Data in the humanities and variations on open data

Nickoal Eichmann-Kalwara, Mississippi State University

Key Points

- **Options for open data**
  As a researcher, you don't have to be completely closed or completely open - you can pick and choose what you want to make available.
- **Different types of data**
  There are many different types of data, especially within history and the humanities.
- **Alternative data sharing**
  There are other methods of sharing your research other than in a journal.

About Nickoal

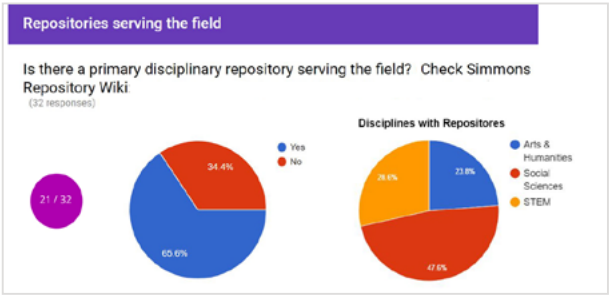
I am a research librarian but I'm focused in history. I use figshare and encourage people to use figshare because of all the different features that are involved in it. It seems to be one of the most popular and easy to use tools for repositories, especially if people don't want to use their institutional repository or don't have one.

I provide workshops on Open Access and that's one of my main roles at our university: to talk about scholarly communication issues.

" I like the control of figshare - that I can edit versions and link to collaborators. I can't do that in my institutional repository. Also, it's pretty and easy to use and the metrics are there! "

That's another reason why figshare will come up a lot: how do you be an open scholar? What are the tools you can use? figshare is great to use for embargoes and control of versions.

I like the control of figshare - that I can edit versions and link to collaborators. I can't do that in my institutional repository. Also, it's pretty and easy to use and the metrics are there!



Investigating Open Access Attitudes by Discipline at Mississippi State University

What is data within history and the humanities?

As a history librarian, people don't necessarily think about what they're using as data. Texts are data and our analyses of those are data. I think that they're just different outputs and because of copyright issues and because of different ideas of what constitutes data - that's why there's a reluctance. But when you look towards more quantitative humanities - computational humanities, digital humanities you'll see more data involved in that. Building a corpus of text and data visualisations becomes the data itself or at least part of it. For me, it's more along that area where an Excel sheet, a very long, structured spreadsheet, becomes one's data. For instance, text and data mining could be built into a data set.

Within history as a subject, data could be a set of oral histories, primary documents, or just spreadsheets of information with coded documents within that. For example, if I did a topic modeling - a computational method to look for co-occurrences of words within a corpus, and I have 1,000 texts that are all machine-readable, I can run them through a programme that looks for co-occurrences of words in order to develop topics. Then, I use the topics to decipher themes or even discourses - something that is humanistic, the computational part that leads into a humanistic enquiry. So, the output that you put into a topic-modelling programme could be data - the set of words, not just the text themselves that you've ingested.

" I think that, at least here, there is a perception that you're either completely closed or you're completely open, but you can pick and choose what you want to make available, especially if you're trying to make an impact. "

The biggest challenges for researchers

Challenges for researchers are a willingness to share research and making it available, at least in the humanities probably because it's such an isolated, intellectual activity - there's not as much collaboration as you see in other disciplines.

Tips for early career researchers

1. I would say it's not all or nothing

I think that, at least here, there is a perception that you're either completely closed or you're completely open, but you can pick and choose what you want to make available, especially if you're trying to make an impact.

2. There are other forms of impact, like Altmetrics, that we can track aside from citation counts.

Because citations are so much lower, using Altmetrics - which I'm so glad to see is integrated into figshare - can tell a broader story by sharing your work early on in different phases, in different formats, and in different venues.

3. It doesn't have to be a journal article, it can be a repository.

It can be a set of slides you put on figshare to demonstrate impact. There are some of us in disciplines, at least in library science, where we publish, but we also have other forms of scholarly output. Our teaching is weighted a lot more in our promotion and tenure and it's good to be able to show the impact of our teaching that's not in a traditional format through a course.

See Nickoal's research on figshare!