



Equality, findability, sustainability: the challenges and rewards of open digital humanities data

Damon Strange, Megan Gooch & Alwyn Collinson

To cite this article: Damon Strange, Megan Gooch & Alwyn Collinson (2023) Equality, findability, sustainability: the challenges and rewards of open digital humanities data, International Journal of Performance Arts and Digital Media, 19:3, 348-368, DOI: [10.1080/14794713.2023.2206286](https://doi.org/10.1080/14794713.2023.2206286)

To link to this article: <https://doi.org/10.1080/14794713.2023.2206286>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 23 May 2023.



[Submit your article to this journal](#)



Article views: 635



[View related articles](#)



[View Crossmark data](#)



[Citing articles: 2 View citing articles](#)



Equality, findability, sustainability: the challenges and rewards of open digital humanities data

Damon Strange , Megan Gooch  and Alwyn Collinson 

Humanities Division, University of Oxford, Oxford, UK

ABSTRACT

The need to sustain data and outputs from research projects well beyond their initial grant-funded period is commonplace within academia, particularly so within the Humanities. We often find that these semi-active or warm data collections have value and require care beyond the life of the original hosting, system, or platform they were conceived upon. With the increasing demand to make data more open, and research funding bodies increasing requirements and time periods for which research data must remain available, institutions need to be ready to offer researchers the tools and platforms to comply. Compliance is one lens to view this particular challenge through, the other is that institutions should also be motivated by celebrating the research they develop, and the continued scholarly benefits and impact gained through being hosted and shared for as long as possible. This paper offers a practical insight into the methods being employed at the University of Oxford to support digital humanities scholars (and others) safeguard their digital legacy for future generations.

ARTICLE HISTORY

Received 20 October 2022

Accepted 13 April 2023

KEYWORDS

Digital sustainability;
repository; research data
management

Introduction

For digital humanists and other scholars, the sustainability of their research data is both a pressing and growing problem with no easy or apparent solution. In this paper, we chart the development of a data repository service at the University of Oxford, which arose in response to researcher demand for a system that could meet the needs of open and sustainable digital data, whilst retaining the flexibility of the (often old) systems they had used to create their research data. We describe the concept of warm data, a particularly tricky category of research data that is neither the hot data of day-to-day use in the heat of a research project nor the cold data deposited in an institutional repository with little expectation that it will change or that it will need to be accessed in anything other than at a dataset level.

CONTACT Damon Strange  damon.strange@humanities.ox.ac.uk  Radcliffe Humanities, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG, UK

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Digital humanities are often, but not exclusively, where this warm data can be found, and often exists within datasets or databases at risk of going offline or being rapidly made inaccessible by the swift development of new software, file formats and web standards.

This is not a problem unique to digital humanities; as over the last 45 years, wider society has become increasingly digitised and digitalised.¹ And whilst anxieties over the permanence of our digitised world have bled between our professional and personal lives, and, for example, the loss of our personal photo archives and data have been growing concerns, there is still no consensus on how to solve the mounting problem of digital sustainability except through the use of major commercial tech companies.²

Despite the challenges of research data management, the growth of researchers using digital tools and methods to collect, analyse, interrogate, and understand humanistic problems has benefitted the study of humanities. However, like in so many fields, the Covid-19 pandemic created a sudden and overwhelming need for digital solutions to ways of working in the higher education sector.³ It also made us wonder what had been preventing us from this digital update before the pandemic hit us in spring 2020. The pandemic helped highlight the need for improved support in the space of research data storage for digital humanities projects.

Such turmoil and rapid transformation can have the effect of making previous efforts seem slow and outdated. Digital humanities projects, and the problems of their preservation have been around for decades, and concern over the sustainability of the research data and outputs has been growing at the same time.⁴ As many of the pioneering digital humanities researchers now reach retirement, they no longer have access to funding or institutional support for their sometimes decades-old projects. In some universities, libraries have been the institutions to offer homes for completed projects or their data, in others, data is stored by departments or IT services.⁵ Despite these best efforts, data horror stories abound of outdated websites, entire databases only accessible on outdated software or on a single laptop, or researchers spending time on a never-ending search for more funding to maintain their life's work.

There is no 'typical' digital humanities project, they vary in terms of content, subject, methodological and theoretical approach, and yet many have shared features comprising images of some form of collection accompanied by rich and linked metadata.⁶ We have seen in-depth studies of texts to form a critical scholarly or genetic edition and many of these text-based projects feature high-quality images of manuscripts, letters, books or legal documents accompanied by encoding such as TEI, as well as diplomatic and normalised texts. Examples of these at Oxford include the Samuel Beckett Digital Library, the Newton Project, and the Correspondence of Catherine the Great.⁷ Projects featuring objects include archaeological artefacts, books and manuscripts, decorative arts and artistic works, and feature a range of images with some including 3D imaging and visualisation. Examples we have worked with include the Novum Inventorium Sepulchrale collection of early medieval grave goods and the Oxyrhynchus Papyri project.⁸ Other projects use audio or visual data in addition to transcripts, timestamps and forms of encoding, for example, MEI. We have worked with the Online language documentation for Biak and Dusner language project and the Åhlfeldt and Johnsson 1968: Activism, Networks, Trajectories project.⁹ Often these projects have communities of users and contributors and have become standard resources in their field or subfield. Sometimes their impact is hard to assess, for example, many people use digital editions

for ease of searching but will reference a print edition as the source, obscuring the work of the team producing the digital version and the critical and scholarly decisions made in transforming a text to a digital format.¹⁰ It is apparent that the community value of a digital humanities resource can be measured through means other than citation. The so-called 'scream test' is the loud response website owners get when a resource goes offline, directed at the lead researcher or the IT team, and is a response that has brought researchers to us as their research data is in danger of loss despite its frequent use.

More recently, discussions at national and international levels have taken place to examine the huge risk of loss of digital research data without investment in technical infrastructures.¹¹ Initiatives driven by the research community such as FAIR and CARE data have shown a drive to make data more open and useful.¹² At the same time, open science or open data criteria driven by governmental or funding body policies have pushed researchers to consider their research data more explicitly. The need for research data management plans, often supported by library functions in universities, is something any researcher in the UK will now be aware of, due to funder mandates, in contrast to the situation 10 years ago when 'technical plans' in humanities projects were optional.¹³ At funder and governmental level too, discussions and projects to secure digital infrastructures are ongoing. In digital humanities in the UK, the Towards a National Collection funding scheme of £18.9 million was launched in 2019, and more recent AHRC calls have targeted finding answers to the research data sustainability problem.¹⁴

It is within this context then, of the growth of digital humanities projects and the risks of loss of research data due to researcher retirement or technological obsolescence combined with the push and pull of data management compliance and the open and FAIR data movements, that institutions are developing a variety of ways to sustain these data, and funders and governments are prioritising wider infrastructures to support this work.

In this paper, we explore the detailed challenges of sustaining data at an institutional level from our work at the University of Oxford. Our first challenge was driven by researcher needs which were not met by existing University systems, so we will explore what kinds of data are used and accessed by digital humanities projects. We ask how a new data repository service to sustain data can itself be sustained financially when the majority of humanities and digital humanities research projects are short-term funded but produce data outputs with long-term access needs. Our paper is grounded in the work that we have done as a team since 2019, and also in the methodologies of the professional services teams we currently work in or have worked in. To understand a research and researcher issue we used standard project management techniques and tools to develop and launch a new service. In this paper we will explain this work within its methodological frameworks but also explore how and why decisions were made and the impact this had on making research data more sustainable.

Sustainable open access: why is this so hard?

Digital preservation is a common area of study in the field of information management, much of which has been driven by libraries and other institutional archival services. This

has traditionally focused on a well-defined records life cycle, in which digital research objects move through discrete and linear stages of collaboration, publication and archival preservation. Other authors have noted the basic incompatibility of this linear process with digital research 'where records are unlikely to reach a definite inactive point but are instead migrated into new formats following developments in technology',¹⁵ but digital humanities projects pose an additional challenge to this model of digital preservation. They are extremely likely to require publication in a semi-active or warm format, one that permits edits, updates, and file-by-file access, stretching the capabilities of traditional archival solutions. In other words, they require digital sustainability rather than digital preservation in the latter's formal archival sense.

In 2020, the University of Oxford commissioned a Research Data Management review, which was carried out by external consultants from Research Consulting and Charles Beagrie.¹⁶ The resulting report identified three categories of research data, roughly corresponding with this traditional breakdown: active data; semi-active or living data; and archived data (Figure 1). Of these three categories, the University has provisioned active IT solutions for the first and last categories, in the form of file-sharing and archiving platforms available across the institution: a range of digital tools for current research projects ranging from Microsoft Office to high-performance computing, and an institutional repository called the Oxford Research Archive (ORA). Only in the 'semi-active or living data' category was there no technical solution available, and it is into this category that

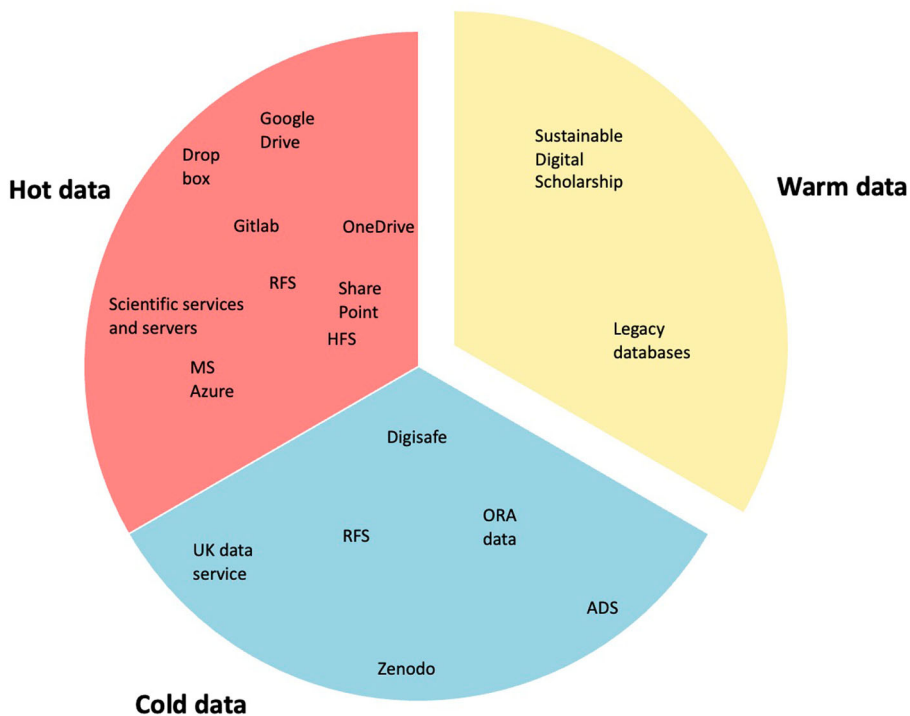


Figure 1. A diagram showing the availability of tools and resources for hot, warm, and cold data at the University of Oxford, as highlighted by the Research Data Management Review, 2020.⁵¹ © University of Oxford.

many digital humanities 'legacy databases' fall. In other words, the digital humanities projects whose precarity was highlighted by the pandemic's digital turn, were precarious because they were difficult to sustain and did not fit into the existing institutional provision of 'in progress' or 'complete' in terms of data management.

The authors of this review noted this semi-active or living data that "there is no widely accepted terminology for this type of data, which is often very long-lived and worked on intermittently by individual scholars or research groups. Our working definition was "used data currently held for possible future re-use or further development",¹⁷ in other words, corpora or databases between periods of grant funding or unfunded. We have come to define this category of digital records as warm data, falling between the hot (pre-publication records undergoing active collaborative review) and cold (completed research preserved in a long-term archival format) ends of the data lifecycle. Whilst the anecdotal and qualitative evidence for a warm data need is strong, we do not have numerical evidence of the scale of the problem beyond Oxford's Research Data Management Review, or from published surveys concerning the precarity of digital humanities data.¹⁸ The heat metaphor is a communicative device, but one that gives immediate understanding to a variety of researchers and one that avoids pitfalls of making value judgements on research data such as calling it a 'legacy' database suggesting obsolescence, or an implied redundancy in 'archived' data. Warm data are characterised by being published digital objects, which are available for open access on the web, but which are also currently, or intended in the future to be:

- Added to, expanding the research collection of which they are a part;
- Updated, with periodic changes to their files or metadata;
- Individually accessed, with each file and its associated metadata being surfaced to the open web;
- Programmatically accessed, supporting the reuse of their research objects elsewhere via APIs.

Perhaps the defining feature of warm data collections so typical of many digital humanities projects is that they are periodically updated, with research outputs of one project enabling future academic research or grant applications. This requirement means that most of these data collections require periods of support which is not covered by short-term research grants. Despite funding bodies' insistence on the long-term open access preservation of the research they support, few are willing or able to finance ongoing costs in perpetuity to safeguard that data.¹⁹

Humanities digital research projects are particularly prone towards falling into this warm data category, as many of them focus on creating a complete corpus of a particular type of digital resource for which new data may be discovered. New data may be uncovered in archaeological excavations, discovered in museum, library or archival collections or created in new publications or fieldwork: for example, digitising every extant cuneiform tablet;²⁰ recording every manifestation of a developing cult of a Christian saint in the Mediterranean in Late Antiquity;²¹ or creating a complete archival record of past and future issues of the *Oxford Gazette* magazine.²² These projects, and many others in the digital humanities, aspire to be comprehensive, reflecting new discoveries; accurate, being updated in line with new translations, reinterpretations, or other recent scholarship;

and accessible, supporting sharing, citation and reproduction of individual records or objects within the collection.

These factors necessitate the use of an actively managed repository (often in the form of an online database) rather than an archival solution. Although the Bodleian Libraries supplies the University of Oxford with a fully featured institutional repository as part of its ORA service, ORA-Data, this is designed to create a funder-compliant and citeable record of completed, published research datasets which are not destined to undergo further development.²³ We encourage all our users to deposit their datasets in ORA-data, and the Sustainable Digital Scholarship (SDS) service allows researchers to extract data and metadata as .csv files for periodic deposit in the institutional repository both for funder compliance and also to allow other forms of use for their data. Traditional digital preservation in institutional repositories creates a single, stable copy of research outputs at the end of each phase of funded research, assuming that this static version reflects the complete, final outcome of the project. But humanities data is as diverse and ever-changing as humanity itself.

In addition to the incompatibility of this snapshot preservation method with some digital humanities research, archival approaches to digital preservation necessarily involve the flattening of complex research collections, compressing files and separating the records representing research outputs from the data structure, links or code which describes their relationship. On a user level, this may manifest as the loss of discovery tools such as search, tags, links or maps, or as a requirement to download entire data collections in order to access the individual records within it. As well as reducing the possibilities of open access to these research outputs, such a process can have an exclusionary effect, making data available only to those with the technical capabilities and computing resources to find, download and reconstruct these collections and the connections therein.

Prioritising access to individual components of research collections is not only an academic concern, but one of societal fairness and intercommunal equality. Digital humanities data collections often consist of objects of cultural, political or religious importance, and providing open access to such material to the general public or publics is a major, if subsidiary, benefit of such research, especially when that research is undertaken with public funding, such as from UKRI. The African Rock Art Image Project at the British Museum, which digitised and disseminated c.24,000 images of rock art from across the continent, prioritised publication formats which would enable users with low-bandwidth internet connections to download and view the resulting data and images. To do otherwise would have made the collections inaccessible to the inhabitants of the very countries where the rock art was physically located.²⁴ To digitise a cultural group's historical artefacts, only to gate the resulting digital objects in inappropriate archival formats, undermines the principles of equal and free 'open access'.

Providing this active repository solution poses significant challenges which are not restricted to the ongoing funding challenge to support file hosting and technical development. The diversity of digital humanities projects is reflected in their differing required file formats, metadata schemas, research structure and user interfaces. In other words, there is a need for a digital sustainability solution, for digital humanities projects, which retains access to the warm data and some of its functionality such as the ability to add to the data collections and the ability to query on a record level. As a result,

this warm digital sustainability is different to the more long-term archival process of digital preservation which is suitable for cold data.

Historically, the hosting of all the associated outputs of these research projects would be a task for individual research projects, their Principal Investigators (PIs) and project teams. Given an individual's association with an institution is finite, whether this be the lead Researcher, research staff or the technical support team, this clearly poses a serious challenge for the sustainability of research outputs that depend on the knowledge of these associated individuals. At the University of Oxford, this challenge is acute due to the scale of digital humanities research undertaken at this research-intensive university and also the early adoption of 'cutting edge' humanities computing which later became known as digital humanities.²⁵

Many of these aforementioned cutting-edge digital humanities outputs, within Oxford and more broadly in the field, have sadly since become condemned to a life as an unreadable file format or media type, archived on internal servers when life on the open web would be more appropriate. Some have even been lost forever due to an insufficient digital sustainability plan in place or a loss of funding. This issue has been at the forefront of recent funder policies and funding initiatives and has been highlighted by a recent report commissioned by the AHRC.²⁶ Identifying the root cause of why projects go offline will often be a result of several factors which are people, process or technology related.

One reoccurring reason why sustaining digital humanities research projects is particularly challenging (in the long-term), within Oxford and no doubt everywhere else too, is the reliance on custom-built databases, platforms, and websites to display and share research outputs. Traditionally, digital humanities researchers have tended to build one-off bespoke databases or repository solutions to provide access to their research data.²⁷ The ERC recommends against this approach, noting that 'in contrast to public data repositories, these are generally not deposition databases, and as long as they depend on a single individual and/or funding source, long-term sustainability is challenging'.²⁸ Nevertheless, in the absence of institutional repositories capable of supporting warm data collections, individual technical solutions have proliferated in the digital humanities as cost effective and easily constructed mini repositories for the data of individual research projects.

To understand how widespread this problem of bespoke online databases was, a review of digital projects was undertaken by the Humanities Division. The Humanities Division is one of four academic divisions at the University of Oxford comprising 14 academic faculties and units. The review, undertaken during the scoping phase of a Digital Humanities Sustainability project which is further detailed below, identified 47 projects which were suitable as 'pilots' for a warm data repository.²⁹ This review was not comprehensive, and can only be taken as reflecting a portion of the digital humanities projects requiring digital sustainability support. They were defined as important research projects that had produced collections of warm data but which were currently outside their grant-funded period, and which faced challenges to their digital sustainability for reasons as diverse as technical obsolescence (hardware or software), lack of continuity in project staff, PIs retiring or leaving the University, or with security risks within their project websites/databases due to changes in web standards and new legislation such as GDPR and the Equalities Act which added additional obligations concerning personal data and digital accessibility.

However, the review of bespoke databases demonstrated the troubling scale of the problem of unsustainable digital humanities projects and this, combined with the need to comply with funder requirements, a desire to make data open and FAIR, and the sudden digital shift of the pandemic highlighting the value of these online (or offline) caches of high-quality research data at risk. The category of bespoke databases and data solutions includes relational database software such as Microsoft Access and Filemaker Pro, as well as open source software such as Omeka. The problem we encounter across proprietary and open source software is sometimes financial and concerns lack of funds for up-to-date software licenses, but is often technical, as even relatively simple open source software solutions such as Omeka can become problematic if they have been developed by technical staff who are no longer funded, and the PI lacks the technical ability to make vital updates to the research data system. A solution was in progress at Oxford in the form of the Digital Humanities Sustainability (DHS) project, which sought a way to utilise the collective knowledge of many projects and collaborators and attempt to find a way to standardise the non-standard.

The challenges of retrofitting sustainability

The journey towards a more standardised approach to offering sustainability to digital humanities scholars at Oxford has been a long, and not a particularly easy one. The foundations were laid by Professor Jonathan Prag (Principal Investigator), assisted by Christine Madsen (Project Manager) and Janet McKnight (Research Associate) when they delivered DHARMa (Digital Humanities Archives for Research Materials), a one-year project (August 2013 to August 2014). The findings within DHARMa provided the catalyst and 'researcher-led' evidence base and a strong recommendation that continuing with historical practices for sustaining digital humanities research needed to be reviewed.³⁰ With these recommendations at its core, the 'Digital Humanities Sustainability (DHS) Project' was approved and commenced in 2017 as an institutionally managed IT project within the University. The project successfully delivered its scoping and analysis and planning phases, run according to a traditional PRINCE2 methodology.³¹ The aim of the DHS project was to find a solution for the sustainability of extant and new digital humanities projects not served by the existing digital preservation options offered within the University. These early project phases increased our understanding of what a future state and infrastructure could look like.

However, these phases of the project were not without their limitations and in early 2020 it was decided that the DHS Project would become a project delivered 'by academics for academics' and the management of the project moved from the professional IT Services to the academic Humanities Division. This transition from a centrally delivered project to one delivered by an academic division was necessitated due to a period of portfolio reprioritisation within IT services and IT budgets. This meant DHS, as a fledgling service was deprioritised in favour of funding existing and established services, or those with an acute need due to legal or other regulatory compliance. This threatened to pause the DHS Project indefinitely, which would have lost the momentum gained from the early stages of the project. The DHS Project was the top IT priority of the Humanities Division to support its researchers and so it was decided that

delaying the delivery of this initiative was not an option. The project governance was therefore moved from IT Services to the Humanities Division to enable it to continue. Another change was a move to an Agile Project Management approach, which seemed wholly complementary.³² Damon Strange, an experienced Agile Project Management consultant created a project approach that placed researchers at the heart of the plans for Delivering the DHS Project.³³ One of the four primary values that the Agile Manifesto describes is the prioritisation of 'individuals and interactions over processes and tools' and this resonated with the approach which the DHS Project followed.³⁴

A rich and extensive network of Oxford-led digital humanities projects and their connected scholars provided the perfect community to gather requirements to deliver a DHS solution, which would go on to serve them and other researchers in the future. As part of the DHS Project, extensive community consultations were held to gather feedback on system and feature requirements, with more than 400 requirements identified in total. These requirements were refined by the DHS Project team, until 109 remained which were then entered a thorough round of MoSCoW Prioritisation, which is a, 'technique for helping to understand and manage priorities', and forces contributors to consider priorities in terms of whether they must, should, could or won't happen.³⁵ Requirements were then broken down by functional categories such as persistent identifiers, metadata creation, metadata enrichment, workflow, API interactions, search, browse, and find, as well as ingest and egress. The bulk of these functional categories were derived from specific examples and consultation with researchers responsible for delivering and managing digital humanities research projects. [Figure 2](#) provides a full overview of these category areas, presented within a Reference Architecture for a future DHS Solution, highlighting primary requirements in scope as part of initial implementation and requirements that may be deemed desirable in the future.

Once we had a detailed understanding of what our requirements were for the DHS Project, we carefully considered the options to self-build or to commission a technical system and solution. A key factor was an ambition that individual researchers, or those within a research project should be able to manage their own research data with minimal intervention from support staff beyond initial migration or on-boarding. The ability to self-manage was considered important for the quality of the research data, the user satisfaction of the researchers, and also as a more sustainable service model that did not rely on a large team of support staff.

After reviewing the cost and risks of a bespoke build for the sustainability solution, the project team sought to understand what existing options could be used to meet the needs of this project. The well-honed requirements catalogue was deployed within a formal Invitation to Tender (ITT) process which sought bids from any system and solution providers, who could potentially present a Software as a Service (SaaS) offer to meet our requirements.

An important element of retaining the 'researcher-led' ethos was ensuring that academic representation on the reviewing panel of any submitted ITT bids. The in-depth digital humanities knowledge, critical eye, and passion to find the right solution were delivered by academic colleagues supporting the review process of the ITT. Once a supplier was appointed, we implemented Principle 7 (Communicate Continuously and

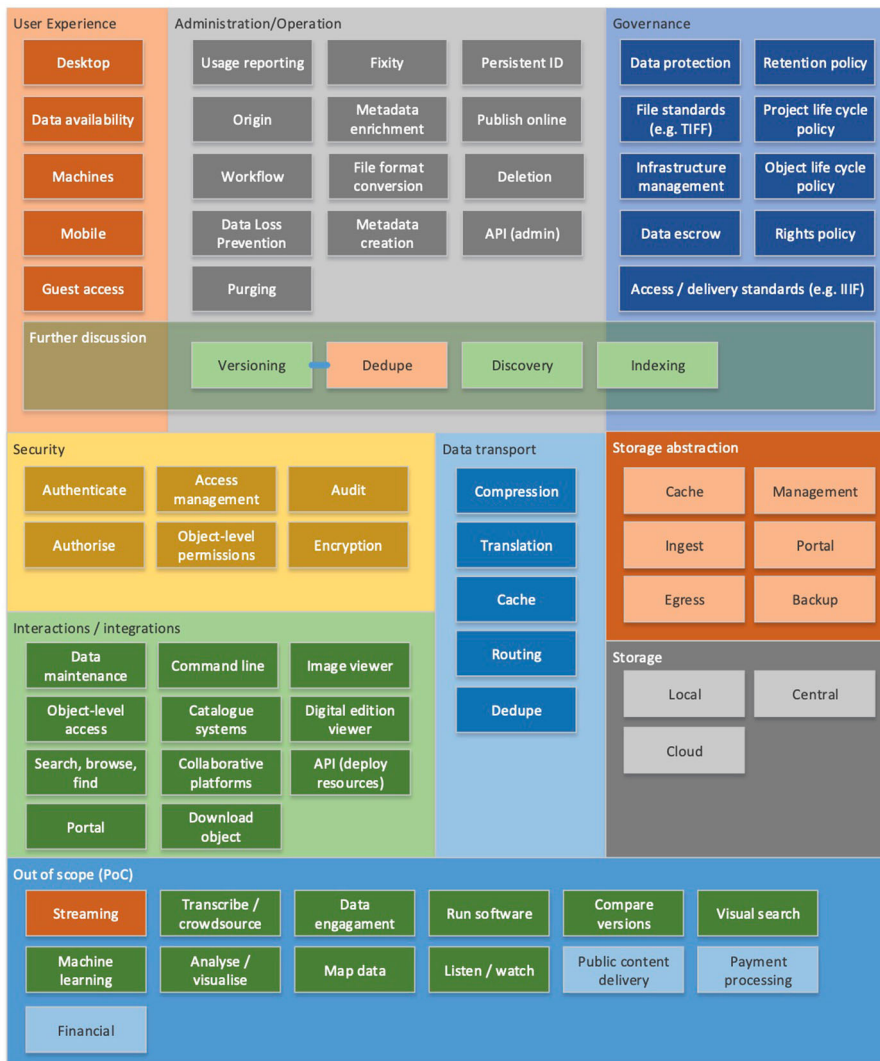


Figure 2. The DHS Reference Architecture which was developed to group features and requirements by key categories © University of Oxford.

Clearly) of the Dynamic System Development Method Project Framework, to allow our academic colleagues to see a solution in practice and understand its strengths and weaknesses with real data as

Modelling and Prototyping make early instances of the solution available for scrutiny. These practices are far more effective than the use of large textual documents, which are sometimes written for reasons other than achieving the business objectives of the project.³⁶

In keeping with our desire to run an Agile project, we opted for this Proof of Concept (PoC) implementation with our preferred bidder, Figshare, to allow us the opportunity to test the platform and showcase its potential to interested early adopter researchers for their digital humanities projects and outputs.³⁷

The PoC was deemed a success by the project board as it delivered the results for its academic users that were intended. This also provided the DHS Project with further feedback and a product roadmap of potential features which digital humanities scholars were interested in seeing developed for the platform in the future. The latter was important as although the Figshare platform allowed us to migrate, store and query data in a sustainable way, there were some areas of functionality that would be lost from many extant, and often pioneering, digital humanities projects. These lost areas of functionality were compiled in a list for future developments. One example was the need for a Text Encoding Initiative (TEI) text editor which was raised by many researchers working on a range of ancient and modern texts. Following the decision to proceed with Figshare, a swift implementation took place in parallel with work in the project team to design service management processes. This allowed us to launch the SDS service in February 2021 as both a support service and a technical platform which can aid digital research scholars at Oxford with existing or new digital humanities projects.³⁸

It is important to pause here to highlight the dual function of the SDS service as both a digital platform for research data as well as a consulting service for researchers' queries about research data. Researchers who contact the service may be referred to colleagues managing other research data repositories within the University or external subject-specific repositories, and this collaboration with colleagues is an essential part of the service. Of course, researchers may also be guided towards the SDS platform run on the Figshare software.³⁹ This combination of service and technology has contributed to the success of the project.

This digital platform is an instance of 'Figshare for Institutions' which ensures digital humanities data are hosted in an 'open by default' manner. This creates inbuilt open access and FAIRness for data hosted on the platform within any research project. Data is 'as open as possible, as closed as necessary' and can be restricted if there are requirements such as embargoes, personal data or copyright restrictions. An early project for the service, Åhlfeldt and Johnsson 1968: Activism, Networks, Trajectories tested our ability as a service and as a platform to deal with restricted data options due to the nature of the recent oral history data including voice recordings of living individuals.⁴⁰

The benefits of using this SaaS approach is that openness is the default option, as is FAIRness, meaning that researchers do not have to design these elements in their research, they can simply use the functionality available to them in Figshare. By contrast, researchers who have sought to cut their own technical path will often face several challenges in building, adding and maintaining their research data in addition to making it open and FAIR. The option to build bespoke systems remains open to researchers at the University of Oxford, as we acknowledge that some research questions will always require specialist software solutions. However, we have found that for many researchers utilising a commercially built and maintained platform will provide all the support they need for their research data.

From the perspective of running and maintaining a service, the cost of implementation and maintenance of a SaaS product such as Figshare, we calculated that it has cost a fraction (c. 10-20%) of how much we estimated developing and building our own infrastructure would cost. This is the case when comparing the Total Cost of Ownership (TCO) over a 10-year initial period for both SaaS and self-built solutions.⁴¹ This SaaS solution is both more cost-effective than developing our own technical platform and has not thus far

led to a restriction of functionality. By contrast, and as you would expect from a paid service, 'Figshare for Institutions' (FFI) offers much more than the free version of Figshare.com. FFI customers are afforded the following as an illustrative but not exhaustive number of features and benefits above the free service; custom URLs and branding for groups implemented on the platform such as badging for Faculties or other academic departments, the ability to define custom metadata fields or schemas for projects, login and authentication handled by institutional Single Sign On, a significant increase of storage (100TB) at the disposal of the institution to distribute across users and projects (compared to 15GB per user for Figshare.com), and the ability to join an active UK and Ireland network of other institutions using the service to seek support and discuss future planned improvements with an engaged supplier. Figshare is of course a proprietary software, not an open source one, but the sustainability of all software is acknowledged as a complex challenge facing researchers and businesses in the modern world.⁴² Many of the digital humanities projects we work with are also on some kind of proprietary software, either database software, or hosted on commercial web platforms. There are indeed risks to both not building our own system and in using open source code. However, we see the use of Figshare as comparable to other major infrastructural software bought and used within many universities and cultural institutions such as museum collections management software, integrated library systems, digital asset management systems, content management systems and digital preservation platforms. As with all software choices, whether they are bespoke, open source or commercial, we have a well-developed 'exit strategy' on our risk register and have already scoped a range of ways to egress data and metadata from all the research projects on the platform should this be required.

As we have mentioned, the Sustainable Digital Scholarship service is part of a range of research data management solutions and repositories within Oxford,⁴³ but so far it is the only one that serves the needs of researchers with warm data projects to deliver open and reliable access to these research data. Digital sustainability is a large and growing issue for many institutions, with challenges that are not just technical but also relate to the cultural adoption of research data management by researchers.⁴⁴ Our project and research have shown that despite a wide variety of research questions, subjects, methodologies and approaches, the introduction of a shared and common infrastructure for warm data can deliver digital sustainability for researchers and their publics.

A practical approach to making research sustainable by design

It is an easy matter to talk of building sustainability but a harder task to implement it. The first and easiest area to address is how the SDS service offers the embedding of sustainability by design through new research projects built on our platform. In these cases, the SDS service is able to offer consultancy and advice and share the burden of scoping, designing and delivering infrastructure for many new research projects. We acknowledge that although every research project will be different, there are often common approaches in structuring data and metadata, and shared expectations from researchers as to how they would like their research data curated and presented, given they have been informed and inspired by extant digital humanities projects. The SDS team, having seen many projects, and their similarities, can therefore consult with researchers

and follow a consistent process of information gathering and scoping. A typical SDS scoping session will establish the key details of a project's planned data collection, including: file types used, estimated storage size required; a guide to custom metadata needs for the project; the nature of the data, and whether there are any limitations such as GDPR, copyright, ethical and anonymization or embargo considerations; the hierarchical and navigational requirements for their data; and any requirements for advanced features, such as a desire to visualise data or link out to other systems.

Once an outline plan has been agreed between the research project and the SDS team, we agree a timeline of activities and acquire data items from the project once they are ready to be handed over from the active or hot phase of research. The raw data and any structured metadata are then prepared for migration onto the SDS platform. For existing projects, there can often be a lengthy phase of data cleansing or data wrangling at this stage to ensure the data and metadata are in a format which can be migrated. This work may be undertaken by the researcher or a member of their team, or by a data specialist within SDS. In the latter case, careful discussions and documentation ensure the researcher is satisfied that their data are not transformed in a way which has implications for their research project, and the process of transformation is recorded so the researcher understands the decisions and processes which affect their data. As a team, we are working to develop, document and share our processes for data wrangling, so that researchers can undertake some routine data cleansing themselves and become less time consuming for the project team. In the case of new projects, the research team mostly do not need this data cleansing phase as they are collecting data in a manner which will work directly within SDS.

The practice of migration involves deploying Figshare's Batch Migration Tool,⁴⁵ or utilising an open-source tool developed by David Banks, at the University of Sussex.⁴⁶ The ability to use these two tools has made it feasible for the SDS team to move research data onto the platform in bulk, creating hundreds of records in minutes with the simple upload of a spreadsheet of metadata. The manual creation and curation of records have often been the alternative for many other digital humanities systems our researchers have used, and the speed at which SDS can import and visualise data online is an advantage for our users.

The second category of research projects SDS offers support to is pre-existing collections of data, some of which may have been created years or even decades ago. The data, as discussed above, are often already hosted on another platform or infrastructure, and researchers are looking to retrofit sustainability. Researchers in this category may be unaffected by recent funder mandates to ensure their data are open and available for a set period, but are driven by the need to sustain the data they are still working with in their research, or that are being actively used by other researchers in their field. The process of migration is essentially the same as for a new project, with an initial scoping meeting and agreement, data structuring or cleaning and upload. However, due to extant data formats, there is potential for more friction at each of these stages as researchers have often invested much time and money in their own infrastructure solutions, and have become used to the bespoke functionality of these systems to aid their research. For some researchers, the loss of functionality or the compromises required to clean their data are not compatible with their research agendas and their projects are simply too complex or unsuitable for migration to SDS. This is to be expected given the range of research and

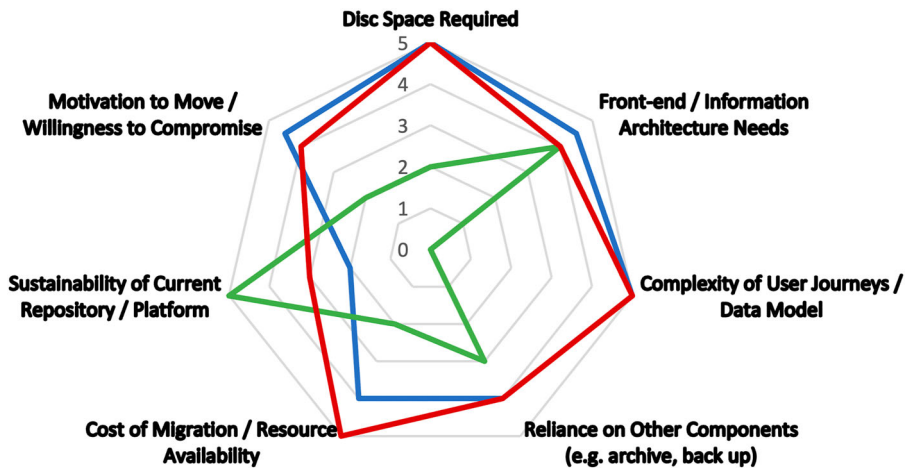


Figure 3. Suitability Assessment Spider Diagram for 3 (anonymised) early adopter projects. Each assessment area for the projects being analysed are apportioned a score of 0-to-5 with 0 representing a poor fit response and 5 being an excellent fit. © University of Oxford.

researchers, and we knew from the start of the project that SDS will not work for everyone as no one system ever will.

To assess the suitability of a project for the SDS platform, the team developed a methodology for assessing research projects. This ‘Suitability Factors Assessment’ methodology was developed with the assistance of employing a professional Business Analyst from IT Services and enabled us to understand the key areas to consider when assessing research projects for suitability (see Figure 3 above for a high-level view of three example projects which were assessed). Projects were assessed on seven facets which were each scored on a scale of 1-5 on areas such as storage space, presentation and front-end requirements, complexity of user journeys, the need to integrate with other systems, estimated migration cost, urgency, and the ability of the researcher to compromise and still align with their research agenda. Figure 4 below captures the detailed Assessment Criteria Guide which was created to offer guidance when analysing the suitability of projects for hosting on the SDS platform. This table provides further information around the circumstances or factors that would elicit a certain score, which was then used as a relative rather than an absolute guide to assess the suitability of projects for migration to the SDS service. This is an attempt to add some more context to the scoring system with a little bit of narrative; for those carrying out the assessments and those reading the outputs. As a professional support team, the SDS team makes no value judgement on the quality of the research or research data in this assessment – it is assumed all projects that come to us are of a high academic standard.

Figshare is not a bespoke system for digital humanities projects, nor has it been built and tailored for a specific research use case. This means that flexibility and expectations management have always been important when discussing the prospect of hosting research data from a new researcher. Open and honest discussions with researchers about what the system offers now, what is on the roadmap for the future and what is unlikely to be on the roadmap, is a key component of ensuring the SDS service offers its



5 = excellent fit
0 = poor fit

ASSESSMENT CRITERIA GUIDE

	0	1	2	3	4	5
Disc space required / available	20TB+	10TB - 20TB	5TB - 10TB	1TB - 5TB	0.25TB - 1TB	Less than 250GB
Front-end / Information Architecture needs	Visualisation-rich web development required to realise existing functionality and / or extensive use of API for data transfer and projection. Potentially requiring custom development work	Some basic visualisation requirements needed to match existing functionality and / or extensive use of API for data transfer and projection. However, delivery should still be possible internally	Likely some basic visualisation requirements needed to match existing functionality and / or extensive use of API for data transfer and projection. However, delivery should still be possible internally	Potential for some basic use of API, or extensive use of linking to data (saved searches, references items, etc.). With a separate front-end web presence required	No need to utilise API for bulk data visualisation / projection onto website. However, a separate web-front end required for the DH Project (with links to Figshare data)	No need to utilise API or limited need to build out a separate web-front end project presence
Complexity of user journeys / data model	Highly complex relational database structure of user journeys, which would be fairly challenging to recreate with custom metadata / Figshare method of structuring data	Significant amount of custom work to map metadata, etc. Likely involving a good deal of support needed from Figshare consultant	Light touch custom work for metadata mapping, with a small amount of support needed from Figshare to realise this	Potential for some general support required from Figshare or internal workshoping to consider how we re-create user journeys to an acceptable level using Figshare as the data repository	Generally all existing user-journeys for the project in question should be straightforward for recreation within Figshare / linked web front-end	Limited number of items and/or the number of relationships they have with each other, and how a user would typically wish to navigate through a collection. I.e. individual records are fairly independent from each other
Reliance on other components (e.g. archive, back up)	Several other complex technical components required to deliver existing functionality / meet being several concerns of a successful outcome (without technical sacrifice)	Several other complex technical components required to deliver existing functionality, to the point that may suggest storing the data within Figshare is adding limited value (other than pure sustainability)	A number of various technological components are required to meet requirements. However, achievable and will operate well following some investment in consultancy support to realise	2-3 other technical components are integrated with the data repository, however, relatively straightforward to achieve via API e.g. periodic additional backing-up to personal AWS S3 bucket.	Another 1 relatively simple technical component is required, however, limited specialist expertise is required to achieve this. I.e. Setting up a Web-page that provides information over and above what is available on Figshare	All data and project interactions are contained and limited to Figshare
Cost of migration / resource availability	Migrating the project would not represent VFM; high consultancy cost, significant amount of additional storage purchase required and / or; No research-based project resource available to support at all	Only PI / Curator of project available or with requisite knowledge about the project, who generally struggles to dedicate time to support other than on an ad-hoc basis. Extensive amount of out-side support required from DHS Project / Figshare consultancy	Very limited amount of resource available from the project. High drain on DHS Project time to support and/or a need for consultancy support from Figshare	Limited amount of resource available from the project, however, some ad-hoc support may be possible. High drain on DHS Project time to support and/or a need for some basic consultancy support from Figshare	Research project can dedicate some resource to supporting the migration. DHS Project needs to dedicate a good deal of time to support (however, no additional consultancy required)	Primarily DIY / Self-sufficient Researcher in their ability to migrate people / expertise to support the project. Very light touch support from the DHS Project. Cost of migration very low (possibly dictated by simplicity of task in hand)
Sustainability of current repository / platform (Technology & People)	Storage arrangements for research data are guaranteed to provide sustainability and/or; Wealth of individuals with project-specific knowledge readily available	Storage arrangements for research data are fairly assured to provide sustainability and/or; Some individuals with project-specific knowledge could be available	Arrangements for the sustainability of data are acceptable, however, could be significantly improved by migrating to Figshare	Minor risk of data loss due to current storage arrangements in place and it would be recommended for the research project to consider a safer home for data to assure longer term sustainability	Significant risk of data loss due to current storage arrangements in place and/or; Only PI / Curator of project available or with requisite knowledge about the project	Imminent risk of data loss due to current storage arrangements in place and/or; Limited human resource with project-specific knowledge
Motivation to move / Willingness to compromise	Strong desire to maintain a unique identity of project, considering that this has to be achieved by being contained within its own database / project environment. Completely unwilling to compromise or deviate from the original project since conceptualisation (primarily due to not believing the DHS Project could replicate like-for-like)	Generally not willing to compromise on the majority of functionality and have a fairly strong attachment to the 'current way of doing things'	Some challenges incurred when asked to consider alternative approaches to realise similar functionality within the new platform	Understands the angle the DHS Project is coming from (securing '80/20' projects etc.), however, struggles to compromise and only really getting involved due to having limited options (and their motivation to move is almost forced)	Generally a positive attitude towards working with the DHS Project exists. However, will potentially have a few 'deal breaker' features / functions that will require some persuasion or additional discussion to ensure a compromise is reached	PI / Researcher fully believes in the vision of the DHS Project and willing to work with the project team / Figshare to consider alternative options to achieve similar (or the same) functional results. Accepting the limitations of Figshare relative to a custom built solution; however, will embrace this opportunity to potentially 'make their DH Project better'

Figure 4. Assessment criteria guide developed to assess suitability of projects for migration to the SDS platform © University of Oxford.

support to the best-fit projects and avoids overpromising to users. An early principle we adopted was the 80/20 rule, meaning that we aim to offer a unified and managed infrastructure for 80% of digital humanities projects, leaving only 20% requiring more detailed care and attention or bespoke solutions to ensure the long-term sustainability of their research data.

The SDS service has engaged with over 120 research projects regarding the potential of offering future support and hosting to date (January 2023). Of these, 17 projects were deemed as not compatible for hosting on the service, which suggests we are servicing more projects than the estimated 20% of projects not be suitable for the platform. For most of these projects who elected not to use the service, the lack of complicated querying and presentation functionality within the Figshare platform was the primary driver for not working with SDS.

Some use-cases will always extend beyond what can be offered by a SaaS product like Figshare, however, we have been able to offer a great deal of functionality within the software. We have utilised curated keyword tagging on individual records and at an item level, which opens the possibility of custom searching not currently available in Figshare's core functionality. The forthcoming Figshare for Institutions Roadmap developments will see further enhancements in the ability to query data held on the platform.⁴⁷ Both the ability to employ 'full-text searching' and search based on 'custom facets' defined by the institution will ensure the discoverability of research data is improved. These planned improvements relating to search, and discovery are of keen interest to many of the digital humanities researchers with whom we have already engaged, but who had previously found Figshare to be unsuitable for their research projects due to the lack of these features.

The practice of conducting a Sustainability Factors Assessment was not one which continued into the general service management processes for SDS, but a great deal was learned from this detailed analysis. As a team, we ensure that when a project is scoped for a potential migration, we understand where challenges could arise. During consultations with researchers, we ensure transparency is maintained and expectations are managed accordingly. This is particularly relevant when we are working with a well-established digital humanities project which may have been at the cutting edge of infrastructure design when it was built many years ago. However, the elapse of time, and the end of a funding period has often led to researchers reporting a degradation of their system as they can no longer afford technical expertise to maintain or update their system as web standards or legislation changes. In these cases, where a system may have once been, or still is, feature-rich, the replication in full of all these complex features is not possible within the SDS platform. As a result, researchers must make a difficult decision as to whether their research project will gain sustainability but lose some functionality by migration to SDS. Unfortunately, for those researchers who are unable to compromise on functionality, the future of their research data is still at risk, and they must continue to host their project and find alternative funding to maintain what is often an important community resource. Whilst SDS offers a level of sustainability for many digital humanities researchers, there are still broader discussions at an institutional and national level as to what the humanities data infrastructure needs of the future are.

The popularity of the SDS service amongst digital humanities researchers at the University of Oxford continues to grow as new projects are funded and extant projects contact us for advice. As a service we advertise a diverse range of projects supported on our website,⁴⁸ to demonstrate to researchers what the SDS platform can achieve. One of these projects that highlights the challenges of sustainable open access is the 'Novum Inventorium Sepulchrale – Kentish and Anglo-Saxon Grave Goods in the Sonia Hawkes Archive' project.⁴⁹ Originally this project had a fully functioning database with searchable

web front-end, showcasing many images and rich descriptive metadata, relating to c. 1,000 graves and the objects found within them. However, since the project went offline indefinitely in 2018, all that remained was access to 2 metadata spreadsheets on a single project webpage. In 2021, following a swift engagement and scoping process with Professor Helena Hamerow at the School of Archaeology comprising the usual cleaning and mapping of metadata steps, the project was fully migrated in around 6 days and the data was once more open and available. This project comprising metadata for c. 4,800 records and over 6,000 images, was described by Professor Hamerow, as ‘a valuable, web-based archaeological resource that had fallen off-line years ago is again being used’,⁵⁰ which clearly articulates that there are challenges but also rewarding solutions to retrofitting sustainability for digital humanities projects.

An ironic challenge and question regularly faced by the Sustainable Digital Scholarship service, both from within and outside the University relates to the financial sustainability of the service and the security and sustainability of the data and projects on the platform. These are valid and important questions which have formed a core part of our service planning

The service is set up as a Small Research Facility at the University, which enables researchers (or research facilitators) to add SDS as a line in research grant applications. This in turn will enable us to recoup some of the running costs of the service (primarily Figshare licences and staff costs). There is also an acceptance that consolidating the costs of running SDS represents greater value for money than the dispersed costs each individual PI, project, department and academic division would have to pay for the upkeep of discrete systems and services for preserving these projects. Furthermore, the University and in particular the Humanities Division have made a long-term commitment to the role the SDS service can play in relation to the sustaining of research data. There is also a growing University-wide acknowledgement that SDS, along with other digital services, may require some central funding to underwrite core digital infrastructures. This will ensure Oxford research is supported in a way which represents value for money for the organisation, and delivers valuable research data management infrastructures and services for researchers.

Conclusion

Systems fail, technology becomes obsolete, people retire, and records are irretrievably lost. These were issues with digital humanities data long before the Covid-19 pandemic, but the lockdown-enforced digital turn has accelerated the adoption of a service for building digital sustainability and open access into digital humanities research projects and their data. Every research project is different and non-standard in its own way; researchers have different research questions, methodologies and approaches. But our work has shown that many researchers have recognised that they can benefit from a level of standardisation in their research data storage and hosting. It is this which the SDS service delivers, within the growing push and pull of open access, data management and FAIR sharing of data.

The SDS service is but one solution in a complex ecosystem of data repositories and infrastructures within and outside the University of Oxford, and we are working in a time of flux with changing research practice, research funder policy and societal expectations of the public benefit of research. As the context of research practice in digital

humanities changes in the years that follow the abrupt and swift digital shift of the COVID-19 pandemic, we hope that the SDS service will be able to adapt to the challenges ahead to ensure the sustainability of research data for our researchers.

Notes

1. Terras, Nyhan, and Vanhoutte (2016, 2–3).
2. Kidd (2009, 139).
3. Watermeyer et al. (2021).
4. Liu, McKay, and Buchanan (2021).
5. Åhlfeldt and Johnsson (2015).
6. Taylor et al. (2022).
7. 'Samuel Beckett: Digital Manuscript Project' (n.d.); 'Welcome to the Newton Project' (n.d.); 'Catherine's Correspondence / Home' (n.d.).
8. 'The Novum Inventorium Sepulchrale' (n.d.); 'The Oxyrhynchus Papyri' (n.d.).
9. 'Online Language Documentation for Biak and Dusner' (n.d.); 'Around n.d.: Activism, Networks, Trajectories' (n.d.); 'Around n.d.: Activism, Networks, Trajectories' (n.d.).
10. Del Turco (2016).
11. McGillivray et al. (2020).
12. Wilkinson et al. (2016); Whyte et al. (2019); Harrower et al. (2020).
13. 'Towards a National Collection | Collections United' (n.d.).
14. Chang (2021).
15. Yusof and Chell (2000).
16. Chiarelli et al. (2022).
17. Chiarelli et al. (2022).
18. Such as Taylor et al. (2022) and McGillivray et al. (2020).
19. For example, AHRC (2022).
20. 'CDLI – Cuneiform Digital Library Initiative' (n.d.).
21. 'The Cult of Saints in Late Antiquity' (n.d.).
22. 'Oxford University Gazette' (n.d.).
23. Partridge (n.d.).
24. Anderson, Galvin, and de Torres Rodriguez (2018).
25. McKnight, Prag, and Madsen (2014).
26. Taylor et al. (2022).
27. Taylor et al. (2022).
28. ERC (2022).
29. Strange et al. (2019).
30. McKnight, Prag, and Madsen (2014).
31. Bentley (2001).
32. Highsmith (2009).
33. 'Ex Nihilo' (n.d.).
34. 'Manifesto for Agile Software Development' (n.d.).
35. Business (n.d.a).
36. Business (n.d.b).
37. 'Figshare for Institutions – a Repository for Research Data of All Types' (n.d.).
38. 'Welcome to the Sustainable Digital Scholarship (SDS) Service' (n.d.).
39. 'University of Oxford Research Repository – Search' (n.d.).
40. 'Around n.d.: Activism, Networks, Trajectories' (n.d.).
41. Gartner (n.d.).
42. For example, this publication from the Software Sustainability Institute: Venters et al. (2018).
43. 'Research Data Oxford' (n.d.).
44. Taylor et al. (2022).
45. 'New Administrative Batch Management in Figshare' (n.d.).

46. Banks [2017] (2022).
47. Figshare Roadmap – Public (n.d.).
48. 'Featured Projects' (n.d.).
49. 'The Novum Inventorium Sepulchrale' (n.d.).
50. 'The Novum Inventorium Sepulchrale' (n.d.).
51. Design based on data from Chiarelli et al. (2022).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was made possible by funding from the University of Oxford.

Notes on contributors

Damon Strange is a freelance Programme and Project Management consultant, with a particular specialism in Research Data Management. Since 2019, playing a vital role at the University of Oxford to design, implement and launch the Sustainable Digital Scholarship service.

Megan Gooch is the Head of the Centre for Digital Scholarship and Digital Humanities Research Support at the University of Oxford. She plays a pivotal role in leading the development and realisation of the long-term vision and strategy for digital scholarship at the University of Oxford.

Alwyn Collinson, Digital Projects Manager on the Digital Scholarship @ Oxford initiative, including the Sustainable Digital Scholarship Service, offering support to research projects and collections across the Humanities Division, Bodleian Libraries, and Oxford's museums.

ORCID

Damon Strange  <http://orcid.org/0000-0002-5851-718X>

Megan Gooch  <http://orcid.org/0000-0002-3190-0509>

Alwyn Collinson  <http://orcid.org/0000-0003-2021-5052>

References

- Åhlfeldt, Johan, and Maria Johnsson. 2015. "Research Libraries and Research Data Management within the Humanities and Social Sciences." <http://lup.lub.lu.se/record/5050462>.
- AHRC. 2022. "Research Funding Guide". UKRI. <https://www.ukri.org/wp-content/uploads/2022/07/AHRC-210722-ResearchFundingGuide.pdf>.
- Anderson, Helen, Elizabeth Galvin, and Jorge de Torres Rodriguez. 2018. "Museological Approaches to the Management of Digital Research and Engagement: The African Rock Art Image Project." *African Archaeological Review* 35 (2): 321–337. <https://doi.org/10.1007/s10437-018-9280-8>.
- 'Around 1968: Activism, Networks, Trajectories'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/around-1968-activism-networks-trajectories>.
- Banks, David. (2017) 2022. "Figshare-Uploader." C++. <https://github.com/amoe/figshare-uploader>.
- Bentley, Colin. 2001. *PRINCE2: A Practical Handbook*. 2nd ed. London: Routledge.
- Business, Agile. n.d.a "Chapter 10: MoSCoW Prioritisation." Accessed 18 October 2022b. <https://www.agilebusiness.org/dsdm-project-framework/moscow-prioritisation.html>.
- Business, Agile. n.d.b "Chapter 4: Principles." Accessed 18 October 2022a. <https://www.agilebusiness.org/dsdm-project-framework/principles.html>.
- 'Catherine's Correspondence / Home'. n.d. Accessed 31 January 2023. <https://catcor.seh.ox.ac.uk/>.

- 'CDLI – Cuneiform Digital Library Initiative'. n.d. Accessed 18 October 2022. <https://cdli.ucla.edu/>.
- Chang, Tao. 2021. 'Arts and Humanities Infrastructure Enabling Knowledge with Impact'. September 2021. <https://www.ukri.org/blog/arts-and-humanities-infrastructure-enabling-knowledge-with-impact/>.
- Chiarelli, Andrea, Neil Beagrie, Lotte Boon, Ruth Mallalieu, Rob Johnson, Amy Warner May, and Rowan Wilson. 2022. "To Protect and to Serve: Developing a Road Map for Research Data Management Services." *Insights* 35: 4. <https://doi.org/10.1629/uksg.566>.
- Del Turco, Roberto Rosselli. 2016. "The Battle We Forgot to Fight: Should We Make a Case for Digital Editions?" In *Digital Scholarly Editing. Theories and Practices*. 1st ed., edited by Matthew James Driscoll, and Elena Pierazzo, 4, 219–238. Cambridge: Open Book Publishers.
- ERC. 2022. 'Open Research Data and Data Management Plans Information for ERC Grantees'. European Research Council. Accessed 18 October 2022 https://erc.europa.eu/sites/default/files/document/file/2022_erc_open_research_data_and_data_management_plans_information_for_erc_grantees.pdf
- 'Ex Nihilo'. n.d. Accessed 18 October 2022. <https://sites.google.com/exnihiloconsulting.co.uk/home/>.
- 'Featured Projects'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/featured-projects>.
- 'Figshare for Institutions – A Repository for Research Data of All Types'. n.d. Accessed 18 October 2022. <https://knowledge.figshare.com/institutions>.
- 'Figshare Roadmap – Public' n.d. Accessed 18 October 2022. <https://eurm.roadmunk.com/publish/50c0cac4ff2d9b46f7c118eb347f7959ffc9f48a>.
- 'Gartner' n.d. Accessed 18 October 2022. <https://www.gartner.com/en/information-technology/glossary/total-cost-of-ownership-tco>
- Harrower, Natalie, Maciej Maryl, Timea Biro, Beat Immenhauser, and ALLEA Working Group E-Humanities. 2020. 'Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities – Digital Repository of Ireland'. <https://repository.dri.ie/catalog/tq582c863>.
- Highsmith, Jim. 2009. *Agile Project Management: Creating Innovative Products*. Pearson Education, London.
- Kidd, Chris T. J. K. 2009. *Digital Sustainability: The Future of Business in a Digital World*. Gower Publishing Ltd.
- Liu, Rui, Dana McKay, and George Buchanan. 2021. "Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage." In *Linking Theory and Practice of Digital Libraries. Lecture Notes in Computer Science*, edited by Gerd Berget, Mark Michael Hall, Daniel Brenn, and Sanna Kumpulainen, 215–226. Cham: Springer International Publishing.
- 'Manifesto for Agile Software Development'. n.d. Accessed 18 October 2022. <https://agilemanifesto.org/>.
- McGillivray, Barbara, Beatrice Alex, David Beavan, Giovanni Colavizza, David De Roure, Federico Nanni, Thierry Poibeau, and Melissa Terras. 2020. 'The Challenges and Prospects of the Intersection of Humanities and Data Science'. A White Paper from The Alan Turing Institute. The Alan Turing Institute. <https://www.turing.ac.uk/research/publications/challenges-and-prospects-intersection-humanities-and-data-science>.
- McKnight, J., Jonathan Prag, and Christine Madsen. 2014. 'DHARMa (Digital Humanities Archives for Research Materials)'. Accessed 18 October 2022. Internal Report.
- 'New Administrative Batch Management in Figshare'. n.d. Accessed 18 October 2022. https://figshare.com/blog/New_Administrative_Batch_Management_in_Figshare/654.
- 'Online Language Documentation for Biak and Dusner'. n.d. Accessed 31 January 2023. <https://www.sds.ox.ac.uk/online-language-documentation-biak-and-dusner>.
- 'Oxford University Gazette'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/gazette>.
- Partridge, Jason. n.d. 'Oxford LibGuides: Oxford Research Archive for Data (ORA-Data): About ORA-Data'. Accessed 18 October 2022. <https://ox.libguides.com/ora-data/about-ora-data>.
- 'Research Data Oxford'. n.d. Research Data Oxford. Accessed 18 October 2022. <https://researchdata.ox.ac.uk/>.
- 'Samuel Beckett: Digital Manuscript Project' n.d. Accessed 18 October 2022. <https://www.beckettarchive.org/>

- Strange, Damon, Blaine Roissetter, Christine Madsen, and Megan Hurst. 2019. 'Digital Humanities Projects Review & Sustainability Analysis'. Accessed 18 October 2022. Internal Report.
- Taylor, Rebecca, Johanna Walker, Simon Hettrick, Philippa Broadbent, and David De Roure. 2022. "Data and Software Policy in the Arts and Humanities Research Community: A Study for the AHRC." Software Sustainability Institute. <https://www.ukri.org/wp-content/uploads/2022/10/AHRC-011122-SSIReport-ShapingDataAndSoftwarePolicyInTheArtsAndHumanities.pdf>.
- Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte. 2016. *Defining Digital Humanities: A Reader*. Routledge, Abingdon.
- 'The Cult of Saints in Late Antiquity'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/cult-saints-late-antiquity>.
- 'The Novum Inventorium Sepulchrale'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/novum-inventorium-sepulchrale>.
- 'The Oxyrhynchus Papyri'. n.d. Accessed 31 January 2023. <https://www.sds.ox.ac.uk/oxyrhynchus-papyri>.
- 'Towards a National Collection | Collections United'. n.d. Accessed 18 October 2022. <https://www.nationalcollection.org.uk/>.
- 'University of Oxford Research Repository – Search'. n.d. Accessed 18 October 2022. <https://portal.sds.ox.ac.uk/search>.
- Venters, Colin C., Rafael Capilla, Stefanie Betz, Birgit Penzenstadler, Tom Crick, Steve Crouch, Elisa Yumi Nakagawa, Christoph Becker, and Carlos Carrillo. 2018. "Software Sustainability: Research and Practice from a Software Architecture Viewpoint." *Journal of Systems and Software* 138: 174–188. <https://doi.org/10.1016/j.jss.2017.12.026>.
- Watermeyer, Richard, Tom Crick, Cathryn Knight, and Janet Goodall. 2021. "COVID-19 and Digital Disruption in UK Universities: Afflictions and Affordances of Emergency Online Migration." *Higher Education* 81 (3): 623–641. <https://doi.org/10.1007/s10734-020-00561-y>.
- 'Welcome to the Newton Project' n.d. Accessed 18 October 2022. <https://www.newtonproject.ox.ac.uk/>
- 'Catherine's Correspondence / Home' n.d. Accessed 18 October 2022. <https://catcor.seh.ox.ac.uk/>
- 'Welcome to the Sustainable Digital Scholarship (SDS) Service'. n.d. Accessed 18 October 2022. <https://www.sds.ox.ac.uk/home>.
- Whyte, Angus, Claudia Engelhart, Daniel Bangert, Gabin Kayumbi-Kabeya, Simon Lambert, Mark Thorley, Ryan O'Connor, Patricia Herterich, and Joy Davidson. 2019. "D3.2 FAIR Data Practice Analysis." <https://doi.org/10.5281/zenodo.5362079>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Yusof, Zawayah M., and Robert W. Chell. 2000. "The Records Life Cycle: An Inadequate Concept for Technology-Generated Records." *Information Development* 16 (3): 135–141. <https://doi.org/10.1177/0266666004240413>.