
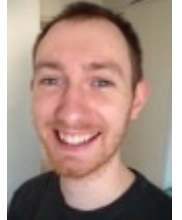


# On the Harvard Dataverse Network Project – an open-source tool for data sharing

 [blogs.lse.ac.uk/impactofsocialsciences/2013/12/31/on-the-harvard-dataverse-network-project/](http://blogs.lse.ac.uk/impactofsocialsciences/2013/12/31/on-the-harvard-dataverse-network-project/)

12/31/2013

*The Harvard Dataverse Network is an open-source platform that facilitates data sharing. [Samuel Moore](#) outlines how this customisable initiative might be adopted by journals, disciplines and individuals.*



I am a huge fan of [grass-roots approaches](#) to scholarly openness. Successful community-led initiatives tend to speak directly to that community's need and can grow by attracting interest from members on the fringes (just look at the success of the arXiv, for example). But these kinds of projects tend to be smaller scale and can be difficult to sustain, especially without any institutional backing or technical support.

This is why the [Harvard Dataverse Network](#) is so great: it facilitates research data sharing through a sustainable, scalable, open-source platform maintained by the Institute for Quantitative Social Sciences at Harvard. This means it is sustainable through institutional backing, but also empowers individual communities to play a part in managing their own research data, especially when coupled with longer-term preservation initiatives.



A Web Application for Publishing, Citing,  
Analyzing and Preserving Research Data

In essence, a Dataverse is simply a data repository, but one that is both free to use and fully customisable according to a community's need. In the project's own words:

*A [Dataverse](#) is a container for research data studies, customized and managed by its owner. A [study](#) is a container for a research data set. It includes cataloging information, data files and complementary files.*

(<http://thedata.harvard.edu/dvn/>)

There are a number of ways in which the Dataverse Network can be used to enable Open Data.

## Journals

A Dataverse can be a great way of incentivising data deposition among journal authors, especially when alongside journal policies of mandating Open Data for all published articles. Here, a journal's editor or editorial team would maintain the Dataverse itself, including its look and feel, which would instil confidence in authors that the data is in trusted hands. In fact, for journals housed on Open Journal Systems, there will soon be a plugin launched that directly links the article submission form with the journal's Dataverse. And so, from an author's perspective, the deposition of data will be as seamless as submitting a supporting information file. This [presentation](#) [pdf] goes into the plugin in more detail (and provides more info on the Dataverse project itself).

## Disciplines

There are some disciplines that simply do not have their own subject-specific repository and so a Dataverse would be great for formalising and incentivising Open Data here. In many communities, datasets are uploaded to general repositories (Figshare, for example) that may not be tailored to their needs. Although this isn't a problem – it's great that general repositories exist – a discipline-maintained repository would automatically confer a level of reputation sufficient to encourage others to use it. What's more, different communities have different preservation/metadata needs that general repositories might not be able to offer, and so the Dataverse could be tailored exactly to that community's need. However, again, it should be noted that this should be done in consultation with data curation experts so best practices are followed.

## Individuals

Interestingly, individuals can have their own Dataverses for housing all their shared research data. This could be a great way of allowing researchers to showcase their openly available datasets (and perhaps research articles too) in self-contained collections. The Dataverse could be linked to directly from a CV or institutional homepage, offering a kind of advertisement for how open a scholar one is. Furthermore, users can search across all Dataverses for specific keywords, subject areas, and so on, so there is no danger of being siloed off from the broader community.

So the Dataverse Network is a fantastic project for placing the future of Open Data in the hands of researchers and it would be great to see it adopted by scholarly communities throughout the world.

*This is an amended version of a blog post originally posted to the [Open Science Working Group blog](#), part of the Open Knowledge Foundation, edited in response to some helpful feedback from the comments sections. Samuel Moore is a 2013/4 Panton Fellow advocating for Open Data across academia, but particularly in the humanities and social sciences. He is currently putting together a multi-author volume on issues in Open Data, due for publication in summer 2014.*

*Note: This article gives the views of the author, and not the position of the Impact of Social Science 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

**Samuel Moore** is a PhD student at King's College London researching Open Access publishing in the humanities and a Panton Fellow in Open Data at the Open Knowledge Foundation. He is also a managing editor at the researcher-led publisher [Ubiquity Press](#). He can be found on Twitter [@samoore\\_](#)

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.