lizenz veröffentlicht: _____
- noch zu klären

Datennutzung / Qualitätssicherung

- Welche Maßnahmen zur Qualitätssicherung werden für diesen Datensatz ergriffen?
- Welche Maßnahmen zur Qualitätssicherung werden für diese Software ergriffen?

- Wie wird die Integration zwischen nachgenutzten und erzeugten Daten gewährleistet?

Datennutzung / Kosten

- Welcher Personalaufwand für das Datenmanagement entsteht im Rahmen der Erhebung, Erstellung oder Akquise der Daten im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten für das Datenmanagement entstehen im Rahmen der Erhebung, Erstellung oder Akquise der Daten im Projekt?
- Welcher Personalaufwand für das Datenmanagement entsteht im Zusammenhang mit der Nutzung der Daten im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten für das Datenmanagement entstehen im Zusammenhang mit der Nutzung der Daten im Projekt?
- Welcher Personalaufwand entsteht im Zusammenhang mit der Speicherung der Daten während des Projekts?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit der Speicherung der Datensätze während des Projekts?

Metadaten und Referenzierung / Metadaten

- Welche Informationen sind für Außenstehende notwendig, um die Daten zu verstehen (d.h. ihre Erhebung bzw. Entstehung, Analyse sowie die auf ihrer Basis gewonnenen Forschungsergebnisse nachvollziehen) und nachnutzen zu können?

Optionen:

- Ort
- Inhalt
- Methodik
- Erzeugungsprozess
- Technologie
- Zeit
- Quellen
- Akteure
- Identifikatoren
- Andere: _____
- Welche Standards, Ontologien, Klassifikationen etc. werden zur Beschreibung der Daten und Kontextinformation genutzt?

Optionen:

- Es werden disziplinspezifische Standards, Klassifikationen etc. genutzt:

- Es wird ein eigenes Beschreibungssystem genutzt. Dieses ist an folgendem Ort dokumentiert: _____
- Es wurde noch nicht entschieden, mit welchem System die Metadaten und Kontextinformationen beschrieben werden
- Es wird kein festgelegtes System zur Beschreibung genutzt
- Sonstiges: _____
- Welche Metadaten werden automatisch erhoben?
- Welche Metadaten werden semi-automatisch erhoben?
- Welche Metadaten werden manuell erhoben?
- Werden Metadaten und Kontextinformation auf Korrektheit und Vollständigkeit geprüft?

Optionen:

- Automatische Prüfung auf Vollständigkeit
- Manuelle Prüfung auf Korrektheit
- Manuelle Prüfung auf Vollständigkeit
- Sonstiges: _____
- Wer ist verantwortlich für die Dokumentation und Prüfung der Metadaten und Kontextinformationen auf Richtigkeit und Vollständigkeit?
- Welcher Personalaufwand entsteht im Zusammenhang mit der Erstellung von Metadaten und Kontextinformation im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit der Erstellung von Metadaten und Kontextinformation im Projekt?

Metadaten und Referenzierung / Objektstruktur, Granularität und Referenzierung

- Wie sind die Daten strukturiert? In welchem Verhältnis stehen die einzelnen Komponenten zueinander? In welchem Verhältnis steht der Datensatz zu anderen im Projekt erhobenen oder genutzten Datensätzen?
- Sollen für diesen Datensatz persistente Identifikatoren (PIDs) genutzt werden?
- Welches System von persistenten Identifikatoren sollen genutzt werden?

Optionen:

- Handle / DOI
- PURL
- ARK
- URN
- ISLRN
- Anderes: _____
- Welche (Sub-)Entitäten / Untereinheiten sollten sinnvollerweise eigene Identifikatoren erhalten? Welche dieser Identifikatoren sollten dauerhaft und zitierfähig sein?
- Wer ist verantwortlich für die Pflege der PIDs und die Objektpflege (d.h. die Langzeitarchivierung des Objekts und somit dafür, dem PID-Service einen Objektumzug und die neue Adresse mitzuteilen)?
- Welcher Personalaufwand entsteht im Zusammenhang mit der Vergabe von persistenten Identifikatoren im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit persistenten Identifikatoren im Projekt?

Rechtliche Fragen / Recht allgemein

- Muss die rechtliche Situation verschiedener Länder berücksichtigt werden?

Rechtliche Fragen / Sensible Daten

- Werden im Projekt personenbezogene Daten genutzt?
- Enthalten die im Projekt genutzten Daten "Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualeben" (BDSG §3, Abs.9)?
- Sind unter den personenbezogenen Daten solche, die im Projekt erhoben werden?
- Werden die Daten anonymisiert oder pseudonymisiert?

Optionen:

- Ja, während der Erhebung
- Ja, vor / zu Beginn der Datenanalyse
- Ja, nach der Datenanalyse / vor der Publikation
- Nein
- In welchem Umfang wird die "informierte Einwilligung" der Betroffenen eingeholt?

Optionen:

- Zur Analyse/Nutzung der Daten im Rahmen des Projektes sowie zur Nachnutzung
- Nur zur Analyse/Nutzung der Daten im Rahmen des Projektes
- Es wird keine "informierte Einwilligung" eingeholt
- Wenn keine "informierte Einwilligung" eingeholt wird, begründen Sie dies bitte.
- Wo und wie sind die "informierten Einwilligungen" abgelegt?
- Welches Gesetz ist bezüglich der Fragen des Datenschutzes für das Projekt maßgeblich?

Optionen:

- Bundesdatenschutzgesetz
- Landesdatenschutzgesetz Baden-Württemberg
- Landesdatenschutzgesetz Bayern
- Landesdatenschutzgesetz Berlin
- Landesdatenschutzgesetz Bremen
- Landesdatenschutzgesetz Brandenburg
- Landesdatenschutzgesetz Hamburg
- Landesdatenschutzgesetz Mecklenburg-Vorpommern
- Landesdatenschutzgesetz Hessen

- Landesdatenschutzgesetz Nordrhein-Westfalen
- Landesdatenschutzgesetz Rheinland-Pfalz
- Landesdatenschutzgesetz Niedersachsen
- Landesdatenschutzgesetz Saarland
- Landesdatenschutzgesetz Sachsen
- Landesdatenschutzgesetz Sachsen-Anhalt
- Landesdatenschutzgesetz Schleswig-Holstein
- Landesdatenschutzgesetz Thüringen
- Sozialgesetzbuch X
- Bundesstatistikgesetz
- EU Datenschutzrichtlinie 95/46/EG
- Verordnung (EU) 2016/679
- Sonstiges: _____
- Gibt es nicht-personenbezogene Daten, die aus ethischen, kommerziellen oder anderen Gründen sensibel sind (z.B. weil sie Betriebs- oder Geschäftsgeheimnisse, Ortsangaben zu bedrohten Tier- oder Pflanzenarten u.a.m. enthalten)?
- Um welche nicht personenbezogenen sensiblen Daten handelt es sich?
- Welche Maßnahmen zur Gewährleistung der Datensicherheit werden getroffen (z.B. Schutz vor unbefugtem Zugriff)?
- Welcher Personalaufwand entsteht für die Anonymisierung von sensiblen Daten im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit der Anonymisierung von sensiblen Daten im Projekt?
- Welcher Personalaufwand entsteht im Zusammenhang mit technischen Sicherheitsmaßnahmen für sensible Daten im Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Wie hoch sind die weiteren Sachkosten für technische Sicherheitsmaßnahmen für sensible Daten im Projekt?

- Rechtliche Fragen / Urheber- oder verwandte Schutzrechte
- Werden Daten oder Software genutzt, die durch Urheber- oder verwandte Schutzrechte geschützt sind?
- Be- oder entstehen an diesem Datensatz Urheberrechte?

Optionen:

- Recht des Herstellers eines Tonträgers
- Recht des Sendeunternehmers
- Recht des Lichtbildners
- Recht des Verfassers sichtender wissenschaftlicher Ausgaben
- Recht des Datenbankherstellers
- Andere Urheberrechte: _____
- Be- oder entstehen für diesen Datensatz andere Schutzrechte?

Optionen:

- Ergänzende Schutzzertifikate
- Halbleiterschutz bzw. Schutz von Topografien
- Gebrauchsmuster
- Patente
- Sortenschutz (Pflanzenzüchtungen)
- Markenrecht
- geografische Herkunftsangaben
- eingetragene Designs (Designs und Modelle)
- geschäftliche Bezeichnungen (Unternehmenskennzeichen und Werktitel)
- Andere: _____
- Wer hält diese Rechte? Welche Nutzungsrechte werden eingeräumt bzw. wurden / werden eingeholt?
- Sind für diese Software Urheberrechte zu beachten?

Optionen:

- Recht des Herstellers eines Tonträgers
- Recht des Verfassers sichtender wissenschaftlicher Ausgaben
- Recht des Sendeunternehmers
- Recht des Datenbankherstellers
- Recht des Lichtbildners
- Andere Urheberrechte: _____
- Sind für diese Software weitere Schutzrechte zu beachten?

Optionen:

- geschäftliche Bezeichnungen (Unternehmenskennzeichen und Werktitel)
 - eingetragene Designs (Designs und Modelle)
 - geografische Herkunftsangaben
 - Sortenschutz (Pflanzenzüchtungen)
 - Patente
 - Halbleiterschutz bzw. Schutz von Topografien
 - Ergänzende Schutzzertifikate
 - Gebrauchsmuster
 - Markenrecht
 - Andere: _____
- Wer hält diese Rechte? Welche Nutzungsrechte werden eingeräumt bzw. wurden / werden eingeholt?
 - Welcher Personalaufwand entsteht im Zusammenhang mit Urheber- oder verwandten Schutzrechten im Projekt?

Bereich:

- Minimum: 0,0
 - Maximum: 12,0
 - Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit Urheber- und verwandten Schutzrechten im Projekt?

Speicherung und Langzeitarchivierung / Auswahl

- Auf Basis welcher Kriterien / Regeln werden die Daten zur Archivierung (nach Projektende) ausgesucht?
- Durch wen erfolgt diese Auswahl?

Speicherung und Langzeitarchivierung / Langzeitarchivierung

- Muss dieser Datensatz langfristig aufbewahrt werden?
- Aus welchen Gründen müssen die Daten langfristig aufbewahrt werden?

Optionen:

- Grundlage einer Publikation / Nachweis guter wissenschaftlicher Praxis
- Nachnutzung in Folgeprojekten oder durch andere
- Aufgrund gesetzlicher Bestimmungen
- Dokumentation aufgrund gesellschaftlicher Relevanz

- Selbstverpflichtung
- Andere: _____
- Wie lange müssen die Daten aufbewahrt werden?
- Wo werden die Daten nach Projektende gespeichert bzw. archiviert?

Optionen:

- Eigene Institution
- Disziplinspezifisches Datenzentrum: _____
- Generisches Datenzentrum: _____
- Wurde noch nicht entschieden
- Sonstiges: _____
- Sollen die Daten erst nach Ablauf einer Sperrfrist zugänglich gemacht werden?
- Welcher Personalaufwand entsteht im Zusammenhang mit Langzeitarchivierung für dieses Projekt?

Bereich:

- Minimum: 0,0
- Maximum: 12,0
- Schrittgröße: 0,1
- Welche Sachkosten entstehen im Zusammenhang mit Langzeitarchivierung für dieses Projekt?

Anlage 3: RDMO-Entitäten/-Attribute

Art	Name
Entität	project.additional_requirements
Attribut	project.additional_requirements.requirements
Attribut	project.additional_requirements.yesno
Entität	project.costs
Entität	project.costs.creation
Attribut	project.costs.creation.expenses
Attribut	project.costs.creation.personnel
Entität	project.costs.ipr
Attribut	project.costs.ipr.expenses
Attribut	project.costs.ipr.personnel
Entität	project.costs.metadata
Attribut	project.costs.metadata.expenses
Attribut	project.costs.metadata.personnel
Entität	project.costs.pid
Attribut	project.costs.pid.expenses
Attribut	project.costs.pid.personnel
Entität	project.costs.preservation
Attribut	project.costs.preservation.expenses
Attribut	project.costs.preservation.personnel
Entität	project.costs.sensitive_data
Entität	project.costs.sensitive_data.anonymization
Attribut	project.costs.sensitive_data.anonymization.expenses
Attribut	project.costs.sensitive_data.anonymization.personnel

Entität	project.costs.sensitive_data.security
Attribut	project.costs.sensitive_data.security.expenses
Attribut	project.costs.sensitive_data.security.personnel
Entität	project.costs.storage
Attribut	project.costs.storage.expenses
Attribut	project.costs.storage.personnel
Entität	project.costs.usage
Attribut	project.costs.usage.expenses
Attribut	project.costs.usage.personnel
Entität	project.dataset
Attribut	project.dataset.collaboration_organisation
Attribut	project.dataset.collaboration_tools
Attribut	project.dataset.collaboration_yesno
Attribut	project.dataset.creation_methods
Entität	project.dataset.creator
Attribut	project.dataset.creator.name
Attribut	project.dataset.description
Attribut	project.dataset.format
Attribut	project.dataset.id
Entität	project.dataset.ipr
Attribut	project.dataset.ipr.copyrights
Attribut	project.dataset.ipr.other_rights
Entität	project.dataset.ipr.owner
Attribut	project.dataset.ipr.owner.name
Entität	project.dataset.metadata
Attribut	project.dataset.metadata.creation_automatic

Attribut	project.dataset.metadata.creation_manual
Attribut	project.dataset.metadata.creation_semi_automat
Attribut	project.dataset.metadata.description
Attribut	project.dataset.metadata.information
Attribut	project.dataset.metadata.quality_assurance
Entität	project.dataset.metadata.responsible_person
Attribut	project.dataset.metadata.responsible_person.name
Attribut	project.dataset.origin
Entität	project.dataset.pids
Entität	project.dataset.pids.responsible_person
Attribut	project.dataset.pids.responsible_person.name
Attribut	project.dataset.pids.subentities
Attribut	project.dataset.pids.system
Attribut	project.dataset.pids.yesno
Entität	project.dataset.preservation
Attribut	project.dataset.preservation.accessibility
Attribut	project.dataset.preservation.duration
Attribut	project.dataset.preservation.motivation
Attribut	project.dataset.preservation.repository
Attribut	project.dataset.preservation.yesno
Attribut	project.dataset.quality_assurance
Attribut	project.dataset.rate
Attribut	project.dataset.reproducibility
Attribut	project.dataset.reuse_scenario
Attribut	project.dataset.sharing_conditions

Attribut	project.dataset.sharing_license
Attribut	project.dataset.sharing_yesno
Attribut	project.dataset.size
Entität	project.dataset.storage
Attribut	project.dataset.storage.backup_frequency
Attribut	project.dataset.storage.naming
Attribut	project.dataset.storage.organization
Attribut	project.dataset.storage.permissions
Entität	project.dataset.storage.responsible_person
Attribut	project.dataset.storage.responsible_person.name
Attribut	project.dataset.storage.type
Attribut	project.dataset.storage.uri
Attribut	project.dataset.structure
Attribut	project.dataset.technical_integration
Attribut	project.dataset.title
Attribut	project.dataset.uri
Attribut	project.dataset.usage_description
Attribut	project.dataset.usage_frequency
Attribut	project.dataset.usage_infrastructure
Attribut	project.dataset.usage_methods
Attribut	project.dataset.usage_support
Attribut	project.dataset.versioning_strategy
Attribut	project.dataset.versioning_technology
Attribut	project.dataset.versioning_yesno
Entität	project.funder

Attribut	project.funder.id
Attribut	project.funder.name
Entität	project.funder.program
Attribut	project.funder.program.title
Attribut	project.funder.requirements
Entität	project.legal_aspects
Attribut	project.legal_aspects.international_yesno
Entität	project.legal_aspects.ipr
Attribut	project.legal_aspects.ipr.yesno
Entität	project.legal_aspects.other_sensitive_data
Attribut	project.legal_aspects.other_sensitive_data.description
Attribut	project.legal_aspects.other_sensitive_data.yesno
Entität	project.legal_aspects.personal_data
Attribut	project.legal_aspects.personal_data.anonymization
Attribut	project.legal_aspects.personal_data.bds_g_3_9
Entität	project.legal_aspects.personal_data.consent
Attribut	project.legal_aspects.personal_data.consent.extent
Attribut	project.legal_aspects.personal_data.consent.record
Attribut	project.legal_aspects.personal_data.consent.statement
Attribut	project.legal_aspects.personal_data.creation
Attribut	project.legal_aspects.personal_data.privacy_laws
Attribut	project.legal_aspects.personal_data.yesno
Attribut	project.legal_aspects.security_measures

Entität	project.partner
Attribut	project.partner.id
Attribut	project.partner.name
Entität	project.partner.contact_person
Attribut	project.partner.contact_person.name
Attribut	project.partner.requirements
Entität	project.preservation
Attribut	project.preservation.criteria
Entität	project.preservation.responsible_person
Attribut	project.preservation.responsible_person.name
Entität	project.coordination
Attribut	project.coordination.name
Entität	project.schedule
Attribut	project.schedule.end_date
Attribut	project.schedule.start_date
Entität	project.research_field
Attribut	project.research_field.title
Entität	project.research_question
Attribut	project.research_question.keywords
Attribut	project.research_question.title
Entität	project.software
Entität	project.software.creator
Attribut	project.software.creator.name
Attribut	project.software.description
Attribut	project.software.id
Entität	project.software.ipr

Attribut	project.software.ipr.copyrights
Attribut	project.software.ipr.other_rights
Entität	project.software.ipr.owner
Attribut	project.software.ipr.owner.name
Attribut	project.software.origin
Attribut	project.software.quality_assurance
Entität	project.software.reuse
Attribut	project.software.reuse_scenario
Attribut	project.software.sharing_conditions
Attribut	project.software.sharing_license
Attribut	project.software.sharing_yesno
Attribut	project.software.sustainability
Attribut	project.software.title
Attribut	project.software.uri
Attribut	project.software.versioning_strategy
Attribut	project.software.versioning_technology

Anlage 4: XML-Export-Programmcode

```
from __future__ import unicode_literals

from django.utils import six

from django.utils.xmlutils import SimplerXMLGenerator

from django.utils.six.moves import StringIO

from django.utils.encoding import smart_text

from rest_framework.renderers import BaseRenderer

class XMLRenderer(BaseRenderer):

    media_type = 'application/xml'

    format = 'xml'

    def render(self, data):

        if data is None:

            return ''

        stream = StringIO()

        xml = SimplerXMLGenerator(stream, "utf-8")

        xml.startDocument()

        xml.startElement('Domain', {

            'xmlns:dc': "http://purl.org/dc/elements/1.1/"

        })

        for attribute_entity in data:

            if attribute_entity['is_attribute']:
```

```

        self._attribute(xml, attribute_entity)
    else:
        self._attribute_entity(xml, attribute_entity)

xml.endElement('Domain')
xml.endDocument()
return stream.getvalue()

def _attribute(self, xml, attribute):
    xml.startElement('Attribute', {})
    self._text_element(xml, 'dc:title', {}, attribute["title"])
    self._text_element(xml, 'dc:description', {}, attribute["description"])
    self._text_element(xml, 'dc:uri', {}, attribute["uri"])
    self._text_element(xml, 'is_collection', {}, attribute["is_collection"])
    self._text_element(xml, 'value_type', {}, attribute["value_type"])
    self._text_element(xml, 'unit', {}, attribute["unit"])

    if 'options' in attribute and attribute['options']:
        xml.startElement('Options', {})

        for option in attribute['options']:
            self._option(xml, option)

        xml.endElement('Options')

    if 'range' in attribute and attribute['range']:
        self._range(xml, attribute['range'])

    if 'conditions' in attribute and attribute['conditions']:
        xml.startElement('Conditions', {})

```

```

    for conditions in attribute['conditions']:
        self._conditions(xml, conditions)

    xml.endElement('Conditions')

    if 'verboseName' in attribute and attribute['verboseName']:
        self._verboseName(xml, attribute['verboseName'])

    xml.endElement('Attribute')

def _attribute_entity(self, xml, attribute_entity):
    xml.startElement('AttributeEntity', {})

    self._text_element(xml, 'dc:title', {}, attribute_entity["title"])
    self._text_element(xml, 'dc:description', {}, attribute_entity["description"])
    self._text_element(xml, 'dc:uri', {}, attribute_entity["uri"])
    self._text_element(xml, 'is_collection', {}, attribute_entity["is_collection"])

    if 'children' in attribute_entity:
        xml.startElement('Children', {})

        for child in attribute_entity['children']:
            if child['is_attribute']:
                self._attribute(xml, child)
            else:
                self._attribute_entity(xml, child)

        xml.endElement('Children')

    if 'conditions' in attribute_entity and attribute_entity['conditions']:

```

```

        xml.startElement('Conditions', {})

        for conditions in attribute_entity['conditions']:
            self._conditions(xml, conditions)

        xml.endElement('Conditions')

    if 'verbose_name' in attribute_entity and attribute_entity['verbose_name']:
        self._verbose_name(xml, attribute_entity['verbose_name'])

    xml.endElement('AttributeEntity')

def _option(self, xml, option):
    xml.startElement('Option', {})
    self._text_element(xml, 'order', {}, option["order"])
    self._text_element(xml, 'text_de', {}, option["text_de"])
    self._text_element(xml, 'text_en', {}, option["text_en"])
    self._text_element(xml, 'additional_input', {}, option["additional_input"])
    xml.endElement('Option')

def _range(self, xml, range):
    xml.startElement('range', {})
    self._text_element(xml, 'minimum', {}, range["minimum"])
    self._text_element(xml, 'maximum', {}, range["maximum"])
    self._text_element(xml, 'step', {}, range["step"])
    xml.endElement('range')

def _conditions(self, xml, conditions):
    xml.startElement('condition', {})
    self._text_element(xml, 'source_attribute', {}, conditions["source_attribute"])

```

```

self._text_element(xml, 'relation', {}, conditions["relation"])

self._text_element(xml, 'target_text', {}, conditions["target_text"])

self._text_element(xml, 'target_option', {}, conditions["target_option"])

xml.endElement('condition')

def _verbosename(self, xml, verbosename):

    xml.startElement('verbosename', {})

    self._text_element(xml, 'name_en', {}, verbosename["name_en"])

    self._text_element(xml, 'name_de', {}, verbosename["name_de"])

    self._text_element(xml, 'name_plural_en', {}, verbosename["name_plural_en"])

    self._text_element(xml, 'name_plural_de', {}, verbosename["name_plural_de"])

    xml.endElement('verbosename')

def _text_element(self, xml, tag, attributes, text):

    xml.startElement(tag, attributes)

    if text is not None:

        xml.characters(smart_text(text))

    xml.endElement(tag)

```

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Masterarbeit mit dem Titel „Datenmodellierung für Forschungsdatenmanagementpläne“ selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus fremden Quellen entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt oder an anderer Stelle veröffentlicht. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Potsdam, 15.08.2016