

"Responding to Asian Diversity"

DH2022

DIGITAL HUMANITIES 2022

Conference Abstracts

*July 25-29, 2022
Tokyo, Japan*

*Online
The University of Tokyo*

another objective of this presentation is to showcase the possibilities of the Digital Humanities project CBDRS as a starting point for research works on Anthropology, Culture, History, Linguistics, and Press, besides the social interest of knowing the birth of journalism, that is, the birth of a way of shaping information, thinking, and international communication.

Bibliography

- Arblaster, P. et al.** (2016). The lexicons of Early Modern News. In Raymond, J. and Moxham, N. (eds.), *News Networks in Early Modern Europe*. Leiden/Boston: Brill, pp. 64-101. <https://doi.org/10.1163/9789004277199>. **EURONEWS Project**. <https://www.euronewsproject.org/>.
- Fernández-Travieso, C. and Garrobo-Peral, M.** (2021). Creación de un corpus digital de textos de relaciones de sucesos: un modelado de datos XML-TEI para acoger diferentes tipos de edición. In Bazzaco, S. (coord.), Congreso Internacional Humanidades Digitales y Estudios Literarios Hispánicos, Verona 22-23 June. Università degli Studi di Verona.
- Folch, D.** (2006). La China en España. Elaboración de un corpus digitalizado de documentos españoles sobre China de 1555 a 1900. <https://www.upf.edu/asia/projectes/che/principal.htm>.
- Pena-Sueiro, N. and Saavedra-Places, Á.** (2019). Obsolescencia y resiliencia en Humanidades digitales. El caso de la Biblioteca Digital de Relaciones de Sucesos. *Artnodes*, 23: 79-88. <http://dx.doi.org/10.7238/a.v0i23.3243>.
- Rojo-Mejuto, N.** (2019). Las relaciones de sucesos españolas sobre Japón en los siglos XVI y XVII. In Torres, L., Tropé, H. and Espejo-Surós, J. (coords.), IX Coloquio de la Sociedad Internacional para el Estudio de las Relaciones de Sucesos (SIERS). Rennes 19-21 September. Université Rennes 2.
- Rojo-Mejuto, N.** (2021a). Japón escondido: cosas muy notables en las cartas de las Indias. In Andrés-Renales, G. and Peñasco-González, S. (eds.), *Buenas noticias: relaciones de sucesos en los siglos XVI-XVIII*. Pesaro: Metauro, pp. 49-64.
- Rojo-Mejuto, N.** (2021b). La literatura informativa sobre Japón en 1621. In Borrego, M. (coord.), Seminario Internacional El mundo en 1621: avisos, relaciones de sucesos y conexiones culturales. Besançon 16-17 September 2021. Université de Franche-Comté.

Mining and Modeling Spaces and Places for Literary History as Linked Open Data

Röttgermann, Julia

roettger@uni-trier.de
Trier University, Germany

Hinzmann, Maria

hinzmannm@uni-trier.de
Trier University, Germany

Dietz, Katharina

dietz@uni-trier.de
Trier University, Germany

Gebhard, Henning

gebhard@uni-trier.de
Trier University, Germany

Klee, Anne

klee@uni-trier.de
Trier University, Germany

Konstanciak, Johanna

konstanciak@uni-trier.de
Trier University, Germany

Schöch, Christof

schoech@uni-trier.de
Trier University, Germany

Steffes, Moritz

steffesm@uni-trier.de
Trier University, Germany

Introduction

In literary history and historiography, places and spaces play an important role – not least in the context of the ‘spatial turn’ (Lafon, 1997; Piatti et al., 2009; Dennerlein, 2009; Weber, 2014). In literary works, narrative locations are particularly relevant, but places of publication as well as further spatial dimensions can also be taken into account (Curran 2018; Burrows et al. 2016). Our contribution presents how we obtained spatial statements from three

different information sources and combined them in a knowledge network based on the Linked Open Data (LOD) paradigm (Berners-Lee, 2006; Hooland und Verbough, 2014; Hitzler, 2021).

Project context

The aim of the project Mining and Modeling Text is to establish an information network for the humanities built from various sources.¹ This aim is closely linked to the finding that, considering the steadily growing digital cultural heritage, the acquisition of knowledge from large amounts of text and data can no longer be handled by individuals. In representing knowledge as LOD, we see untapped potential that we are exploring in the current project phase on the French Enlightenment novel (Delon/ Malandain, 1996; Mylne, 1981).

Creating spatial statements

Information on spatial statements relevant to our domain is extracted from three different types of sources: 1. bibliographic metadata, 2. primary sources, 3. scholarly publications. After focusing on thematic statements in an earlier phase of our project (Schöch et. al., 2022; Röttgermann et al., 2022), we are currently addressing spatial statements.

Mining

Bibliographic metadata: For our domain, the *Bibliographie du genre romanesque français 1751-1800* (BGRF, Mylne et al., 1977) is central, as it defines the population of about 2000 French Enlightenment novels. The BGRF has been extensively analysed and modeled (Lüschow 2020) and contains rich metadata (including places of publication, narrative locations, narrative form, characters, themes, style).

Primary sources: The *Collection of Eighteenth-Century French Novels* (Röttgermann, 2021) is analysed via SpaCy's (Honnibal und Montani, 2017) named entity recognition and reconciliation pipeline supported by OpenRefine (Huynh, [2012] 2010). Our pipeline requires human intervention (Hinzmann et al., 2022) concerning the challenges of ambiguity, fictionality and historicity (Heuser et al., 2016; Jockers, 2016; Nielsen, 2016).

All	Column 1	bgrf	LOC1	Wikidata Label fr	URL	Quantity LOC1
1.	Abbes_Voyage	58.5	Palais	palais	https://www.wikidata.org/wiki/Q16560	3
2.	Anonym_Suzon	83.9	Couvent	couvent	https://www.wikidata.org/wiki/Q1128397	21
3.	Anonyme_Zoloe	00.37				
4.	Amaud_Epoux	83.15	Paris	Paris	https://www.wikidata.org/wiki/Q60	27
5.	Amaud_Matinees	99.43				
6.	Amaud_Sentiment	70.21	Nancy	Nancy	https://www.wikidata.org/wiki/Q40896	111
7.	Barthelemy_Voyage	88.27	Grèce	Grèce	https://www.wikidata.org/wiki/Q41	440

Figure 1: Reconciliation of narrative locations with OpenRefine

Modeling

Our approach relies on importing both the text strings as found in the information source (green) and the abstract spatial items (blue) into our Wikibase instance. Combined with the fact that we reference all statements uniquely via the "stated in" property (orange), this ensures a high degree of verifiability of our data (see fig. 2).

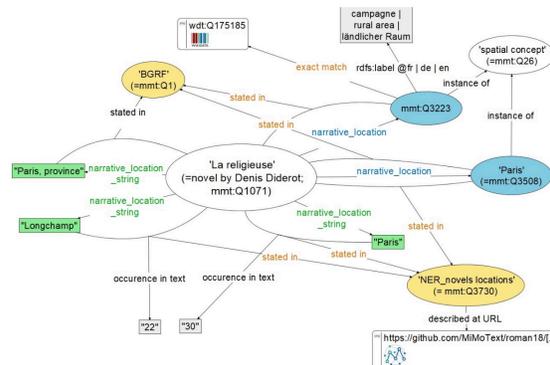


Figure 2: 'Narrative locations' of Diderots *La religieuse* from NER and BGRF data

Our spatial vocabulary was built up incrementally and provides items (=spatial concepts) for 7 properties, of which 5 are currently mapped with Wikidata (see fig. 3). 2

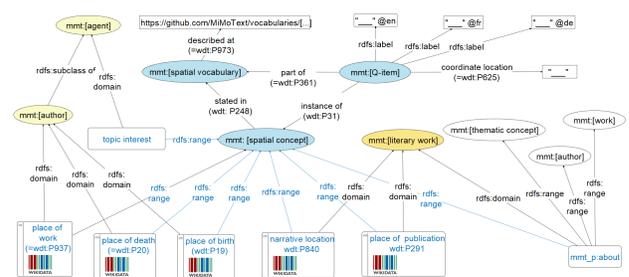


Figure 3: Ontology of 'spatial statements' within the domain of literary history/historiography

Infrastructure

For the provision of data, we follow Open Science principles, such as the publication of FAIR data in open access as well as the use of open source software – in particular Wikibase (see fig.4).³ We created a custom bot using the Python library Pywikibot to import and update the RDF triples into our Wikibase instance from TSV files.⁴

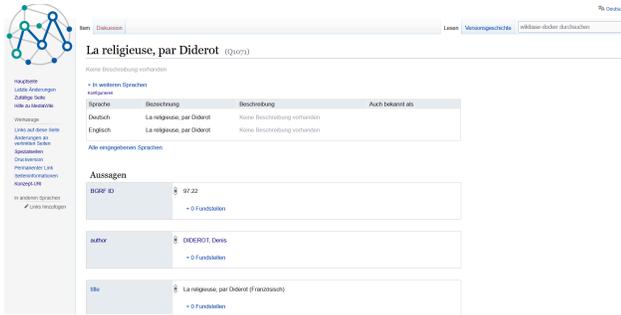


Fig. 4:
Local Wikibase instance

Spatial Querying

Having all the spatial triples stored in our Wikibase allows us to query and visualize them using the DockerWikibaseQueryService interface. We can gain an overview of the entire set of metadata (see fig. 5, query 1), see places of publication appearing and disappearing over time (see fig. 6, query 2) or explore narrative locations linked to specific thematic concepts such as ‘miracle’ (see fig. 7, query 3). Via 'federated queries' (see fig. 8, query 4), information from other knowledge bases (here Wikidata) can be used.

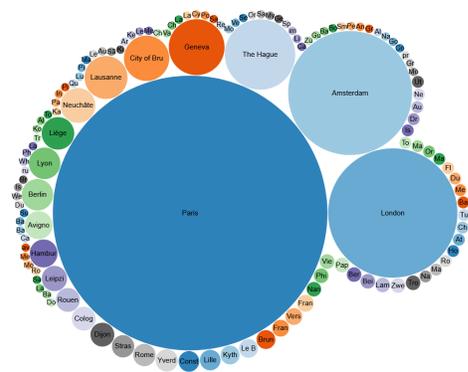


Figure 5:
Overview of the most frequent places of publication

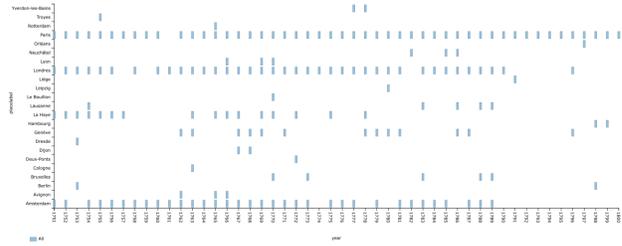


Figure 6:
Places of publication over time



Figure 7:
Narrative locations linked to the thematic concept 'miracle' (excerpt)



Figure 8:
Cluster of dominant publication places based on a federated query (Wikidata)

Conclusion

We showcase key aspects of spatial information extraction and modeling in our project. Our presentation will show how we link spatial statements from three sources into a multilingual knowledge network. Triples on publication dates, themes, locations and authors can be combined and be differentiated by their source, something which allows new perspectives for literary history, book history and other domains. Future work concerns extracting and adding statements from scholarly publications to the Wikibase instance.

Appendix

Query 1: Items (novels) and their place of publication (fig. 5)

```

1 #defaultView:BubbleChart
2 SELECT ?topLabel (count(*) as ?count)
3 WHERE {
4   ?item wdt:P8 ?top . #P8 = 'place of publication'
5   ?top rdfs:label ?topLabel .
6   filter(lang(?topLabel) = "en")
7 }
8 GROUP BY ?topLabel
9 ORDER BY desc(?count)

```

Query 2: Places of publication over time (fig. 6)

```

1 #defaultView:BarChart
2
3 Select ?year ?placelabel(count(*) as ?countyear)
4 WHERE{
5   ?item wdt:P8 ?place.
6   ?place rdfs:label ?placelabel .
7   ?item wdt:P7 ?date .
8   FILTER(lang(?placelabel) = "fr")
9   BIND(str(year(?date)) as ?year)
10  SERVICE wikibase:label {bd:serviceParam wikibase:language "{AUTO_LANGUAGE}","fr" .}
11 }
12
13 GROUP BY ?year ?placelabel
14 having (?countyear > 1)
15 #ORDER BY asc(?year) desc(?countyear)

```

Query 3: Narrative location of novels with theme “miracle” (fig. 7)

```

1 #defaultView:TreeMap
2 SELECT
3   ?value ?valueLabel ?work ?workLabel
4 WHERE
5 {
6   ?work wdt:P52 ?value; #exploring narrative location of works
7     wdt:P25 wd:Q2970 #works with miracle as theme
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" . }
9 }
10

```

Query 4: Places of publication with geocoordinate location via federated query (fig. 8)

```

1 #defaultView:Map["markercluster":true]
2 PREFIX wd: <http://www.wikidata.org/entity/> #wikidata wd
3 PREFIX wdt: <http://www.wikidata.org/prop/direct/> #wikidata wdt
4
5 Select DISTINCT ?item ?itemLabel ?loc ?locLabel ?wikidataEntityLink ?coordinateLocation ?theme ?themeLabel
6
7 ?item wdt:P8 ?loc. # loc = publication place
8 ?loc wdt:P30 ?wikilink.
9
10 BIND(STR(STRCONCAT(STR(?wikilink),"https://www.wikidata.org/wiki/",http://www.wikidata.org/entity" )) AS ?wikidataEntityLink).
11 #federated Query -> Wikidata
12 SERVICE <https://query.wikidata.org sparql> {
13   ?wikidataEntityLink
14     wdt:P625 ?coordinateLocation
15 }
16
17 SERVICE wikibase:label {
18   bd:serviceParam wikibase:language "en" .
19 }
20 }
21

```

Bibliography

- Berners-Lee, T.** (2006). *Linked Data – Design Issues*. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Burrows, S. et al.** (2016). Mapping Print, Connecting Cultures. *Library & Information History*, 32(4), pp. 259–71. 10.1080/17583489.2016.1220781.
- Curran, M.** (2018). *The French Book Trade in Enlightenment Europe I: Selling Enlightenment*. London: Bloomsbury Publishing.
- Delon, M. and Malandain, P.** (1996). *Littérature française du XVIIIe siècle*. Paris: Presses universitaires de France. <https://gallica.bnf.fr/ark:/12148/bpt6k48060529>.
- Dennerlein, K.** (2009). *Narratologie Des Raumes*. Berlin, New York: De Gruyter.
- Heuser, R., Algee-Hewitt, M. and Lockhart, A.** (2016). Mapping the Emotions of London in Fiction, 1700–1900: A Crowdsourcing Experiment. In *Literary Mapping in the Digital Age*. London: Routledge.
- Hitzler, P.** (2021). A Review of the Semantic Web Field. *Commun. ACM*. 10.1145/3397512.
- Honnibal, M. and Montani, I.** (2017). *SpaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing*.
- Hooland, S. van. and Verborgh, R.** (2014). *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Facet Publishing.
- Huynh, D.** (2010). *OpenRefine*. OpenRefine. <https://github.com/OpenRefine/OpenRefine>.
- Jockers, M. L.** (2016). The Ancient World in Nineteenth-Century Fiction; or, Correlating Theme, Geography, and Sentiment in the Nineteenth Century Literary Imagination. *Digital Humanities Quarterly*, 10(2).
- Lafon, H.** (1997). *Espaces Romanesques Du XVIIIe Siècle, 1670-1820: De Madame de Villedieu à Nodier*. 1. éd. Paris: Presses Universitaires de France.

Martin, A., Mylne, V. and Frautschi, R. L. (1977). *Bibliographie du genre romanesque français, 1751-1800*. London: Mansell.

Mylne, V. (1981). *The Eighteenth-Century French Novel: Techniques of Illusion*. 2nd Edition. Cambridge, New York: Cambridge University Press.

Nielsen, F. A. (2016). Literature, Geolocation and Wikidata. In *Wiki@ICWSM*. Proceedings of the International AAAI Conference on Web and Social Media. Cologne, pp. 61–4. <https://ojs.aaai.org/index.php/ICWSM/article/view/14833>.

Röttgermann, J. (ed.) (2021). Collection de romans français du dix-huitième siècle (1750-1800) / Eighteenth-Century French Novels (1750-1800) [dataset]. *Release v0.2.0*. 10.5281/zenodo.5040855.

Schöch, C. et al. (2022). Smart Modelling for Digital Literary History. *IJHAC: International Journal of Humanities and Arts Computing [Special Issue on Linked Open Data]*, 16(1), pp. 78–93. <https://doi.org/10.3366/ijhac.2022.0278>.

Suber, P. (2012). *Open Access*. Cambridge, Mass: The MIT Press.

Weber, A.-K. (2014). *Mapping Literature: Spatial Data Modelling and Automated Cartographic Visualisation of Fictional Spaces*. Zurich: ETH Zurich. 10.3929/ETHZ-A-010106067.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(1), p. 160018. 10.1038/sdata.2016.18.

Notes

1. See <https://mimotext.uni-trier.de/en/>.
2. See for the vocabulary: https://github.com/MiMoText/vocabularies/blob/main/spatial_vocabulary.tsv.
3. Generally, see Suber 2012 and Wilkinson et al. 2016; related to the project, see Röttgermann and Schöch 2020 and Schöch, 2021. We expect the Wikibase instance to become publicly available in mid-2022: <https://www.mimotext.uni-trier.de/en>
4. See <https://github.com/MiMoText/Wikibase-Bot>.

The Silences in Archives: A Case Study of Annual Reports from Archives at the National Centre for Biological Sciences, India

Roy, Dibyadyuti

dibyadyutir@gmail.com
Indian Institute of Technology (IIT) Jodhpur, India

Taparia, Kanishtha

taparia.1@iitj.ac.in
Indian Institute of Technology (IIT) Jodhpur, India

Srinivasan, Venkat

venkats@ncbs.res.in
Archive at the National Centre for Biological Sciences (NCBS)

Contemporary Digital Humanities Scholarship has placed considerable emphasis on the power of archives that has its etymological root in the Greek word "Arkhe"—simultaneously connoting commencement and commandment—as not only sites of preservation but also as opportune sites of exclusion. Institutional archives in particular, through their exclusion of certain records and narratives, create epistemic gaps that enable dominant narratives while concurrently silencing minority voices, events, and subjectivities. In particular for scientific and research institutions, institutional archives and repositories provide crucial evidence of the challenges overcome in the pursuit of scientific knowledge as well as the subtle social changes precipitated by the growth of science and technology in local contexts. Contextually in India, colonial legacies combined with the vagaries of postcolonial statecraft ensure that many of the leading scientific and research institutes in India do not have any institutional archives. A notable outlier is the Archives at The National Center for Biological Sciences (NCBS), which is a collecting archive chronicling the growth of the NCBS along with being a site to document the history of contemporary biology in India. The NCBS was established in 1992 and is a part of the Tata Institute of Fundamental Research, Mumbai (TIFR) which is a National Centre of the Government of India, under the umbrella of the Department of Atomic Energy.

While considerable importance is given in archival and DH scholarship to first person accounts of scientists and similar dominant stakeholders, there is hardly any scholarship that looks at the annual reports of scientific institutions as knowledge systems that generate new questions or inaugurate new models of inquiry. Further existing scholarship has drawn attention toward the underrepresentation of female scientists in archival spaces: but have rarely parsed how institutionalized gender biases mediate such silences. In addressing these lacunae, this project will narrate the results, complications and future directions of an exploratory study which analyses the digitized annual reports for a period of 25 years (1992-2017)