Data publication models: benefits, risks and peer review

Sarah Callaghan

#preparde

sarah.callaghan@stfc.ac.uk @sorcha_ni

Journals – a 17th century technology

- The first scientific journal, Journal des sçavans (later renamed Journal des savants), was first published on Monday, 5 January 1665.
- It also carried a proportion of material that would not now be considered scientific, such as obituaries of famous men, church history, and legal reports.
- It still exists, but is more of a literary journal

- The first edition of the Philosophical Transactions of the Royal Society of London was on 6 March 1665. That still exists, and continues to publish scientific information to this day.

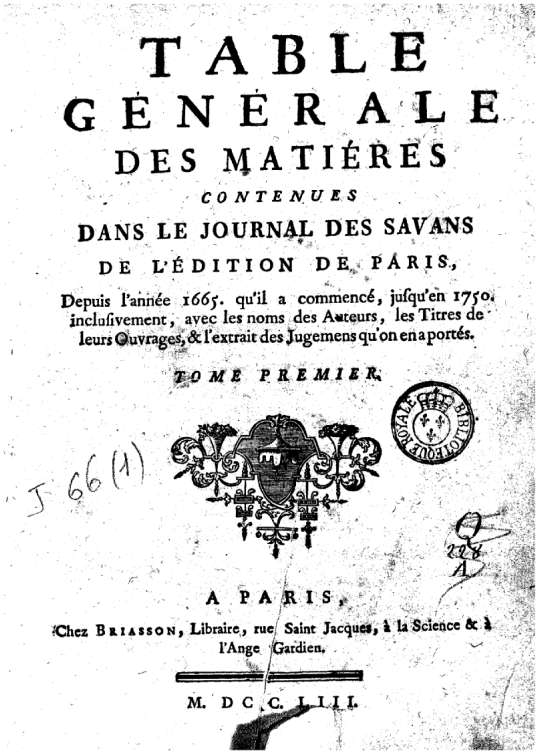
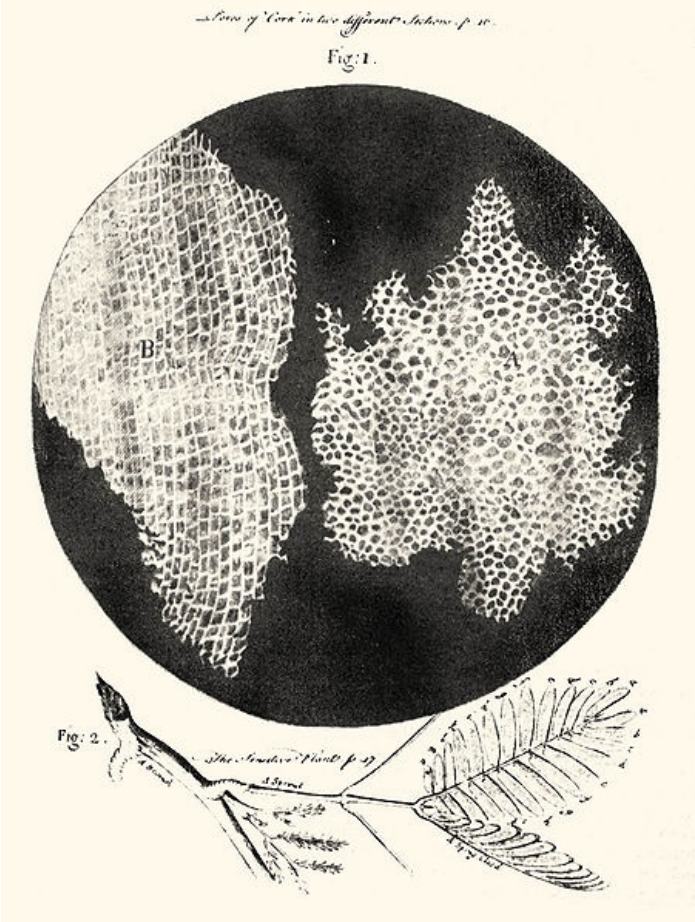


TABLE
GÉNÉRALE
DES MATIÉRES
CONTENUES
DANS LE JOURNAL DES SAVANS
DE L'ÉDITION DE PARIS,
Depuis l'année 1665. qu'il a commencé, jusqu'en 1750. inclusivement, avec les noms des Auteurs, les Titres de leurs Ouvrages, & l'extrait des Jugemens qu'on en a portés.
TOME PREMIER.
A PARIS,
Chez BRIASSON, Libraire, rue Saint Jacques, à la Science & à l'Ange Gardien.
M. DCC. LII.

Source gallica.bnf.fr / Bibliothèque nationale de France

Journals have always published data



Suber cells and mimosa leaves. Robert Hooke, Micrographia, 1665

[*Observations of Stars in the Spiral Nebula.* H. 1622.

The spiral form of this nebula is very distinctly seen in the Pulkova refractor. Unfortunately in the month of March, the best season for the observation of this object, the sky was constantly cloudy; so that I could only get three nights' observations in the months of April and May, when the twilight did not cease for the whole night. It must be attributed to this unfavourable circumstance that the following list of determinations is not so complete as it probably would have been without the twilight. The observations have been made alternately with powers of 138 and 207.

Observations.

Date.	Object.	Magnitude.	Ang. Pos.	No. of measures.	Distance.	No. of measures.
1851, April 7.	N *n*	.....	14° 55′	5	267″·1	4
	N *a*	*a* =(11)	229 24	3	88·0	3
	N *b*	*b* =(11.12)	169 12	3	242·6	3
	a b	.....	93 42	3	298·6	3
April 28.	*a b*	.....	94 23	3	300·8	4
	N *a*	.....	228 36	4		
	N *b*	.....	168 54	4		
	n a	.....	283 42	3		
	n b	.....	153 30	3		
	a d	*d* =(12.13)	323 51	3		
	N *d*	.....	277 27	3		
	a e	*e* =(13)	112 13	3		
	N *e*	.....	161 56	3		
	N *f*	*f* =(12.13)	309 18	3		
	n f	.....	237 31	3		
	a f	.....	335 23	3		
	a g	*g* =(12.13)	215 17	3	115·5	4
	a h	*h* =(12.13)	193 29	3		
	g h	.....	87 5	3		
May 3.	N *h*	*h* =(13.14)	51 47	3		
	n h	.....	173 29	4		
	b h	.....	317 23	3		
	b l	*l* =(11.12)	27 20	4		
	n l	.....	83 17	4	355·2	4
	a e	.....	112 56	4		
	N *e*	.....	161 39	3		
	a m	*m* =(12.13)	172 43	5		
	N *m*	.....	190 44	4		
	b m	.....	238 50	4		
	N *a*	.....	229 12	4	87·0	3
	N *n*	.....	14 47	4	264·2	3

The Scientific Papers of William Parsons, Third Earl of Rosse 1800-1867

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Benefits of data publication

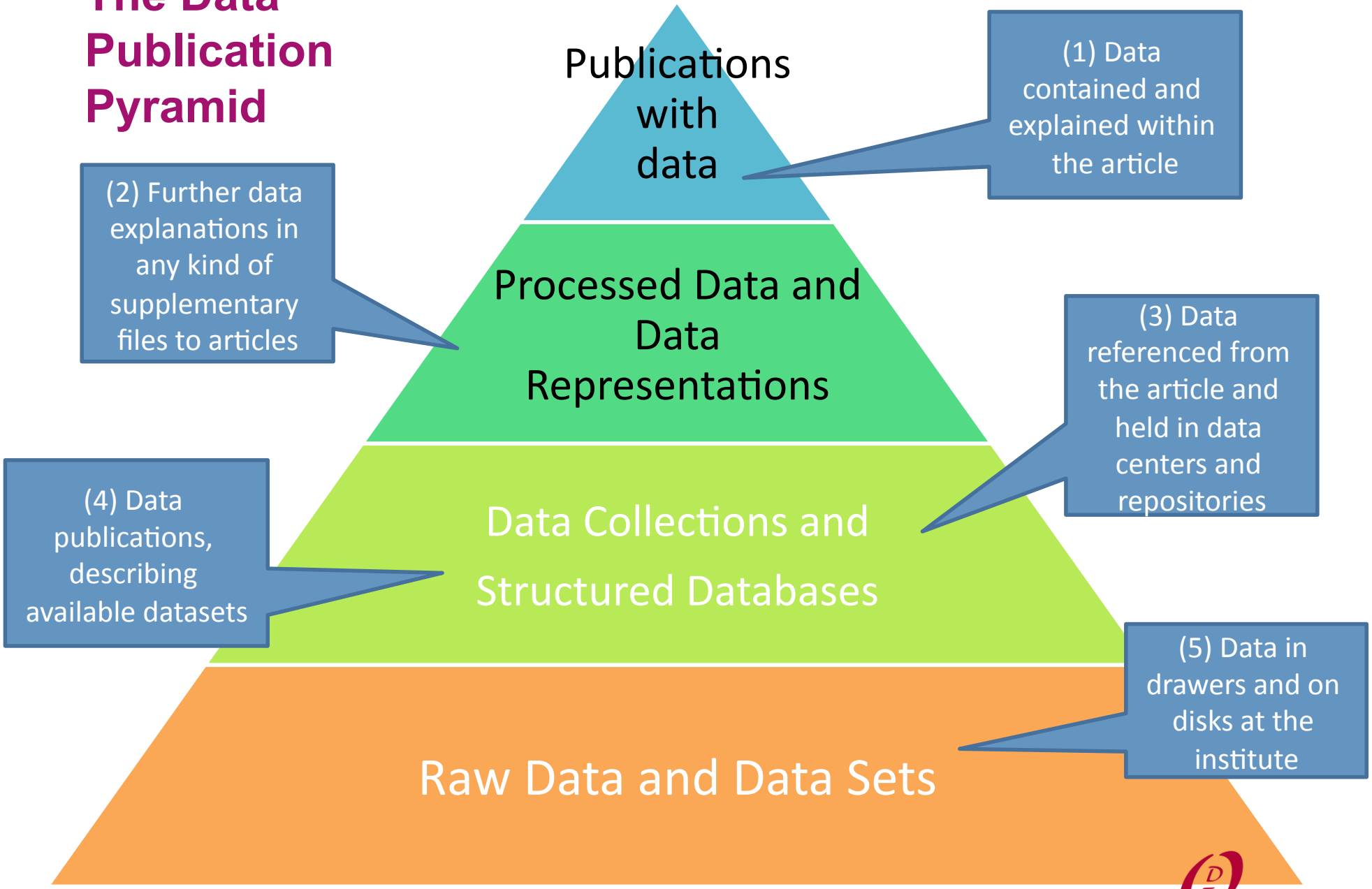
- Gives academic credit to data producers and curators for their efforts in creating, documenting and managing datasets
- Encourages reuse of datasets and discourages duplication of effort in re-creating already existing datasets
- Encourages proper curation and management of data
- Gives indication of data quality and impact
- Helps ensure the completeness of the scientific record, and the reproducibility of research.
- Shows transparency of the research process
- Improves discoverability

What it all comes down to:

Composite image from Flickr user bnilsen and Matt Stempeck (NOI), shared under Creative Commons license

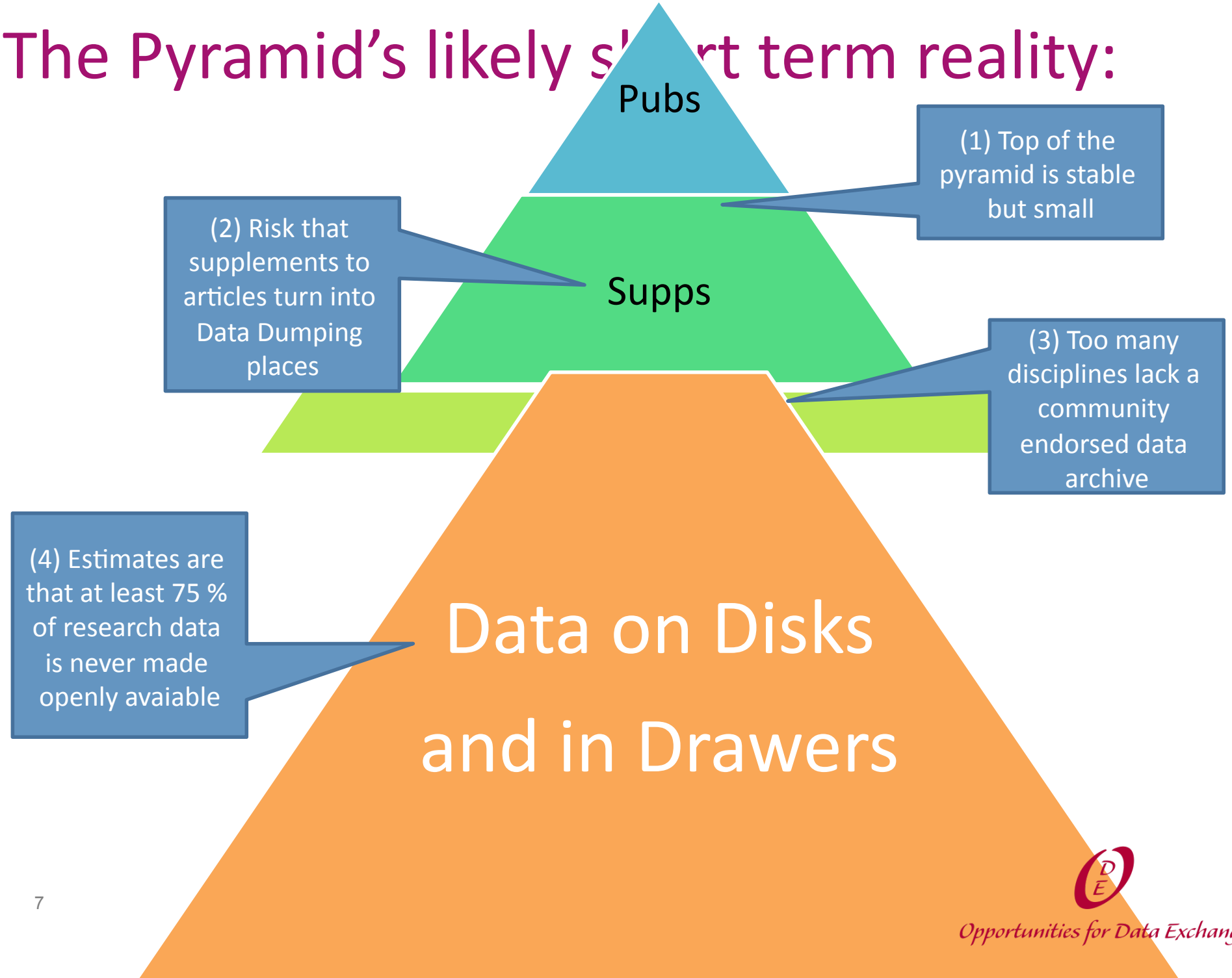
- Encourage and provide credit to researchers and institutions for managing and disseminating their data properly.

The Data Publication Pyramid



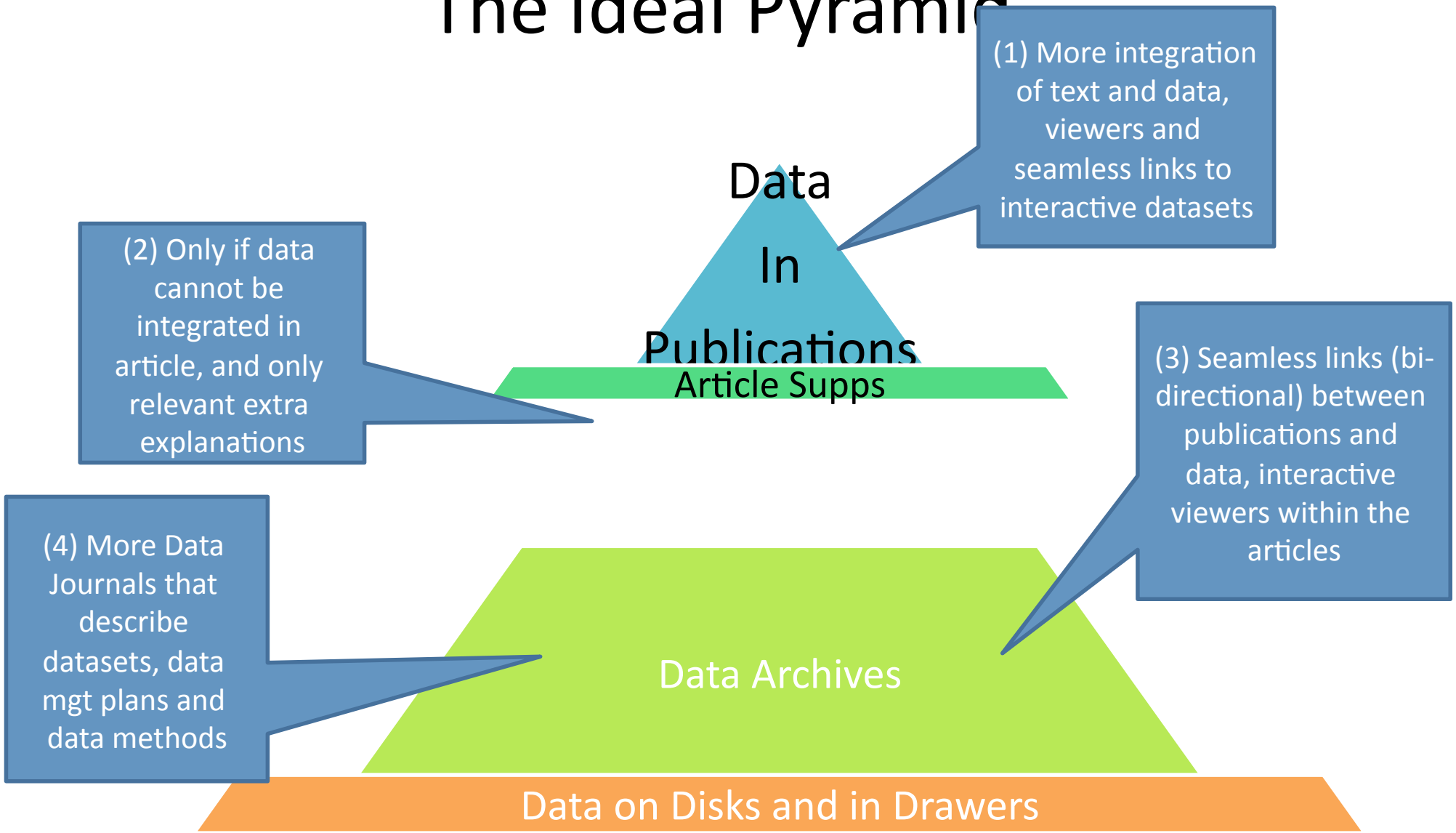
Opportunities for Data Exchange

The Pyramid's likely short term reality:



Opportunities for Data Exchange

The Ideal Pyramid



Opportunities for Data Exchange

How to publish data

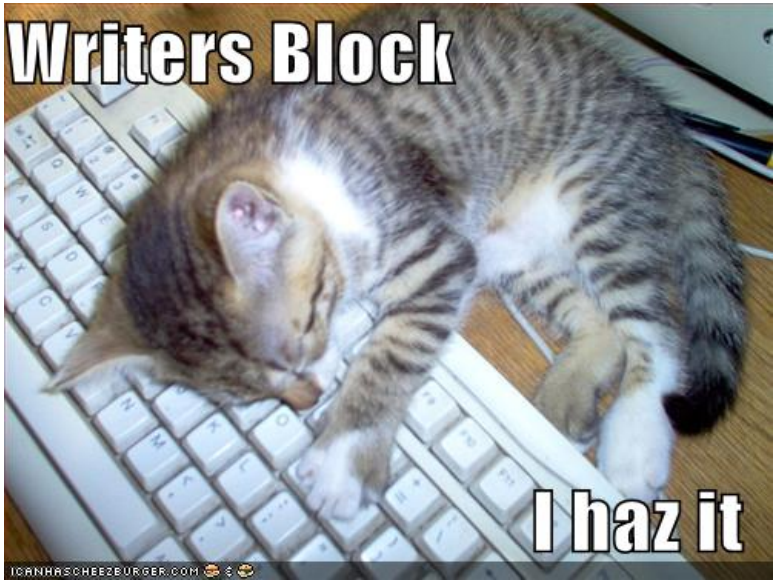
- Stick it up on a webpage somewhere
  - Issues with stability, persistence, discoverability…
  - Maintenance of the website
- Put it in the cloud
  - Issues with stability, persistence, discoverability…
- Attach it to a journal paper and store it as supplementary materials
  - Journals not too keen on archiving lots of supplementary data, especially if it's large volume.
- Put it in a disciplinary/institutional repository
- Write a data article about it and publish it in a data journal

By David Fletcher
http://www.cloudtweaks.com/2011/05/the-lighter-side-of-the-cloud-data-transfer/

What is a data article?

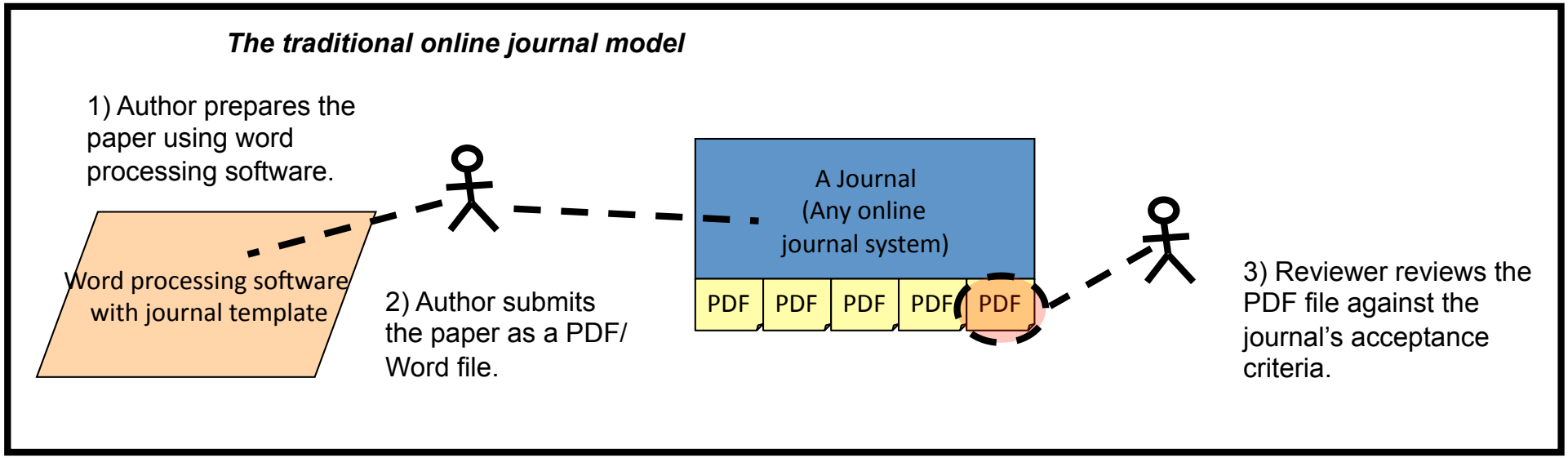
A data article describes a dataset, giving details of its collection, processing, software, file formats, etc., without the requirement of novel analyses or ground breaking conclusions.

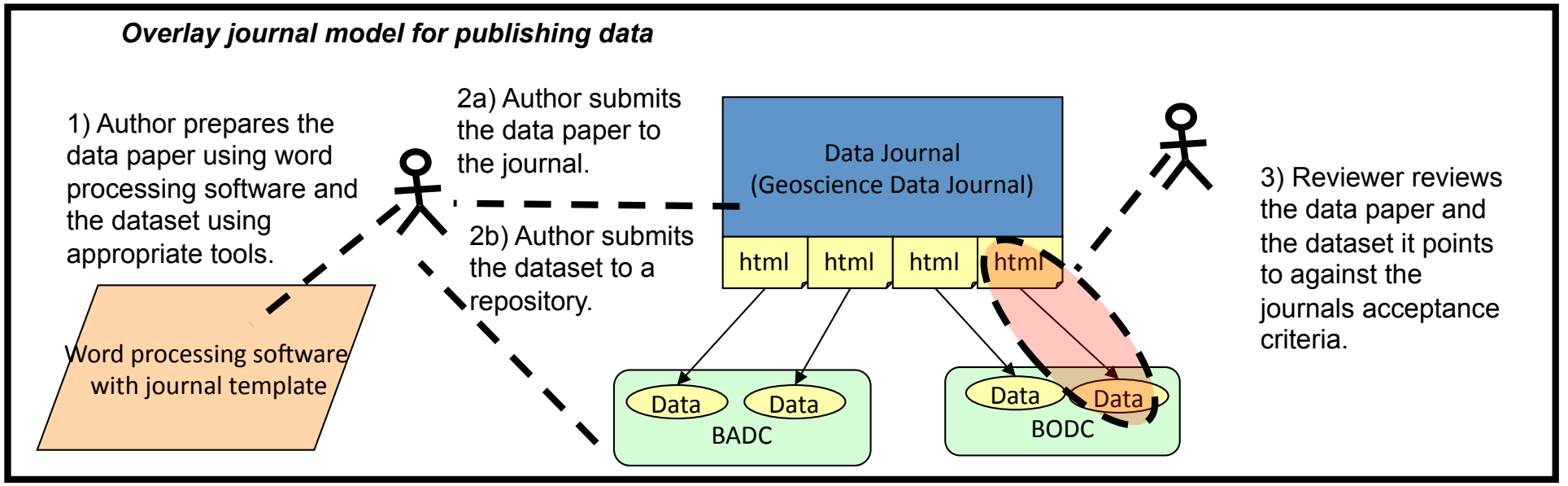
- the when, how and why data was collected and what the data-product is.



JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH





JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

A list of data journals (in no particular order)

- On PREPARDE blog at: http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList
- Headings:
  - Name of Data Journal (and URL to main journal page)
  - Aims and Scope
  - Repository Criteria
  - Other notes

The list

- Geoscience Data Journal http://www.geosciencedata.com
- Earth System Science Data http://earth-system-science-data.net/
- Ecological Archives - Data Papers http://esapubs.org/archive/archive_D.htm
- Hindawi publishing: http://www.datasets.com/
  - Dataset Papers in Agriculture, Dataset Papers in Biology, Dataset Papers in Chemistry, Dataset Papers in Ecology, Dataset Papers in Geosciences, Dataset Papers in Materials Science, Dataset Papers in Medicine, Dataset Papers in Nanotechnology, Dataset Papers in Neuroscience, Dataset Papers in Pharmacology, Dataset Papers in Physics
- Journal of Chemical and Engineering Data http://pubs.acs.org/journal/jceaax
- GigaScience http://www.gigasciencejournal.com/
- Journal of Physical and Chemical Research Data http://jpcrd.aip.org/resource/1/jpcrbu
- Biodiversity Data Journal http://www.pensoft.net/journals/bdj/
- F1000 Research http://f1000research.com
- International Journal of Robotics Research http://ijr.sagepub.com/
- CODATA's Data Science Journal http://www.codata.org/dsj/index.html
- Ubiquity Press: http://www.ubiquitypress.com/
  - Journal of Open Archaeology Data: http://openarchaeologydata.metajnl.com/, Journal of Open Public Health Data: http://openpublichealthdata.metajnl.com/, Journal of Open Psychology Data: http://openpsychologydata.metajnl.com/, Journal of Open Research Software: http://openresearchsoftware.metajnl.com/
- BMC Research Notes http://www.biomedcentral.com/bmcresnotes/

- Geoscientific Model Development (GMD) http://www.geoscientific-model-development.net/
  - Not a Data Journal as such, but serves a similar function for the modelling community.

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Very varied!

- Cover a wide range of subject areas – Earth Sciences well covered, but plenty of missing subject areas
- Repository criteria vary from not specified to very specific requirements (e.g. what data repository data must be submitted to)
- Hindawi data journals hold the dataset as well as the data paper.
- Majority are Open Access for the data article. ESSD mandate that the published dataset needs to be OA too.

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

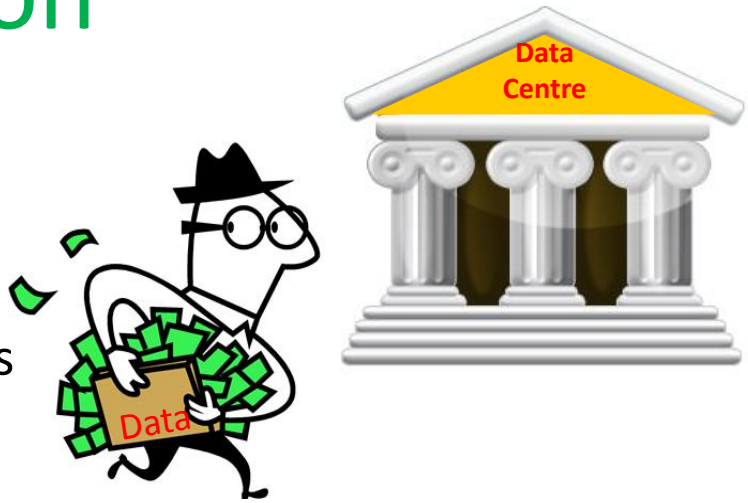
Risks – what happens if the links break?

- Data articles with broken links to the datasets they describe lose all credibility (bad for the journal)
- Datasets where the main documentation is the data article lose crucial metadata (bad for the repository).
- Everyone loses trust and reputation!

House of Cards by peterjroberts

Repository accreditation for data publication

Different people have different definitions of "trustworthy"



- Dark archives can be trusted by researchers to store the data
  - Not useful for publication
- Journal editors need a quick and easy way to determine if a repository hosting a published dataset will meet the publication requirements:
  - Landing pages, permanent ids, access control, persistence guarantees, …

Who gets to decide if a repository is trustworthy?

Trust is reciprocal!

JISC · University of Leicester · British Atmospheric Data Centre (National Centre for Atmospheric Science, Natural Environment Research Council) · Wiley-Blackwell · University of Reading · DCC · University of California CDL California Digital Library · F1000 Faculty of 1000 Post-Publication Peer Review · NCAR National Center for Atmospheric Research

What are the right requirements?

Keep the focus on data publication

- Existing repository accreditation schemes (TRAC, Data Seal of Approval, WDS,…)?
- Low level checklist of requirements?
- Community acceptance?
- Institutional reputation?
- Social networking approval (TripAdvisor)?
- …?

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Are these the right requirements?

Repositories should:

- Have long term data preservation plans in place for their archive.
- Actively manage and curate the data in their archive.
- Provide landing pages giving extra information about the dataset (metadata) and information on how to access the data.
- Use persistent, actionable links (e.g., DOIs, ARKs) to cite data held in their archive
- Resolve cited dataset links to landing pages

Over to you!

Image Credit: http://bit.ly/9H4qBX

The project is led by the University of Leicester and the support of JISC and NERC in funding the PREPARDE project is gratefully acknowledged.

JISC · University of Leicester · British Atmospheric Data Centre (National Centre for Atmospheric Science, Natural Environment Research Council) · Wiley-Blackwell · University of Reading · DCC · University of California CDL California Digital Library · F1000 Faculty of 1000 Post-Publication Peer Review · NCAR National Center for Atmospheric Research


JISC

University of
Leicester

British Atmospheric
Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

WILEY-
BLACKWELL

University of
Reading

DCC


University of California
CDL
California Digital Library

F1000
FACULTY of 1000
POST-PUBLICATION PEER REVIEW

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

JISC · University of Leicester · British Atmospheric Data Centre (National Centre for Atmospheric Science, Natural Environment Research Council) · Wiley-Blackwell · University of Reading · DCC · University of California CDL California Digital Library · F1000 Faculty of 1000 Post-Publication Peer Review · NCAR National Center for Atmospheric Research