



RESEARCH DATA MANAGEMENT AND LIBRARIES

yatrik@inflibnet.ac.in

Indian Librarian Appreciation Day
14th August 2015 ,Pune

TYPES OF RESEARCH DATA

Observational:

- data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neurological images.

Experimental:

- data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.

Simulation:

- data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.

Derived or compiled:

- data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.

Reference or canonical:

- a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

MOTIVATION FOR SHARING RESEARCH DATA

When data sharing is an essential part of the research process

Direct career benefits, derived from sharing through greater visibility of one's work, reciprocal data exchanges, and the reassurance of having one's data recognised as valuable by others;

The norms that researchers are exposed to within their research circle or discipline

A framework of funder and publisher expectations, policies, infrastructure and data services as external drivers

THE INCENTIVES

Direct benefits

- for the research itself (more robust)
- for the career of the researcher (recognition)
- for discipline (get wiser)
- for science (better science)

External drivers:

- policies and expectations from research funders and publishers
- Norms of the project, research group, and/or discipline

GLOBAL SCENARIO

Researcher's Data Sharing insights (2014) : Wiley

A survey conducted in March 2014, Over 2200+ Respondents worldwide .

SOCIETIES RESEARCH LIBRARIES // NOVEMBER 3RD, 2014

How and why researchers share data (and why they don't)



Liz Ferguson

Publishing Solutions Director, Wiley

15 Comments

Tags: data management, data sharing, Dryad, journals, Liz Ferguson, open access, research

<http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/>

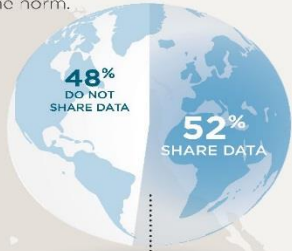
RESEARCHER DATA SHARING INSIGHTS

WILEY

- Wiley's Researcher Data Insights Survey was launched earlier this year to understand how and why researchers make their research data publicly available. The study's results, highlighted below, are intended to advance the global conversation about data sharing and help Wiley better meet the needs of our researchers, authors, and partners in the rapidly evolving landscape of scientific research and communications.
- The survey was deployed in March 2014 and received more than 2,250 responses from researchers around the world.

GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.

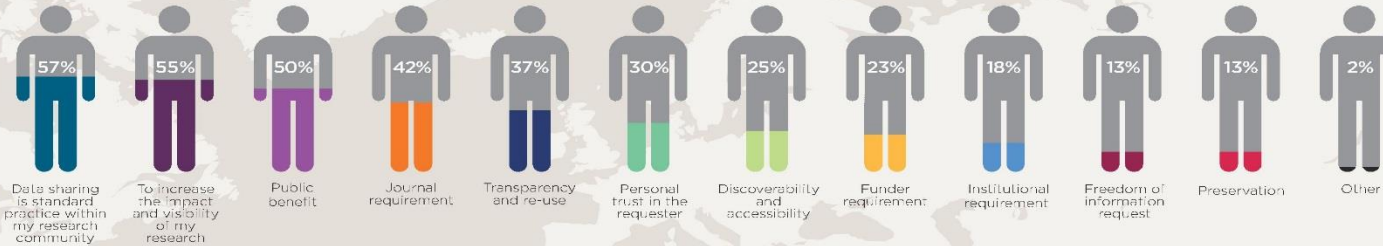


WAYS DATA IS SHARED

- 67% As supplementary material in a journal
- 37% Personal, institutional or project webpage
- 26% Institutional data repository (i.e. university or institute-sponsored)
- 19% Discipline-specific data repository
- 6% General-purpose data repository (e.g. Dryad, figshare)
- 5% Other

Globally, researchers also report sharing their data in limited and non-permanent ways: 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).

RESEARCHER MOTIVATIONS FOR SHARING DATA



DATA SHARING TRENDS BY COUNTRY



REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

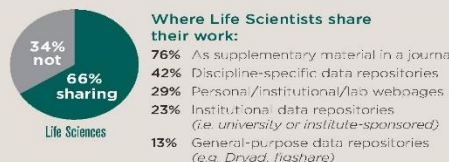
- 42% Intellectual property or confidentiality issues
- 36% My funder/institution does not require data sharing
- 26% I am concerned that my research will be scooped
- 26% I am concerned about misinterpretation or misuse
- 23% Ethical concerns
- 22% I am concerned about being given proper citation credit or attribution
- 21% I did not know where to share my data
- 20% Insufficient time and/or resources
- 16% I did not know how to share my data
- 12% I don't think it is my responsibility
- 12% I did not consider the data to be relevant
- 11% Lack of funding
- 7% Other

DATA SHARING BY DISCIPLINE

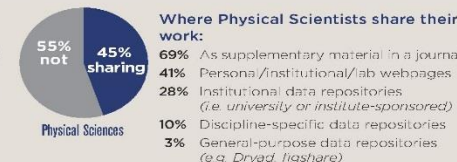
Data sharing, specifically by way of data repositories, is most prevalent amongst life scientists, particularly those in the earth and environmental and agriculture and food sciences.



A typical *Health Science researcher* says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assuaged.



A typical *Life Science researcher* says she would be motivated to share more of her data in the future if she was guaranteed proper credit.

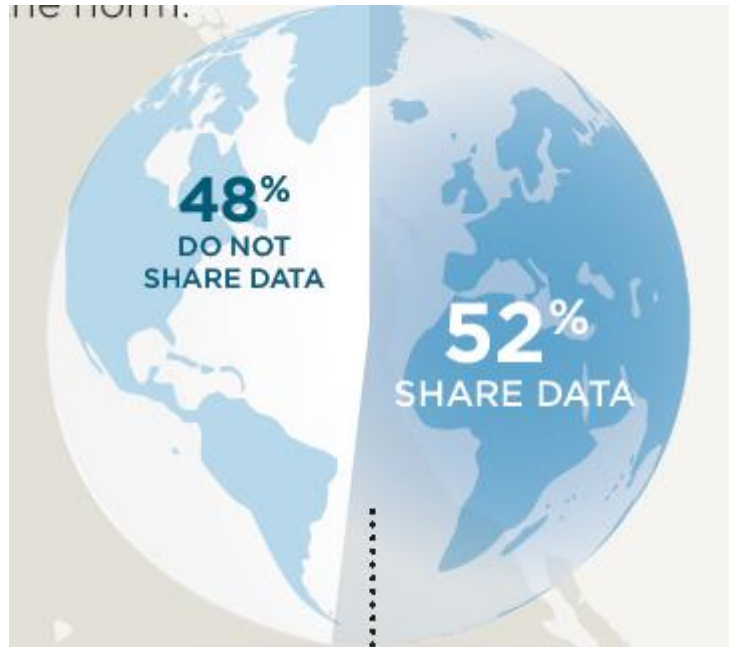








A typical *Physical Science researcher* says she would be motivated to share her data in the future because it is standard practice within her research community and because it increases the impact and visibility of her work.



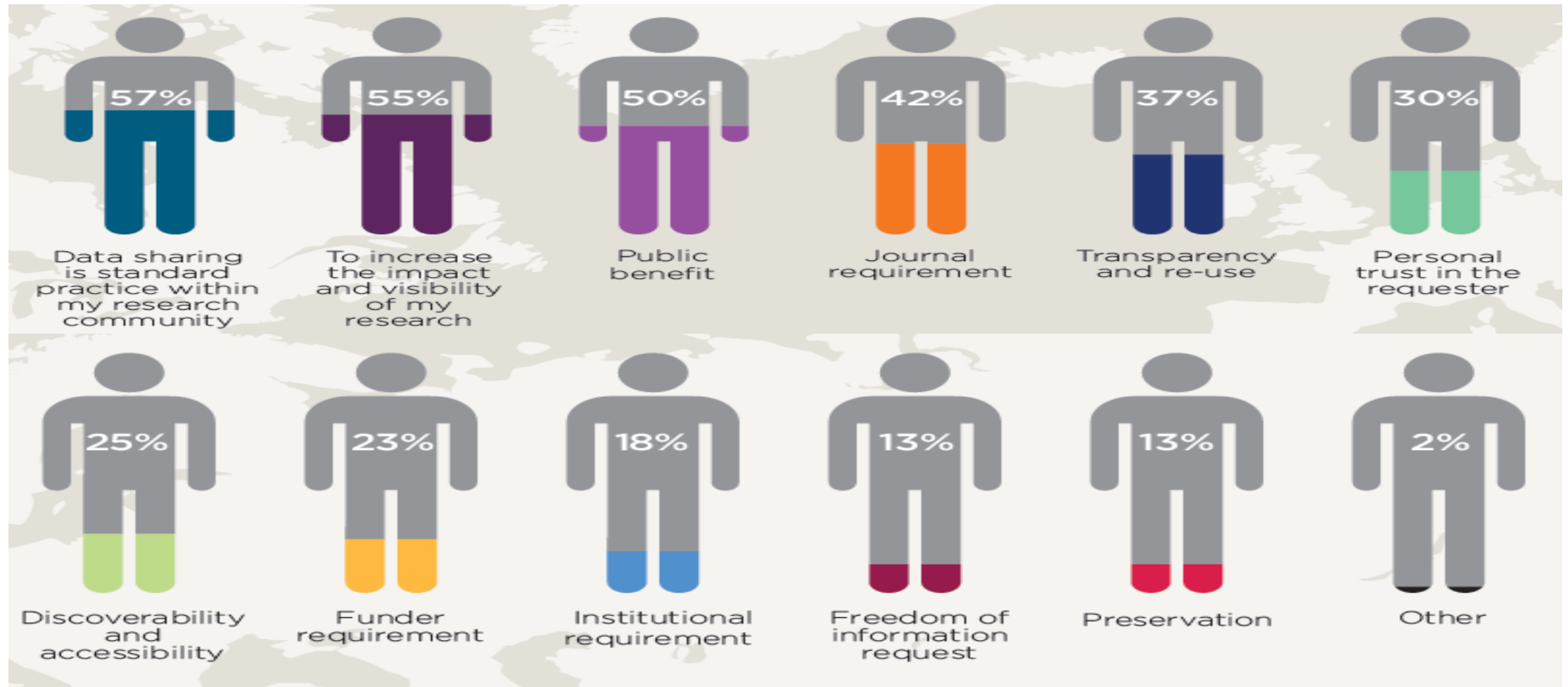
A typical *Social Science and Humanities researcher* says she would be motivated to share her data in the future if it increased the impact and visibility of her work or if she was required to by her funder.

DATA SHARING SCENARIO.



-  **67%** As supplementary material in a journal
-  **37%** Personal, institutional or project webpage
-  **26%** Institutional data repository
(i.e. university or institute-sponsored)
-  **19%** Discipline-specific data repository
-  **6%** General-purpose data repository
(e.g. Dryad, figshare)
-  **5%** Other

RESEARCHER'S MOTIVATION



COUNTRY TRENDS.



United States

- two out of three do so because it is standard practice
- they believe it benefits the public.
- share data to increase the impact or visibility of their research.



United Kingdom

- Only about 14% are using discipline-specific or other public repositories (Dryad and figshare.)
- Motivation : the prospect of gaining increased impact or visibility for their work
- Motivation : to satisfy funder requirements.



Japan

- Five out of Ten worried about being scooped as a reason for not sharing data more frequently.
- roughly double the global average.

COUNTRY TRENDS.



36%

China

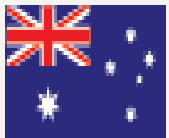
- Nearly five in ten say they are not sharing data because not required to do so by their funders or institutions.
- They do not see data sharing as a personal responsibility



52%

Brazil

- Two out of three say that a guarantee of proper credit or attribution would compel them to share more of their data publicly in the future



41%

AUSTRALIA

- would be most incentivized to make their data accessible in the future to ensure preservation as well as transparency and re-use.
- The majority also ranked funder requirements among top reasons to share in the future.



55%

Germany

- Three out of four are believes increase the visibility of their research and want to ensure public transparency and re-use.
- About 20% making use of general purpose repositories (like figshare and Dryad), more than their counterparts around the world

HESITANCE IN DATA SHARING

- 
- 42%** Intellectual property or confidentiality issues
 - 36%** My funder/institution does not require data sharing
 - 26%** I am concerned that my research will be scooped
 - 26%** I am concerned about misinterpretation or misuse
 - 23%** Ethical concerns
 - 22%** I am concerned about being given proper citation credit or attribution
 - 21%** I did not know where to share my data
 - 20%** Insufficient time and/or resources
 - 16%** I did not know how to share my data
 - 12%** I don't think it is my responsibility
 - 12%** I did not consider the data to be relevant
 - 11%** Lack of funding
 - 7%** Other

ANOTHER “IN SIGHT”



Research Policy

Volume 43, Issue 9, November 2014, Pages 1621–1633



Open access to data: An ideal professed but not practised

Patrick Andreoli-Versbach^{a, b},   , Frank Mueller-Langer^{a, c} 

^a Max Planck Institute for Innovation and Competition

^b University of Munich, Department of Economics

^c International Max Planck Research School for Competition and Innovation

Received 4 April 2013, Revised 11 April 2014, Accepted 15 April 2014, Available online 9 June 2014

 CrossMark

 Show less

doi:10.1016/j.respol.2014.04.008

[Get rights and content](#)

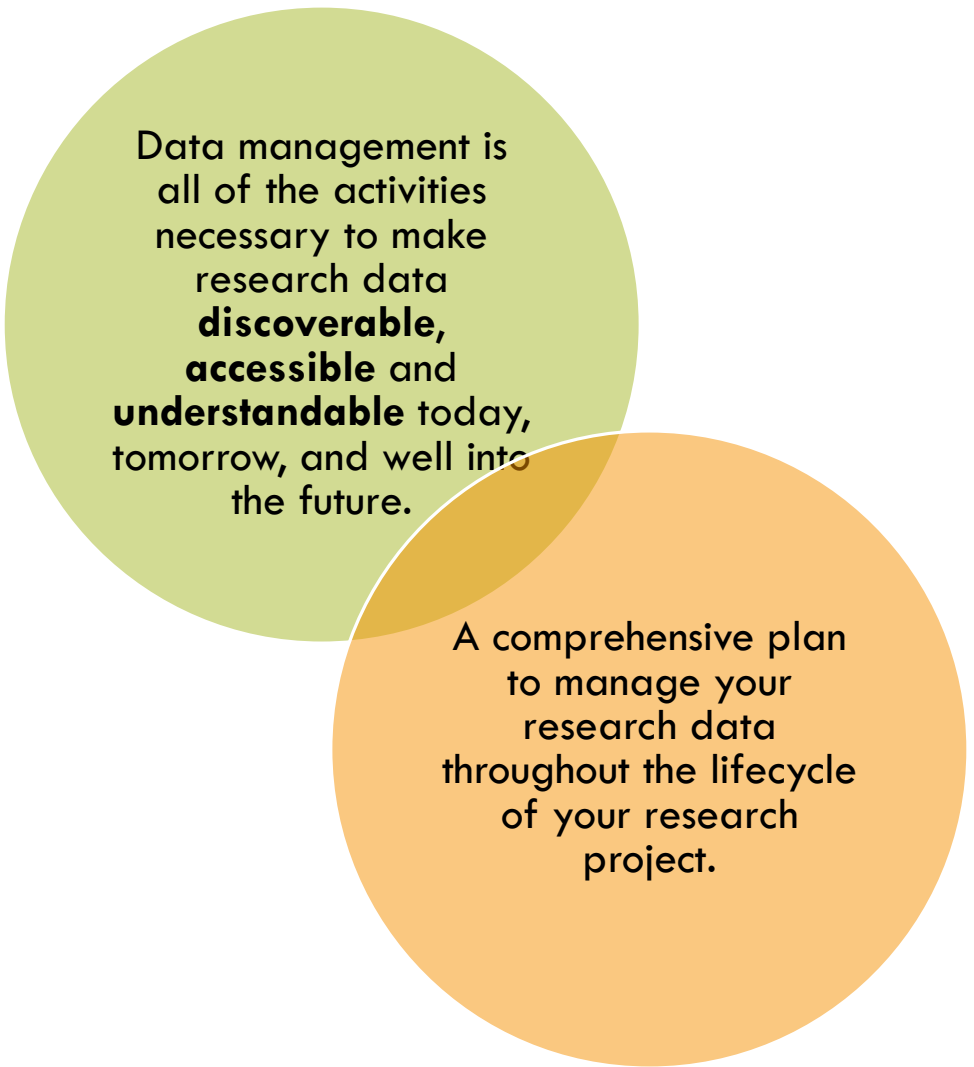
Highlights

- Data-sharing in economics is often professed but seldom practised.
- **We find that 80.74% of researchers do not voluntarily share their data.**
- We derive five testable hypotheses based on the literature on information-sharing.
- We find four significant predictors of voluntary data-sharing.
- Tenure, author quality, extent of mandatory data-disclosure and personal attitudes.

WHY ?



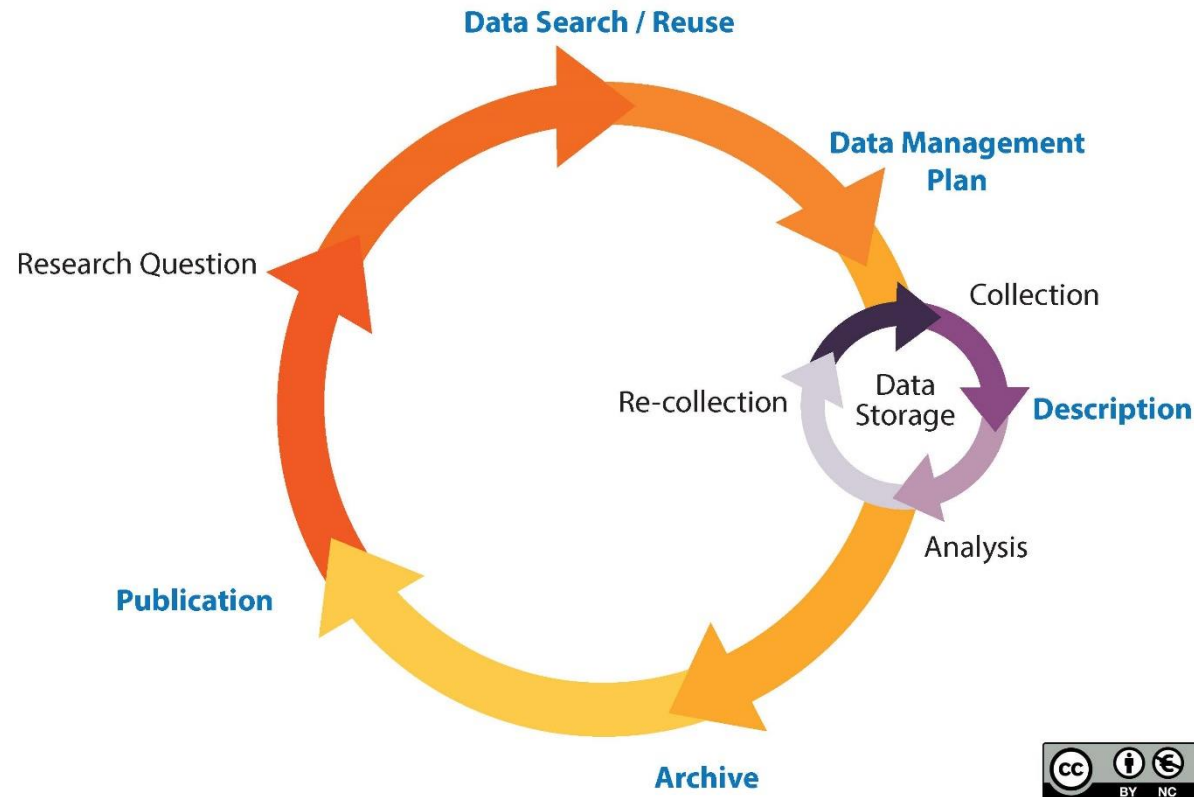
DATA MANAGEMENT.



Data management is all of the activities necessary to make research data **discoverable, accessible** and **understandable** today, tomorrow, and well into the future.

A comprehensive plan to manage your research data throughout the lifecycle of your research project.

RESEARCH DATA MANAGEMENT LIFECYCLE



Choosing file formats

File organization & naming conventions

Version control

Document all project/file details

Access control & security

Backup & storage

File format conversions

Sharing and preservation



COMPONENTS : GENERIC DATA MANAGEMENT PLAN

Products of the Research

- The types of data, samples, physical collections, software, curriculum materials, and other **materials to be produced in the course of the project.**

Data Formats

- The **standards to be used for data and metadata format** and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

Access to Data and Data Sharing Practices and Policies:

- Policies for **access and sharing** including provisions for **appropriate protection** of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Policies for Re-Use, Re-Distribution, and Production of Derivatives.

Archiving of Data:

- Plans for archiving data, samples, and other research products, and for **preservation of access** to them.

ONLINE DATA MANAGEMENT PLANNING TOOL

<https://dmptool.org>



[Home](#)

[DMP Requirements](#)

[Public DMPs](#)

[News](#)

[Help](#)

[Contact Us](#)

[About](#) ▼

 [Log In](#)

Data Management Planning Tool

Create, review, and share data management plans that meet institutional and funder requirements.

[Get Started](#)

MANDATE BY FUNDING AGENCIES

Require a Data Management Plan (DMP)

- National Science Foundation (NSF)
- National Institutes of Health (NIH)
- National Oceanographic and Atmospheric Research (NOAA)
- Institute of Museum and Library Services (IMLS)
- National Endowment of Humanities – office of digital humanities (NEH)

Require Sharing of Results – per a Data Policy

- Andrew W. Mellon
- Bill & Melinda Gates Foundation
- NASA
- NEH – Preservation & Access
- IES – Institute of Education Sciences
- Wellcome Trust

DATA SHARING POLICIES BY PUBLISHERS

The screenshot shows the Nature.com website with the URL www.nature.com/authors/policies/availability.html. The page title is "authors & referees". The main content area is titled "Availability of data, material and methods". The text states: "An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must **also** be disclosed in the submitted manuscript." It further states: "After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings." Below this, it says "See sections below for details on:" followed by a list of links: [reporting requirements](#), [availability of data](#), [availability of materials](#), [availability of computer code](#), [experimental protocols](#), [clinical trials](#), and [further reading](#). The left sidebar contains a "Site content" menu with items like "Homepage", "Policies", "Publication ethics", "Bioethics", "Availability of data & materials", "Peer-review policy", "Embargo", "Corrections", "License to publish", "Feedback", "Author resources", "Peer review", "Nautilus blog", and "Peer-to-peer blog". A "SUBMIT NOW" button is visible in the top right, and an advertisement for "nature SPECIAL BUILDING THE 21ST CENTURY" is at the bottom right.

..authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications

Unstructured repositories like figshare and Dryad are suitable alternatives if no structured public repositories exist.

As a less desirable alternative, data sets can be made available as Supplementary Information files, which will be freely accessible on nature.com upon publication

DATA SHARING POLICIES BY PUBLISHERS

The screenshot shows the Royal Society of Chemistry website. The browser address bar displays the URL: www.rsc.org/journals-books-databases/journal-authors-reviewers/prepare-your-article/experimental-data/. The page features the RSC logo, navigation links for 'About us', 'Membership & professional community', 'Campaigning & outreach', 'Journals, books & databases' (highlighted), 'Resources & tools', 'News & events', and 'Locations & contacts'. A search bar with 'Google™ Custom Search' is visible. The breadcrumb trail reads: Home > Journals, books & databases > Journal authors & reviewers > Prepare your article > Experimental data. The main content area has a background image of an open book and contains the following text:

Experimental data

Information about our data policy and experimental data you need to include

On submission of a manuscript authors should provide all data required to understand and verify the research presented in the article.

we encourage authors to deposit as much data as possible that is related to the research in their article.

This should be in appropriate and publically available repositories

RESEARCH DATA MANAGEMENT SOFTWARE

Online Repositories/Infrastructures created to manage a researcher's data (sharing, archiving, preservation, metadata)

May be hosted or installed on a university's server

Each software contains different ranges of management/collaborative options

Open source and proprietary options

DATA ARCHIVING PLATFORMS

Institution Repository with Data

- DSpace
- Fedora
- BePress Digital Commons
- Hydra
- Drupal

Data Specific Repositories

- Dataverse
- HubZero
- NADA (Social Science and Survey Data)
- CKAN/DKAN
- Custom.

CLOUD BASED INDIVIDUAL/INSTITUTIONAL PLATFORM



search figshare (titles, tags, authors, etc.)



[Browse](#)

[Upload](#)

[Sign up](#)

[Login](#)

Raw data for Bitam et al., 2015 'An unexpected effect of TNF α on F508del-CFTR maturation and function.'

| | |
|--|--|
|  Uncropped_gel_images_for_Figures_1a...tif  | preview download |
|  Figure_1B,_D,_Figure_8B_and_Supplem...xlsx  | preview download |
|  Fig_3_B_raw_data.xlsx  | preview download |
|  Fig_3_C_raw_data.xlsx  | preview download |
|  Fig_3_D_raw_data.xlsx  | preview download |
|  Fig_4_raw_data.xlsx  | preview download |
|  Fig_7_B_raw_data.xlsx  | preview download |

3
views

0
shares

cites
coming
soon

Published on 11 Aug 2015 - 08:34 (GMT)

Filesize in total is 8.55 MB

This data is part of the peer reviewed publication:

An unexpected effect of TNF- α on F508del-CFTR maturation and function

[FICOOResearch](#)

[Browse more](#)

[Download all](#)

CLOUD BASED INDIVIDUAL/INSTITUTIONAL PLATFORM



The screenshot shows the top portion of the DataDryad.org website. At the top left is the DataDryad logo, which consists of a green tree icon surrounded by a grid of dots, with the word "DRYAD" in green capital letters below it. To the right of the logo is a navigation menu with the following items: "About", "For researchers", "For organizations", "Contact us", "Log in", and "Sign up". In the top right corner, there are social media icons for Twitter, Facebook, and RSS. Below the navigation menu is a large white box with a green border containing the following text: "DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome." Below this text are four small black dots. To the right of this box is a green button labeled "Submit data now" and a link "How and why?". Below these is a search section titled "Search for data" with a search input field containing the placeholder text "Enter keyword, author, title, DOI, et" and a green "Go" button. Below the search field is a link for "Advanced search". At the bottom of the screenshot, there are two sections: "Browse for data" on the left and "Latest from @datadryad" on the right.

When using this data, please cite the original publication:

Quirk J, Leake JR, Johnson DA, Taylor LL, Saccone L, Beerling DJ (2015) Constraining the role of early land plants in Early Palaeozoic weathering and global cooling. *Proceedings of the Royal Society B* 282(1813): 20151115. <http://dx.doi.org/10.1098/rspb.2015.1115>

Additionally, please cite the Dryad data package:

Quirk J, Leake JR, Johnson DA, Taylor LL, Saccone L, Beerling DJ (2015) Data from: Constraining the role of early land plants in Early Palaeozoic weathering and global cooling. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.6dh6g>

[Cite](#) | [Share](#)




RESEARCH DATA REPOSITORY REGISTRY



[Home](#) [Search](#) [Browse](#) [Suggest](#) [FAQ](#) [About](#) [Schema](#) [API](#) [Contact](#) [Imprint](#)

Search for Repositories (1309 reviewed repositories)

▼

| | | |
|---|--|--|
| Subject <input type="text" value="Add subjects"/> | Content Type <input type="text" value="Add content types"/> | Country (of the responsible institutions) <input type="text" value="Add countries"/> |
| <input type="checkbox"/>  Certificates | <input type="checkbox"/>  Open Access | <input type="checkbox"/>  Persistent Identifier |

WHY LIBRARIES

Significant expertise

- Metadata
- Archival management
- Policy development

Organizational experience and stability

- Process and results driven

Culture of trust

- Responsible guardians of the cultural record
- Service oriented
- Respectful of privacy and intellectual property

LIBRARIES : THE HIGH SCORER

| Parameter | Researcher/Faculty | Library | IT Support (Systems) |
|----------------------------------|--------------------|---------|----------------------|
| Rapid Response on Research Front | Green | Yellow | Green |
| Grabbing Funds | Green | Yellow | Green |
| Metadata and IT | Red | Green | Green |
| Sustainability | Red | Green | Red |
| Attitude to work collaboratively | Red | Green | Red |
| Heritage of Preservation | Red | Green | Red |

LIBRARIES CAN

Data acquisition,
ingest layer

- Selection, taxonomy, ontology, metadata, workflow

Preservation layer

- Archival retention, format migration, quality assurance, trust

Physical layer

- Storage, network security, reliability standards

Service layer

- Discovery, retrieval, data mining, data visualization

Management layer

- Administration, budget, policy, development



