

## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

### Abstract

This project gathered data from print bibliographies and digital databases in order to develop the first comprehensive bibliographical database of women's contributions to print culture from 1750-1830. The purpose of this publicly available database is to enable new quantitative understandings of women's involvement in print culture. This data management plan is accompanied by a metadata schema developed by the researchers to describe the unique contents of the database (link provided within).

**Principal investigator:** Michelle Levy (Simon Fraser University)

### Administrative Details

**Project Name:** Women's Print History Project (1750- 1830)

**Grant Title:** 31 - R639951

**Principal Investigator / Researcher:** Michelle Levy

**Project Data Contact:** [cgraebne@sfu.ca](mailto:cgraebne@sfu.ca)

### Description:

The Women's Print History Project, 1750-1830 will be the first comprehensive bibliographic database of women's contributions to print for one of the most convulsive periods in both women's and print history. Data from print bibliographies and digital databases is being aggregated and normalized to enable new quantitative understandings of women's involvement in print culture. The database, which will be made publicly available for use by other researchers, will enable rigorous quantitative analysis of patterns in women's print history, building upon the extensive body of qualitative scholarship on women's literary history that has been produced over the last half century. Some of the questions to be posed of the data include: How did women's print output change over time? How likely were women to publish more than one book in their lifetimes, and how often did books reach second editions or more? Did particular publishers print more women than others? How often did women change



Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Portage Network | [portage@carl-abrc.ca](mailto:portage@carl-abrc.ca) | [portagenetwork.ca](http://portagenetwork.ca)

## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

publishers over the course of their careers? What social networks were formed through women's engagement with publishers, printers and each other? How might we better understand women's contributions to the history of print in a variety of roles (as authors but also as co-authors, editors, translators, patrons, engravers, or booksellers)?

**Institution:** Simon Fraser University

### Data Collection

**What types of data will you collect, create, link to, acquire and/or record?**

*When gathering data from written sources, it is helpful to provide the reader with information surrounding the material(s) from which you'll collect, as well as a few examples of the resources you'll draw upon to find those materials.*

We will be collecting bibliographic metadata about titles of written works, persons and firms such as title of work, name of author, publisher, date published, etc. Currently we are collecting data from about 50 different sources, some print, most digital. Examples include the British Library online catalogue and specialized print bibliographies like the English Novel 1770-1830.

*It is useful to include where these data are being collected from, using examples of the print bibliographies and databases from which you'll be aggregating this information.*

**What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?**

Data is coming from some sources provided to us (MARC<sup>1</sup> records, for example, from ESTC<sup>2</sup>), and from a few other libraries/projects. Most of the data however is being found and entered by

<sup>1</sup> Machine-Readable Cataloguing

<sup>2</sup> English Short Title Catalogue



## **Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)**

us, consulting print and Digital bibliographies and catalogues, as well as full-text resources like ECCO<sup>3</sup> and google books. Sometimes we have bibliographical data, often we are creating the data by looking at other sources. We have also 'hand checked' or visibly examined a much smaller number of books.

All data is saved in a MySQL<sup>4</sup> database from which the data can be outputted as .csv files.

### **What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?**

The project team has developed a set of tables within our MySQL database. Each record has a unique identifier and tables are linked through these keys. For example, title #3 is "Life of Carsten Niebuhr". It was printed in location #2643743 (London, GB). Sarah Austin (person #9) is the author.

## **Documentation and Metadata**

### **What documentation will be needed for the data to be read and interpreted correctly in the future?**

We will develop a data map structure with details about data for each field. A detailed description of the methodology will be documented and updated when needed.

### **How will you make sure that documentation is created or captured consistently throughout your project?**

We have regular team meetings to ensure that consistent standards are applied. Within the title record, we have both an "attempted final check" (when we are unable to confirm 2 sources for our title entries, but we have attempted to do so) a 'final check' mechanism (when we have been

<sup>3</sup> Eighteenth Century Collections Online

<sup>4</sup> MySQL is an open-source database management system



## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

able to confirm two independent sources) to make sure our data standards are met. When we have hand-checked a title (meaning we have physically examined it), we note that and final check the entry; in these cases, we do not require two sources. We also have a “final check” for our persons table data, also to ensure that records have been checked against reliable sources.

**If you are using a metadata standard and/or tools to document and describe your data, please list here.**

*If you and your research team have created a metadata standard to apply to your data and documentation, it is beneficial to the reader to indicate you've done so and provide some rationale for this decision.*

We have developed our own metadata standard appropriate to the nature of the project. The description of our metadata standards, which also provides contextual information regarding how the tables are linked, is provided as an Appendix to this document.

### Storage and Backup

**What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?**

Currently our project programmer is maintaining the server space we have with SFU's<sup>5</sup> Dean of Arts. We will migrate the server to SFU Library, having arranged for appropriate server capacity earlier. The current codebase of the Women's Print History Project is approximately 1.285 GB of disk space--this includes copies of the source data files contributed by ESTC and Osborne.

<sup>5</sup> Web-service stack, named for Linux, Apache, MySQL, PHP/Perl/Python



## **Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)**

The database requires just under 1 GB of disk space.

The system resources of the current server:

CPU: 2XIntel(R) Xeon(R) CPU E5-4650 0 @ 2.70GHz

RAM: 4GB

Disk:>250GB

### **How and where will your data be stored and backed up during your research project?**

Data is currently stored on the server space maintained by our project programmer, and then in the SFU Library server, post- migration.

### **How will the research team and other collaborators access, modify, and contribute data throughout the project?**

Our programmer has built a data-entry web interface on a LAMP<sup>6</sup> platform.

As the server access is controlled by an SFU-administered access list, people may be added or removed on request. There is no need to share server administration passwords as full access is available to anyone on the list. The major barrier is Linux administration knowledge. Website access passwords (wphp/WPHP...) may be regenerated on the command line if lost or forgotten. Passwords may also be stored in previous emails to you and others in the course of discussing the website.

For general storage of administrative information, the Canadian instance of OSF<sup>7</sup> will ensure complete redundancy and allow for version control as files are updated.

<sup>6</sup> Web-service stack, named for Linux, Apache, MySQL, PHP/Perl/Python

<sup>7</sup> An open-source collaborative data sharing platform created by the Center for Open Science



## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

### Preservation

#### Where will you deposit your data for long-term preservation and access at the end of your research project?

At the end of the project we will deposit our data in SFU's research data repository, Radar, as .csv files with a data map to show our table structure. We are also planning to make all of our data public via a web interface currently being developed by SFU's Digital Humanities Innovation Lab.

Radar is supported by Islandora, an open-source repository platform. Data deposited in Radar receive unique identifiers (DOIs), searchable subject headings and keywords, relevant licensing information and other applicable metadata.

**Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.**

*It is useful to indicate clearly what standards you're referencing (institutional, gov. principles, etc.) and mention where the supporting documentation will be and in what format. If you're unsure of applicable standards at the time of creating your DMP, indicate how you intend to align your data management with best practices moving forward.*

We will use best standards, i.e. csv files plus data map as it appears Radar does not take MYSQL databases. Consultation with librarians or information professionals will ensure our adherence to any applicable preservation guidelines and best practices.

### Sharing and Reuse

#### What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

We will be sharing final data, according to your descriptions, i.e. regularized data that users can convert to csv.



## **Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)**

### **Have you considered what type of end-user license to include with your data?**

We have acknowledged the source providers of our bibliographical metadata, which we will make freely available. Sources are included/acknowledged in our data model.

### **What steps will be taken to help the research community know that your data exists?**

*Describing how you intend to communicate about your data can help the reader to more concretely envision your plan for dissemination. This also helps to solidify your commitment to sharing your data.*

There will be a public-facing website, we will present the project at conferences, will cite our data in published analyses, and will send email announcements to relevant list-servs upon release.

## **Responsibilities and Resources**

### **Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.**

Jenn Ross, our programmer, is managing our data. After the project development stage concludes, it is anticipated that SFU will manage the data according to the terms of a hosting agreement currently being formulated.

### **How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?**

My project manager and PhD student will handle the project's data when I, as the current PI, am no longer able to do so.



## **Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)**

**What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?**

Currently updating and maintaining our data costs about \$100/month (two developer hours). I expect these costs to continue, and to be borne by SFU Library if a hosting agreement is reached. I am also investigating other sources of ongoing funding.

### **Ethics and Legal Compliance**

**If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?**

N/A

**If applicable, what strategies will you undertake to address secondary uses of sensitive data?**

N/A

**How will you manage legal, ethical, and intellectual property issues?**

N/A





## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

### Appendix

#### Title Table

Field	Description
<b>Title ID</b>	Unique ID number for this entry
<b>Title</b>	Full title as it appears on the title page, including subtitle, signed author, and edition statement where applicable
<b>Attribution</b>	Linked to the <b>Persons Table</b> . Attribution roles include name of author(s) (if known) and, where appropriate, editors, translators, authors of introductory material, engravers, and women involved in printing, publishing, bookselling, etc. If any of the roles are attributed to a woman whose name is unknown, they will be listed as “Unknown.”
<b>Signed Author</b>	Author attribution as it appears on the title page or at the end of the preface. (Ex. “By a lady,” “By the author of . . . ”)
<b>Pseudonym</b>	If the author has signed the work with a name that has been proven false, it will be listed here.
<b>Firm</b>	Linked to the <b>Firm Table</b> . Includes all printers, publishers, and booksellers listed in the imprint and colophon.
<b>Volumes</b>	Number of volumes that the edition was published in, indicated in Arabic numerals.
<b>Pagination</b>	Pagination of each volume. Each volume indicated by a capital letter roman numeral, prefatory matter indicated by lowercase roman numerals, regular pagination indicated in Arabic numerals.
<b>Publication Date</b>	Date (year) as it appears in the imprint. Where imprint information is unavailable, this will reflect the best estimate we have.



## Data Management Plan Exemplar #1: Digital Humanities

### Women's Print History Project (1750- 1830)

<b>Edition</b>	Edition as it appears in the edition statement on the title page; 1st editions are not indicated on title pages so in the case of an earliest known printing 1st edition is assumed.
<b>Date of First Publication</b>	Date (year) that the work was first published.
<b>Imprint</b>	Information about printers, publishers, booksellers as represented on the title page
<b>Location of Printing</b>	Geotagged location as indicated by the imprint
<b>Format</b>	The way that sheets have been folded and gathered into pages. For example, folio, quarto, octavo, duodecimo, sextodecimo, octodecimo, vicesimo-quarto, etc.
<b>Size</b>	Length and Width measured in cm
<b>Price</b>	Price as listed on the title page or in catalogues and periodicals expressed in pounds, shillings, and pence
<b>Genre</b>	Category that the work falls into. [Link to full list of genres?]
<b>Source/Source ID</b>	Sources consulted to populate the entry fields. Source ID will vary according to the source, but will correspond to the
<b>Shelfmark</b>	Call numbers for location in various libraries.
<b>Notes</b>	Any important information that does not fit in the above fields will be indicated here, as well as links to sources that do not have a unique identifying number.
<b>Hand Check</b>	Indicates that a physical copy of the text has been consulted.
<b>Final Check</b>	Indicates that either two sources have been consulted or the text has been hand-checked.



## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

### Attribution/Person Roles

Role	Description
<b>Author</b>	The person responsible for the composition of the work. See Foucault.
<b>Publisher</b>	Include if the firm is female-run. Indicates the person running the firm for whom the work was printed.
<b>Bookseller</b>	Include if the firm is female-run. Indicates the person running the firm that sold the work.
<b>Printer</b>	Include if the firm is female-run. Indicates the person running the firm that printed the work.
<b>Editor</b>	The person who selected, revised, and arranged the work for print.
<b>Translator</b>	The person who is responsible for translating the text from one language to another.
<b>Engraver</b>	The person responsible for turning an illustration into an engraving.
<b>Introduction</b>	The person who composed the introduction or any prefatory material for the work. Must be a different person than the author.
<b>Illustrator</b>	The person who created the image from which any engravings were created. Usually different than the Engraver.
<b>Compiler</b>	The person who selected the works contained within a text. Usually applies to anthologies and collections.
<b>Composer</b>	The person responsible for setting a piece to music.
<b>Performer</b>	

### Firm Roles

Role	Description



## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

<b>Printer</b>	The firm responsible for compositing and printing the work. Usually indicated by the phrase “printed by” in the imprint and/or the colophon.
<b>Publisher</b>	The firm that takes on the financial responsibility for the editing, printing, and distribution of the work. Usually indicated by the phrase “printed for” in the imprint. In the event that a work has been self-published, this will be indicated in the “Self-Published field.”
<b>Bookseller</b>	The firm from which the work can be purchased. Indicated by the phrase “sold by” in the imprint.
<b>Unknown</b>	No firm role indicated in the imprint.

### Persons Table

Field	Description
<b>Person ID</b>	Unique ID number for this entry
<b>Last name</b>	Person’s last name (preference will be given to the name by which he or she is most commonly known)
<b>First Name</b>	Person’s first name
<b>Title</b>	Where appropriate, full aristocratic title
<b>Gender</b>	Either M or F
<b>Date of Birth</b>	If known, YYYY-MM-DD
<b>Date of Death</b>	If known, YYYY-MM-DD
<b>City of Birth</b>	If known; geotagged



## Data Management Plan Exemplar #1: Digital Humanities Women's Print History Project (1750- 1830)

<b>City of Death</b>	If known; geotagged
----------------------	---------------------

### Firm Table

Field	Description
<b>Name</b>	Most complete name of the firm known
<b>Street Address</b>	Street address of the firm, if known. Each time a firm moves to a new location, a new firm entry is made. Addresses usually drawn from imprints. Geotagged.
<b>City</b>	City/town/village in which the firm's street address is located; geotagged
<b>Start Date</b>	If known, YYYY-MM-DD
<b>End Date</b>	If known, YYYY-MM-DD

