# Publishing and sharing data papers can increase impact and benefits researchers, publishers, funders and libraries

10/12/2016

The process of compiling and submitting data papers to journals has long been a frustrating one to the minority of researchers that have tried. **Fiona Murphy**, *part of a project team working to automate this process, outlines why publishing data papers is important and how open data can be of benefit to all stakeholders across scholarly communications and higher education.*
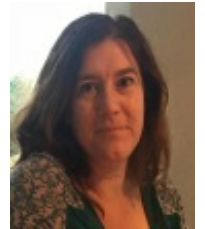
Giving Researchers Credit for their Data – or 'Data2Paper' as we're now more snappily calling it – is a cloud-based app which uses existing DataCite and ORCID-derived metadata to automate the process of compiling and submitting a data paper to a journal without the researcher having to leave the research space or wrestle directly with the journal's submission system (an occasional source of frustration):

Sound too good to be true? Well, at the time of writing it is. However, a small project team – led by Bodleian Libraries and F1000Research – is working hard to make it reality.

Part of Jisc's Research Data Spring initiative, Data2Paper is now in Phase 3. We've built on work done by the WDS-RDA Publishing Data Workflow Working Group on data publishing, run a survey for stakeholders to establish the baseline demand, and produced a (so far silent) demonstration video. Now we're building a live end-to-end workflow for testing with real authors, data sets, repositories and journals. Partners for this phase include the University of Manchester, Mendeley Data and Elsevier, but we've also had helpful input from ORCID, Figshare, Project THOR and SURF as well as expressions of interest from a wide range of publishers and repositories.

What are we hoping to achieve with the app? As well as improving the lives of researchers wishing to publish data papers using data sets, we believe it could prove beneficial to a range of key stakeholders:

- *Funders* – this service encourages better research data management
- *Researchers* are more likely to engage with the repositories if they are likely to derive a citable research object at the cost of a few minutes' work. There would be additional metrics available, as well as better information about re-use. It should also encourage better data citation practices than are currently in evidence
- *Publishers* – can secure a pipeline of (better quality) data papers directly to journal submission systems
- *Higher education institutions* – this is an additional opportunity to demonstrate research impact and derive metrics
- *Repositories* – improves their range of services and represents an opportunity to engage researchers to not only comply but also engage with data management and deposition
- ORCID – this is also an opportunity to enhance ORCID's value proposition by increasing its directly useful function for both researchers and HEIs (for instance, ORCID can automatically inform the researcher/institution directly if a data paper is published).

And what's the wider context for publishing data papers? Those who have been keeping an eye on this topic will be well aware that the debate as to whether the 'data paper' and 'data journal' are more than a transitional or transient scholarly communication format and medium is still ongoing (see, for instance, the session at SciDataCon in September: 'Do we need data journals?'). And currently very few researchers are publishing their data – it simply hasn't been integrated into their training, workflows or incentive schemes. Funders, publishers and other organisations such as DataCite have been working hard to raise awareness of the benefits in general terms to 'science', but it's been difficult to make the case to the individual for taking the time to pull together a data paper.

However, in recent years, evidence has been amassing which appears to correlate increased impact of primary research with the discoverability of its underlying data (e.g. Piwowar and Vision's analysis that specifically concentrated on micro-array data) and the research landscape has been adapting accordingly. For instance:

- The Research Data Alliance fosters a number of working groups designed to provide practical and scientifically rigorous support to encourage and enable researchers to share their data (e.g. Data Citation WG Recommendations, WDS-RDA Publishing Data Workflows WG and RDA-CODATA Summer Schools in Data Science WG)

- Thomson Reuters has been developing its Data Citation Index with a view to building the analytics and services it anticipates will be needed for future research assessment and evaluation

- In June 2016, Earth System Science Data became the first data journal to achieve an impact factor. At 8.286, it already ranks 2nd in Meteorology & Atmospheric Sciences and 3rd in Geosciences, Multidisciplinary. This is a significant event in data publishing communities as it has implications for perceived – and measurable – value, publisher interest and potential revenue streams (as data paper publishing itself starts to gain traction via Article Publication Charges).

Finally, the UK Concordat on Open Research Data was published on 28 July 2016 with a foreword by Jo Johnson, Minister of State for Universities, Science, Research and Innovation. Although not an official state document as such, it has been drawn up by a wide range of stakeholders and it makes strong representations about the significance of open data by way of its Ten Principles. Principle Five, for instance, includes:

> *"Production of open research data should be acknowledged formally as a legitimate output of the research process and should be recognised as such by employers, research funders and others in contributing to an individual's professional profile in relation to promotion, research assessment and research funding decisions. Such formal recognition should be accompanied by the development and use of responsible metrics that allow the collection and tracking of data use and impact. In general, data citations should be accorded appropriate importance in the scholarly record relative to citations of other research objects, such as publications."*

As these initiatives and policy influences further permeate the research community ecosystem, it does feel as though some real transformations will begin to take effect. It remains the case, however, that both social and technical drivers and barriers need to be understood and addressed in order for the majority of researchers to take the view that sharing their open data is – usually – the right thing to do.

To that end, we'd love to hear from anyone who would like more information about our app or is keen to work with us – so do get in touch!

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy *if you have any concerns on posting a comment below.*

About the author

After completing a DPhil in English Literature, **Fiona Murphy** *held a range of scholarly publishing roles with Oxford University Press, Bloomsbury Academic and Wiley. At Wiley, she specialised in emerging scholarly communications with particular emphasis on open science and open data. She is a past and current member of research projects including PREPARDE (Peer Review of Research Data in the Earth Sciences), Data2Paper (a cloud-based app for automating the data article submission process) and the Belmont Forum (a multi-national, multi-agency global environmental change project – in association with the IEA). She is Co-Chair of the WDS-RDA Publishing Data Workflows Working Group, and on the Organising Committee for the Force11 Scholarly Commons Working Group. An independent publishing consultant advising institutions, learned societies and commercial publishing companies, Fiona is an Associate Fellow at the Institute for Environmental Analytics (University of Reading) and has written and presented widely on data publishing, open data and open science.*