Referencing data in publications: principles, policy, and practice
The Australian Academy of Science, The Shine Dome, Canberra
28 October 2015

Data Citation: framing the discussion and global context

Dr Simon Hodson
Executive Director, CODATA
www.codata.org

Building a Culture of Data Citation

http://www.ands.org.au/guides/data_citation_poster.pdf

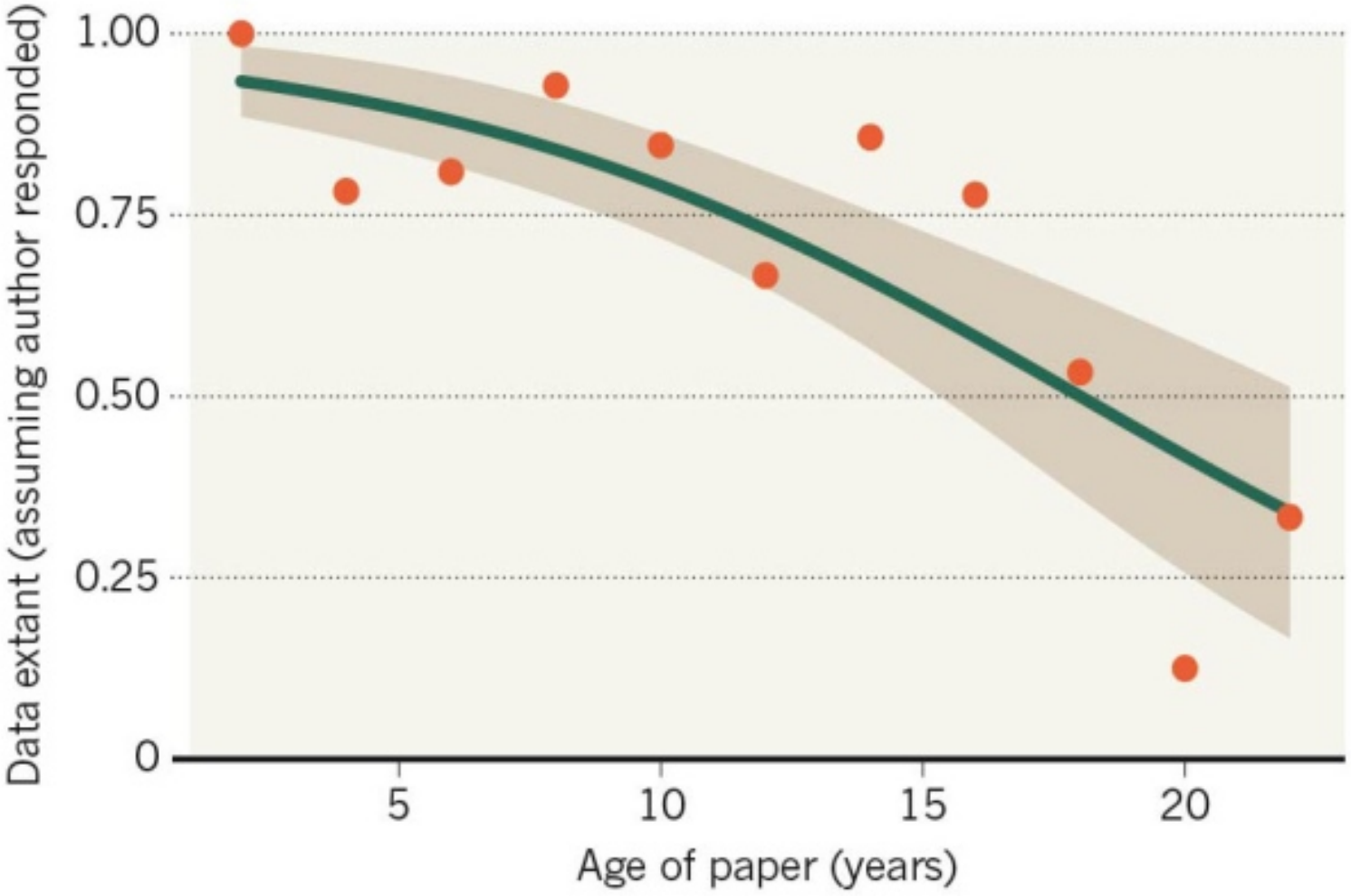
Barriers to Data Sharing

Researchers concerns:

- Concern that data may be misused or misunderstood.
- Concern that will lose scientific edge if sharing before fully exploited.
- Desire to retain control of a professional asset.
- **Concern that will not be credited.**
- **Lack of career rewards for data publication.**
- See ODE report, using Parse.Insight findings: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf
- Culture in particular research disciplines; availability of infrastructure.
- **Fundamentally, researchers are reluctant to expend effort sharing data because they do not feel that data is adequately exposed or credited.**

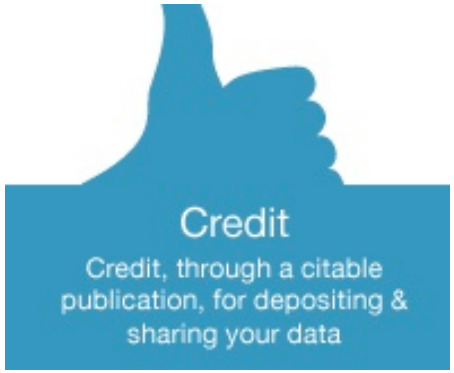
Nature special issue on data sharing: http://www.nature.com/news/specials/datasharing/index.html

80% of ecology data irretrievable after 20 years



Vines TH *et al*. (2013) Current Biology DOI:10.1016/j.cub.2013.11.014

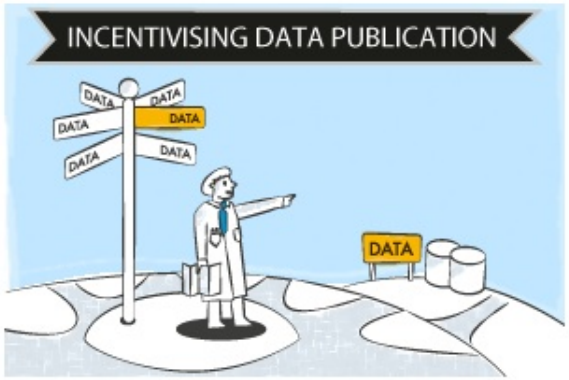
Individual Benefits?



get credit for all your research

citable

- All published research is citable
- Cloud based service
- Always available



Ubiquity Press Metajournals feature peer reviewed papers that promote discovery, reuse and citation of research outputs such as data and software. Publish your data and software to create new collaborations and to increase the impact of your work.

THE
DATA CITATION INDEX℠
CONNECTING THE DATA TO THE RESEARCH IT INFORMS

What is it?
VIEW VIDEO

Professional Benefits of Data Sharing and Citation

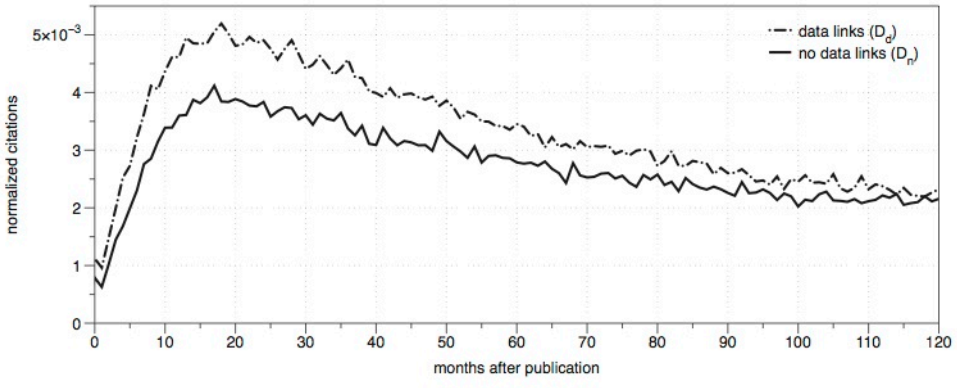


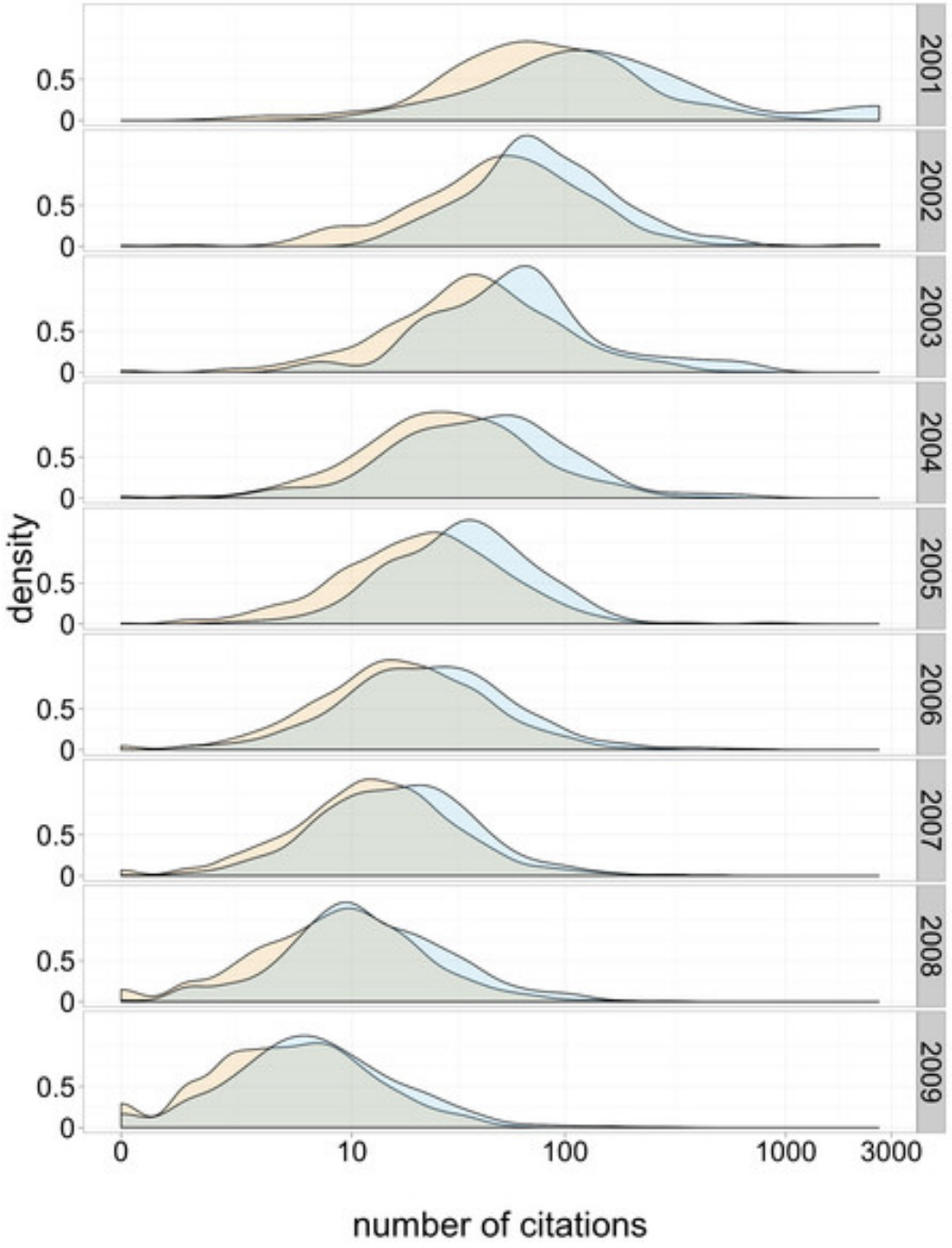
Figure 2. The normalized number of citations for data sets D_d and D_n. The citations have been normalized by the total number of citations.

'articles with links to data result in higher citation rates than articles without such links.'

Henneken, Accomazzi, 'Linking to Data - Effect on Citation Rates in Astronomy' http://arxiv.org/pdf/1111.3618v1.pdf

'We find strong and consistent evidence that data sharing, both formal and informal, increases research productivity across a wide range of publication metrics. **Data archiving, in particular, yields the greatest returns on investment with research productivity (number of publications) being greater when data are archived.** Not sharing data, either formally or informally, limits severely the number of publications tied to research data.'

Pienta, Alter, Lyle (2010) *The Enduring Value of Science Research: The Use and Reuse of Primary Research Data* http://hdl.handle.net/2027.42/78307



Citation advantage of having archived Gene Expression Omnibus data

Examined **10,555** studies that created gene expression microarray data, comparing those that made data available and those that didn't.

Studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available.
Increased citation of 30% for those published 2004-5.

Piwowar and Vision (2013), PeerJ DOI:10.7717/peerj.175

Developments in Data Citation

- **ICSU, *International Council for Science*, Statement on 'Open access to scientific data and literature and the assessment of research by metrics', Sept 2014 http://bit.ly/icsu-OA-statement**
- Endorses the OECD Principles and Guidelines on Access to Data from Publicly Funded Research (2007)
- Recommendation 4: '**Science publishers and chief editors of scientific publications should require authors to provide explicit references to the datasets underlying published papers, using unique persistent identifiers.** They also should require clear assurances that these datasets are deposited and available in trusted and sustainable digital repositories. **Citing datasets in reference lists using an accepted standard format should be considered the norm.**'

ICSU
International Council for Science

Need for Improved Metrics

- DORA, the San Francisco ***Declaration on Research Assessement*** http://dmm.biologists.org/content/early/2013/05/16/dmm.012955.full.pdf
- **Research assessment should include value and impact of all research outputs (including data and code) as well as qualitative indicators.**
- **Shift to article-based metrics, rather than journal-based metrics.**

- ICSU, International Council of Science Statement: http://bit.ly/icsu-OA-statement
- Endorses DORA **(Recommendation 11)**
- **Recommendation 10:** In research evaluation and assessment, metrics should be regarded as an aid, and not a substitute, for good decision-making. They should not normally be used in isolation to assess the performance of researchers, to determine appointments, or to distribute funds to individuals or research groups, for which expert review is indispensable.

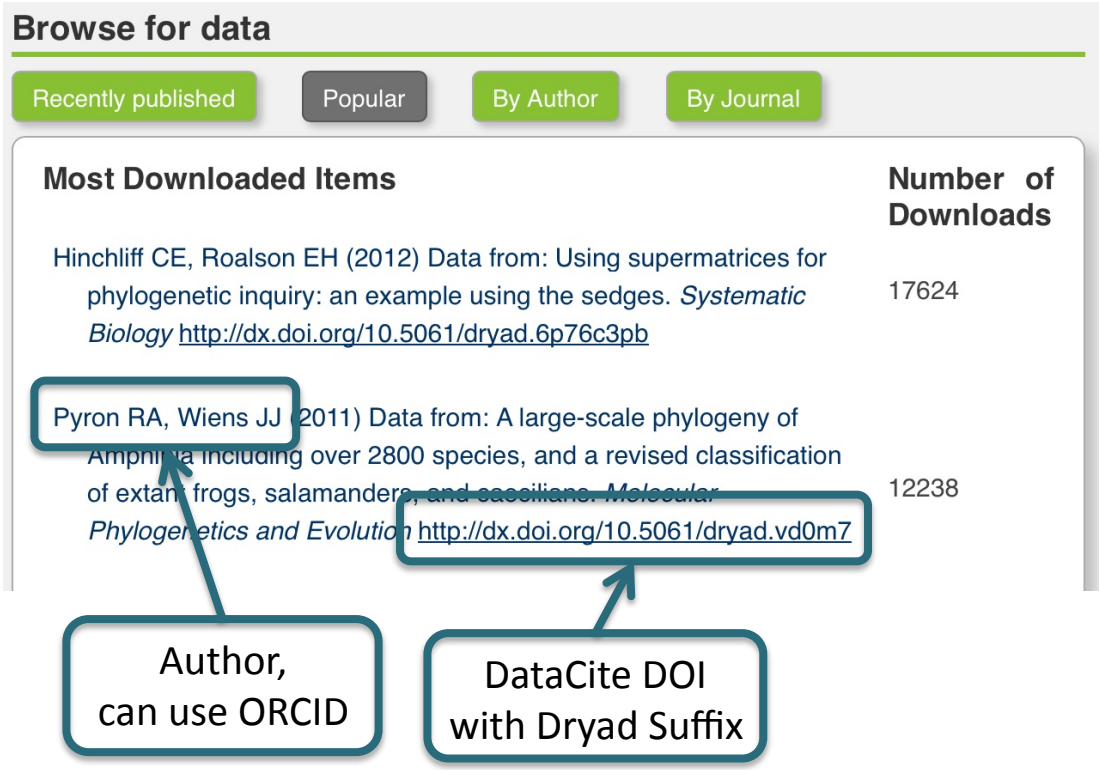
'Do not use journal-based metrics, such as journal impact factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion or funding decisions.'

Developments: Journal Data Policies

- **Dryad Joint Data Archiving Policy, Feb 2010: http://datadryad.org/jdap**
- This journal **requires, as a condition for publication**, that data supporting the results in the paper should be archived in an appropriate public archive, such as GenBank, TreeBASE, Dryad, or the Knowledge Network for Biocomplexity.

- **PLOS Data Availability Policy, revised Feb 2014:** http://www.plosone.org/static/policies.action#sharing
- PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exceptions.

Data Citations, DOIs and ORCIDs



- Assists with associating data citations unambiguously with a given author.
- Increases visibility of data sources.
- **ODIN ORCID/DataCite Claim tool allowing authors to link dataset records with DataCite DOIs to their ORCID Profile.**
- Adoption of PIDs and interoperability of those PIDs is an important enabler of data citation, data sharing.

ORCID Auto-Update

Researchers can now set their ORCID profiles to automatically update with any published article or dataset associated with their ORCID.

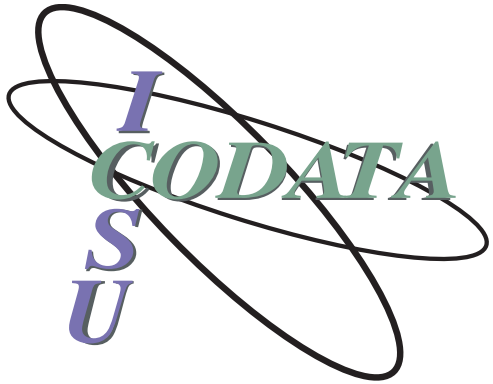
http://blog.datacite.org/auto-update-has-arrived/

FORCE11 Data Citation Implementation Group (DCIG)

- Clark, Starr et al, 'Achieving human and machine accessibility of cited data in scholarly publications', *PeerJ Computer Science* 1:e1 https://dx.doi.org/10.7717/peerj-cs.1
- Responsibilities of data archives in making the components for a citation available and ensuring persistence of source, landing page.
- Responsibilities of journals in ensuring data is credited through citation and including data citations and included in the reference list and so visible to citation indexes.

What should publishers do?

- **JATS 'Journal Article Tag Set', XML tags for journal articles.** This has recently been modified to better facilitate citing data and including data citation in the reference list: Citing Data in Journal Articles using JATS https://www.force11.org/sites/default/files/d7/project/882/citing-data-in-jats-2015-06.pdf
- Main issues for publishers are (thanks Tim Clark):
  - Adopt JATS 1.1d3 or a later revision, to support data citation based on the JATS model;
  - Ensure that new JATS elements translate properly into publisher HTML and PDF representations;
  - Adopt a standard list of appropriate repositories for varying kinds of data;
  - Require authors to provide a valid accession number from an approved repository, as a condition for completing peer review;



ICSU

International Council for Science

Thank you for your attention!

Slide credits: Louise Corti

Simon Hodson

Executive Director CODATA

www.codata.org

http://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org

Email: simon@codata.org

Twitter: @simonhodson99

Tel (Office): +33 1 45 25 04 96 | Tel (Cell): +33 6 86 30 42 59

CODATA (ICSU Committee on Data for Science and Technology), 5 rue Auguste Vacquerie, 75016 Paris, FRANCE