

# Entity Management in JudaicaLink

DCMI Panel “Entity Management”  
Oct. 14th, 2021

Kai Eckert

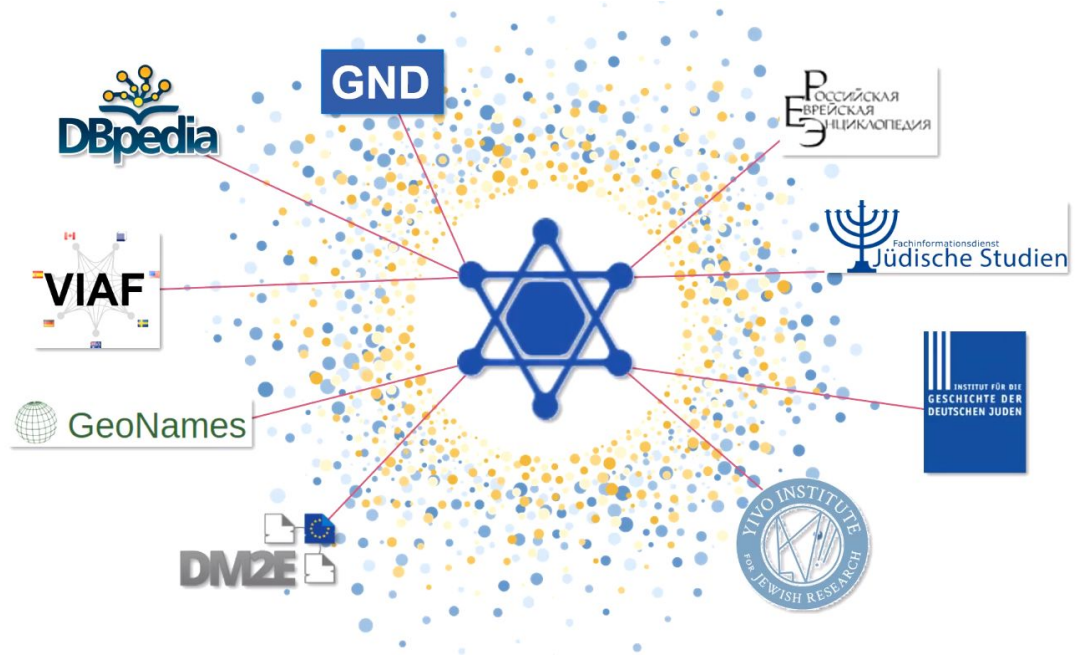
# JudaicaLink

A **Knowledge Graph** for Jewish Studies (in its broadest sense).

We collect and integrate data from **various sources**.

**Entities** described in multiples sources are **linked**.

Links can already **exist** in a source, can be **added intellectually** or **automatically**.



# Data Integration Workflows

We collect from various formats and sources.

In decreasing appeal and fun:

- RDF / Linked Data
- Web-Scraping
- APIs
- Databases
- “Databases” (think Excel-Files and worse)
- Text
- PDFs

GO TO: YIVO INSTITUTE FOR JEWISH RESEARCH VISIT A CJP PARTNER WEB SITE

ייוועניקלאָפּעדיע  
פֿון די ייִדן אין מזרח־אייראָפּע

Search the Encyclopedia SEARCH

HOME BROWSE MAPS IMAGES, AUDIO & VIDEO DOCUMENTS FOR EDUCATORS FOR RESEARCHERS HELP

PREVIOUS ARTICLE: Minorities Treaties NEXT ARTICLE: Minskoi, Nikolai Maksimovich

## Minsk

Contents Hide

- \* Suggested Reading
- \* Author

Capital of the Republic of Belarus since 1991. From the beginning of the fourteenth century, Minsk was part of Lithuania; from the mid-sixteenth century it belonged to the Polish-Lithuanian Commonwealth. In 1793, the city was annexed to the Russian Empire and became the capital of Minsk province. From 1920 to 1991, Minsk served as the capital of the Belorussian Soviet Socialist Republic (BSSR).

In 1489, a Jew named Mikhl Danilevich of Troki leased the rights to the custom duties of Minsk, and Jews began to settle in the city during the sixteenth century. In 1579, King Stefan Batory granted Minsk's Jews a privilege or charter allowing them to engage in commerce in the city; in 1606, at the request of the Christian population, King Sigismund III invalidated this charter but by 1629 reinstated the Jews' commercial rights, allowing them to open shops. In 1633, King Wladyslaw IV granted the Jewish community permission to buy land for a new cemetery and acquire real estate on the market square. In 1653, the leaders of the Jewish community attended the first meeting of the Lithuanian Va'ad (Council) as representatives of an independent community; this despite the fact that until 1631 the Minsk community was subordinated to the Brest community. During the Russian-Polish war (1654-1667), Minsk was occupied by Russian troops and the majority of Jews left in 1655; but as soon as Minsk was returned to the commonwealth (in 1668), the Jewish community was reestablished.

Members of the Bund on a picnic, Minsk, 1910. (YIVO)

SEE MEDIA RELATED TO THIS ARTICLE

EXTERNAL LINKS

Find more information about

Title: **Minsk**

Description: Capital of the Republic of Belarus since 1991. From the beginning of the fourteenth century, Minsk was part of Lithuania; from the mid-sixteenth century it belonged to the Polish-Lithuanian Commonwealth. In 1793, the city was annexed to the Russian Empire and became the

Links in text:  
"surface form" => concept

Images:

Members of the Bund on a picnic, Minsk, 1910. (YIVO)

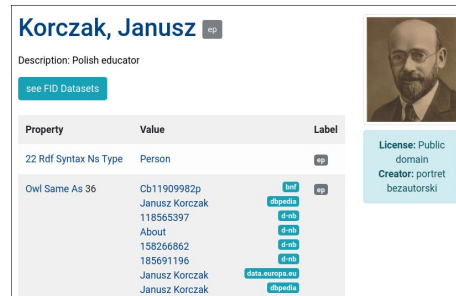
# Entity Pages: one URI per entity

## Janusz Korczak

- 36 internal and external links
- Internal:
  - [http://data.judaicalink.org/data/dbpedia/Janusz\\_Korczak](http://data.judaicalink.org/data/dbpedia/Janusz_Korczak)
  - [https://data.judaicalink.org/data/ubffm/Janusz\\_Korczak](https://data.judaicalink.org/data/ubffm/Janusz_Korczak)
  - [http://data.judaicalink.org/data/yivo/Korczak\\_Janusz](http://data.judaicalink.org/data/yivo/Korczak_Janusz)
  - <http://data.judaicalink.org/data/gnd/118565397>
  - ...

**Challenge:** Give each cluster of linked resources a persistent URI and maintain the URIs when links get added or deleted.

Basically: how to maintain an authority file without human resources?

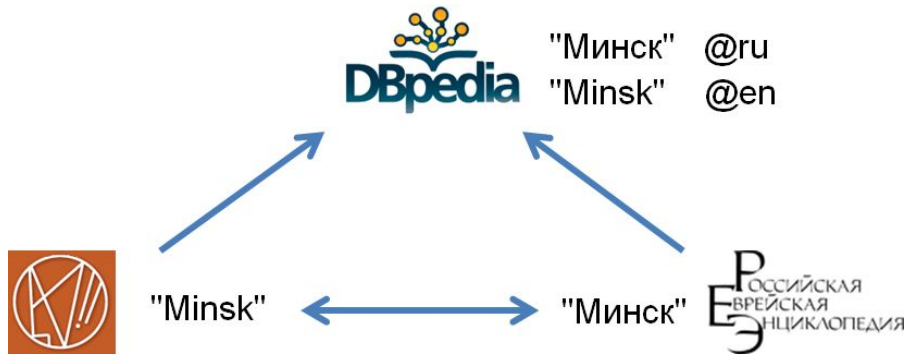


The screenshot shows a Wikidata entity page for 'Korczak, Janusz'. It includes a description 'Polish educator', a 'see FID Datasets' button, and a table of properties. The table lists '22 Rdf Syntax Ns Type' as 'Person' and 'Owl Same As 36' with various identifiers like 'Cb11909982p', '118565397', and '185691196'. It also shows a portrait and license information.

Property	Value	Label
22 Rdf Syntax Ns Type	Person	
Owl Same As 36	Cb11909982p Janusz Korczak 118565397 About 185691196 Janusz Korczak Janusz Korczak	

# Example: Cross-lingual linking

Cross-lingual integration: The **more sources** we have, the **more links** we get - also indirectly.



Automatic linking inevitably leads to **mistakes**. Even worse: wrong links might propagate via several data sources.

# How to create entity URIs?

**Step 1:** Calculate **transitive closure** of owl:sameAs links to identify all clusters of same entities.

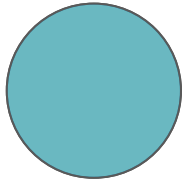
**Step 2: Assign a URI** to each cluster.

**Step 3: Never** touch your data again!

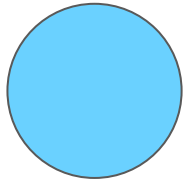
# How to maintain entity URIs?

Step 1: Calculate transitive closure of owl:sameAs links to identify all clusters of same entities.

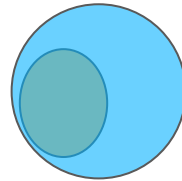
Step 2: Iterate over all clusters and apply rules according to overlaps with **existing clusters with URI**:



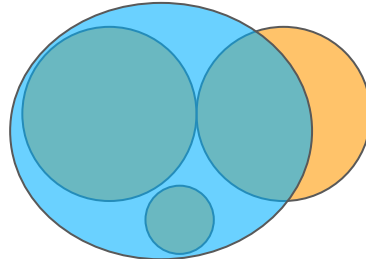
Cluster is **unchanged**: Keep URI



Cluster only contains **new resources**: Assign new URI



Cluster contains **one** existing cluster: Add new resources to existing cluster, keep URI.



**Partly overlapping clusters**: Split.

**Containing multiple old clusters**: Merge.

# Split and Merge

**Merge:** Keep the URI of the largest cluster, create redirects (jl:wasMerged) for other cluster URI.

**Split:** Create new URIs for all clusters, create redirect (jl:wasSplit) for old cluster URI.

## Strategy:

- **Minimize overall changes** of assigned entity (cluster) URI.
- When **semantic ambiguity** can not be resolved (e.g. when a cluster is split), assign new URIs, keep the old one as ambiguous URI that requires disambiguation for downstream applications.

# Lessons learned so far

- For consistency and piece of mind, each integrated resource needs to be associated with an entity URI.
  - We wanted a consistent way all our resources are presented (easier to search and to navigate).
  - This leads to many single resource clusters with one dataset specific resource and a entity page.
  - If desired, it can be avoided: handle unlinked resources like clusters. If multiple unlinked resources get merged: assign a new URI.
- Avoid wrong links when you start; to avoid cluster splitting... but:
- Create links first; to avoid countless merges later.

# Our Team



Kerstin von der Krone

Lead FID Jewish Studies  
k.vonderkrone@ub.uni-frankfurt.de



Kai Eckert

Lead JudaicaLink  
Project partner FID  
eckert@hdm-stuttgart.de



Sara Fischer

Research Assistant



Dov Winer

Co-Founder  
JudaicaLink,  
Advisory



Rachel Heuberger

Former Lead FID Jewish Studies  
r.heuberger@ub.uni-frankfurt.de



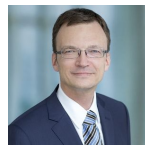
Patricia Kaluza

Research Assistant



Marko Knepper

Head of IT, University Library Mainz  
Former Co-Lead FID, Advisory



Thomas Risse

Head of IT, Frankfurt  
t.risse@ub.uni-frankfurt.de



Annette Sasse

Librarian  
a.sasse@ub.uni-frankfurt.de



Jonathan Kaplan

Research Assistant  
j.kaplan@ub.uni-frankfurt.de



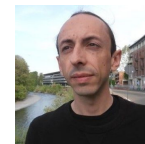
Hanna-Barbara Rost

Research Assistant



Benjamin Schnabel

Research Assistant  
schnabel@hdm-stuttgart.de



Marco Rovera

Former Research Assistant  
rovera@hdm-stuttgart.de



Aaron Christianson

Research Assistant



Maral Dadvar

Former Research Assistant  
dadvar@hdm-stuttgart.de