Plan Overview

A Data Management Plan created using DMPTuuli

Title: Dark Ages, Old and New

Creator: Aleksi Peura

Affiliation: University of Helsinki

Template: General Finnish DMP template

Project abstract:

My PhD project focuses on the interwar period writings of H. G. Wells and others. The research is in the field of intellectual history.

ID: 27513

Last modified: 03-04-2025

Dark Ages, Old and New

1. General description of the data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

Type	Where from?	Format	Size	Storage	Publish
Books published by Wells and others	Libraries, online repositories	Physical, .pdf	max. 10 GB	Physical book shelves, Zotero and UH's Z-drive	N/A
Notes	Created	Physical, .docx	max. 10 MB	Physical notebook, UH's Z-drive	N/A
Wells' correspondence	Library of the University of Illinois	Physical	N/A	Library of University of Illinois	N/A
Photos of letters	Created	.jpeg	max. 1 GB	UH's Z-drive	N/A
Table of Wells bibliography	Created	.csv	max. 10 MB	UH's Z-drive	Zenodo

1.2 How will the consistency and quality of data be controlled?

For books published by Wells and others, the quality will be controlled by using as original as possible editions of the works. Especially e-books will be checked carefully, since some e-book formats will change page numbers, hence affecting referencing and citations.

The notes will be firstly written down in a notebook, from where they'll be transcribed to a .docx-file. All quotations from books or letters will be double-checked for accuracy.

The clarity of the photos of the letters will be checked asap. after taking said pictures, since a slight tremor of hands might cause the photos to become useless. Even a small mistake in picture taking might cause harm for the research, if it leads to misinterpretation later in the research project.

There are no plans for using OCR, since there are edited collections of the correspondence and these can be used as backups, though in a limited capacity. In case the handwriting turns out to be unreadable, aid in deciphering will be asked from project's supervisors or other more experienced researchers. Should the handwriting turn out to be too difficult even for experienced researchers to decipher, it is unlikely that OCR could be of assistance.

Table of bibliography will be compiled using several different available versions of Wells' bibliographies. The bibliography will include all book-length works by Wells, but since the line between a book and a pamphlet is a blurred one with Wells, careful review of each entry is required. In the bibliography a classification of the works will also be included, again compiled from various sources and also enhanced by the PI's own, more detailed classification.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

No major legal issues are anticipated. The most critical ones are IPR-related issues, but since most of Wells' works are nowadays in the public domain, even these are limited in scope.

2.2 How will you manage the rights of the data you use, produce and share?

The photos of the Wells' correspondence will be limited by the Library of the University of Illinois' IPR, so they cannot be shared or published. As the letters are at least 85 years old and the authors have been, at the time of writing of said letters, adults, it is highly unlikely anyone is still alive - therefore GDPR's requirements are not relevant, since they do not apply to deceased persons. The bibliography created during the project is comprised of metadata about the books and can therefore be freely shared.

3. Documentation and metadata

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata

standards, README files or other documentation will you use to help others to understand and use your data?

Mostly not applicable, since the main research material used (the letters and the books) are already available in libraries and online archives. The photos of the letters will be named using the naming system used by the Library of the University of Illinois to ensure correct referencing.

A reference guide to the notes will be created detailing used acronyms to ensure the usability and understandability of the notes. This is functions similarly to a README.txt-file.

The notes will also include, in addition to page numbers of the books, information about which chapter of the book the quotation is from: this information will take a form of book number in Roman numerals / chapter number / section number, e.g. I/4/§2. This will increase the transparency the quotations, since these are not dependent on edition used.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will the data be backed up?

The physical books and notebooks are stored on a physical bookshelf located at the PI's office. The electronic transcriptions of the notebooks are stored in the PI's Z-Drive and they function as backup for the physical notebooks.

E-books are stored mainly in Zotero, which is also used for reading the e-books, and as backups on the PI's Z-Drive. Storing the same e-books in two different locations should not cause issues regarding disk space limitations, since the files are not particularly large.

The photos taken by the PI are stored in the UH's Z-Drive, which creates hourly an automatic backup. The photos will be transferred asap. from the phone used to Z-Drive to ensure their safety - after transferring, the photos are removed from the phone.

The bibliography table will be stored in the PI's Z-drive as well.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

Not applicable, since no sensitive data is handled and the research project only includes the PI.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

The books and the letters are already publicly available - many of them are in public domain as well - and publishing them is pointless. The letters are also affected by the IPR of the Library of the University of Illinois, and cannot be shared.

Sharing the notes is most likely a moot practice: they are unlikely to be relevant to anyone else, since their creation is tied to research questions of this particular research project. It is difficult to foresee a situation where anyone else would be interested in seeing the notes, since these are mostly just quotations or observations from the material they themselves can easily access. Though, should someone ask, the notes can be shared.

The bibliography will be published in Zenodo in a .csv format. This will grant the bibliography a DOI, making referencing it easier.

5.2 Where will data with long-term value be preserved, and for how long?

The books and the letters are already preserved for future generations in libraries and online archives. The notes created will not have long-term value for others, but will be stored by the PI for the foreseeable future. The bibliography will be available in Zenodo, which does not curate the data, but since the table will be stored in an open .csv-format, it will in all likelihood be usable with future software(s) and operating system(s).

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management?

The PI will be responsible for everything, except maintaining the servers where the files of the Z-Drive or the Zotero are stored. Should issues arise, the PI will ask UH's It Center and UH's Library for help.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

Two hours per week is set aside for transcribing the notes created during the week. Should this amount end up not being adequate, it will be scaled up to make sure all the notes are transcribed on a weekly basis.

Z-Drive is free and accessible for all UH-affiliated researchers and therefore no resources will be required.

Zotero gives free storage space for up 300 MB. Should this amount end up not being adequate, the storage upgrades range from 20 to 120 $ per year.