

# Linking the 20th century paper history to the sum of all knowledge

---

*Joachim Neubert*

ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg

DCMI Virtual 2020, 22.9.2020

# 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outdated and  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

HOME PERSONEN FIRMEN WAREN KONTAKT IN EIGENER SACHE IMPRESSUM DATENSCHUTZHINWEIS ENGLISH

SACH

DIGITALISIERUNG DER PRESSEARCHIVE VON HWWA UND IFW

Pressemappe  
20. Jahrhundert

**Tipps**

Publikationen

Firmenorte und  
Wirtschaftszweige

Personen-  
Wirkungsbereiche

Links

Blog "Archive in der  
Leibniz-Gemeinschaft"

HamburgWissenDigital

Deutsche Biographie

**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

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

The screenshot shows a Wikidata item page for 'pe/005823'. On the left, there is a label 'PM20 folder ID'. The main content area displays the following information:

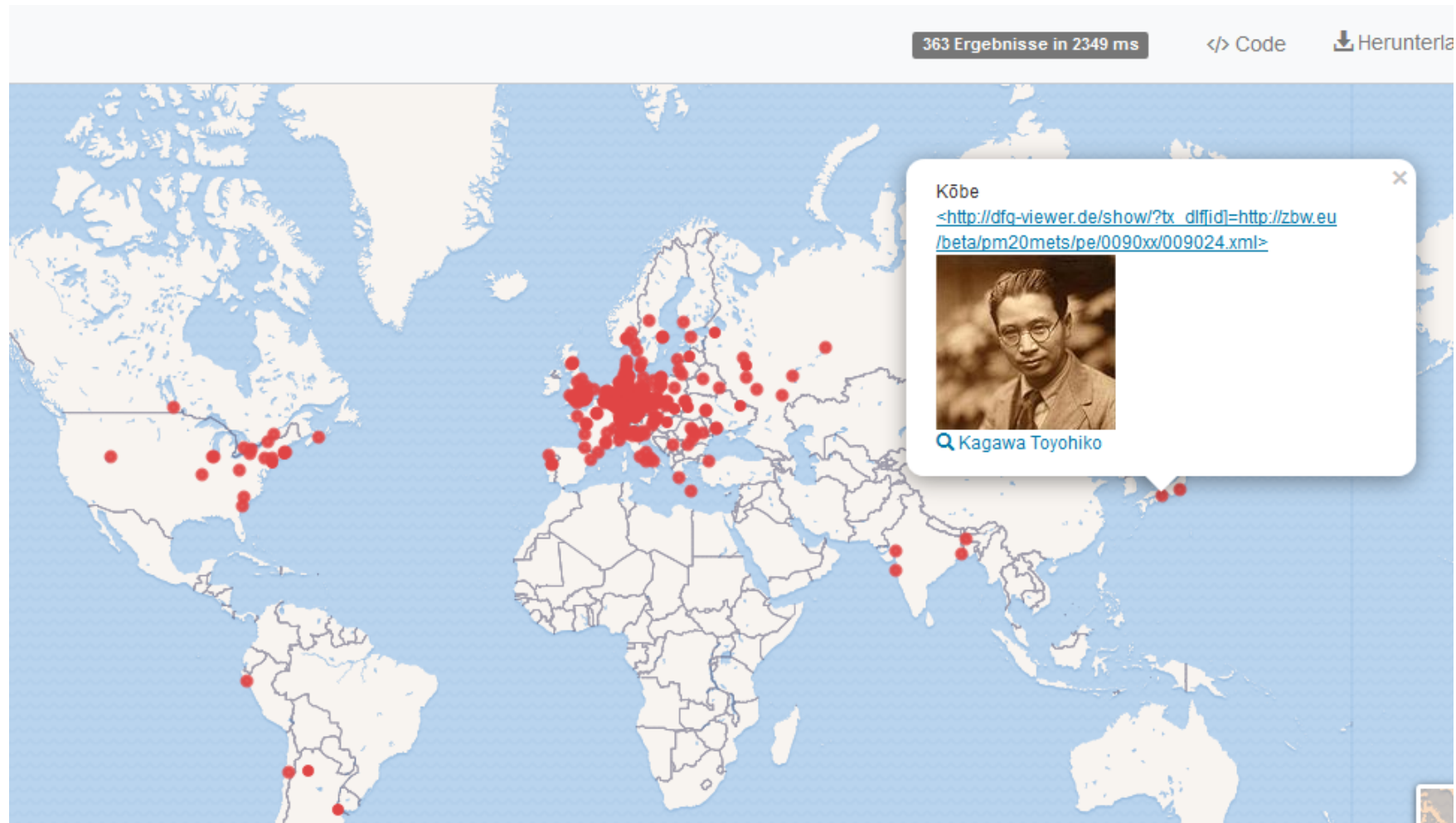
<b>number of works</b>	669
<b>number of works accessible online</b>	377
<b>named as</b>	Gandhi, Mohandas Karamchand

Below this, it indicates '0 references'. At the bottom right, there are two buttons: '+ add reference' and '+ add value'. An 'edit' button is located at the top right of the main content area.

- More than 6000 WD statements (“birth date”, e.g.) sourced in PM20
  - including complex relations (e.g., a person was board member of a company for a certain timespan)

# Wikidata as discovery and access layer for PM20: map of economists

---



[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)

# A PM20 folder in Wikidata

---

## Japan : Literature, general (Q99395498)

folder of the 20th Century Press Archives  
B111 a



 edit

[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](#)

# A PM20 folder in Wikidata (cont.)




---

main subject	 Literature, general  edit
	PM20 subject code      a
	 0 references
	<a href="#">+ add reference</a>
<a href="#">+ add value</a>	

facet of	 Japan  edit
	PM20 geo code      B111
	 0 references
	<a href="#">+ add reference</a>
<a href="#">+ add value</a>	

[+ add statement](#)

## Identifiers

PM20 folder ID	 sh/141272,142393  edit
	 0 references
	<a href="#">+ add reference</a>
	<a href="#">+ add value</a>

## 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 \(6 documents\) \(folder\)](#)

## 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)

# Folder view

The screenshot displays the DFGviewer interface. On the left, a dark blue sidebar contains the 'DFGviewer' logo, language options 'DE / EN', and a menu icon. Below this, the title 'Titel: Japan : Literature, general' is shown. A scrollable list of items is visible, including 'Japan : Literature, general' and several 'Press article' entries with their respective page counts. The main area features the 'ZBW' logo and a set of navigation icons (link, download, book, grid). A document preview is shown in the center, displaying a page with dense text and handwritten annotations. The preview includes a large number '7' at the top, a date '1899', and a signature. The document title is 'THE JAPAN REVIEW' and the subtitle is 'LITERATURE OF JAPAN'. Navigation arrows are present on the left and right sides of the preview area. At the top right of the main area, there are icons for page selection (showing '[1]'), refresh, zoom in, zoom out, and share.

# A PM20 folder in Wikidata (cont.)



---

main subject	 Literature, general  edit
	PM20 subject code a
	▼ 0 references
	+ add reference
+ add value	

facet of	 Japan  edit
	PM20 geo code B111
	▼ 0 references
	+ add reference
+ add value	

+ add statement

## Identifiers

PM20 folder ID	 sh/141272,142393  edit
	▼ 0 references
	+ add reference
	+ add value

Rich historically grown knowledge organization system on index cards and in SKOS

## Skosmos

### HWWA subject category system

Alphabetical

Hierarchy

- g4a Economic relations with individual countries
- g5 Colonial policy, general
  - g5 Sm1 . Claims to foreign colonies
  - g5 Sm100 Colonial policy, history
  - g5 Sm20 Colonial policy, conferences
  - g5 Sm40 Colonial policy; education
  - g5 Sm50 Colonial policy, institutions
  - g5 Sm501 (A10) . Chartered Company question
  - g5 Sm501 (A22) . British Colonial Conference
  - g5 Sm501 (H) . Internationalization of the colonies
  - g5 Sm502 (A10) . Colonial mandates
  - g5 Sm502 (H) . Right of the natives to self-determination

g<sup>5</sup> 506:  
1 Ansprüche auf fremde Kolonien.

g<sup>6</sup> 506:  
1 Internationalisierung (Zentral)  
2 Verhandlungen und Abmachungen fremder Staaten über... (Hinweise g<sup>5</sup> 506 18)  
3 Interessen einzelner Staaten in... (Hinweise g<sup>5</sup> 506 18)

A10 Deutschland  
501 „Chartered Company“ Frage  
502 Kolonialmandate

A22 England:  
501 Britische Kolonialkonferenz

g<sup>5</sup> Kolonialpolitik, allg.



# Subject category hierarchy in Wikidata

- Items of class „PM20 subject category“, e.g.
  - a Literature, general
    - a1 Map literature
- Hierarchy via „part of“ / „has part“ properties

**Literature, general** (Q99312062) [edit](#)

subject category of the 20th Century Press Archives [edit](#)  
a Literature, general

[In more languages](#)  
Configure

Language	Label	Description	Also known as
English	Literature, general	subject category of the 20th Century Press Archives	a Literature, general
German	Literatur, Allgemein	Systematikstelle des Pressearchiv 20. Jahrhundert	a Literatur, Allgemein
French	No label defined	No description defined	
Bavarian	No label defined	No description defined	

**Statements**

instance of PM20 subject category [edit](#)  
[0 references](#)  
[+ add reference](#)  
[+ add value](#)

part of PM20 subject category system [edit](#)  
[0 references](#)  
[+ add reference](#)  
[+ add value](#)

main subject literature [edit](#)  
[0 references](#)  
[+ add reference](#)  
[+ add value](#)  
[+ add statement](#)

**Identifiers**

PM20 subject code a [edit](#)

# 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 query output to input for Wikidata's QuickStatements bulk data loading tool
- Use of Mix-n-Match and of OpenRefine for matching to existing items

*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 community

---

- Everybody can participate; no „central committee“ making decisions
- Become part of the community with an individual user account, and disclose your affiliation
- Learn about the rules, in particular Wikidata’s Notability policy
- Follow events and discussions via mailing list or the „Weekly Summary“
- Discuss data imports beforehand, e.g. in the „Project chat“ online forum
- Consider creating a Wikiproject, inviting others to join

# 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

[https://www.wikidata.org/wiki/Wikidata:WikiProject 20th Century Press Archives](https://www.wikidata.org/wiki/Wikidata:WikiProject_20th_Century_Press_Archives)

# 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>