

Science Europe Research Data Working Group

Initial insight into the evaluation of RDM Plans

Research Data Management (RDM) Task Force

Patricia Clarke, Health Research Board and Karl Gertow, Swedish Research Council

WG Meeting 29 January 2018

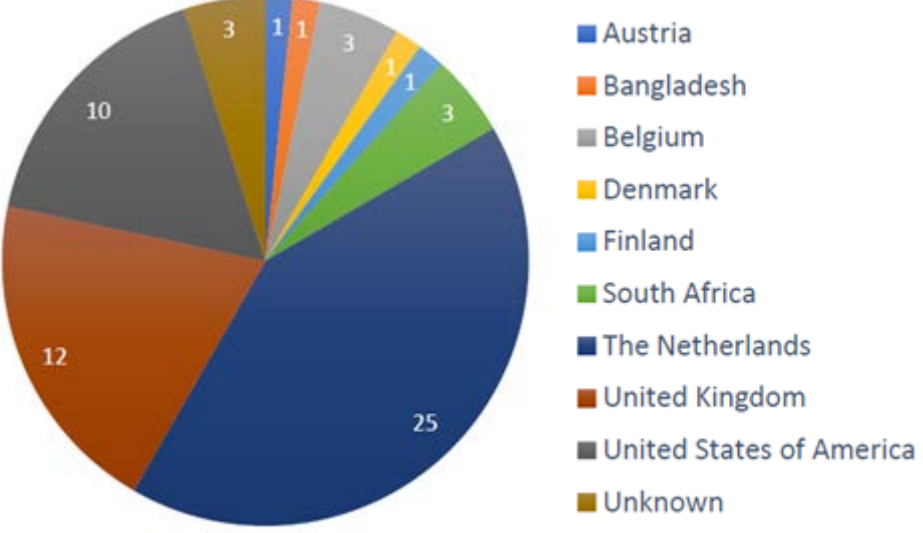
An under-developed area

Recent discussions on reviewing and monitoring of DMPs

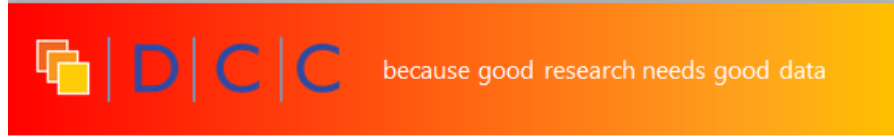


Report: survey of DMP reviewer experiences
Marjan Grootveld¹ and Mariëtte van Selm², June 2017

Number and origin of responses (n=60) [1, 2]



1 Data Archiving and Networked Services (DANS), The Hague.
2 University of Amsterdam (UvA) /Amsterdam University of Applied Sciences (AUAS).



Reviewing Data Management Plans

30 November 2016
Friends House, London
Digital Curation Centre

www.dcc.ac.uk/resources/data-management-plans

Guidance for Reviewers



	Research product	Source	Format	Size	Preserved (how?)	Shared (how?)
	Tables, images, computer code, curriculum items, physical samples	Data repository, Instrument, interviews	JPG, MATLAB, Excel table	>1TB, 20K files	Discarded, PI retains Archiving service	By request, Website, Archive/Repository
1						
2						
3						
4						
5						

Get it at bit.ly/DMPworksheet

Data management during project:

- Storage: has a backup plan
- Location & media used:
- ☆ 2+ copies with 1 off-site
- ☆ Specifies who is responsible
- ☆ Data security / access controls
- ☆ Has conventions for naming & organizing files
- ☆ Version control
- ☆ Collaboration coordination

Data retention after the project:

- Where is data preserved?
- How long?
- Who administers?
- ☆ Gives reasons for preserving data (especially raw data)
- ☆ Using an archive service or repository?

Data Sharing

- Is data publically accessible?
- When will data be shared?
- Who administers?
- ☆ Describes audience to benefit.

Preparation of shared data

- Uses their research field's metadata standards
- AND/OR creates description sufficient for re-use
- Metadata or supplementary files explaining: content/ file structure/ procedures/ codebook or variable-level detail
- ☆ Metadata associated with digital files

Data sharing policy

- Gives conditions for re-use
- Accounts for:
 - ☆ privacy (personal identifiers)/security issues
 - ☆ intellectual property (copyrights, patents)
 - ☆ delays for sharing (e.g., embargos)

Services of archive (if specified for preservation and/or sharing data)

Archive type		Preservation activities		Data sharing services
<input type="checkbox"/> PI's Institutional repository (documents)	<input type="checkbox"/>	Data integrity checking	<input type="checkbox"/>	Public access to data files
<input type="checkbox"/> ☆ Data repository	<input type="checkbox"/>	☆ Migration to new formats, media	<input type="checkbox"/>	☆ Persistent data citation

- If a plan states there is no data to manage or share, have they justified it?

Assessment/ Evaluation rubrics



DART
PROJECT

DMPs as A Research Tool

US [DART project](#)

- analytic rubric for assessing DMP content and quality
- Amanda L. Whitmire IDCC 2016 presentation – [Analysing DMPs to inform research data services. Lessons from the DART project](#)
- See http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/Workshop8/Whitmire_DARTPres.pdf

UK community rubrics initiative

- Inspired by the DART project and led by Mary Donaldson, University of Glasgow
- 1st phase involves developing rubrics to be used when evaluating DMPs against UK funder requirements.
- See Research Data Network UK (Folder Data Management Plans/ Compliance Tools) <https://research-data-network.readme.io/docs/compliance-tools>

Example – Wellcome Trust Rubric V2.0

Performance Criteria	Performance Levels		
	Detailed	Addressed but incomplete / unsatisfactory	Not addressed
What types of data outputs will your research generate?	Data types clearly defined. Eg experimental measurements, models, recordings, video, images, machine logs, source code, databases, physical samples etc.	Some data types are mentioned, but not all.	No information provided.
Which data will have value to other people and why?	Data types of potential value to others clearly identified and justification about the value is provided (indication of likely user base/demand).	Valuable data types merely listed, but no justification of the value provided.	No information provided.
Will file formats in which data will be stored and shared allow long-term preservation?	A clear statement that data will be stored and shared in open formats, or in formats widely used by the community. If proprietary formats are used for data storage and sharing, information is provided justifying why open formats are not suitable and reference to software necessary to open and read these files is provided.	File formats for different data types are mentioned, but there is no indication of their suitability for long-term data preservation and sharing.	File formats and their suitability for sharing are not mentioned.
How will you describe and document your data? Are there any metadata standards that you can adhere to in order to aid comprehension and make your data intelligible to re-users?	Clear outline of documentation and metadata strategy with references to existing good practice in the community or detailed project-specific approach where community standards do not exist.	Some mention of documentation or metadata standards without detail about community standards or a project-specific approach.	No mention of documentation or metadata.

The evaluation of FAIRness

Efforts to define metrics to assess the FAIRness of a digital resource.

- Metrics page: <http://fairmetrics.org>
- Paper: <https://www.biorxiv.org/content/early/2017/12/01/225490>
- Github: <https://github.com/FAIRMetrics/Metrics>
- Human readable description of metrics: <https://github.com/FAIRMetrics/Metrics/tree/master/Distributions>

14 FAIR indicators (November 2017)

FM-1A Identifier Uniqueness

FM-F1B Identifier Persistence

FM-F2 Machine readability of met-data

FM-F3 Resource identifier in metadata

FM-F4 Indexed in searchable resource

FM-A1.1 Access Protocol

FM-A1.2 Access Authorisation

FM-A2 Meta-data Longevity

FM-I1 Use a knowledge representation language

FM-I2 Use FAIR vocabularies

FM-I3 Use qualified references

FM-R1.1 Accessible Usage Licence

FM-R1.2 Detailed Provenance

FM-R1.3 Meets Community Standard

Template

Metric Descriptor

Metric Identifier

Metric Name

To which principle does it apply?

What is being measured?

Why should we measure it?

What must be provided?

How do we measure it?

What is a valid result?

For which digital resource(s) is this relevant?

Examples of their application across types of digital resources

Comment

Horizon2020 Commission Expert Group Turning FAIR data into reality

FAIR-Data-EG / consultation

Watch

19

★ Star

17

Fork

3

<> Code

🔔 Issues 33

🔗 Pull requests 0

📁 Projects 0

📊 Insights

🔍 is:issue is:open label:metrics

Labels

Milestones

New issue

✕ Clear current search query, filters, and sorts

🔔 5 Open ✓ 0 Closed

Author ▾






Labels ▾

Projects ▾

Milestones ▾

Assignee ▾

Sort ▾

- 🔔 FAIR Maturity Model **FAIRness** **metrics**  1
#31 opened on Jul 31 2017 by CaroleGoble
- 🔔 FAIR metric form **FAIRness** **metrics**  3
#25 opened on Jul 12 2017 by micheldumontier
- 🔔 A proposal for assessing the FAIRness of data in Trusted Digital Repositories **FAIRness** **metrics**  5
#23 opened on Jul 4 2017 by etsoupra
- 🔔 measures of FAIRness as a guide to data providers **FAIR extensions** **evaluation** **metrics**  2
#13 opened on Jun 30 2017 by mellybelly
- 🔔 FAIRmetrics score on the profile of each published dataset? **FAIRness** **evaluation** **metrics** **questions** **standards** **visualisation**  6

Plan-Europe - Platform of National eScience Centers in Europe

PLAN-E meeting, April 27 & 28, 2017, Poznan, PSNC, Poland

DANS: FAIR badge scheme



2 User Reviews
1 Archivist Assessment
24 Downloads

- First Badge System based on the FAIR principles: proxy for data quality assessment
- Operationalise the original principles to ensure no interactions among dimensions to ease scoring
- Consider Reusability as the resultant of the other three:
 - the average FAIRness as an indicator of data quality
 - $(F+A+I)/3=R$
- Manual and automatic scoring



5 ★ DATA RATING TOOL

5 ★ DATA RATINGS

The CSIRO 5-star Data Rating tool provides a self-assessment rating scheme against the social, technical and informational attributes of data. This tool provides implementations of the [FORCE 11 FAIR data principles](#). The 5-star scheme aims to help users understand how mature some data or a service is.

More details about the [CSIRO 5-star data rating scheme](#) can be found [here](#).

Findable	★★★★★
Accessible	★★★★★
Interoperable	★★★★★
Reusable	★★★★★

Self-assessment tool (version 1)

CSIRO tool: <https://research.csiro.au/oznome/tools/oznome-5-star-data>

Other issues

- Funders and reviewers need training in DMP evaluation
- Open DMPs would support evaluation and monitoring use cases
- How can DMP evaluation and reporting be automated (especially for large scale reviews)?