

Introduction to the Data Management Plan: Specifics of Data in the Arts and Humanities

Dr Deborah
Thorpe

Research Data
Steward



Overview

- Arts and humanities research data?
- The FAIR principles
- The Data Management Plan (DMP)

5 minute break

- Data papers and publicising your shared arts and humanities data
- Some Tips
- Q&A

Arts and humanities
research data?



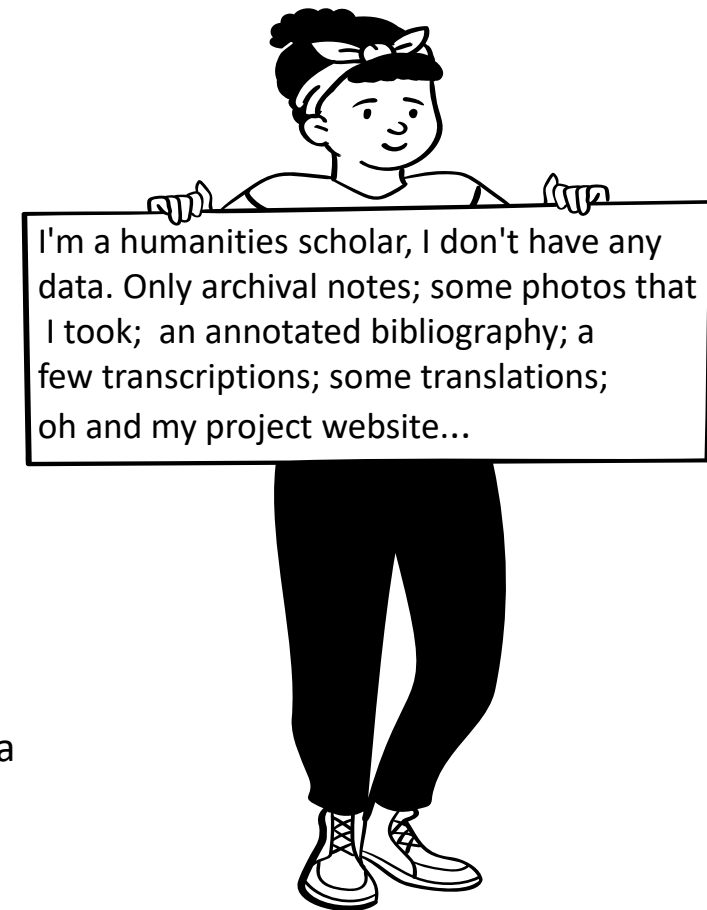
'All materials and assets scholars collect, generate and use during all stages of the research cycle'

TIP: 'Think of **all** your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data'

A [spreadsheet](#) of historical information about marriage banns registration between 1580 and 1810 in Amsterdam: www.doi.org/10.25397/eur.14049842

[Audio recordings](#) of interviews conducted as part of a project on the history and future of Irish and Welsh ports: www.doi.org/10.25397/eur.14049842

[Video files](#) of film-poems by Irish authors: www.doi.org/10.25397/eur.14049842 (a thematic research collection)



A note about a particular feature of arts and humanities data:

‘[The] creation and presentation of resources (including but not necessarily limited to research data) are always, and particularly in the humanities, **acts of curation**. Data are ‘always already’ **marked** by both the epistemic and organisational frameworks **of the creator**’

Michelle Doran, Jennifer Edmond, Georgina Nugent Folan, Seeing Shapes in the Cloud: Perspectives from the Humanities on Interdisciplinary Data Integration, Data Science Journal CODATA, 2022. <http://www.tara.tcd.ie/handle/2262/98529>



Research Intern

University of Oxford

Jun 2012 - Aug 2012 · 3 mos

I worked at the Ashmolean Museum, in the Winton Institute for Monetary History. The Research Intern role involved extracting information about the rental of Oxford University properties in the Early Modern period, contributing towards the Property Rents and Prices in Urban England project.

EHS Internship Report 2012, Deborah Thorpe

My role as an intern with the Winton Institute/ Economic History Society was to contribute to a pilot study for a bigger project that aims to understand property rents and prices in urban England from the 1400s up until the 1900s. The intention of the project is to collect and interpret data for several towns and time periods in order to determine how rents varied over time and what factors affected the rent that property could commandeer (ie. location, state of repair, national policies, the use of the property). For my internship, I was working specifically on Oxford, and even more specifically, on Balliol College and its properties. My task was

	A	B	P	Q
1	Property ID	Date	Tenant name	Residential use
2	BAL 1 1600	1600	Richard Lloyd	1
3	BAL 1 1608	1608	Mrs Langley	1
4	BAL 1 1627	1627	Mrs Langley	1
5	BAL 1 1649	1649	Dr Richard Bayley (Pres. St. John Baptist College)	1
6	BAL 1 1661	1661/1662	Mr John Bayley	1

Questions about spreadsheet entries

How do we record leases that are for several properties to the same lessee (ie. assume that they are sublet)?

If the number of tenements within a single lease was specified, it has been split into several records. For the fine and rent value, the fraction is stated in the first column (eg. 1/3 of...) and then the amount is calculated in the next two columns. The record name has been kept the same for each tenement.

How do we record the 'sale of lease' section in the spreadsheet, as the 'sale value' column refers to sale of property, not of the lease of the property. Eg. BAL 2 1681.

We created a new column

How do we deal with 'acquittances' It seems that they are an amount that is deducted from the rent, but how do we record it? Eg BAL 26 1794 (deduct in shillings and pence column but mark with 'acquittance deducted')

Timeline of events in sixteenth, seventeenth, and eighteenth-century Oxford

1566: Elizabeth I visits Oxford

1571: Outbreak of plague

1577: Outbreak of gaol fever at the assizes at Oxford castle

1592: Elizabeth I's second visit, preceded by extensive cleaning and repairing of public and private buildings.

1603: Outbreak of plague

1603, 1625, 1636: James I visits Oxford

1642 and 1643: Charles I visits Oxford

What about publications?

Both sources for and outputs for arts and humanities research

Publications as research outputs are not generally counted as 'data' for a DMP, but there is an argument that they should be (especially in the arts and humanities):

[in contrast with STEM] 'publications emerge as the most important type of data in the humanities both in terms of research output, but also in terms of input, be it primary or secondary sources'

= explorations of the concepts of 'digital object' ; 'research object' ; 'research materials'; 'sources'

What are the FAIR principles?

Findable

Does the dataset have a Persistent Identifier?

Is there metadata or documentation available? Is this sufficient to fully understand the data content?

Accessible

Is it clear how you access the data? Are the metadata accessible?

Interoperable

Does the dataset have a user licence? Are there clear conditions for reuse of the data? Do user restrictions apply?

Reusable

Are the data files in a proprietary format (not so good)? OR a well supported proprietary format (better) OR an open format (best!)

Are the data linked to other data, and if so, how?

Open
where possible
(not all FAIR data is open)

How can the FAIR Principles be addressed in the arts and humanities?

Download asset

DOI
10.7486/DRI.gh93wr62w

Title
Applying for an Artistic Research PhD – A Prospective Student Guide

Description
This set of resources for artistic doctorates is a major output of the research project Visioning the Future led by Jools Gilson (principal investigator) with Inês Bento-Coelho (postdoctoral researcher) in 2020 / students, staff, and institutions to enhance their doctoral provision for Artistic Research PhDs. This document (for Artistic Research PhD Students) focuses on: Applying for an Artistic Research PhD

Creators
Jools Gilson
Inês Bento-Coelho
Patrick Egan

Publishers
University College Cork

Date
2021

Subjects
Applying for an Artistic Research PhD – A Prospective Student Guide
University College Cork

Formats
PDF

Types
Text

Language
English

Rights
Copyright Inês Bento-Coelho and Jools Gilson

Licence
 Creative Commons Attribution-NonCommercial 4.0 International License

All objects within the DRI are assigned a **unique identifier** within the repository on ingest and automatically assigned a persistent identifier (PID), a DataCite Digital Object Identifier (DOI)

Datasets or Digital Objects ingested into DRI are described by **rich metadata**. DRI supports ingest of metadata in numerous formats including Dublin Core, Qualified Dublin Core, EAD, MODS or MARC XML formats. (Researchers can use a template to gather metadata)

All objects in the Repository require a **Rights statement** which provides information about the rights holder and permitted reuse. Each object also has a Licence associated with it

Above: <https://doi.org/10.7486/DRI.gh93wr62w>

How can the FAIR Principles be addressed in the arts and humanities?

The Eighth Amendment
Constitutional amendments
Referendum
Women's rights
Personal narratives
Social history
Medical History

Subjects (Places)
Dublin
Ireland

Types
Collection

Rights
The photographic material, press clippings, art works and design works are subject to copyright restrictions. For material published on the website please note the attached copyright statements. Please consult NIVAL staff re other material in the collection.

Licence
Please see individual objects for licensing information

Depositing Organisation
National Irish Visual Arts Library (NIVAL), NCAD

Related Materials

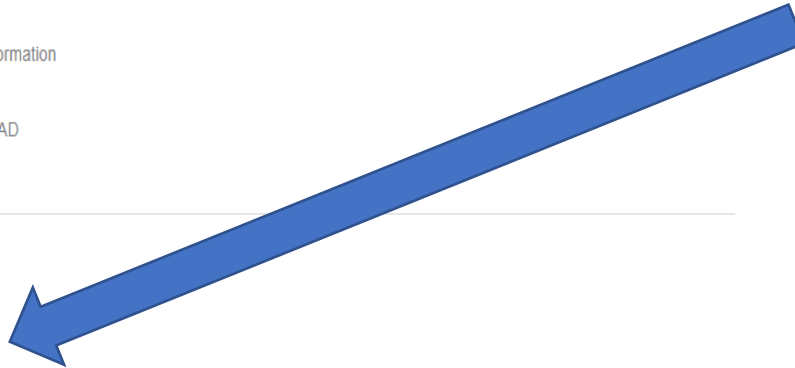
Finding Aid - Artists' Campaign to Repeal the Eighth Amendment Archive

 Has Documentation

Metadata are accessible, **even when the data are no longer available**

The internal DRI unique identifier can also be used to **create relationships between multiple objects**. These relations enable researchers to, for example, link a publication to its data, or to link a dataset to documentation describing how the dataset was collected.

Where appropriate, **information about how the data was gathered (e.g. a ReadMe file)** is also ingested along with the main dataset. This is uploaded as documentation for a collection, or sub-collection of digital objects. A link to the Documentation is displayed on the object or dataset's landing page.



Above: <https://doi.org/10.7486/DRI.gx421568w>

Challenges for data sharing in the arts and humanities

- I don't recognise my research / this was not designed for me / still not sure about 'data'
- Arts and humanities research and data are very diverse
- Do I have to share data? Does this policy apply to me?
- What can I / should I share?
- Ownership and licensing of the data – third party materials, creative works/commissions
- I need training / who can help me with my data?

Source: Cannon, Matthew, & Grant, Rebecca. (2021, September 8). Academic publishers and sharing humanities data: the interface between theory and practice. Interfaces - DARIAH Annual Event 2021, Online. Zenodo.

<https://doi.org/10.5281/zenodo.5495625>

'It is better to **keep the unpublished archival original to yourself**, otherwise somebody can publish it before you.

The material is shared with very few people and only partially, not in its entirety, to **protect the originality of the research and its novelty**, which has to be made public only when the research is finished and complete'

'It is **extra work**, and I feel we should understand that, if that is the case, all projects need a bit more workforce and even people with specific expertise. **We have very small projects**' (see also 'The Invisible Workload of Open Research', <https://doi.org/10.36850/mr5>)

Interviews by Gualandi, Bianca. (2022). What do we mean by "data" in the arts and humanities? Final Dissertation in Open Science [Zenodo]. <https://doi.org/10.5281/zenodo.6385132>

The more we all can organise and share **the more we all benefit**

- But you don't have to share everything / share your data 'right now'

The more you organise and share the more **you benefit**

- Supports transparency, understanding, and trust in your research
- Recognition of the different elements of your research – there is growing interest in this!
- 'Your first collaborator is yourself' – it's difficult to revisit data in the future
- Making data FAIR removes the admin burden on yourself
- Longevity of research – publications date faster than the data

... And (increasingly) meeting the requirements of funders and publishers!

The Data Management Plan can help to overcome the challenges of identifying, working with, and sharing arts and humanities data

[See: Open Data for Humanists, A Pragmatic Guide – Digital Humanities@TCD](#)

What is a Data Management Plan (DMP)?

A formal document that outlines the management and stewardship of data both **during a research project (active management)**, and **after the project is completed**. It describes the **organization, storage, preservation and potential sharing** of research outputs and it is an integral part of good research practice

A **living document** that should be kept up to date



Why should I produce a DMP?

- Essential part of project management
- Establishes clear workflows and guidelines surrounding data management
- Day to day, you can find and understand your data when you need to use it
- There is continuity if project staff leave or new researchers join
- The data underlying publications are maintained, allowing for validation of results
- The data can become another output from a project



Well organised data from the outset! What file formats you use; how you organise your folders; how you name your files; creating versions of files – these decisions will serve you well *if made early on!*



What if you or a colleague leave the university, or leave academia? What if you forget details about your research? *Make your data future proof, understandable, citable, reusable!*



Make the most of your effort!

Structure of Data Management Plans



- 1. Data Collection**
- 2. Documentation, metadata and data quality**
- 3. Ethics and Legal Compliance**
- 4. Storage and Back-up during research**
- 5. Data Sharing and Long-term Preservation**
- 6. Resources and Responsibilities**

<https://doi.org/10.5281/zenodo.4915861>

Source: Aoife Coffey slides

What does the DMP require me to think about?

Data Reuse and Collection

- What data are you reusing?
- What data are you collecting – a name for your dataset; what source/type of data?
- What file formats? What volumes?

Create a table!

Data Collection

Data collected and created during this project will include:

Dataset name	Type	Format of data	Origin/source
Bibliographic materials	Literature review	Zotero group library	Bibliographic data from published materials
Template timesheet and activities record	Template spreadsheet	.xlsx	Desk based planning

File formats

- What format is best suited for data creation?
- What format is best suited for your planned uses?
- What format is best suited for long-term sustainability and sharing of data?
- Should you choose an open versus a proprietary format?
- Should the format be lossy or not?
- Is the format suitable for conversion?

[File formats — UK Data Service](#)

Photo by [Andras Vas](#) on [Unsplash](#)



**LOC Recommended Formats Statement
2022-2023**

ii. Photographs – Digital		
	Preferred	Acceptable
A. Faithful representation of the work	<ul style="list-style-type: none"> • Equal in quality to the published version, best edition or master copy • In the same format as the master copy 	
B. Technical Characteristics	<ul style="list-style-type: none"> • Highest resolution available, not rescaled or interpolated • Highest bit depth available, 16 bits per channel if available • Embedded color profile or specified color space used in published version • Uncompressed • Unlayered 	<ul style="list-style-type: none"> • Lossless compression or lower compression ratios • Discrete wavelet transform (DWT) preferred to discrete cosine transform (DCT) • Layered, if supported by preferred or acceptable format
C. Formats	<ul style="list-style-type: none"> • TIFF (*.tif) • JPEG2000 (*.jp2) • PNG (*.png) • JPEG/JFIF (*.jpg) • BMP (*.bmp) 	<ul style="list-style-type: none"> • Photoshop (*.psd, *.psb) • JPEG2000 Part 2 (*.jpf, *.jpx) • Digital Negative DNG (*.dng) • Proprietary Camera Raw formats (*.nef, *.crw) • GIF (*.gif)
D. Metadata	<ol style="list-style-type: none"> 1. As supported by format: <ol style="list-style-type: none"> a. Title b. Creator c. Creation Date 	Metadata provided separately in external text of XML-based file

What data quality control measures do you use?

- **Map and document your workflow** from the point of collection to the final format dataset
- **Standardise** data capture, data entry or recording methods
- Create and maintain **well organised folder structures**
- **Naming conventions** for your files: choose meaningful but brief names

What data quality control measures do you use?

- Ensure your actions **are reversible**:
 - Make your raw data 'read only'
 - Version control: save modified files with a new name with the version number
- Make use of any **data validation** functions in your software, e.g. Excel allows you to specify permitted values for a cell or range of cells
- **10% checks on the accuracy of transcripts**, especially if as PI you are using a transcription service or otherwise not doing it yourself

Untitled 138.docx
Untitled 241.doc
Untitled 138 copy.docx
Untitled 138 copy 2.docx
Untitled 139.docx
Untitled 40 MOM ADDRESS.jpg
Untitled 242.doc
Untitled 243.doc
Untitled 243 IMPORTANT.doc
Untitled 41.jpg



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Documentation and metadata

- **What metadata will you collect** that is suited for the arts and humanities and your chosen repository? For example, you could use or create a template based on the standards of your chosen repository. This will help you to state which metadata standard you will use
- **What documentation** do you need to store with your data to ensure that people (and yourself) can understand and reuse it?

Metadata is a love note to the future photograph courtesy of Flickr user sarah0s / <https://www.flickr.com/photos/sarahseverson/6245395188>



What **metadata** will you collect to help people to identify and discover the data?

- Metadata = 'Information about the data'
- Made up of a number of elements and each supports a different function
- **Metadata Standards** provide specific data fields or elements to be used in describing data for a particular use
- Help to ensure metadata quality

Dublin Core

E.g.

A basic standard which can be easily understood and implemented

Title:

One of the best known metadata standards

Creator:

Did comprise 15 'elements' and has since been expanded to better represent digital objects

Date:

Description:

[Useful advice from DublinCore about creating metadata](#)

Data is easier to organise, find, understand and use if we use consistent to describe it. This is known as a 'controlled vocabulary'

Use existing **controlled vocabularies** to describe data, e.g.

- The [Library of Congress Subject Headings](#);
- [AAT](#) (Art and Architecture Thesaurus)
- Getty Union List of Artists Names – [ULAN](#)

You can also build upon these if the terms/words that you need are not there

These controlled vocabularies also provide a consistent way of dealing with uncertainty

‘These set vocabularies provide a standardised approach to the indeterminate or unknown, using words such as ‘circa’ for uncertain dates and terms such as ‘anonymous’ for uncertainty regarding authorship. This is a further example of the manner in which adaptation to the messiness of humanistic research data results in accommodation between the needs for common standards and a looser hold on the precision of what is known, such as the human brain is easily able to assimilate’

(Michelle Doran, Jennifer Edmond, Georgina Nugent Folan, Seeing Shapes in the Cloud: Perspectives from the Humanities on Interdisciplinary Data Integration, Data Science Journal CODATA, 2022)

How do I collect metadata?

- **Describe your datasets using a README text file** for logical 'clusters' of data (e.g. a collection of photographs, or just one spreadsheet) stored with the file(s) with information like:

e.g.

- The title of the dataset; creators; dates; keywords; funding sources
- A description of what data each filename contains
- Methods for data collection/generation/curation
- A list of all the column names used in a spreadsheet
- explanations of abbreviations

← This file becomes part of your 'documentation'

See: [Guide to writing "readme" style metadata | Research Data Management Service Group \(cornell.edu\)](#)

Collect batches of discovery metadata to form the basis of the repository 'object' using a template for each data deposit

- Based on the chosen repository, a collection of the metadata that will ensure the findability, accessibility etc of your data
- Could be provided by your chosen repository, e.g. Digital Repository of Ireland [Batch Metadata Template](#)
- Try the [Dublin Core Metadata Generator](#)

Filename	dc:Identifier	dc>Title	dcterms:Alternative	dc:Creator	dc:Creator	dc:Creator	dc>Date	dcterms:Created
MotherShirley.txt	001_Parent	Growing Up in Ireland (9 Years Old) Interview with the mother of Shirley, Child001Shirley		Williams, J,	Greene, S	Harris, E.		2008
Shirley.txt	GUI9YearsOldShirley	Growing Up in Ireland (9 Years Old) Interview with Shirley, Child001Shirley		Williams, J,	Greene, S	Harris, E.		2008

CALENDAR OF LONDON CONCERTS 1750-1800

Overview

Description

Data files (2)

Cite as:

McVeigh, Prof S. (Goldsmiths, University of London) (2014): *Calendar of London Concerts 1750-1800*. DANS. <https://doi.org/10.17026/dans-znv-3c2j>

2014-05-01 | McVeigh, Prof S. (Goldsmiths, University of London) | [➤ 10.17026/dans-znv-3c2j](https://doi.org/10.17026/dans-znv-3c2j)

The calendar contains records of over 4000 concerts, advertised in London daily newspapers from 1750 to 1800: subscription concerts, benefits, oratorio performances, meetings of musical societies, concerts at the principal gardens. Information on concert programmes is provided where known, with editorial standardization of names, genres and instruments.

McVeigh, Prof S. (Goldsmiths, University of London) (2014): *Calendar of London Concerts 1750-1800*. DANS. <https://doi.org/10.17026/dans-znv-3c2j>

CALENDAR OF LONDON CONCERTS 1750-1800

Overview

Description

Data files (2)


Persistent identifier	DOI: 10.17026/dans-znv-3c2j URN: urn:nbn:nl:ui:13-jn-riz1
Title	Calendar of London Concerts 1750-1800
Creator	McVeigh, Prof S. (Goldsmiths, University of London), Project leader
Date created (ISO 8601)	2014-05-01
Description	The calendar contains records of over 4000 concerts, advertised in London daily newspapers from 1750 to 1800: subscription concerts, benefits, oratorio performances, meetings of musical societies, concerts at the principal gardens. Information on concert programmes is provided where known, with editorial standardization of names, genres and instruments.
Audience	Musicology
Subject	Music London concert
Temporal coverage	1750-1800 eighteenth century
Spatial coverage	London England United Kingdom (UK)
Identifier	Fedora Identifier: easy-dataset:131100
Relation	is version of <ul style="list-style-type: none">McVeigh, Simon. 2014. Calendar of London Concerts 1750-1800. [Dataset] at Goldsmiths University of London, Research Online
Type (DCMI resource type)	Dataset


CALENDAR OF LONDON CONCERTS 1750-1800

Overview

Description

Data files (2)






 Download

 View details



Dataset Contents

Dataset Contents

<input type="checkbox"/>	Name 	Size 	Accessible 
<input type="checkbox"/>	 MUS-McVeigh_Calendar_csv Ver_02 29_Dec_2014.csv	1154297	Yes
<input type="checkbox"/>	 MUS_McVeigh_Calendar_Information and Abbreviations_pdf Ver_02 29_Dec_2014.pdf	590939	Yes

SIMON McVEIGH

CALENDAR OF LONDON CONCERTS 1750–1800

advertised in the London daily press

<http://research.gold.ac.uk/10342/>

Department of Music, Goldsmiths, University of London

s.mcveigh@gold.ac.uk

Acknowledgment

Acknowledgment should be made in the form: 'Simon McVeigh, *Calendar of London Concerts 1750–1800*, Goldsmiths, University of London, <http://research.gold.ac.uk/10342/> (Version __, accessed DATE)'. Alternatively: '[Source], located from ...'

Extent of Search

One daily newspaper title has been checked for the entire period (up to 1785 the *General/Public Advertiser*). For most of the period a second title has been checked in addition, and for the later years many more titles were also consulted. The majority of newspapers cited are available on microfilm in the series Early English Newspapers (from Research Publications) and in digitised form in the 17th and 18th Century Burney Collection Database (see

www.bl.uk/reshelp/findhelprestype/news/newspdigproj/burney/index.html).

The Times (Daily Universal Register) is available through *The Times Digital Archive* (see gale.cengage.co.uk/times.aspx/). Both series can be viewed via major research and university libraries. A few newspapers are to be found only in hard copy in the British Library, London; the Bodleian Library, Oxford; or the Beinecke Rare Book Library, Yale University.

Over 4000 records of advertised concerts are included in the database: subscription concerts, benefits, oratorio performances, meetings of musical societies, readings and music, concerts at the principal gardens (Ranelagh, Vauxhall, Marybone). The geographical boundary is that of present-day Greater London. While further searches may reveal further concerts, it is unlikely that there will be a significant number of these. I am indebted to the researches of H.C. Robbins Landon and Thomas B.

The Lothian Diary Project - Oct 2020

No Thumbnail

Date Available

2021-03-30

Type

Dataset

Data Creator

Hall-Lew, Lauren

Cowie, Claire

McNulty, Stephen

Markl, Nina

Liu, Sarah

Lai, Catherine

Llewellyn, Clare

Fang, Nini

Elliott Slosarova, Zuzana

Citation

Hall-Lew, Lauren; Cowie, Claire; McNulty, Stephen; Markl, Nina; Liu, Sarah; Lai, Catherine; Llewellyn, Clare; Fang, Nini; Elliott Slosarova, Zuzana; Klingler, Anita. (2021). The Lothian Diary Project - Oct 2020, 2020 [Dataset]. University of Edinburgh. <https://doi.org/10.7488/ds/3009>.

Description

The ongoing Lothian Diary Project consists of 125+ audio/video recordings collected since May 2020 from residents of Edinburgh and the Lothian counties in Scotland. The diaries comprise self-recorded monologues or semi-structured interviews in which participants discuss their experiences during different stages of the Covid-19 pandemic. Recordings were uploaded to an online survey that also collected consent, demographic information, and opinion regarding Covid-related policies. This archive contains all fully transcribed Lothian Diaries with maximum sharing consent (internet, TV, radio etc.) submitted to The Lothian Diary Project prior to 4 October 2020. There are 24 diaries submitted by participants who agree for their names to be made public (in folder 'Name_Public') and 21 by participants who would like to remain anonymous (in folder 'Anon_Public'). All diaries are identified by a unique participant ID (number). Diaries are stored in WAV format (audio) or MP4 format (video) and time-aligned transcriptions are stored in tab-delimited plain text files. All transcriptions were auto-generated and have been hand-corrected.

The University of Edinburgh, DataShare, DOI: [10.7488/ds/3009](https://doi.org/10.7488/ds/3009).

Download all files



- Anon_Public_recordings.zip (5.204Gb)
- Anon_Public_txt.zip (57.97Kb)
- Lothian_Diaries_Data_Collection_Processing.pdf (81.06Kb)
- Lothian_Diaries_Main_Survey_30062020.pdf (573.8Kb)
- Name_Public_recordings.zip (6.786Gb)
- Name_Public_txt.zip (67.05Kb)
- README.txt (2.392Kb)
- [Preview file]
- license_text (17.00Kb)

PDF that describes the data collection process: contains info on how the data were collected and stored; data processing and how the data has been selected for the repository

.txt file that gives a description of the study/the archive, the difference between 'Anon_Public' and 'Name_Public' in terms of consent, and an explanation of the files in the repository

Creative Commons License: Attribution 4.0 International
<http://creativecommons.org/licenses/by/4.0/legalcode>

Ethics and Legal Compliance

- How will you manage ethical issues and codes of conduct?

Will you need to get consent from participants for preservation and sharing?

Will your data be sensitive? How will it be handled?

Do you have participants whose identity you need to protect?

- How will you manage copyright and Intellectual Property Rights (IPR) issues?

Who 'owns' your data?

What reuse licenses will you apply?

Are there any restrictions on reuse of third party data?

Some tips

- **Establish your sharing rights while you are at the archive**, or otherwise accessing third party materials

If I am taking photographs myself, may I preserve and share them in a repository? If so, under what conditions? Under what license(s)?

Arrange to put in place a 'Memorandum of understanding' for e.g. the relationship between the researcher and data owner; details on the research including estimated completion dates; agreement the terms of the reuse of data that is provided to me; the details of deposit in a repository

- Even if it's not immediately obvious, think about whether there are **any sensitivities in the data** (e.g. personal data; topically sensitive content) and what are the best practices for using this data in research? Just because you are not doing participatory research, doesn't mean you are not responsible for the ethical management of your data

[See: Open Data for Humanists, A Pragmatic Guide – Digital Humanities@TCD](#)

Storage and backup (during the research process)



How will data be stored and backed up during the research?



How will you take care of data security and personal data protection?



How will you manage access and security?

Be aware of your university's storage policies / facilities / advice

Data sharing & Long-term Preservation

(in relation to FAIR / data sharing rather than active data storage)

- How will you preserve the data?
- How will you select which data needs to be preserved?
- Which repository(ies) will you use
- What file formats will you use, that will be easily accessible?
- What licenses will you apply?

What is a trustworthy digital repository?

- Designed to **meet the challenge of storing data for the long term**
- Can accommodate for different levels of access: **restricted access; embargoes**
- **Are certified** against a range of criteria by organisations such as the CoreTrustSeal
- **Provides access** to the data – preservation is only deemed valuable if the data can be accessed
- The digital objects need to be provided with a **Persistent Identifier**, e.g. a DOI. This is different from a URL because it means that data can be located even if their location on the internet changes (a stable link)

Find a **discipline-specific repository** where it exists

Some resources to find them:

- The Registry of Research Data Repositories: Re3data.org
- Fairsharing.org (manually curated information on standards, policies, and repositories)
- OpenDOAR: the quality-assured, global Directory of Open Access Repositories: <https://v2.sherpa.ac.uk/opensoar/>
- Talk to your local experts: the library; data stewards...
- Ask a more experienced colleague
- Look at data papers and/or papers that reference shared data (latter can be difficult in the humanities)

Institutional data repositories

Widely-used **multi-disciplinary / generic repositories**

- DANS (has developed domain area 'data stations', e.g. Data Station Arts and Humanities), open to all, though is a NL org.
- Zenodo (no certification, but has a long standing and solid user base and is operated by CERN)
- Figshare (good findability, excellent content presentation, DOIs, but in terms of long term preservation bear in mind - operated by Figshare LLP so a commercial entity)

A close-up photograph of a person's hand holding a folder in a library or archive. The hand is wearing a ring and a bracelet. The background is filled with rows of folders, creating a sense of depth and organization. The lighting is soft and focused on the hand and the folder it is holding.

How to select data for preservation/sharing

- Not all data needs/should be preserved (or preserved right now)!
- Data that you are **allowed to preserve/share** (your data / you have permission)
- Preserve ... data that **directly supports the result/arguments/conclusions** that are presented in your publications
- Data that was **generated under unique circumstances or at great cost** – e.g. archival visits
- **Any other data** – this is down to your judgement (e.g. other curated data including the associated metadata)

See also: '**How to Appraise and Select Research Data for Curation**' | DCC
This guide will help you develop a managed approach to appraising and selecting datasets for curation.

<https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data#1>

For data that has been reused / third party sources

- For **existing datasets that you have reused**
 - Provide a 'data availability statement' within your publication describing where the dataset is located and how it can be accessed by others
 - Provide a citation for each dataset and add it to your reference list
- For **third party sources that you have accessed but which you will not be sharing in a repository (e.g. data from an archival collection)**
 - In your 'data availability statement' describe where the source material is located and how it can be accessed by others
 - Include a citation for your source in your reference list

(Rebecca Grant, Reusable, FAIR Humanities Data: Creating Practical Guidance for Authors at Routledge Open Research, <https://doi.org/10.2218/ijdc.v17i1.820>) <-- Contains sample data availability statements

+ **preserving your Zotero** annotated bibliographies for each project

Web archiving

A website may be a) the primary output of your funded project, i.e. an interactive online tool or resource and/or b) a dissemination platform for your project

Consider the ephemeral nature of web resources and think about web archiving services such as [The Internet Archive/the Wayback Machine](#) as well as more [DIY solutions to capturing websites for preservation](#)

Consider [preservation by design](#) when making / updating your website. E.g. digitally preserve outputs (e.g. reports; presentations etc) and link to them using PIDS; avoid proprietary formats; design navigation of your website carefully etc

Blog post on this forthcoming!

Responsibilities & Resourcing

- Identify **who is responsible** for **what** in relation to data management
- **What resources** and costs do you need if relevant for data management?
- What **training and support** is needed?



5 minute break!



Data papers and publicising your shared data



- Data papers

Journal of Humanities Data

Research Data Journal for the Humanities and Social Sciences

- Europeana

- ORCID

- Humanities Commons

Reading: The Lothian Diary Project: Investigating the Impact of the COVID-19 Pandemic on Edinburgh and Lothian Residents

[Download](#)

A-

A+

“

Alt.
Display

Share:



Collection: [Humanities Data in the time of COVID-19](#)

Data Papers

The Lothian Diary Project: Investigating the Impact of the COVID-19 Pandemic on Edinburgh and Lothian Residents

Authors: Lauren Hall-Lew , Claire Cowie, Stephen Joseph McNulty, Nina Markl, Shan-Jan Sarah Liu, Catherine Lai, Clare Llewellyn, Beatrice Alex, Nini Fang, Zuzana Elliott, Anita Klingler

[Contents](#)[Annotations
& Comments](#)[Related
Articles](#)[Abstract](#)[\(1\) Overview](#) ▾[\(2\) Method](#) ▾[\(3\) Dataset Description](#) ▾[\(4\) Reuse Potential](#)[Notes](#)[Acknowledgements](#)

Rescuing a Heritage Database: Some Lessons from London Concert Life in the Eighteenth Century

Arts and Media

In: *Research Data Journal for the Humanities and Social Sciences*


Author: [Simon McVeigh](#)

Online Publication Date: 04 Nov 2020

[Abstract](#) [Full Text](#) [PDF](#) [Metadata](#) [References](#) [Figures](#) [Metrics](#)

Abstract

The paper outlines the genesis and subsequent transformation of the database *Calendar of London Concerts 1750–1800*, now available as a dataset at <https://www.doi.org/10.17026/dans-znv-3c2j>. Originally developed during the 1980s, the database was used as a primary research tool in the preparation of articles and a 1993 monograph: the first comprehensive study of London's flourishing public concert life in the later eighteenth century, which culminated in Haydn's London visits in 1791–5. The database itself, extending to over 4000 records, was derived from an exhaustive study of London newspapers. Following the obsolescence of the relational database in which the material was initially stored, it has recently been transferred to a spreadsheet in csv format, publicly available with free open access. Issues arising out of the standardisation of concert data are explored, especially regarding the layout of complete concert programmes, and the strengths and

 Open Access



[Download PDF](#)

[Download Citation](#)

[Get Permissions](#)



[Abstract](#)

[Keywords](#)

- [1. Introduction: The History of a Dataset](#)
- [2. Methods and Coverage](#)
- [3. The Dataset](#)
- [4. Conclusion: Advantages and Disadvantages](#)

[References](#)

Exhibit of Datasets

Research Data Journal for the Humanities and Social Sciences



◀ Home

More Showcases ▶

McVeigh, S. (2020)

Goldsmiths, University of London

Rescuing a Heritage Database: Some Lessons from London Concert Life in the Eighteenth Century

Showcase



DATASET



DATAPAPER



WEBSITE



EXPLORE DATA



HOW TO CITE



Vauxhall Gardens, one of London's well-known concert venues, with visitors and the orchestra in the background. Engraving by Edward Rooker (1712-1744) after Giovanni Antonio Canaletto.

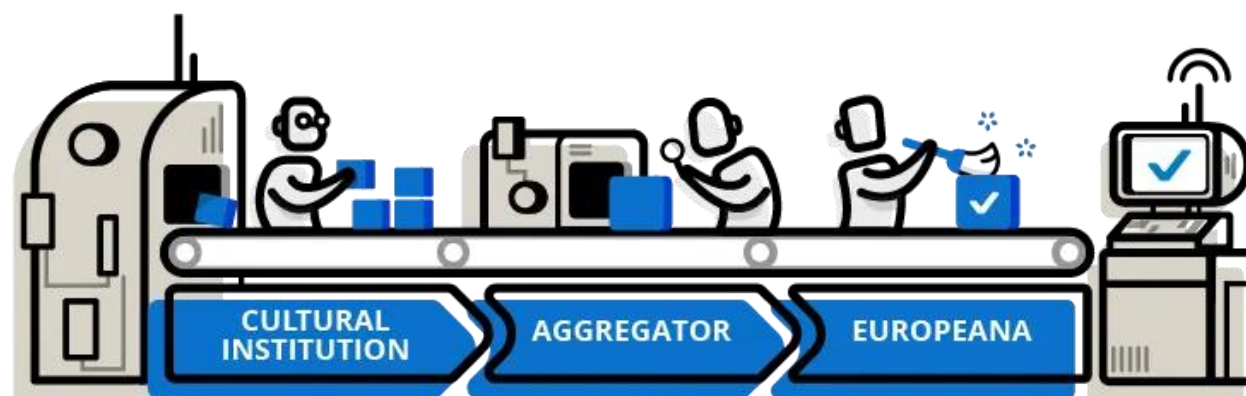
The database *Calendar of London Concerts 1750–1800* is about the flourishing public concert life in London in the later eighteenth century.

This city had a pioneering role in the development of a public concert culture. The commercial opportunities provided by an ever-expanding and increasingly wealthy metropolis were linked to an openness towards new music and a welcome towards international virtuosi that attracted top musicians from across Europe. The increasing public prominence of concerts in the second half of the century placed the symphony concert alongside the Italian Opera at the heart of the fashionable calendar of the leisured classes.

[McVeigh 2020 - Showcase \(dansdatajournal.nl\)](https://dansdatajournal.nl)





Europeana provides cultural heritage enthusiasts, professionals, teachers, and researchers with digital access to European cultural heritage material. Why? To inspire and inform fresh perspectives and open conversations about our history and culture. To share and enjoy our rich cultural heritage. To use it to create new things.



LGBTQ+ lives and heritage

This gallery presents queer histories and cultural heritage collections relating to LGBTQ+ lives, communities and campaigns across Europe.

Curated by  europeana  Published gallery

 SHARE

43 ITEMS



Laurie Steele ICTU
Conference Cork 1981

Kieran Rose

Cork LGBT Archive



Dokumentation av Örebro
Pride 2013, den 31 augusti
2013. Förträff vid Scandic
Grand Hotel.

Torgén, Per

Örebro County Museum



Rockmärke

RFSL Riksförbundet för
homosexuellas, bisexuellas,
transpersoners och queeras
rättigheter

Bohuslän Museum



photograph - Queer fish.
Front and back

Battersea Arts Centre



ORCID



**The ORCID iD: a unique, persistent identifier
free of charge to researchers**

You can use the same iD throughout your whole career, even if your name changes or you move to a different organization, discipline, or country.

You do not need to be in an academic position to create / keep an iD - anyone who participates in research, scholarship, or innovation can have one

▼ Works (31)

☰ Sort

Creative Connections subcollection, DRI

The Ports, Past and Present Project
2023 | Data set
DOI: [10.7486/DRI.DN409T50V](https://doi.org/10.7486/DRI.DN409T50V)
CONTRIBUTORS: The Ports Past

[Show more detail](#)

Source: James L. Smith

Documentary Film Still Images subcollection, DRI

The Ports, Past and Present project
2023 | Data set
DOI: [10.7486/DRI.FN10MQ90G](https://doi.org/10.7486/DRI.FN10MQ90G)
CONTRIBUTORS: The Ports Past

[Show more detail](#)

Source: James L. Smith

Documentary Films subcollection, DRI

The Ports, Past and Present project
2023 | Data set
DOI: [10.7486/DRI.DR27BQ340](https://doi.org/10.7486/DRI.DR27BQ340)
CONTRIBUTORS: The Ports Past

[Show more detail](#)



- Create a profile
- Create a WordPress website
- Share and preserve your work in their Repository
- Discover open access scholarship

as an alternative to commercial networks such as Academia.edu

"the first thing to note is that, despite its misleading top level domain (which was registered by a subsidiary prior to the 2001 restrictions), Academia.edu is not an educationally-affiliated organization, but a dot-com, which has raised millions in multiple rounds of venture capital funding."

Kathleen Fitzpatrick, ['Academia, Not Edu - Kathleen Fitzpatrick \(kfitz.info\)'](http://kfitz.info)

Some final tips...

- If you are applying for funding / have funding, **find out if there are requirements or recommendations on DMP templates** and use that one. If there is not one, look at the existing templates and use one of them
- Create your DMP **as early as possible** and **keep updating it**
- **Plan for data sharing as early as possible** – collecting the highest quality data (e.g. photographs); agreements regarding third party data; consent for sharing data open access within a repository if relevant; ethical issues, sensitivities...
- Consider the **resources** needed for data management, including your **time** and the work needed at the end of the project

Some tips...

- Plan **how to collect your metadata and documentation**
- Speak **to those who are in the know**: the archivists when you are accessing sources; your library; data steward; repository support staff
- Be **aware of your processes and tools** and how they support open research: use Zotero for your literature reviews to create annotated bibliographies; use persistent identifiers in citations; consider learning how to use tools to clean your data such as OpenRefine
- Consider what you, as an arts / humanities researcher **can bring to the conversation** – don't be passive, be active in the discussions / be an ambassador/leader; CARE; Climate issues



Q&A

Dr Deborah Thorpe
Research Data Steward
University College Cork Library

Twitter: @DebsEThorpe

Mastodon: @debsethorpe@hcommons.social
dthorpe@ucc.ie



Further resources / referenced here

DHCC Information, Measurement and Practice Action Group. (2022). A Researcher Guide to Writing a Climate Justice Oriented Data Management Plan (v0.6). Zenodo. <https://doi.org/10.5281/zenodo.6451499>

Gualandi, Bianca. (2022). What do we mean by "data" in the arts and humanities? Final Dissertation in Open Science [Zenodo]. <https://doi.org/10.5281/zenodo.6385132>

Harrower, Natalie, Maryl, Maciej, Biro, Timea, Immenhauser, Beat, & 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 [Distributor], Digital Repository of Ireland [Depositing Institution], <https://doi.org/10.7486/DRI.tq582c863>

Knazook, Beth. Introduction to research data management (slideshare.net), https://www.slideshare.net/dri_ireland/introduction-to-research-data-management-251777098

DARIAH Pathfinder to Data Management Best Practices in the Humanities | DARIAH-Campus, <https://campus.dariah.eu/resource/posts/dariah-pathfinder-to-data-management-best-practices-in-the-humanities>

Manage, Improve and Open up your Research and Data – Parthenos training (parthenos-project.eu), <http://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/>

Health at the writing desk of John Ruskin: a study of handwriting and illness

Deborah E Thorpe,^{1,2} Jane E Alty,^{3,4,5} Peter A Kempster^{6,7}

¹Trinity Long Room Hub Arts & Humanities Institute, University of Dublin Trinity College, Dublin, Ireland

²The Department of Electronic Engineering, University of York, York, UK

³Department of Neurology, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

ABSTRACT

Though John Ruskin (1819–1900) is remembered principally for his work as a theorist, art critic and historian of visual culture, he wrote exhaustively about his health in his correspondence and diaries. Ruskin was prone to recurring depressive and hypochondriacal feelings in his youth and adulthood. In 1871, at the age of 52 years, he developed an illness with relapsing

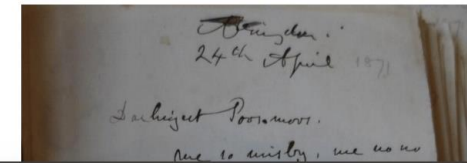
inner feelings and outward appearance. He often made such references to well-being ‘in body and mind’ in his correspondence.² This letter, though, exemplifies the acuteness of Ruskin’s awareness of his decline, which he thought evident in his physical demeanour. Ruskin was dismayed by his thinning beard. Rather than physically ill, he thought he looked ‘poo’—‘bad’ in his character-

Ruskin Library Visit 22nd - 25th January

Notes on documents already viewed, with photographs

Box L 36 - John Ruskin to Joan Severn, 1871:

24th April 1871 - nonsense letter from Ruskin to Joan Severn, beginning ‘me so misby me no no fot to do...’, also mentioning being ‘very ill’ (see Rachel Dickinson, *John Ruskin's Correspondence with Joan Severn: Sense and Nonsense Letters*, p. 131)



20th April: ‘eyes not right yet’;

23rd April: ‘No, not yet - neither, and I was curiously and alarmingly ill in general feeling - indescribable...eyes swimming in my head after lunch...Cut off my cheese and port and shortbread - and physic, and am better by a deal therefore this morning’)

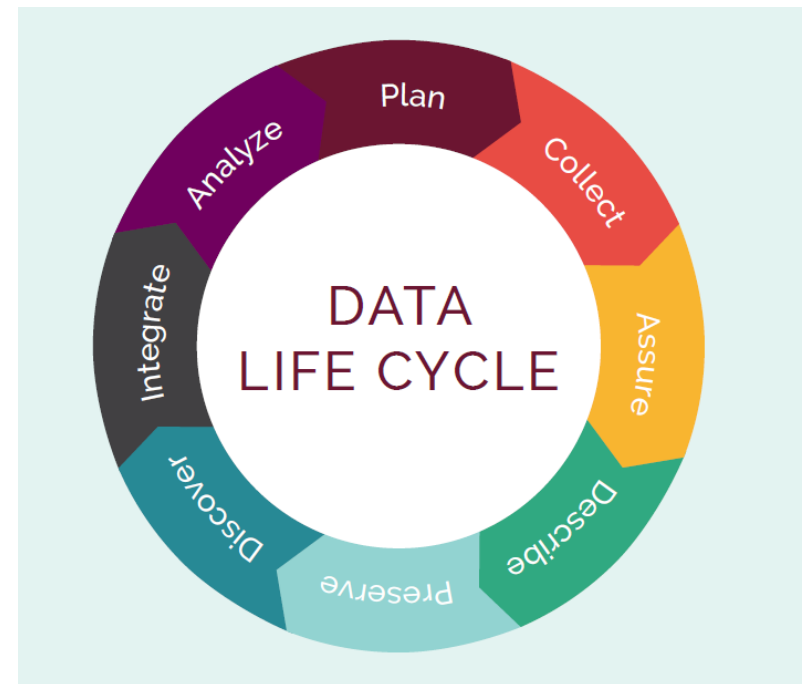
24th April: ‘not sleeping since one, I am not languid, nor in any depth depressed but dull - joyless - effectless, and discomforted as not often in life before’

29th April: ‘I slept well, but eyes not right yet, and I can’t eat rightly for my teeth’

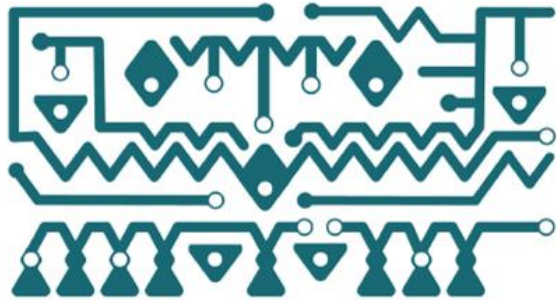
30th April: ‘Teeth and eyes still bad - and - it may be - soul and heart - worse - compared with what they ought to be’

What is data management?

- It is the **active** management of research data during the lifecycle of a project and also includes decisions about treatment of the data post-project
- It involves organization, storage, preservation and potential sharing of the data
- It's an integral part of good research practice



Building on FAIR



CARE Principles for Indigenous Data Governance

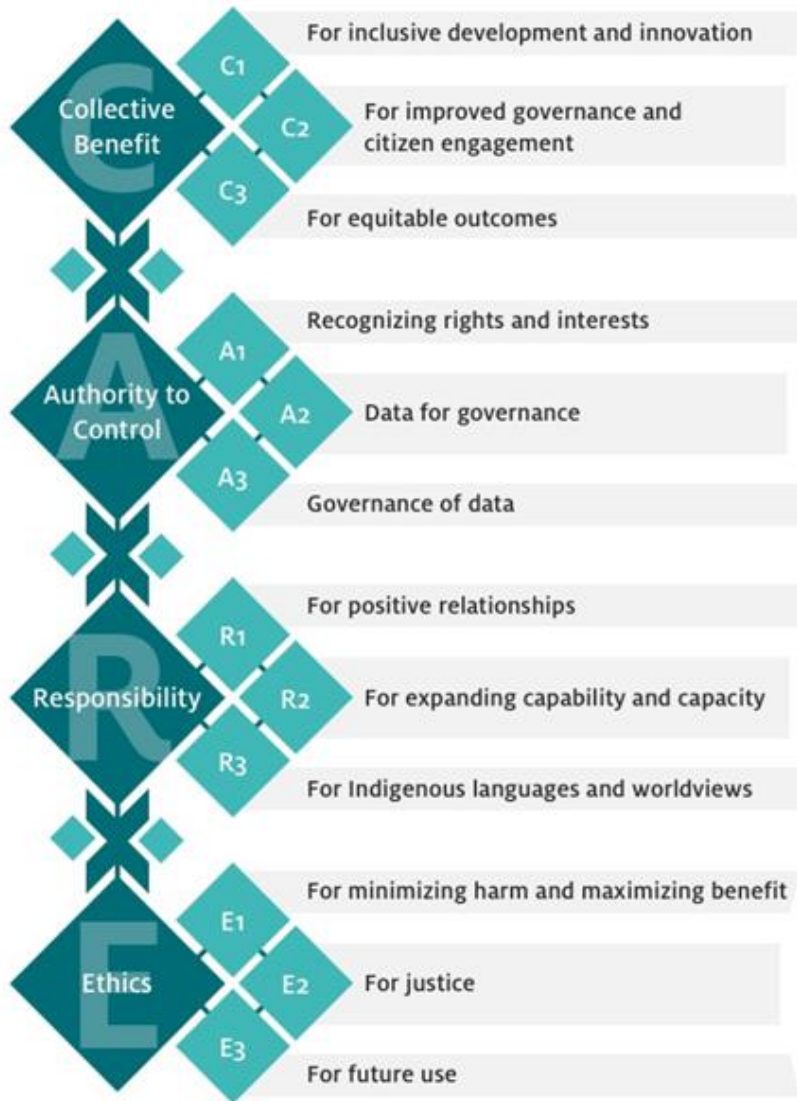
'Existing principles within the open data movement (e.g. FAIR: findable, accessible, interoperable, reusable) primarily focus on characteristics of data that will facilitate increased data sharing among entities while ignoring power differentials and historical contexts'

'The CARE Principles address important considerations for modern data ecosystems and across data lifecycles that support both innovation and Indigenous self-determination'

'The Principles define rights, interests, and concepts to be employed in facilitating Indigenous control in data governance and reuse'

'The implementation of the CARE Principles in tandem with the FAIR Principles will result in data that reflect the realities of Indigenous Peoples, be useful for Indigenous purposes, and remain under Indigenous control, while promoting knowledge discovery and innovation'





See also:

[Thinking about the CARE Principles in the Digital Humanities | DARIAH-Campus](#)

'Prof. Dan O'Donnell (University of Lethbridge) discusses the CARE principles, how they sit alongside the FAIR Principles, and how (digital) humanists can apply them in their research. He presents examples from his own research, particularly around studies of historical artefacts in small rural communities in Scotland.'



A Researcher Guide to Writing a Climate Justice Oriented Data Management Plan



- Addressing the environmental impacts of research
- Managing data sustainably
- Preparing for future funding frameworks
- Developing and sharing best

Demonstrate that the research design seeks to ensure that the resources used (e.g. hardware purchases, compute time, data storage) will be proportional to the results produced (e.g. outputs, anticipated findings, impacts)

Justify the environmental costs of new device purchases, to demonstrate alignment with institutional policies on device recycling

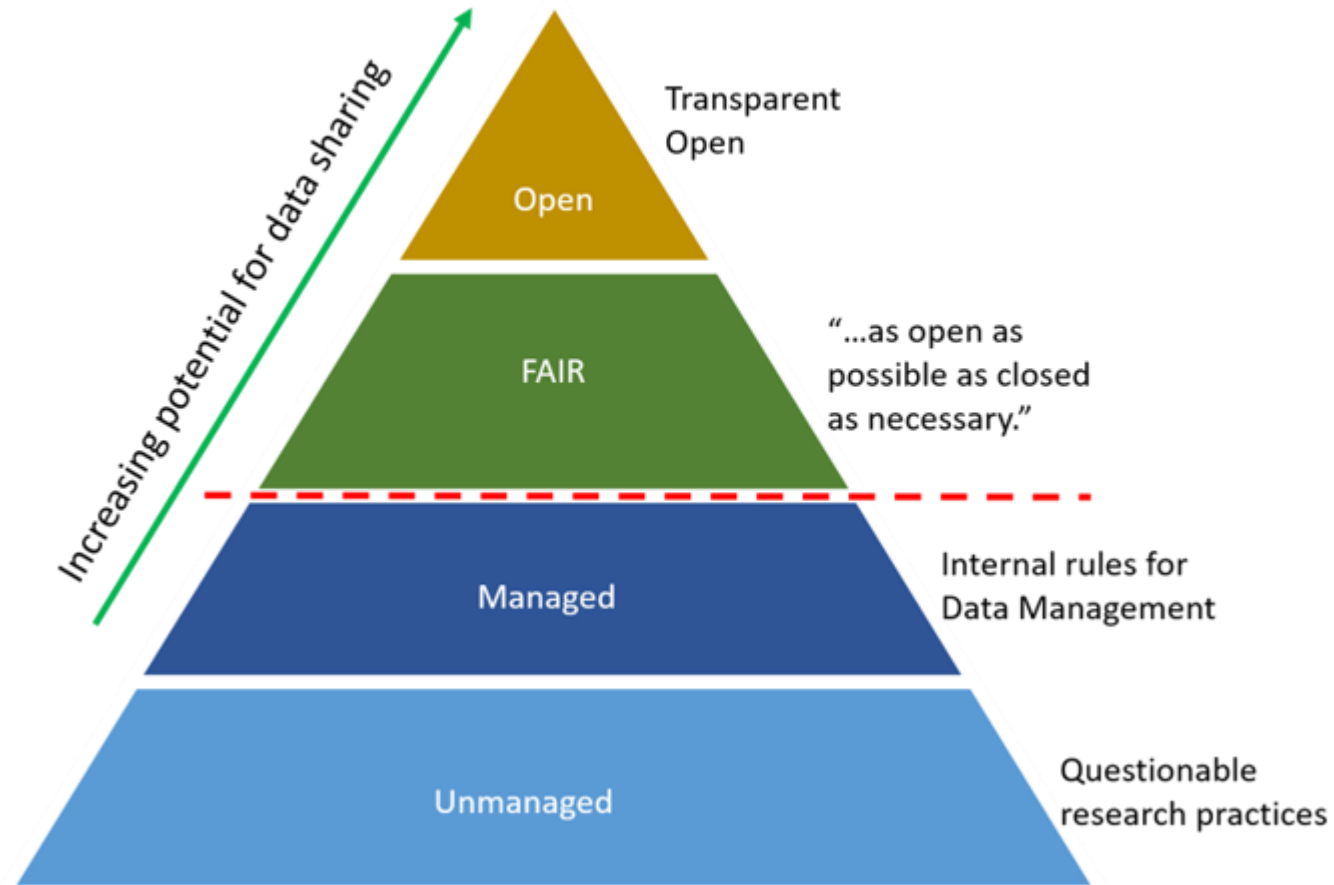
Carefully select data for preservation, knowing that preserving data without purpose is not energy proportionate

A Researcher Guide to Writing a Climate Justice Oriented Data Management Plan, <https://doi.org/10.5281/zenodo.6451498>

Archival notes
Text with markup
Interviews and transcriptions
Text corpora
Photographs: e.g. of
manuscripts
3D scans
Maps
Timelines
Visualisations
Spreadsheets
Videos
Drawings
Websites



Degrees of Data Management



Modified from slide by [@sjDCC](#)

FAIR and open (where possible) should underlie research data management

Though publishers are increasingly requiring data sharing, their data sharing policies **do not account for the specifics of humanities data**

Mandatory aspects of guidelines, e.g. 'your dataset(s) must be deposited in an appropriate data repository' has potential to confuse authors who, e.g.:

- Reuse existing datasets
- Work with third party digital sources e.g. from a gallery or museum
- Work with bibliographical sources only