



*Data Archiving and Networked Services*

# **World Wide Data Management: chaos of harmonie?**

**Peter Doorn, DANS**

**Landelijke bijeenkomst datamanagement van NWO en  
SURFacademy**

**10 September 2014, De Reehorst, Ede**



# What is data management?

- Data management is a general term covering how you organize, structure, store, and care for the information used or generated during a research project
- It includes:
  - How you deal with information on a day-to-day basis ***over the lifetime of a project***
  - What happens to data in the longer term – what you do with it ***after the project concludes***
- Acronyms:
  - DMP = Data Management Plan
  - RDM = Research Data Management

First 3 slides based on “Research Data Management Training Materials”, University of Oxford [http://damaro.oucs.ox.ac.uk/training\\_materials.xml](http://damaro.oucs.ox.ac.uk/training_materials.xml)



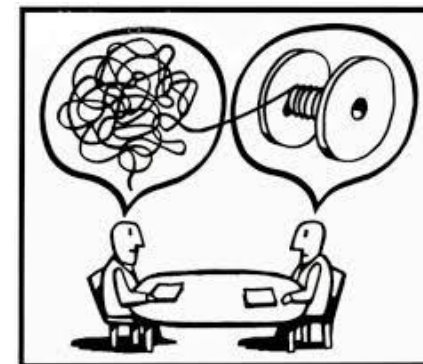
# Chaos of harmonie?

## Wie vereisen RDM?

- financiers, nationaal en internationaal, particulier en overheid
- onderzoeksorganisaties, nationaal en internationaal (bijv. ESFRI)
- universiteiten
- tijdschriften (DAP)

## Wat vereisen ze?

- welke criteria
- bewaartermijn
- inclusief archivering
- data delen
- subsidiabel?
- inclusief software?
- aangewezen datacentrum/archief?



## Hoe stringent?

- eis
- sterke aanbeveling
- advies

## Hoe vereisen ze het?

- via projectvoorstel
- bij evaluaties
- bij functioneringsgesprekken
- via gedragscode

## Hoe uitvoerig?

- data management plan
- paragraaf
- eenmalig of periodiek
- wel/geen tool voor DMP

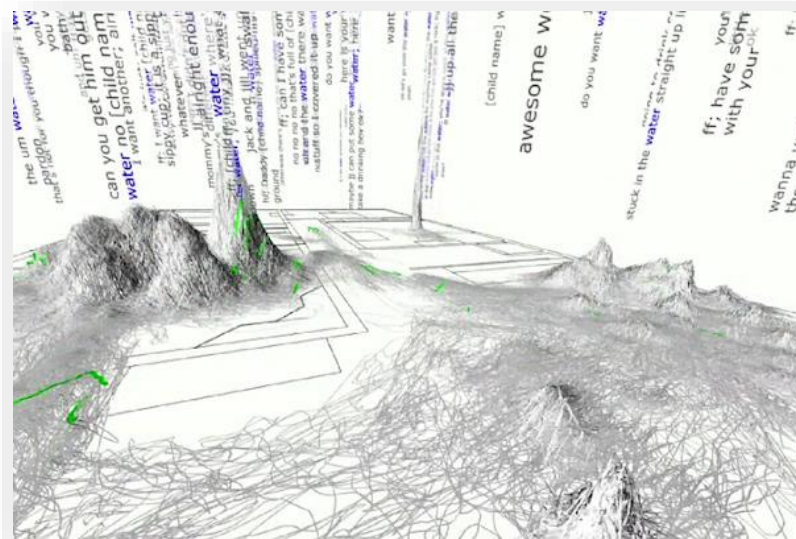


# SIM4RDM European Landscape Study of Research Data Management (May 2013)

- SIM4RDM: a two year ERA-NET project funded by FP7
- The SIM4RDM project strives for better policies in the area of managing research data on the national and European level
- Authors: Wilma Mossink, Magchiel Bijsterbosch, Joeri Nortier



**SIM4RDM**  
SUPPORT INFRASTRUCTURE MODELS  
FOR RESEARCH DATA MANAGEMENT

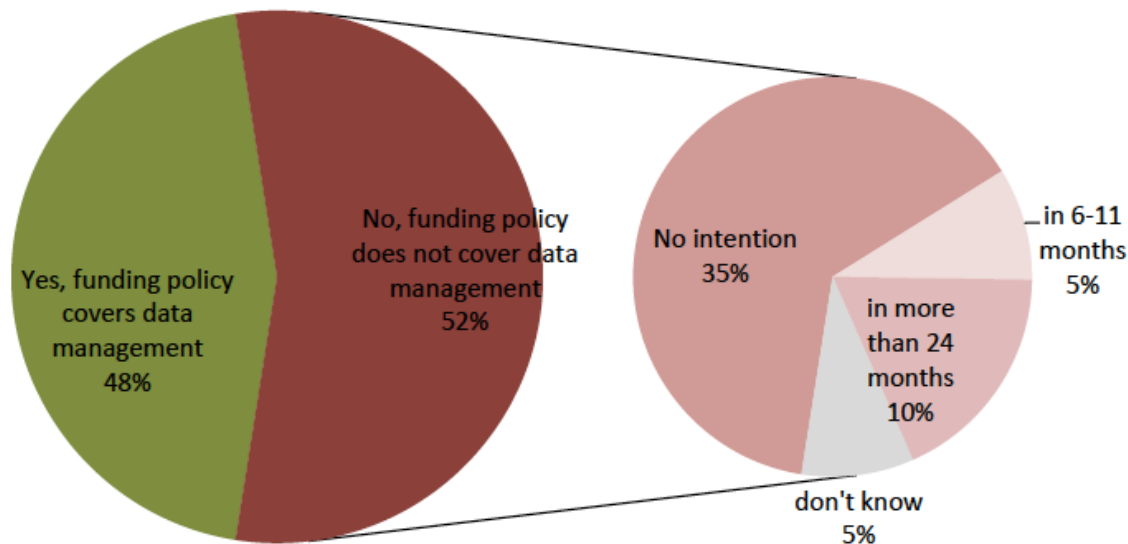




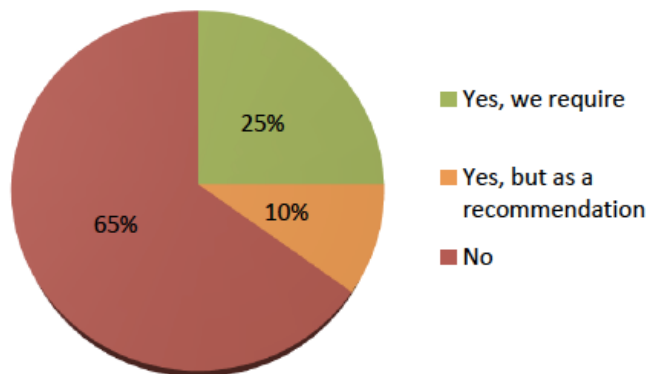
# SIM4RDM Landscape Study (N = 18; 2013)

COUNTRY	RESPONDENTS
FINLAND	3
NORWAY	3
IRELAND	2
SWEDEN	2
BULGARIA	1
HUNGARY	1
UK	1

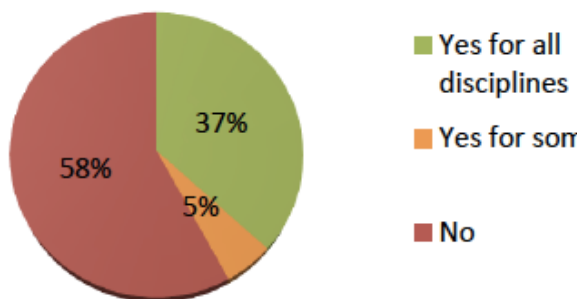
**Table 1:** Research funders



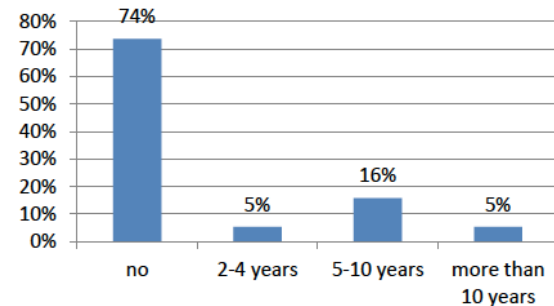
**Figure 1:** Policy for data funding and data management (n=18)



**Figure 2:** Data Management Plan (n=18)



**Figure 4:** Data retention (n=18)



**Figure 5:** Term for how long data should be preserved

*Driven by data*



# Data Management Policies in the United States

- The Office of Management and Budget sets forth standards for obtaining consistency and uniformity among federal agencies in the administration of grants to and agreements with institutions of higher education, hospitals, and other non-profit organizations.
- A number of US funding agencies have drawn up data management policies based on that circular.
- Since 2011, the National Science Foundation in the United States has made a data management plan mandatory when submitting grants proposals. Proposals must now include a supplementary document of no more than two pages labelled 'Data Management Plan' which should include the following information:





# NSF requirements of DMPs

- **Products of the Research:** types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.
- **Data Formats:** standards for data and metadata format and content
- **Access to Data, Data Sharing Practices and Policies:** including protection of privacy, confidentiality, security, intellectual property, or other rights
- **Policies for Re-Use, Re-Distribution and Production of Derivatives**
- **Archiving of Data:** preservation, access



*Driven by data*



# H2020 Guidelines on RDM

## Annex 1: Data Management Plan (DMP) template

- The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.
- The DMP is not a fixed document, but evolves during the lifespan of the project.
- The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)





## H2020 DMP requirements

- **Data set reference and name:** Persistent Identifier
- **Data set description:**
  - to whom useful;
  - related scientific publication;
  - integration with existing data and reuse.
- **Standards and metadata:**
  - suitable standards of the discipline;
  - metadata.
- **Data sharing:**
  - access procedures, levels and restrictions (e.g. intellectual property, commercial, privacy-related, security-related);
  - embargo periods (if any);
  - mechanisms for dissemination and necessary software for re-use;
  - repository where data will be stored.
- **Archiving and preservation:**
  - long-term-storage and backup procedure;
  - retention period;
  - data volume, how are the associated costs covered.



## Annex 2: Additional guidance for Data Management Plans

Scientific research data (and associated software) should be easily:

- 1. Discoverable:** persistent identifier
- 2. Accessible:** licenses
- 3. Assessable and intelligible:** metadata, documentation
- 4. Useable beyond the original purpose for which it was collected:** safely stored in certified repositories for long term preservation
- 5. Interoperable to specific quality standards:** data exchange, formats





# Australian Code for the Responsible Conduct Of Research (A2: management of research data)

- Responsibilities of institutions
  - Retain research data and primary materials
  - Provide secure research data storage and record-keeping facilities
  - Identify ownership of research data and primary materials
  - Ensure security and confidentiality of research data and primary materials
- Responsibilities of researchers
  - Retain research data and primary materials
  - Manage storage of research data and primary materials
  - Maintain confidentiality of research data and primary materials

ARC funding requirement for NHMRC and ARC

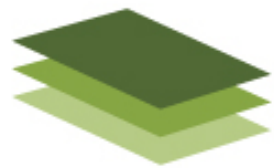


# Comparison of criteria US, EU, Australia

NSF (US)	H2020 (EU)	ANDS (Australia)	
Products of the Research	Data set reference and name; Data set description	What data will be created?	Who is responsible?
Data Formats	Standards and metadata	What data management practices will be used?	
Access to Data, Data Sharing	Data sharing	Who will own and have access to the data?	
Re-Use, Re-Distribution and Production of Derivatives	Archiving and preservation	What policies will apply to the data?	
Archiving of Data		What facilities and equipment will be required?	

## Summary guide to funder Requirements for Data Management and Sharing

Funder	Data covered	Data Plan Requirement	Funding for Data Management	Data retention requirements	Data sharing requirements	Designated Data Centre	Compliance Monitoring
Biotechnology & Biosciences Research Council	All research outputs	Yes	DM costs may be requested	10 yrs minimum	Publish data at same time as research findings or in-line with best practice in field	No specific archive mentioned	Yes, Final Report assessment procedure
Bill & Melinda Gates Foundation	Final datasets & supporting information	Yes, if grant is \$500,000+	DM costs may be requested	Not stated	Publication in 'timely manner'	No specific archive mentioned	None stated
Breast Cancer Campaign	All research outputs	No	Not stated	Not stated	Encourage publication as early as possible	No specific archive mentioned	None stated
Cancer Research UK	All research outputs, including non-published results	Yes	No	5 yrs minimum	Publish data at same time as research findings or in-line with best practice in field	No specific archive mentioned	Monitoring through grant review and grant closure report
Department of Health UK	All research outputs	No	Not stated	Retention should comply with appropriate legislation	Publication in 'timely manner'	Data.gov and OpenLocally for open data	Institutional monitoring
Dept for International Development	All research outputs	Yes	DM costs may be requested	5 yrs minimum	Deposit data in disciplinary or institutional repository within 12 months	Submit MD, research papers & 'small	None stated
Drugs for Neglected	All research outputs	No	Not stated	Not stated	Results should be "readily available and		



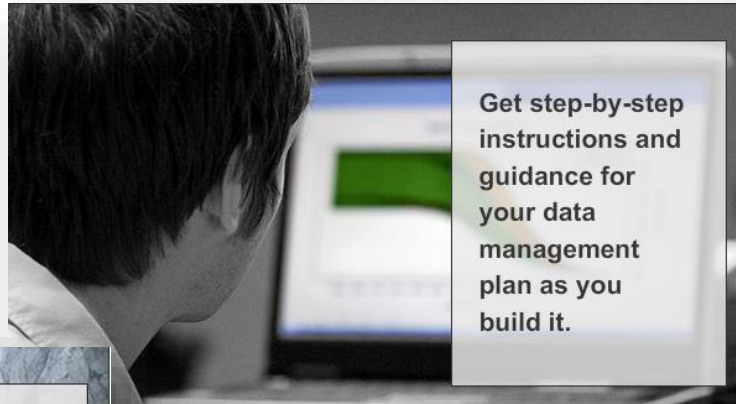
# DMPTool

Guidance and Resources for your Data Management Plan

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



Create ready-to-use data management plans for specific funding agencies.

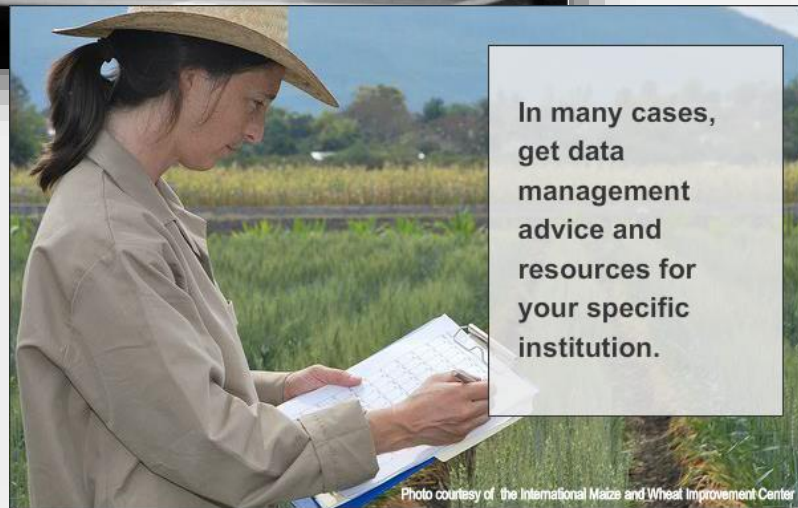


Get step-by-step instructions and guidance for your data management plan as you build it.



Meet funder requirements for data management plans.

Photo courtesy of Glacier National Park

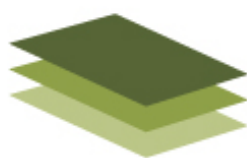


In many cases, get data management advice and resources for your specific institution.

Photo courtesy of the International Maize and Wheat Improvement Center

**DANS**

*Driven by data*



# DMPTool

Guidance and Resources for your Data Management Plan

## Funder Requirements

Funder	Funder Link	Sample Plan	Funder Requirements Template	Supported in DMP Tool
Gordon and Betty Moore Foundation	<a href="#">Guidelines</a>		Template <a href="#">[RTF]</a>	✔
Gulf of Mexico Research Initiative			Template <a href="#">[RTF]</a>	✔
Institute of Education Sciences (US Dept of Education)	<a href="#">IES Data Sharing Implementation Guide</a>		Template <a href="#">[RTF]</a>	✔
Institute of Museum and Library Services	<a href="#">Guidance</a>		Template <a href="#">[RTF]</a>	✔
Joint Fire Science Program			Template <a href="#">[RTF]</a>	✔
National Institutes of Health	<a href="#">Guidance</a>		Template <a href="#">[RTF]</a>	✔
National Oceanic and Atmospheric Administration	<a href="#">Procedural Directive</a>		Template <a href="#">[RTF]</a>	✔
NEH-ODH: Office of Digital Humanities	<a href="#">Guidelines</a>	<a href="#">NEH-ODH Sample</a>	Template <a href="#">[RTF]</a>	✔
NSF - Mathematical Sciences	<a href="#">Advice to PIs</a>			
NSF-AGS: Atmospheric and Geospace Sciences	<a href="#">AGS Advice</a>		Template <a href="#">[RTF]</a>	✔
NSF-AST: Astronomical Sciences	<a href="#">Advice to PIs</a>		Template <a href="#">[RTF]</a>	✔

As new Data Management Plan Guidelines become available, they will be added to this page and eventually to the DMPTool.

### Disclaimer

The sample plans provided here may or may not be associated with successful grant applications. They may contain details not relevant to your specific project. They are provided only to illustrate representative responses.

### Funder Key

**NSF** = National Science Foundation  
**NEH** = National Endowment for the Humanities  
**IMLS** = Institute of Museum and Library Services

### Additional DMP Tool Help

- [DMP Tool Guide](#)
- [Video Demo](#)



Signed in as Peter Doorn -

- View plans
- Create plan
- About
- News
- Help

## My project (DCC Template)

No questions have been answered

- Plan details
- Generic DMP**
- Share
- Export

<b>Data Collection</b> (2 questions, 0 answered)	+
<b>Documentation and Metadata</b> (1 question, 0 answered)	+
<b>Ethics and Legal Compliance</b> (2 questions, 0 answered)	+
<b>Storage and Backup</b> (2 questions, 0 answered)	+
<b>Selection and Preservation</b> (2 questions, 0 answered)	+
<b>Data Sharing</b> (2 questions, 0 answered)	+
<b>Responsibilities and Resources</b> (2 questions, 0 answered)	+





# Final remarks & some questions

- Tendency: DMP will become mandatory (eg. Science Europe Working Group on Research Data)
- Agreement on basic principles
- Variations in the degree of detail:
  - depending on scientific domain, size of project, funding instrument/organisation, institution, country, ...
- Procedures are under construction
  - data management section in proposal stage
  - DMP as part of contract or in first phase of project?
  - DMP as a living document?
- Not just a paper (or digital) tiger...
  - who will review? Sanctions?
- To what degree will RDM be fundable?
- A DMP tool helps to harmonize & make requirements transparent





# DMPTool.org



90 second video

# RESEARCH DATA ALLIANCE FOURTH PLENARY MEETING

22 – 24 September 2014  
Amsterdam, the Netherlands | Meervaart conference centre

[www.rd-alliance.org/rda-fourth-assembly-meeting.html](http://www.rd-alliance.org/rda-fourth-assembly-meeting.html)

The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data to address cross-border & cross-disciplinary challenges.

The current global research data

landscape is highly fragmented, by disciplines or by domains, from oceanography, life sciences and health, to agriculture, space and climate, social sciences and humanities.



Johnny saw some fine plums hanging,  
Oh, like eggs, so very large...  
(Reaping the fruits from the first Dutch children's book "Kleine gedigten voor kinderen" by Hieronymus van Alphen, 1778).

## Reaping the fruits:

The Research Data Alliance Fourth Plenary Meeting in Amsterdam showcases the first concrete outputs from the RDA Working Groups

RDA PLENARY 4 MEETING IS CO-ORGANISED BY:

3TU.Datacentrum **DANS** Data Archiving and Networked Services **net.nl** netherlands Science center



UNIVERSITY OF AMSTERDAM

For more information on the Research Data Alliance – [www.rd-alliance.org](http://www.rd-alliance.org)

## RDA FOURTH PLENARY MEETING CO-LOCATED EVENTS

<https://www.rd-alliance.org/assembly4-co-located-events.html>



50 years Data Archiving and The Dutch Data Award



APARSEN: Alliance Permanent Access to the Records of Science in Europe Network



Creative-B: Workshop "Global Roadmap for biodiversity research infrastructures"



Data Science Research Center Amsterdam: new opportunities for research Data



Data Science of Approval Conference 2014



EUDAT 3rd Conference - Bringing data infrastructures to Horizon 2020



ORCID and DataCite: Towards Holistic Open Research

## WHAT DATA SHARING PROBLEM ARE YOU TRYING TO SOLVE?

**Get involved by becoming an RDA Member and signing up to the Working & Interest Group.** Participation in the RDA is free and open to anyone who agrees to the RDA principles. Scientists come together through focussed Working Groups, exploratory Interest Groups to exchange knowledge, share discoveries, discuss barriers and potential solutions, explore and define policies and test as well as harmonise standards to enhance and facilitate global data sharing. Coupled with this RDA boasts a broad, committed membership of individuals and organisations dedicated to improving data exchange. <https://www.rd-alliance.org/user/register>

# *Data Archiving and Networked Services*

## **Thank you for your attention**

[www.dans.knaw.nl](http://www.dans.knaw.nl)

[www.narcis.nl](http://www.narcis.nl)

[peter.doorn@dans.knaw.nl](mailto:peter.doorn@dans.knaw.nl)

