Walk softly and carry a large carrot: how to give credit for academic work

Sarah Callaghan[1], Todd Carpenter,[2] John Ernest Kratz[3]

[1]British Atmospheric Data Centre, STFC Rutherford Appleton Laboratory, Harwell Oxford, Didcot, OX12 7DQ, UK. sarah.callaghan@stfc.ac.uk
[2]National Information Standards Organization (NISO). tcarpenter@niso.org
[3]California Digital Library. John.Kratz@ucop.edu

Abstract

Researchers want to know how their work impacts their communities, and the wider world, including research outputs other than peer-reviewed journal publications. The journal article provides a way of claiming and defining an area of intellectual work, and citation of articles allows others to acknowledge that work. Yet the paper can only give an overview of the work - it is not possible to fit into a paper everything needed to make the work fully reproducible.

For providing credit (and for making recruitment and promotion decisions) we abstract the paper further. Instead of reading every citing paper, we instead count the citations, reckoning this an appropriate proxy for the quality of the paper, and hence the described work.
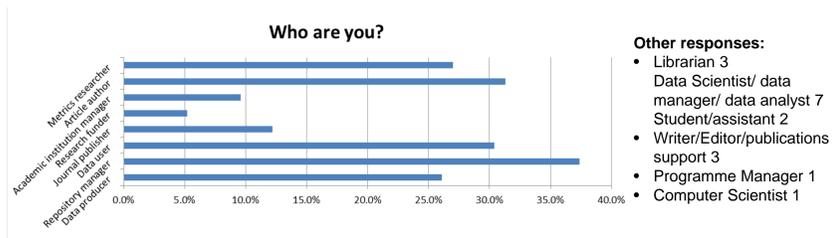
Citation counts for datasets are one of the "carrots" promised to researchers for their efforts in citing and publishing data, also producing a metric by which the quality of a dataset can be evaluated. Quality is an especially slippery concept when it comes to data, which can be of good quality for one purpose, and bad for another. Measuring the impact of research directly is difficult, so we resort to measuring what we can (i.e., number of citations). Care must be taken with indirect measurements to ensure that they map appropriately to what we really want to measure.
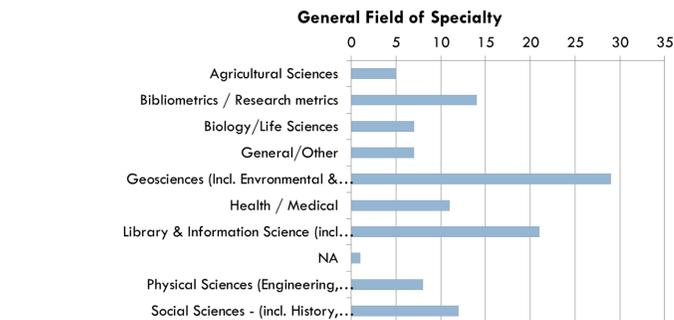
Survey results

This survey was carried out by the RDA/WDS Publishing Data Bibliometrics Working Group. It asked interested parties what they currently used to evaluate the impact of data, and what they would like to use in the future.

The survey was carried out via a web-based survey system (SurveyMonkey) and invitations to participate were distributed widely, mainly through mailing lists for interested parties. There were 115 respondents.
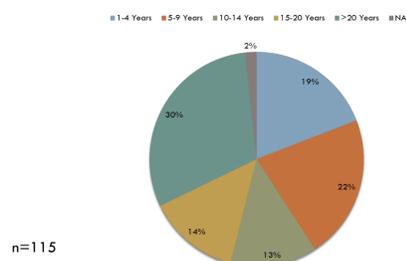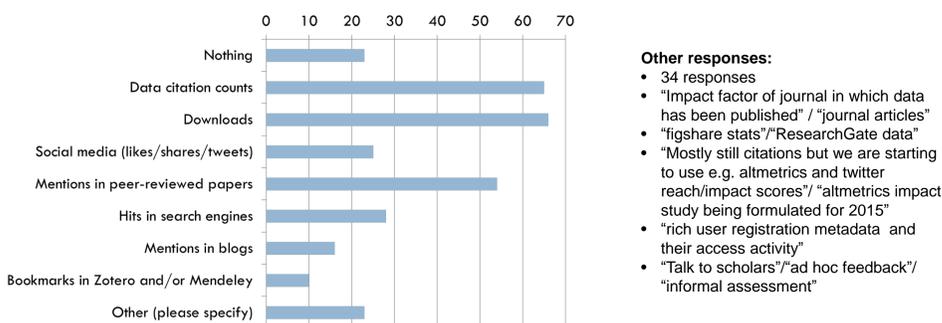
1) Who Responded?

Who are you?

Other responses:
- Librarian 3
- Data Scientist/ data manager/ data analyst 7
- Student/assistant 2
- Writer/Editor/publications support 3
- Programme Manager 1
- Computer Scientist 1

2) What is your field of specialty?

General Field of Specialty

3) How long have you been working in this field?

n=115

4) What do you currently use to evaluate the impact of data?

Other responses:
- 34 responses
- "Impact factor of journal in which data has been published" / "journal articles"
- "figshare stats"/"ResearchGate data"
- "Mostly still citations but we are starting to use e.g. altmetrics and twitter reach/impact scores"/ "altmetrics impact study being formulated for 2015"
- "rich user registration metadata and their access activity"
- "Talk to scholars"/"ad hoc feedback"/ "informal assessment"

5) If you don't use anything to evaluate the impact of data, why?

33 responses
- **Majority opinion**: current metrics not good enough, no standards, don't know what to do
- Other opinions:
  - Impact metrics not important for respondee
  - Interest in quantifying impact, but repository/policies still under development
  - Metrics are too easily gamed, or too complicated

5) Are the methods you use to evaluate impact adequate for your needs? If not, why not?


Are the methods you use to evaluate impact adequate for your needs?
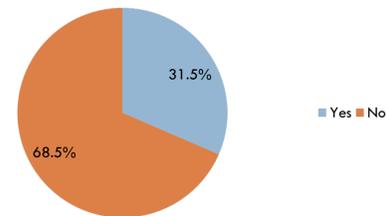
75 responses to "if not, why not?:
- Lack of tools and standards
- Limited data citation
- Available measures not good enough. Too difficult/time consuming
- Need to focus on other (non-scientific) impacts (e.g. planning/educational use of data)
- Hoping increased use of DOIs will help situation

6) Why do you want to evaluate the impact of data?

n=111
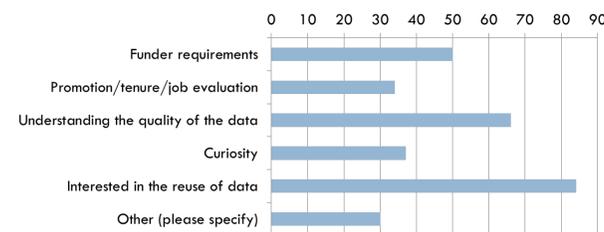
Other responses:
- 30 responses – mostly different
- Measure organisational impact
- Encourage data openness/publication (and provide benefit for data producers)
- Knowledge of who uses data and why – improve access for users - inform data retention decisions – improve user experience – justify repository investment – discover impact of data centre policies
- Evaluate employees performance/institutional requirement

7) In the future, what would you like to use to evaluate the impact of data?

- 96 responses
- **Majority opinion:** Data citations (and better tools to track them)
- Other opinions:
  - Downloads
  - Altmetrics / "anything and everything"
  - Peer review / community feedback
  - Use outside scholarly literature (e.g., in patents)
  - Reuse / "actual use"

9) What is currently missing and/or needs to be created for bibliometrics for data to become widely used?

- 92 responses
- **Majority opinion:** Standards
- Other opinions:
  - Data Citation
  - Consistent use of PIDs / DOIs
  - Culture Change / "A belief that they are valid."

10) What difference would it make to your work to be able to evaluate the impact of data?

- 91 responses
- Promote data sharing/publication/reuse/data stewardship – provide credit for data producers
- Justify funding for data activities
- Other criteria for evaluation of research impact
- Inform and prioritise data access systems – improve services
- Influence in public-policy decisions making

Further Work

Obviously, there is a lot of interest in this area, and for bibliometrics for data to be accepted and used widely, standards and tools must be designed to work for the community. Concerns were raised about the possibility of "gaming" metrics, and that solutions adopted should be open and free to use.

If you are interested in contributing to this work, please join the RDA/WDS Publishing Data Bibliometrics Working Group: https://rd-alliance.org/groups/rdawds-publishing-data-bibliometrics-wg.html