

Data Management Plan

Eine Anleitung zur Erstellung von Data Management Plänen

Projekt e-Infrastructures Austria

e-infrastructures
austria

Version 2.0
Mai 2015



Dieses Werk bzw. dieser Inhalt steht unter einer Creative Commons Namensnennung 4.0 International Lizenz.
<http://creativecommons.org/licenses/by/4.0/legalcode>

Dieses Dokument entstand im Rahmen des Projektes e-Infrastructures Austria.

Folgende Personen waren an der Entstehung des Dokumentes (Print- und Online Version) beteiligt:

Susanne Blumesberger, Universität Wien
Paolo Budroni, Universität Wien
Raman Ganguly, Universität Wien
Tomász Miksa, Technische Universität Wien
Ralf Pausz, Universität Wien
José Luis Preza, Universität Wien
Barbara Sánchez Solís, Universität Wien

In Zusammenarbeit mit den Mitarbeiterinnen und Mitarbeitern des Clusters C - Aufbau eines Wissensnetzwerks: Erarbeitung eines Referenzmodells für den Aufbau von Repositorien

Veronika Gründhammer, Universität Innsbruck
Hannes Kulovits, Österr. Staatsarchiv
Karin Lackner, Universität Graz
Edith Leitner, Mozarteum Salzburg
Joachim Losehand, Creative Commons Austria
Linda Ohrtmann, Universität Salzburg

Diese Version wird auf folgenden Plattformen veröffentlicht und zur Verfügung gestellt:

www.univie.ac.at/phaidra

www.univie.ac.at/phaidraplus

Kontakt: office@e-infrastructures.at

□ Gründe für einen Data Management Plan (DMP)

Data Management Pläne (DMP) entstehen hauptsächlich, um Ressourcen, vor allem finanzielle Ressourcen, effizient einzusetzen und um digitale Daten langfristig auffindbar, verständlich und nachnutzbar zu machen. Die sorgfältige Planung eines Projekts bringt eine Reihe von Vorteilen mit sich: Kostentransparenz von Beginn an sowie einen optimalen Einsatz von Know-how, Infrastruktur und Dienstleistungen. DMP fördern darüber hinaus das Verständnis für die eigenen Arbeitsprozesse bei der Durchführung von Projekten. In diesem Sinne helfen sie auch, eine unserer wichtigsten Ressourcen, nämlich „Zeit“, effektiv zu nutzen.

DMP bringen nicht nur Vorteile für die Projektleitung bzw. das Management, sie sind auch für folgende Stakeholder von Bedeutung: den Projektantragsteller, das Repository-Management, die Institution, den Fördergeber und die Policy-Verantwortlichen. Aus diesem Grund können DMP in ihrer konkreten Form variieren, je nachdem in welchem Kontext sie entstehen.

In der vorliegenden Anleitung richten wir uns vor allem an zwei Nutzergruppen: den Projektantragsteller und das Projektmanagement. So werden bei der Erstellung eines DMPs die wichtigsten Phasen der Durchführung und damit verbundene Arbeitsprozesse sichtbar.

Wir weisen darauf hin, dass wir uns dieser Thematik und auch jenem wichtigen Teil der Arbeit, der sich hinter dem Akronym „PMP“ (Process Management Pläne) verbirgt, im Rahmen unserer Repositorien-Beratung widmen.

□ Die Phasen des Projekts aus der Sicht eines DMPs

Im Allgemeinen kann die Planung wie folgt unterteilt werden:

- **Idee** und Konzept: Projektfindungsphase
- **Projektantrag**, Projektbeschreibung und Projekteinreichung
- **Bewilligung** des Projekts
- **Pilotphase**: Beginn der Arbeit, Produktion und Ordnen der ersten Daten. Clustering der Daten
- **Pre-Ingest-Vorgänge**: Ordnen der Daten, Festlegung der Datenformate, Konzeption und eventuell Vergabe von Metadaten (deskriptive, technische, kontextuelle, Provenienzmetadaten, Langzeitmetadaten), eventuell Klassifizierungen, Übersetzungen, Klärung von rechtlichen Fragen, Monitoring usw.
- **Ingest-Vorgänge**: Tatsächlicher Einspeisungsvorgang in ein Repository. Bedienung der Schnittstellen. Kann als Einzel- oder Bulkupload durchgeführt werden, Push- oder Pull-Modalitäten. Anschließend Vergabe von Permanent Identifiers (PIDs). Einzelne technische Vorgänge, wie Monitoring, Checking, Validation usw.
- **Management von Daten im Repository**: Damit sind die Prozesse rund um die Archivierung der Daten gemeint. Hier ist auch schon der Beginn der Nachnutzung bzw. Mehrfachnutzung festgelegt, die Verknüpfung von digitalen Objekten, die Bildung von Kollektionen usw.
- **Extraktion der Daten und Dissemination**: Gemeint ist hier die Bedienung der Schnittstellen, die Weitergabe und das Teilen der Daten (Data Sharing). Dieser Punkt ist



von enormer Bedeutung, denn Fördergeber und Verleger verlangen häufig, dass die Ergebnisse der Forschungsprozesse und die dazu gehörigen Daten für die „Scientific Communities“ zur Verfügung gestellt werden.

- **Nachnutzung:** Die Nachnutzung von Daten betrifft folgende Bereiche: Nachnutzung für bestimmte Zielgruppen oder auch Maschinen, das Weiterreichen der Daten an andere Repositories, die Verknüpfung von Daten und die Visualisierung der Daten in unterschiedlichen Kontexten. Die Visualisierung kann durch eine zielgruppenorientierte Präsentation der Daten dazu dienen, Forschungsräume zu eröffnen und „Storytelling“ zu ermöglichen. Dadurch kann, z.B. in der forschungsorientierten Lehre, eine kontextbezogene Darstellung einer „Forschungsgeschichte“ betrieben werden, die sich, je nach Publikum, immer wieder neu artikuliert.

Jede dieser Phasen weist einzelne Prozesse auf, die mit Managementaufgaben verbunden sind und die die Manipulation von Daten erforderlich machen. Eine möglichst vorausschauende Beschreibung eben dieser Aktivitäten ist Kernbestandteil von DMP.

□ Prozesse hinter DMP

■ Pre-Ingest

Idee, Projektbeschreibung, Ordnen der Daten, Konzeption, Vergabe von Metadaten, Klassifizierungen, Übersetzungen, Klärung rechtlicher Fragen, Monitoring

■ Ingest

Upload der Daten, Bedienung der Schnittstellen. Einzel- oder Bulkupload, Push- oder Pull-Modalitäten, Vergabe von Permanent Identifiers (PIDs). Monitoring, Checking, Validierung usw.

■ Management

Prozesse rund um die Archivierung der Daten. Beginn der Nachnutzung bzw. Mehrfachnutzung, Verknüpfung von digitalen Objekten, Bildung von Kollektionen usw.

■ Dissemination

Bedienung der Schnittstellen, Weitergabe der Daten. Teilen der Daten in den „Scientific Communities“

■ Nachnutzung

Für bestimmte Zielgruppen oder Maschinen, Datenaustausch mit anderen Repositories, Verknüpfung von Daten, Visualisierung der Daten in unterschiedlichen Kontexten, „Storytelling“

Ein ebenso wichtiger Aspekt ist die Bedeutung eines DMPs als solcher, nämlich dann, wenn er zur Grundlage eines Nachfolgeprojekts wird. Die Qualität eines DMPs kann ausschlaggebend für die Bewilligung weiterer Projekte werden, z.B. bei internationalen Vorhaben.

□ Formen von DMP

Es gibt keine formalen Kriterien um einen DMP zu erzeugen. Manche Institutionen geben digitale Formen vor, andere arbeiten mit vorgefertigten Formularen. Für alle gilt jedoch:

- Formulieren Sie präzise
- Fixieren Sie Dinge, die Ihnen wichtig erscheinen, schriftlich, vermeiden Sie jedoch Überflüssiges
- Halten Sie Ihre Aussagen kurz und prägnant, vermeiden Sie Redundanz
- Verfassen Sie Beschreibungen stets zielgruppenorientiert
- Vermeiden Sie Copy-and-paste aus anderen Dokumenten
- Passen Sie Ihren DMP immer wieder den aktuellen Gegebenheiten an

□ Hauptbestandteile eines DMPs

Wir empfehlen, den DMP nach den oben genannten Phasen zu strukturieren, wobei folgende Aspekte besonders zu beachten sind:

- Formate und Art der Daten, Infos über angewendete Standards
- Angaben über bzw. Klärung von ethischen und rechtlichen Fragestellungen
- Angaben über den beabsichtigten Zugang zu den Daten, vorgesehene Modalitäten bzgl. Teilen und Nachnutzung der Daten
- Kurzzeit- und Langzeitarchivierung der Daten
- Ressourcen und deren Aufteilung
- Zeitplan
- Zuständigkeiten
- Handlungsplan

□ Grundsätze zum Umgang mit Forschungsdaten – Schwerpunktinitiative „Digitale Information“

Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und können unabhängig von ihrem ursprünglichen Erhebungszweck vielfach Grundlage weiterer Forschung sein. Dies gilt für die Aggregation von Daten aus unterschiedlichen Quellen zur gemeinsamen Nutzung. Die nachhaltige Sicherung von Forschungsdaten und deren Bereitstellung dienen daher nicht nur der Prüfung früherer Ergebnisse, sondern in hohem Maße auch der Erzielung künftiger Ergebnisse. Sie bildet eine strategische Aufgabe, zu der Wissenschaft, Politik und andere Teile der Gesellschaft gemeinsam beitragen müssen. Mit dem Ziel, die Qualität, Produktivität und Wettbewerbsfähigkeit der Wissenschaft zu fördern, wurden von der Allianz der deutschen Wissenschaftsorganisationen daher folgende Grundsätze für ein koordiniertes weiteres Vorgehen verabschiedet:

<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze.html>



Weiterführende Links:

(COAR 2012) Data Management Plans – How to Treat Digital Sources

The imminent future for repositories and their management , Paolo Budroni , Library and Archive Services, University of Vienna, Tomasz Miksa, Secure Business Austria, Andreas Rauber, Vienna University of Technology - https://www.coar-repositories.org/files/7_DMP_Vienna.pdf

EC guidelines on data management in Horizon 2020 - http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Open Data Access Policies and Strategies in the European Research Area and Beyond vom Science Metrix - http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf

National Science Foundation (NSF) in US, DataRes Project - <http://datamanagement.unt.edu/>

Australian National Data Service (ANDS) in Australia - <http://ands.org.au/guides/data-management-planning-awareness.html>

<http://ands.org.au/resource/data-management-planning.html>

<http://ands.org.au/datamanagement/guide-to-filing-in-the-data-management-section-for-the-arc-march2014.pdf>

Digital Curation Centre (DCC) in Great Britain - <http://www.dcc.ac.uk/resources/data-management-plans>

Digital Curation Centre: How to Develop a Data Management and Sharing Plan - <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>

Digital Curation Centre: Checkliste für Data Management Pläne - http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf

Academy of Finland research plan requirements - <http://www.aka.fi/en-GB/A/Funding-and-guidance/How-to-apply/Appendices/Research-plan>

Deutsche Forschungsgemeinschaft (2009) Recommendations for Secure Storage and Availability of Digital Primary Research Data - www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901_en.pdf

University of Michigan - <http://www.lib.umich.edu/research-data-services/nsf-data-management-plans>

Requirements for Projects that Develop Digital Content - http://www.ims.gov/applicants/projects_that_develop_digital_content.aspx

Data Management Plan Templates - <http://scholcomm.columbia.edu/data-management/data-management-plan-templates/>

Process Management Plans

Tomasz Miksa, Stephan Strodl and Andreas Rauber, Process Management Plans. International Journal of Digital Curation, Vol 9, No 1 (2014), pp. 83-97. DOI:10.2218/ijdc.v9i1.30 - <http://www.ijdc.net/index.php/ijdc/article/view/9.1.83>

Tomasz Miksa and Andreas Rauber. Increasing preservability of research by process management plans, in Proceedings of the 1st International Workshop on Digital Preservation of Research Methods and Artefacts (DPRMA '13). 2013, ACM, New York, NY, USA, 20-20. DOI:10.1145/2499583.2499591-
<http://dl.acm.org/citation.cfm?id=2499591>



Ein Use Case an der Universität Wien in Zusammenarbeit mit 6 Fakultäten im Rahmen des Projekts e-Infrastructures Austria

Mustertext, der an jeder Universität angepasst wird.

Die Universität Wien verfügt seit 2008 mit Phaidra (<https://phaidra.univie.ac.at>) über ein Repository, in dem Ihre wertvollen digitalen Bestände dauerhaft gesichert und nachnutzbar gemacht werden.

Das Phaidra-Team verfügt über das notwendige Know-how für die Durchführung von komplexen Projekten sowohl im technischen als auch im nicht-technischen Bereich. Phaidra entstand aus einer Zusammenarbeit zwischen Experten aus der größten wissenschaftlichen Bibliothek des Landes, der Universitätsbibliothek Wien, und dem ZID (<http://zid.univie.ac.at/>) der Universität Wien.

Wenn Sie ein Projekt mit Phaidra durchführen möchten, würden wir Sie bitten, zunächst mit uns Kontakt aufzunehmen (phaidra@univie.ac.at). Auf der Serviceseite von Phaidra finden Sie erste Informationen über Phaidra: <http://phaidraservice.univie.ac.at>

Ein Data Management Plan dient dazu, allen Beteiligten die Arbeit zu erleichtern: Ihnen, um so früh wie möglich zu erfahren, welche Informationen die Betreiber des Repository von Ihnen benötigen, den Betreibern, um Ressourcen bestmöglich zu strukturieren.

Halten Sie den DMP stets aktuell. Ein Projekt ist zumeist nicht statisch und wird sich in seiner Laufzeit auch immer wieder verändern, daher ist es wichtig, den DMP immer wieder anzupassen und kritisch zu hinterfragen.

Bitte ergänzen Sie Hinweise, Kommentare usw. wo es notwendig erscheint. Es wird eventuell nicht jeder Punkt unserer folgenden Checkliste für Sie relevant sein, bitte antworten Sie auf unsere Fragen, soweit es Ihnen möglich ist. Fragen Ihrerseits beantworten wir natürlich jederzeit gerne.

Bevor Sie beginnen, möchten wir Sie noch auf die Policy von Phaidra und auf die Nutzungsbedingungen hinweisen:

<http://phaidraservice.univie.ac.at/das-system-phaidra/policy>

<http://phaidraservice.univie.ac.at/juristisches/nutzungsbedingungen>

Die Universität Wien unterstützt als Unterzeichnerin der Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen die Forderung nach freiem und uneingeschränktem Zugang zu wissenschaftlicher Information im Internet (Open Access). Mehr über die Open Access Policy der Universität Wien:

<http://openaccess.univie.ac.at/policy>



Checkliste

Allgemeines

- ▶ Titel des Projekts
- ▶ Autor/in des vorliegenden Dokuments
- ▶ Version und Datum
- ▶ Projektleitung an der Universität Wien (Bitte Namen, Tel.-Nr. und E-Mail-Adresse angeben)
- ▶ Projektverantwortung (Bitte Namen, Tel.-Nr. und E-Mail-Adresse angeben)
- ▶ Projektauftraggeber/in (Bitte Namen, Tel.-Nr. und E-Mail-Adresse angeben)
- ▶ Technische/r Ansprechpartner/in (Name, Funktion, E-Mail-Adresse)
- ▶ Kontaktperson für die Kommunikation mit dem Repository (Bitte Namen, Tel.-Nr. und E-Mail-Adresse angeben)
- ▶ Projektteam
- ▶ Namen
- ▶ Aufgaben während des Projektverlaufs

Kooperationspartner

- ▶ Universitätsintern
- ▶ Universitätsextern (*Weitere wissenschaftliche Einrichtungen? Firmen?*)

Projektbeschreibung

- ▶ Kurzbeschreibung (max. 10 Zeilen, Kurzzusammenfassung für Laien, Entstehung, Disziplinen, Ziele, Datenbestand ...)
- ▶ Projektbeginn und Projektende (Bitte möglichst genau angeben)
- ▶ Detailbeschreibung des Projekts (Bezüglich der digitalen Bestände, die dauerhaft gesichert werden sollten)
- ▶ Materialauswahl (Haben Vorarbeiten stattgefunden? Wer hat diese durchgeführt? Wann erfolgten diese Arbeiten? Wo liegen die Daten? Werden schon vorhandene Daten genutzt? Wie sieht der Zugang aus? Sollten sämtliche Daten langzeitarchiviert werden?)
- ▶ Liegen die Daten bereits vor (wenn ja, in welchem Verhältnis stehen diese Daten zu neuen Daten? Detailreicher, größere Abdeckung, ...) oder werden sie erst im Laufe des Projekts erzeugt?
- ▶ In welchen Formaten liegen die Daten vor? (Offene Formate werden bevorzugt)
- ▶ Art der Daten (Um welche Daten handelt es sich? Rohdaten, bzw. Texte, Videos, Bilder?)
- ▶ Datenproduktion (Beschreiben Sie den Prozess, wie die Daten produziert werden. Wie werden die Daten produziert? Welche Soft- und Hardware wird verwendet? Welche Formate werden produziert?)
- ▶ OCR: Wurde ein OCR-Scan durchgeführt? Ist ein OCR-Scan erwünscht?
- ▶ Umfang der Daten (Bitte geben Sie das ungefähre Volumen Ihrer Daten an, die im Repository archiviert werden sollten)



- ▶ Wie groß sind die größten Daten?
- ▶ Ist eine Datenmigration nötig, d.h. sollen die Daten aus einem „Altsystem“ in ein anderes System übernommen werden?
- ▶ Sind die Daten bereits katalogisiert? (Bitte geben Sie dazu Hinweise, bzw. ist eine Katalogisierung erwünscht?)
- ▶ Qualitätskontrolle der Objekte (An welcher Stelle des Prozesses wird die Kontrolle durchgeführt? Wer ist dafür verantwortlich? Wer dokumentiert?)

Speicherung der Daten

Short-Term-Storage

- ▶ Ort und Art des Speicherplatzes (Wo werden die Daten gespeichert, bevor sie ins Repository kommen? Ist genügend Speicherplatz vorhanden? Sind die Daten bis zum Upload ins Repository ausreichend gesichert? Werden Back-ups durchgeführt? Finanzierung?)
- ▶ Zugänglichkeit (Wer hat Zugang zu den Daten? Gibt es einen Passwort-Schutz? Wie sind die Files benannt?)
- ▶ Zuständigkeit (Name, Tel.-Nr. und E-Mail-Adresse? Gibt es dafür bereits eine Finanzierung?)
- ▶ Datenimport ins Repository.
- ▶ Zuständigkeit (Name, Tel.-Nr. und E-Mail-Adresse? Gibt es dafür bereits eine Finanzierung?) Wie könnte dieser Weg technisch aussehen? (Bitte beschreiben Sie, welche technischen Voraussetzungen hier vorhanden sind)

Rechtliche Fragen

- ▶ Urheberrecht, Verwertungsrecht: Ist die Rechtslage geklärt? Darf der/die Projektverantwortliche die Objekte ins Repository hochladen? Dürfen diese Objekte im Internet angezeigt werden?
- ▶ Verantwortung für Rechtsfragen auf der Projektseite: Wer trägt die Verantwortung für die Ergebnisse dieser Arbeit (Offenlegung im Internet)?
- ▶ Bei Projekten mit mehreren Partnern: Ist die rechtliche Situation untereinander geklärt?
- ▶ Lizenzen: Welche Lizenz wird vergeben? Link? Gibt es Embargozeiten?
- ▶ Zugriffsrechte: Müssen Ihre Daten zum Beispiel „Open Access“ sein?
- ▶ Wiederverwendung der Daten: Ist die Wahl der Technik, der Formate und der Metadaten geeignet, um eine Wiederverwendbarkeit zu garantieren? Ist ethisch bedenkliches Material vorhanden? Sind personenbezogene Daten enthalten? Haben Sie bei audiovisuellen Materialien die Zustimmung zur Veröffentlichung von den vertonten bzw. abgebildeten Personen?

Metadaten

- ▶ Sind bereits Metadaten vorhanden? (Wenn ja, in welcher Form liegen diese vor?)
- ▶ Wo liegen die Metadaten?
- ▶ Wie sieht es mit der Zugänglichkeit zu den Metadaten aus?
- ▶ Welche Metadatenstandards werden verwendet? (z.B. Dublin Core, MODS, ...)



- ▶ Sind die Rechte an bereits bestehenden Metadaten vorhanden?
- ▶ Wer ist verantwortlich für die Metadaten? (Wer gibt vor, welche Metadatenfelder wie ausgefüllt werden? Wer füllt die Vorlagen im Repository aus?)
- ▶ Welche Metadatenfelder werden verwendet? (Technische, deskriptive, kontextuelle Metadaten, Metadaten für Langzeitarchivierung, Provenienzmetadaten?)
- ▶ Qualitätskontrolle der Metadaten (Wer ist dafür zuständig? An welcher Stelle des Prozesses wird die Qualitätskontrolle durchgeführt?)

Kosten für die Langzeitarchivierung

- ▶ Personalkosten (Gibt es Personen, die die Objekte langfristig pflegen können?)
- ▶ Sachkosten (Speicherplatz, Kosten für Short-Term-Preservation)

Dokumentation und Dissemination

- ▶ Wer dokumentiert Projektergebnisse?
- ▶ Wo werden Projektergebnisse voraussichtlich publiziert? (Journals, Websites, Forschungsberichte ...)

Visualisierung der Objekte

- ▶ Visualisierung im Repository (Bookviewer, Collections, Container, Vorder- und Rückseite von Bildern, Verbindungen, ...)
- ▶ Visualisierung außerhalb des Repositories (diese muss von Projektseite erfolgen – Finanzierung, technisches Know-how, Software usw.)
- ▶ Weiternutzung der Daten nach Projektabschluss
- ▶ Gibt es Vereinbarungen zwischen den Partnern? (z.B. aus rechtlicher Sicht)
- ▶ Ist ein Folgeprojekt oder eine Weiterfinanzierung geplant?
- ▶ Wer ist dafür zuständig? (z.B. Pflege der Metadaten?)

Projektplan

Erstellen Sie eine Zeitleiste, um die Projektschritte von den Vorarbeiten bis hin zur eventuell angestrebten Nachnutzung der Daten abzubilden. Dabei sollten die jeweiligen Meilensteine, die Workpackages und die Verantwortlichkeiten eingetragen werden. Wichtig ist auch festzulegen, wer wann mit wem kommuniziert.

Geben Sie bitte genau an, an welchen Stellen Ihrer Meinung nach das Team eingebunden werden muss. Verwenden Sie für diese Darstellung eventuell ein frei verfügbares Tool.

Unterstützung finden Sie u.a. in:

http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_368.pdf

Wir empfehlen, die einzelnen Schritte der Langzeitarchivierung zu überlegen und schriftlich festzuhalten:

- ▶ **Pilotphase** (Überprüfen der Objektauswahl, der Tools, Formate, Visualisierung usw.)
- ▶ **Pre-Ingest-Phase** (Welche Arbeiten müssen vor dem eigentlichen Upload ins Repository erledigt sein? Wer ist dafür zuständig? Welche Informationen werden benötigt?)
- ▶ **Ingest-Phase** (Mit dem Team besprechen, wie der Ingest der Objekte und der Metadaten erfolgen soll. Wie kann das Übergabeszenario aussehen?)
- ▶ **Controlling** (An welchen Stellen muss Controlling durchgeführt werden?)
- ▶ **Feedback** (Wer gibt wann an wen Feedback?)
- ▶ **Dissemination** (Welche Formen der Dissemination sind geplant? Wer ist dafür verantwortlich?)

Sie finden laufend weitere Infos online unter

<http://e-infrastructures.at/das-projekt/work-package-cluster/cluster-c/>

