Identifying and overcoming barriers to the effective use of research data

17 March 2023 via webinar

Dr Dugald McGlashan Dr Caroline Hadley

dugald.mcglashan@inlexio.com caroline.hadley@inlexio.com

www.inlexio.com @inlexio



TODAY'S TALK

- Framing the issue
- Exploring barriers to the effective use of data
- What journals are doing and can do



Framing the issue



For a more effective research system, scholars need to be able to easily and consistently access goodquality research data that is reusable.



For a more effective research system, scholars need to be able to easily and consistently access good quality **research data** that is reusable.

OECD definition (2007):

"Research data" are defined as **factual records** (numerical scores, textual records, images and sounds) used as **primary sources** for scientific research, and that are commonly accepted in the scientific community as necessary to **validate** research findings."

Materials, code, and analytical techniques are under increasing focus e.g. PLOS Biology Staff (2020) Low availability of code in ecology: A call for urgent action. *PLOS Biology* 18(12): e3001048

Source: www.oecd.org/science/inno/38500813.pdf



Exploring barriers to the effective use of data

RESEARCHERS

For a more effective research system, **scholars need** to be able to easily and consistently access good quality research data that is reusable.

The focus must be on researchers as the generators and main users, but other parties can be, or are, influential:

- Funders
- Institutions
- Data repositories
- Open science and data proponents, scholars of research data
- Journals and publishers, Editors-in-Chief
- Technologists and production providers
- Libraries
- Downstream public users



RESEARCHERS

For a more effective research system, **scholars need** to be able to easily and consistently access good quality research data that is reusable.

Challenges

- Lack of knowledge of the issue(s)
- Time, complex processes, cost, inertia/apathy: change!
- Ethics issues
- Balancing other priorities (misplaced or missing incentives)
- Loss of control and misuse of data; security issues
- Licensing confusion
- Lack of persistence of data repositories; preservation concerns
- No citation standards
- Can't find or re-use data.

Initiatives

Many, as we will explore in this talk.

RESEARCH SYSTEM

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is reusable.

Challenges

What is "effective"?

- trustworthy
- more efficient
- faster
- better return on investment.

Requirement for data to be:

- reproducible
- repeatable
- replicable



9 Effective use of research data

RESEARCH SYSTEM

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is reusable.



Open Science Image (CC BY): www.fosteropenscience.eu/taxonomy/term/139



10 Effective use of research data

EASILY ACCESS

For a more effective research system, scholars need to be able to **easily** and consistently **access** good quality research data that is reusable.

Challenges

- Data are not shared or made available
 - "Available upon request" often means unavailable in practice
- Data are not in an easily accessible format or location
- Data are not discoverable (no metadata or poor metadata, lack of proper citation standards, lack of integration with e.g. Scholix)
- Ethics and risks of sharing
- Data are unlicensed or have restrictions
- Lack of tools to find data and datasets.



EASILY ACCESS

For a more effective research system, scholars need to be able to **easily** and consistently **access** good quality research data that is reusable.

Initiatives

- Funder initiatives: Nelson memo, EU directives
- Policy development
- Date repository ecosystem is developing
- Metadata standards
- Citation standards
- Discoverability –

mapping science: e.g. OpenAIRE Graph



https://graph.openaire.eu/



CONSISTENTLY ACCESS

For a more effective research system, scholars need to be able to easily and **consistently access** good quality research data that is reusable.

Challenges

- Persistence and quality of data repositories
- Financial sustainability
- Technology challenges.



CONSISTENTLY ACCESS

For a more effective research system, scholars need to be able to easily and **consistently access** good quality research data that is reusable.

Initiatives

- Agreed standard for repositories
- Approved lists and quality assessment of repositories
- Large investment by governments.

https://fairsharing.org/; https://www.re3data.org/; https://www.repositoryfinder.datacite.org



CONSISTENTLY ACCESS

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is reusable.

Initiatives

• Global investment in infrastructure, policy development, product development, etc.





GOOD QUALITY

For a more effective research system, scholars need to be able to easily and consistently access **good quality** research data that is reusable.

Challenges

- Data are not in an accessible or appropriate format
- Data are not curated or reviewed
- Data curation ("stewardship") takes expertise, extra costs
- Data and/or data repositories can have uncertain provenance, traceability or ethics standards
- Published standards are not implemented by stakeholders (authors, data repositories, publishers, etc.)
- Frequent proposals of new standards, high turnover of (potential) standards.



GOOD QUALITY

For a more effective research system, scholars need to be able to easily and consistently access **good quality** research data that is reusable.

Initiatives

- Basic review of data and associated data during publication process
- FAIR and TRUST initiatives
- Projects and training around data curation (e.g. data stewardship diploma at the University of Vienna)
- Research Data Alliance working groups
- Quality control by repositories (metadata and files are checked).

TRUST: https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories-0 https://www.openaire.eu/openaire-survey-who-is-managing-research-data https://fairsharing.org/; https://www.re3data.org/; https://www.repositoryfinder.datacite.org



REUSABLE

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is **reusable**.

Challenges

- Unusable formats (proprietary, in PDF)
- No metadata, poor metadata
- No licensing information
- Not using standard community-accepted citation standards.



REUSABLE

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is **reusable**.

Initiatives

• FAIR principles

"If data is Findable, Accessible and Interoperable, it tends to be Reusable"

To be reusable:

- (meta)data are released with a clear and accessible data usage license
- (meta)data are associated with detailed provenance
- (meta)data meet domain-relevant community standards.

https://www.nature.com/articles/sdata201618



REUSABLE

For a more effective research system, scholars need to be able to easily and consistently access good quality research data that is **reusable**.

Initiatives

- Development of standards during publication and other processes, e.g.
 - Authors should use proprietary-free, open data formats
 - Metadata should comply with the latest standards (and publishers and others should update older records as necessary)
- Appropriate licensing should be used
- Citation information should be full and appropriate
- Journals are undertaking peer review of data sets to increase repeatability.



SUMMARY – JOURNAL PERSPECTIVE

Barriers for journals in the effective use of research data include:

- Awareness and understanding of the issues
 - Is the imperative there?
 - Language barriers
 - Lack of subject and technical expertise
- Extra time, costs, staff needed to:
 - Consider and introduce new policies
 - Implement policies (which is ongoing)
 - Assess data repositories
 - Monitor and respond to change.
- Increasing burdens on authors can affect submissions
- Data is and are complicated!



What journals are doing and can do

TRENDS IN DATA POLICY

~2010-2015

What policies are in place? What are needed?

- Some publishers mandate data sharing
- Unclear, difficult to find descriptions
- No standards, lack of harmony between publishers
- Research Data Alliance
 (RDA) launched (2013)

~2015-2020

Policies defined and standardized

- Meetings, workshops, studies
- Two key articles:
 o Cousijn et al (2018)
 - o Hrynaszkiewicz et al. (2020)

2020–present~ Implementation, review, refinement

- 'Big 4' publishers implement data policies
- Continuing work on metadata/JATS standards
- Work now focused on implementation efficacy:
- e.g. quality of DASs
- e.g. template effectiveness
- author compliance



GLOBAL INITIATIVES: POLICIES

Journals now have a research data policy as a standard practice

- 'Big 4' publishers have research data policy directives
- Studies show majority of journals have research data policies
 - e.g. 10 of 14 studies showed >50% journals have policies
 - Fewer journals mandate data sharing than encourage
 - Differences between commercial publishers and independent publishers
- Spans most, if not all sciences; work is active in the humanities and social sciences
- As an aside, journals with higher Journal Impact Factors are more likely to have data sharing policies.

https://link.springer.com/article/10.1007/s11192-020-03467-9 https://doi.org/10.1080/08989621.2019.1591277



JOURNAL POLICY DEVELOPMENT

Hrynaszkiewicz et al. (2020) Developing a research data policy framework for all journals and publishers. *Data Sci J* 19:5

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	0	0	0	0	0	0
Exceptions to policy	0	0	0	0	•	•
Embargoes	0	0	0	•	٠	•
Supplementary materials	0	0	0	•	•	•
Data repositories	0	0	0	•	•	•
Data citation	0	0	0	0	•	•
Data licensing	0	0	0	0	0	0
Researcher/ author support	•	•	•	•	•	•
Data availability statements		0	•	•	٠	•
Data formats and standards				0	0	•
Mandatory data sharing (specific data types)				٠	•	٠
Mandatory data sharing (all papers)				0	•	•
Peer review of data				0	0	•
Data Management Plans (DMPs)				0	0	0

The article provides templated standard texts dependent on the policy position chosen by the journal

See also JST Research Data webinar (2020, link below)

https://doi.org/10.5334/dsj-2020-017 https://www.youtube.com/watch?v=EdF4fAfPo80



GLOBAL INITIATIVES: WORKFLOWS

Journals are working to improve workflows for authors

- Integration with repositories
 - assessing data repositories
 - providing a list of endorsed repositories
- Provision of policies and instructions, consistent with agreed standards
- Provision of text templates
- Prompts from journal staff at relevant times of the publication process.

Availability of data	Template for data availability statement
Data openly available in a	The data that support the findings of this study are openly available in
public repository that	[repository name e.g "figshare"] at http://doi.org/[doi], reference
issues datasets with DOIs	number [reference number].

Wiley



GLOBAL INITIATIVES: DATA CITATION

Open Access Published: 20 November 2018

A data citation roadmap for scientific publishers

Helena Cousijn 🖂, Amye Kenall, Emma Ganley, Melissa Harrison, David Kernohan, Thomas Lemberger,

Fiona Murphy, Patrick Polischuk, Simone Taylor, Maryann Martone & Tim Clark

Scientific Data 5, Article number: 180259 (2018) Cite this article

enabling authors to cite datasets in the same way as they cite articles

https://doi.org/10.1038/sdata.2018.259



- Help authors share data easily, efficiently and in compliance with other stakeholders' requirements
- Reduce barriers for authors to be able to share (knowledge, time, processes, cost, etc.)
- e.g. deposition of data sets into third-party repositories as part of the submission process (PLOS)



Bloom et al (2015): Workflows for Research Data Publishing

https://datascience.codata.org/articles/10.5334/dsj-2021-031/ https://theplosblog.plos.org/2021/10/plos-dryad-integration/ https://zenodo.org/record/20308



Establish a research data policy

- Be consistent with global best practice in your subject area
- Allow for appropriate exceptions on ethics grounds
- Align with funder and institution policies and mandates
- Have a list of recommended repositories
- Mandate DASs, data sharing, data peer review
- Check for compliance



https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230416



Implement and use emerging publishing technologies

- Strongly encourage, if not require, use of ORCID
- Introduce other persistent identifiers (PIDs) (e.g. ROR, upcoming grant ID system)
- Update citation processes and practices
- Update journal technology to include all relevant metadata in JATS compliant formats
- Collaborate with JST with to introduce a J-STAGE Data channel (and Jxiv)



- Prepare materials to help authors navigate the new issues and procedures
 - e.g. brochures, toolkits, how-to's, etc.
 - Include any incentives, such as:
 - evidence for an increase in article citations if data are shared
 - o compliance with funder mandates
 - o common good.
- Actively prompt authors at submission or acceptance to share their research data.

Open Data for Open Science

Introducing the JMSJ channel on J-STAGE Data



- Enhancing data sharing and research reproducibility
- Making data more transparent, accessible and reusable
- Fostering stronger research collaboration

The *Journal of the Meteorological Society of Japan* (JMSJ) has created a channel on J-STAGE Data to enable authors to deposit, share and link to the data within their articles.

JMSJ encourages authors to deposit the data underlying their JMSJ articles to the JMSJ channel on J-STAGE Data.

What is J-STAGE Data?

J-STAGE Data is a repository provided by JST, the Japan Science and Technology Agency.

J-STAGE Data is free to use, and user registration is not required to download data.

Each data item on J-STAGE Data and the corresponding article on J-STAGE are linked to each other.

Each dataset is assigned a DOI, allowing it to be integrated into the publishing ecosystem. Why use J-STAGE Data?

Data sharing and availability underpin Open Science, allow for the innovative reuse of data, and enhance research collaboration.

Open Science drives progress in reproducibility, transparency and therefore science itself.

J-STAGE Data enables JMSJ authors to participate in the development of Open Science.

https://jstagedata.jst.go.jp/jmsj



SUPPORT FOR J-STAGE JOURNALS

Research data policies

- Text and examples are available for use
- JST has been providing advisory services

Infrastructure via JST

- J-STAGE an online platform for Japanese academic journals
- J-STAGE Data a data repository for research data related to articles published on J-STAGE
- Jxiv a preprint server for disseminating research results before formal publication.







Jxiv

JOURNAL TO DO LIST

Priority actions for journals:

- Implement an appropriate data policy
- Update your data citation policies and practices
- Produce and provide materials for authors
- Ensure data and article metadata standards are up to date (work with your production service providers)
- Work with JST to launch a J-STAGE Data channel
- Work to improve data sharing rates, quality of materials (e.g. DASs, datasets).



FUTURE RESEARCH ECOSYSTEM?



34 Effective use of research data



ありがとうございました! THANK YOU!

Contact us: dugald.mcglashan@inlexio.com caroline.hadley@inlexio.com <u>inlexio.com</u>

Disclaimer

INLEXIO provides no warranties or guaranties, express or implied, with respect to the content and analysis provided in this document and in any discussions or to any omissions. Any implied warranties of fitness for any particular purpose are expressly disclaimed. None of the advice should be construed as formal legal advice and the client should obtain appropriate legal advice relevant to their situation.