

# DATA MANAGEMENT PLAN

March 2026 - V2.0  
DOI: 10.5281/zenodo.19333379

## Mapping and Translating Spaces, Cultures and Languages. Experiences from the Missions connected to the Portuguese Empire (1540-1700)

Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN 2022)  
Ref.: 2022SY2K7

### PROJECT ABSTRACT

The project “Mapping and Translating Spaces, Cultures and Languages” (MAT) is a two-years [PRIN 2022](#) funded project (ref. 2022SY2K7) which combines the methods of history with those of linguistics and translation studies to promote an innovative interdisciplinary analysis of the processes of cultural (mis)communication and (mis)translation among communities across the Portuguese Empire and Royal Patronage between 1540 and 1700.

The project has three main objectives:

- 1) Draw up a **comprehensive analytical catalogue of overlooked, dispersed metalinguistic and multilingual sources** (reports, letters, Christian doctrines, maps, word lists, lexicons, grammars, which describe linguistic practices and/or display bilingual or three-lingual evidence) **produced mostly in missionary contexts and in large part held in Italian and Portuguese religious and state archives.**
- 2) **Study the emergence of multilingual communities in early modernity** involving cultures and languages that were previously unknown in Europe.
- 3) Within this broad “horizontal” survey, **highlight two case studies to carry out an in-depth “vertical” comparative analysis of cultural-linguistic contacts and translations in Sub-Saharan Africa and China**, specifically chosen because they are two paradigmatic, coeval but antithetical cases detailing the different shades of cultural translations in colonialism and missions.

The integration of two working strategies - the extensive mapping of intercultural multilingual sources and the analysis of two antithetical case studies - will allow to undertake a **comparative analysis of the processes related to the learning, imposing or rejection of cultures and languages in the “troubled pasts” of missionary and colonial contexts**. MAT connects historians, linguists and specialists of Chinese, Sub-Saharan African, Portuguese languages and Data Management, to **document the largest possible corpora of translations in early modernity** and offers **new ideas on the relevance of linguistic interactions and on our multicultural and multilingual “troubled present”**.

**Project ID:** ref. 20222SY2K7

**Start date:** 28-09-2023

**End date:** 27-09-2025

## DATA SUMMARY

Will you re-use any existing data and what will you re-use it for?

Data was generated and reused from both primary and secondary sources (manuscripts and printed works) from the 15th to 20th centuries. Most of these works are available in digitised form (.PDF) in open-access repositories. Relevant data sources have been digitised as needed.

The collected data has been ingested into the WebGIS platform to allow for interactive analysis and geospatial visualisation.

The data collected consists in:

- grammatical and lexical information
- linguistic descriptions
- cultural information
- metadata (authors and place names)
- visual data (maps, images)

Extracted data has been processed in various formats, including:

- software tools
- catalogues of works (in excel, or database)
- synoptic tables to compare metalanguages (in excel, or database)
- tables relating dates, places, languages, authors, provenance and editions (in excel, or database)
- maps (dynamic and interactive) illustrating the geospatial position of the works and languages described (image files)
- publications (monograph)

What types and formats of data will the project generate or re-use?

The project generated and/or reused the following data:

- Textual data: text documents which have been digitised and are available in PDF format. (narrative, qualitative, non-numerical data). Some documents have also been transcribed into OpenDocument Text (.odt).
- Visual data: raster images which have been digitised.
- Geospatial data: Geospatial data will be generated and later be available for extraction through the WebGIS platform. For now it is stored in a reserved-access database.
- Structured data: data on authors, outcomes, interactions, and Chinese toponyms will be generated and made available through the WebGIS. For now, the database allows only authorised users to extract data in a structured format.

Data comes in the following forms:

Text documents	<ul style="list-style-type: none"> <li>• .PDF</li> <li>• some may be transcribed in .ODT</li> </ul>
Raster images	<ul style="list-style-type: none"> <li>• .TIFF</li> <li>• .JPEG</li> </ul>
Geospatial data	Geospatial data is not yet available through the WebGIS platform. In the future, the data will be made available through the WebGIS in the following formats: GeoJSON.
Structured data	In the future, the structured data will be available in .xls format. Currently, it can be exported in .csv format through the database (for authorised users only).

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of the project is to **provide access to data on multilingual sources and practices in missionary contexts associated with the Portuguese Empire (1540–1700) through the WebGIS platform.**

The WebGIS functions as the primary interactive platform for the study forgotten metalinguistic and multilinguistic sources<sup>1</sup>, including maps, missionary reports and letters, religious texts, word lists, lexicons, and grammars, based on Portuguese as a translational contact language. This corpus represents a neglected world heritage resource of universal importance.

---

<sup>1</sup><https://mapping-translating.cnr.it/mappa.php>

As much as possible, data will be made openly available through the WebGIS. When their use is legally restricted to research purposes, they will not be available in open access.

The data is needed to map multilingual and metalinguistic sources in order to shed light on the intricate interplay between languages, cultures and spaces within the given framework.

Specifically, the data enables the project to:

1. Map places where contacts with non-European languages occurred and identify the fundamental aspects of these contacts.
2. Understand how knowledge exchange between Europe and other continents occurred, particularly through the mediation of the Portuguese language.
3. Make available information relating to the pedagogical and linguistic policies of the missionary religious orders.
4. Examine the way in which exchanges took place between the Portuguese and the Africans in Sub Saharan Africa and assess the data on African languages reported by the authors of grammars, lexicons and catechisms in African languages.
5. Create a corpus of toponyms with historical linguistic data from the Sino-Jesuit atlases, together with historical biographical information on the sources.

What is the expected size of the data that you intend to generate or re-use?

The following table provides an overview of the estimated sizes of the different types of data generated and reused by the project.

Text documents (PDF, ODT)	Each file on the WebGIS has a maximum size of 50 MB. The CNR server provides a total maximum storage capacity of 95 GB.
Raster images (TIFF, JPEG)	Each file on the WebGIS has a maximum size of 50 MB. The CNR server provides a total maximum storage capacity of 95 GB.
Geospatial data (GeoJSON)	The expected size will be known once the data is made available in the WebGIS.
Database (PostgreSQL)	Database file size varies depending on the number of records exported.
Structured data (.xls, .csv)	The size of the data depends on the size of the record exported.

The total estimated size of all data generated and reused by the project remains below the maximum capacity of the CNR server (95 GB).

What is the origin/provenance of the data, either generated or re-used?

The project has generated and reused data from manuscript sources and historical materials which have been digitised and made available in open access repositories and archives. These include grammars, vocabularies, letters, travel reports, historiographical works, and Christian doctrines for the catechesis of non-European peoples.

Sources of the data include:

- Public libraries, such as the Biblioteca Nacional de Portugal, Arquivo digital, Centro de História, Faculdade de Letras, Lisbon University.
  - Example: <https://bndigital.bnportugal.gov.pt/records/item/90937-doutrina-christiana>
  - Example: Monumenta Missionaria Africana  
[https://repositorio.ulisboa.pt/browse/author?value=Br%C3%A1sia,%20Ant%C3%B3nio%20\(org.\)](https://repositorio.ulisboa.pt/browse/author?value=Br%C3%A1sia,%20Ant%C3%B3nio%20(org.))
- Private libraries, such as the digital repository of the Archivum Romanum Societatis Iesu (ARSI).
  - Example: a selection of Monumenta Historica Societatis Iesu (MHSI) volumes available in open access as PDFs: <https://arsi.jesuits.global/arsi-digitale/monumenta-historica-societatis-iesu/>
- Publishers who have made specific works available in open access
  - Example: Peter Lang: <https://www.peterlang.com/document/1190560>

Data has also been collected from manuscript sources from the mendicant orders, in particular Franciscans, Dominicans, Augustinians, Carmelites, Mercedarians, Trinitarians and Theatines.

Finally, maps and a wooden globe have been included as data sources.

To whom might your data be useful ('data utility'), outside your project?

The data will be most useful to scholarly communities connected to:

- History (early modern history, history of the missions, connected and global history),
- Linguistics (Diachronic Linguistics, Portuguese Linguistic, History of Linguistic, Missionary Linguistic),
- African, Chinese and Portuguese Studies
- History of the Missions and Religious orders
- Doctoral students working on the topics above

The data will also be useful to stakeholders beyond academia, including schools and museums, through specific dissemination activities. Specifically, the project has signed an agreement with the Museum Casa delle Lingue to contribute to its work.

The WebGIS platform will allow future access to the data for additional users and communities.

## FAIR DATA

2.1. Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?

Yes, persistent identifiers will be used, in particular DOI for data and project outputs, and ORCID for project members.

2.1. Making data findable, including provisions for metadata: Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The metadata that will be provided will include: Author, title, date, place of writing or publication, place(s) to which the document refers, language used in writing, languages referred to or described in the document, document location, document format, document genre.

2.1. Making data findable, including provisions for metadata: Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Yes, search keywords will be included in the metadata using controlled vocabularies relevant to the Digital Humanities. Keywords appear in the subject and keywords fields following a standard Metadata Schema. This approach ensures datasets are easily discoverable by both humans and machines, supporting potential reuse.

2.1. Making data findable, including provisions for metadata: Will metadata be offered in such a way that it can be harvested and indexed?

Yes, metadata will be provided in a machine-readable format following a standard Metadata Schema. Metadata will be deposited in a certified repository and made available via OAI-PMH endpoints to enable harvesting and indexing by search engines and research catalogs. All datasets and metadata are assigned DOIs to ensure persistent identification and improve discoverability.

2.2. Making data accessible - Repository: Will the data be deposited in a trusted repository?

Yes, data will be deposited in the trusted repository Zenodo.



2.2. Making data accessible - Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?

The policies of the Zenodo<sup>2</sup> repository suit our needs. Zenodo ensures long-term preservation of deposited items for the lifetime of the repository (currently at least 20 years) and assigns persistent identifiers.

2.2. Making data accessible - Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes. Data deposited in Zenodo are assigned a persistent DOI that resolves to a landing page providing access to the digital object and its metadata.

2.2. Making data accessible - Data: Will all data be made openly available?

As much as possible, the data generated by our work will be made openly available. Some datasets, however, may be legally restricted to research purposes and cannot be shared openly.

This is the case, for example, with certain facsimile reproductions of documents when the holding archives do not permit open distribution. We will, however, seek to negotiate with these institutions to enable broader access whenever possible. In all cases where data access is restricted, the corresponding metadata will be openly available through the repository to ensure discoverability.

2.2. Making data accessible - Data: If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

At this time, there are no plans to apply any kind of embargo.

2.2. Making data accessible - Data: Will the data be accessible through a free and standardized access protocol?

Yes, the data are accessible through the standard open web protocol (HTTPS) via Zenodo.

2.2. Making data accessible - Data: If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Where restrictions on use arise for legal reasons, only the metadata will be made available through the repository.

---

<sup>2</sup> <https://about.zenodo.org/policies/>

2.2. Making data accessible - Data: How will the identity of the person accessing the data be ascertained?

For the data that is openly available, it is not necessary to ascertain the identity of the person accessing the data, as it is openly accessible via Zenodo without authentication requirements.

2.2. Making data accessible - Data: Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

A data access committee is not required, as the project data are made available in open access via Zenodo and do not include personal or sensitive data.

2.2. Making data accessible - Metadata: Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Metadata will be made openly available, under a CC0 license.

2.2. Making data accessible - Metadata: How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data and related metadata will be stored in Zenodo, which guarantees long-term availability.

2.2. Making data accessible - Metadata: Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The project provides the WebGIS platform as a platform providing open access to data for analysis and visualisation. The raw data is available in open formats that do not require proprietary software.

2.3. Making data interoperable: What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

The intention is to use shared and widely used vocabularies in our sectors, they will be identified in a successive phase of the project.



2.3. Making data interoperable: In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Yes. In the eventuality that the project uses uncommon or creates specific ontologies or vocabularies, mappings to more commonly used ontologies will be provided and made openly available so as to allow for their reuse, refinement or extension.

2.3. Making data interoperable: Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

[1]A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

Yes, the data will include qualified references to other data.

2.4. Increase data re-use: How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

The documentation needed to validate data analysis and facilitate data re-use will be provided using README files.

2.4. Increase data re-use: Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes, for data that is freely accessible, we will use open licenses which allow the widest possible reuse, such as Creative Commons licenses.

Data not made available in open access is due to legal restrictions, typically when the holding archives do not permit open distribution and limit use to research purposes.

2.4. Increase data re-use: Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes, the data that is openly available will be usable by third parties.

2.4. Increase data re-use: Will the provenance of the data be thoroughly documented using the appropriate standards?

Yes, the provenance of data will be documented using the appropriate standards.

2.4. Increase data re-use: Describe all relevant data quality assurance processes.

We aim to clearly define data quality criteria and document quality controls

## OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Outputs generated by the project will include:

- WebGIS platform: a digital interactive open-access catalog, containing an atlas and a timeline of documented and spatialized cultural and linguistic contacts between 1540-1700, made available via (<https://mapping-translating.cnr.it/>) in open formats. Total size: maximum 95 GB (maximum size the CNR server can host)
- Scholarly publications data:
  - Book of abstracts are deposited in Zenodo, under a Creative Commons Attribution 4.0 International license (Accessible at: <https://doi.org/10.5281/zenodo.12571707> (2024), <https://doi.org/10.5281/zenodo.14655808> (2025), <https://doi.org/10.5281/zenodo.18255435> (2026)).
  - Scientific publications will be deposited in Zenodo, under an open license.
  - A scientific journal ("Status Quaestionis") will be deposited as a collection of articles in Zenodo, under an open license.
- Events (workshops): the videos will be made available on Youtube in the institutional channel (<https://www.youtube.com/@ufficiostudieuropei4857>) and embedded in the project website.
- Project website, accessible at: <https://sites.google.com/uniroma1.it/mapping-translating-prin22-cnr>

All outputs are accompanied by appropriate metadata or documentation to enable reuse, validation, and long-term discoverability.

## ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

The use of Zenodo guarantees free storage and archiving. In the eventuality costs related to making data FAIR arised, the project will ensure that these are covered. They should cover the data storage, archiving, re-use and security of data.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

For the first two years, the costs were covered by PRIN funds.

For the following years, the members of the research group are identifying further funding and financing to give continuity to the project.

Who will be responsible for data management in your project?

The following team members were responsible for data management in the project:

Lottie Provost, CNR-ILC, [lottiemiaprovost@cnr.it](mailto:lottiemiaprovost@cnr.it)

Francesca Di Donato, CNR-ILC, [francesca.didonato@cnr.it](mailto:francesca.didonato@cnr.it)

Angelo Cattaneo, CNR-ISEM, [angelomaria.cattaneo@cnr.it](mailto:angelomaria.cattaneo@cnr.it)

Simone Celani, Sapienza Università di Roma, [simone.celani@uniroma1.it](mailto:simone.celani@uniroma1.it)

Giulia Maggiore, Sapienza Università di Roma, [giulia.maggiore@uniroma1.it](mailto:giulia.maggiore@uniroma1.it)

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Data preservation and site maintenance are ensured by the institutions involved (Sapienza and CNR).

The use of the Zenodo repository to deposit data will ensure long term preservation.

## DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

The use of a certified repository such as Zenodo will enable us to ensure that provisions are in place regarding data security.

In addition, the WebGIS server (hosted by CNR/Sapienza) follows institutional protocols for secure transfer (HTTPS).

Is the data safely stored in certified repositories for long term preservation and curation?

Yes the data will be safely stored in certified repositories for long term preservation and curation.

## ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

We do not foresee any ethical issues related to the data. However, some datasets are subject to known legal restrictions due to ownership by private and public archives and libraries. These restrictions affect digital reproductions of certain sources, which cannot be shared openly without permission. We aim to negotiate conditions with each holding institution to clarify the terms under which these datasets can be shared.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

We ensured informed consent for data sharing and long term preservation were included in questionnaires dealing with personal data.