



e-Infrastructures Austria

Data Management Plans

Fortbildungsseminar für Forschungsdaten und e-Infrastrukturen

Tomasz Miksa, Barbara Sánchez Solís, Cornelia Schrauf

Universität Wien, 8. Juni 2016

e-infrastructures
austria

Diese Unterlagen wurden im Rahmen der e-Infrastructures Austria Veranstaltung
Fortbildungsseminar für Forschungsdaten und e-Infrastrukturen
erstellt und stehen im Web unter folgender CC-by Lizenz zur Verfügung:



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

Seminar-Website: <http://e-seminar.univie.ac.at/>

Projekt-Website: <http://e-infrastructures.at/>

Agenda

- What is a Data Management Plan (DMP)?
- Where does it come from?
- What is inside of a DMP?
- What are the funder requirements?
- How to create a DMP?
- Who benefits and how from DMPs?
- How to establish data management practices?
- Lessons learned
 - University of Vienna
 - WWTF

What is a Data Management Plan?

What is a Data Management Plan?

*“... Data Management Plans (DMPs) detailing **what data** the project will generate, whether and how it will be **exploited or made accessible** for verification and re-use, and how it will be **curated and preserved.**”*



Guidelines on Data Management
in Horizon 2020
Version 1.0
11 December 2013



[European Commission, Guidelines on Data Management in Horizon 2020]

What is a Data Management Plan?

- DMP is a formal document
- It outlines what you will do with your data **during** and **after** you complete your research
- It ensures your data is safe for the **present** and the **future**

[from University of Virginia Library]

What is a Data Management Plan?

“Plans typically state **what data will be created and how**, and outline the **plans for sharing and preservation**, noting what is appropriate given **the nature of the data** and any **restrictions** that may need to be applied.”

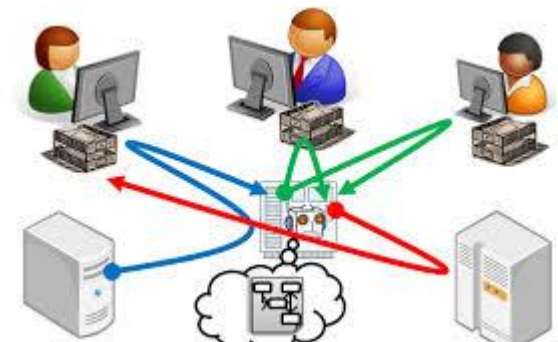
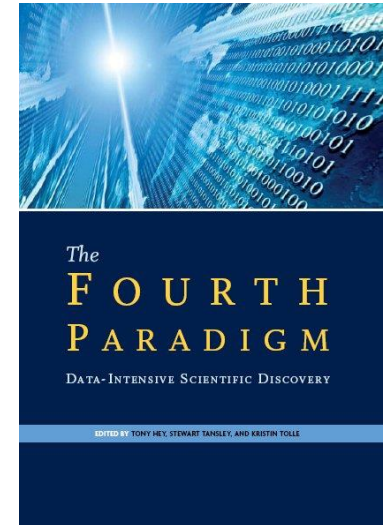


[Digital Curation Centre, GB]

Where does the idea come
from?

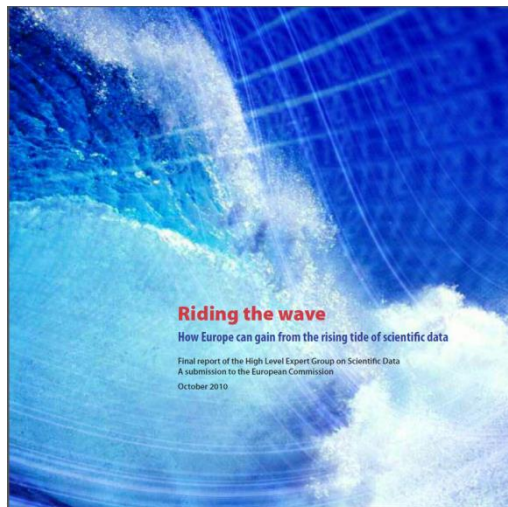
eScience

- data intensive science
- computational science
- Data
 - is fuel for research
 - is the result of processes such as
 - capturing
 - pre-processing
 - transformation
 - integration
 - analysis



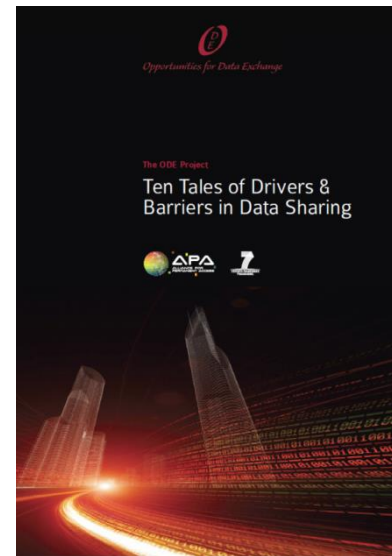
Importance of scientific data

- *Riding the wave: How Europe can gain from the rising tide of scientific data (2010)*
 - The report describes long term scenarios and associated challenges regarding **scientific data access, curation and preservation** as well as the strategy and actions necessary to realise the vision.



Success Stories

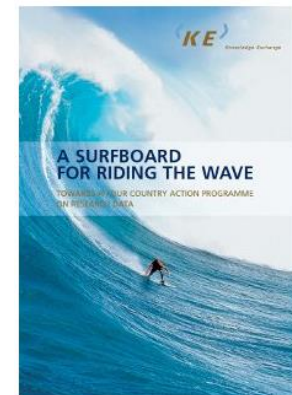
- *Ten Tales of Drivers & Barriers in Data Sharing (2011)*
- Collection of success stories and lessons learned in the area of
 - data sharing
 - re-use
 - preservation



Importance of exchanging data

- *A surfboard for riding the wave: Towards a four country action programme on research data (2012)*
 - overview of the situation with regard to research data in Denmark, Germany, the Netherlands and the United Kingdom
 - outlines a **possible action programme** for the four countries in realising the envisaged **collaborative data infrastructure**.

<http://www.knowledge-exchange.info/surfboard>



Awareness of data importance

“Because research in genomics, pharmacology or the fight against cancer increasingly depends on the availability and sophisticated analysis of large data sets. **Sharing such data means researchers can collaborate, compare, and creatively explore whole new realms.** We cannot afford for access to scientific knowledge to become a luxury, and **the results of publicly funded research in particular should be spread as widely as possible.**”



Neelie Kroes,
Vice President
European Commissioner
responsible for the Digital
Agenda

Variety of solutions

- In response to these needs many solutions were proposed and are being implemented
 - open access to scientific publications and data
 - research data repositories to host the data
 - data citation to reference the datasets
 - DATA MANAGEMENT PLANS



DMPs are obligatory

- Required by funding institutions
 - National Science Foundation in US
 - European Commission for Horizon 2020 proposals
 - Australian Universities (Australian National Data Service)
 - Research Councils in GB
 - ...
- Required by some universities and their libraries
- Required by some publishers
- Current state in Austria – next slides

What is inside of a DMP?

What is inside of a DMP?

- It depends...
 - on the institution requiring a DMP
 - field of research
- DMP is
 - usually a written document
 - usually has an enforced structure
- DMP creation facilitated by
 - questionnaires, guidance documents, checklists, etc.
- There are common parts and overlaps between various templates
- Level of details varies

What are the funder requirements?

FWF – The Austrian Science Fund

FFG – Austrian Research Promotion Agency

European Commission - Horizon 2020

FWF Forschungsdaten-Policy

- Absatz zu Forschungsdaten in der Open Access-Policy:

<http://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/>

FWF

Der Wissenschaftsfonds.

IV. Open Access bei Forschungsdaten

Forschungsdaten und ähnliche Materialien, die im Rahmen von FWF geförderten Projekten erhoben und/oder analysiert werden, sollen sofern dies rechtlich und ethisch möglich ist, Open Access zur Verfügung gestellt werden. Entweder sollten die Daten, die publizierten Forschungsergebnissen zugrunde liegen, zum Zeitpunkt der Publikation frei zur Verfügung gestellt werden. Oder wenn die Daten nicht in Publikationen verwendet wurden, sollten sie spätestens zwei Jahre nach Projektende frei zugänglich gemacht werden. Die folgenden Bedingungen müssen erfüllt werden:

- Ein geeignetes Repository muss gewählt werden, siehe [hier](#)
- Die Forschungsdaten müssen in einem zitierbaren Format vorliegen, siehe [hier](#)
- Die Forschungsdaten müssen für eine uneingeschränkte Wiederverwendung archiviert werden, siehe [hier](#)

FWF Open Research Data Pilot

FWF

Der Wissenschaftsfonds.

- Gültig für FWF-Projekte der letzten fünf Jahre
- Antrag: Letter of Interest bis 03/2016 - Einreichung des Vollantrags bis 07/2016 - Finale Förderentscheidungen bis Anfang 03/2017
- Kein verpflichtender DMP, aber Anforderungen:

„Die Forschungsdaten müssen (1) nach den neuesten technischen Standards, (2) frei zugänglich (Open Access), (3) reproduzierbar, (4) maschinenlesbar, (5) zitierfähig sowie unter (6) Verwendung möglichst offener Lizenzen zur uneingeschränkten Weiterverwendung (7) in registrierten Repositorien publiziert werden“

[<https://www.fwf.ac.at/de/forschungsfoerderung/fwf-programme/open-research-data/>]

FFG - Status und Ausblick

- Derzeit keine Forschungsdaten-Policy
- **Policy** zu Open Access und **Open Data** bis Ende 2016 geplant!
- FFG ist **National Contact Point** für **Horizon 2020**
- Informationen und Schwerpunkt-Veranstaltungen zu Open Research Data in H2020 unter: <https://www.ffg.at/europa/ncp>



Europe: EC Horizon 2020

- Alle wissenschaftlichen Publikationen mit **Peer Review** müssen in H2020 **Open Access** veröffentlicht werden – Verpflichtung!
- Open Research Data (ORD) Pilot: Verpflichtende Veröffentlichung von Forschungsdaten **für ausgewählte Programmbereiche** („Work Programme“)
- Opt-in/opt-out möglich – kein Einfluss auf Evaluierung
- Datenmanagementkosten sind **förderfähig**

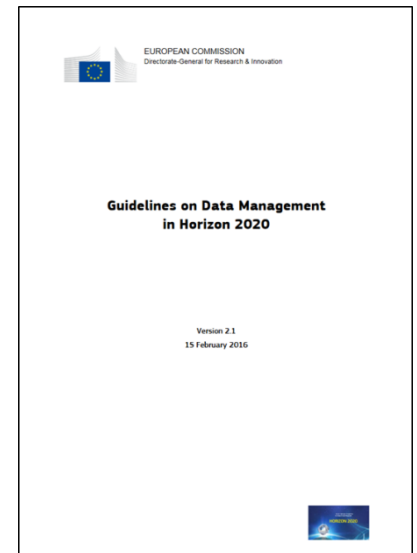
[http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf]



DMP in Horizon 2020

- Bei Teilnahme im ORD-Pilot ist DMP verpflichtend!
- DMP muss ***up-to-date*** gehalten werden (*initial/mid term/final*)
- 1. Version ist innerhalb der **ersten 6 Monate** des Projekts zu liefern, danach aktualisierte Versionen
- Definition FD: ***Underlying Data*** und ***Other Data***
- Dzt. **kein** bestimmtes Formular von der EC vorgeschrieben. DMPOnline wird vorgeschlagen
- Bei Beachtung bestimmter Anforderungen auch andere DMP Formen möglich

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



How to create a DMP?

Users' perspective

How to create a DMP?

- Most cases by filling out a document template or by answering questions from a checklist
- Using software tools
 - users choose appropriate funders template
 - only relevant questions and guidance is presented
 - results can be exported or directly submitted
- DMPTool <https://dmp.cdlib.org>
- DMPOne <https://dmponline.dcc.ac.uk/>



My plan (Horizon 2020 DMP)





0/5 questions answered

approx. 20% of available space used

[Plan details](#)[Initial DMP](#)[Mid-term Review DMP](#)[Final review DMP](#)[Share](#)[Export](#)

Scientific research data should be easily: 1. Discoverable (1 question, 0 answered) –

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?

B *I*    

[Save](#)[Not answered yet](#)[Guidance](#)[Share note](#)

DCC guidance on Discovery by Users –

Questions to consider:

- How will potential users find out about your data?
- Will you provide metadata online to aid discovery and reuse?

Guidance:

Indicate how potential new users can find out about your data and identify whether they could be suitable for their research purposes. For example, you may provide basic discovery metadata online (i.e. the title, author, subjects, keywords and publisher).

2. Accessible (1 question, 0 answered) +

3. Assessable and intelligible (1 question, 0 answered) +

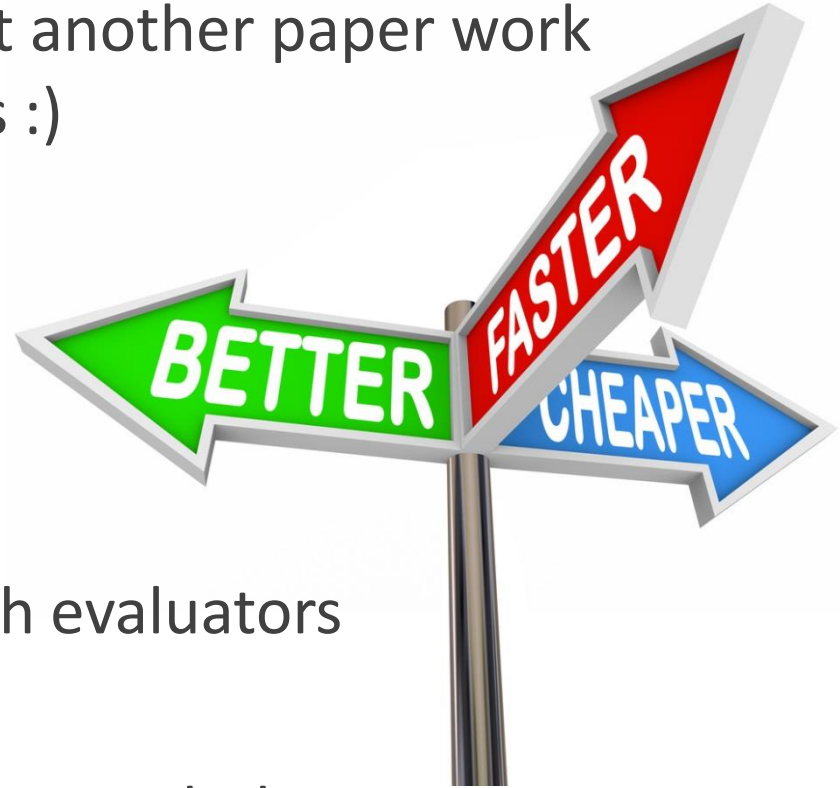
4. Usable beyond the original purpose for which it was collected (1 question, 0 answered) +

5. Interoperable to specific quality standards (1 question, 0 answered) +

Who benefits from DMPs?

Who benefits from DMPs?

- DMPs are NOT meant to be yet another paper work imposed on (poor) researchers :)
- Beneficiaries
 - Researchers
 - Projects applicants
 - Funding bodies and research evaluators
 - Repositories
 - Reusing parties and society as a whole



Who benefits from DMPs?

- Researchers
 - save time
 - data well organised from the beginning
 - less effort when publishing data
 - lower risk of losing data
 - increase efficiency
 - easier location of appropriate data
 - easier sharing of data with others thus enabling more contributions
 - higher visibility of research and more credit
 - well documented data is easier to understand and therefore easier to reuse by others
 - gain credit as data providers

Who benefits from DMPs?

- Repositories
 - ingest process in terms of OAIS is quicker (but may still require checks)
 - metadata is already provided
 - format of data is known
 - legal rights and ethics issues are already clarified
 - which data cannot be disclosed to public
 - which preservation actions can be performed
 - which data and for how long has to be retained

Who benefits from DMPs?

- Projects applicants
 - not always the same person who is conducting the research
 - can better estimate costs
 - storage and backup
 - software licenses
 - plan for long-term preservation
 - assign responsible party
 - allocate money to cover the expenses

Who benefits from DMPs?

- Funding bodies and research evaluators
 - increase trust that results are real and trustable
 - enhanced cooperation between projects and scientists
- Reusing parties and society as a whole
 - higher confidence in claims made by scientists
 - having the data the verification is easier (but not necessarily easy)
 - less repetition of efforts by reusing the data

What is needed to implement a DMP solution?

Institutional perspective

Do you really need your own DMP template?

- 1. Does your institution encourage/require that researchers write Data Management Plans, even if their funder does not require one? If so, what information are researchers asked to provide?
- 2. Do DMPs submitted to research funders supply the information your institution asks for (if any), or are there additional questions that you want to ask?
- 3. What guidance, examples and suggested answers can you provide to help researchers write DMPs?

[<http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMPonline-customisation-guidelines.pdf>]

If you need one...

- Identify who and why will need this information
- Browse existing checklists for help
- Provide a template
 - questions
 - guidance
 - possible answers
- Identify overlaps of your template (e.g. with Horizon 2020)
- Test it and incorporate feedback

It is not only about DMPs...

- Design the whole data management process
 - do you need different DMP phases?
 - who receives the DMPs?
 - who reviews them?
 - who provides assistance?
 - etc.
- Provide a reviewers' guide
 - clear judgment criteria
- Offer default data location
 - if you require people to deposit data and describe in a DMP, provide a place for doing that

It is not only about DMPs...

- Establish a policy
 - make a use of DMPs obligatory
 - consider incentives and penalties
- Provide support
 - tools
 - examples
 - trained staff for assistance in DMP creation
 - related problem: Where do you find people with expertise?
- Awareness and understanding
 - offer training material
 - organise workshops
 - related problem: How to teach correct data management practices at schools...

Lessons learned

Universität Wien

DMPs an der Universität Wien

- Vermehrt Anfragen seit H2020 und FWF Research Data Pilot
- Seit März 2016 **Schulungen** im Rahmen der Personalentwicklung
- Lessons learned:
 - (1) **Forschungsservices** miteinbeziehen
 - (2) **Forschungsförderer** miteinbeziehen
 - (3) **Kontextualisierung** mit vorhandenen Infrastruktur-Angeboten und Services
 - (4) **Templates** (DE & EN) an eigene Einrichtung anpassen: <http://phaidra.univie.ac.at/o:424283>
 - (5) **Kostenmodelle** anbieten
- To dos
 - Optimierung von DMP-Templates und Online-Tool
 - Institutionelle Verankerung
 - Stay in the loop! (FFG Policy, WWTF Policy)



Lessons learned

WWTF

WWTF Status

- Im Zuge der Precision Medicine Ausschreibung DMP erstmalig verlangt
- Die Bedeutung des DMP (im Speziellen die Nachnutzbarkeit der Daten nach Projektende) wurde von der Jury besonders hervorgehoben
- Für die Vollantragsphase wurden darausfolgend entsprechende Fragen an die AntragstellerInnen formuliert (mit Unterstützung der Jury und Barbara Prainsack (Kings College London sowie Bioethikkommission beim Bundeskanzleramt))

Guide for writing a full proposal

Data Management Plan (max. 1 page)

- First, a data set description is required detailing what types and amount of data the project will generate/collect.
- Additionally, the proposal should delineate how the data will be exploited and/or shared/made accessible for verification and re-use (standards and metadata).
- Finally, please provide details on a sustainability plan in relation to the data you generate.

Peer questionnaire (sent out for external review)

- Have the applicants devised a sensible and viable concept for the collection of data and respective data analysis?
- Does the data management plan in the grant proposal consider how the data should be handled after the end of the granting period (e.g., storage for potential internal re-use or public availability)?
- Keine Bewertung mit Notensystem!

Kontakt

- **Tomasz Miksa**
SBA Research und Institut für Softwaretechnik und Interaktive Systeme,
TU Wien
miksa@ifs.tuwien.ac.at
- **Barbara Sánchez Solís**
Bibliotheks- und Archivwesen der Universität Wien
e-Infrastructures Austria
barbara.sanchez.solis@univie.ac.at
- **Cornelia Schrauf**
WWTF
cornelia.schrauf@wwtf.at