This version is Read Only. Please provide your feedback by November 15th, 2025 using the form at this link:

https://forms.gle/Ry3o95iczbPKipvx7

Research Sustainability Workflow for the Humanities

Read Me First (please!)

Introduction

Earlier this year DARIAH-IE partnered with Digital Repository of Ireland (DRI) for a day-long workshop on the topic of Research Sustainability at the Royal Irish Academy in Dublin. An output of this workshop was the development of a workflow to support researchers in thinking about sustainability from the very outset of their projects. As part of the collaborative development process we invite the wider DARIAH-IE and DRI community to provide feedback on this workflow, before it is finalised for publication in the Social Sciences and Humanities Open Marketplace (SSHOMP).

What are workflows ?

Workflows are a series of steps or tasks performed in a specific, sequential order to complete a goal. Each step in a workflow consists of a short descriptive text that may include links out to other resources. Workflows are one of the five types of items you can find in the Social Sciences and Humanities Open Marketplace (SSHOMP).

> The goal of a workflow is to be brief and to the point, and let the links provide the added richness.

This Research Sustainability Workflow was drafted using a pre-existing template which is particularly useful when working collaboratively, and it has allowed us to collect all the information needed to create a workflow suitable for sharing in SSHOMP. If you are interested in sharing your own workflows you can find tips on how to create and describe a workflow in this tutorial and you can see an example of a workflow here: Bibliographical Data Science Workflow.

What is SSHOMP ?

The Social Sciences and Humanities Open Marketplace (SSHOMP) is a resource for Social Sciences and Humanities research communities. It contains tools, services, training materials, datasets, publications and workflows in order to highlight and showcase solutions and research practices for every step of the SSH research data life cycle.

The SSH Open Marketplace is:

- a discovery portal, to foster serendipity in digital methods
- an aggregator of useful and well curated resources
- a catalogue, contextualising resources
- an entry point in the European Open Science Cloud (EOSC) for the Social Sciences and Humanities researchers

Contributing to SSHOMP

DARIAH-IE is a registered contributor to SSHOMP and as such can propose content for inclusion, as well as suggest edits to enrich existing entries. In its role as the national node for Ireland, DARIAH-IE can also act as a mediator to submit content and metadata to SSHOMP on your behalf, in such cases you will always be listed as the Contributor (owner) of the resource. The Virtual Record Treasury is one example of an Irish resource in SSHOMP. If you have any resources you would like to add to SSHOMP then please get in touch with DARIAH-IE's National Manager: joan.y.murphy@tcd.ie.

(Reminder) Please provide your feedback on this workflow in the form at this link:

https://forms.gle/Ry3o95iczbPKipvx7

Workflow Template

Field name	Introduction
Label	Research Sustainability Workflow for the Humanities
Description (max *1500 words*)	This workflow was developed with the generous input of researchers, arts professionals, data stewards, cultural heritage practitioners and HSS repository managers across Ireland to help Humanities researchers think through the problem of designing and maintaining reusable datasets for long-term sustainability. Humanities data may include research materials collected from many different sources including: research notes, archival collections, interviews, websites, performances, design processes, etc. Sustainability may be thought of as distinct from shareability or even preservability in that it is not an achievable end goal for a project, but a strategic investment by both creators and users in generating ongoing engagement with the research. We consider that sustainable research data are data which: 1. Have been created and shared as openly as possible in alignment with the FAIR principles, 2. Demonstrate appreciable ongoing value for an audience, and

	3. Benefit from responsible stewardship, often managed through agreements with established organisations capable of ensuring the appropriate continuity of resources. This workflow therefore considers functionality, audience, and environment in addition to more technical considerations such as open file formats and appropriate documentation. We anticipate that this workflow will be revisited during the research lifecycle as part of an iterative process. It should be used in combination with a data management plan (DMP) and other guidance on the preparation and documentation of FAIR (findable, accessible, interoperable, and reusable) research outputs.
Actors	
accessibleAt	AUTO-REFERENCE
externalId	
media	
Related items	

CATEGORISATION metadata

Activity (based on TaDiRAH)	Archiving, Preserving, Disseminating
keyword	Data Management, Sustainability, Bibliodiversity
Discipline (based on ÖFOS 2012)	Social Sciences, Humanities
Language (based on ISO 639-3)	English
Standard (based on SSK Standards List)	

CONTEXT metadata

See-also (url)	
Helpdesk-url	

ACCESS metadata

License (based on Software	Creative Commons Attribution 4.0 International

License vocab)	
terms-of-use-url	
access-policy-url	

(Reminder) Please provide your feedback on this workflow in the form at this link:

https://forms.gle/Ry3o95iczbPKipvx7

Step 1

label	**Set up a research project for sustainability**
Description	The question of how to sustain resources developed in the course of a research project is often left to the 'evaluate and archive' phase of a research lifecycle, which typically occurs after data collection, analysis, and collaboration, but before publication and dissemination of results. By deciding at the start of a project which outputs need effort to support continued reuse and engagement you can increase your ability to effectively manage your resources and expectations. We recommend utilising this research sustainability workflow in combination with a Data Management Plan (DMP). A DMP is a document created at the outset of your project, often as part of a funding proposal, which asks a series of questions about how you will gather, manage, store, and ultimately share, preserve, or dispose of data at the end of a project. When answering these DMP questions, consider sustainability issues - some outputs may be effectively used in the short-term and allowed to fade, while others might need to be sustained through repositories, partnerships, or other collaborative agreements. There are many resources and guides available to follow in setting up a DMP and your funder might have a specific template that they want you to use if preparing a plan to accompany a grant application. DMPOnline provides templates and examples tailored to various funders in the EU and UK and their requirements. We suggest also looking at resources such as the DARIAH-EU research catalogue and the Journal of Open Humanities Data to see examples of datasets and data papers which describe the creation, management, structure, and format of Humanities datasets. This will give you an idea of the kinds of data outputs that other researchers in your field have produced and shared.
Related items	Data Management Plans DMPOnline Funder Requirements Digital Curation Centre Example DMPs and Guidance Journal of Open Humanities Data DARIAH-EU Making Qualitative Data Reusable - A Short Guidebook For Researchers...

Step 2

label	**Document your sources**
Description	This step is concerned with mapping the provenance of the data and processes that will inform your outputs. As a Humanities researcher it is highly likely that your research draws on a variety of sources, historical or contemporary, which you intend to transform by refining, narrowing, or otherwise curating in preparation for analysis, interpretation, or display. Documenting the processes and decisions behind these transformations is a critical part of preparing your research to answer potential questions from future researchers and readers. A reusable data output is one that clearly communicates what sources have been included, what has been added or changed, who is responsible, and also, what was left out and why. By documenting these decisions you can highlight issues that need to be addressed for others to responsibly reuse your data, which might include anything from copyright clearance information to harmful language warnings. Consider the following activities in this step: 1. Record the names of all contributors, either to your project directly or those found in the original sources. Note contributor roles and any relevant personal or organisational identifiers (such as ORCID or ROR numbers) as well as copyright restrictions and citation obligations, if relevant. 2. Consider how you will capture contextual information, particularly around your research processes. There may be things you have not thought of sharing that could enhance the reusability of your resources e.g. data modelling decision you made, or sources that were not available. This information could be presented later as a datasheet or narrative dossier, or form the basis of a data paper. 3. Decide on a system of version control early on, especially if your project will involve collaboration with others. If Artificial Intelligence becomes a part of the workflow, plan ahead to record how and where it was used to make appropriate disclosure statements later if needed. Contextual information should be recorded in the most accessible format possible to ensure its own sustainability. Consider supplying this in a simple spreadsheet or downloadable text file. It may be linked to locations on the web where the corresponding data is currently stored or provided as a descriptive list, depending on the design of the main data output. It is okay if names are repeated if there are multiple roles. For example, a datasheet provided for the Munster Women Writers Recovery Project dataset in the Digital Repository of Ireland includes the following activities, described using the CRediT taxonomy: 1. [Conceptualization](https://credit.niso.org/contributor-roles/conceptualization/): Joan Murphy

	2. [Methodology](https://credit.niso.org/contributor-roles/methodology/): Joan Murphy 3. [Data Curation](https://credit.niso.org/contributor-roles/data-curation/): Joan Murphy Simply recording the names of all contributors and collaborators is a powerful tool for inclusivity, accessibility and equity in any data practice. Early Career Researchers, students, collaborating professionals in other fields and industries, cultural contributors or Indigenous communities may not receive credit for their intellectual labour and traditional knowledge through more traditional publishing pathways.
Related items	The Archiving Dossier Narrative Copyright issues in Secondary Data Reuse Hidden REF CRediT Taxonomy The Collaborators' Bill of Rights Datasheets for Cultural Heritage Datasets Open Researcher and Contributor ID (ORCID) Research Organization Registry (ROR) Traditional Knowledge (TK) labels An Introduction to Version Control Using Github Desktop

Step 3

label	**Appraise your data**
Description	At some point in the research process, you will need to consider how best to organise your data for sharing with others. This organisation may differ from internal structures devised during the active phase of research. There are useful resources designed to help you decide on file structures and naming conventions, but fewer guides to selecting which data to share. Researchers often have a high degree of autonomy from most funders and institutions when it comes to determining which data are shared and how those data are communicated. When preparing an output, you may wish to show all the data you collected or only a subset of data that you have enhanced and refined. You may choose not to publish all the data collected or find that some versions of the data are more useful than others. Some data collected will be useful mainly for shorter-term needs, like raising awareness, managing processes, opening a discussion, or introducing hypothetical questions. Consider the following in this step: 1. Regarding primary data that supports and enhances your research outputs, ask yourself two questions: Were the objects, files, or code useful

	to you in answering your research question? Do you see value in the resource beyond its original intended use by yourself or another researcher or member of the public? 2. Consider how the organisation of the data (and files) may reflect changes in your research over time. It may be helpful to draft a timeline of major decisions (when you stopped archival research, when you decided to add another focus group, when a project member left for other employment, etc.). As you tell your data story, it may also help to determine which data are most critical to that narrative. 3. Set aside any of the data sources (which you identified in the first step of this workflow) for which reuse licences or credit is unclear so that you can follow-up with individuals and organisations about permission to publish. 4. For secondary data sources, such as those taken from cultural heritage institutions, consider whether it would be more appropriate to provide links to digital sources available online already, or whether any access conditions need to be set, e.g. for sensitive or embargoed data. It may be that at the end of this exercise you have one or several potential datasets to share reflecting different approaches, methods, and outcomes. Will you share them all at once or in the same way? Do each of these need to be sustained or is there a critical reusable piece that should be prioritised? Don't be afraid to put yourself in the position of both the creator and the audience in this step, as good appraisal techniques employed now may often inspire further research later.
Related items	Organising Data Checklist for Appraising Research Data for Long-term Preservation How to Appraise and Select Research Data for Curation Bringing Data to Life: Interactive Data Stories

Step 4

label	**Simplify the presentation**
Description	Your research outputs may include books, articles, websites, conference presentations, performances, events, software, geospatial data and code as well as social media posts, nano-publications, or data visualisations. The possibilities are as wide-ranging as your research goals and audience needs. Consider, however, whether the format, layout and design of each of these outputs is as important to sustain as the underlying data. Simple structures, such as spreadsheets or file directories, are generally recommended as they represent the most straightforward way to manage and preserve your data. Repositories widely accept these formats and they can ensure longevity with little maintenance required. Researchers can download

	them, query them, and output their results into other formats without compromising the integrity and provenance of the original information. If more complexity is required to communicate your research, however, you will need to consider how you can maintain both the data and the presentation format. For example, if your research output is in the form of a website, code, database, visualisation, or other complex artefact, it will be more likely to suffer from obsolescence due to software changes which prevent your code from running, or outright loss (e.g. if you leave your institution where you were hosting a website). Additionally, if you use proprietary tools in your research or to facilitate publication of an output, be aware that you may not be able to archive/deposit that tool with a third party (i.e. a repository or institutional archive) at the end of a project, as this may violate the licence or terms of use. Document your choices carefully so both the data and the platform for communication can be maintained. Some helpful tips to consider: 1. Start with community-maintained open source tools and methodically document any customisations you make so that updates to software can be carried out if needed. The Endings Project offers advice specific to the sustainability of Digital Humanities websites, such as avoiding links which can break when platforms are migrated to different hosting environments. 2. Utilise an existing data model or format to guide how you present the data. This could be a metadata standard (e.g. Dublin Core), a common data format (e.g. ISO 8601), or a linked data vocabulary (e.g. Wikidata). This ensures the data may be more easily imported into other systems if needed. 3. Consider alternative ways to present your data using those simple structures mentioned above, which would be less likely to suffer from these problems. For example, if your research output is in the form of a web-hosted data visualisation, you might also provide the data as a spreadsheet or csv file for download, along with static images of some of the more interesting charts or visualisations.
Related items	The Endings Project Software Sustainability Institute Training Hub Dublin Core Metadata Initiative (DCMI) Terms Wikidata FAIRsharing Standards Registry International Standards Organisation (ISO) 8601: Date and Time Format

Step 5

label	**Provide for ongoing stewardship and care**

Description	Once the scholarly contribution is designed and published, who will responsibly manage continued access or engagement? Many researchers turn to data repositories or archives at the end of a research project to take over the management of their data and provide ongoing access for future public benefit. Repositories are not just for storage, they act as key players in scientific dissemination by keeping the data online, active, and reusable, and they will benefit from early knowledge of both your data and your expectations. Disciplinary repositories are uniquely positioned to guide and advise on the design of research products within a specific field, which can help you in the appraisal process and even the research study design. It may be helpful to map out what you imagine to be the necessary short, medium, and long-term timeframes for your outputs before considering where and how to maintain them. ● Short-term data may be data that supports a conference paper, which you anticipate might be most relevant for 2-5 years after dissemination. ● A medium-term dataset could be one which is presented in a custom website that lasts beyond the initial length of the project, and serves as a key engagement resource for exhibitions, talks, or other public events for 10-20 years. ● A long-term dataset may be a compilation of data gathered from a variety of sources and annotated with unique observations that needs to be archived and made available for as long as possible. Consider who your potential partners are that can help you to manage both costs and responsibilities: 1. Researchers at higher education institutions may have resources in place locally to ensure medium-term or long-term storage of their outputs through the institution. Find out early whether these options come with a cost, or are available only for a fixed timeframe. 2. Have a conversation with other potential stewarding organisations as early on in the planning process as possible. Ensure that those best able to invest ongoing time and effort into your data are positioned to understand it and manage it in line with your expectations. Thinking about long-term stewardship also raises questions about environmental and ethical responsibility, including maintaining open communication with communities or people reflected in your research long after the data is published. Discussions with repository and data management experts at this early stage can also, for example, help you to reduce the carbon cost of your data and its presentation format.
Related items	Digital Curation Centre - Where to keep research data guide FAIRsharing repository listings Understanding and using data repositories Digital Humanities Climate Coalition List of statements on bias in library and archives description

	The CARE Principles for Indigenous Data Governance

Step 6

label	**Encourage broad reuse of your data**
Description	If we imagine sustainability as the product of a relationship between data and an audience, then an important consideration in ensuring its ongoing relevance is to encourage reuse. A broad range of people using your data can be a significant asset in terms of securing continued funding, petitioning for technical maintenance or upgrades, and ensuring prioritisation for long-term preservation. Research data don't benefit from the same outreach support as publications, where publishers are involved in the distribution and promotion of the content to relevant audiences, and may require some additional investment. One method to ensure continued engagement with your data is to write a data paper. This will put the data into the publications workflow and make it more discoverable through literature searches. There are also various approaches to seeking community endorsement which may be borrowed from traditional publishing, such as peer review, submission to awards programs, or collaboration with established entities (e.g. University departments, libraries, local museums, community groups). Consider the following in this step: 1. Solicit feedback from your peers using the channels and platforms that helped you most when you were developing your output, encouraging them to share with their networks. 2. If your output exists on a standalone platform is open peer review an option? This may include publishing the review alongside the identities of the reviewers. 3. Publish research outputs as Open Educational Materials, workflows or datasets on platforms that increase visibility such as DARIAH-CAMPUS, SSHOMP or EOSC. 4. Add your research outputs to relevant Zenodo communities and Zotero libraries. Provide a detailed user guide that explains in plain language how to navigate the data or resource, access information about it, or seek help. Ensure that any code outputs are stored in a version-controlled repository such as GitHub or GitLab. These platforms not only provide a place to share your code after your project has ended, but allow others to review your code or to collaborate in its development by submitting enhancements and bug fixes. Where feasible, consider making your data available as Open Access, meaning that others can easily view and download it without the need to sign up or pay access fees. Clearly licence your data for reuse as well, so that researchers can incorporate and cite your data without the need to contact you for permission

	each time. A good licence will ensure any reuse is appropriately credited and provides that any further publication must be licensed as well. You'll not only see your work benefit new research, you'll be able to reuse that research yourself.
Related items	Creative Commons licences DARIAH-CAMPUS Data papers Datasheets for Digital Cultural Heritage Datasets EOSC EU Node GitHub GitLab Journal of Open Humanities Data Open Peer Review SSHOMP Zenodo DARIAH Community Zotero DARIAH-IE library