Data Management Planning: where are we and where do we want to be?

Joint session of Active Data Management Plans IG and DMP Common Standards WG and Discipline-specific Guidance for Data Management Plans WG

20th RDA Plenary
Agenda

➤ Part 1 – Getting to know each other and introduction for newcomers

➤ Part 2 – Updates and topics for discussion
  ➤ Software Management Plans
  ➤ Maintenance of maDMP specification
  ➤ Santosh
  ➤ Cross-fertilization workshop results

➤ Part 3 – Discussion in groups and reporting

➤ Part 4 – Wrap up
Participants

▷ Let’s get to know each other
   ▷ Go to slido.com

#maDMPs

#activeDMPs
Where are you from?

- France
- Norway
- Netherlands
- United States
- Sweden
- UK
- Japan
- Nancy
- Italy
- Costa Rica
- Eskilstuna
- Trondheim
- Canada
- Ottawa
- Oslo
- Belgium
- INDONESIA
- Finland
- Germany
- Colombia
- Austria
- England
- Amsterdam
- Poland
- Berlin
- Tromsø
- Bergen
- Dublin
- Serbia
- Belgium
- Austria

Join at slido.com #4225 137
Are you a member of the WG and/or the IG?

- None: 65%
- Active DMPs IG: 15%
- DMP Common Standards WG: 13%
- Don't remember: 13%
- Domain Specific Guidance for DMPs WG: 11%
How familiar are you with maDMPs?

- Briefly following developments 55%
- Considering adoption 21%
- Never heard of! 12%
- Hands-on experience in using maDMPs 12%

Join at slido.com #4225137
Why joint session?

Active DMPs *Interest Group*
- Place for discussion on all topics related to DMPs
- Can trigger new WGs
  - Like it happened in the past with the DMP Common Standards WG

DMP Common Standards *Working Group*
- Maintenance mode
  - Updates the recommendation when necessary
  - Supports adopters of the recommendation

Domain-Specific Guidance for DMPs *Working Group*
- Ends in May
  - Paper + Jupyter Notebook
Objective for today

Identify new hot topics relevant to this community
Introduction for newcomers on maDMPS

Part 1
# Data Management Plans (DMPs)

<table>
<thead>
<tr>
<th>Data Officer</th>
<th>Who is responsible for the data management and the DMP of the project (permanent address)?</th>
</tr>
</thead>
</table>

## 1. Data Description

| 1.1 Description of the data | What kind of data/source code will be generated or needed (data, format, volume)? How will the research data be generated and when will results be sold? How will the data be structured to ease handling and versioning? |

## 2. Documentation and Metadata

<table>
<thead>
<tr>
<th>2.1 Metadata standards</th>
<th>What metadata standards if any will be used and why? (See OpenCitations Coordination)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Documentation of data</td>
<td>How will information be recorded for the data to be findable, accessible, interoperable and re-usable (FAIR)? Is the data versionable? How are you planning to document this information?</td>
</tr>
</tbody>
</table>

## 3. Data quality control

| 3.1 Data quality control | How will the consistency and quality of data collection be controlled and documented? (This may include processes such as regular audits or measurements, standardized data capture, peer review of data or representation via controlled vocabularies.) |

## 4. Data availability and storage

<table>
<thead>
<tr>
<th>4.1 Data sharing strategy</th>
<th>How and where will the data be shared and made accessible? Other researchers will be interested?</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 Data storage strategy</td>
<td>What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and cleaned up during the lifetime of the project? Are there any restrictions for sharing or access? Will the data be stored in a central or distributed manner? How will the data be stored after the project ends? Are there any technical barriers to making the research data fully or partially accessible?</td>
</tr>
</tbody>
</table>

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreward by Dr Thierry Damerval</td>
<td>2</td>
</tr>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>GUIDANCE FOR ORGANISATIONS: Core Requirements for Data Management Plans</td>
<td>7</td>
</tr>
<tr>
<td>GUIDANCE FOR ORGANISATIONS: Criteria for the Selection of Trustworthy Repositories</td>
<td>11</td>
</tr>
<tr>
<td>GUIDANCE FOR RESEARCHERS: Evaluating the Core Requirements of a DMP Template: Guiding the Selection of Trustworthy Repositories</td>
<td>15</td>
</tr>
<tr>
<td>GUIDANCE FOR REVIEWERS: Evaluation Rubrics for Data Management Plans</td>
<td>31</td>
</tr>
<tr>
<td>Notes and References</td>
<td>51</td>
</tr>
<tr>
<td>Annex: Compatibility with the FAIR Data Principles</td>
<td>52</td>
</tr>
</tbody>
</table>

---

4 For procedural elements of implementing DMPs, see the RDA DMP Common Standards Working Group: [https://www.rda-alliance.org/groups/dmp-common-standards-wg](https://www.rda-alliance.org/groups/dmp-common-standards-wg)
Machine-actionable DMPs (maDMPs)

- Machine-actionable DMPs
  - Living documents
  - automate data management
    - collect information from systems
    - trigger actions in systems
  - facilitate validation
- This requires
  - well-defined RDM workflows
  - data management infrastructure
  - common standard to represent information
Official RDA Recommendation on maDMPs

RDA DMP Common Standard for Machine-actionable Data Management Plans

The Challenge:
Data Management Plans are free-form text documents describing the data that is used and produced during the course of research activities. They specify where the data will be archived, which licenses and constraints apply, and to whom credit should be given, etc. The workload and bureaucracy often associated with traditional DMPs can be reduced when they become machine-actionable.

Produced by: DMP Common Standards WG
https://www.rd-alliance.org/groups/dmp-common-standards-wg

RDA DMP Common Standard for Machine-actionable Data Management Plans

maDMPs - documentation

## Properties in 'dmp'

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Data Type</th>
<th>Cardinality</th>
<th>Example Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>contact</td>
<td>Contact person for a DMP</td>
<td>Nested Data Structure</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>contributor</td>
<td>To list people that play a role in data management related to this DMP, e.g. responsible for performing actions described in this DMP.</td>
<td>Nested Data Structure</td>
<td>0..n</td>
<td></td>
</tr>
<tr>
<td>cost</td>
<td>To list costs related to data management. Providing multiple instances of a ‘Cost’ allows to break down costs into details. Providing one ‘Cost’ instance allows to provide one aggregated sum.</td>
<td>Nested Data Structure</td>
<td>0..n</td>
<td></td>
</tr>
<tr>
<td>created</td>
<td>Date and time of the first version of a DMP. Must not be changed in subsequent DMPs.</td>
<td>DateTime</td>
<td>1</td>
<td>2019-03-13 13:13</td>
</tr>
<tr>
<td>dataset</td>
<td>To describe data on a non-technical level.</td>
<td>Nested Data Structure</td>
<td>1..n</td>
<td></td>
</tr>
</tbody>
</table>

NOT a questionnaire! NOT a template!

Most fields are optional!

Machine-actionable DMP

Example: https://doi.org/10.5281/zenodo.6467730

"contributor" : [ {
  "contributor_id" : {
    "identifier" : "0000-0002-5164-2690",
    "type" : "orcid"
  },
  "mbox" : "moritz.staudinger@tuwien.ac.at",
  "name" : "Moritz Staudinger",
  "role" : [ "Data Manager" ]
}

maDMPs use PIDs and controlled vocabularies.

Example shows that Moritz is the one responsible for data management.
Each dataset has a title and a human readable description. It is also clear what the format, size and the location of the dataset are. License and mode of access, including any exact embargo periods, are specified as well.
maDMPs are the ‘glue’ between different systems

Automate getting information in and out
Adoptions (selected)
Slides from all our sessions are in the repository

Today’s presentations will also be there

https://www.rd-alliance.org/node/56938/file-repository
Read more in...

- Describes the full story of developing the recommendation
- Example of a minimal maDMP
- Presents adoptions
  - Haplo
  - Open Research Publishing Platforms
  - DMP Tool
  - DMPonline
  - DMP OPIDoR
  - Data Stewardship Wizard
  - NSD DMP
  - Argos
  - Research infrastructure at TU Wien
  - Easy DMP

http://doi.org/10.5334/dsj-2021-032
Read more in...

ACM Transactions on Management Information Systems
https://doi.org/10.1145/3490396

Practice Paper
Interconnecting systems using machine-actionable Data Management Plans - hackathon report

Data Science Journal
http://doi.org/10.5334/dsj-2021-035

Contains:
- Enterprise Architecture that uses maDMPs
- Examples of tasks automation at institutions using maDMPs

Contains:
- Summary of results from hackathon
Introduction for newcomers on domain-specific guidance for DMPs

Part 1
Discipline-specific Aspects for DMP WG

Daniela Hausen  Ivonne Anders, Santosh Ilamparuthi, Yasemin Türkyilmaz-van der Velden, Shannon Sheridon, Briana Wham
Motivation

• Researchers should
  • Think about RDM in advance
  • Create a plan
  • Identify gaps

• Researchers want to
  • Have clear and unique concepts
  • Have direct support by structure, helping texts and examples
  • Support from the discipline

Create Guidance and Examples
adopt Terminology

source: http://doi.org/10.5281/zenodo.1120245
RDA Online Survey on discipline-specific aspects for DMP templates

- 358 participants
- 21 questions:
  - (1) Demographics,
  - (2) Data Description,
  - (3) Data Documentation & Quality,
  - (4) Data Archiving, Publishing & Sharing After the Project,
  - (5) Guidelines, Principles, & Best Practices
Results from the Online Survey

Disciplines of the participants

- Natural Sciences
- Engineering Sciences
- Life Sciences
- Humanities and Social Sciences

Position of the participants

- Professor: 11%
- PostDoc/Senior Researcher: 25%
- Promovierende: 14%
- Technische Angestellte: 5%
- Studierende: 2%
- Andere: 5%
- keine Angabe: 3%
- Forschungsunterstützung: 35%
Results from the Online Survey

Issues/areas with discipline-specific differences:
• Metadata and metadata standards
• Data types
• Data generation and data collection
• Post-use of data types
• Data documentation
• Quality control
• Framework conditions (legal, contractual and ethical)
• Types of data publication

Issues/areas with large overlaps:
• Discipline-specific guidelines
• Implementation of FAIR principles
• Data storage incl. storage media
• Use of naming conventions

Issues/areas without discipline-specific focus:
• Allocation of persistent identifiers
• Allocation of licences
• Use of learning materials

Prototype

Introduction

Welcome to the Discipline Specific Guidance for data management plans (DMPs) compendium. This was created as part of the output of the RDA working group of the same name. While guidelines for some disciplines were created during the period of the working group the goal here is to create a collection through the contributions of researchers and research support staff belonging to different disciplines.

What is this book about?

In this book you will find the template used for gathering the guidelines for different disciplines along with the already curated guidelines for some disciplines. The disciplines are broadly categorized into four groups,

- Life Sciences
- Natural Sciences
- Humanities and Social Sciences
- Engineering Sciences

The goal is that, over time, the different disciplines would have very specific guidelines on

https://santoshilam.github.io/Discipline_Specific_Guidance_for_DMPs/intro.html
https://github.com/santoshilam/Discipline_Specific_Guidance_for_DMPs
Updates and topics for discussion

Part 2
“Jupyter Notebook” live demo - Santosh
Software Management Plans - Kerstin (virtual)
Maintenance of the maDMP specification - Marek
Cross-fertilization workshop results - Elli
FAIRification of software development – (machine-actionable) software management plans

Discussion at RDA Plenary 20
suggested by Kerstin Helbig (Humboldt-Universität zu Berlin, Germany)
What is a software management plan?

Definition by DINI/nestor Working Group Research Data:
A software management plan (SMP) includes general and technical information about the software project, information about quality assurance, release and public availability, as well as legal and ethical aspects affecting the software.

The SMP summarizes information that sufficiently describes and documents the creation, documentation, storage, versioning, licensing, archiving and/or publication of the software generated or used in a project. Related hardware and necessary other resources, as well as related other software and software libraries, text and data publications, must also be described and are a feature of the SMP.

The purpose of an SMP is first of all to support the traceability and, if necessary, the long-term usability of the software (for direct application as well as for further processing) and to facilitate the support of the users in case of queries. The SMP therefore also serves the purpose of quality assurance (cf. FAIR4RS Principles).

The SMP can be linked to one or more data management plans (DMP) if the software is used for data generation or processing. SMP and DMP can be combined as output plans (cf. Software Sustainability Institute).

Source: https://forschungsdaten.info/praxis-kompakt/english-pages/glossary/#c499503
Funder requirements and integration of SMP in DMP tools

• Most funders treat research software as data -> no specialized requirements
• SMP is not a standard tool for software management planning
• With few exceptions, DMP tools do not offer up-to-date, comprehensive templates for software
• Machine actionable integration of SMP (maSMP) into GitHub could support FAIR4RS
Points for discussion

• Do we need SMPs or should we work towards more generalized output plans that combine data and software management?
• How can FAIRification of software be supported via SMPs?
• To what extent can maSMP help with this?
• ...
Maintenance of the maDMP specification

Marek Suchánek

0000-0001-7525-9218

marek.suchanek@fit.cvut.cz
Current State and Issues

- GitHub repository, Google Spreadsheet, GitHub Pages (README-like)
- Not a single source of truth (figures, JSON schema, spreadsheet, generated README)
  - Harder to contribute (and adopt)
  - Possible inconsistencies
  - Issues with versioning
- No community content (adoption stories/experience)
- No automation (except the one with Google Spreadsheet)
Current State and Issues

- RDF/OWL + examples
- JSON Schema + examples
- Diagram
- Entities + links
- README
- GitHub Pages

rd-alliance.org  @resdatall | @rda_europe | @RDA_US
Desired State

- RDF/OWL + examples
- JSON Schema + examples
- Diagrams
- Experience
- Standard specification
- Releases with DOI
- GitHub Pages
- README
- CODEOWNERS
- CONTRIBUTING
- Documentation etc.
Work Plan

EOSC Future support through Optimising Frameworks and Guidelines in EOSC #2 grant

Tasks and partial goals in the project:
- Conduct user/adopter research, gather feedback, document issues
- Develop tools and other resources supporting use and adoption as well as maintainence of the standard
- Foster contributions and community development by providing proper way of crediting contributors and contributing guidelines
- Enhance standard documentation (website, implementation recipes, sharing experience, clarify versions of standards, extensions, etc.)
Part 2

- Software Management Plans - Kerstin (virtual)
- Maintenance of the maDMP specification - Marek
- “Jupyter Notebook” - Santosh
- Cross-fertilization workshop results - Elli
RDA for Data Management Planning Community Cross-fertilisation Workshop Summary

Community cross-fertilisation workshop: RDA for Data Management Planning

RDA for Data Management Planning

Celebrating A Decade of Data
RDA community cross-fertilisation workshop

Version December 2022

PARTICIPATING GROUPS & WORKSHOP LEADS

- DAMP Common Standards Wiki
  - Cross-community DAMP Report
  - Data Management Forum & Wiki

See community group card

Active Data Management Plans (ADMP) Forum
- DAMP Cross-Community Wiki

See community group card

DAMP-specific guidance for Data Management Plans (DMP)

See community group card

Guye Odgers, School of Business, Economics and Information Technology, The University of Queensland, Australia

W.: workshop@rda-alliance.org

Workshop topics: collect evidence of best practices and work to use them to develop best practice data management plans, and help support and standardise data management plans across the community.

W.: DAMP@rda-alliance.org

RDA Cross-Community Wiki

W.: wiki@rda-alliance.org

RDA for Data Management Planning

RDA website: https://rda-alliance.org

INITIATIVES & RESOURCES OF INTEREST

- ISMMP (Integrating & Supporting Metadata & Management)
  - Data Management Project (DMP) - Call for Proposals
  - DataLink - Data Management Framework

See community group card

WORKSHOP PARTICIPANTS

1. Antonio Sarratea, National Research Council, CNR - IREA, Italy
2. Angela Waith, Penn State University Libraries, USA
3. Daniel Mann, Forschungszentrum Jülich, Germany
4. Donald Haxel, WUSTL, Washington University, USA
5. Eric Ely, FIMBIO Ltd and CSIRO MEROI, Darwin, Australia
6. Erla Blood, Zuercher Bibliothek, Switzerland
7. Désiré Papotti, CONCEA, University of Antwerp, Belgium
8. Fredrik Ekblom, University of Gothenburg, Sweden
9. Francesco Genoves, Università degli Studi, Firenze, Italy
10. Helga Gjeng, HiNorBut, Jülich, Germany
11. Jennifer Le, University of Wisconsin Madison, USA
12. Jonathan West, University of Oxford, UK
13. Jouko Pekkanen, University of Helsinki, Finland
14. Josephine Winkler, DFG Germany, Germany
15. Jürgen Schulz, Technische Universität Berlin, Germany
16. Judith Dore, Eurekadata, USA
17. Lindsey Anderson, Pacific Research Laboratory, USA
18. Madye Mckean UIC, USA
19. Mary McMillan, Colorado State University, USA
20. Michael Serpa, Oak Ridge National Laboratory, USA
22. Michelle Edwards, Agri-Data Strategy - University of Guelph, Canada
23. Mira van de Wijgerde, University of Amsterdam, The Netherlands
24. Mute Zhang, RDA-Latin America Group - University of Chile, Latin America
25. Paulette Liégeois, Société Nouvelles Catégories Thématiques, France
26. Rodericades, Netherlands Scientifc Data Service - Netherlands Institute for Advanced Research (NIWA), Netherlands
27. Sarah Gough, School of Library and Information Studies, University of British Columbia, Canada
28. Sarah L. Gabriel, National Library of Medicine, USA
29. Sarah Weirens, National Library of Medicine, USA
30. Sarah Weirens, National Library of Medicine, USA
31. Sarah Weirens, National Library of Medicine, USA
32. Sarah Weirens, National Library of Medicine, USA
33. Sarah Weirens, National Library of Medicine, USA
34. Sarah Weirens, National Library of Medicine, USA
35. Sarah Weirens, National Library of Medicine, USA
36. Sarah Weirens, National Library of Medicine, USA
37. Sarah Weirens, National Library of Medicine, USA
38. Sarah Weirens, National Library of Medicine, USA
39. Sarah Weirens, National Library of Medicine, USA
40. Sarah Weirens, National Library of Medicine, USA

To become a member of the RDA, register at https://rda-alliance.org/
CHALLENGES TO BE ADDRESSED WITHIN THE THEME OF DATA MANAGEMENT PLANNING

Raising awareness & understanding about DMPs:
- Lack of awareness that DMPs are integral to good research data management (RDM).
- Creating a DMP is often considered an extra task.
- Creating a DMP requires prerequisite knowledge and skills related to RDM, tools and services.
- Insufficient incentives, reward and credit for RDM and the creation of DMPs.

Creation, adoption & implementation of DMPs:
- Lack of clarity about roles and responsibilities for supporting DMP creation and limited staff.
- Lack of automated and machine-actionable processes and workflows to harmonise DMP creation and implementation across stakeholders, tools, services and policies.
- Unclear definitions and language (vocabularies and terminologies) make DMPs challenging to understand and complete.
- Diversity of data means generic DMPs are unsuitable for specific research disciplines.
- As primarily text documents and parts of research proposals, DMPs are not FAIR research objects that enable their utility.

Review, evaluation & assessment of DMPs:
- No clear process or assessment criteria for DMP review and evaluation.
- Lack of accountability on following through on the implementation of DMPs that support funded research.

COLLECTIONS FOR THE RDA COMMUNITY

Collect ‘gold star’ DMP case studies. Real-world examples of DMPs from different institutions, disciplines and projects to demonstrate benefits of DMPs for various stakeholders. Case studies and infographics showcase best practices for how to create DMPs and integrate them into the research data lifecycle. This output supplements the ‘Engaging Researchers with Data Management: The Cookbook’ by the RDA Engaging Researchers with Data IG.

Construction of DMP typologies. Define information, language and terminology to be included in DMPs that unambiguously describe RDM concepts and tasks. Typologies leverage existing ontologies and controlled vocabularies (e.g., T4FS, FIP ontology).

Create framework(s) for DMP evaluation and assessment. Collaborate with funding agencies (RDA Funders Forum and Research Funders and Stakeholders on Open Research and Data Management Policies and Practices IG) to create a scalable DMP evaluation framework that defines content to be assessed and evaluation criteria for DMPs dependent on institutional, disciplinary and project contexts.

Develop a toolkit for how to engage researchers in DMPs. DMP professionals, domain experts and RDM supporters co-create a step-by-step guide for how best to drive adoption and implementation of DMPs.
Discussion

Part 3
Wrap up

Part 4
Summary

➢ Bring topics that we could tackle together
➢ You can always
   ➢ Contact group chairs
   ➢ Write directly to the mailing list
➢ Further activities can be organised based on the needs
➢ We hope to see you all in person in Salzburg!
Please cite the recommendation:

Publications about maDMPs


- Tomasz Miksa, João Cardoso, José Luis Borbinha: Framing the scope of the common data model for machine-actionable Data Management Plans. BigData 2018: 2733-2742