

DataONE Webinar Series

Analyzing, interpreting, and implementing data management plans

Amanda Whitmire

Heidi Imker

Sarah Jones





*Enable new science and knowledge creation
through universal access to data about life on earth and the environment
that sustains it*

*DataONE network of Member Nodes:
www.dataone.org/current-member-nodes*

*DataONE Search tool:
<https://search.dataone.org>*

DataONE

Lesson 3: Data Management Planning

View all Education Modules at <https://www.dataone.org/education-modules> CC 0

The Data Management Plan (DMP)

A DMP outlines what you will do with your data during and after you complete your research project. It is a formal document that details how you will manage your data throughout its lifecycle.

Component 2: Metadata content & format

- What metadata are needed?
- How will metadata be created and/or captured?

Lesson 2: Data Sharing

Address data sharing throughout the data lifecycle

Describe data content, character, and process.

Deposit in a location from which it can be accessed.

Preserve in a location for long-term access.

Publish info to discover it easily.

Concerns about data sharing

Concern	Solution
Dependence on a single location	Use multiple locations and formats
Format and schema differences	Standardize or convert to common formats
Lack of metadata	Include detailed metadata
Access restrictions	Provide clear access policies
Version control	Use versioning tools
Storage costs	Optimize storage and consider cloud options

Lesson 1: Data Management

View all Education Modules at <https://www.dataone.org/education-modules> CC 0

Why manage data: the researcher perspective

- Keep yourself organized => find your own files!
- Track your processes for reproducibility
- Better version control of data
- More efficient data quality control
- More backups to avoid data loss
- Format your data for reuse by yourself & others
- Document your data for understandability and reuse
- Prepare it to share it & gain credibility and recognition for your scientific efforts

Information vs. Available Storage

Gainz, The Expanding Digital Universe

Causes of data loss

- Natural disasters
- Facilities infrastructure failures
- Storage failure
- Server hardware or software failure
- Application software failure
- Human errors
- Malicious attack
- Format obsolescence
- Loss of competencies
- Loss of funding
- Loss of institutional commitment

Costs of not doing data management can be very high!

The Data Lifecycle

The stages through which well-managed data passes from project inception to conclusion.

```

graph TD
    Plan --> Collect
    Collect --> Assure
    Assure --> Describe
    Describe --> Preserve
    Preserve --> Discover
    Discover --> Integrate
    Integrate --> Analyze
    Analyze --> Plan
    
```

The Case for Data Management

If data are:

- Well-organized
- Documented
- Preserved
- Accessible
- Verified as to accuracy & validity

The results are:

- High quality data
- Data that is easy to share and reuse
- Citation & credibility to researcher
- Cost savings to further science

Local contact information

DataONE

Tutorials on Data Management

Lesson 10: Analysis and Workflows

Analysis and Workflows

1 of 34

Get content → check → store → make available

© Jerry & Sherri Liggett

Presentations 10

Tutorials on Data Management

Documents 0

Infographics 0

Videos 0

#DWS2016
@DataONEorg
www.DataONE.org



DataONE Webinar Series

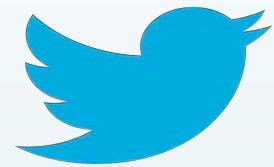
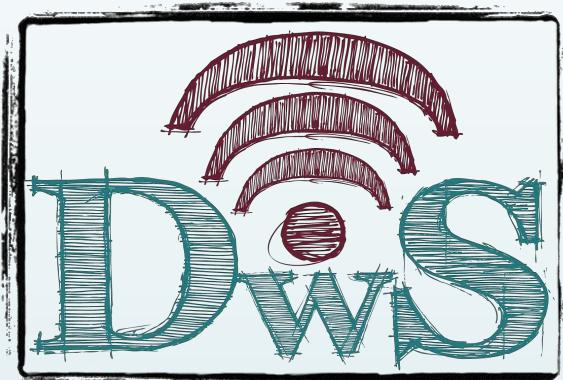
www.dataone.org/webinars

Upcoming Webinar Event

www.dataone.org/upcoming-webinar

Previous Webinar Events (Recording and Discussion)

www.dataone.org/previous-webinars



#DWS2016
@DataONEorg



Upcoming Webinar Event

www.dataone.org/upcoming-webinar

DPM “Stack”: A Management Infrastructure Frame for
Digital Preservation that Parallels Technical Infrastructure

December 13

Nancy McGovern

Massachusetts Institute of Technology



If you attending
as part of a group,
please enter the number of people
listening within
the “questions” box
Thanks!

Analyzing, Interpreting, and Implementing Data Management Plans



Amanda Whitmire
Stanford University



Heidi Imker
University of Illinois, Urbana-Champaign



Sarah Jones
Digital Curation Centre



DataONE Webinar Series

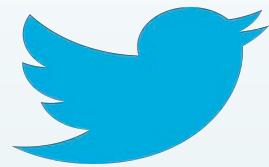
www.dataone.org/webinars

Upcoming Webinar Event

www.dataone.org/upcoming-webinar

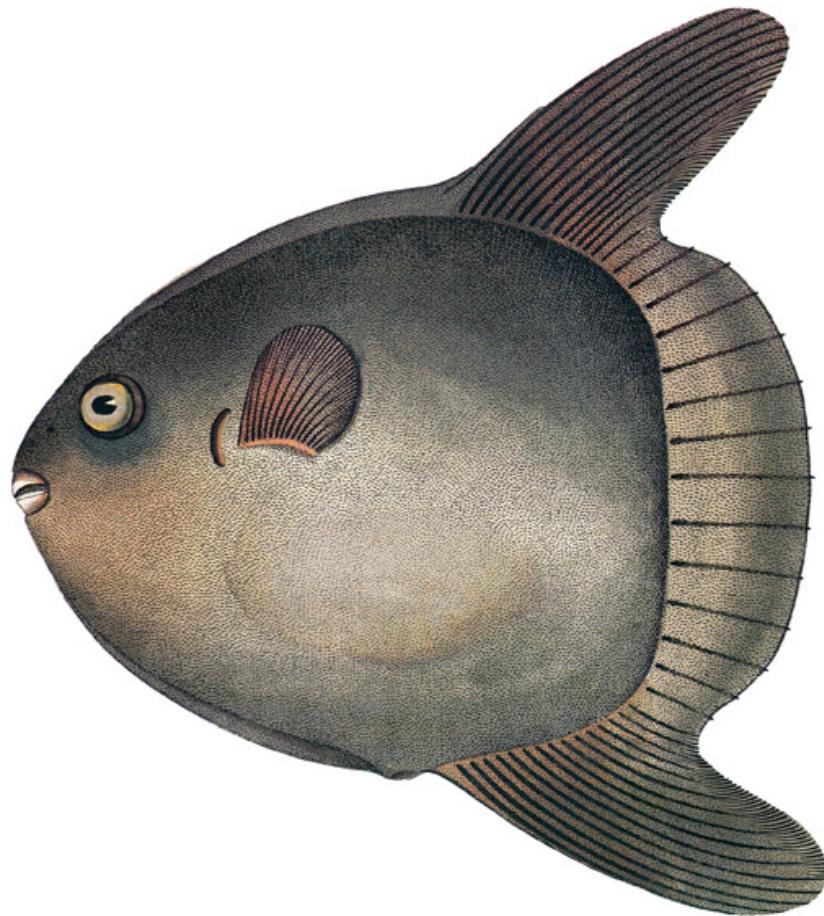
Previous Webinar Events (Recording and Discussion)

www.dataone.org/previous-webinars



#DWS2016
@DataONEorg





Data Management Plans

What good are they to us?

Amanda L. Whitmire, Ph.D.
Head Librarian & Bibliographer
Harold A. Miller Library
Hopkins Marine Station
Stanford University

@AWhitTwit

What are DMPs good for?

1. Information about researcher habits for data services development -**Amanda**
2. DMP consultation as gateway service to launch more meaningful interactions -**Heidi**
3. Overseas perspective; how universities have embedded DMP services into existing workflows & systems -**Sarah**



DMPs as source of researcher intel



Data management plan As Research Tool (DART Project)

Amanda Whitmire | Stanford University Libraries

Jake Carlson | University of Michigan Library

Patricia M. Hswe | Update

Lizzy Rolando | MailChimp

**Susan Wells Parham | Georgia Institute of Technology
Library**

Brian Westra | University of Oregon Libraries

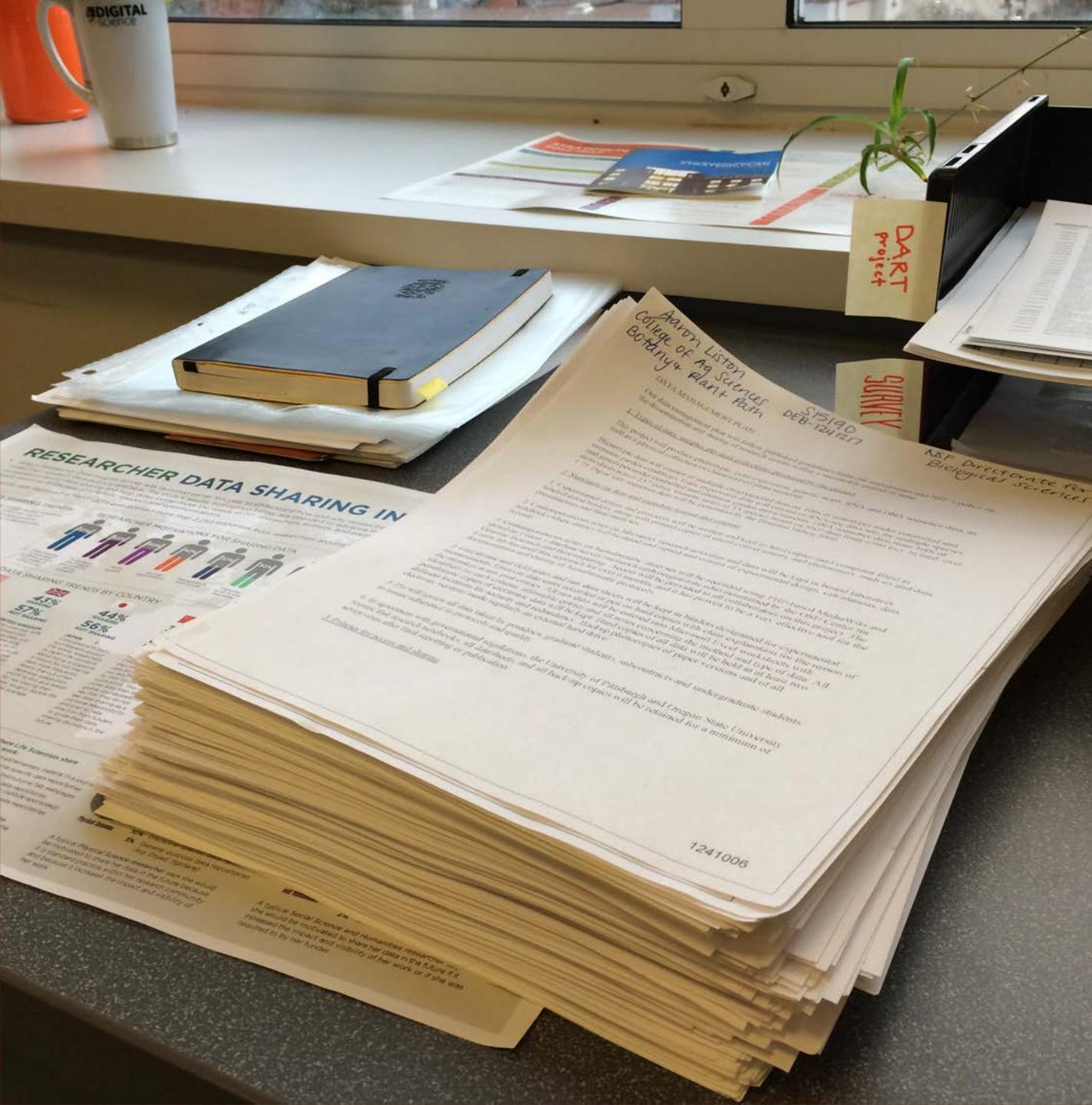


This project was made possible in part by the
Institute of Museum and Library Services
grant number LG-07-13-0328.

@DMPResearch

Performance Level

Performance Criteria	Performance Level			Directories	
	Complete / detailed	Addressed issue, but incomplete	Did not address issue		
General Assessment Criteria	<i>Describes what types of data will be captured, created or collected</i>	Clearly defines data type(s). <i>E.g. text, spreadsheets, images, 3D models, software, audio files, video files, reports, surveys, patient records, samples, final or intermediate numerical results from theoretical calculations, etc. Also defines data as: observational, experimental, simulation, model output or assimilation</i>	Some details about data types are included, but DMP is missing details or wouldn't be well understood by someone outside of the project	No details included, fails to adequately describe data types.	All
Direktorate- or division-specific assessment criteria	<i>Describes how data will be collected, captured, or created (whether new observations, results from models, reuse of other data, etc.)</i>	Clearly defines how data will be captured or created, including methods, instruments, software, or infrastructure where relevant.	Missing some details regarding how some of the data will be produced, makes assumptions about reviewer knowledge of methods or practices.	Does not clearly address how data will be captured or created.	GEO AGS, GEO EAR SGP, MPS AST
	<i>Identifies how much data (volume) will be produced</i>	Amount of expected data (MB, GB, TB, etc.) is clearly specified.	Amount of expected data (GB, TB, etc.) is vaguely specified.	Amount of expected data (GB, TB, etc.) is NOT specified.	GEO EAR SGP, GEO AGS



500 DMPs
100 from each
institution

Used Qualtrics
survey to collect
data

Distribution across
NSF directorates
followed distribution
of funded proposals

The DART Project: using data management plans as a research tool

Contributors: Amanda Whitmire, Jacob Carlson, Brian Westra, Patricia Hswe, Susan Parham

Date created: 2015-10-15 02:44 PM | Last Updated: 2016-06-22 04:36 PM

Category: Project

Description: This is a three-year National Leadership Grant for Libraries Demonstration Project to facilitate a multi-university study of faculty data management plans (DMPs).

The primary outputs of this project will be an analytic rubric to standardize the review of data management plans as a means to inform targeted expansion or development of research data services at academic libraries; and a study utilizing the rubric that presents the results of data management plan analyses at five universities. This project was made possible in part by the Institute of Museum and Library Services grant number LG-07-13-0328.

License: CC0 1.0 Universal

Wiki



ABOUT THE PROJECT

This two-year NLG-Libraries Demonstration Project proposal, led by Oregon State University in collaboration with the University of Oregon, the University of Michigan, the Georgia Institute of Technology and Pennsylvania State University, will facilitate a multi-university study of faculty data management plans (DMPs). **The primary output of this Read More**

Files



Citation

osf.io/kh2y6

Components

Rubric & related files

Whitmire, Carlson, Westra & 2 more

6 contributions

Data from: Using data management plans to explore variability in research data management practices across domains

Whitmire, Carlson, Westra & 2 more

20 contributions

Presentations

Whitmire, Carlson, Westra & 2 more

38 contributions

<https://osf.io/kh2y6/>

Find the rubric

See the survey we used to collect assessment data

Look at our DMP assessment data

Describes what type(s) of data produced

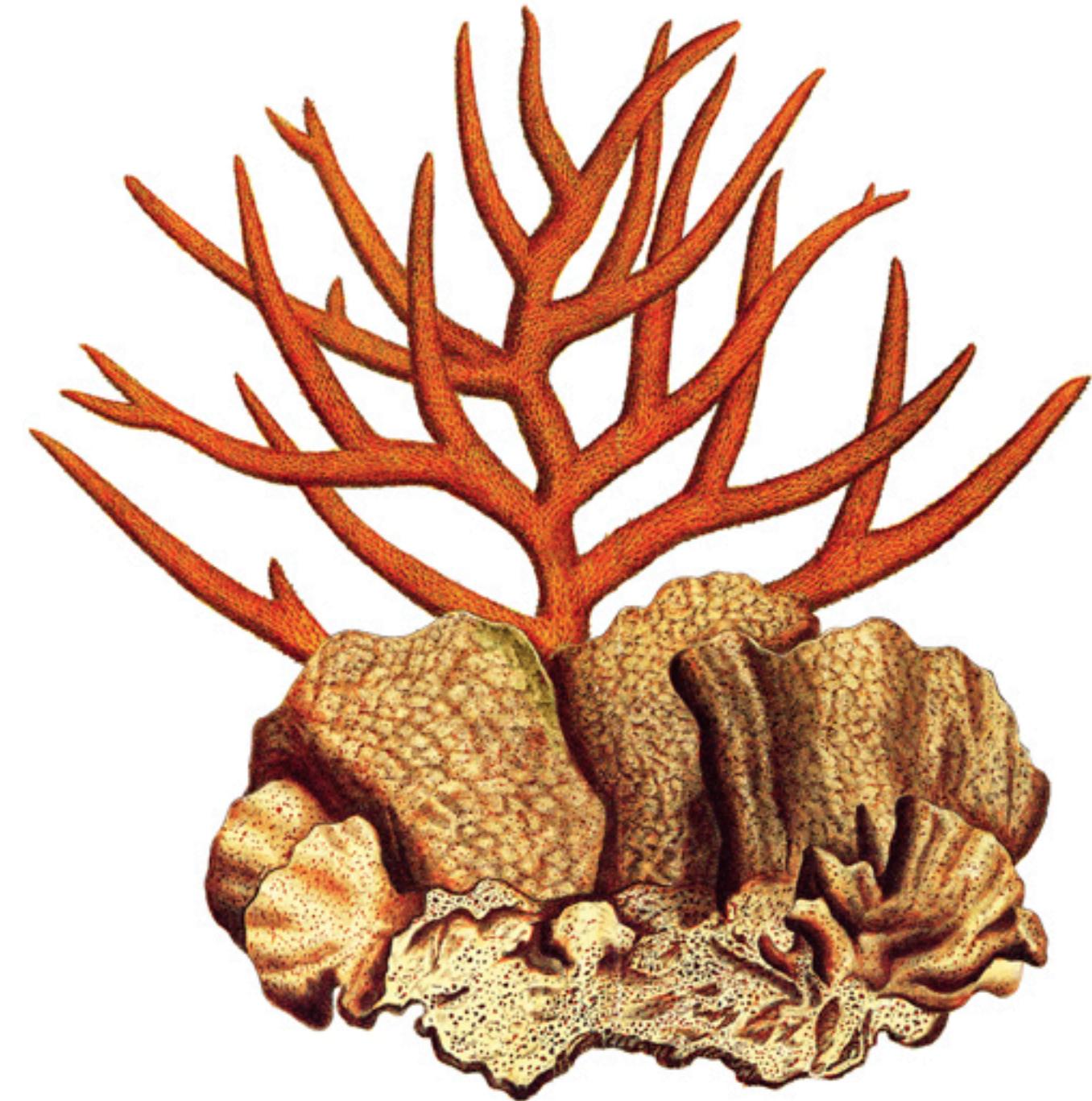
	Complete/ detailed	Addressed issue, but incomplete	Did not address	Scale
All	68	22	10	90
BIO	83	11	6	75
CISE	53	32	15	60
ENG	74	20	7	45
GEO	64	29	7	30
MPS	61	24	15	15
SBE	82	12	6	0

Describes how data will be shared

	Complete/ detailed	Addressed issue, but incomplete	Did not address	Scale
All	50	41	9	70
BIO	67	31	2	60
CISE	41	44	15	50
ENG	37	53	10	30
GEO	61	35	4	20
MPS	48	44	8	10
SBE	58	34	8	0

Where will they share data?

	All	BIO	CISE	ENG	GEO	MPS	SBE	Scale
Journal / supplement	36	27	23	45	35	54	18	80
Data center or repository	34	75	14	8	66	25	42	70
On request	30	23	30	38	24	34	26	60
Personal website	25	13	44	31	25	20	12	50
Other method	22	27	30	15	23	18	22	40
Institutional repository	17	6	12	20	6	28	20	35
Conference / proceedings	13	8	11	23	8	13	8	30
Did not specify	8	0	18	9	2	8	4	25
Thesis / Dissertation	3	0	0	5	2	6	2	20
Not planning to share	3	0	5	3	0	1	10	10
Book	2	2	2	3	2	2	2	0



**What is going
on with
Biology?**

They have infrastructure!

BIO: Repositories mentioned (frequency)

GenBank (14) Knowledge Network

Dryad (12) for Biocomplexity (3)

SRA (11) MorphBank (3)

iDigBio (3) NCBI (3)

TreeBASE (2)



-Question-



RESEARCH ARTICLE

Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia

Steven Van Tuyl¹*, Amanda L. Whitmire²

1 Center for Digital Scholarship and Services, University Libraries, Oregon State University, Corvallis, Oregon, United States of America, **2** Harold A. Miller Library, Hopkins Marine Station, Stanford University, Pacific Grove, California, United States of America

* These authors contributed equally to this work.

* steve.vantuyl@oregonstate.edu



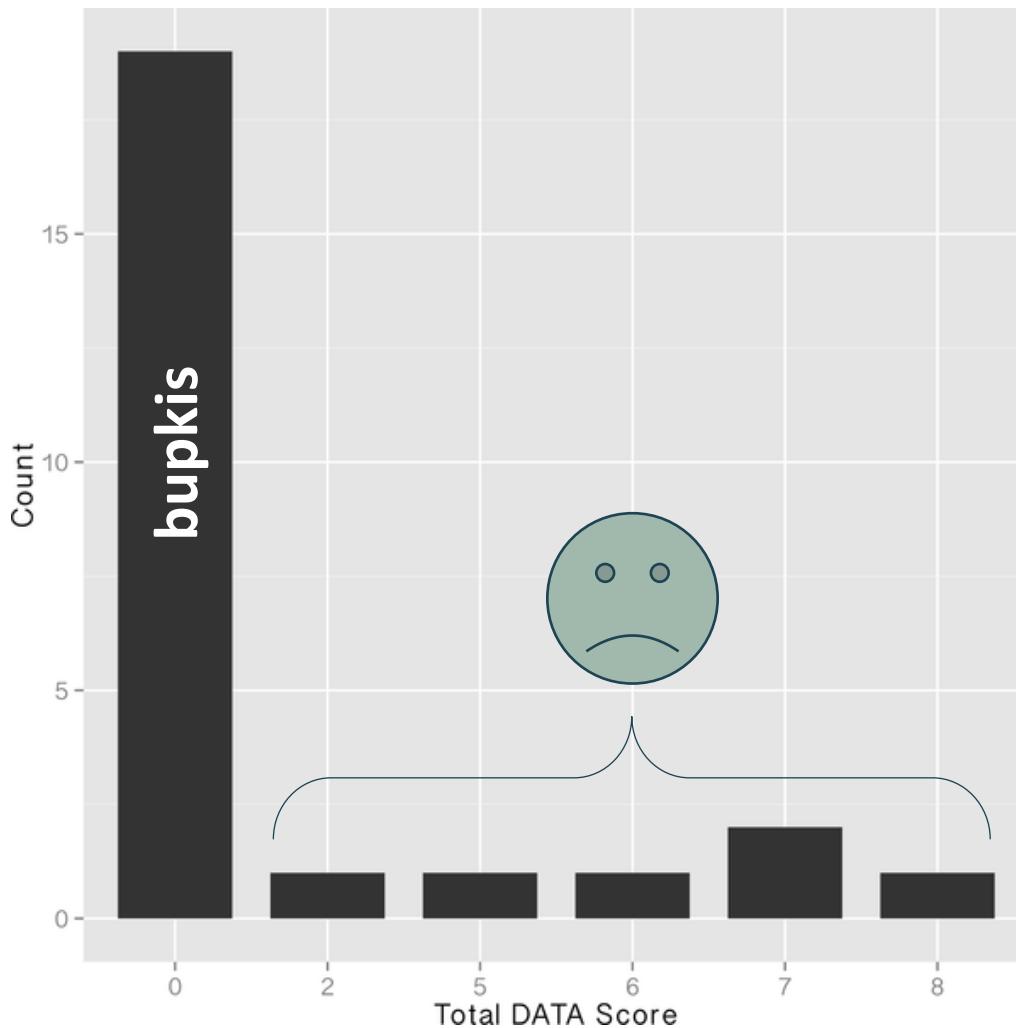
Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. PLoS ONE 11(2): e0147942. doi: [10.1371/journal.pone.0147942](https://doi.org/10.1371/journal.pone.0147942)

Van Tuyl, Steve and Amanda L. Whitmire (2015). Data from: Water, water everywhere: Defining and assessing data sharing in academia. Dataset. Oregon State University Libraries. <http://dx.doi.org/10.7267/N9W66HPQ>

What we did

1. Define criteria for assessing the effectiveness of data sharing
 1. Discoverable?
 2. Accessible?
 3. Transparent?
 4. Actionable?
2. Used DMPs & publications from NSF-funded work to look for associated datasets

Fig 1. Total DATA scores from 25 NSF-funded projects, as located via data management plans



Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. PLoS ONE 11(2): e0147942. doi: 10.1371/journal.pone.0147942

<http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0147942>



**DMP review deepens
our understanding &
allows for more
targeted support.**



A Case Study of DMP Implementation

- Peg Burnette
- Biomedical Librarian
- Social Sciences, Health, and Education Library
- Sarah Williams
- Life Sciences Data Services Librarian
- Funk ACES Library
- Heidi Imker (presenting)
- Director
- Research Data Service
- University Library

Illinois Research Data Service

- Funded by campus administration in 2013 and based at the University Library
- Core staff of 4 FTE + “voluntary” efforts of many others
- Regular interactions with data-related campus groups, e.g. central IT, supercomputing, IRB, security, OVCR, etc.
- Depend on the expertise of our library colleagues for on-the-ground interactions, including DMP reviews and data management consultations

Set-up

- Contacted in 2013 for a DMP review (pre-RDS)
 - Contacted in 2014 when the grant was funded
 - Did one preliminary consultation and based on the questions, pulled in 3 others and followed-up with a report and several “check-point” consultations
-
- **Seemed to be going well! How? Why?**
 - <http://dx.doi.org/10.7191/jeslib.2016.1101>

Theme - Basic Elements

- Communication
- Planning for data management
- Documentation
- Failsafe efforts and checks

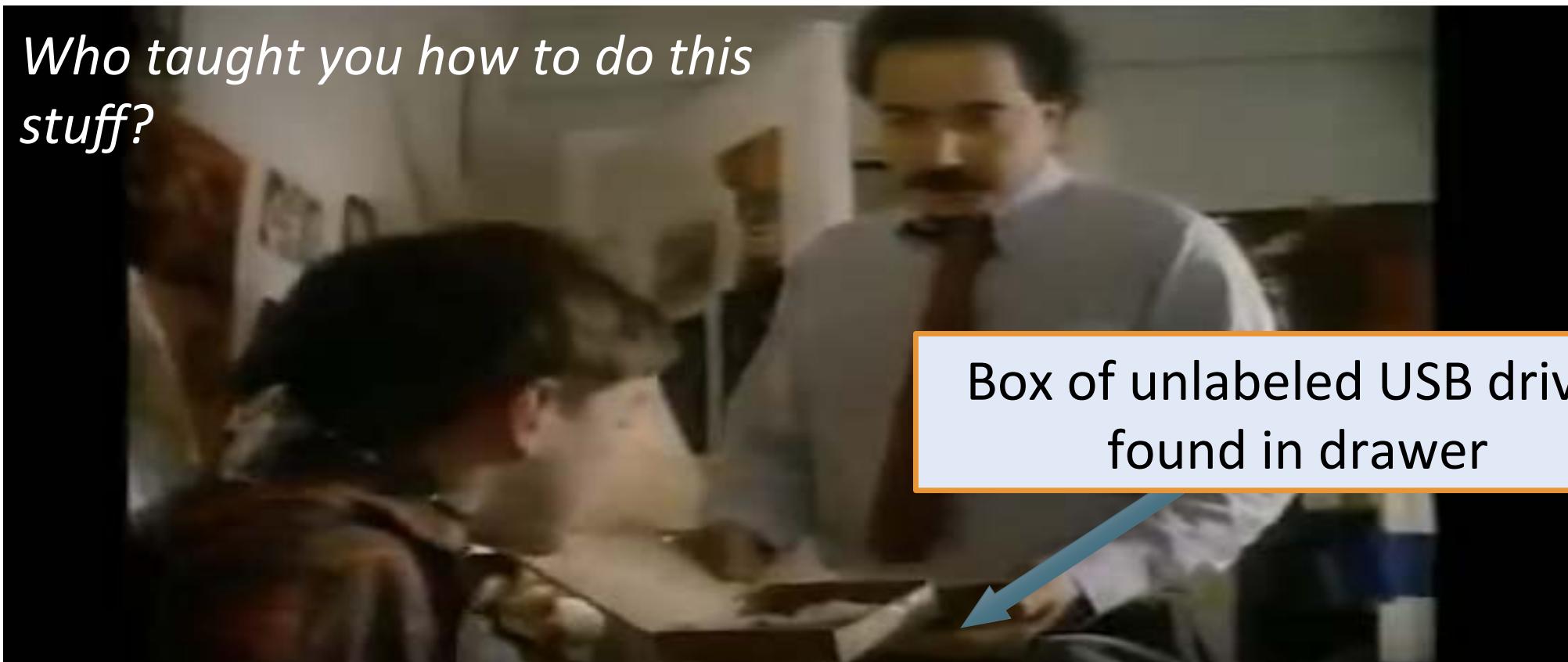
Theme - Staffing

- Hired a project coordinator that served as the data point person
- Effort on setting up protocols and documentation for data management was front loaded, and thereafter was just checking.
- This person served many roles!
 - *Smaller projects (e.g. not enough need or funds for a dedicated project manager), explicitly assign duties and expectations.*

Theme - Data Quality

- Good data is core good science
- Getting better at data management is part of getting better at research itself
- A sort of experiential learning

Theme - Mentoring



<https://www.youtube.com/watch?v=KUXb7do9C-w>

Theme - Mentoring

- Leader came from a back ground in strong data management
 - Co-PIs both expressed that they're learning from colleagues and as they go
 - Project manager actually seemed baffled when I asked if she has trouble getting people to follow data management protocols
- **Creating those expectations and accountability *is* work**

Theme - Peace of Mind

- Expression of emotional distress over possibility of poorly managed data
- High need to be efficient as faculty
- Thought of as an investment with hard-to-measure returns – and that was okay

What have we learned?

- Basic elements are present and core
- Some of the “softer” skills were a little surprising
- Hadn’t occurred to us to think so explicitly about the role of mentoring in data management
 - Can we use mentorship as a way to frame how to set up proactive data management practices?
 - Write into DMPs that researchers on the grant will attend data management training from the RDS.

What have we gained?

- New perspectives
- Validation for what we're trying to accomplish
- Ammo: if you don't believe us – believe *them*, your own colleagues at your own institution. It can be done.

Thank you!

Heidi Imker imker@illinois.edu

Peg Burnette phburn@illinois.edu

Sarah Williams scwillms@illinois.edu



Supporting DMPs: lessons from Europe

Sarah Jones

Digital Curation Centre, Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC

Heavy requirements landscape

- Research Councils and charity funders require DMPs.
- 74% of uni RDM policies also mandate DMPs*



Arts & Humanities
Research Council



BBSRC
bioscience for the future



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

EPSRC
Engineering and Physical Sciences
Research Council



wellcome trust

NERC
SCIENCE OF THE
ENVIRONMENT

* See slides at www.dcc.ac.uk/resources/policy-and-strategy/institutional-data-policies

DMP trends

- Increasing drive towards openness
- DMPs as living documents
- FAIR data management



My plan (Horizon 2020 DMP)

0/9 questions answered
approx. 15% of available space used

Plan details Initial DMP Detailed DMP Final review DMP Share Export

1. Data summary (1 question, 0 answered) +
2. FAIR data (4 questions, 0 answered) +
3. Allocation of resources (1 question, 0 answered) -

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

B I H E S C

Guidance Share note

EC Guidance

Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Costs are eligible for reimbursement during the duration of the project under the conditions defined in the H2020 Grant Agreement, in particular [Article 6](#) and [Article 6.2.D.3](#), but also other articles relevant for the cost category chosen.

Glasgow Uni guidance on Resourcing +

DCC guidance on Responsibilities +

A screenshot of a DMP planning interface for Horizon 2020. The top navigation bar includes 'Plan details', 'Initial DMP' (which is selected), 'Detailed DMP', 'Final review DMP', 'Share', and 'Export'. Below this, three main sections are listed: '1. Data summary' (1 question, 0 answered), '2. FAIR data' (4 questions, 0 answered), and '3. Allocation of resources' (1 question, 0 answered). A progress bar at the top right shows '0/9 questions answered' and 'approx. 15% of available space used'. The 'Allocation of resources' section contains instructions to explain the allocation of resources, address issues like cost estimation and responsibility, and includes a rich text editor toolbar with buttons for bold, italic, headings, etc. To the right, there's a 'Guidance' sidebar with sections for 'EC Guidance' (mentioning grant conditions and Article 6), 'Glasgow Uni guidance on Resourcing', and 'DCC guidance on Responsibilities'.

Basic uni support

- Many unis offer custom guidance (at institutional and in some cases school / department level) as well as example answers

Ethics and Legal Compliance (2 questions, 0 answered)

How will you manage any ethical issues?

Example of answer

It is not envisaged that there will be any ethical or privacy issues with respect to the data as there is no personal data as defined by the Data Protection Act.

B I

Save

Not answered yet

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Example of answer

The research data from this project will be issued under an Attribution Non-commercial Share Alike (by-nc-sa) licence, meaning others can use, adapt and build upon our work non-commercially, as long as they credit us and license their new creations under the identical terms.

B I

University of Strathclyde guidance on Data Security

Outline security and any sharing/access issues relating to data during active research.

Include a link to the [University Information Security](#) pages.

The various University storage platforms, i.e. [Filestore \(H: and I: drives\)](#); Strathcloud-Sharefile; and Sharepoint can only be accessed by *bona fide* staff and students and authorised [collaborators/visitors](#).

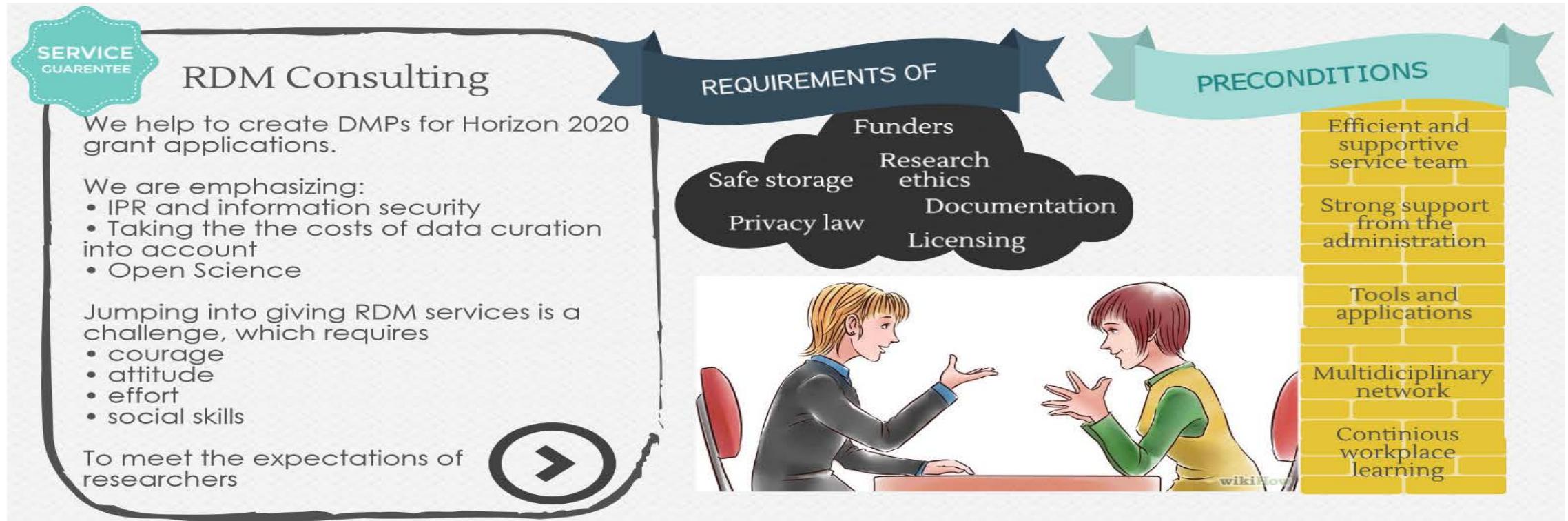
Data stored on the University provided storage is dual sited and replicated between two data centres which are physically separated by several hundred meters. Data links between data centres are provided by dual disparate fabrics, providing added resilience. Additionally the central I.T. service provides tape based backup to a third and fourth site. Data from each of the two data centres is backed up to one or other of the tape backup locations. The University also implements file system snapshots.

Mathematics & Statistics guidance on Data Security

Outline the specific access and security arrangements for data held by Maths & Stats.

In addition to the University's central storage platforms ([Filestore H: and I: drives](#); StrathCloud-Sharefile; and Sharepoint), the Department of Maths & Stats has an in-house server. To gain access and security arrangements for this server, please contact the Maths & Stats IT Officers, [Ronnie Wallace](#) and [Ian Thurlbeck](#).

DMP consultations



Consulting, supporting and networking with researchers & all other interest groups

DMP feedback and review

Rules for using the service:

You need to submit your data management plan to us at least **7 working days before your grant submission deadline**. Otherwise we might be unable to help you with your data plan.

Send us your plan by filling in the form below:

Funder's name *

- Select -

Deadline for the submission of the grant proposal *

Day ▾ Month ▾ Year ▾

Your name *

Your e-mail address *

Please enter your e-mail address carefully – we will use it to get in touch

Upload your draft plan *

Choose file No file chosen

Upload

Submit



UK adoption of DART approach

- Community-led initiative to develop evaluation rubrics based on key funder requirements
- Emphasis on funder specifics to check compliance and provide feedback pre-submission
- Example BBSRC rubric:
- <https://research-data-network.readme.io/docs/bbsrc-dmp-compliance-rubric>

Review functionality in DMP tools

- Revising the review functionality in DMPRoadmap (joint codebase for DMPonline & DMPTool)
- Opting for simpler, less-formal process
- Use of substance editor to annotate text directly
- Share what you need!

roadmap

Collaboration with research offices

- Research offices play a key role as the first point of contact when PIs are preparing grants
 - Include links to DMP tools in research office mailings e.g. replies to costing requests
 - RO provides list of new awards so RDM team can contact PIs about DMPs at outset
 - Hot-desking / co-location so teams work more closely and share expertise
 - Collaboration on costing data management

Integrating DMPs into workflows

Example of embedding flags into the grant costings system at the University of Leicester

Modify Costing No. 9 (Application Version)

Research Data Management

IMPORTANT: The University and funding bodies require that researchers demonstrate that they have planned for the management of research data over the project life cycle. Please indicate if:

1. Your research may use or generate research data that is sensitive or confidential (including Commercial in Confidence)? Yes

2. The project will require advice in planning and costing your IT and research data management requirements (including secure and reliable storage) either before, during or after completion? Yes

Responding 'Yes' to the above questions will be logged with IT Services for the appropriate team to make contact. Please tick this box if you wish to hide the title of the research project in this notification Hide Title

Please contact the [IT Academic and Research Liaison team](#) for advice in good time before submission.

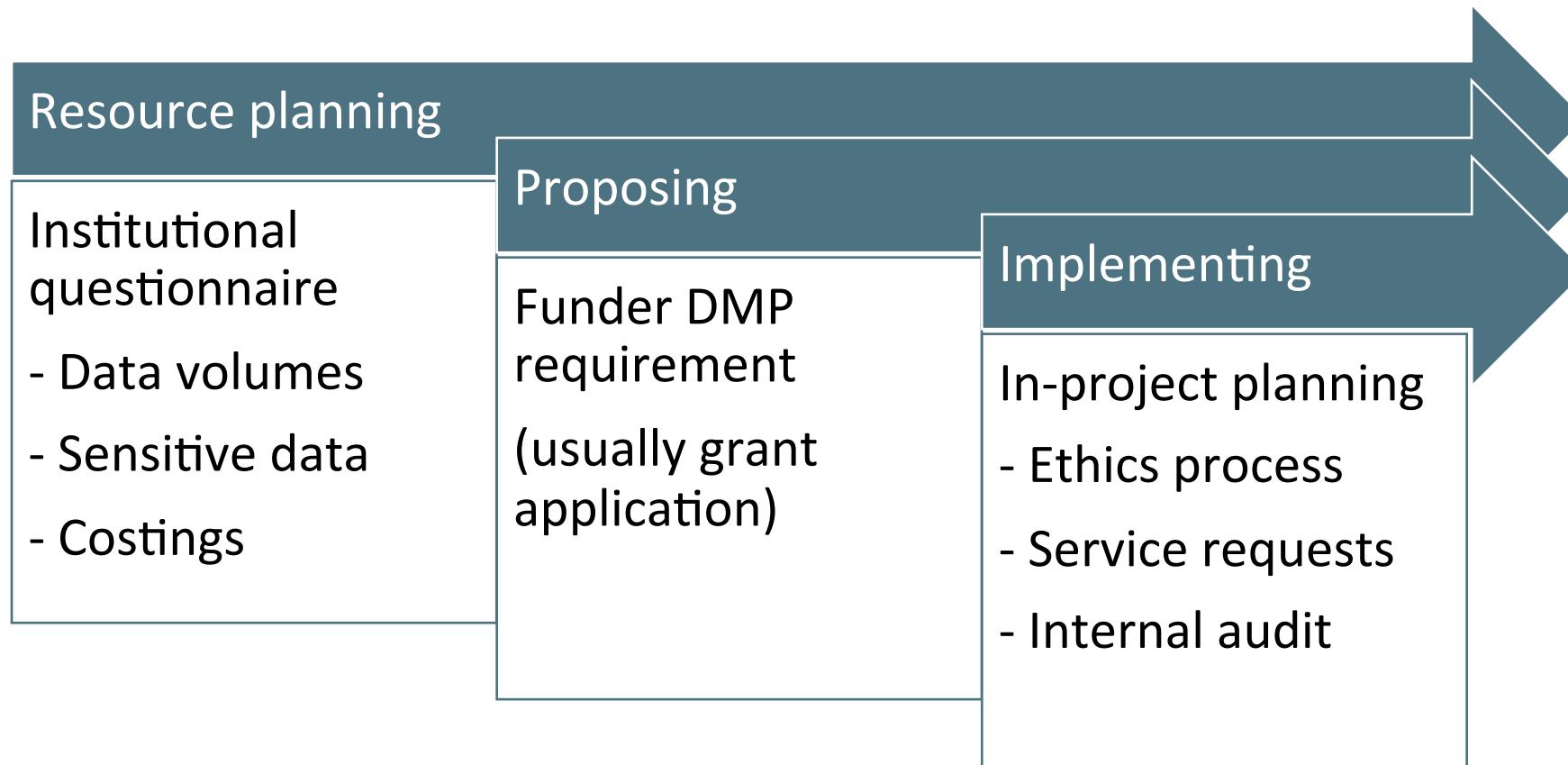
Continue Cancel

DMPs to define / allocate storage

- The University of Manchester requires an outline DMP **prior** to a grant application
- This asks questions about the storage requirements to plan resourcing and allocate space
- The outline DMP generates an RDM Plan Reference Number to include in the Research Application form. Proposals can't proceed without this.
- www.library.manchester.ac.uk/services-and-support/staff/research/services/research-data-management/data-management-planning-tool

Desire for pre- or post- phases

Idea to blend institutional and funder requirements, ensure costings are included and plans implemented



Thanks for listening

- DCC resources on DMPs:
 - [www.dcc.ac.uk/resources/
data-
management-plans](http://www.dcc.ac.uk/resources/data-management-plans)
-
- Follow us on twitter:
 - @DMPonline and #ukdcc