

# Linked Open Data for Literary History: Constructing, Querying and Using the MiMoTextBase



Maria Hinzmann and Tinghui Duan

with Matthias Bremm, Anne Klee, Johanna Konstanciak,  
Julia Röttgermann and Christof Schöch

Project: <https://mimotext.uni-trier.de/en>

Slides: <https://mimotext.github.io/lod-lithist/berlin23.html>

Workshop FU Berlin, 10th October 2023  
Wikipedia, Wikidata and Wikibase: Usage Scenarios for Literary Studies

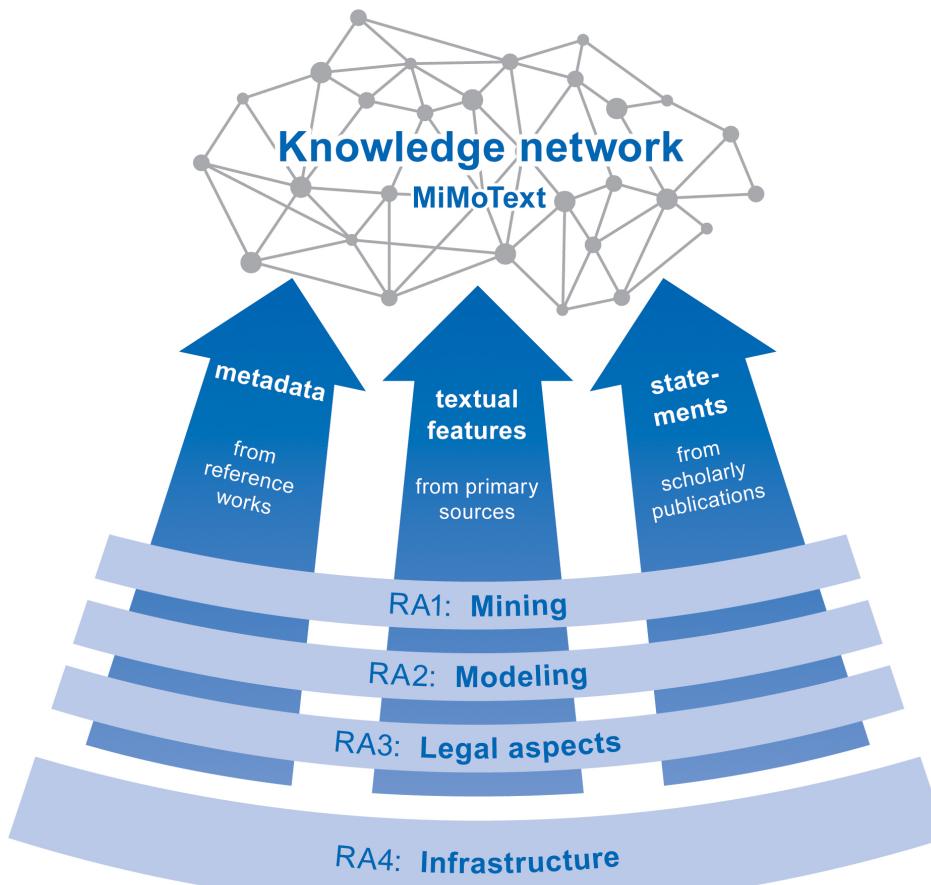


# Structure

1. Introduction
2. Mining
3. Modeling
4. MiMoTextBase & Wikidata
5. Examples

# (1) Introduction

# MiMoText in a nutshell



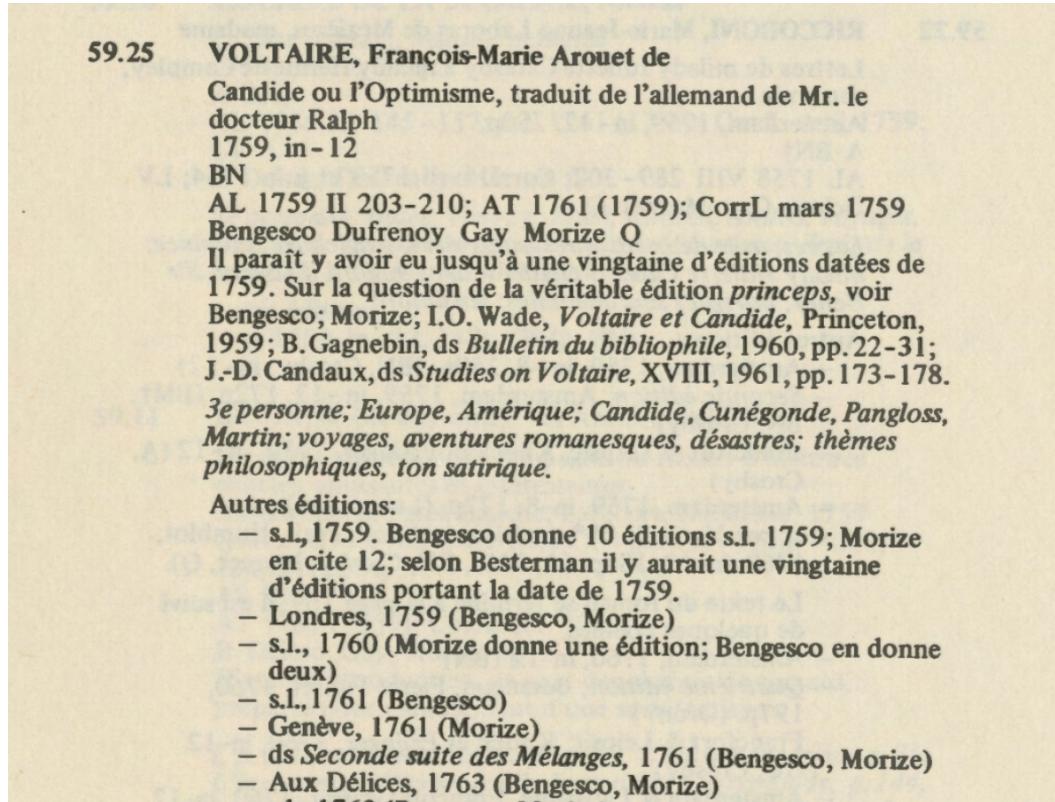
<https://mimotext.uni-trier.de/en>

# Aims of the project

- Our goal: "Wikidata for literary history"
  - An information system for literary history
  - LOD-based, with exploratory interface and SPARQL-endpoint
  - Sort of an "atomization" of literary history into many small statements
  - Held together by taxonomies, ontologies, authority files
- Compared to Wikidata:
  - Much more focused on one domain (French novel 1750-1800)
  - Better coverage for this domain
  - Higher density of assertions for this domain
  - Based on more explicit data modeling

## (2) Mining

# Source type 1: Bibliographic data



Martin / Mylne / Frautschi: *Bibliographie  
du genre romanesque français, 1751-1800*, 1977

# Source type 2: Primary literature (novels)

- corpus of 205 french novels (1750-1800)
- in XML-TEI, with metadata, following the ELTeC schema
- methods: e.g. Topic Modeling, NER, stylometry

The screenshot shows a GitHub repository page for 'roman18'. The left sidebar contains a 'README.md' file, a 'DOI' link (10.5281/zenodo.4061903), and a 'Code' button. The main content area includes:

- Contributors**: 7 contributors represented by icons.
- Languages**: A chart showing the distribution of code files:
  - HTML: 72.3%
  - Jupyter Notebook: 26.6%
  - Python: 1.1%

**README.md**

DOI 10.5281/zenodo.4061903

## roman18

Collection de romans français du dix-huitième siècle (1750-1800) / Collection of Eighteenth-Century French Novels (1750-1800)

### Introduction

This collection of Eighteenth-Century French Novels contains digital texts of novels created or first published between 1751 and 1800. The collection is created in the context of Mining and Modeling Text, a project which is located at the Trier Center for Digital Humanities (TCDH) at Trier University. Work on the collection is ongoing.

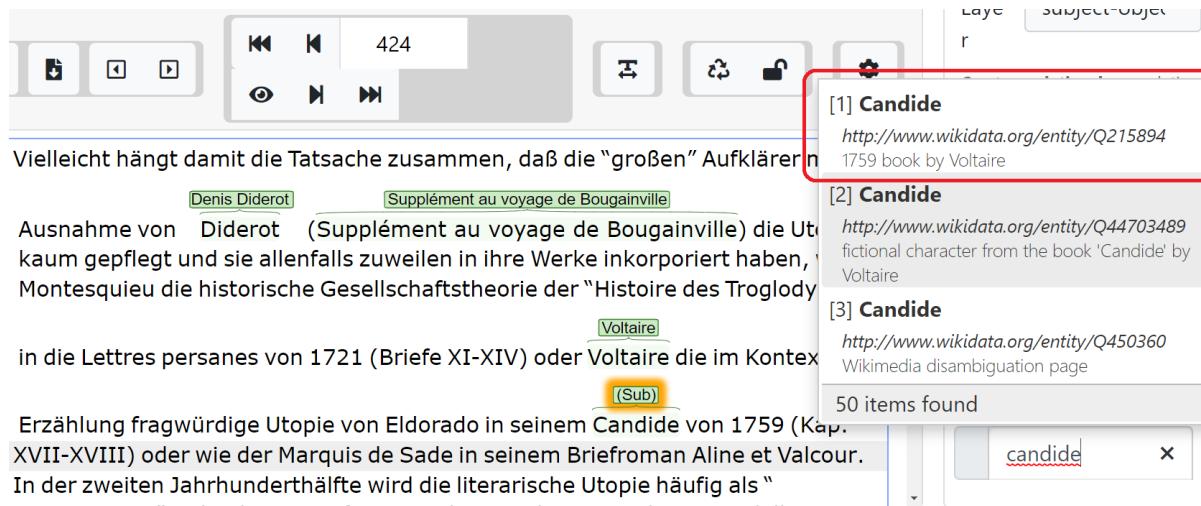
Contributors 7

Languages

Language	Percentage
HTML	72.3%
Jupyter Notebook	26.6%
Python	1.1%

Collection of Eighteenth-Century French Novels (1750-1800)

# Source type 3: Scholarly publications



- manual annotation (e.g. statements about themes)
- annotation tool INCEpTION linked with Wikidata and MiMoTextBase
- pipeline (export INCEpTION annotations => import Wikibase statements)

# (3) Modeling

# Ontology

- Überblick

- Module 1: theme
- Module 2: space
- Module 3: narrative form
- Module 4: literary work
- Module 5: author
- Module 6: mapping
- Module 7: referencing
- Module 8: versioning & publication
- Module 9: terminology
- Module 10: bibliography
- Module 11: scholarly work

(<https://github.com/MiMoText/ontology>)

# Reification (1)

about

libertinism

▼ 2 references

stated in

Bibliographie du genre romanesque  
français

stated in

BGRF\_matching-table (03-2022)

correspondence

▼ 2 references

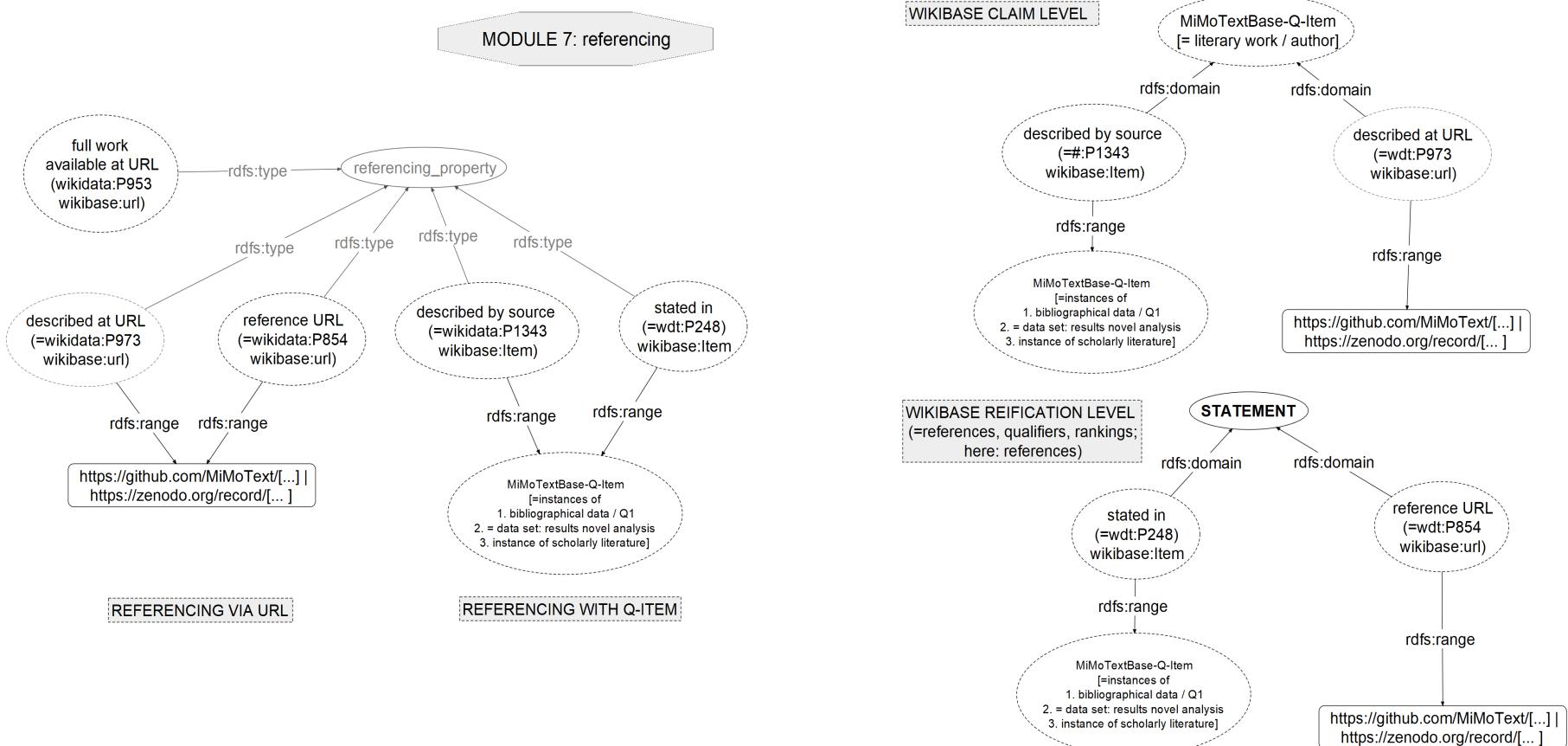
stated in

Topic Model MMT 11-2020

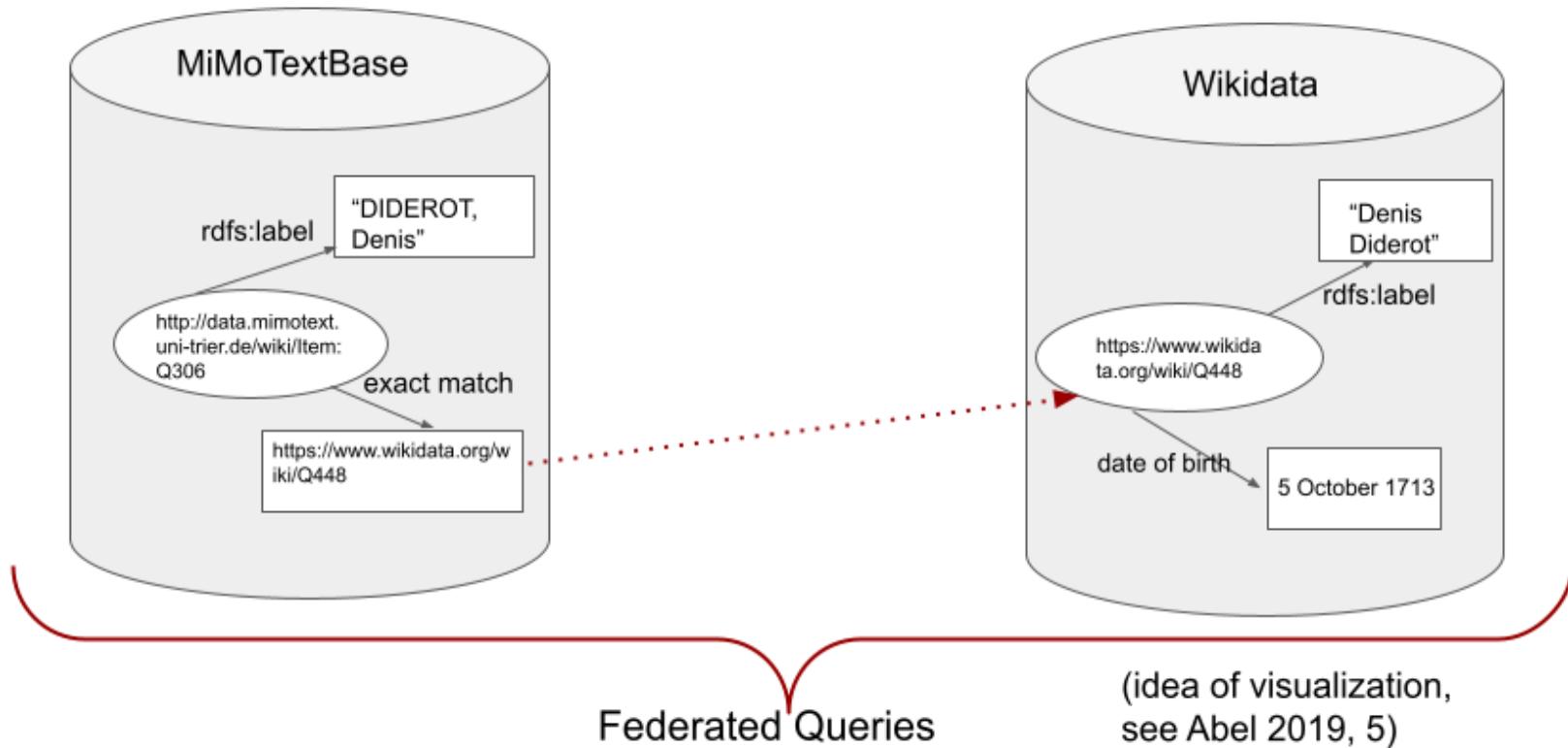
stated in

topic labels and concepts (11-2020)

# Reification (2)



# Alignment with Wikidata: enabling 'federated queries'

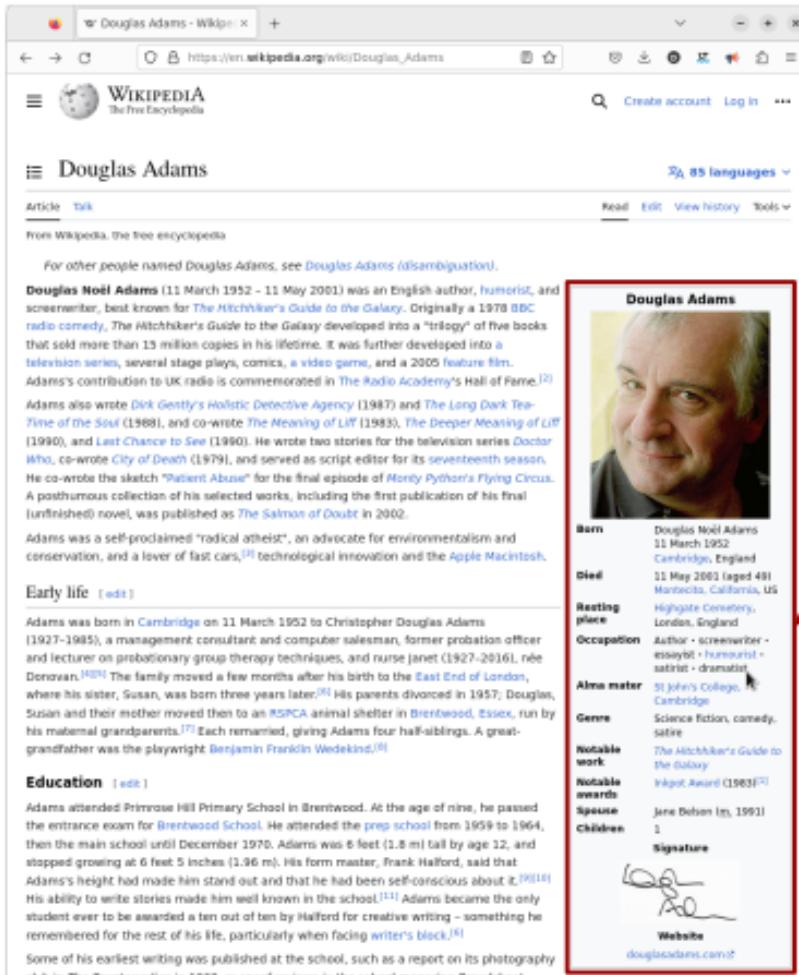


# (4) MiMoTextBase & Wikidata

# What is Wikibase?



since 2001

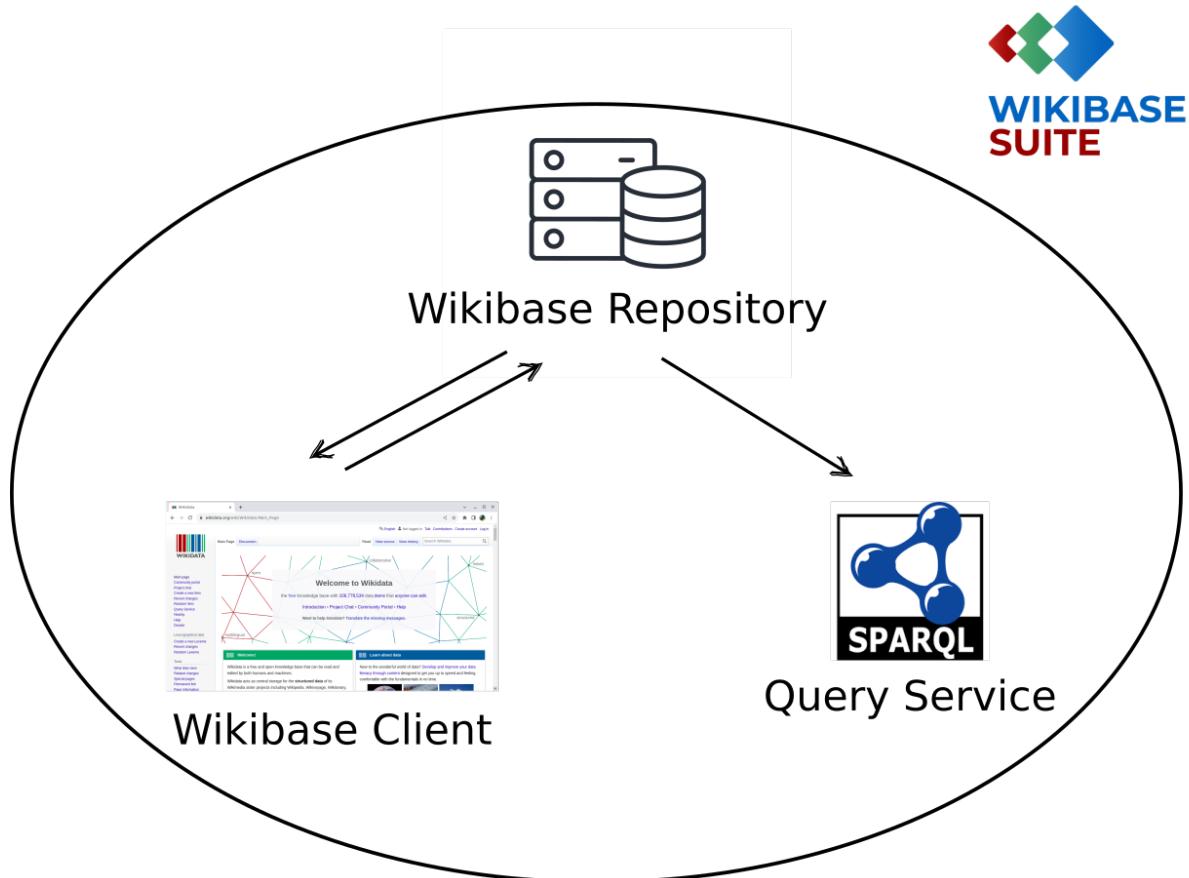


A screenshot of a Wikipedia article page for "Douglas Adams". The page title is "Douglas Adams". The main content discusses the life and work of Douglas Adams, mentioning his books like "The Hitchhiker's Guide to the Galaxy" and "The Meaning of Life", and his television series like "Doctor Who". A large portrait of Douglas Adams is centered on the page. To the right of the portrait is a red-bordered box containing his biography, which includes details such as birth (11 March 1952), death (11 May 2001), resting place (Highgate Cemetery, London, England), occupation (Author, screenwriter, essayist, humorist, satirist, dramatist), alma mater (St John's College, Cambridge), genre (Science fiction, comedy, satire), notable work ("The Hitchhiker's Guide to the Galaxy", "Dirge Award 1983"), and awards (Dirge Award 1983). Below this box is his signature and a link to his website: dougadams.com/stuff.

since 2012



# Wikibase as infrastructure



# Why using Wikibase?

- It's a FOSS! (free open source software)
- Flexible data model
- User-friendly interface and different import possibilities
- Easily linkable to other Wikibase-instances
- Multilingualism

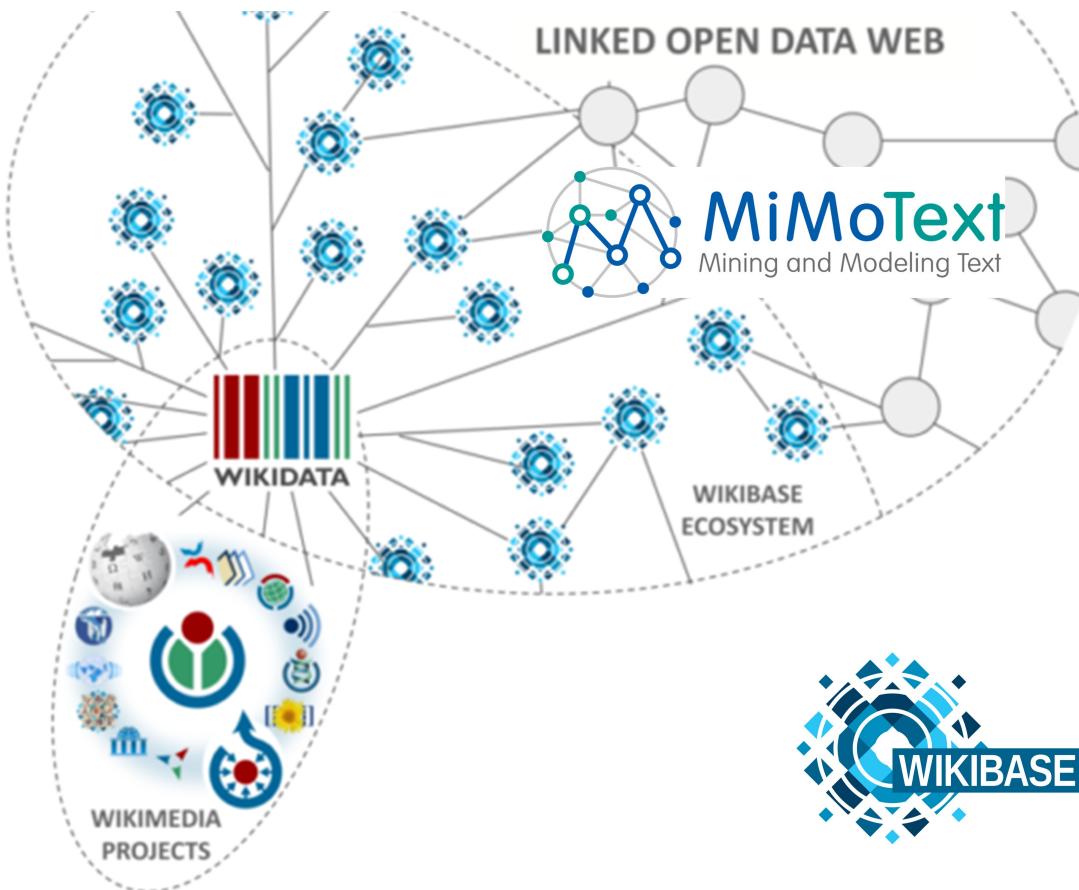
# Some example instances of Wikibase



RHIZOME



# MiMoTextBase as part of the Wikibase ecosystem



A view of the MiMoTextBase within the Wikimedia Linked Open Data web.  
Credit original visualization: [Dan Shick \(WMDE\)](#) / CC-BY-SA 4.0

# Multilingualism facilitates international collaboration

The screenshot shows three Wikidata pages for the same entity, Douglas Adams (Q42), displayed in three different languages: Chinese (简体), English, and Arabic. The pages are part of a multilingual interface where the same data is presented in multiple languages simultaneously.

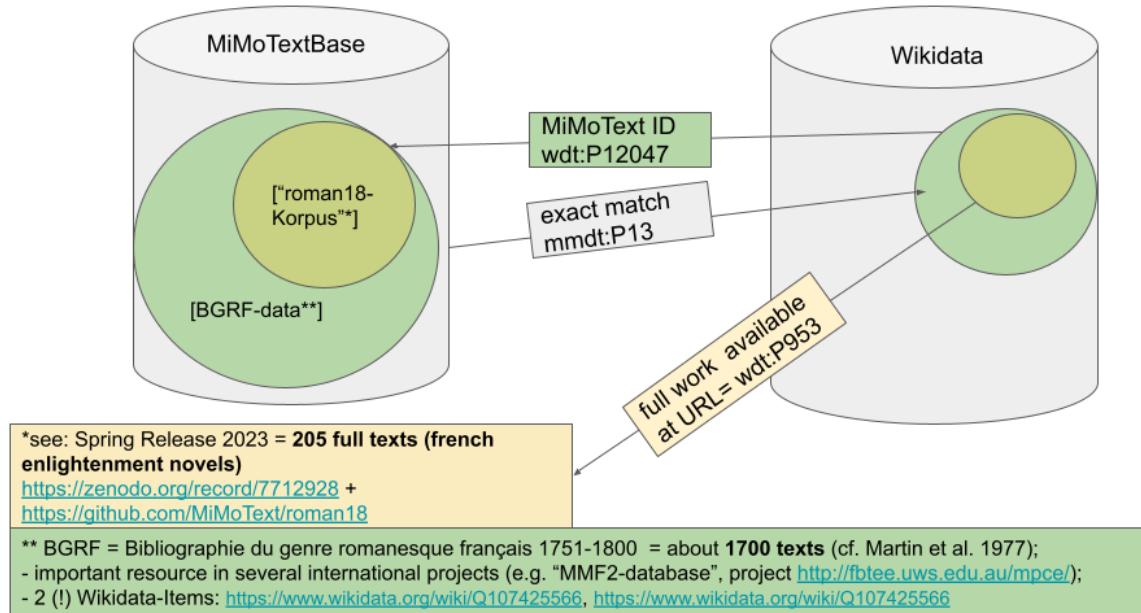
**Left Panel (Wikidata Home):** Includes links for 首页 (Home), 社群首页 (Community Home), 互助客栈 (互助客栈), 新建项目 (New Project), 最近更改 (Recent Changes), 随机项目 (Random Page), 查询服务 (Query Service), 附近 (Nearby), 帮助 (Help), 资助 (Sponsorship), 词典编纂数据 (Dictionary Data), 新建词位 (New Word), 最近更改 (Recent Changes), 随机词位 (Random Word), 工具 (Tools), 链入页面 (Linked Pages), 相关更改 (Related Changes), 特殊页面 (Special Pages), 固定链接 (Permanent Link), 页面信息 (Page Information), 概念URI (Concept URI), 引用本页 (Reference This Page).

**Central Panel (Douglas Adams (Q42) Page):** The main page displays the entity's label, description, statements, and other properties. A specific statement about his education at St John's College is highlighted with a green box. This statement includes qualifiers like end time (1974), academic major (English literature), and academic degree (Bachelor of Arts). It also lists references from Encyclopædia Britannica Online and NNDB.

**Right Panel (Douglas Adams (Q42) Page):** The right panel shows the entity's identifier, aliases, and other details. It includes sections for بحث (Search), تفاصيل (Details), وسائط (Media), and مراجع (References). A sidebar on the right provides links to various Wikidata features and tools.

# Connecting MiMoTextBase & Wikidata

## Future work / plans (10/2023)



- Contribution of MiMoText to Wikidata
  - Adding new Items to Wikidata
  - **MiMoText ID** as a new property (external identifier) to Wikidata
  - Linking novel items on Wikidata to fulltexts on Github

# (5) Examples

# Result: the MiMoTextBase



Main page  
Recent changes  
Random page  
Help about MediaWiki

Tools  
What links here  
Related changes  
Special pages  
Printable version  
Permanent link  
Page information

In other languages  
Add links

Main Page Discussion Read View source View history Search MiMoText

## Main Page

**Mining and Modeling Text: Interdisciplinary applications, informational development, legal perspectives (MiMo Text)**

The acquisition of knowledge from large amounts of text and data which can no longer be handled by individuals is becoming increasingly important due to the possibilities of digitisation. For the humanities, this means in particular that digital full texts and rich metadata must not only be available, but must also be available in a form that promotes knowledge in the humanities.

The aim of the MiMoText project is therefore to establish an information network for the humanities fed from various sources, which, by making it available as Linked Open Data, is not only freely available and can be linked to other knowledge resources of the Semantic Web, but also offers innovative and efficient access possibilities to scientific information.

MiMoTextBase was built as part of the project "Mining and Modeling Text" (2019-2023). It is implemented using a Wikibase infrastructure and integrates various data from heterogeneous sources. Note that the project is ongoing and the contents and structure of the MiMoTextBase will continually be further developed.

### Key starting points

- Tutorial and further information about the MiMoTextBase: <https://docs.mimotext.uni-trier.de>
- SPARQL endpoint: <https://query.mimotext.uni-trier.de>
- MiMoTextBase (current page): <https://data.mimotext.uni-trier.de>
- Project homepage: <https://mimotext.uni-trier.de/en>
- Ontology repository: <https://github.com/MiMoText/ontology>

### Example pages

- An author entry (Réveroni Saint-Cyr): <http://data.mimotext.uni-trier.de/wiki/Item:Q851>
- A title entry (Liaisons dangereuses): <http://data.mimotext.uni-trier.de/wiki/Item:Q1053>
- A thematic concept (travel): <http://data.mimotext.uni-trier.de/wiki/Item:Q3126>
- A spatial concept (Geneva): <http://data.mimotext.uni-trier.de/wiki/Item:Q3478>

- <http://data.mimotext.uni-trier.de>

# The SPARQL endpoint

A screenshot of a web-based SPARQL endpoint interface. At the top, there are navigation icons (refresh, help) and a status bar indicating "1366 results in 454 ms". To the right are links for "Code", "Download", and "Link". Below the header is a search bar with a magnifying glass icon. The main area is a table with the following columns: bgrf, item, authorlabel, itemLabel, year, narrpers, tonality, pages, and normalized. The table contains six rows of data, each corresponding to a book entry with details like title, author, year, and genre.

bgrf	item	authorlabel	itemLabel	year	narrpers	tonality	pages	normalized
51.42	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1010&gt;</a>	VOISENON, Claude-Henri de Fusée, abbé de	Histoire de la félicité	1751	Deux récits 1re personne	but moralisateur	136p.	Deux récits 1re personne
51.41	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1237&gt;</a>	TOUSSAINT, François-Vincent	Histoire des passions	1751				unbekannt
51.39	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1235&gt;</a>	MÉHÉGAN, Guillaume-Alexandre, chevalier de	Zoroastre	1751	3e personne	satire de la religion chrétienne, du luxe, etc.	14 + 60p.	3e personne
51.38	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1234&gt;</a>	MARTIGNY, comte de	Voyage d'Alcimédon	1751	3e personne	ton satirique	ix + 144p.	3e personne
51.37	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1233&gt;</a>	MARCHAND, Jean-Henri	Les avis d'un père à son fils	1751	1re personne, avec récits intercalés 3e personne	intentions moralisatrices, thème de l'éducation	151p.	1re personne, avec récits intercalés 3e personne
51.36	<a href="#">Q &lt;http://data.mimotext.uni-trier.de/entity/Q1232&gt;</a>	MAINVILLIERS, Genu Soalhat, chevalier de	Le petit-maitre philosophe	1751	1re personne	autobiographie romancée, avec traits satiriques	viii + 162, 216, 161p.	1re personne

- SPARQL = SPARQL Protocol and RDF Query Language
- Used to formulate complex queries on LOD
- <https://query.mimotext.uni-trier.de>

# Some example queries

- Simple queries
  - List of novels with information from BGRF
  - The number of works written by each author (first 25)
  - The themes of the novels, in French and in English
- Queries with visualization
  - Number of novels published per year
  - The authors (by date of birth, with portrait)
  - The narrative form of the novels (and their prevalence)
  - Book history: formats per year
- Federated queries
  - The narrative locations in all novels (map)
  - Alternative authorlabels via skos:altLabel
  - Linking with catalogue data via 'BNF identifier'
  - Linking relations between authors via 'influenced by'
- Compare information from two sources
  - Themes derived from topic modeling compared to themes according to BGRF
  - Combined: themes by BGRF vs. from topic modeling

# Thank you!



## To learn more

- Tutorial: <https://docs.mimotext.uni-trier.de>
- SPARQL endpoint: <https://query.mimotext.uni-trier.de>
- MiMoTextBase: <https://data.mimotext.uni-trier.de>
- MiMoText Ontology: <https://github.com/MiMoText/ontology>
- Reference publication: 'Smart Modeling for Digital Literary History'
- Overview visualizations WDQS

**Link to this page** <https://mimotext.github.io/lod-lithist/berlin23.html#/6/4>