L'elaboració del Pla de Gestió de Dades.
L'experiència de tres projectes de la UAB

24 Octubre 2022

WINK

Women's Invisible Ink:

Trans-Genre Writing and the Gendering of Intellectual Value in Early Modernity
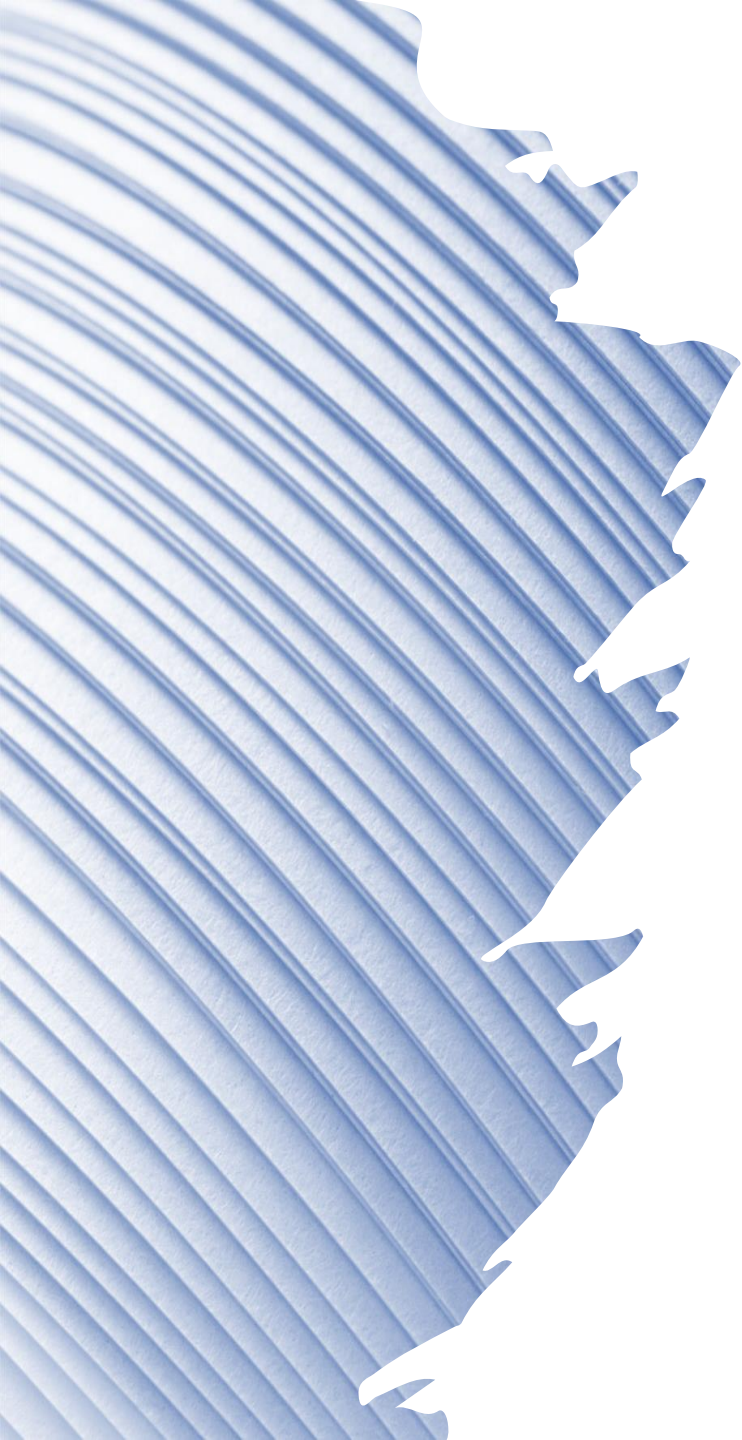
Carme Font Paz

Universitat Autònoma de Barcelona

https://www.projectwink.eu/

Carme.Font@uab.cat

UAB
Universitat Autònoma de Barcelona

- WINK locates, identifies and examines neglected written production of women in early European modernity in order to modify the single-gender paradigm of intellectual value.

- WINK suggests a critical and methodological approach to the study of intellectual value as a category of socio-cultural, literary, and gender analysis, de-problematizing how and why equal-value can be produced differently.
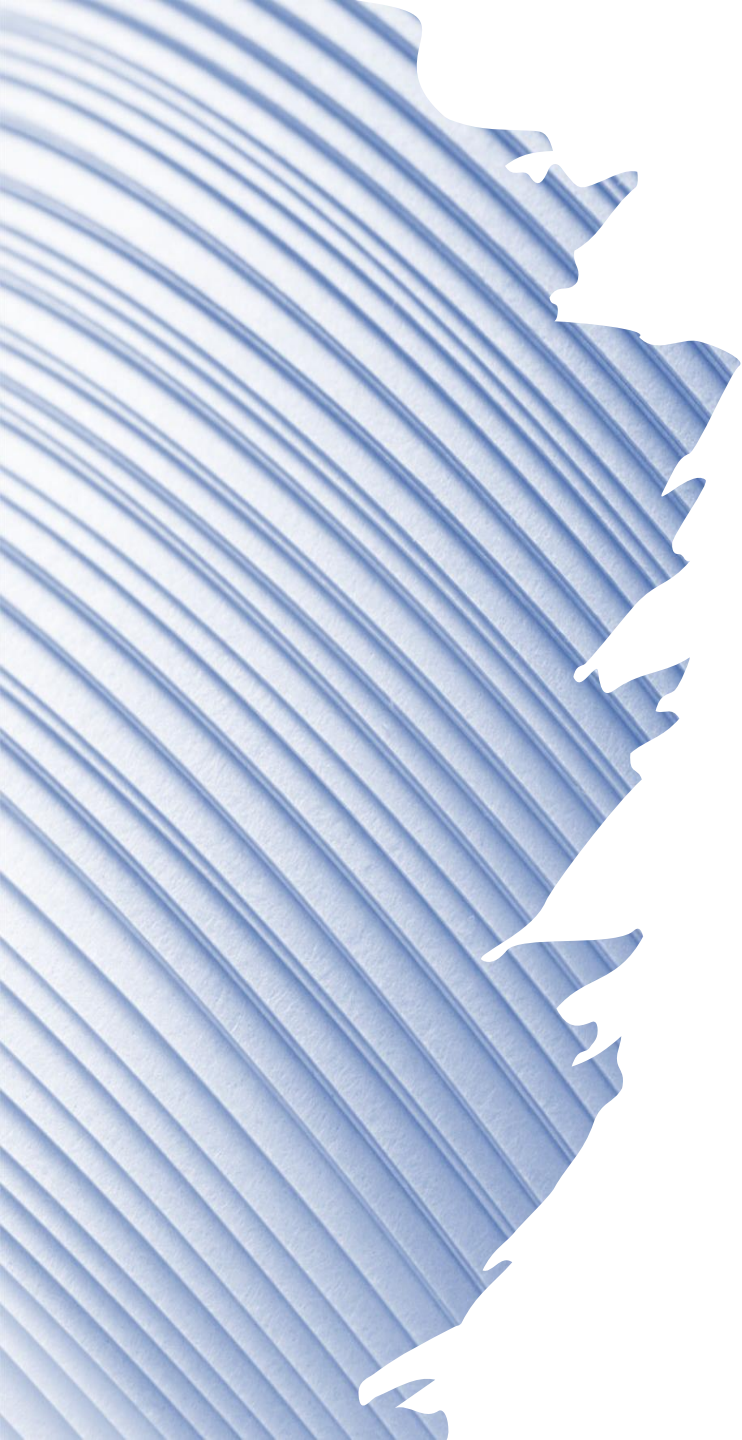
- The project surveys sources in five languages through a methodology based on trans-genre writing rather than on close genre types, allowing patterns of persuasive argumentation to emerge as intellectual input, while exposing the rhetorical models that have impinged on the social and cognitive processes identifying intellectual value as being androcentric.

- It examines textual and stylistic factors that have promoted the invisibility of intellectual value in women's writing across genres from the early modern period, a process of "male gendering" of epistemological paradigms that the project defines as "textual misogyny".

- The emphasis is on how the specifics of gender modify and inform the system of value production in intellectual thinking, instead of studying gender as opposing this system.
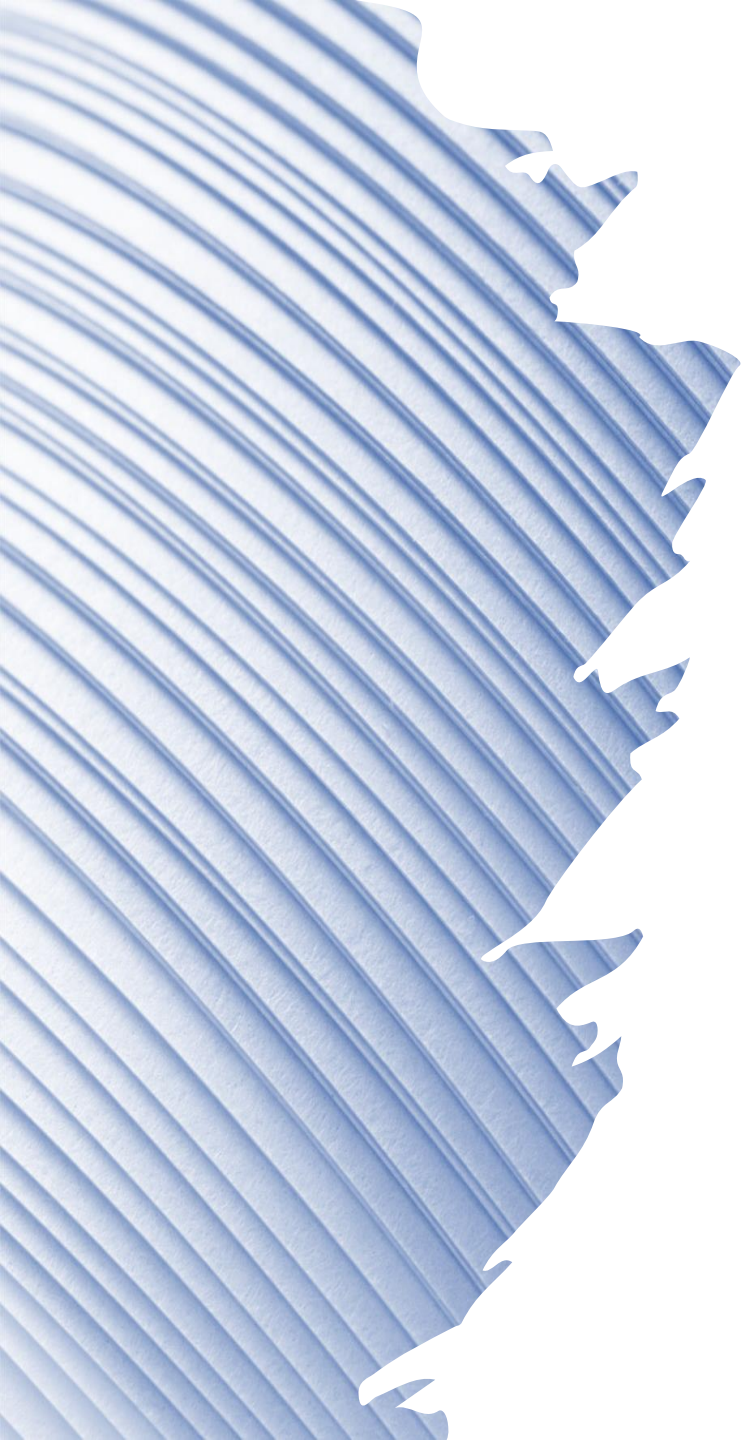
- Extract Value → see whether neglected textualities respond well to trans-genre analysis and provide evidence of sustained valuable thought (intellectual value).

- Trans-genre methodology allows visibility of intellectual value: taking all sources by a single woman author, her authorial ecology and analyzing them beyond the borders of genre and text.

- WINK approaches intellectual value as a category of gender analysis, bringing to light transformative thinking from understudied and underrepresented women authors.

Un pla de gestió de dades no és tràmit burocàtic ni un objectiu en sí mateix, però ens permet accedir i descobrir coneixement, integrar-lo als resultats del projecte, i permetre que aquest pugui ser consultat i reutilitzat per la comunitat després del procés de publicació.

+ **_Project data should be Findable, Accessible, Interoperable and Reusable (FAIR)._**

Open Research Data and Data Management Plans, ERC, August 2021.

Data should be:

--identified in a persistent manner using community conventions, and described using sufficiently rich metadata;

--stored in such a way that they can be accessed by humans and machines;

--structured in such a way that they can be combined with other datasets;

--licensed or having terms-of-use that spell out how they can be used by others.

A DMP should provide information on:

+ 1. *Dataset description:*

Grantees should provide a sufficiently detailed description, including the scientific focus and technical approach, to allow association of their datasets and derived data products with specific research themes.

+ 2. *Standards and metadata:*

Grantees should describe the protocols used to structure their data and indicate the metadata standards applied. This will allow other scientists to make an assessment, to attempt to reproduce the conclusions derived from the dataset (and possibly even the dataset itself), and potentially reuse the data for further research. If available, grantees should provide a reference to the community data standards with which their data conform and that make them interoperable with other datasets of similar type.

3. *Name and persistent identifier for the datasets:*

Grantees should plan to use repositories that will provide a unique and persistent identification (an identifier) of their datasets and derived data products, and a stable resolvable link to where they (or, as a minimum, their metadata) can be directly accessed.

4. *Curation and preservation methodology:*

Grantees should provide information on the standards that will be used to ensure the integrity of their datasets, and the period during which they will be maintained. Grantees should also explain whether and how their datasets will be preserved and kept accessible in the longer term. If applicable, they should detail the criteria for prioritisation, appraisal and selection of the datasets to be retained. If raw data cannot be stored (e.g. because they are too large or modified in (quasi-)real-time), grantees should describe what data products will be derived, and how these will be preserved and kept accessible. If available, grantees should provide a reference to the public data repository in which their datasets or data products will reside.

5. *Data sharing methodology*

Grantees should provide information on how their datasets and/or data products can be accessed, including the terms-of-use or the license under which they can be accessed and re-used, and information on any restrictions that may apply. It is also important to specify and justify the timing of data sharing. This could be, for example, as soon as possible after the data collection, or at the end of the project. For data that underlie publications it could be, for example, at the time of publication or prepublication.

Characteristic features of the disciplines that together make up the ERC's SH domain is their variety, in terms of topics, epistemologies, and methodologies.

This is reflected also in the data that SH projects produce: quantitative datasets; experimental data; observational data; interviews; archival data; human artefacts; medical and genetic data; and so forth. In addition, the various kinds of data cross-cut the disciplinary divisions, as several disciplines produce different kinds of data, depending on the methodologies used.

Data may include copyrighted material, such as literary texts or images, or archival materials to which access is restricted.

A generally accepted minimum standard for describing information on the web, including research data, is Dublin Core. Further information on this metadata standard is available at: https://www.dublincore.org/

Data set	Identifier	Access
Books	Book_	Repository
Book Chapters	Book_Chapter_	Repository
Articles in journals	Journal_Name_	Repository
Transcripts of original texts (without philological analysis WINK)	Transcript_original_text_	Repository
Transcripts of original texts (with philological analysis WINK)	Transcript_original_text_with_analysis_	Repository
Transcripts of edited texts (without philological analysis WINK)	Transcript_edited_text_	Repository
Transcripts of published texts (with philological analysis WINK)	Transcript_published_text_with_analysis_	Repository
PhD theses	PhD_thesis_	Repository
Titles of seminars and conferences	Titles of seminars and conferences	Web
List of cataloged books from libraries and collections	Cataloged_Books_List_	Repository
List of editors	Editors_List_	Repository
List of authors / authors	Authors_List_	Repository
List of cataloged manuscripts from libraries and collections	Cataloged_Manuscripts_List_	Repository
List of non-cataloged manuscripts of private collections (with authorization)	Non_Cataloged_Manuscripts_List_	Repository
Free access images	Images_	Repository
Summaries of extracted information and conclusions	Summary_ Conclusions_	Repository
Registries of bibliographic data MARC 21	MARC21_Registries_	Repository
References to published biographical information of authors	References to published biographical information of autors	Web
Digital visualization of philological analysis WINK	Visualizations_	Repository
Keyword clusters	Keywords_	Repository
Instructional videos on philological analysis WINK	Videos_	Web
Data about upcoming activities	Upcoming activities	Web

Table 1: Data set description, identifier and storage location (i.e. project website - https://www.projectwink.eu/ - or institutional repository - https://ddd.uab.cat/ -).

Expected size of datasets?

Transcripts of original texts (without philological analysis WINK): Transcripts of original texts accessed under the WINK project without philological analysis. Persons interested in it: Group members, project partners and the general public. Type of research data: observational, experimental. Origin of the data: Group members, open access library repositories.

Keyword clusters. Persons interested in it: Group members, project partners and the general public. Type of research data: descriptive, observational. Origin of the data: Group members.

List of authors / authors: List of authors of books consulted during the WINK project. Persons interested in it: Group members, project partners and the general public. Type of research data: observational, experimental and interpretative. Origin of the data: library catalogues and repositories.

Summaries of extracted information and conclusions: Summaries of extracted information and conclusions of the WINK project. Persons interested in it: Group members, project partners and the general public. Type of research data: descriptive. Origin of the data: Group members.

FAIR Data and Resources

1. Making Data Findable

NAMING CONVENTIONS

Files will be named using the identifier presented in Table 1 and using the following convention:
<identifier>_<short_text>_v<version number>, where the short text will define the data. For example, author's names will use the name initial and the three first letters from surname: Transcript_original_text_ASMI_v1

DISCOVERABILITY OF DATA (METADATA PROVISION)

The metadata standard used to describe the dataset will be the Dublin Core Schema (DC), as it is a flexible and common used standard and is also the one adopted by the European OpenAIRE repository. If any extensions are needed the Data Document Initiative (DDI) will be included.

IDENTIFIABILITY OF DATA AND STANDARD IDENTIFICATION MECHANISM

The files will be stored in the institutional repository (Dipòsit Digital de Documents, DDD, https://ddd.uab.cat/) which provides a unique URL to access the document with the format https://ddd.uab.cat/record/XXXXXX where XXXXXX is the identification code. Or Zenodo: https://zenodo.org/record/5628463 - https://zenodo.org/record/5629199 - https://zenodo.org/record/5634285

DC Element	DDI Element	Notes
Title	\<titl\> 2.1.1.1	Title of Data Collection
Creator	\<AuthEnty\> 2.1.2.1	Authoring Entity of Data Collection
Subject	\<keyword\> 2.2.1.1	Keyword(s)
	\<topcClas\> 2.2.1.2	Topic Classification
Description	\<abstract\> 2.2.2	Abstract
Publisher	\<producer\> 2.1.3.1	Producer of Data Collection
Contributor	\<othId\> 2.1.2.2	Other Identification/Acknowledgements - Data Collection
Date	\<prodDate\> 2.1.3.3	Production Date - Data Collection
Type	\<dataKind\> 2.2.3.10	Kind of Data
Format	\<fileType\> 3.1.5	Type of File
Identifier	\<IDNo\> 2.1.1.5	ID Number - Data Collection
	\<holdings location="" callno="" URI=""\> 2.1.8	Holdings Information - Data Collection
Source	\<sources\> 2.3.1.8	Sources - Used for Data Collection
Language		
Relation	\<othrStdyMat\> 2.5	Other Study Description Materials
Coverage	\<timePrd\> 2.2.3.1	Time Period Covered
	\<collDate\> 2.2.3.2	Date(s) of Data Collection
	\<nation\> 2.2.3.3	Country
	\<geogCover\> 2.2.3.4	Geographic Coverage
Rights	\<copyright\> 2.1.3.2	Copyright - Data Collection

Table 2: Elements of the DC elements and its correspondence to the DDI elements (Source: http://www.ddialliance.org/resources/ddi-profiles/dc).

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences.

DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving.

2. Making Data Openly Accessible

DATA ACCESS PROVISION

Data will be openly shared in the DDD once it has been analyzed and published.
In case it is decided to temporarily restrict the access to a particular data set it will be possible to ask for individual access by contacting the principal investigator.

DATA ACCESSIBILITY

Data will be made accessible to other group members and project partners in open access, and to the general public (data sets on title, author, edition, location, document scans) when copyright permits.

SOFTWARE REQUIREMENTS TO ACCESS THE DATA

Data will be stored in CSV, TSV or TXT to ensure flexibility for data reuse.
CSV (comma separated values) files are used to store tabular data in plain text format. Most often the fields in this data are separated by commas but other delimiters can be used such as |. TSV (tab separated values) files are similar but breaks are delimited by tabs. Both formats are widely supported and are often used to exchange data across multiple different computers and systems that support the format.

Images will be shared in PNG or PDF format since they are commonly used formats with widely distributed visualization software.

3. Making Data Interoperable

INTEROPERABILITY OF DATA

The metadata standard used to describe the dataset will be the Dublin Core Schema, if any extension are needed the Data Document Initiative will be included. Where possible, standard codes will be followed, for example country identification will be registered following the ISO 3166-1-alpha-2 codes and language identification will use the ISO 639-2 code.

STANDARD VOCABULARY TO ALLOW INTER-DISCIPLINARY INTEROPERABILITY

The vocabulary used in the data will be as standard as possible to ensure its comprehension by non-experts in the field.

4. Increase Data Reusability

DATA AVAILABILITY FOR RE-USE

The data will remain re-usable after the end of the project by anyone interested in it, with no access or time restrictions for at least 10 years.

DATA AVAILABILITY TO THIRD PARTIES AFTER THE END OF THE PROJECT

Each archived data set will have its own permanent repository ID and will be easily accessible. Most of the data generated will be made available without restrictions and only data sets subject to IPR and confidentiality issues will be restricted. Where access to data is restricted, agreements will be made based on the individual data sets. Requests for the use of the data by externals will be approved by the project consortium.

DATA QUALITY ASSURANCE PROCESSES

The data quality will be ensured by different measures. These include validation of the documents through comparison and collation with similar documents from the same period, language, origin and community of authors.

DATA LICENSING

The deliverables associated to the dataset will be licensed through an *All rights reserved* license as they are working papers not intended to be re-used. Nevertheless, the database will be shared when possible as a reusable dataset. For this reason, an Attribution-NonCommercial license (By-Nc) will be used. The data will be available for re-use from the repository at the latest by the end of the project.

LENGTH OF TIME FOR WHICH THE DATA WILL REMAIN RE-USABLE

The data will be stored for re-usability on the repository for at least ten years.

5. Allocation of Resources and Data Security

The repository at UAB (DDD, https://ddd.uab.cat/) will provide free of charge the necessary resources to store the data files in an accessible and secure manner for at least ten years after the end of the project.

6. Ethical Aspects

No ethical aspects are expected to arise from this project.
A warning message will inform of any potential data in those works whose content might be offensive to readers by dint of its explicit violent or sexual content.
In the event of any other ethical issue arising, this would be dealt with by the principal investigator with the support of the ethical committee from her host institution, if necessary.