Der BERD@BW-Wissensgraph für offene Firmendaten mit Wikibase

FAIR knowledge-graph-based research data management

Dr. Renat Shigapov, Dr. Irene Schumm, Markus Herklotz & Lars Oberländer

Mannheim University Library, 04.03.2021

Outline

Motivation

Part 1: Creating

Part 2: Scaling

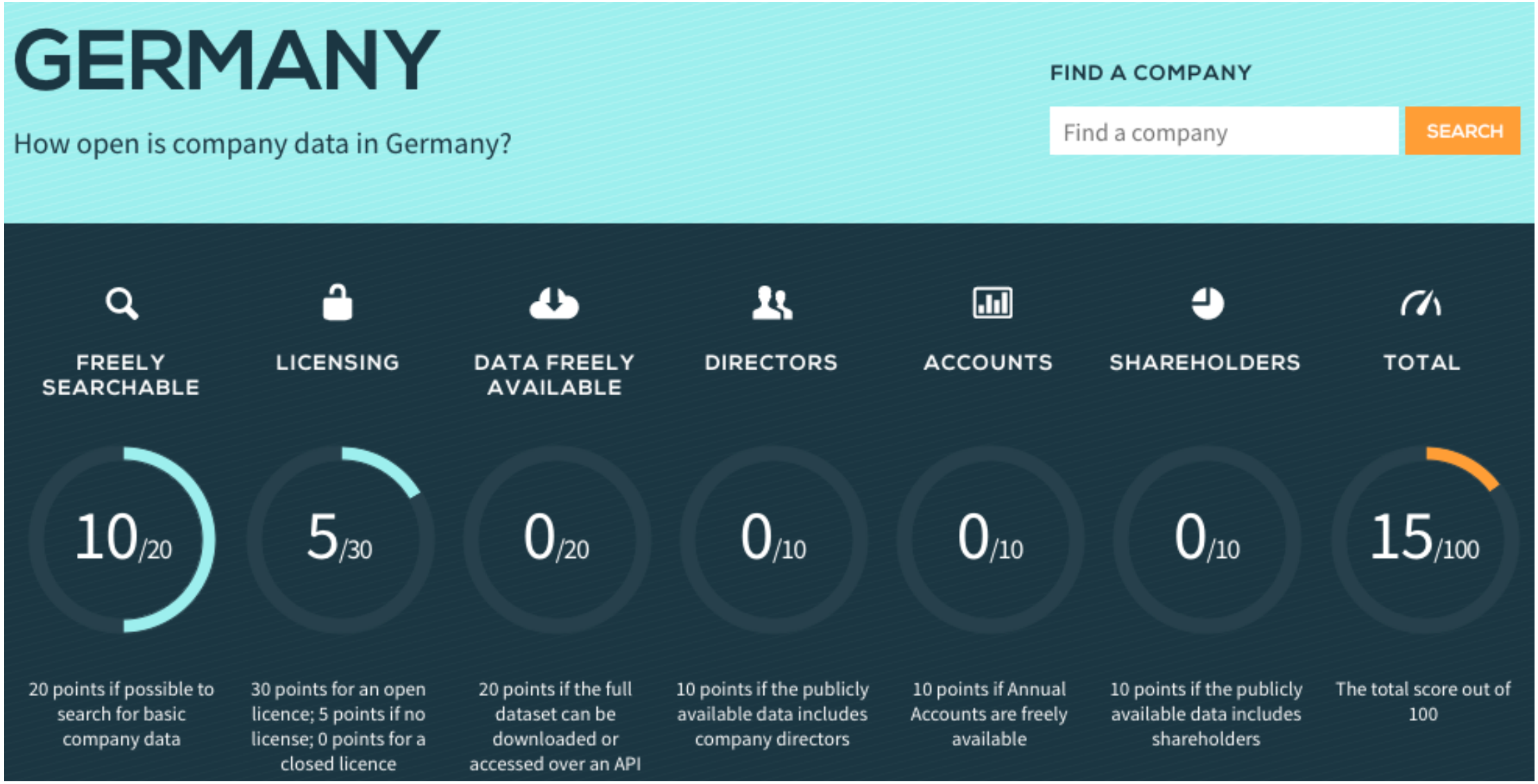
Part 3: Using

Conclusions

instance of

15 100

The open company data index for Germany is 15/100

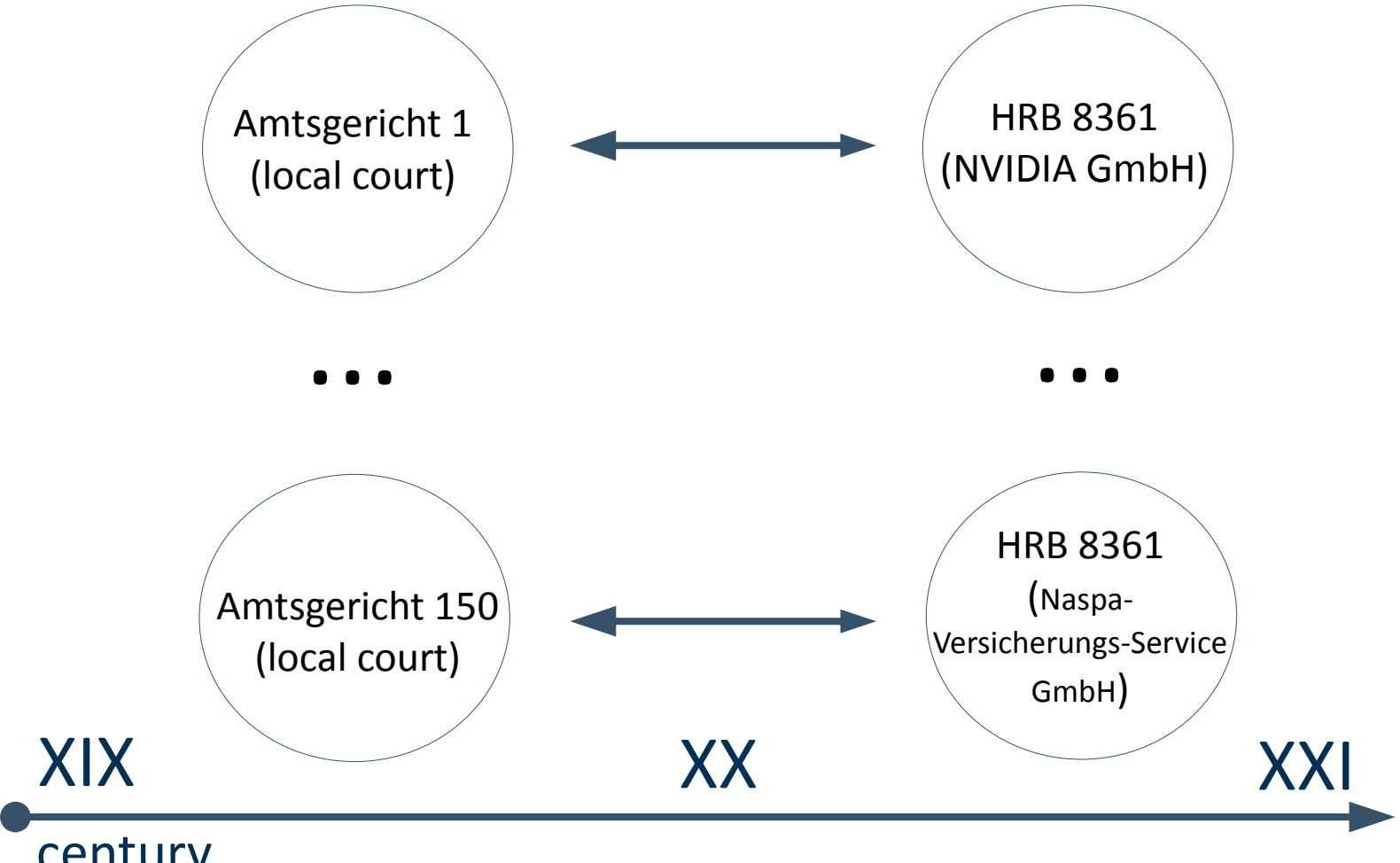


Source: http://registries.opencorporates.com/jurisdiction/de

The curse of German company data

Many registration authorities

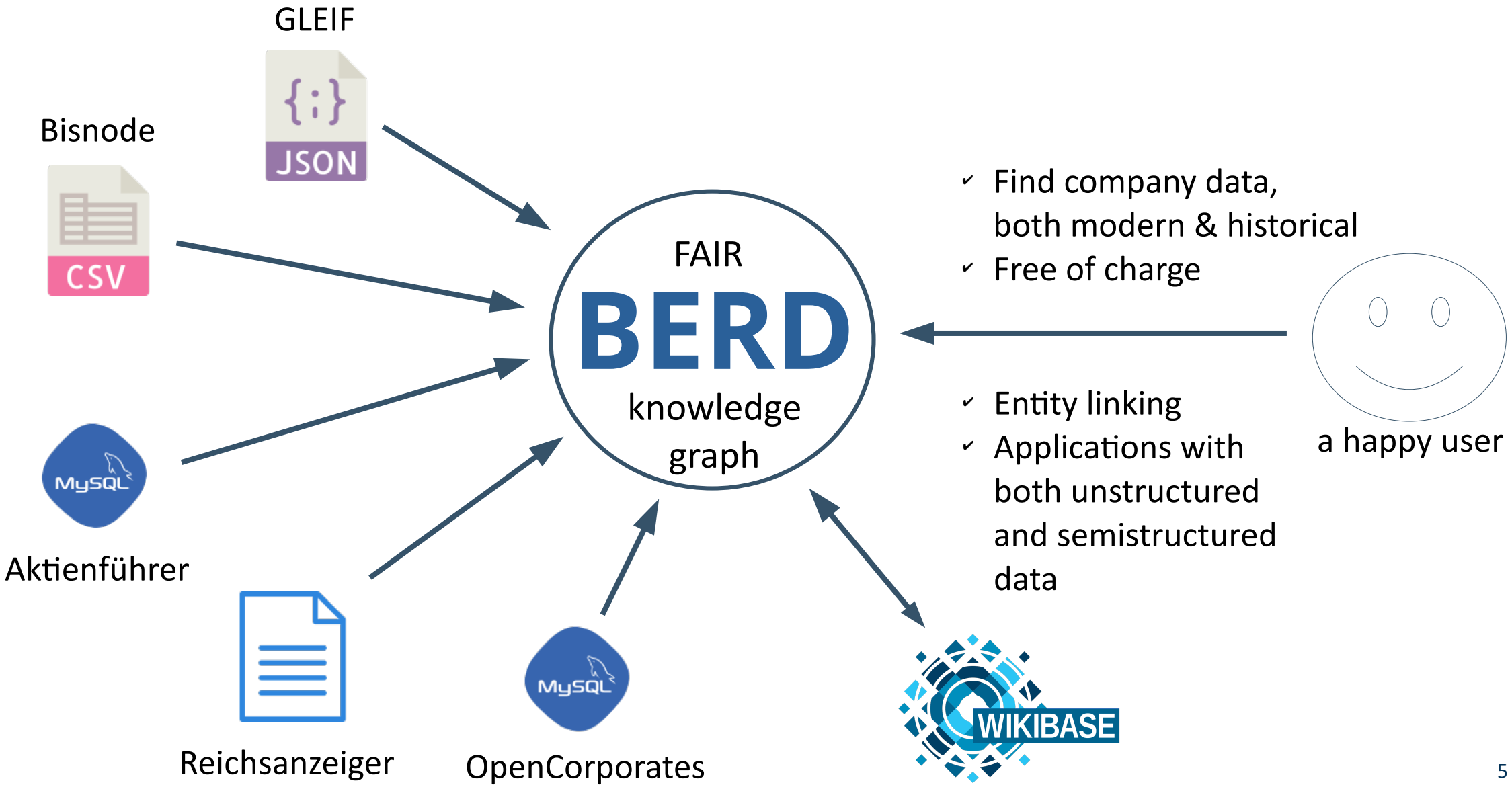
Non-unique company identifiers



The historical & modern German company data ARE NOT:

- ☹ Findable
- ☹ Accessible
- ☹ Interoperable
- ☹ Reusable

Solution: BERD knowledge graph of German companies



Outline

Motivation

Part 1: Creating

Part 2: Scaling

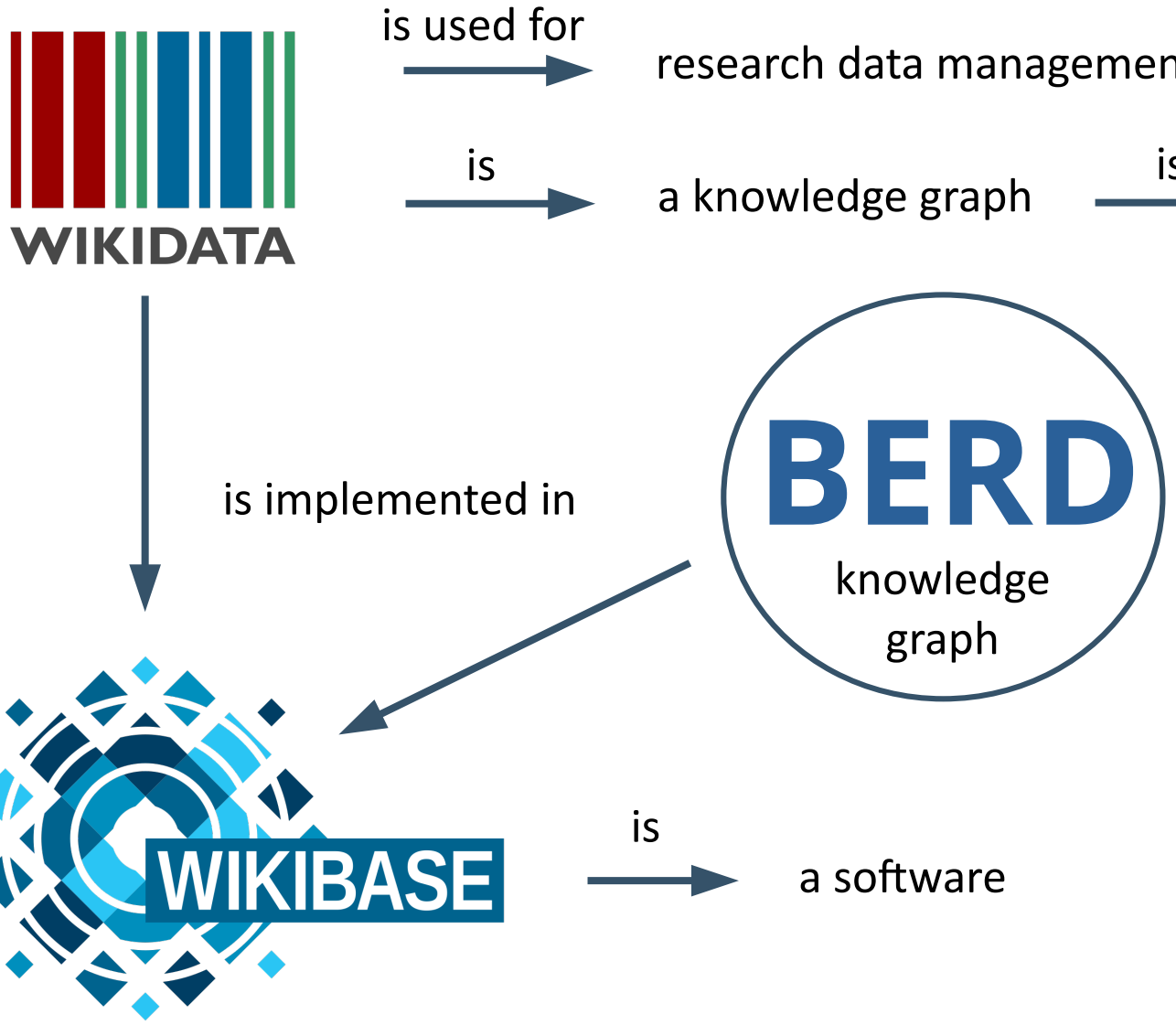
Part 3: Using

Conclusions

instance of

15 100

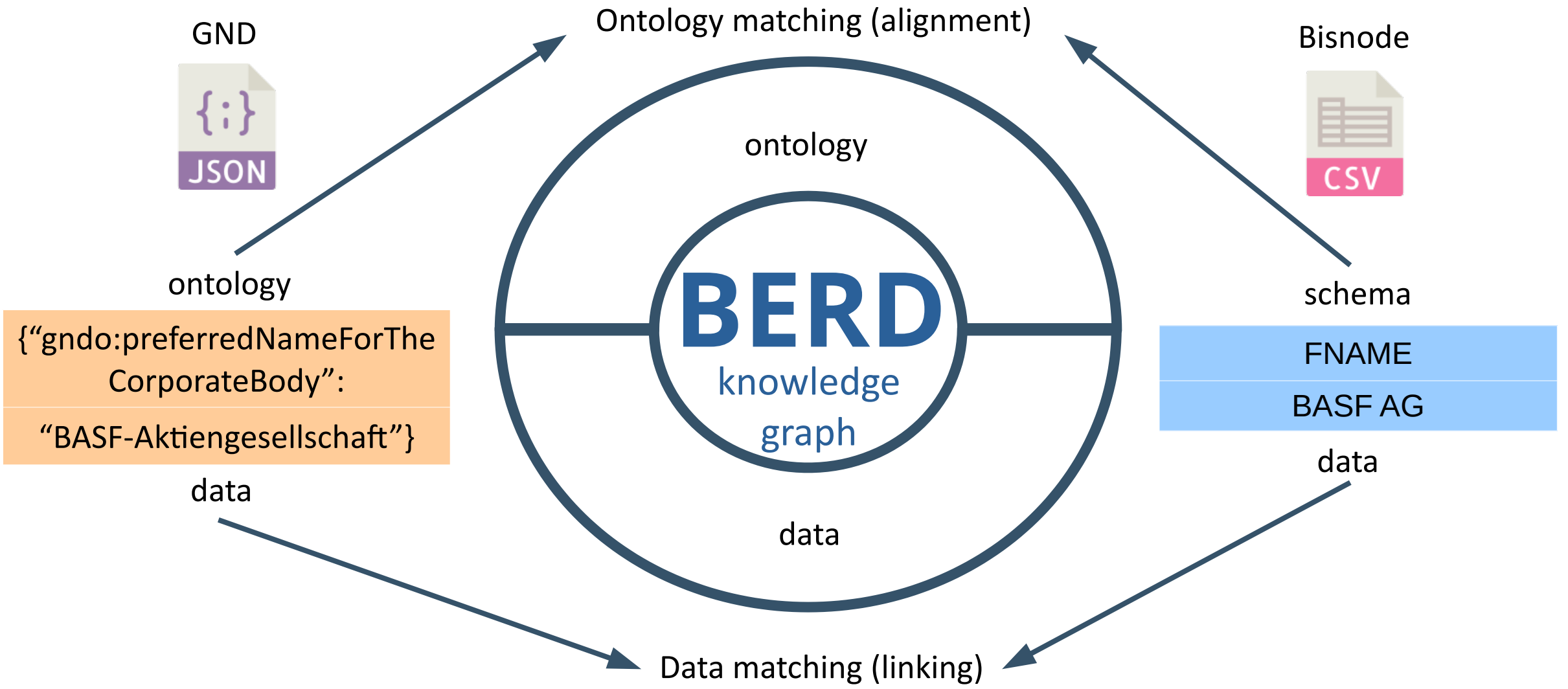
Wikibase, Wikidata & knowledge graphs



Why knowledge-graph-based research data management?

- ✔ State-of-the-art technology
- ✔ Reach data modeling
- ✔ Statement-level provenance
- ✔ Unstructured & structured data
- ✔ 18 datatypes
- ✔ Data science applications

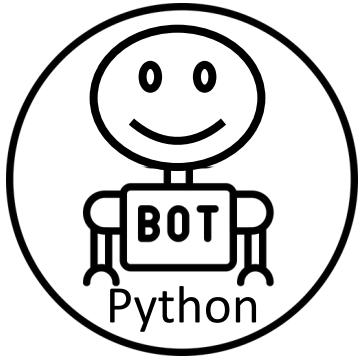
Knowledge graph = ontology + data



Filling BERD ontology and BERD data via Wikibase API

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

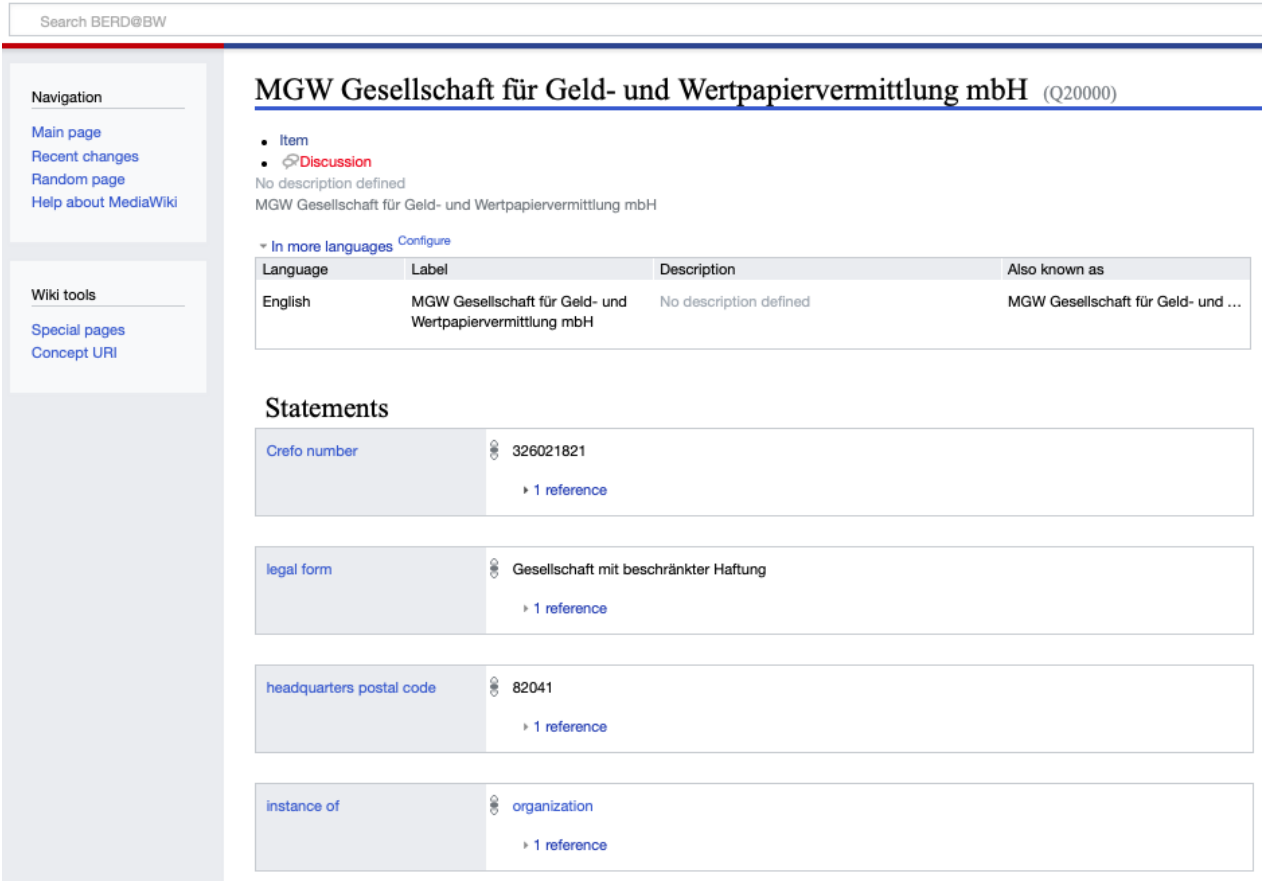
Company page
RDF export



Wikibase API

BERD knowledge graph

SPARQL endpoint
Data analysis and export



Outline

Motivation

Part 1: Creating

Part 2: Scaling

Part 3: Using

Conclusions

instance of

15 100

Towards a large-scale knowledge graph

UB
UNIVERSITÄTSBIBLIOTHEK
MANNHEIM

BERD
knowledge
graph

10000-100000 entities

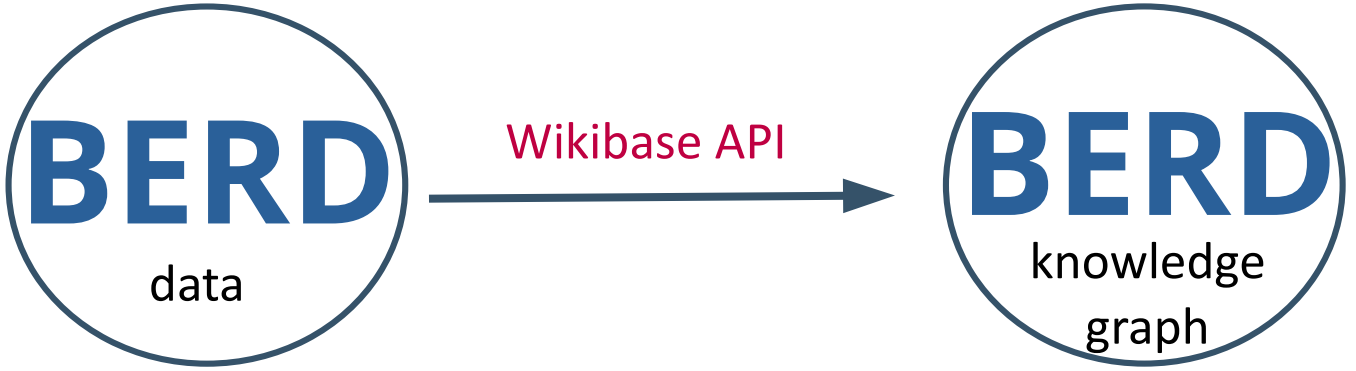
1000000-10000000 entities

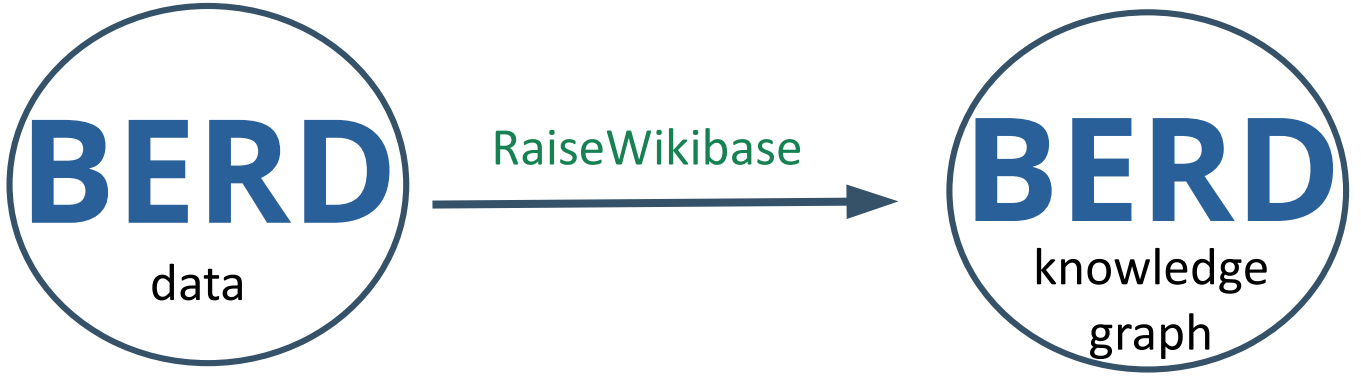
- More ontology alignment
- More data matching
- More data integration
- More data filling

A small problem with big data

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

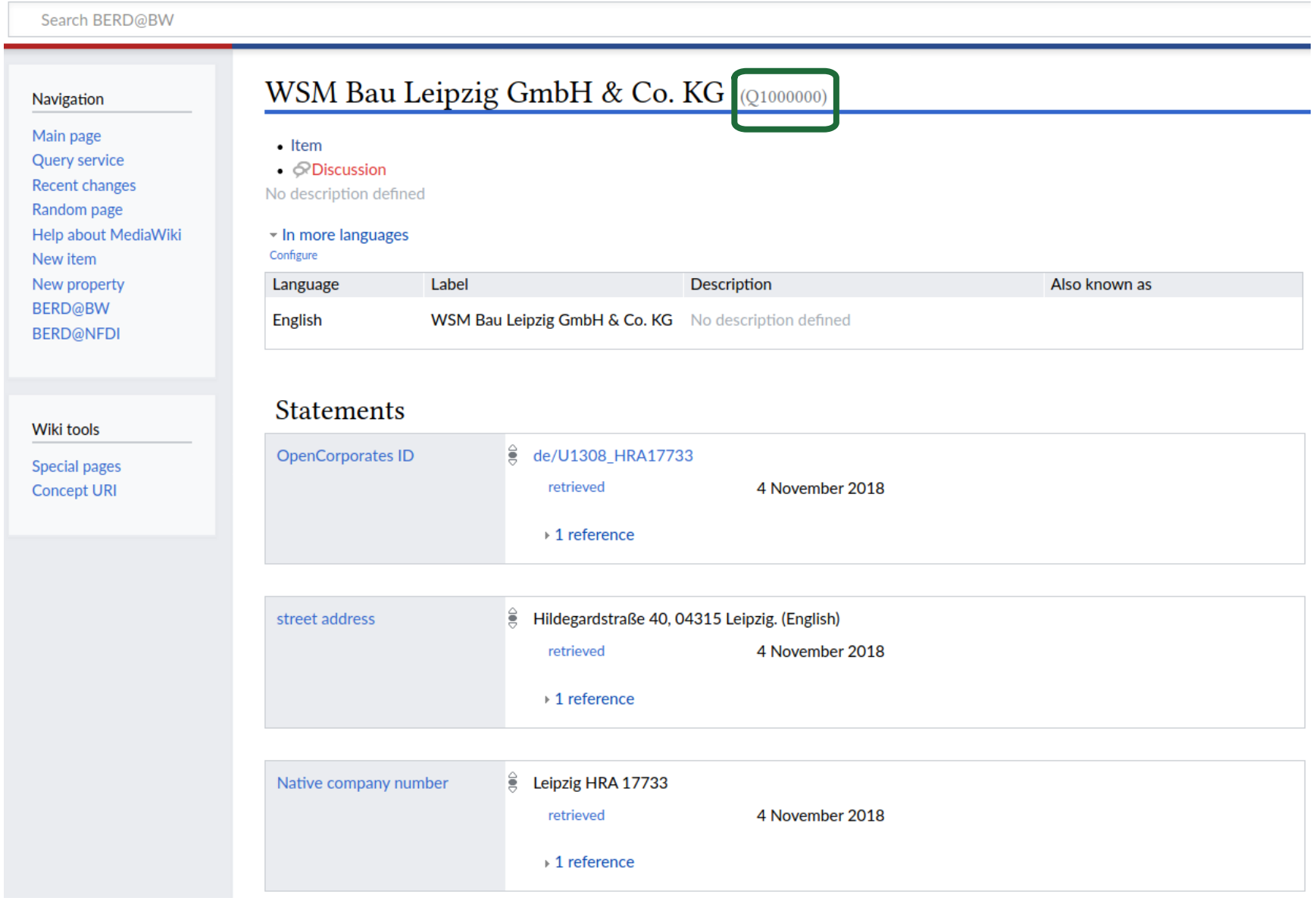
A million edits of entities per 2-12 days





A million edits of entities per 1-2 hours

Towards the large-scale BERD knowledge graph



Outline

Motivation

Part 1: Creating

Part 2: Scaling

Part 3: Using

Conclusions

instance of

15 — 100

Wikibase as a backend for search services

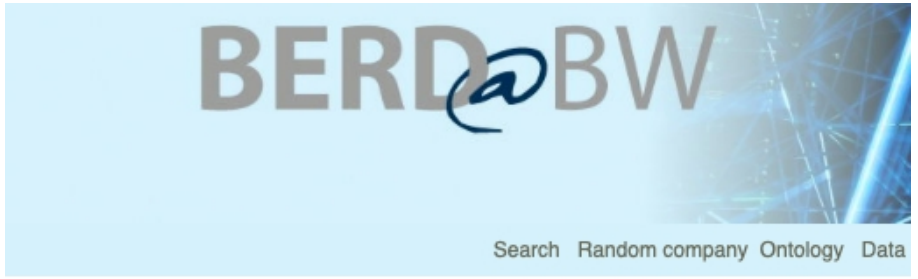
UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

BERD
knowledge graph

company | state | city | address | Search

LEI | ISIN or WKN | Creditreform-ID | Handelsregister

min. employees | max. employees | legal form | SIC industry



Graubner Industrie-Beratung

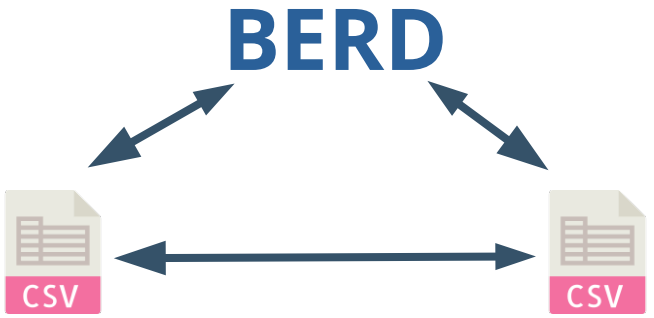
Copy | CSV | Excel | PDF | Print | JSON

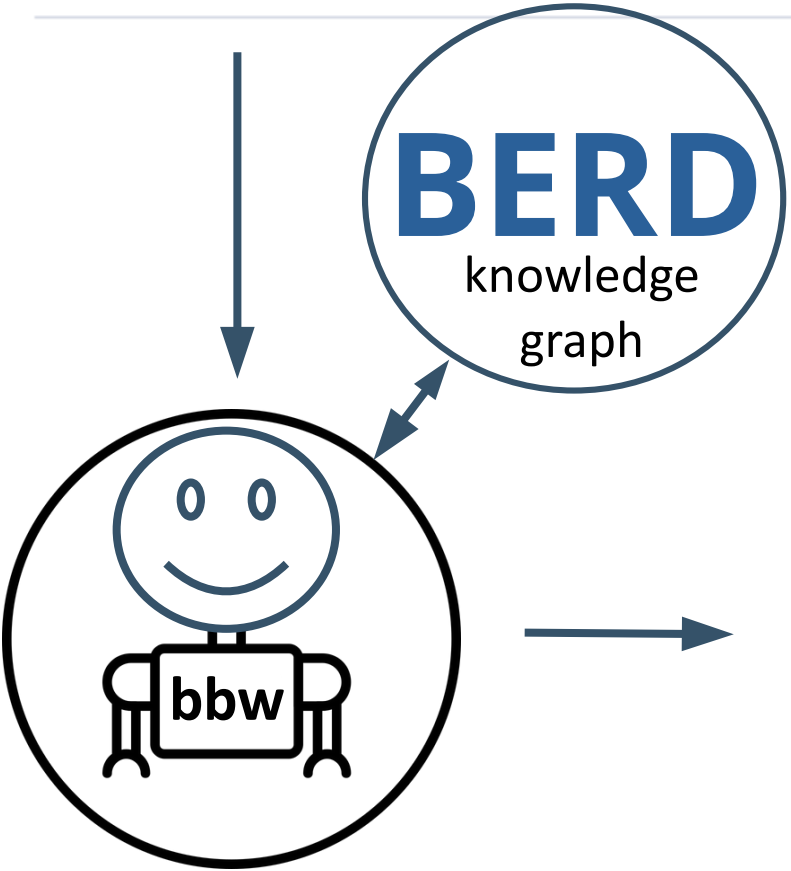
Property	Value
also known as	Graubner Industrie-Beratung GmbH
Crefo number	318147083
Handelsregister: court	Amtsgericht Stuttgart
Handelsregister: court, code & number	Amtsgericht Stuttgart HRB 331527
headquarters address	Schwimmbadstr. 26
headquarters city	Bad Herrenalb
headquarters fax number	(07083) 524914

Semantic annotation, entity linking & data integration

UB UNIVERSITÄTSBIBLIOTHEK MANNHEIM

1	Deutsche Telekom	mobile phone industry	Bonn	1994-01-01	https://www.telekom.com/
2	Lufthansa	air transport	Cologne	1953-01-01	https://www.lufthansa.com
3	SAP SE	software development	Walldorf	1972-04-01	https://www.sap.com





1	Deutsche Telekom	mobile phone industry	Bonn	1994-01-01	https://www.telekom.com/
2	Lufthansa	air transport	Cologne	1953-01-01	https://www.lufthansa.com
3	SAP SE	software development	Walldorf	1972-04-01	https://www.sap.com
property		industry	headquarters location	inception	official website
type	business	industry	urban municipality of Germany		
datatype		WikibaseItem	WikibaseItem	Time	Url

Outline

Motivation

Part 1: Creating

Part 2: Scaling

Part 3: Using

Conclusions

instance of

15

100

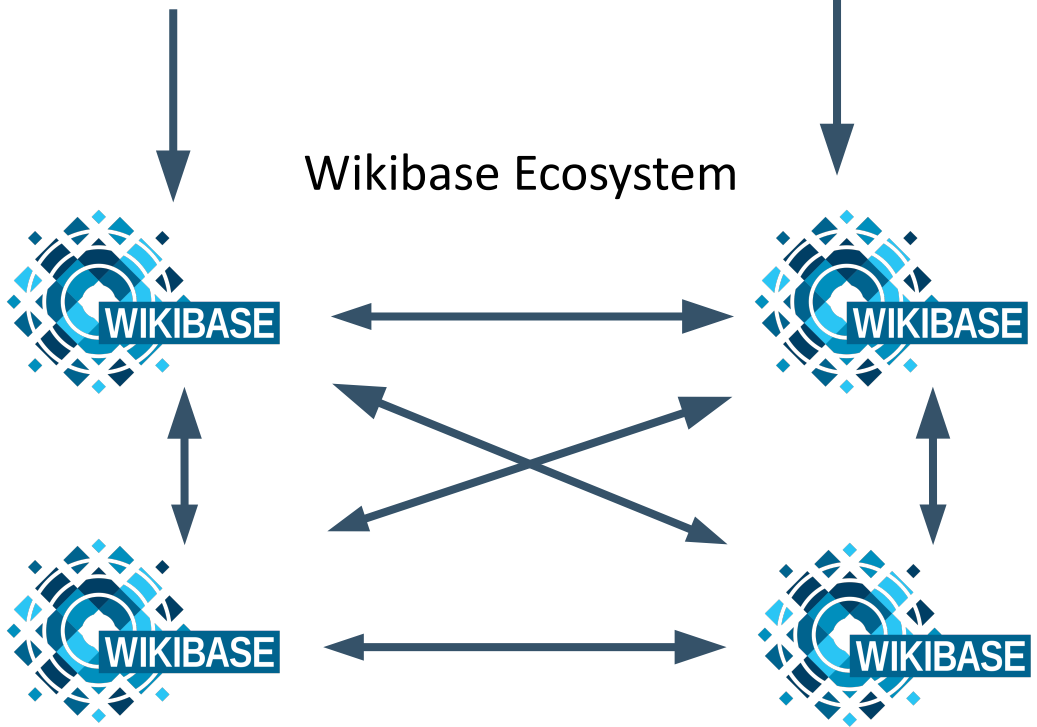
Towards the open company data index = 100

A domain-specific knowledge graph of German companies

BERD

A general-purpose knowledge graph

WIKIDATA



Conclusions

☺ FAIR BERD knowledge graph

☺ State of the art research data management

☺ FAIR **B**usiness, **E**conomic and **R**elated **D**ata

FAIR BERD software:
https://www.berd-bw.de/services-tools