

# Pressemappe 20. Jahrhundert goes Wikidata: Offene Daten, offene Infrastruktur

---

*Joachim Neubert*

ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg

InnoCamp, 4.11.2020



Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics

ZBW is member of the Leibniz Association

# 20th Century Press Archives



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

# Digital 20th Century Press Archives (PM20)

---

25.000 thematic folders (c. 1908-1949) about

- Persons
- General subjects and events
- Companies
- Products

More than 2 million scanned pages in these folders are available online  
(plus about 10 million scanned pages, which are due to intellectual  
property rights accessible on ZBW premises only)

---

# Long term sustainability?

Project 2004-2007:  
Specialized  
application for  
discovery and  
access,  
architecturally  
outdated and  
expensive to  
maintain

<http://webopac.hwwa.de/pressemappe20>

The screenshot shows the homepage of the Pressemappe 20. Jahrhundert website. At the top, there's a navigation bar with links to HOME, PERSONEN, FIRMEN, WAREN SACH, KONTAKT, IN EIGENER SACHE, IMPRESSUM, DATENSCHUTZHINWEIS, and ENGLISH. Below the navigation is the title "DIGITALISIERUNG DER PRESSEARCHIVE VON HWWA UND IFW". The main content area features a portrait of Mahatma Gandhi under the heading "PERSONENARCHIV". Below it is a snippet of text about public figures. To the right, a large red-bordered box contains sections on "AKTUELLE ARBEITEN", "PERSONEN- UND FIRMENARCHIV", and "FIRMENARCHIV". The "AKTUELLE ARBEITEN" section highlights the DFG-Viewer feature. The "PERSONEN- UND FIRMENARCHIV" section discusses search functionality and persistent links. The "FIRMENARCHIV" section notes the completion of British company coverage. On the right side of the page, there are sections for "Tipps", "Publikationen" (with a thumbnail of a publication cover), "Firmenorte und Wirtschaftszweige" (with a map of Berlin), "Personen-Wirkungsbereiche" (with a list of categories), and "Links" (with links to "Archive in der Leibniz-Gemeinschaft", "HamburgWissenDigital", and "Deutsche Biographie").



Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics



# Wikidata ...

---

... is a *free knowledge base* that *anyone can edit*, forming a global open knowledge graph and supporting Wikipedia's claim to allow everybody to access and to share into the sum of all knowledge

... is multilingual, has currently ~ 90 million items and information in thousands of properties (items and properties extendable by users)

... provides a mechanism to link to other open databases and web resources via „external identifier“ properties

# Data donation to Wikidata: Moving PM20 to the Linked Open Data Cloud

---

PM20 metadata (CC0) is currently integrated into Wikidata, aiming at:

- 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

# First part of ZBW's data donation: Persons Archive – completed 2019



- 5266 links from Wikidata items to all existing person folders

PM20 folder ID	pe/005823	 edit
	number of works	669
	number of works accessible online named as	377
		Gandhi, Mohandas Karamchand
	 0 references	 + add reference
		 + add value

- of these items, 1037 have no other external identifier links

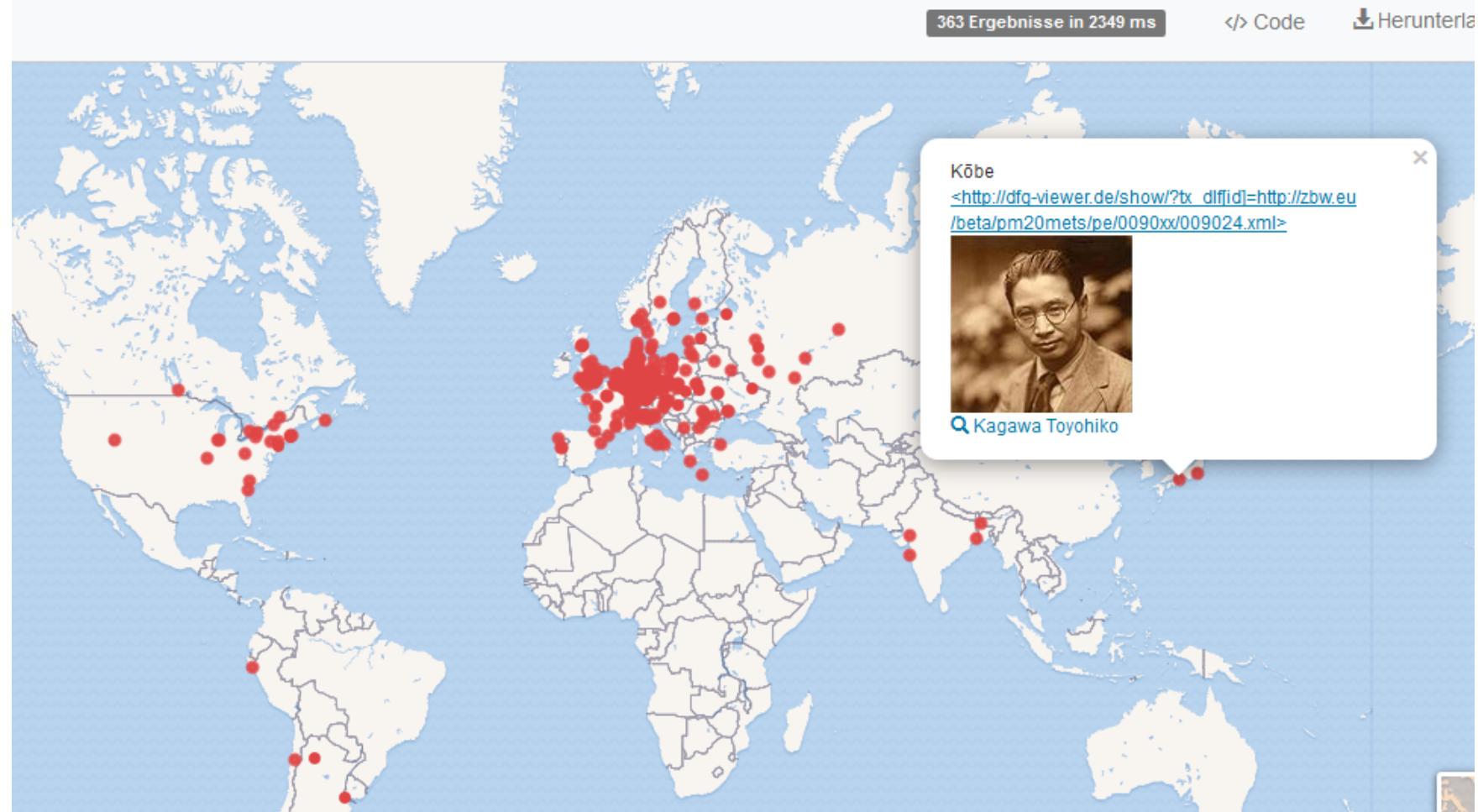
# Wikidata enrichment from PM20 metadata

---

- More than 6000 WD statements sourced in PM20; from simple facts, like “birth date”, to complex relations between pre-existing items, e.g., for Friedrich Krupp AG (Q679201):

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

# Wikidata as discovery and access layer for PM20: map of economists (WD open query interface)



[Query link](#) (on Wikidata SPARQL endpoint – see also [list of all PM20 economists](#))

# Second part of ZBW's data donation: Countries/subjects archive – currently underway

---

- This archive covered all countries and subcontinents
- ~ 9200 folders of it are digitally accessible
- each folder is defined by
  - a geographical facet (one of ~460)
  - a subject facet (one of ~1400)

Live Demo

# A PM20 folder in Wikidata

## Japan : Literature, general (Q99395498)

folder of the 20th Century Press Archives

 edit

B111 a

 In more languages

Configure

Language	Label	Description	Also known as
English	Japan : Literature, general	folder of the 20th Century Press Archives	B111 a
German	Japan : Literatur, Allgemein	Mappe aus dem Pressearchiv 20. Jahrhundert	B111 a
French	No label defined	No description defined	
Bavarian	No label defined	No description defined	

## Statements

instance of

 PM20 country/subject folder

 edit

 0 references

 + add reference

 + add value

part of

 20th Century Press Archives

 edit

 0 references

 + add reference

 + add value

*Live Demo*

# A PM20 folder in Wikidata (cont.)

main subject

Literature, general

edit

PM20 subject code

a

0 references

+ add reference

+ add value



This screenshot shows a Wikidata statement card for the main subject 'Literature, general'. The card includes fields for the subject, object (a), and type (PM20 subject code). It also features buttons for editing, adding references, and adding values.

facet of

Japan

edit

PM20 geo code

B111

0 references

+ add reference

+ add value

+ add statement



This screenshot shows a Wikidata statement card for the facet of 'Japan'. The card includes fields for the subject, object (B111), and type (PM20 geo code). It also features buttons for editing, adding references, and adding values.

## Identifiers

PM20 folder ID

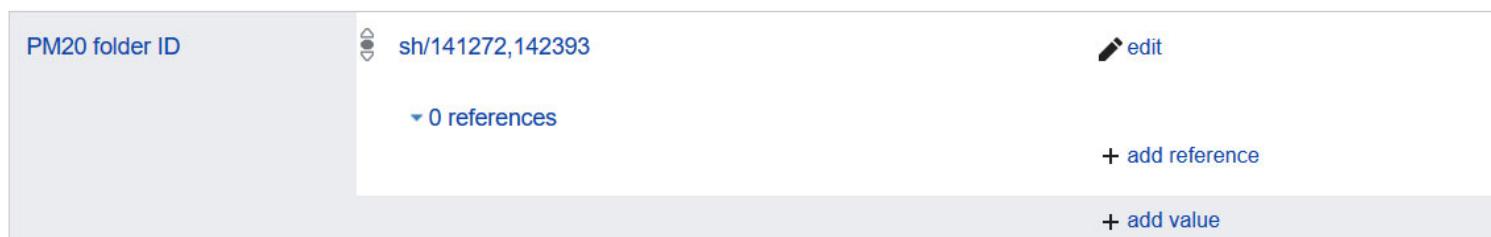
sh/141272,142393

edit

0 references

+ add reference

+ add value



This screenshot shows a Wikidata statement card for the PM20 folder ID 'sh/141272,142393'. The card includes fields for the subject, object (sh/141272,142393), and type (PM20 folder ID). It also features buttons for editing, adding references, and adding values.

*Live Demo*

# A PM20 folder in Wikidata (cont.)

main subject

Literature, general

PM20 subject code a

+ 0 references

+ add reference

+ add value



facet of

Japan

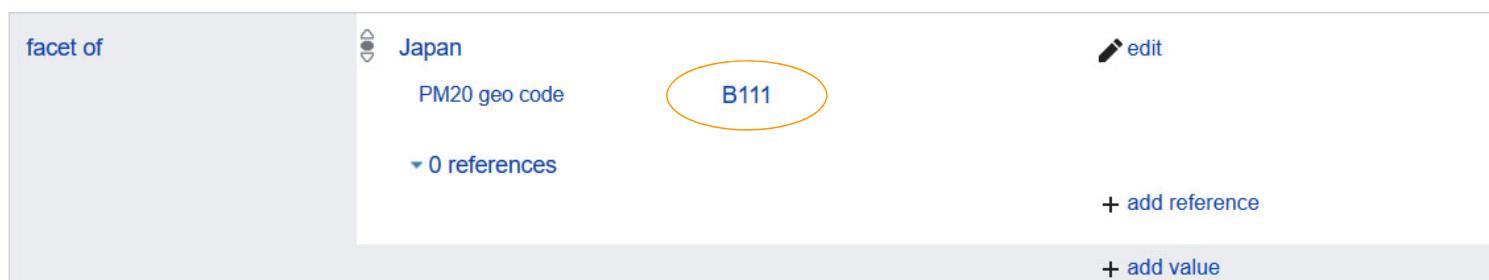
PM20 geo code B111

+ 0 references

+ add reference

+ add value

+ add statement



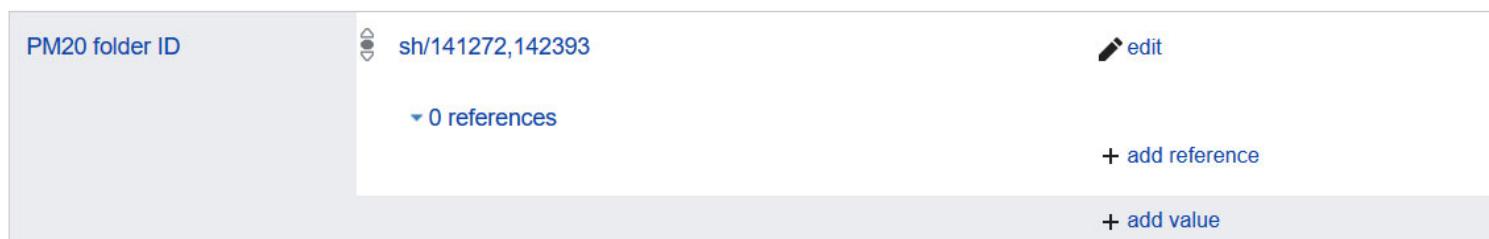
## Identifiers

PM20 folder ID sh/141272,142393

+ 0 references

+ add reference

+ add value



[back to Category Overview](#)[en](#) | [de](#)

## Hamburgisches Welt-Wirtschafts-Archiv (HWWA)

### B111 Japan

In total 731 subject folders, 36040 documents - folders complete.

*For direct access to the documents, click the “(xy documents)” link, for folder information use the “(folder)” link.*

## Subject archives

### Literature

- [a Literature, general \(15 documents\) \(folder\)](#)
- [a1 Map literature \(1 documents\) \(folder\)](#)

### The country and its people, politics and economy

- [b The country and its people, politics and economy, general \(137 documents\) \(folder\)](#)
  - [b0 Sm1 \(alt\) Country and people, series of articles \(36 documents\) \(folder\)](#)
- [b1 Country and people, travel descriptions \(7 documents\) \(folder\)](#)

### Country studies

- [c Country studies, general \(40 documents\) \(folder\)](#)
  - [c0 Sm1 \(alt\) nature reserve \(2 documents\) \(folder\)](#)
  - [c0 Sm2.I \(alt\) Cartography, general \(1 documents\) \(folder\)](#)
- [c1 Expeditions \(6 documents\) \(folder\)](#)

[back to Category Overview](#)[en](#) | [de](#)

## Hamburgisches Welt-Wirtschafts-Archiv (HWWA)

### B111 Japan

In total 731 subject folders, 36040 documents - folders complete.

*For direct access to the documents, click the “(xy documents)” link, for folder information use the “(folder)” link.*

#### Subject archives

##### Literature

- a Literature, general ([15 documents](#)) ([folder](#))
- a1 Map literature ([1 documents](#)) ([folder](#))

##### The country and its people, politics and economy

- b The country and its people, politics and economy, general ([137 documents](#)) ([folder](#))
  - b0 Sm1 (alt) Country and people, series of articles ([36 documents](#)) ([folder](#))
- b1 Country and people, travel descriptions ([7 documents](#)) ([folder](#))

##### Country studies

- c Country studies, general ([40 documents](#)) ([folder](#))
  - c0 Sm1 (alt) nature reserve ([2 documents](#)) ([folder](#))
  - c0 Sm2.I (alt) Cartography, general ([1 documents](#)) ([folder](#))
- c1 Expeditions ([6 documents](#)) ([folder](#))

Live Demo

# Folder view

The screenshot shows the DFGviewer interface. On the left, a sidebar displays a list of items under the title 'Japan : Literature, general': 'Press article 00001 (2 p.)', 'Press article 00002 (3 p.)', 'Press article 00003', 'Press article 00004', 'Press article 00005 (2 p.)', 'Press article 00006 (2 p.)', and 'Press article 00007'. The main area shows a document titled 'THE JAPAN TIMES' dated 'London, Nov 21, 1892' with the heading 'ASIAN'S GUEST OF STATE'. The document contains several columns of text. Handwritten annotations are visible on the left margin, including circled numbers 1, 2, and 3. The ZBW logo is at the top left, and a toolbar with various icons is at the top right.

# Wikidata:WikiProject 20th Century Press Archives

[Home](#)[Data Structure](#)[Typology](#)[Data Sources](#)[Use Cases](#)[Tools & Tasks](#)[Statistics](#)

The heritage of the [20th century press archives](#) (Q36948990) (PM20) is a large collection of folders with millions of press clippings and other material about persons, organizations, wares, events and general subjects. Founded in 1908 in Germany, it soon extended its coverage to newspapers all over the world. To our best knowledge, it is the world's most comprehensive public press archives and thus a unique source for politics, business and social life in the 20th century.

## Contents [hide]

- [1 Aim and scope](#)
- [2 Background](#)
- [3 Presentations](#)
- [4 Next steps](#)
- [5 How to contribute](#)
- [6 Participants](#)
- [7 Related WikiProjects](#)

## Aim and scope [edit]

This WikiProject strives to

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

# Wikidata als offene Infrastruktur – beitragen und nutzen

---

- Die Pressemappen und ihre Artikel ergänzen Wikidata mit Fakten und zahlreichen zeitgenössischen Quellen (auch für WP-AutorInnen)
  - Wikidata bietet Kontext für die einzelnen Mappen – vertrauenswürdig, umfassend aktualisiert und multilingual (mit Links zu Wikipedia in der Sprache der User)
  - Die Last der Bereitstellung eines Discovery- und Accesslayers verlagert sich von der ZBW, wo PM20 eine Spezialanwendung war, zu Wikidata, wo PM20 ein kleiner Teil einer offenen Infrastruktur ist
  - PM20 wird damit Teil eines Global Knowledge Graph – jede Mappe ein Knoten im Web of Linked Open Data
-

# Thanks!

---

Joachim Neubert

ZBW – Leibniz Information Centre for Economics

j.neubert@zbw.eu

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

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