

New affordances through sustaining digital humanities data

Authors

Damon Strange, Ex Nihilo Consulting, <https://orcid.org/0000-0002-5851-718X>

Dr Megan Gooch, University of Oxford, <https://orcid.org/0000-0002-3190-0509>

Category of Proposal

Short Paper, Word Count - 852

In this practice-based paper we explore how a project to sustain digital humanities projects through the provision of shared infrastructure, has opened the possibility for new uses and reuses of data.

At the University of Oxford, as at many institutions, the issue of sustaining access to digital humanities projects and their data has been of great concern to researchers and digital humanities practitioners.¹ In 2013-14, a team of researchers, librarians and IT staff investigated the scale and nature of the digital humanities data projects at the University which led to the development of the [Sustainable Digital Scholarship](#) (SDS) service in 2021.²

The problem was a common one; research projects beyond their funding period lacked technical support for their digital interfaces, and these projects had often become a community resource within or between humanities disciplines. In addition, the growth of university data repositories, including the Oxford University Research Archive (ORA), did not meet the researcher needs of allowing data to be accessed at a record (rather than file) level, updated in the light of new data or interpretations, or to be worked on in a collaborative manner. SDS, which is based on software provided by Figshare, allows researchers the flexibility they need, whilst sustaining their data until it is ready to be fully archived in ORA. SDS facilitates data to be FAIR and open by default and allows researchers to customise access and licensing where they need to.

The SDS corpus now contains 32 research projects, over 60,000 records, and the data stored within the system has (since its launch in 2021) received over 5.9 million views and nearly 1 million downloads. Projects include corpora of Anglo-Saxon grains, grave goods and saints' names, Papua New Guinean language data and oral histories of activism in 1968, ancient landscape data and Egyptian papyri.³ Some of these data are from websites that had ceased to function, and their

¹ Taylor, Rebecca, Johanna Walker, Simon Hettrick, Philippa Broadbent, and David De Roure. 'Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC'. Software Sustainability Institute, July 2022.

<https://www.ukri.org/wp-content/uploads/2022/10/AHRC-011122-SSIRReport-ShapingDataAndSoftwarePolicyInTheArtsAndHumanities.pdf>.

² McKnight, J, Jonathan Prag, and Christine Madsen. 'DHARMa (Digital Humanities Archives for Research Materials)', November 2014.

³ 'Featured Projects'. Accessed 7 February 2024. <https://www.sds.ox.ac.uk/featured-projects>.

DH Benelux 2024 – Short Paper Abstract

publication on SDS makes research data available again for human researchers and for computational access such as the Common Crawl.⁴ For other researchers, SDS is a ‘back up plan’ for a more complex website or infrastructure to ensure their underlying data are sustainable.

The impact of this service is only just beginning to emerge, and the benefits could reach beyond simple research data management functions. The migration of projects to SDS has reduced the technical debt and at-risk digital infrastructures within the University at a time when out of date or legacy websites pose a serious threat to institutions, given recent cyber-attacks on higher education and library institutions.⁵ Every project migrated to SDS is one less bespoke system to maintain, security patch, update, and fund for a researcher and their department. The collation of projects on one platform also offers unprecedented opportunities for cross-searching data at Oxford, which could spark new research project ideas and outputs through use of collections as data.⁶ The combination and compilation of data within SDS helps us breakdown traditional data silos, allowing for new searches, new methods and new understandings of a range of research questions and approaches. The benefits for digital humanities research are not showy or immediate, but this work addresses a long-known need to ‘shore up’ digital humanities data so that it is available to the researchers who created it throughout their careers, as well as to researchers of the future.

A unified platform also opens possibilities for further computational access to research data, where researchers are able and have chosen to make their data open. Data in SDS as well as in other research data repositories, library and museum systems across the University are well-structured data and metadata and discussions are underway as to how best to make use of these rich resources in an age of AI hype and angst. There are huge possibilities for bringing together high-quality data from diverse disciplinary areas, but there are also fears from researchers, librarians, and others that as a university, we understand the value of our data. They are concerned that we do not undervalue data nor that we ignore the potential risks in building or using large language or multi-model models with our data. Our experience of setting up and managing the SDS service has also shown us that there are tempting promises of automation in the datasphere, but that a lot of human hours are needed to process, clean and preserve these data. And these hours are often undertaken by technician roles such as IT services, library, or professional services staff.

⁴ ‘Common Crawl - Open Repository of Web Crawl Data’. Accessed 7 February 2024. <https://commoncrawl.org/>.

⁵ Learning Lessons From The Cyber-Attack. Accessed 2 May 2024. <https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf>

⁶ Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke, and Stewart Varner. ‘Final Report --- Always Already Computational: Collections as Data’, 22 May 2019. <https://doi.org/10.5281/zenodo.3152935>.