

DataONE Webinar Series

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

Jens Klump

Heidi Laine

Fiona Murphy





*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> CC0

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 outlines how you will lay out a plan for your data management needs.

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> CC0

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 format that is media and location neutral. Publish information so that others can discover it easily.

Concern	Solution
Concern	Solution

DataONE

Lesson 1: Data Management

View all Education Modules at <https://www.dataone.org/education-modules> CC0

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

Year	Information (Petabytes)	Available Storage (Petabytes)
2005	~100	~100
2006	~150	~150
2007	~200	~200
2008	~250	~250
2009	~300	~300
2010	~350	~350

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[Plan] --> Collect[Collect]
    Collect --> Analyze[Analyze]
    Analyze --> Integrate[Integrate]
    Integrate --> Discover[Discover]
    Discover --> Preserve[Preserve]
    Preserve --> Describe[Describe]
    Describe --> Assure[Assure]
    Assure --> Local[Local contact information]
    Local --> DataManagement[Data Management]
    DataManagement --> DataSharing[Data Sharing]
    DataSharing --> DataReuse[Data Reuse]
    DataReuse --> DataManagement
  
```

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

DataONE

Tutorials on Data Management

Lesson 10: Analysis and Workflows

Analysis and Workflows

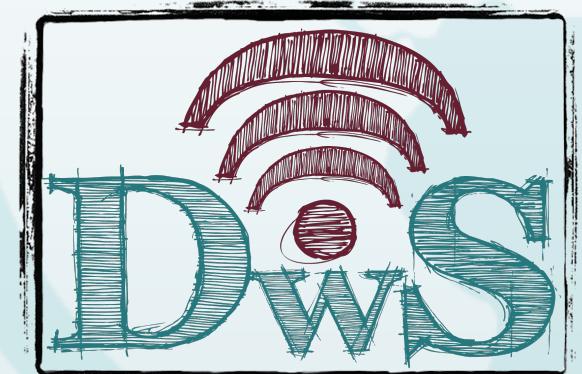
CC-BY-SA 4.0 International

DataONE

Documents 0
Infographics 0
Videos 0



#DWS2016
@DataONEorg
www.DataONE.org



DataONE Webinar Series

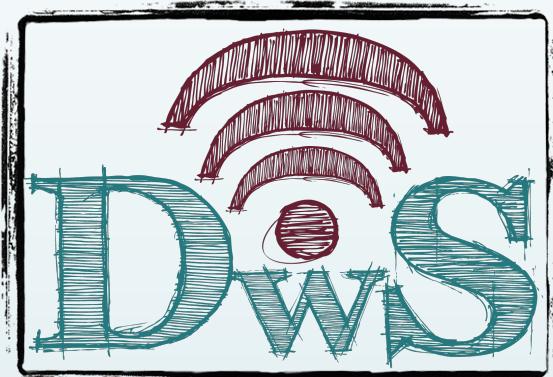
[**www.dataone.org/webinars**](http://www.dataone.org/webinars)

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

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

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?



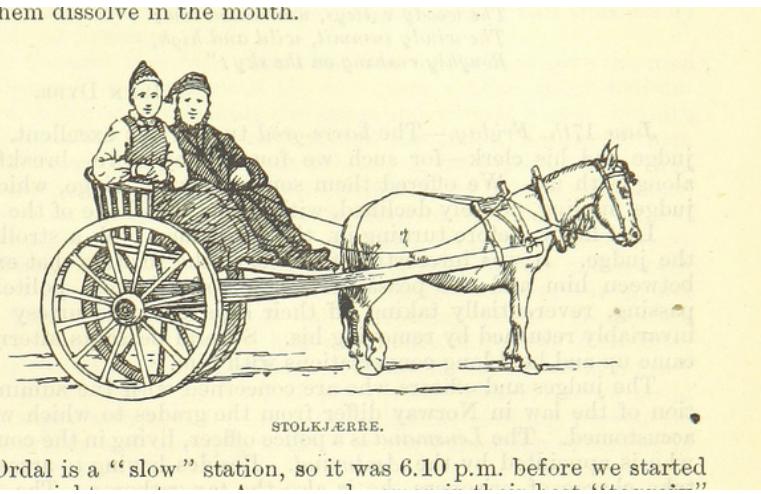
South Atlantic (Namibia)



Lake Baikal (Russia)

Are we getting the incentives right?

Image: British Library (public domain)



Ordal is a “slow” station, so it was 6.10 p.m. before we started

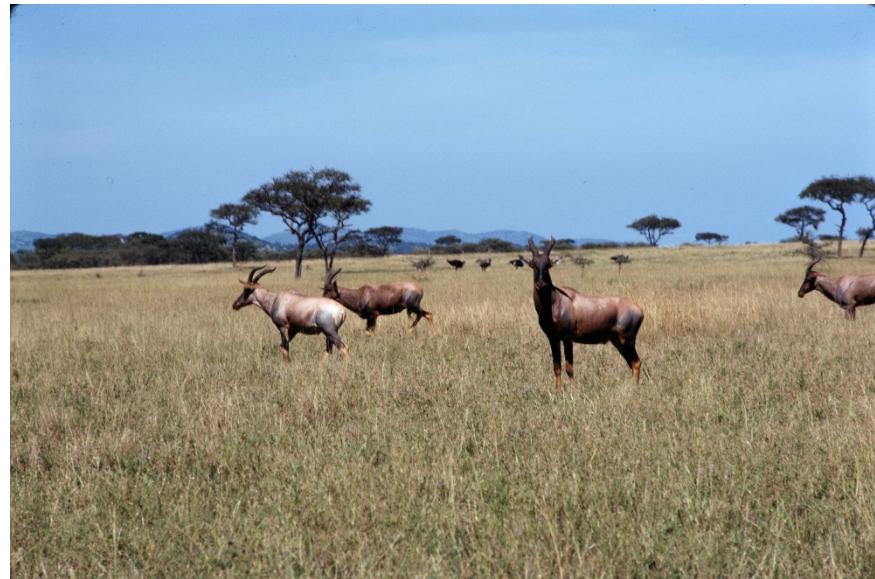


Image: ETH Zürich Library (public domain)

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.

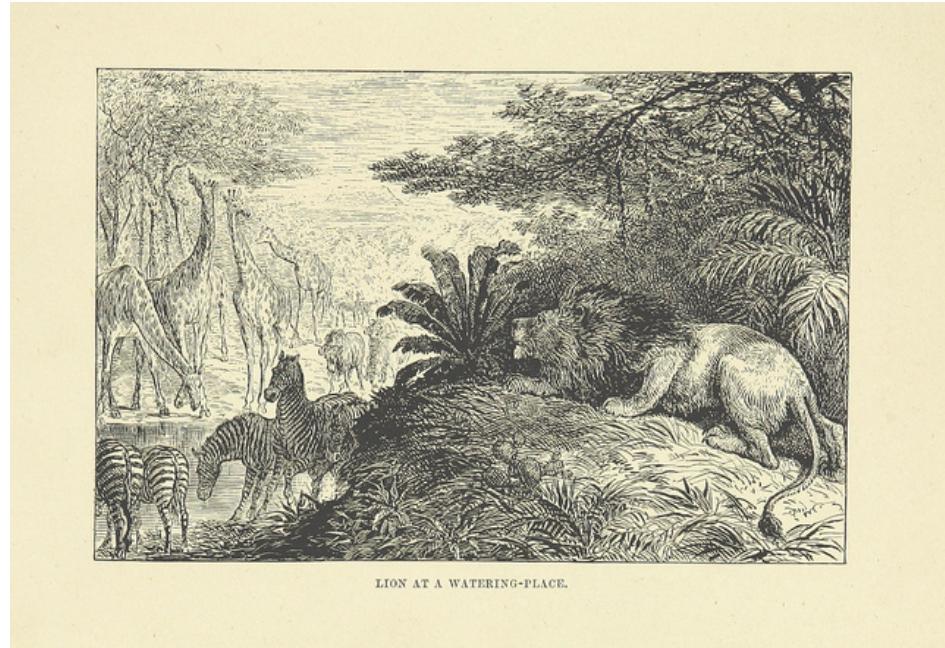
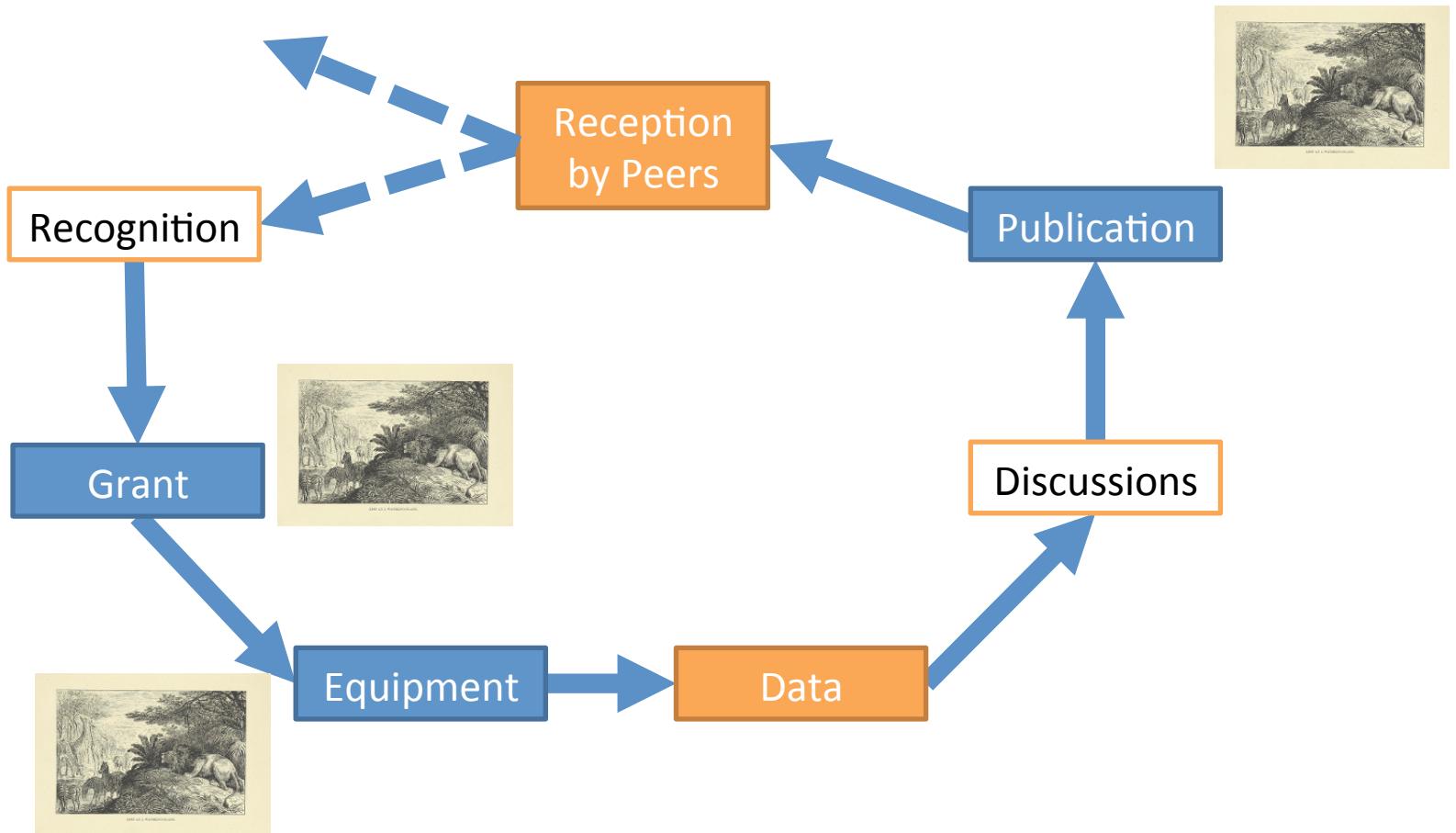


Image: British Library (public domain)

Reputation Economy



Mineral Resources

Jens Klump

Science Leader Earth Science Informatics

t +61 8 6436 8828

e jens.klump@csiro.au

w www.csiro.au/en/Research/MRF

AFRAID OF SCOOPING?

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

HEIDI LAINE / heidi.laine@helsinki.fi /
@heidiklaine



This work is licensed under a

[Creative Commons Attribution 4.0 International
License.](#)



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?

RADICAL OPENNESS

ORS / SOMUS	NMRLP
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

Paper will be submitted for the Data Science Journal SciDataCon special collection and will be available as a pre-print on zenodo.org as of 1 October 2016.

Follow or contact me:

@heidiklaine / 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



DataOne Webinar, Denver, 13 September 2016



The Business Case

 Jon Tenn@nt
@Protolothedgehog

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

[Follow](#)

 Andre Nantel
@nanted

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

[Follow](#)

 Pat Schloss
@pat.schloss

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

[Follow](#)

 Susan Johnston
@SusanJohnston

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

[Follow](#)

 Anne Carpenter
@AnneCarpenter

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

[Follow](#)

 Benjamin Saunders
@BenSaunders

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

[Follow](#)

 Katherine James
@kJames_imbio

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*

[Follow](#)

 Matthew Cobb
@matthewcobb

@alisonatkin Peer-reviewed. Journal submission systems arrgggh

[Follow](#)

 Nick Loman
@pathogenomenick

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

[Follow](#)

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 +

 OPEN ACCESS ITEM

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

Identifiers

Urn: [uuid:e6083f46-cfbc-4c99-a929-b7183e980cc2](http://ora.ox.ac.uk/objects/uuid:e6083f46-cfbc-4c99-a929-b7183e980cc2)
DOI: [10.5072/bodleian:nZaoXzEm5](http://doi.org/10.5072/bodleian:nZaoXzEm5)

Create data paper

 Visit Data2Paper

Contextual and Provenance Metadata.txt - TXT - 408.2 kB

CSV Bibliographic Data
<http://ora.ox.ac.uk/objects/ora:11878>

Data2Paper

The screenshot shows the Data2Paper website running in a web browser. The header features a logo with a book and a graduation cap, followed by the text "Data2Paper". Below the header is a navigation bar with links for Home, About, Help, Contact, and a search bar with placeholder text "Enter search terms". A user profile is shown on the right. The main content area displays a table of data papers, each with a thumbnail, title, journal, status (all shown as "Submitted"), date uploaded, and an "Action" dropdown menu. On the left, there's a sidebar titled "Filter your data papers" with dropdown menus for Status, Author, Keyword, Subject, and Repository.

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
@DrFionaLM

Jisc: Research Data Spring Grant

JISC

DataONE Webinar Series

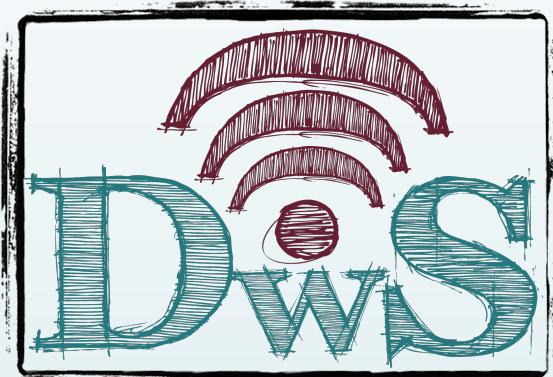
[**www.dataone.org/webinars**](http://www.dataone.org/webinars)

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