

# Donating data to Wikidata: First experiences from the „20th Century Press Archives“ (PM20)

---

*Joachim Neubert*

ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg

WikidataCon, Berlin, 25./26.10.2019

GLAM panel

# PM20 - largest public press archives worldwide



[https://commons.wikimedia.org/wiki/File:ZBW-Personenarchiv\\_2015.jpg](https://commons.wikimedia.org/wiki/File:ZBW-Personenarchiv_2015.jpg) by Max-Michael Wannags

# 20th Century Press Archives

---

25.000 thematic dossiers (c. 1908-1949) about

- General subjects and events
- Persons
- Companies
- Products

with more than 2 million pages scanned during a DFG funded project 2004-2007 and available online

# Long term sustainability?

Specialized application for discovery and access, architecturally outdated and expensive to maintain

HOME PERSONEN FIRMEN WAREN SACH KONTAKT IN EIGENER SACHE IMPRESSUM DATENSCHUTZHINWEIS ENGLISH

DIGITALISIERUNG DER PRESSEARCHIVE VON HWWA UND IFW

**Personenarchiv**

Hierbei handelt es sich um Personen des öffentlichen Lebens, etwa um Politiker, Wirtschaftsführer, Wissenschaftler, Staatsmänner, Vertreter des Militärs, des Adels, der Kirchen usw., Künstler wie Maler, Musiker, Dichter, Schriftsteller etc.

**Firmenarchiv**

Presseauschnitte zu einzelnen Firmen und Institutionen, wie Universitäten, Vereine, Institute usw. Sie wird durch Geschäftsberichte, Festschriften, Zeitschriftenaufsätze und weitere Materialien ergänzt.

**AKTUELLE ARBEITEN**

**PERSONEN- UND FIRMENARCHIV**

Beta : Dokumentenansicht über alle Archive jetzt auch über **DFGviewer** DFG-Viewer möglich. (Beispiel - Mappe zu Mahatma Gandhi)

**PERSONENARCHIV**

Suche nach Personen-Wirkungsbereiche über Baumstruktur. Siehe [Personenarchiv](#)

**PERSISTENTER LINK**

jetzt über PURL.org. Siehe [Zitierlinks und GND-Beacons](#)

**FIRMENARCHIV**

Britische Firmen abgeschlossen. Französische bis Buchstabe N

**ZBW - LEIBNIZ-INFORMATIONSZENTRUM WIRTSCHAFT**

Hervorgegangen aus der Bibliothek des Instituts für Weltwirtschaft (IfW) und des Informationszentrums des Hamburgischen Welt-Wirtschafts-Archivs (HWWA) ist die ZBW heute das weltweit größte Informationszentrum für wirtschaftswissenschaftliche Literatur.

In ihren Sammlungen beherbergt die ZBW das europaweit größte frei zugängliche Presseauschnittsarchiv, dessen Aktualisierungen Ende des Jahres

**Tipps**

**Publikationen**

**Firmenorte und Wirtschaftszweige**

**Personen-Wirkungsbereiche**

**Links**

Blog "Archive in der Leibniz-Gemeinschaft"

HamburgWissenDigital

Deutsche Biographie

# Moving PM20 to the Linked Open Data Cloud

---

## Aims of WikiProject 20th Century Press Archives

- link all folders of the collection to Wikidata, providing access to sources to Wikimedia projects and the general public
- add metadata from the folders (e.g., persons' birth dates) to the linked Wikidata items

## ZBW will maintain

- Storage of digitized images, accessible via DFG Viewer or, in future, IIIF viewers
- Static „landing pages“, which will serve as reference for the metadata integrated into Wikidata

# LOD-based infrastructure for the data donation

---

- Custom SPARQL endpoint with PM20 data
- Federated queries on that endpoint and the Wikidata Query Service
- Scripts transforming the output to QuickStatements input
- Data loading via QuickStatements, complemented with reference statements

*Blog post on ZBW Labs, with links to all query and script code:*

<http://zbw.eu/labs/en/blog/20th-century-press-archives-data-donation-to-wikidata>

# Wikidata enrichment from PM20 metadata

---

includes relations between pre-existing items,  
e.g., for Friedrich Krupp AG (Q679201):

supervisory board member	 <a href="#">Gustav Krupp von Bohlen und Halbach</a>  edit
	start time 1908
	end time 1942
	<a href="#">▼ 1 reference</a>
	 copy
	named as <a href="#">Fried. Krupp AG</a>
	stated in <a href="#">20th century press archives</a>
	PM20 folder ID <a href="#">co/009441</a>
	retrieved 5 October 2019

# First part of ZBW's data donation: The Persons Archive - completed

---



- **5266 links** (P4293) to all existing person dossiers from Wikidata items of these, 1037 items have no other external identifier
- More than **6000 statements sourced** in PM20 (“birth date”, e.g.)



# The next big challenge

---

*„Map a historic system for the organization of knowledge – materialized as newspaper clippings – **about the whole world to Wikidata**“*

# The PM20 countries/topics archive

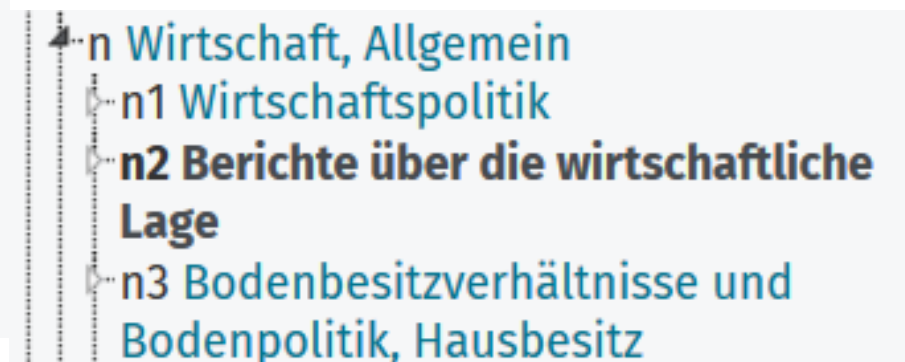
---

Folders are defined as a combination from two hierarchies

- Countries (and a few other geo entries – c. 450)
- Topical categories (c. 1,400)



C58 Togoland  
C60 Colonial Nigeria, incl. British  
Cameroons  
**C65 Cameroon**  
C70 Spanish Guinea  
C71 Sao Tome and Principe



n Wirtschaft, Allgemein  
n1 Wirtschaftspolitik  
**n2 Berichte über die wirtschaftliche  
Lage**  
n3 Bodenbesitzverhältnisse und  
Bodenpolitik, Hausbesitz

# Thanks!

---

Joachim Neubert

ZBW – Leibniz Information Centre for Economics

[j.neubert@zbw.eu](mailto:j.neubert@zbw.eu)

<http://zbw.eu/labs>

<https://www.wikidata.org/wiki/User:Jneubert>

[\*\*WikiProject 20th Century Press Archives\*\*](#)