

# DataONE Webinar Series

Incentives, Challenges, Barriers:  
Exploring social, institutional and  
economic reasons for sharing data

Jens Klump

Heidi Laine

Fiona Murphy



# 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](http://www.dataone.org/current-member-nodes)*

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

### 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 in which you lay out a plan for the present

#### Component 2: Metadata content & format

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

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 used for long-term

Publish information to discover it

#### Concerns about data sharing

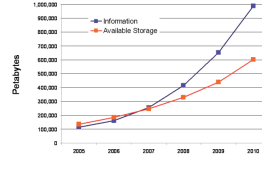
Concern	Solution

### 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



Source: The Expanding Digital Universe

### 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



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.



### 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

#### To the scie

- Recei
- autho
- Impr
- Great
- Impr
- poten

#### To the res

- Enhanc
- ing th

#### To the res

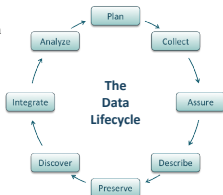
- Better repea
- Ability
- Incre
- comp
- Ability recom
- Better most

#### To the pu

- Better unde
- effect

### The Data Lifecycle

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




#DWS2016  
@DataONEorg  
[www.DataONE.org](http://www.DataONE.org)



*DataONE Webinar Series*

***[www.dataone.org/webinars](http://www.dataone.org/webinars)***

*Upcoming Webinar Event*

***[www.dataone.org/upcoming-webinar](http://www.dataone.org/upcoming-webinar)***

*Previous Webinar Events (Recording and Discussion)*

***[www.dataone.org/previous-webinars](http://www.dataone.org/previous-webinars)***



***#DWS2016***

***@DataONEorg***



*Upcoming Webinar Event*

*[www.dataone.org/upcoming-webinar](http://www.dataone.org/upcoming-webinar)*


## **How not to collect data: organizing data for long-term use and re-use**

**October 11**

**Stephanie E. Hampton**

Washington State University



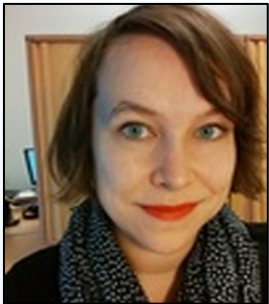


If you attending  
as part of a group,  
please enter the number of  
people listening within  
the “questions” box  
Thanks!

# Incentives, Challenges, Barriers: Exploring social, institutional and economic reasons for sharing data



**Jens Klump**  
CSIRO



**Heidi Laine**  
University of Helsinki



**Fiona Murphy**  
University of Reading







# Data as Social Capital and the Gift Culture in the Scholarly Community

Jens Klump | OCE Science Leader Earth Science Informatics  
13 September 2016

MINERAL RESOURCES  
[www.csiro.au](http://www.csiro.au)



# Why am I interested in data sharing?

---

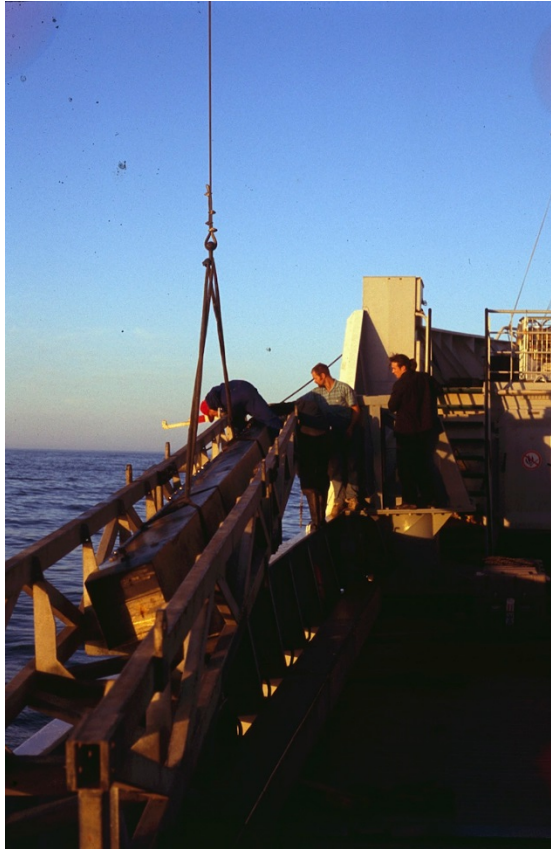


**I am not a  
sociologist.**

- I am a geochemist, my field of research is Earth Science Informatics.
- Research data infrastructures have been part of my work since 1999.
- When I switched from marine geology to limnology I was puzzled by the difference in attitudes towards data sharing.
- Over the years I made some observations in this respect.

# Why do communities behave so differently?

---



South Atlantic (Namibia)



Lake Baikal (Russia)



# Are we getting the incentives right?

---

Image: British Library (public domain)

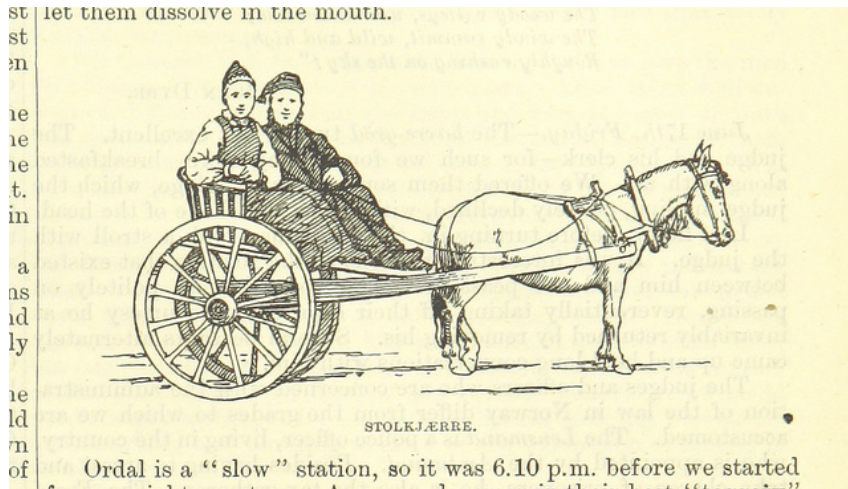


Image: ETH Zürich Library (public domain)

# Waiting at the watering hole

---

- Sometimes waiting at the watering hole can be a successful strategy.
- The art is to identify suitable watering holes.
- Which resources do researchers need to access for their distinction gain?
- This is not only an opportunity to coerce compliant behaviour but also to develop better services for researchers.

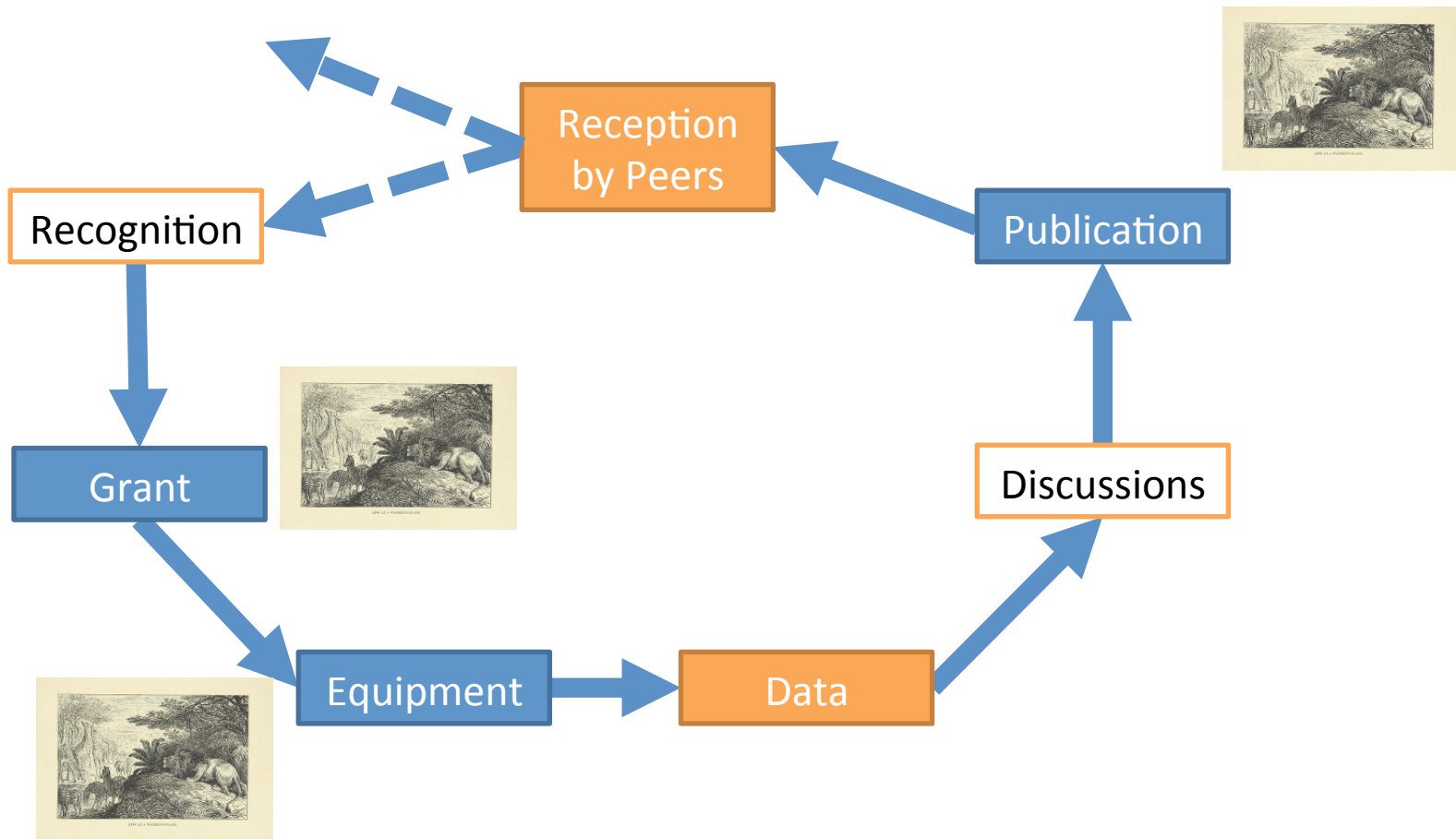


LION AT A WATERING-PLACE.

Image: British Library (public domain)

# Reputation Economy

---



## Mineral Resources

Jens Klump

Science Leader Earth Science Informatics

**t** +61 8 6436 8828

**e** [jens.klump@csiro.au](mailto:jens.klump@csiro.au)

**w** [www.csiro.au/en/Research/MRF](http://www.csiro.au/en/Research/MRF)

**MINERAL RESOURCES**

[www.csiro.au](http://www.csiro.au)



# AFRAID OF SCOOPING?

## Case Study on Perceived vs. Actualized Risks of Sharing Research Outputs

HEIDI LAINE / [heidi.laine@helsinki.fi](mailto:heidi.laine@helsinki.fi) /  
@heidiklaine

DataONE



This work is licensed under a

[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).





# SCOOPING AS MISCONDUCT

---

**Misappropriation:** unlawful presentation of another person's result, idea, plan, observation or data as one's own research

- Finnish Advisory Board on Research Integrity: *Responsible conduct of research and procedures for handling allegations of misconduct in Finland - RCR guidelines*, available at <http://www.tenk.fi/en/responsible-conduct-research-guidelines>

# RESEARCH QUESTIONS

---

**Descriptive level:** How two pioneering open collaboration research projects coped with the risk of research misconduct?

**Interpretive level:** What allowed the researchers of the two case projects to ignore the risk of research misconduct?

# OPEN COLLABORATION CASES

Open Research Swarm / Social Media for Citizen Participation (SOMUS), 2007-2010



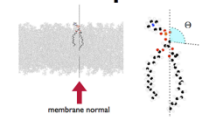
NMR Lipids Project (NMRLP), 2013-ongoing

Open collaboration that uses **NMR** data to judge the correctness of phospholipid glycerol and head group structures in **MD** simulations

[nmrlipids.blogspot.fi](http://nmrlipids.blogspot.fi) — Patrick F. J. Fuchs<sup>1,2</sup>, Matti Javanainen<sup>3</sup>, Antti Lamberg<sup>1</sup>, Markus S. Miettinen<sup>1</sup>, Luca Monticelli<sup>4</sup>, Jukka Määttä<sup>5</sup>, O. H. Saruuli Ollari<sup>6</sup>, Marius Reitegaard<sup>6</sup>, Hubert Santuz<sup>2,7</sup>

**Abstract** We compare the C-H order parameters measured in Nuclear Magnetic Resonance (NMR) experiments to those predicted by 13 different molecular dynamics (MD) simulation models. We focus on the order parameters of the lipid headgroups and glycerol backbones in phospholipid bilayers. For a fully hydrated bilayer, only three models show reasonable agreement with experiment: CHARMM36, Nucleoside-Rig, and GROMACS. We then compare for these three models and the Berger model, the most used lipid model in the literature, to NMR experiments the changes in the order parameters as a function of hydration level and cholesterol content. The results clearly show that the glycerol and headgroup structures in the Berger model are not realistic, and the cholesterol-induced structural changes are over-estimated. CHARMM36, Nucleoside-Rig and GROMACS perform somewhat better. This research is openly progressed at [nmrlipids.blogspot.fi](http://nmrlipids.blogspot.fi), all the results and discussions are available at that address.

## C-H order parameter

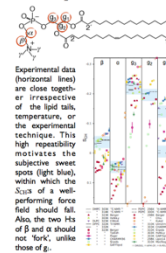


Distribution  $P(\theta)$  of the angle  $\theta$  between membrane normal and a given C-H vector, can be calculated from MD and is directly related to the C-H order parameter

$$S_{CH} = \frac{3}{2} (\langle \cos^2 \theta \rangle - \frac{1}{3})$$

that can be accurately measured with various NMR experiments. A valid MD model would be able to reproduce all the  $S_{CH}$  — and in return it would provide an accurate atomistic resolution interpretation of the NMR experiment.

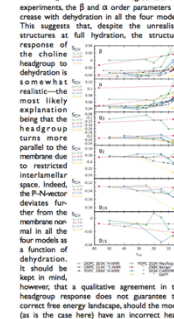
## Full hydration



Rough ranking of MD models through carbon-wise assessment of their ability to reproduce experimental  $S_{CH}$ . 'E' indicates a magnitude problem, 'F' a forcing problem. Letter size shows the level (0-4) of severity; the  $\Sigma$  column shows the sum of these, i.e. the 'total severity'.

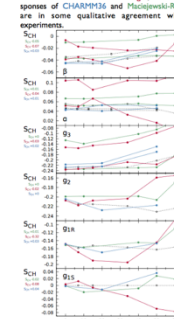
Model	PC	PE	PS	PI	PI1	PI2	PI3	PI4	PI5	PI6	PI7	PI8	PI9	PI10	PI11	PI12	PI13	PI14	PI15	PI16	PI17	PI18	PI19	PI20	$\Sigma$	
CHARMM36																									0	
Nucleoside-Rig																										0
GROMACS																										0
Berger	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	20	

## Dehydration



It should be kept in mind, however, that a qualitative agreement in the headgroup response does not guarantee the correct free energy landscape, should the model (as in the case here) have an incorrect headgroup structure.

## Cholesterol



We find it to be quite interesting that although the glycerol backbone dihedral potentials of CHARMM36 have been fitted to reproduce the correct order parameters at full hydration, this does not seem to have led to overfitting, which would manifest as wrong responses to dehydration or increasing cholesterol content.

# RADICAL OPENNESS

---

<b>ORS / SOMUS</b>	<b>NMRLP</b>
Open collaborative writing of research proposal to a funding call	Participants required to give credit for ideas also from “hallway” discussions
Funding allocated to an open online community	Data shared via Github without embargo
Open collaborative writing of conference papers, meeting notes, wiki	Co-authorship determined by input through blog comments

# LESSONS TO BE LEARNED?

---

- Research ethics and integrity training for researchers is vital also for advancement of open science.
- Multidisciplinarity fosters innovation: get physicists reading philosophy, engineering scientists working with sociologists.
- Data citation practices and data publication workflows should be implemented to ensure that openness pays off.
- Lack of gender balance indicative of women researcher's lower tolerance of risk? Targeted efforts for encouraging women to share.
- Pioneers should be recognised and rewarded with funding and positions!
- Openness as an empowering experience should be promoted.

# FIND OUT MORE

---

Longer version of this presentation: [bit.ly/fearofscooping](http://bit.ly/fearofscooping)

Paper will be submitted for the Data Science Journal SciDataCon special collection and will be available as a pre-print on [zenodo.org](http://zenodo.org) as of 1 October 2016.

Follow or contact me:

@heidiklaine / [www.thehonestbrokerblog.org](http://www.thehonestbrokerblog.org) / heidi.laine@helsinki.fi

# Giving Researchers Credit for their Data (or 'Data2Paper')

Fiona Murphy, Neil Jefferies, Anusha Ranganathan,  
Thomas Ingraham, Hollydawn Murray

JISC


DataOne Webinar, Denver, 13 September 2016

DataONE




# The Business Case


---

 **Jon Tenn@nt**  
@Prot hedgehog Follow


@dupuisj @Science\_Open Hey, probably can't make it tonight. Fighting journal submission systems. Have a great one though :)

 **Andre Nantel**  
@nantel Follow


Trying to submit a systems biology paper with lots of modelling data but the journal submission system will only accept Excel or PDF files.

 **Pat Schloss**  
@pschloss Follow


data should not be made available upon acceptance to a journal. they should be made available upon submission.

 **Susan Johnston**  
@SuseJohnston Follow


Submission to journal took several hours with limitations on size of supplementary data. Re-uploading to @biorxivpreprint took about 5 mins.

 **Anne Carpenter**  
@AnneCarpenter1 Follow


Our supplemental data just crashed the journal's submission servers. Whoopsies. #BigData

 **Nick Loman**  
@pathogenomick Follow


Ahh, come here journal submission site -- let me feel your warm embrace. What a lovely long form!

 **Benjamin Saunders**  
@benSaunders Follow

It's not the most frustrating part of science, but navigating the online submission system of a journal has to break the top 10 for me

 **Katherine James**  
@KJames\_IntBio Follow

Well, this journal submission form is utterly delightful, I'm not losing the will to live at all. Honest... \*curls up under desk and weeps\*

 **Matthew Cobb**  
@matthewcobb Follow

@alisonatkin Peer-reviewed. Journal submission systems arrgggh

JISC



# The App

---


- Researcher logs in to app via institutional repository
- Automated data paper compilation within the research space (*no journal submission systems*)
- Article packages submitted to publisher system via SWORD
- If published, ORCID updates the record

# Institutional Repository View

data2paper.org

My Drive - Google Drive Meeting Plan... Time Zones Dr Fiona Mu...ook Web App timeanddate.com (12) Webmail :: Inbox Webmail :: W...e to Webmail MyMaths - Br...live - Home Show My Ho... Homework.

ORA Dataset: "Contextual and provenance metadata in the Oxford Uni..." Data papers How to Take a Screenshot on a Mac | Digital Trends



53 views

16 downloads

Reference: **Tanya Gray Jones, Lucie Burgess, Neil Jefferies et al., Contextual and provenance metadata in the Oxford University Research Archive. Communications in Computer and Information Science. 274-285.**

Citable link to this page:  
<http://ora.ox.ac.uk/objects/uuid:e6083f46-cfbc-4c99-a929-b7183e980cc2>

**Title: Contextual and provenance metadata in the Oxford University Research Archive**

**Abstract:** Context and provenance are essential for understanding the meaning and significance of an artefact. In this paper we describe how scholarly outputs deposited in a long-term data repository, the Oxford University Research Archive (ORA), are described with contextual information and provenance. In addition, the digital objects in ORA that act as proxies to the scholarly outputs are also described with contextual information and provenance. The ORA data model is presented, together with a description of the relationships in context.

**About The Authors**

- Tanya Gray Jones**  
institution University of Oxford  
faculty The Bodleian Libraries [More by this author](#)
- Lucie Burgess**  
institution University of Oxford  
faculty The Bodleian Libraries [More by this author](#)
- Neil Jefferies**  
institution University of Oxford  
faculty The Bodleian Libraries [More by this author](#)
- Anusha Ranganathan**  
faculty The Bodleian Libraries [More by this author](#)
- Sally Rumsey**  
institution University of Oxford  
faculty The Bodleian Libraries [More by this author](#)

**Bibliographic Details**

Publisher: University of Oxford  
Date published: 2015  
Copyright Date: 2015  
DOI: 10.5072/bodleian:nZaoXzEm5

[Create data paper](#)

[Visit Data2Paper](#)

**Identifiers**

Urn: [uuid:e6083f46-cfbc-4c99-a929-b7183e980cc2](http://nbn-resolving.org/urn:uuid:e6083f46-cfbc-4c99-a929-b7183e980cc2)  
DOI: [10.5072/bodleian:nZaoXzEm5](https://doi.org/10.5072/bodleian:nZaoXzEm5)

Contextual and Provenance Metadata.txt - TXT - 408.2 kB

CSV Bibliographic Data

<http://ora.ox.ac.uk/objects/ora:11978>

# Data2Paper

The screenshot shows the Data2Paper web application interface. The browser address bar displays "data2paper.org". The page features a dark blue header with the Data2Paper logo and a user profile for "Anusha Ranganathan". Below the header is a navigation menu with "Home", "About", "Help", and "Contact" links, along with a search bar and a "My Data papers" dropdown. The main content area is divided into two tabs: "My data papers" (active) and "All journals". The "My data papers" tab shows a list of papers with columns for "Title", "Journal", "Status", "Date uploaded", and "Action". The list includes papers such as "Ephemera and the British Empire - colour illustrations" and "The Elton Archive", both with a "Submitted" status. A sidebar on the left offers options to "Upload File", "Create data paper", and "Create journal account", along with a "Filter your data papers" section with dropdown menus for "Status", "Author", "Keyword", "Subject", and "Repository".

data2paper.org

My Drive - Google Drive Meeting Plan... Time Zones Dr Fiona Mu...ook Web App timeanddate.com (12) Webmail :: Inbox Webmail :: W...e to Webmail MyMaths - Br...live - Home Show My Ho... Homework.

Taking a Screenshot on a Mac and Pasting It - PC Unleashed ORA Dataset: "Contextual and provenance metadata in the Oxford U... Data papers

Data2Paper

1 Anusha Ranganathan

Home About Help Contact

Enter search terms  My Data papers

Filter your data papers

Status > Author > Keyword > Subject > Repository >

My data papers All journals

Sort By: relevance Show 10 per page

Title	Journal	Status	Date uploaded	Action
Ephemera and the British Empire - colour illustrations >		Submitted		Select an action
The Elton Archive >		Submitted		Select an action
Sermons of John Donne Collation Results >				Select an action
Contextual and provenance metadata in the Oxford University Research Archive >				Select an action
Contextual and provenance metadata in the Oxford University Research Archive >				Select an action

# Next Steps

---

- Feedback, UI/UX work
- Researchers – domain and social factors
- Outreach and trials
- Sustainable long-term strategy
- All suggestions welcome!

# Thank you!

---

[fionalm27@gmail.com](mailto:fionalm27@gmail.com)  
@DrFionaLM

Jisc: Research Data Spring Grant

JISC

*DataONE Webinar Series*

***[www.dataone.org/webinars](http://www.dataone.org/webinars)***

*Upcoming Webinar Event*

***[www.dataone.org/upcoming-webinar](http://www.dataone.org/upcoming-webinar)***

*Previous Webinar Events (Recording and Discussion)*

***[www.dataone.org/previous-webinars](http://www.dataone.org/previous-webinars)***



***#DWS2016***

***@DataONEorg***