

Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: <https://doi.org/10.48321/D1H62F>

Title: eMunch meets Linked Open Data. Extracting, Enriching, and Exchanging Correspondence Metadata

Creator: Annika Rockenberger - **ORCID:** [0000-0001-9515-8262](https://orcid.org/0000-0001-9515-8262)

Affiliation: University of Oslo (uio.no)

Principal Investigator: Annika Rockenberger

Data Manager: Hilde Bøe, Loke Sjølie

Funder: University Of Oslo, Teksthub+digital Humanities

Grant: <https://www.uio.no/tjenester/it/forskning/kompetansehuber/teksthub/utviklingsprosjekter/index.html#toc14>

Template: Digital Curation Centre

Project abstract:

Pilotprosjektet skal undersøke og praktisk gjennomføre i hvorvidt det kan automatisk hentes ut (extracting) metadata som beskriver korrespondanser (brev, telegram, postkort) fra en digital utgivelse laget i XML/TEI P5. eMunch har publisert alle kjente brev fra og til Edvard Munch som del av Munchs tekster. Her kan en søke etter enkle brev, lese dem, studere faksimile og søke i selve brevteksten. Men for å kontekstualisere Munch som brevskriver og -mottaker er det nødvendig å koble sammen hans korrespondanse med andres, dvs. å knytte korrespondansemetadata fra eMunch til andre utgivelser av brev og korrespondanser, ikke bare i Norge og Norden, men i Europa og den internasjonale verden. Utgivelsens brevmetadata skal berikes (enriching) med koblinger til Linked Open Data-entiteter som Virtual International Authority File og GeoNames. Det lages dermed en ny XML-fil i Correspondence Metadata Interchange Format (CMIF) som er egnet til å integrere og tilgjengeliggjøre (exchanging) eMunch i den internasjonale søke- og analysetjenesten CorrespSearch.

Start date: 03-27-2023

End date: 01-27-2024

Last modified: 07-08-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

eMunch meets Linked Open Data. Extracting, Enriching, and Exchanging Correspondence Metadata

Data Collection

What data will you collect or create?

Metadata from letters in the published digital scholarly edition of Edvard Munch's Writings, eMunch.no. Metadata includes names of senders and receivers, a VIAF reference or Wikidata reference (if it exists), date(s) of sending/receiving the letter, and place of sending/receiving the letter, including a geoNames reference (if it exists). A general reference to the URL of the letter in the digital edition.

The file format is XML, according to the standard CMIF (Correspondence Metadata Interchange Format), based on the TEI P5 standard (Text Encoding Initiative, P5).

CMIF documentation: <https://github.com/TEI-Correspondence-SIG/CMIF/blob/main/doc/documentation-en.md>

TEI P5 documentation: <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

How will the data be collected or created?

Data will be collected using a script developed specifically for the project. Code for the script on GitHub: <https://github.com/norkorr/MXMLM>, developed by Loke Sjølie.

Data is extracted from the eMunch.no digital scholarly edition and harmonized with updated data on selected letters from an unpublished spreadsheet in the Munch Museum.

Documentation and Metadata

What documentation and metadata will accompany the data?

Documentation of the script (README file).

Documentation of the standards used for the data, CMIF and TEI P5 in their published versions.

A project README file.

Project metadata.

Ethics and Legal Compliance

How will you manage any ethical issues?

The data collected (and created) is so-called "green" data, i.e. no personal data or data of financial or other interest. The collaborating parties (Munch Museum and the University of Oslo Library) have agreed to use and share the data and their subsequent sharing with the CorrespSearch service at the Berlin Brandenburg Academy of Sciences and archiving in the dataverse.no data repository.

Access to the data in the CorrespSearch service is mid-term, depending on the Berlin Brandenburg Academy of Sciences funding situation.

Access to archived data on dataverse.no is long-term (5 years), ensured by the University of Oslo Library.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The metadata used from the eMunch.no digital scholarly edition is public domain.

The collected/created correspondence metadata for the project will be a Creative Commons public domain dedication / CC0 license.

The script for creating the data is licensed under an MIT license.

Publications related to the project will be CC BY 4.0 where possible.

Storage and Backup

How will the data be stored and backed up during the research?

Script/code is stored on GitHub.

Data from eMunch.no is stored by Munch Museum (specifications needed!).

Munch Museum stores spreadsheet data.

Project documentation is stored on a public GitHub repository: https://github.com/norkorr/edvard_munchs_tekster.

Article and other drafts are stored in a UiO-approved G-Suite folder, with restricted access to only the PI and the Data Manager.

References and bibliography are kept in a public, closed Zotero group library:
<https://www.zotero.org/groups/2214573/norkorr>.

How will you manage access and security?

Access to unpublished data material is kept secure at the Munch Museum, with access only granted by the data manager.

Access to project documents, incl. drafts and bibliography, is granted only by PI.

All other data is publicly available.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The collected/created dataset is of long-term value and will thus be archived on an institutional, open data repository, dataverse.no.

Data will be published and archived in accordance with the FAIR principles.

The data has been shared with the public, open-access CorrespSearch service.

The project uses open standards for letter metadata and Linked Open Data standards for person names and place names (VIAF/Wikidata and geoNames).

What is the long-term preservation plan for the dataset?

Archiving on dataverse.no

Archiving of script/code on Zenodo.org

Data Sharing

How will you share the data?

On GitHub (public repository) of script/code, and created CMIF files.

On dataverse.no and Zenodo.org of archived and documented data and script/code.

With a publication in open access about the creation of the data set.

By ingestion into the CorrespSearch service.

Are any restrictions on data sharing required?

The only data that cannot be shared in its "raw" form is the updated metadata of selected letters collected in a spreadsheet kept by the Munch Museum.

Responsibilities and Resources

Who will be responsible for data management?

The PI, together with the data manager.

What resources will you require to deliver your plan?

All resources needed are provided for by the University of Oslo at no additional costs.

The creation of the script/code to create and collect the correspondence metadata was funded by the University of Oslo Teksthub+Digital Humanities with paid hours for a developer.

Planned Research Outputs

Data paper - "From eMunch to CorrespSearch"

Data paper / article describing the creation and challenges of extracting correspondence metadata from the existing digital scholarly edition eMunch.no and its preparation for the ingestion into the CorrespSearch service.

Dataset - "Correspondence Metadata from eMunch.no"

Dataset with CMIF files from the eMunch.no digital scholarly edition, created with the XMLMuncher script.

Software - "MunchXMLMuncher"

Script to extract correspondence metadata from the eMunch.no digital scholarly edition and enrich said data with Linked Open Data resources like VIAF/Wikidata for person names and geoNames for place names. Updates selected data from the edition based on an unpublished spreadsheet.

Planned research output details

| Title | Type | Anticipated release date | Initial access level | Intended repository(ies) | Anticipated file size | License | Metadata standard(s) | May contain sensitive data? | May contain PII? |
|--|------------|--------------------------|----------------------|--------------------------|-----------------------|--|---|-----------------------------|------------------|
| From eMunch to CorrespSearch | Data paper | 2024-01-30 | Open | None specified | | Creative Commons Attribution 4.0 International | TEI (Text Encoding Initiative) Guidelines | No | No |
| Correspondence Metadata from eMunch.no | Dataset | 2024-01-30 | Open | None specified | | Creative Commons Zero v1.0 Universal | TEI (Text Encoding Initiative) Guidelines | No | No |
| MunchXMLMuncher | Software | 2024-01-30 | Open | Zenodo | | MIT License | None specified | No | No |

Related Works

Articles

- <https://norkorr.hypotheses.org/235>

Datasets

- Rockenberger, Annika, Loke Sjølie, and Hilde Bøe. 2023. “Correspondence Metadata from the Digital Scholarly Edition of Edvard Munch's Writings.” [Dataset]. DataverseNO. <https://doi.org/10.18710/TAFUSV>.

Softwares

- <https://github.com/norkorr/MXMLM>