

SUB

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



CERL resources as Linked Open Data

Andreas Walker | Alexander Jahnke
LIBER 2020 Online - Linked Open Data Working Group
2020-06-23



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

CERL – The Consortium of European Research Libraries

- Consortium to **share resources** and **expertise** between research libraries to improve access to and preservation of the European printed heritage
- Originally founded to facilitate copy cataloging of Early Modern prints across European research libraries through the Hand Press Book database
- One main area of activity today is the maintenance and development of an array of **database resources** on Early Modern printing:
 - The **Heritage of the Printed Book** database (**HPB**), formerly Hand Press Book database, with ~ 8 Mio. bibliographic records for items of European printing of the hand-press period (c. 1455-1830) held at European and North American research libraries
 - The **CERL Thesaurus** (**CT**), with ~ 1.4 Mio. records containing multi-lingual information on names of printers, persons, corporations and places found in imprints and beyond, originally as a tool for search term expansion in the HPB
 - **Material Evidence in Incunabula** (**MEI**), with ~ 54.000 records of detailed provenance information for incunabula copies
 - And various others: <<https://www.cerl.org/resources/main>>

The CERL Thesaurus

CERL Search Browse History Bookmarks More... Logout

Göttingen

Record ID	cnl00029316
URI	http://thesaurus.cerl.org/record/cnl00029316
Geographic Coordinates	51.5338888888889, 9.93555555555556
Last Edit	2020-04-20

General Note

Bundesrepublik Deutschland, Bundesland Niedersachsen


More Information

Geographic Note	DE (iso3166) DE915 (nuts)
-----------------	------------------------------

Names

Heading	Göttingen <i>used in: Austrian Academy of Sciences: Bernstein Paper Atlas, Wien (Austria); Bavarian State Library, Munich (Germany)</i>
Variant Name	Gttingen <i>vermutlich Schreibfehler für Göttingen</i> Gttingen <i>vermutlich Schreibfehler für Göttingen</i> öttingen <i>Druckfehler</i> Gutingi Chöttingen Göttinge Goddinga Vicus Gotinga Göttinghe Goddinga [Sources: Orbis]

Map



General Information

[Wikipedia](#)

Related Records

- 2925 Related Records altogether
- 4 taught here (at the University, School etc.)
- 2507 were active here
- 4 lived here
- 40 studied here
- 1 Diocesan Sees / Headquarters etc.
- 362 Persons were born here
- 379 Persons died here

The CERL Thesaurus


- Established in 1999 to improve searching in the multilingual HPB, openly accessible from the beginning
- Originally in a UNIMARC-derived internal format
 - Migrated to JSON as internal format around 2018
 - First RDF/XML serialization available in 2012
 - Today: RDF/XML, RDF/TTL, JSON-LD
- Regular enrichment with links to other data providers
 - e.g. original source records, authority files (such as the German National Library's *Gemeinsame Normdatei*, VIAF)
 - Most recently: added CT identifiers to Wikidata, imported Wikidata identifiers into CT
 - Usually in the course of regular maintenance and updates from data suppliers (both harvesting and active acquisition of data sets)

The CERL Thesaurus – Challenges & Insights

- Ontology choices
 - RDF export based on best practices in 2012: mostly RDA vocabulary with domain-specific additions, has largely fallen out of use today
 - Standards only **emerge over time**, early adopters need to revisit their initial decisions
- Continuity
 - No particular LOD project for the CT – making our data accessible through current standards and technologies is an **incremental process** that is positioned within regular database enrichment, maintenance and migration
 - As technologies become available and mature, we assess researchers' needs and iterate based on their requirements, making sure to keep our technology stack **stable and supported long-term**
- Accessibility
 - Serving a largely **non-technical audience** of domain experts, we aim to add new access points to our data without losing easy-to-use interfaces integrated with their workflows

The Heritage of the Printed Book database



Search | **Results** | Advanced | Privacy | Help | Home Copyright © 2020 OCLC


search [and] [ALL] all words sort by year of publication all libraries fuzzy search 

search history shortlist **full title** first | previous | 4 | 5 | 6 | 7 | 8 | **9** | 10 | 11 | 12 | 13 | 14 | next | last | go to

results search [and] aut Schedel tit Liber | 98 hits save/print info

default Marc 21

 **HPB Identifier:** [GB-Uk.ISTC.is00311000](#) 
Cataloguing Source: The British Library, London

Title: [Liber chronicarum \[German\] Das Buch der Croniken und Geschichten](#)
Author: [Schedel, Hartmann](#)
Published: Augsburg : Johann Schönsperger, 1500
 [Augsburg \(place of printing\) in the CERL Thesaurus](#)
Country: Unknown
Language/s: German
Extent: f°
Note: Electronic facsimile : Bayerische Staatsbibliothek, München <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00079543-5>
Digital images: Bayerische Staatsbibliothek, München http://inkunabeln.digitale-sammlungen.de/Exemplar_S-199,1.html
Bibliogr. citation:
Goff S311
HC 14512*
Klebs 890.3
Schr 5207
IDL 4063
IBP 4945
CCIR S-37
Sajó-Soltész 3043
Madsen 3638
Günt(L) 84
Voull(B) 250

The Heritage of the Printed Book database

- Launched in 1994 as CERL's first major project, first hosted by RLIN, then OCLC, now VZG in Göttingen
- Initially access only for members, opened to the public in 2018
- „Traditional“ aggregated library catalogue in PICA+ format, with MARC used as import/export format
- Accessible through OPAC, Z39.50 and SRU
- 2019: First prototype for a full conversion of the database to RDF
 - Based on BIBFRAME 2.0, the emerging standard for bibliographic data
 - Interface run on top of SPARQL endpoint
 - Planned and evaluated with input from CERL's community
 - Prototype not maintained after evaluation

The Heritage of the Printed Book database

HPB

Home

Search

Results

Help

CERL →

Advanced Search Options

<input type="text" value="v"/>	Title	<input type="text" value="Luther"/>	<input type="button" value="- Remove field"/>
<input type="text" value="v"/>	Published near river	<input type="text" value="Rhine"/>	<input type="button" value="- Remove field"/>
<input type="text" value="v"/>	Title	<input type="button" value="+ Add search field"/>	

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](#) →



The Heritage of the Printed Book database

HPB Home Search Results Help CERL →

← 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 🇳🇱 🇧🇪
Holdings	St Asaph See the description of this library at Wikidata
Holdings	Salisbury Cathedral 🔍 🇬🇧
Holdings	Worcester Cathedral 🔍 🇬🇧

www.wikidata.org/entity/Q390150

Geography Export Related

The HPB – Challenges & Insights

- Ontology choices
 - Again based on current best practices, may need to be revisited
 - BIBFRAME is **geared towards librarians** as users (with an eye towards cataloging in it), may not meet the needs of researchers
- SPARQL endpoint
 - Useful for more complicated search queries, but presents technical difficulties in having to **reconstruct records** from it when building „traditional“ interfaces
 - Better to keep **record-based database infrastructure for record-based operations**, treat RDF representation as an additional layer
 - No standards for presenting SPARQL endpoints to human end users yet, questions around **usability for non-technical audiences**

The HPB – Challenges & Insights

- Data quality
 - Less of an issue in record-based systems intended for human consumption (intellectual understanding of errors, correcting search strategies)
 - Much more fatal in attempts to construct linked data, e.g., generating entities from strings
 - **Multiple languages** in the HPB due to aggregated nature
 - **Different cataloging traditions**, changing cataloging rules
 - In the CT: **manual corrections** by human editors, this is not available for HPB
- Linked data sets
 - Need to **distinguish in the interface** between links enabling searches across databases and links providing more information to human end users
 - Choice between **linking and ingestion** in an aggregated system

Future plans

- **Integration of a SPARQL endpoint** in the existing architecture, beginning with the CERL Thesaurus and then making the technology available to other databases
- Continued work on the **conversion of the HPB** with the aim of making it available as a data set in that SPARQL endpoint
- Increasing accessibility and use of our LOD resources by showcasing their abilities and collecting resources, scripts, tools etc. that can be **re-used by researchers**
- Increased **outreach and cooperation** with other data providers to find useful links between data sets and systems

SUB

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



Thank you

Contact: Andreas Walker (walker@sub.uni-goettingen.de)
More on CERL: <https://www.cerl.org>



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN