From Domain Experts to Data Scientists

Connecting the Heritage of the Printed Book Database and its users

CERL Annual Seminar 2019. Collections and Networks. Reconstructing the Historical Context of Texts, Publications and Objects with Digital Methods. Wednesday, Oct 9.

Andreas Walker

SUB Göttingen / Data Conversion Group







The HPB Database (previously called the Hand Press Book Database) is a steadily growing collection of files of catalogue records from major European and North American research libraries covering items of European printing of the hand-press period (c.1455-c.1830) integrated into one file. This makes it possible for information to be retrieved in one single search across all files. As the digitisation of collections in contributing libraries progresses, more and more catalogue records point to digital presentations of the early printed books.

The HPB Database is of interest to librarians and any one else with academic pursuits across many fields of study that use printed books as source material. It is especially valuable for research in intellectual history, social history, and transmission of thought as well as in the history of printing and the history of the book. It may be accessed for information retrieval and downloading by CERL member institutions, their staff and users.

Notice

- This database is presented in unicode (UTF8). Possible character coding problems might be related to locally missing character fonts (Fonts) in your operating system. The following links could be helpful: Display Problems (Unicode Consortium), UTF8 problems (Wikipedia)
- Usage of MSC (Mathematics Subject Classification) in line with Creative Commons Attribution Noncommercial Share Alike License.



Search





CERL - Heritage of the Printed Book Database. The HPB Database (previously called the Hand Press Book Database) is a steadily growing collection of files of catalogue records from major European and North American research libraries covering items of European printing of the hand-press period (c.1455-c.1830) integrated into one file. This makes it possible for information to be retrieved in one single search across all files. As the digitisation of collections in contributing libraries progresses, more and more catalogue records point to digital presentations of the early printed books.

Project context

- My background: Librarianship / Linguistics / Media Studies
- 2017-2019: Bibliotheksreferendariat (Academic Librarian Training) at the SUB Göttingen, M.A. (LIS) at Humboldt University
- This project = final project + thesis
- Aim: To explore the potential of a linked open data representation of the bibliographic data in CERL's Heritage of the Printed Book Database (HPB), and the development of a web interface for this representation based on the needs of current HPB users.

About the HPB

- **Scope:** bibliographic records of early European prints (c. 1450-1830), aggregated from member libraries
- Roughly eight million records stored in a Pica database hosted in Göttingen and made accessible through a traditional OPAC interface at http://hpb.cerl.org (to humans) and through Z39.50 and SRU (to machines)
- Open to non-members since January 2018
- CERL's aim: increase the visibility and usefulness of this data to researchers and other interested users



The HPB Database (previously called the Hand Press Book Database) is a steadily growing collection of files of catalogue records from major European and North American research libraries covering items of European printing of the hand-press period (c.1455-c.1830) integrated into one file. This makes it possible for information to be retrieved in one single search across all files. As the digitisation of collections in contributing libraries progresses, more and more catalogue records point to digital presentations of the early printed books.

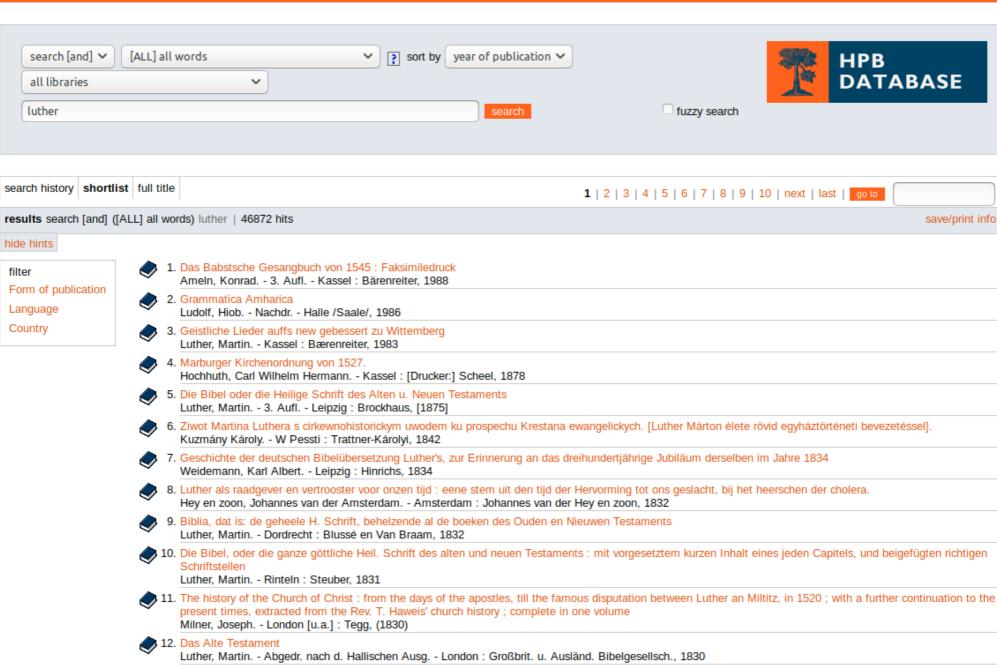
The HPB Database is of interest to librarians and any one else with academic pursuits across many fields of study that use printed books as source material. It is especially valuable for research in intellectual history, social history, and transmission of thought as well as in the history of printing and the history of the book. It may be accessed for information retrieval and downloading by CERL member institutions, their staff and users.

Notice

- This database is presented in unicode (UTF8). Possible character coding problems might be related to locally missing character fonts (Fonts) in your operating system. The following links could be helpful: Display Problems (Unicode Consortium), UTF8 problems (Wikipedia)
- Usage of MSC (Mathematics Subject Classification) in line with Creative Commons Attribution Noncommercial Share Alike License.



search fill	ter												
search	Fill out one or more words in the search form below and add the desired settings											clear form	
			[TIT] title (keywo	rds)			~		luther				
	and	٠ ل	PER] person/au	thor			~						
	and	٠ ل	/ [ALL] all words				~						
	and	and					√ (sb						
		sort by						y [year of publication 🗸				
	language country year of publication						nguage	e	All languages All countries for example: 1948-1980 or 1948- or 1955				
							countr	y					
							lication	n [
									fuzzy searc	h			
	material selection all none												
		Books	Music	Music 🖺 💋			Periodicals/Series (printed)						
	<u> </u>	 ✓ Maps ✓ Microforms ✓ Games, Sculpture, etc. 							Articles				
	4												
	Library all none												
		✓ [BE] Brussels, Bibliothèque royale / Koninklijke Bibliotheek							✓	ſ	CZ] Prague, National Library of the Czech Republic		
	_	 ☑ [DE] GBV Union Catalogue (GVK) ☑ [DE] Bavarian Union Catalogue ☑ [EE] Tallinn, Academic Library of Tallinn University ☑ [ES] Madrid, Universidad Complutense ☑ [FI] Helsinki, National Library of Finland 						•			DE] VD 16		
	_										DK] Copenhagen, Kongelige Biblioteket		
											EE] Tartu, University of Tartu Library		
									✓		ES] Madrid, Biblioteca Nacional de España		
	\checkmark								$ \mathbf{\nabla}$		FR] Sudoc Catalogue		



🃣 13. Geschichte der vornehmsten Reformatoren und der Folgen ihrer Bemühungen von Jesus Christus an bis auf Martin Luther und den dreißigjährigen Krieg : ein

▲ 15 Evagation apprelation & Continues Engrationes in Consein Can 20, 20

Oertel, Eucharius Ferdinand Christian. - Nürnberg : Campe, 1830

14. Exegetica opera latina. 6 Continens Enarrationes in Genesin, Cap. 24 - 26
Luther, Martin. - [Erlanger Ausg.]. - Erlangae : Heyder, 1830

belehrendes und unterhaltendes Volksbuch zur Vorbereitung auf die Konfessionsfeier am 25. Juni 1830



HPB Identifier: NO-0030100.01.910360928 Cite

Cataloguing Source: National Library of Norway (Nasjonalbiblioteket, Oslo)

Title: Das Babstsche Gesangbuch von 1545: Faksimiledruck / mit einem Geleitwort herausgegeben von Konrad Ameln

Collaborator: Ameln, Konrad; Babst, Valentin; Luther, Martin *1483-1546*

Edition: 3. Aufl.

Published: Kassel: Bärenreiter, 1988 Place of printing in the CERL Thesaurus

Language/s: German

Extent: [554], 16 s. : mus.

Series: Documenta musicologica. Reihe 1, Druckschriften-Faksimiles ; 38

Note: Også med original tittelside: Geystliche Lieder / mit einer newen vorrhede Mart. Luth.

Contains: Babst, Valentin: Geystliche Lieder Contains: Luther, Martin: Geystliche Lieder

Related title: Geystliche Lieder

ISBN: 3-7618-0909-3

Subject headings: *Tyskland

*koraler *salmebok

Subject (DDC): DDC4/nor: 782.5250943

Former identifier: NO-0030100.01.910360928

Record in source system

Copies

Holding Institution: Nasjonalbiblioteket

Shelfmark: TZ Doc:38

Location: Seksjon Biblioteket, Utlånet, NB/BRU Musikk

Country: Norway

Our users

- The "typical" HPB users is a domain expert, but not necessarily a tech user (distinctions due to Dadzie & Rowe 2011)
- New tools should integrate with researchers' existing workflows
- My methodology: qualitative interviews with five researchers to find out about their current use of the HPB and their desiderate for future version

What we learned

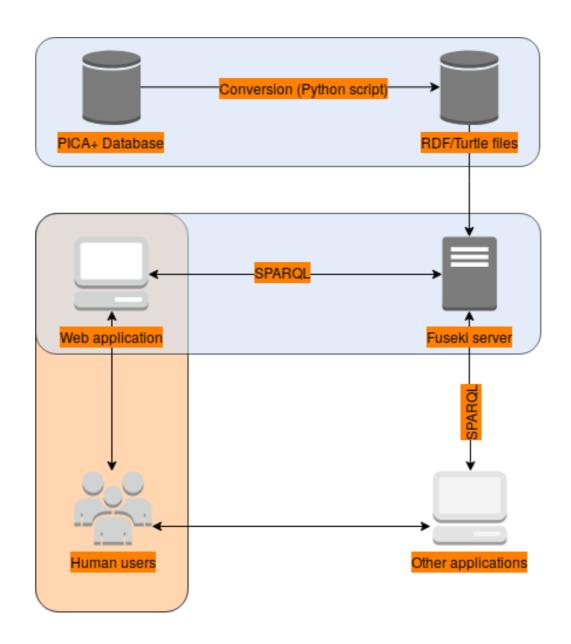
- Participants see the HPB as a useful tool but had encountered some problems working with it (e.g., incomplete data, difficulties in accessing the data, no clustering of copies by edition)
- Participants are interested in the possibility of manipulating and enriching data directly within a future web interface of the HPB, but also need their workflows based on standard office software to be supported
- Participants mostly mention maps and timelines as visualizations they would like to be able to generate from the data

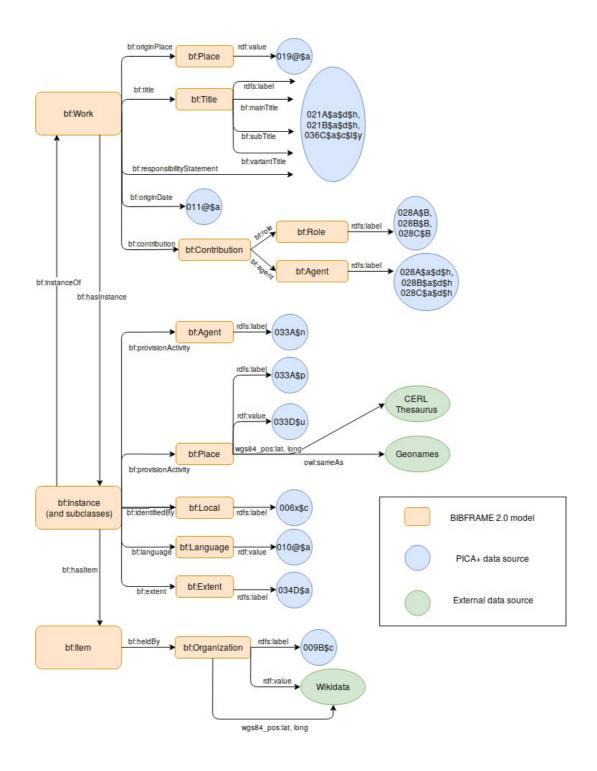
Design decisions

- Develop an interface that is aimed at domain experts and does not require them to interact directly with technologies like SPARQL
- Let the data presentation be guided by **researchers' workflows** rather than by the internal data structure
- Maintain a coherent visual identity with the current HPB
- Make generous use of map visualizations to quickly generate added value to the new interface

The underlying technology

- Data conversion from Pica+ to RDF, using BIBFRAME 2.0 as the data model to increase interoperability, implemented in Python
- Enrichment of the data from external data sources, including the
 CERL Thesaurus and Wikidata
- Provision of a SPARQL endpoint through Apache Fuseki
- Web application built in Python / Flask, with the triple store serving as the database queried through the SPARQL endpoint







Advanced Search



CERL - Heritage of the Printed Book Database. The HPB Database (previously called the Hand Press Book Database) is a steadily growing collection of files of catalogue records from major European and North American research libraries covering items of European printing of the hand-press period (c.1455-c.1830) integrated into one file. This makes it possible for information to be retrieved in one single search across all files. As the digitisation of collections in contributing libraries progresses, more and more catalogue records point to digital presentations of the early printed books.

CERL →

SPARQL Interface

This version of the HPB is implemented as Linked Open Data in RDF. It allows complex queries using the powerful SPARQL language. Note that this requires an understanding of the underlying data model, BIBFRAME.

To the SPARQL interface →

Demonstrations

Combining multiple data sources through linked data allows for a number of interesting queries.

[1] Items held outside their country of origin (matches HPB country data to || Wikidata's country data for holding institutions)

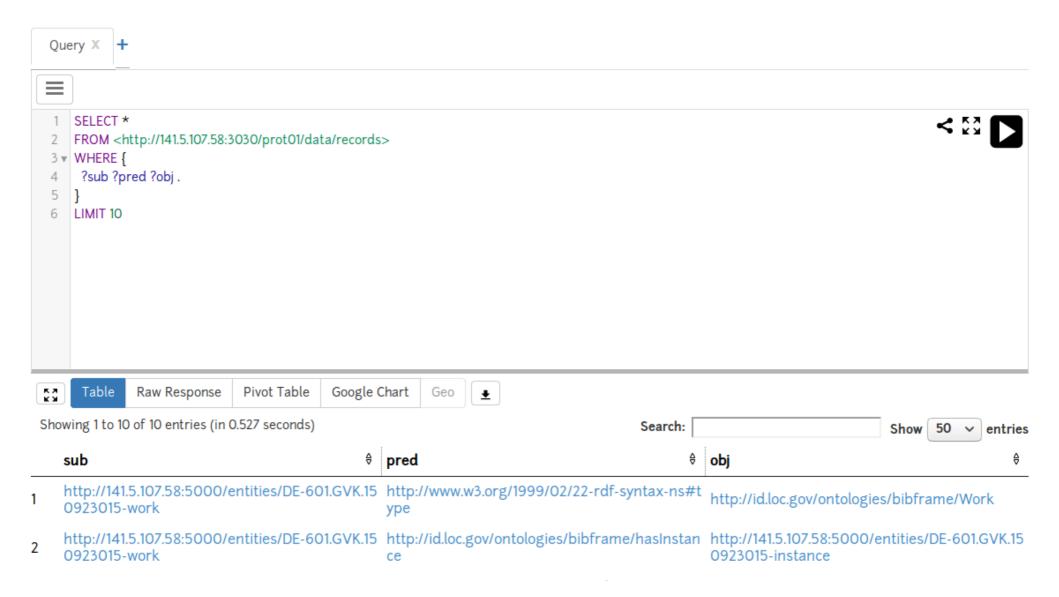
Traditional interface



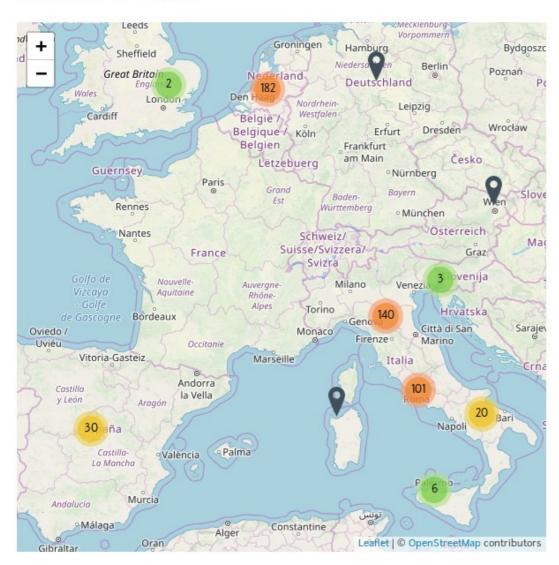
SPARQL Interface

Please note that the HPB data is stored in a named graph. You can query it by including the following statement in your SPARQL query:

FROM http://141.5.107.58:3030/prot01/data/records



Your results searching for items held outside their country of origin



11 records not displayed because they lack geographical data for their holdings. See list view for all records.



Your results searching for luther

- M. Lutheri Catechesis major Luther, Martin (1483-1546)
- 2. [Bildnis Margarethe Luther]
- 3. Kurtze Christliche Fragstücke / Martin Luther Luther, Martin (1483-1546)
- 4. Jesus Syrach Teutsch / Martin Luther [Vorr.]
- Quæ de genuino systemate Lutheranorum solicita est, et quindecim priores §§ complectitur Mörl, Gustav Philipp (1673-1750)
- Vera fides, seu Romanae ecclesiae evidens credibilitas et Lutheranae aliarumque sectarum certa falsitas solide demonstrata Pichler, Vitus (1670-1736)
- 7. Ein newer Moses: Nemblich ein offentlich Zeugniß, daß die Papisten mit diesem ... Krieg nicht allein die genandten Calvinisten, sondern auch die Lutheraner ... außzureutten gemeinet seind
- 8. D. Mort, Luthers liiden Catechismus
- 9. De egressu et ingressu Lutheri Wormatiam carmen Philomusus, Decius
- 10. Decem Precepta witt-e||bergensi predicata || populo || per P. Mar||tinum Luther || augustinia||num.|| Luther, Martin
- 11. Der ganze Psalter des Königes und Propheten Davids / verdeutscht von D. Martin Luther. Mit jedes Psalms kurzen Summarien und nöthigsten Parallelen.
- 12 Instructio Pro Vero Lutherano eleganti carmine scripta / ah Ernesto



List view | Map view

Filters

Language: All (113)

List options

Sort by author

Sort by title

Sort by year

Edit list manually

Your results searching for luther



6 records not displayed because they lack geographical data. See list view for all records.



HPB

Search

Advanced

Results

Help

CERL →

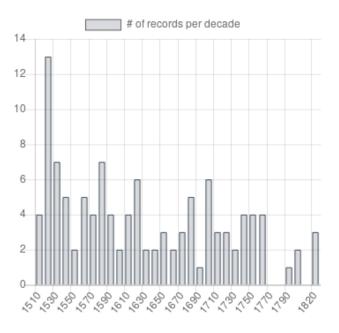
Kettner, Leonnart (10. jn.)

- 19. Reinchristlicher Religions-Unterricht nach Luthern : Ein Katechismus zum Gebrauch der Schullehrer / (C. L. Krüger) Krüger, Carl Ludwig
- 20. Auserlesenes Gesangbuch: Darinn Eine gute anzahl geistreicher So wol Alten und Neuen Gesängen/ Aus D. Martino Luthero/ Und andern Evangelischen Lehrern und gottseligen Christen Mit fleiß zusammen getragen; Und samt Angehängten Christlichen Gebethen/ Zu allgemeiner Kirch- und Haus-Andacht/ In richtige ordnung verfasset
- 21. Das Benedictus odder weyssagung des heyligen Zacharie, Luce. j. / durch D. Mart. Luther gepredigt vnd ausgelegt Luther, Martin (1483-1546)
- 22. Looke beyond Luther: or an answere to that question proposed by our adversaries where our religion was before Luthers time Bernard, Richard (1568-1641)
- 23. Eyn gutte trostliche predig vo[n] der wirdigen bereytu[n]g zu de[m] hochwirdigen Sacrament / Doctor Martini Luther Augustiner tzu Wittenberg Luther, Martin (1483-1546)
- 24. Vermanung zum Sacrament des leibs und bluts unsers Herrn. Mart. Luther. Wittemberg. 1531.
 Luther. Martin
- 25. Grund und Kerrn des Catechißmi, D. Martini Lutheri : Einfeltig u. kürtzlich auß Gottes Wort, u. reiner Lehrer, sonderlich aber D. Lutheri Schrifften ...
- Comment und Außlegung weiland des Herren Doctor Martin Luthers uber den CXXVII. Psalm: Nisi dominus aedificaverit domum Luther, Martin (1483-1546)
- 27. Utrum et quantum quove consilio et successu theologi recentiores, qui omnem scripturae sacrae interpretationem ad rationem revocent, a Lutheri mente atque legibus, quas ille sequendas putavit, defecerint. 1 / [Ioan. Frid. Krause]

Krause, Johann Friedrich (1770-1820)



113 items in list.



← Back

Schmalkaldische Artickel, So da hetten s ...

CID DE-604.VK.BV010631077 →

Author Luther, Martin (1483-1546)

Title Schmalkaldische Artickel, So da hetten sollen auffs

Concilium zu Mantua, oder wo es würde sein, von unsers Theils wegen uberantwortet werden ... / Geschrieben durch D. Martin Luth. und von vielen Gelarten auff den Tag zu Schmalkalden unterschrieben ... Anno 1537

Place of

Wittenberg Q 🝕

Publication

Publisher Krafft

Year 1576

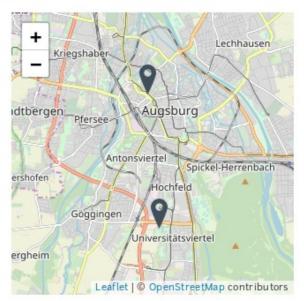
Language German Q 🝕

Holdings Bayerische Staatsbibliothek

Holdings Staats- und Stadtbibliothek Augsburg IIII

Holdings Universitätsbibliothek Augsburg **Ⅲ**





Evaluation of the first prototype

- Methodology: qualitative, semi-structured **follow-up interviews** with the previously interviewed researchers after providing them with a link to the prototype and giving them about a month for testing it
- Questions asked were focused on:
 - Their **overall visual impression** of the interface
 - Their opinion on the presentation of search results
 - Their opinion on the presentation of single records
 - Any remaining thoughts on usability

Evaluation of the first prototype

• Researchers responded in writing. The following is a free summary of the points most salient to me; see Walker (2019) for a more complete list:

Visual impression:

- Colours and layout mostly judged as attractive and well-organized
- Search interface too small and cluttered

Search results:

- Participants liked the map visualization and the options for filtering and manually editing result lists
- Result counter not obvious enough
- Result list should display different metadata

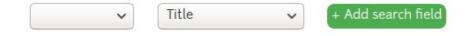
Evaluation of the first prototype

Single records:

- Participants liked the map visualization and the overall organization of the presentation
- Wikidata and RDF links were judged as not useful
- Interface displays less data than the original HPB
- Other identifiers should be added (e.g., GdW, MEI, VD16)

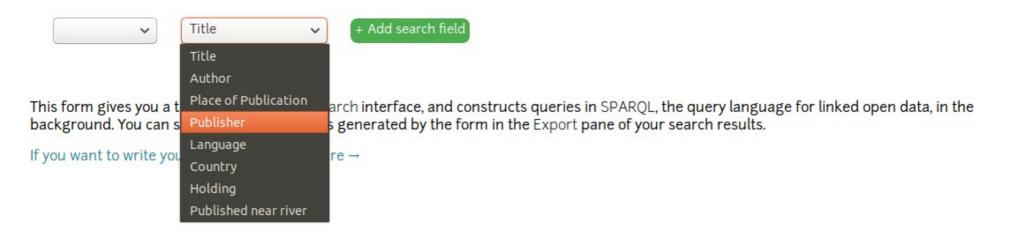
Usability:

- Participants liked the additional integration of the HPB with other CERL resources (i.e., the CERL Thesaurus) and suggested adding more such links
- The search functionality of the prototype was judged as inadequate

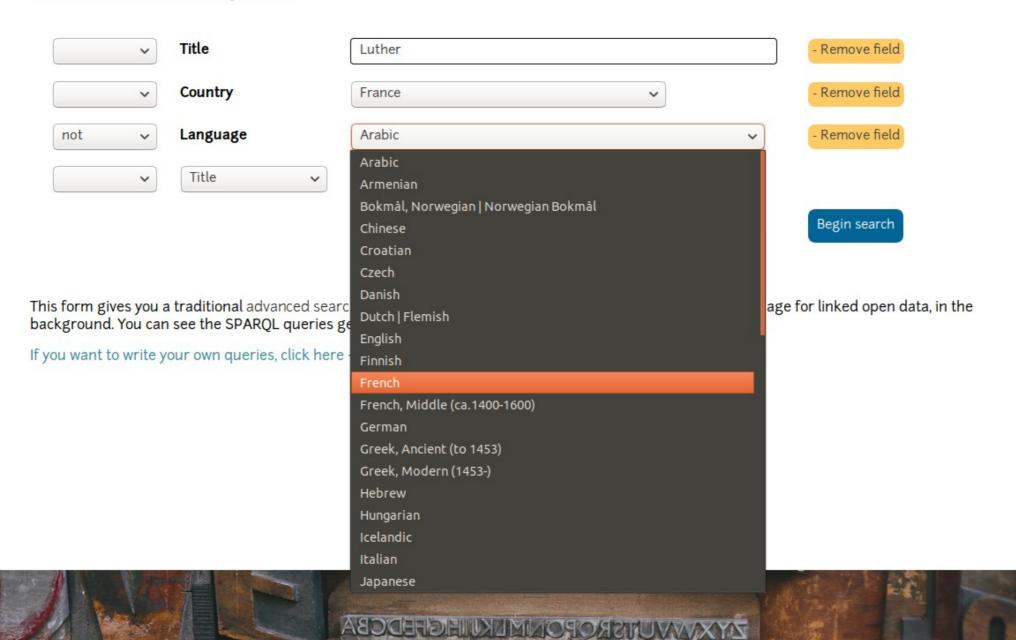


This form gives you a traditional advanced search interface, and constructs queries in SPARQL, the query language for linked open data, in the background. You can see the SPARQL queries generated by the form in the Export pane of your search results.

If you want to write your own queries, click here →









This form gives you a traditional advanced search interface, and constructs queries in SPARQL, the query language for linked open data, in the background. You can see the SPARQL queries generated by the form in the Export pane of your search results.

If you want to write your own queries, click here →

Your 3 results searching for your SPARQL query

- Exposition de l'histoire des dix lépreux [par Martin Luther]. Translatée de latin en françois... [par Antoine Du Pinet] [Texte imprimé] Luther, Martin (1483-1546)
- 2. Histoire du Lutheranisme. Maimbourg, Louis
- 3. Histoire du Lutheranisme, par le P. Louis Maimbourg, de la Compagnie de lesus.

Maimbourg, Louis (1610-1686)

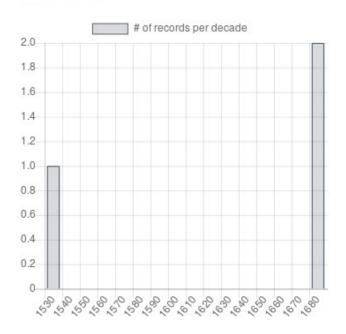


Your 3 results searching for your SPARQL query





3 items in list.



← Back

Histoire du Lutheranisme.

Author Maimbourg, Louis

Title Histoire du Lutheranisme.

CID GB-UkCyUK.CLC.CLC24878 (view in classic HPB)

Place of Publication [Rotterdam?] ♀ ❖

Publisher Suivant la copie imprimée à Paris

Year 1681

Language French ♥ ❖

Country Netherlands Q III

Holdings St Asaph Cathedral, Wales 🔍 💵

Holdings Salisbury Cathedral Q IIII

Holdings Worcester Cathedral Q IIII





← Back

Histoire du Lutheranisme.

Author Maimbourg, Louis

Title Histoire du Lutheranisme.

CID GB-UkCyUK.CLC.CLC24878 (view in classic HPB)

Place of Publication [Rotterdam?] Q ❖

Publisher Suivant la copie imprimée à Paris

Year 1681

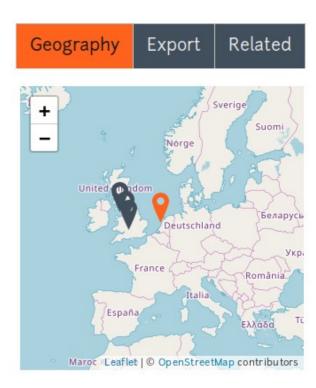
Language French Q €

Country Netherlands O

Holdings St Asaph library at Wikidata

Holdings Salisbury Cathedral Q 🛍

Holdings Worcester Cathedral Q IIII



Beyond prototypes

- Scaling up: A full database conversion currently runs at ~4 hours. We haven't tested the web application with the full database yet, as it already runs slowly with ~15.000 records (mostly due to the effort of reconstructing records from the graph)
- Improving search: Full text search in SPARQL is not very efficient.
- **Potential solution:** Combine the triple store with a document-based database to build on their respective strengths.

Beyond prototypes

- Clustering records by editions: Partially already underway, independently of the conversion to RDF
- Visualization of results: Currently implemented through third party libraries that do not entirely match our requirements (e.g., permalinks, export)
- Enrichment of data: Identify useful data sets to link to, but also demonstrate the value of large data hubs like Wikidata to researchers

Beyond prototypes

- Towards a research environment: Allow users to create their own LOD data sets that build on HPB data and edit, search and access them through the same interface
- Expanding to new audiences: Make the RDF data available to researchers who want to run large-scale analyses on our bibliographic data (cf. Lahti et al. 2019)

Thank you for your attention

I would also like to thank the DCG and CERL for hosting this project, as well as the five researchers who kindly provided their expertise.

Contact me at:

walker@sub.uni-goettingen.de

Bibliography

Dadzie, A.-S. and M. Rowe (2011). Approaches to visualising linked data. A survey. In *Semantic Web* 2(2), 89-124.

Lahti, L., J. Marjanen, H. Roivainen, and M. Tolonen (2019). Bibliographic data science and the history of the book (c. 1500-1800). In *Cataloging & Classification Quarterly* 57(1), 5-23.

Walker, A. (2019). Improving access to bibliographic data. Representing CERL's Heritage of the Printed Book database as Linked Open Data. MA thesis, Humboldt University Berlin.

