
The art of code-switching: Latin and Greek in Erasmus' (net)works

A Data Management Plan created using DMPonline.be

Creator: Raf Van Rooy

Affiliation: KU Leuven (KUL)

Funder: Bijzonder Onderzoeksfonds

Template: KU Leuven BOF-IOF

Grant number / URL: <https://research.kuleuven.be/portal/en/project/3H230275>

ID: 205971

Start date: 01-10-2023

End date: 30-09-2027

Project abstract:

This project studies the phenomenon of classical bilingualism in Erasmus' (ca. 1469-1536) works and networks. Classical bilingualism involves the competence of actively speaking and writing both Classical Latin and Ancient Greek. This project argues that the great age of written classical bilingualism should be situated in the so-called Renaissance rather than in antiquity. To this end, it compares Erasmus' classical bilingualism with that of Cicero. This comparison requires the creation of a corpus with full-texts of Erasmus' works, including his correspondence. In order to efficiently analyze the linguistic make-up of texts, the project develops the novel 'tongueprint' pipeline. In view of gaining insight into the thematic associations of Latin and Greek, topic modeling experiments are conducted. The results of these automated analyses are checked through a textual microanalysis and supplemented with a social network analysis of Erasmus' classical bilingual networks. The project will lead to a new text corpus, innovative digital tools, and a wide-ranging synthesis of Erasmus' classical bilingualism.

Last modified: 26-03-2024

The art of code-switching: Latin and Greek in Erasmus' (net)works

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		<i>Indicate: N(ew data) or E(xisting data)</i>	<i>Indicate: D(igital) or P(hysical)</i>	Indicate: Audiovisual Images Sound Numerical Textual Model SOftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Erasmus corpus / 1	OA images (out of copyright) / transcriptions of text	E / N	D	I / T	PDF / TXT and others	<100GB	NA
Transkribus models / 2	New Transkribus layout model and further trained OCR model	N (E)	D	M (AI)	OA Transkribus models	<100GB	NA
Topic model / 3	Classical bilingual topic model	N	D	M (AI)	Python code	<100GB	NA
'Tongueprint' pipeline / 4	New 'tongueprint' pipeline to recognize linguistic make-up of texts	N	D	SO (AI)	Python code	<100GB	NA
Statistical data tongueprint / 5	New 'tongueprint' pipeline statistical data	N	D	N	CSV	<1GB	NA
FileMaker database / 6	FileMaker relational database to organize our data	N / E (building on existing DaLeT)	D	T / N / I	FMP12 (many export possibilities)	<100GB	NA
Publications / 7	Publications resulting from C1-project	N	D / P	T / N / I	PDF / EPUB	<1GB	couple of kilos / 2 dissertations, 2 handfuls of papers, 1 conference proceedings
Networks / 8	Social network analysis images	N	D	N / I	JPG and other export possibilities from Gephi / NodeGoat	<1GB	NA

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Google Books PDFs, collected on OneDrive:

https://kuleuven-my.sharepoint.com/:f/g/personal/raf_vanrooy_kuleuven_be/EtrY1X4uhF9Ok5r9xiDSw2sBuEHWeBUF-MqAwf_YIcqK2A?e=fyHepu

NOSCEMUS GM 6.0:

<https://readcoop.eu/model/print-latin-texts-15-th-19th-htr/>

DaLeT:

<https://zenodo.org/doi/10.5281/zenodo.8305741>

<https://www.dalet.be/>

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

Due to a collaboration with publisher Brepols, the texts will be licensed OA but non-commercial and non-derivative.

CC BY-NC-ND 4.0

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

There was with our initial set of PDFs from the modern ASD edition (20th-21st centuries, rights owned by Brill), even though this is shaky at best: it concerns the edited text of Erasmus, but just to make sure we have taken text out of copyright.

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

1: 'About' section of DaLeT (<https://www.dalet.be/about.php>)

2: OA availability in Transkribus

3 & 4: Jupyter Notebooks

5: In publications and on DaLeT website

6: To be converted to a website through MYSQL if sufficient funding acquired, otherwise to be published on Zenodo with README.txt files

7: Open access publication as much and as soon as possible

8: In publication and on the DaLeT website

Will a metadata standard be used to make it easier to find and reuse the data?

If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

We will create our own metadata and link them as much as possible to existing databases, including the USTC (Universal Short Title Catalogue), Wikidata, and GeoNames.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- ManGO

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

Dropbox may be used for personal back-ups.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

By controlling carefully who accesses the (mostly historical) data. This will be a bit of a challenge for our citizen science plans, but Transkribus and FileMaker -- the platforms we use for this -- have options to monitor the contributions of the public. We also train our student assistants to handle the data carefully.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

There are no costs: the OneDrive space provided suffices. I pay for my secure Dropbox backup personally. Part of the data (OCR and lay-out models) are stored on the Transkribus servers, which I have paid for already with project funding, as budgeted.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- KU Leuven RDR
- Other (specify below)

Zenodo; Github; on www.dalet.be; the most important data also on Dropbox.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

There are probably no costs for data preservation, as it concerns relatively limited data sizes. In the improbable event we need funding for data preservation, I can apply for money with the Latin Literature department or the Seminarium Philologiae Humanisticae.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Yes, as open data

But with an NC-ND qualification, probably toward the end or after the project.

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

Zenodo / Github (see above)

When will the data be made available?

- Upon publication of research results
- Other (specify below)

Toward the end of just after the project (see above).

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

But NC-ND

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository
- Yes, my dataset already has a PID

Look at DaLeT practices for examples:

<https://www.dalet.be/course/3>

What are the expected costs for data sharing? How will these costs be covered?

Data shared through the DaLeT website will be paid for by project funding (in this project and another FWO project).

Responsibilities

Who will manage data documentation and metadata during the research project?

The PI Raf Van Rooy together with the PhD students Manou Vermeire and Mariia Timoshchuk.

For advice we consult co-PI Toon Van Houdt.

Who will manage data storage and backup during the research project?

The PI Raf Van Rooy together with the PhD students Manou Vermeire and Mariia Timoshchuk.

Who will manage data preservation and sharing?

The PI Raf Van Rooy.

Who will update and implement this DMP?

The PI Raf Van Rooy.