DataONE Webinar Series

Analyzing, interpreting, and implementing data management plans

Amanda Whitmire

Heidi Imker

Sarah Jones

DataONE

DataONE

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

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 (but evolving) document in which you lay out a pla
the present

- It saves t
the plan
keep eve
- It increa
and othe
the futur
- Prevents
vious wh
(and hov
- Makes it
later pre
- It may b

Componen

- Descript
- How dat
- How dat
- File form
- Quality a
used du
cessing

Metadata is
contains co
and using t
details, inst

Component 2: Metadata content & format

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

DataONE

Lesson 2: Data Sharing

View all Education Modules at https://www.dataone.org/education-modules

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 formats & on media good for long term
Publish info
discover it e

Inte

Concerns about data sharing

Concern	Solution

To the sci
- Recei
autho
- Impro
- Creat
- Impro
poten

To the res
- Enhan
ing th

To the res
- Better
repea
- Abilit
- Increa
comp
- Abilit
recom
- Better
most

To the pu
- Bette
under
effect

DataONE

Lesson 1: Data Management

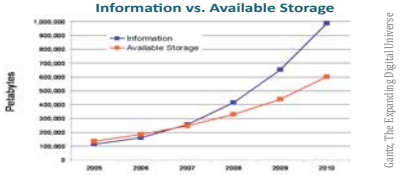
View all Education Modules at https://www.dataone.org/education-modules

The world of data around us

The data deluge has created a surge of information that needs to be well-managed, discoverable, and accessible.

The amount of available storage is not keeping pace with the amount of data being produced.

Information vs. Available Storage



Gantz, The Expanding Digital Universe

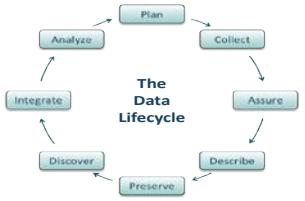
Causes of data loss

- 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 insitutional 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.



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 understability and reuse
- Prepare it to share it & gain credibility and recognition for your scientific efforts

Data management facilitates sharing and reuse.

Data Reuse Example

Researchers reused and aggregated data from several different sources to determine migration routes for specific bird species.

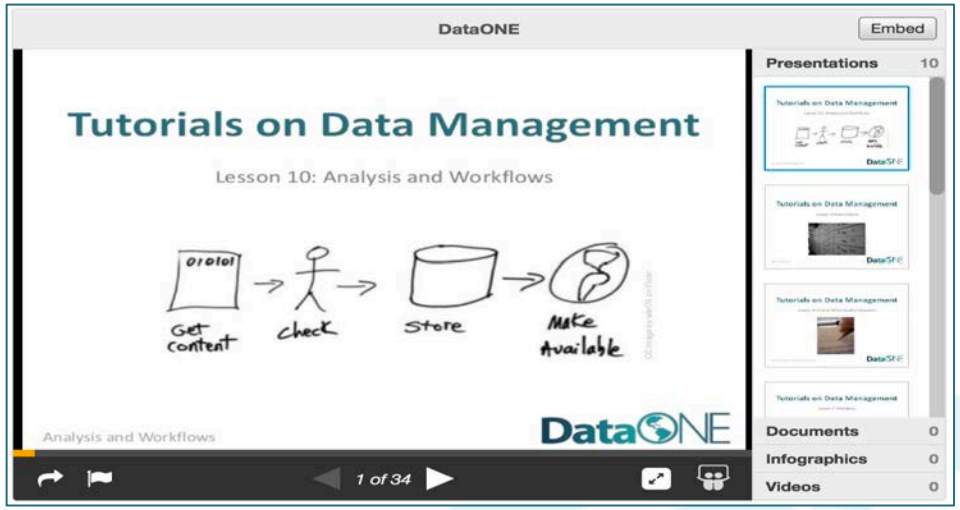
© Jerry & Sherry Liguori

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



#DWS2016
@DataONEorg
www.DataONE.org

DWS

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

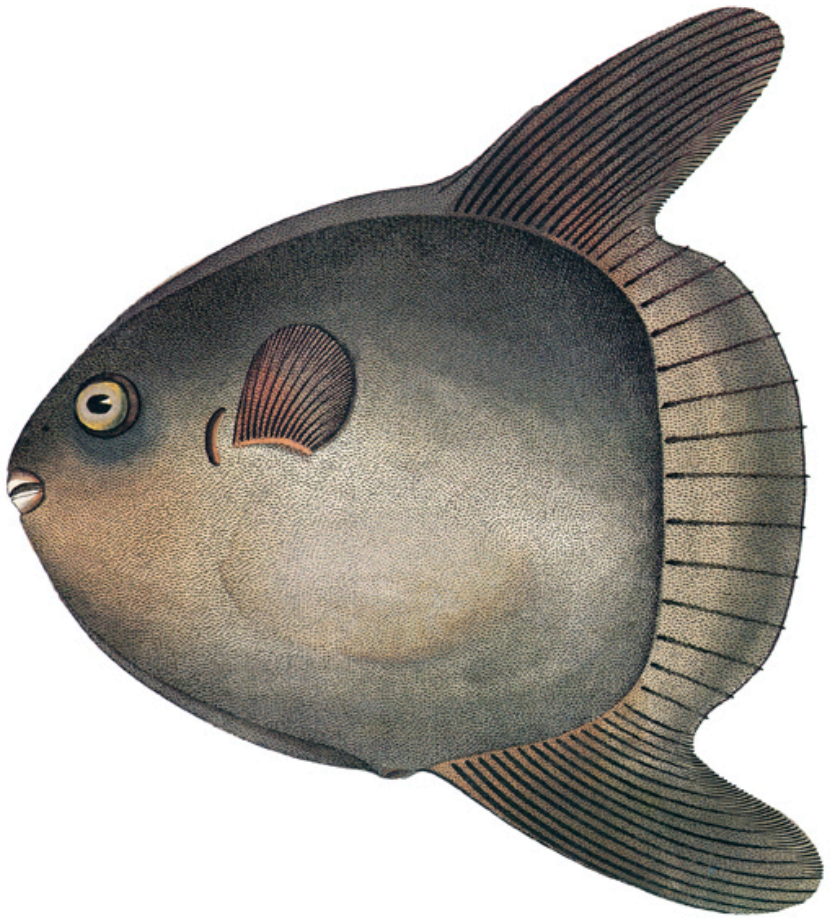
www.dataone.org/upcoming-webinar

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

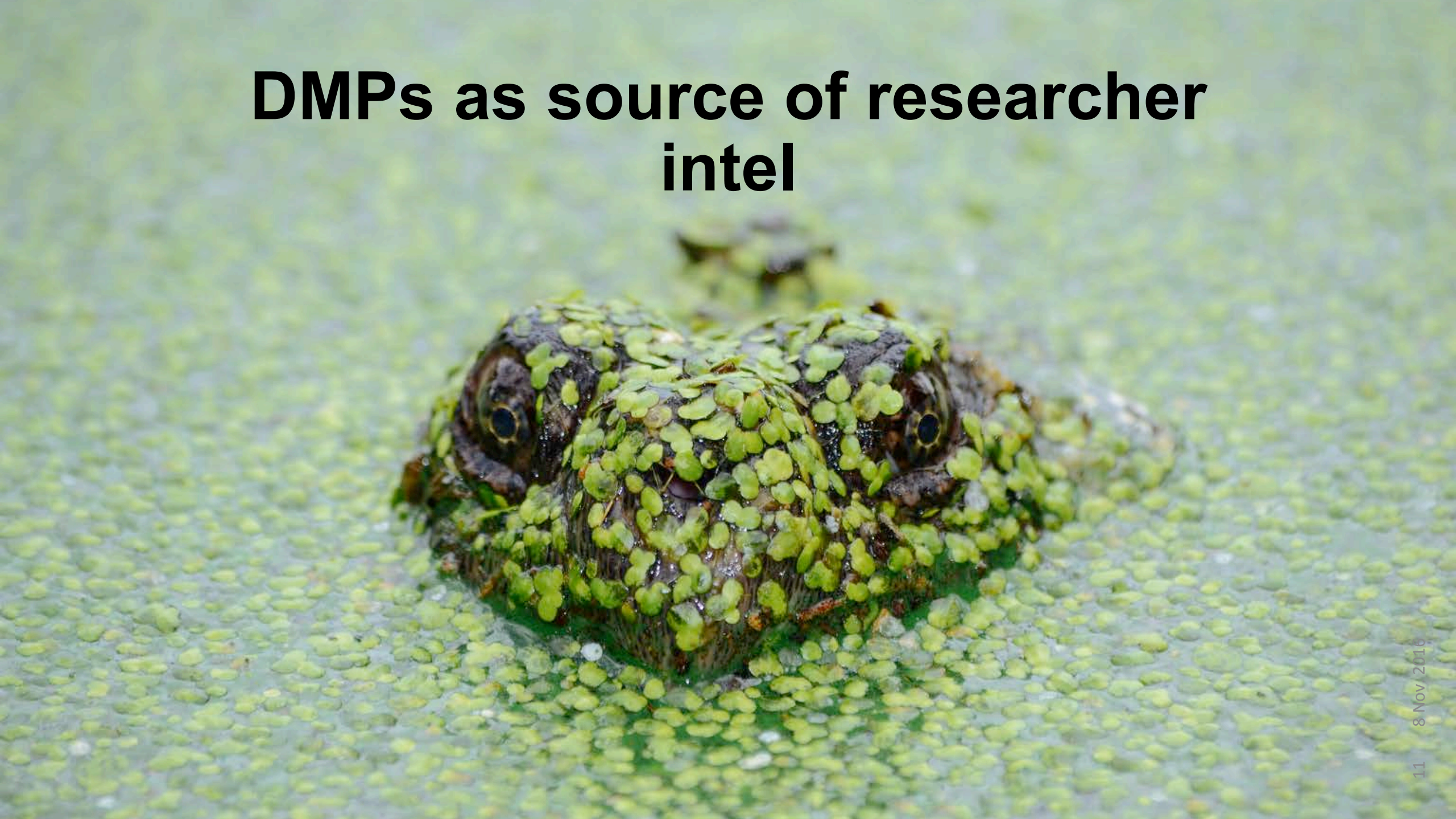
What are DMPs good for?

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)

DART Team

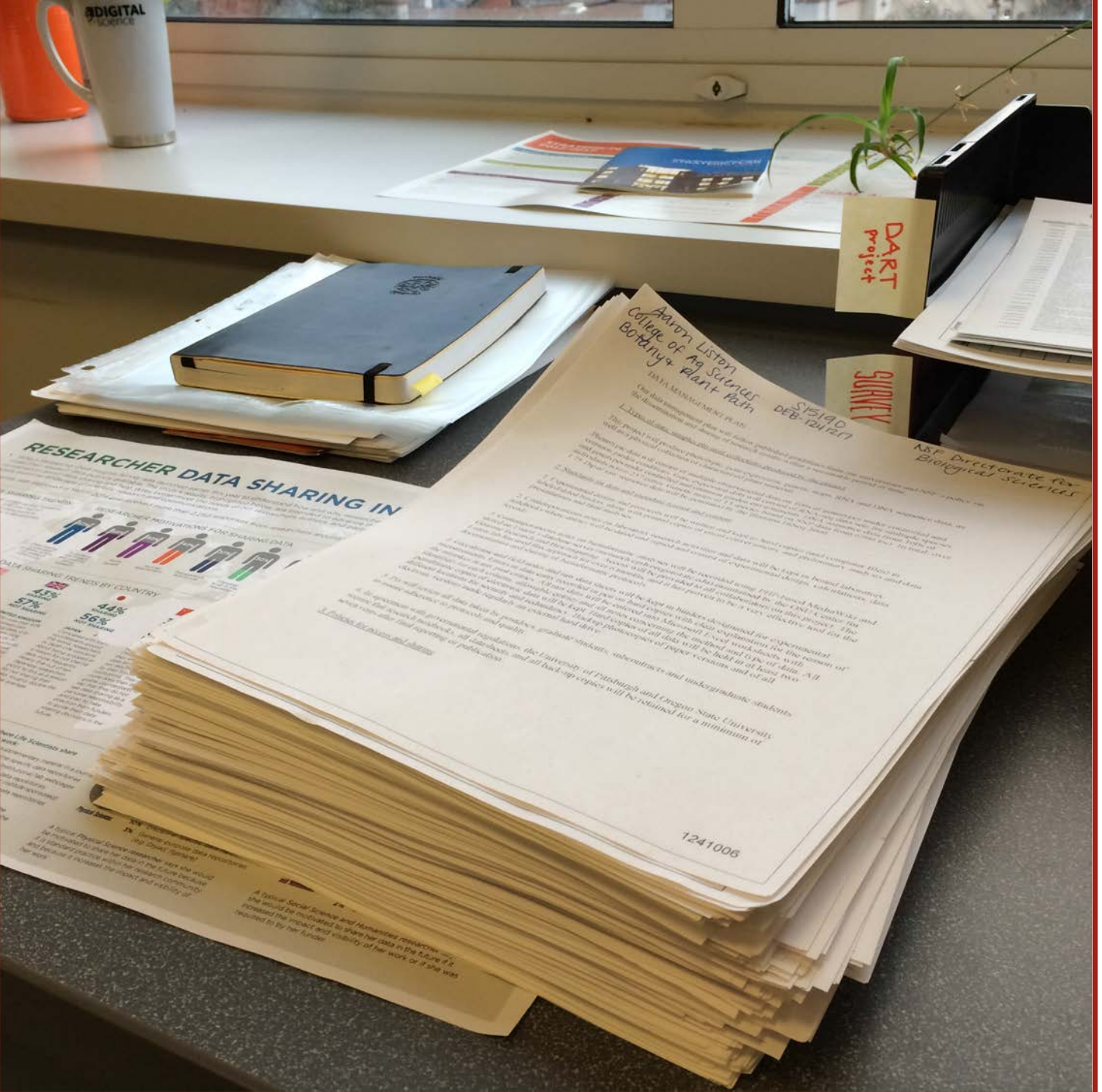
Amanda Whitmire | *Stanford University Libraries*
Jake Carlson | *University of Michigan Library*
Patricia M. Hswe | *Update*
Lizzy Rolando | *MailChimp*
Susan Wells Parham | *Georgia Institute of Technology Library*
Brian Westra | *University of Oregon Libraries*

INSTITUTE of Museum and Library SERVICES

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

@DMPResearch

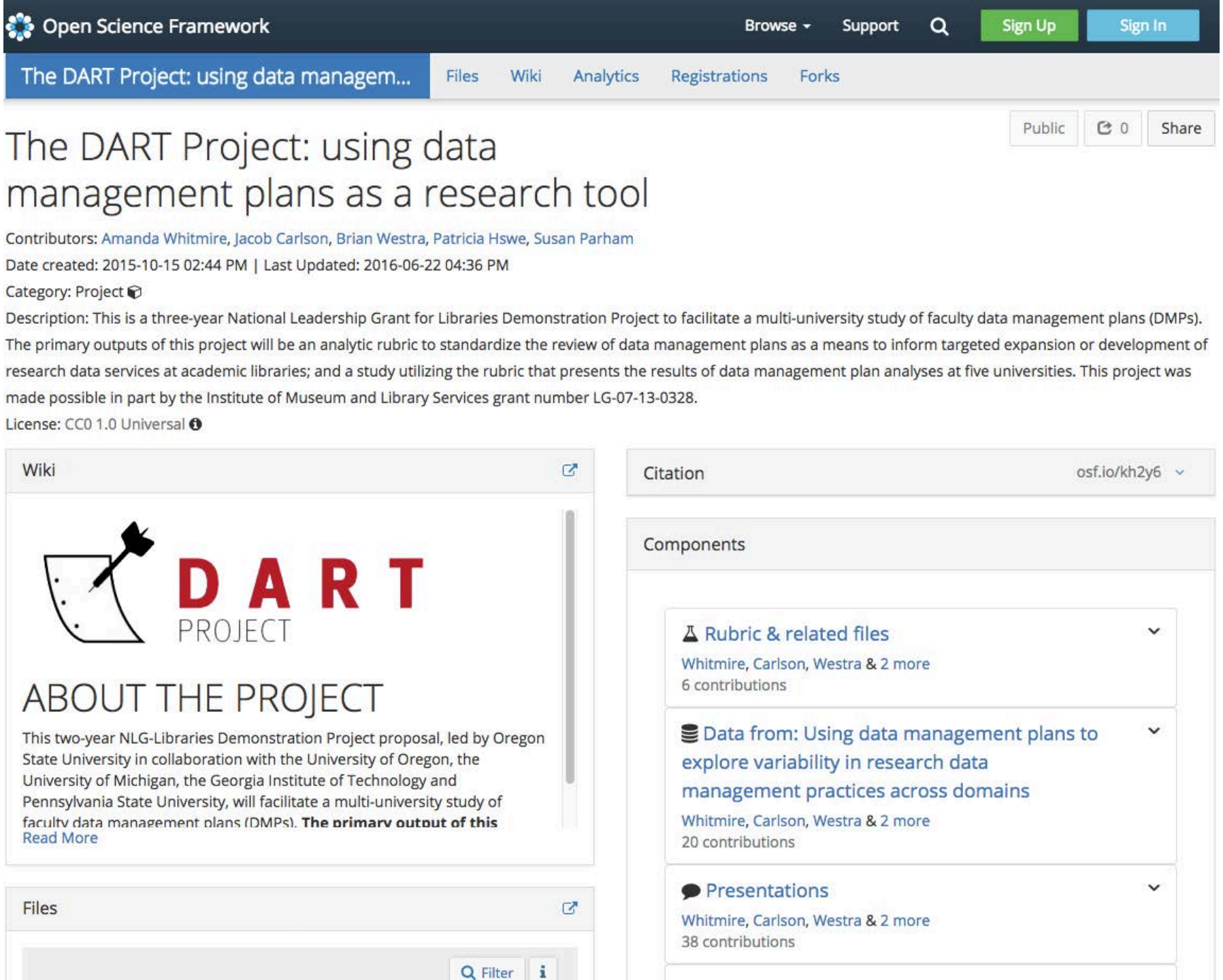
		Performance Level			
	Performance Criteria	**Complete / detailed**	**Addressed issue, but incomplete**	**Did not address issue**	**Directorates**
General Assessment Criteria	*Describes what types of data will be captured, created or collected*	Clearly defines data type(s). *E.g. text, spreadsheets, images, 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*	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
Directorate- or division-specific assessment criteria	*Describes how data will be collected, captured, or created (whether new observations, results from models, reuse of other data, etc.)*	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
	Identifies how much data (volume) will be produced	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



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

Susan Wells Parham et al., 'Using Data Management Plans to Explore Variability in Research Data Management Practices across Domains', *International Journal of Digital Curation* 11, no. 1 (10 May 2016): 53–67. doi:10.2218/ijdc.v11i1.423

Describes how data will be shared

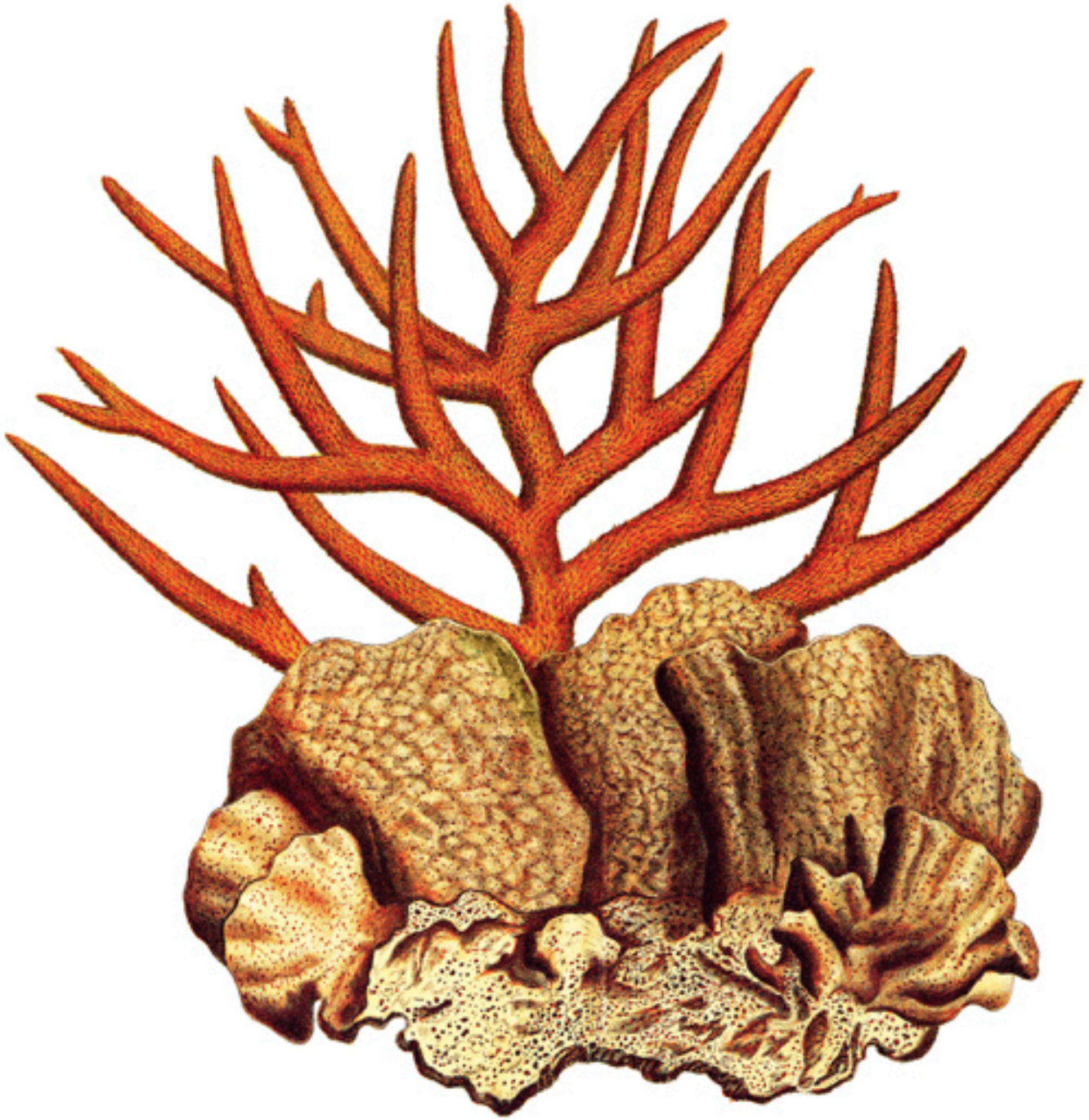
	Complete/ detailed	Addressed issue, but incomplete	Did not address	Scale
All	50	41	9	70
BIO	67	31	2	60
CISE	41	44	15	50
ENG	37	53	10	30
GEO	61	35	4	20
MPS	48	44	8	10
SBE	58	34	8	0

Susan Wells Parham et al., 'Using Data Management Plans to Explore Variability in Research Data Management Practices across Domains', *International Journal of Digital Curation* 11, no. 1 (10 May 2016): 53–67. doi:10.2218/ijdc.v11i1.423

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

Susan Wells Parham et al., 'Using Data Management Plans to Explore Variability in Research Data Management Practices across Domains', *International Journal of Digital Curation* 11, no. 1 (10 May 2016): 53–67. doi:10.2218/ijdc.v11i1.423



What is going on with Biology?

They have infrastructure!

BIO: Repositories mentioned (frequency)

GenBank (14)
Dryad (12)
SRA (11)
iDigBio (3)
Knowledge Network for Biocomplexity (3)
MorphBank (3)
NCBI (3)
TreeBASE (2)

-Question-

PLOS ONE

RESEARCH ARTICLE

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

Steven Van Tuyl[1☯]*, **Amanda L. Whitmire**[2☯]

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

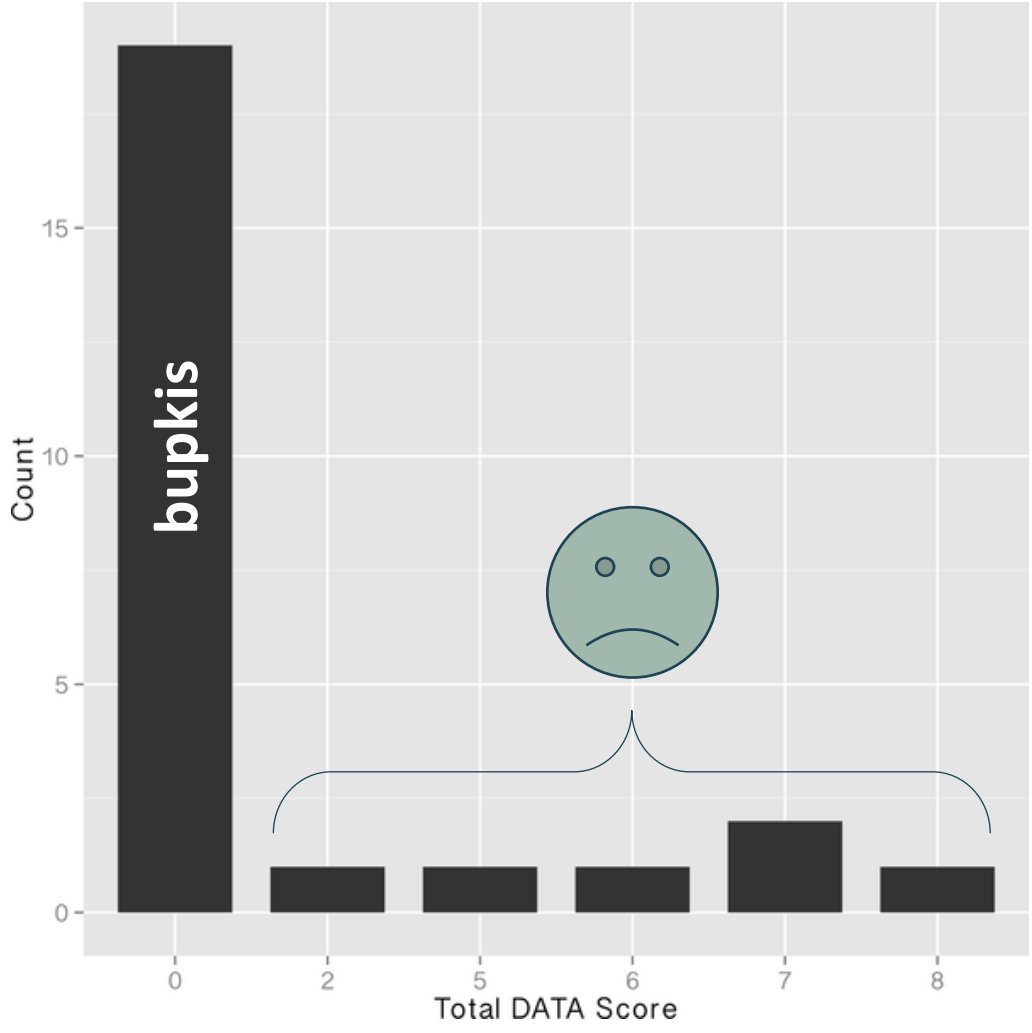
Van Tuyl, Steve and Amanda L. Whitmire (2015). Data from: Water, water everywhere: Defining and assessing data sharing in academia. Dataset. Oregon State University Libraries. http://dx.doi.org/10.7267/N9W66HPQ

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

RESEARCH DATA SERVICE

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

RESEARCH DATA SERVICE

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

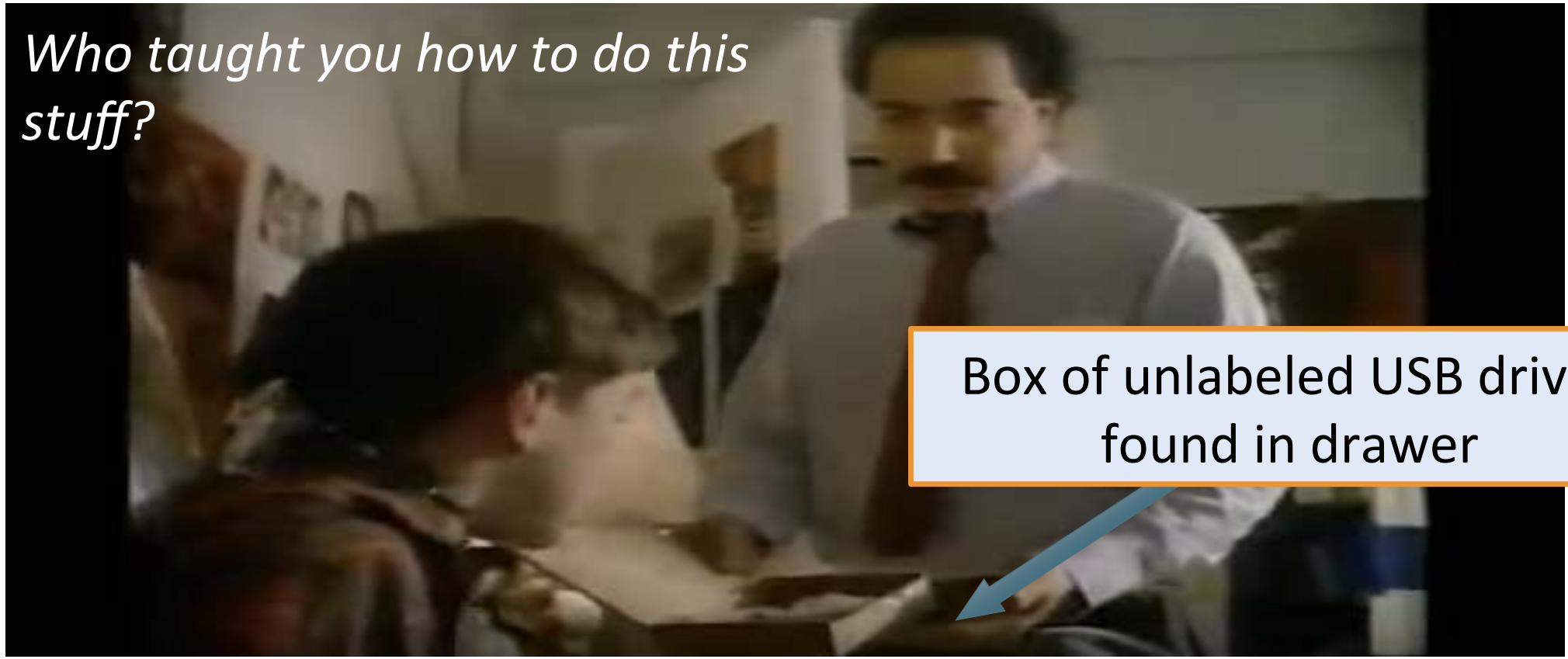
- Hired a project coordinator that served as the data point person
- Effort on setting up protocols and documentation for data management was front loaded, and thereafter was just checking.
- This person served many roles!

➢ *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

RESEARCH DATA SERVICE

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

RESEARCH DATA SERVICE

Supporting DMPs: lessons from Europe

Sarah Jones

Digital Curation Centre, Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC

Analyzing, interpreting & implementing DMPs, DataOne webinar, 8 November 2016

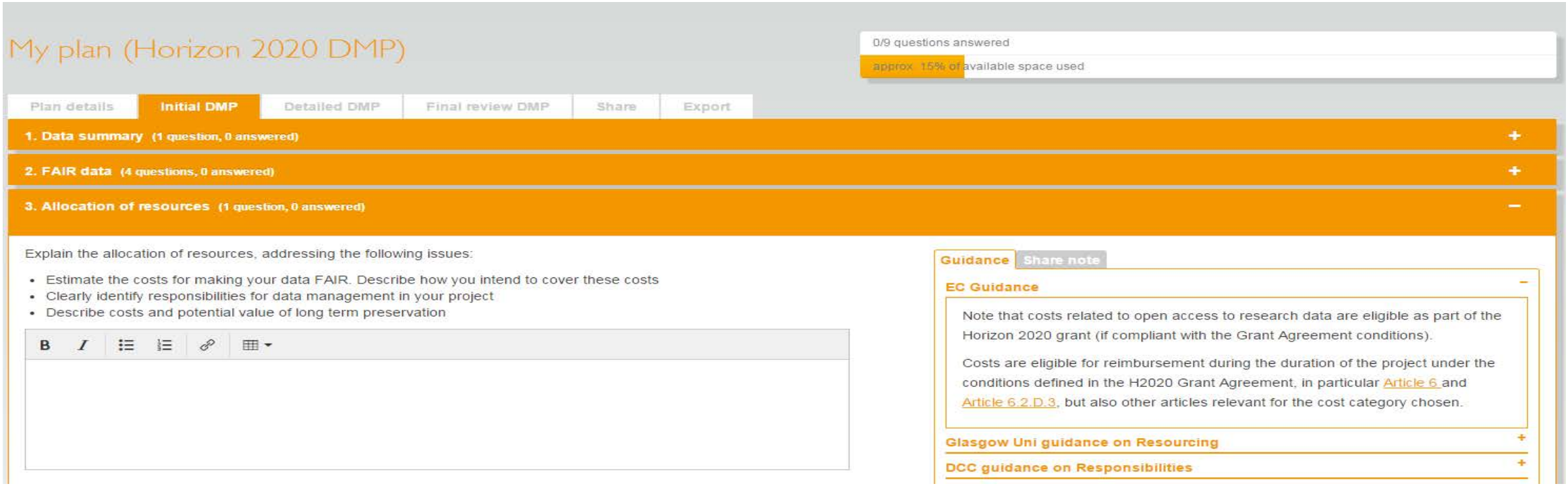
Heavy requirements landscape

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

- sis at www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies

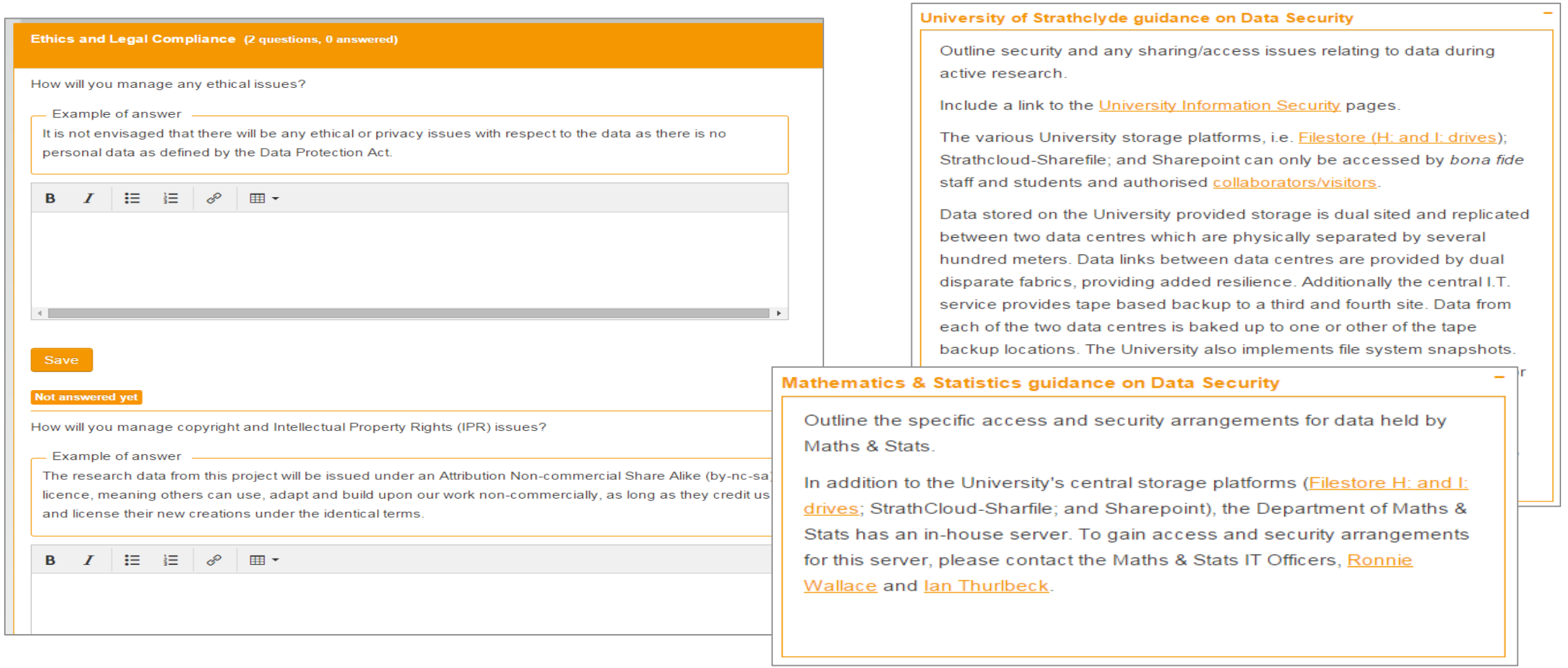
DMP trends

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

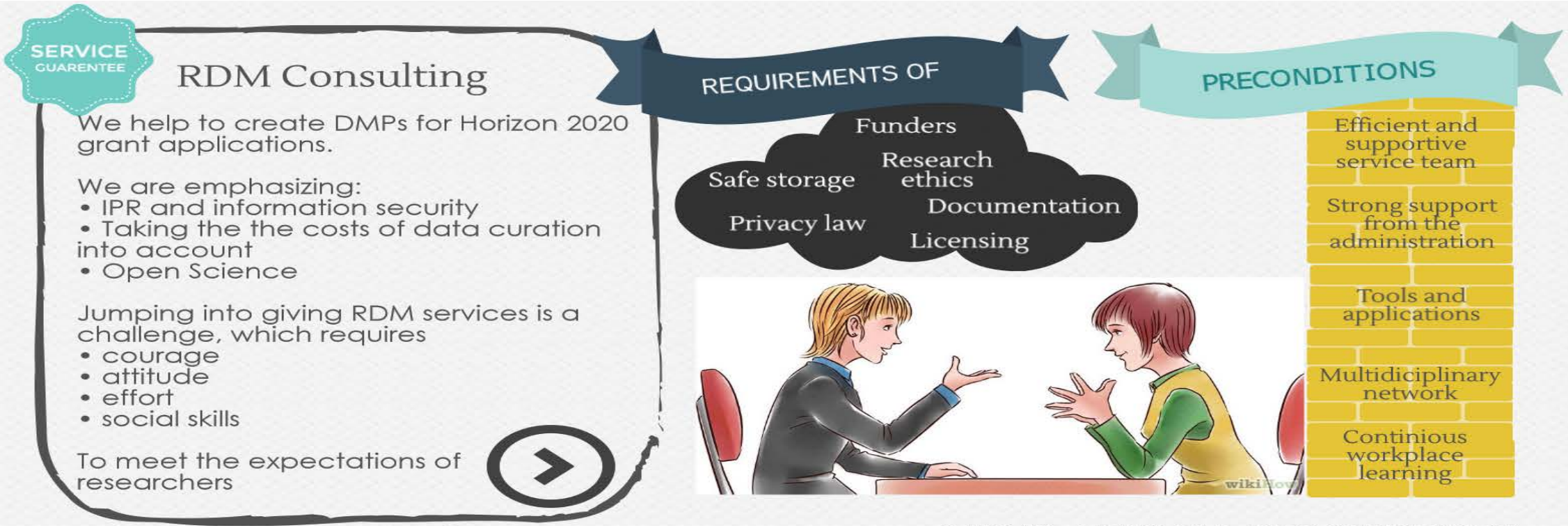


Basic uni support

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



DMP consultations



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

Slide content courtesy of Mari Elisa Kuusniemi (MEK), University of Helsinki Library

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

www.data.cam.ac.uk/DMPsupport

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

DART PROJECT

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

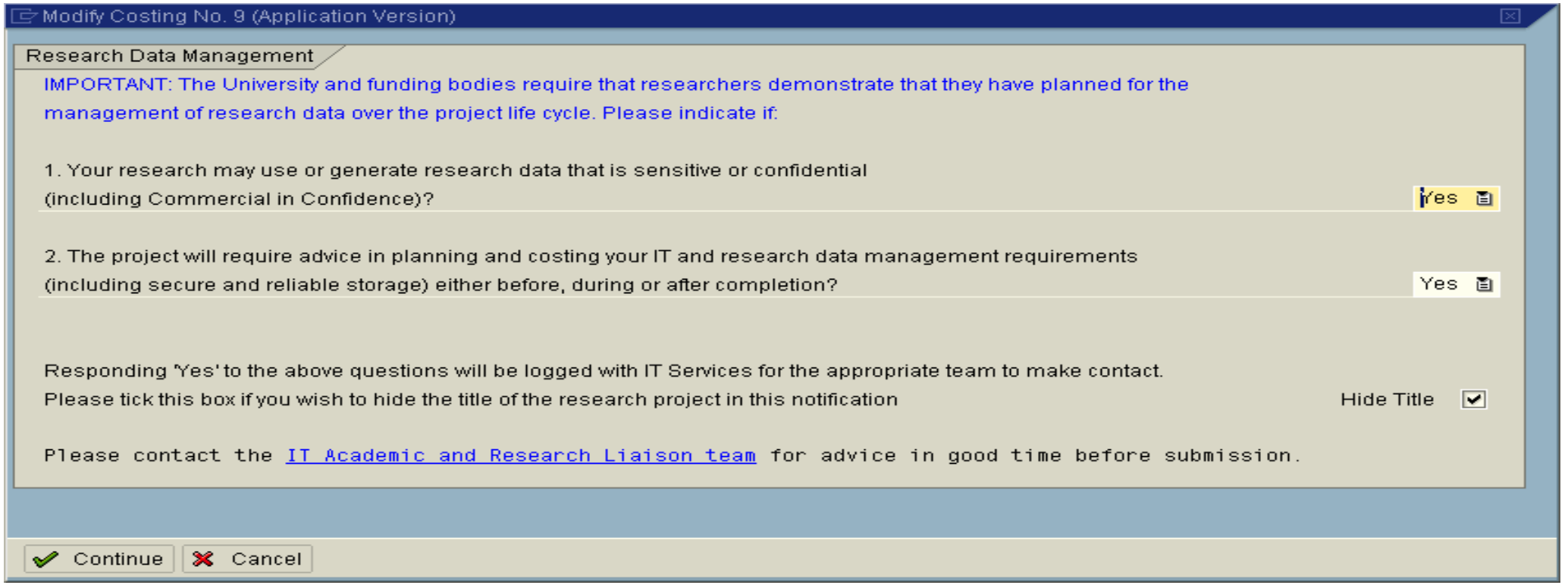
- RO provides list of new awards so RDM team can contact PIs about DMPs at outset

- 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



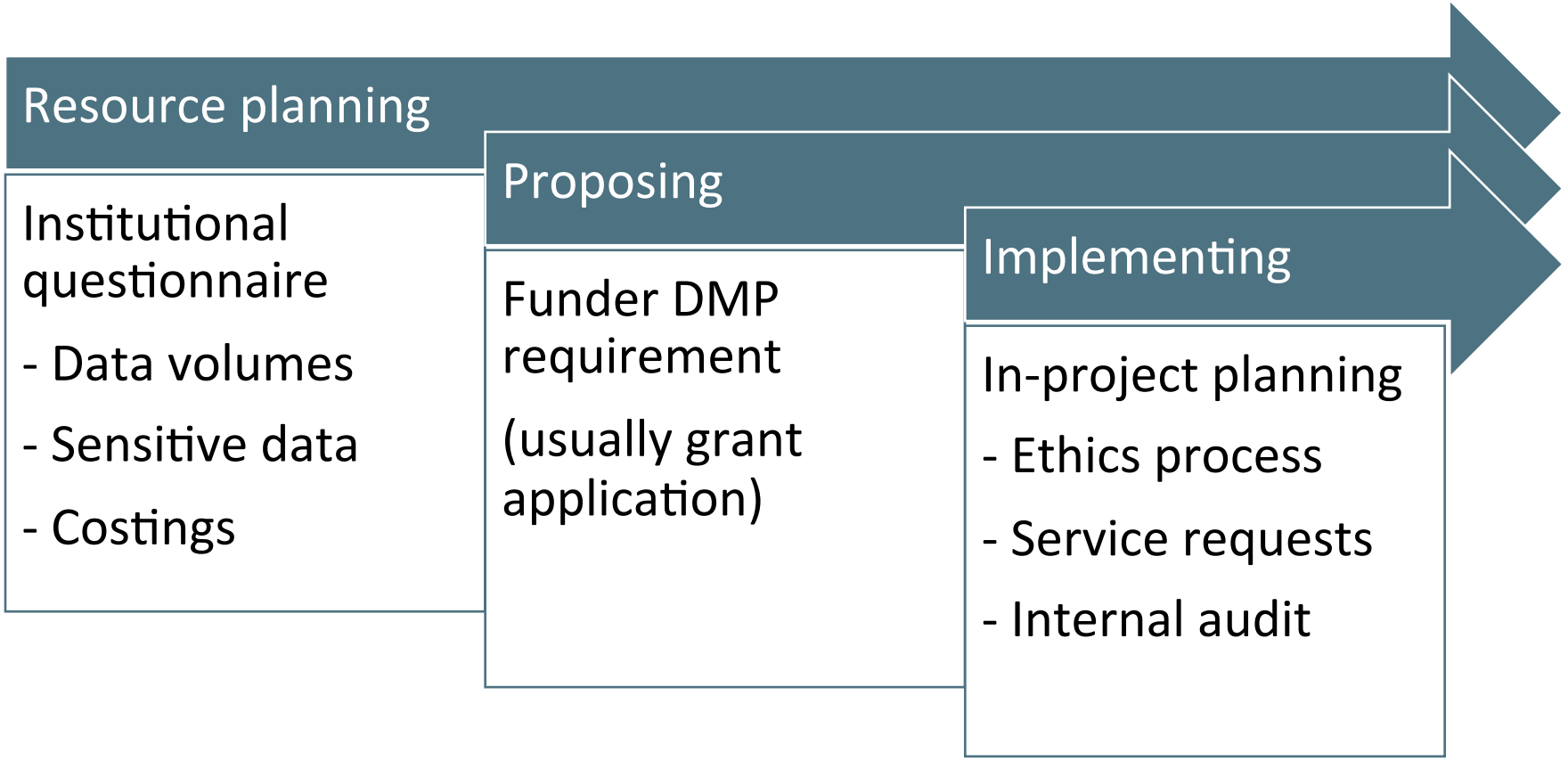
www2.le.ac.uk/offices/itservices/about/news/old-news/2010/July/lucre-08-07-10

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

- Follow us on twitter:
- @DMPonline and #ukdcc

because good research needs good data