MetaBelgica:

A Linked Data Infrastructure Between Federal Scientific Institutes in Belgium

CoRDI conference, 12 September 2023

https://www.kbr.be/en/projects/metabelgica/

Sven Lieber	**0000-0002-7304-3787**
Ann Van Camp	0000-0002-1915-5956
Dieter De Witte	0000-0001-8480-5719
Eva Coudyzer	0000-0002-5985-7231
Erik Buelinckx	0000-0003-1831-158X
Els Angenon	0000-0002-8888-9662
Hannes Lowagie	0000-0002-0671-3568
Julie Birkholz	0000-0003-1193-0847
Karine Lasaracina	0009-0006-1732-6607

KBR

IRPA KIK Royal Institute for **Cultural Heritage**

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

FAIR DATA

The year is **2023**, leading central-European heritage institutions of *Dutch*, *French* or *German* language **entirely provide FAIR data** about **entities**.

Map by NordNordWest, CC BY-SA 3.0 via Wikimedia Commons
Font "Obelix Pro" by Valentin Antonov

FAIR DATA

Well, not entirely… One small country of indomitable Belgians still hold their metadata. And life is not easy for those who have to **garrison the fortified data silos** either.

dutch digital heritage network

KB〉nationale bibliotheek

DEUTSCHE NATIONAL BIBLIOTHEK

GND

GND

{BnF | Bibliothèque nationale de France

Österreichische Nationalbibliothek

GND

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Data gap - problem and envisioned solution

MetaBelgica

Data model - RDF and Wikibase

Data integration

Data governance

Data access

From Silos to FAIR: Strengthening Belgian Cultural Heritage with a Wikibase-driven Research Data Infrastructure

KBR **Koester de tijd**
Protégeons le temps

Data gap - problem and envisioned solution

Data model - RDF and Wikibase

Data integration

Data governance

Data access

KBR 𝟴 **Koester de tijd**
Protégeons le temps

Data about Belgian entities related to cultural heritage?
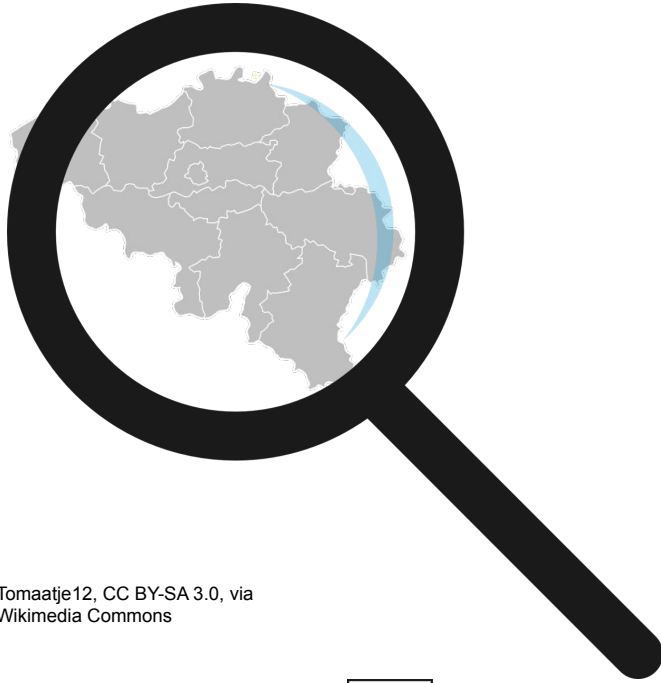
Regional:
 Different cultural heritage institutions in
 both Dutch-speaking *Flanders* and
 French-speaking *Fédération Wallonie-Bruxelles*

National:
 Ten Federal Scientific Institutes with multilingual
 data from which five in the field of Cultural Heritage

International:
 Institutions in neighboring countries
 with the same languages
 Data aggregators like OCLC or Europeana

KBR

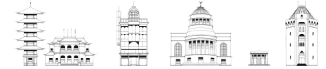IRPA KIK Royal Institute for **Cultural Heritage**

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

Example of entities: metadata of the French comic Asterix

Writer René Goscinny:

http://data.bnf.fr/ark:/12148/cb11905455f
https://isni.org/isni/0000000121235579
https://opac.kbr.be/LIBRARY/doc/AUTHORITY/14119767
https://www.wikidata.org/wiki/Q192214

Illustrator Albert Uderzo:

http://data.bnf.fr/ark:/12148/cb11927309w
https://isni.org/isni/0000000120959006
https://opac.kbr.be/LIBRARY/doc/AUTHORITY/14147813
https://www.wikidata.org/wiki/Q206685

Locations

France: https://data.bnf.fr/ark:/12148/cb152383070
Paris: https://data.bnf.fr/ark:/12148/cb152821567

Dates

1926: https://data.bnf.fr/date/1926/

Different cultural institutions with data of *different domains*

KBR

IRPA KIK · Royal Institute for **Cultural Heritage**

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

ART HISTORY

Each institution has its own type of collections and speciality:
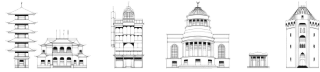good use case for *decentralized data* due to **different domain models**

Duplicate data of shared entities leads to problems

Duplicate efforts to curate data about entities

Varying quality and correctness of data

+ efforts for data consumers to integrate and identify/fix inconsistencies

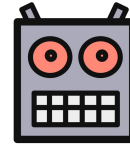# Shared entity management with open source software and FAIR data
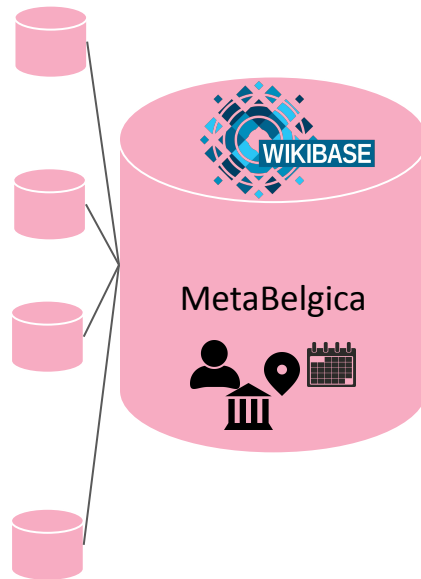

KBR

IRPA KIK Royal Institute for Cultural Heritage

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

ART HISTORY

WIKIBASE

MetaBelgica

Coordinated data management
for enriched data of high quality

Persistent identifiers and a single
trustworthy source for users

MetaBelgica: A shared entity management between Federal Scientific Institutes

Entity types — persons, organizations, time/events, locations

Project duration — 2023 - 2026

Technologies — Resource Description Framework (RDF, Linked Data), Wikibase

Follow-up committee

Metadata experts

Domain experts / Stakeholders

Technology experts

Archives and Museum of Literature
CegeSoma at the Belgian State Archives
Dutch Heritage Network
FID Benelux
KBR ICT department
National Library of Germany
National Library of Luxembourg

OCLC
Open Science Lab at the German National Library of Science and Technology (TIB)
University of Antwerp
Wikibase Stakeholder Group
Wikimedia Belgium
Wikimedia Germany

KBR

IRPA KIK Royal Institute for Cultural Heritage

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

belspo

Data gap - problem and envisioned solution

Data model - RDF and Wikibase

Data integration

Data governance

Data access

KBR **Koester de tijd**
Protégeons le temps

Desk research, workshops and data analysis

Currently used data standards and guidelines

Which data formats are used, any third-party identifiers?

E.g. 53% of KBRs 900k person authorities have an International Standard Name Identifier (ISNI), 70% of our known Belgians

Current cataloging practices

When and how are new entities created?

How are entities linked?

Analyze data (samples) to determine data integration strategy

Legacy data of only text, duplicate data, pseudonyms, …

KBR

Koester de tijd
Protégeons le temps

Reuse existing RDF-based data models (examples)

AGENTSCHAP
DIGITAAL
VLAANDEREN

https://joinup.ec.europa.eu/collection/oslo-open-standards-linked-organisations-0

BELGIF
BELGIAN INTEROPERABILITY FRAMEWORK

https://www.belgif.be/page/specification/fedservicesplatform.en.html

schema.org

https://schema.org/

+ mapping to a Wikibase-based data model

WIKIBASE

KBR
Koester de tijd
Protégeons le temps

Data gap - problem and envisioned solution

Data model - RDF and Wikibase
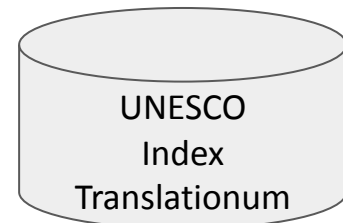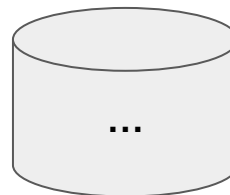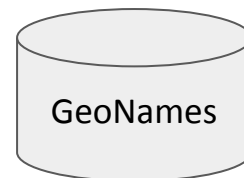
Data integration

Data governance

Data access
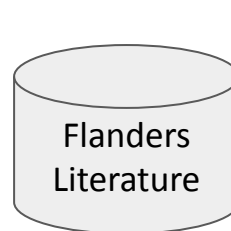
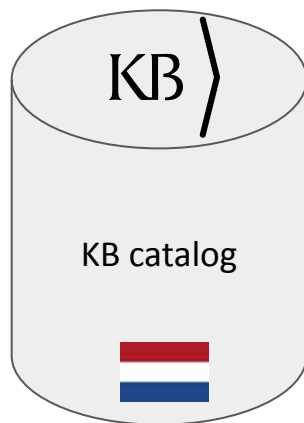**KBR**
Koester de tijd
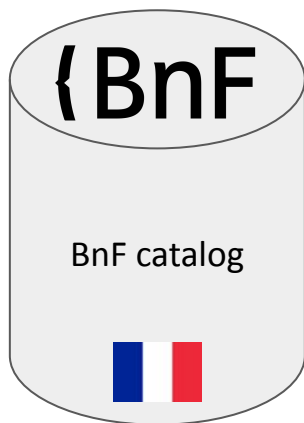**Protégeons le temps**

Building on top of experiences with the BELTRANS project:
Different data *sources* in different *formats* of different *quality*
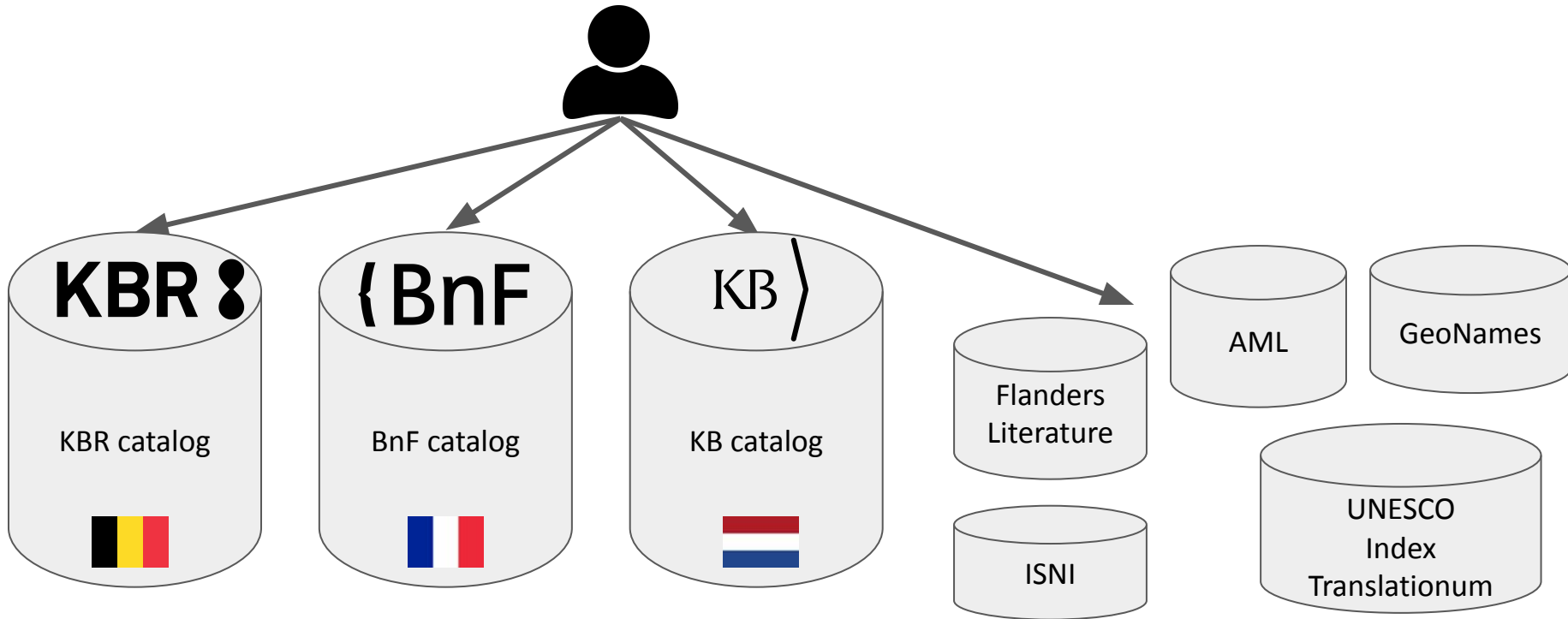
We developed a pipeline with *different* processing per data source to create integrated data

RDF Mapping Language (RML)
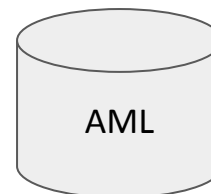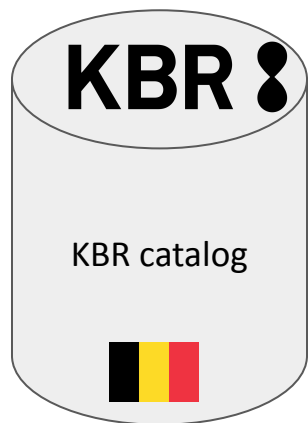
Integrate data by interlinking it with *standard identifiers*

schema:sameAs

Integrate data by interlinking it with *standard identifiers*

schema:sameAs

schema:sameAs

KBR Koester de tijd
Protégeons le temps

Data quality is use case specific - quality difference

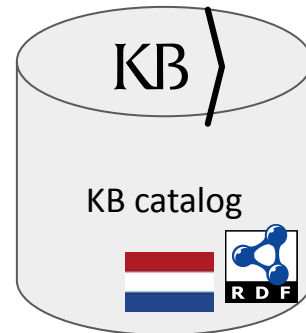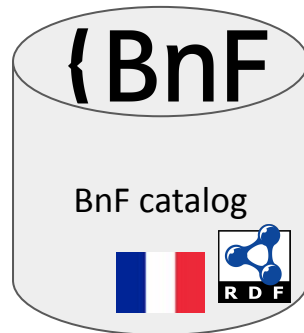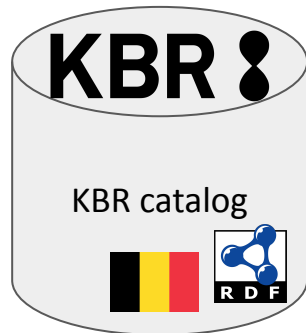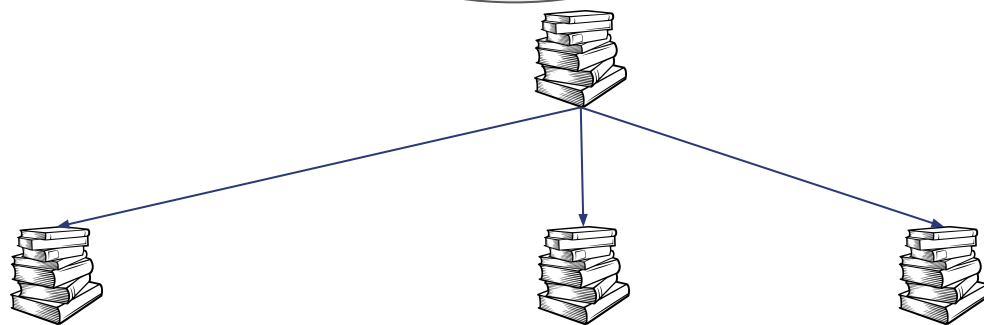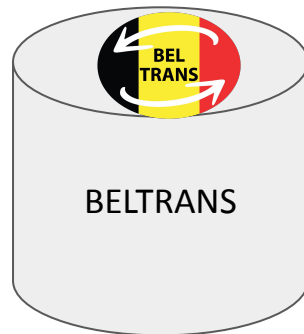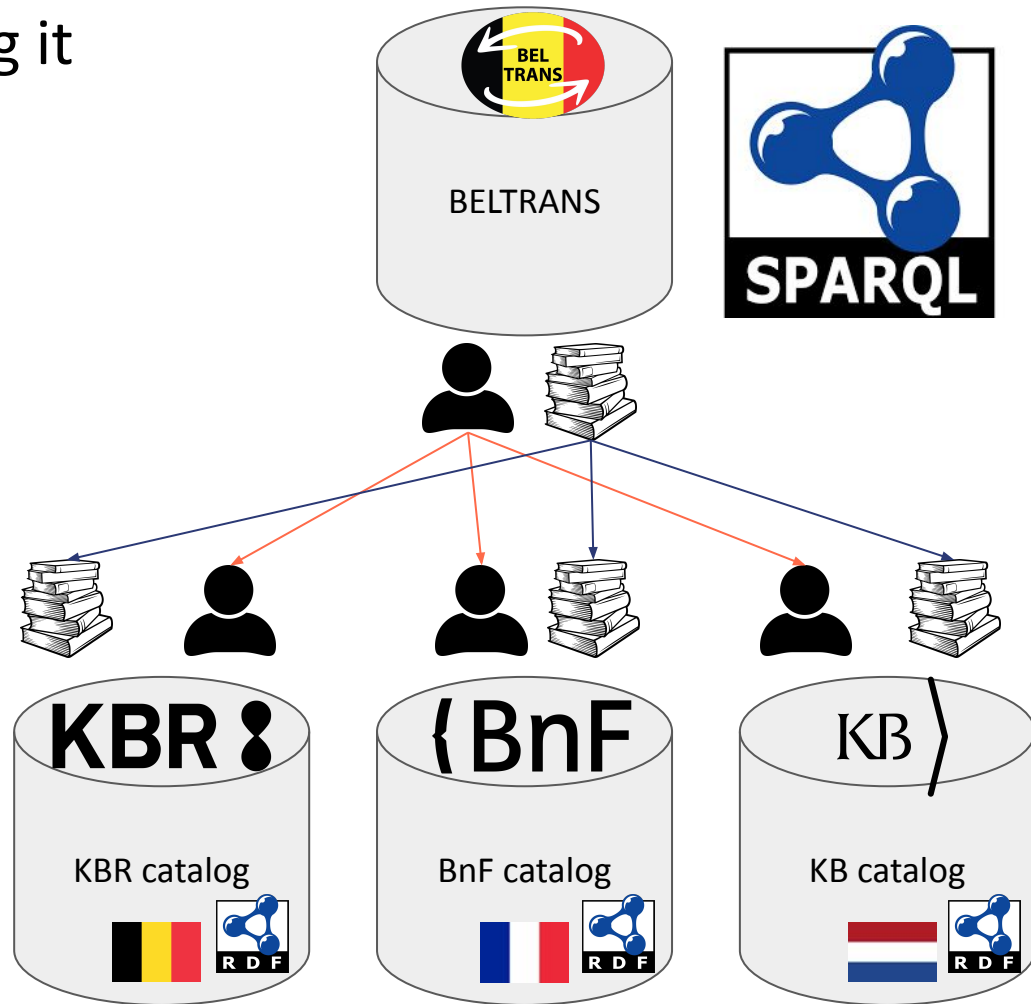Als in een troebele spiegel : brieven / Stefan Hertmans, Gilles Pellerin ; Nederlandse vertaling : Katelijne De Vuyst. - Leuven : Uitgeverij P, 2007. - 96 p. ; 24 cm.

IDN	15749415
ISBN en vorm van uitgave	978-90-77757-58-1
Wettelijk Depotnummer	D/2007/5658/0018
Auteur	Pellerin, Gilles. Auteur
	Hertmans, Stefan (1951-...). Auteur
	De Vuyst, Katelijne (1958-). Vertaler
Uitgever	Uitgeverij P

Geautoriseerde vorm	Uitgeverij P -- Leuven
Land of Regio	BE-VLG
Adres	Sint-Antoniusberg 9, Leuven (3000), België

700(1#) $*14203216 $aPellerin, Gilles $4aut

700(1#) $*14159426 $aHertmans, Stefan $d1951-... $4aut

700(1#) $*14506990 $aDe Vuyst, Katelijne $d1958- $4trl

710(2#) $*14445894 $aUitgeverij P $gLeuven $4pbl $@nl-BE

Ga waarheen je hart je roept : boven alle yoga uit / Jean Déchanet ; met een woord vooraf van Jean Sullivan ; vertaald door Ed Herkes. - Brugge : Uitgeverij Emmaus, 1972. - 147 p. ; 20 cm.

IDN	15151941
Auteur	Déchanet, Jean-Marie (1906-1992) - O.S.B.. Auteur
	Sullivan, Jean. Auteur van voorwoord, inleiding, etc
	Herkes, Ed. (19..-). Vertaler
Uitgever	Uitgeverij Emmaus

246(1#) $aVa où ton coeur te mène $iTitre original / oorspronkelijke titel

264(#1) $aBrugge $bUitgeverij Emmaus $c1972

300(##) $a147 p. $c20 cm

700(1#) $*14109150 $aDéchanet, Jean-Marie $cO.S.B. $d1906-1992 $4aut

700(1#) $*14647172 $aSullivan, Jean $4aui

700(1#) $*14122319 $aHerkes, Ed. $d19..- $4trl

Planned iterative data integration based on data quality

(1) Knowledge Graph Construction with a Human-In-The-Loop for high quality data

Integrated Data (RDF)

WIKIBASE

MetaBelgica

Royal Library of Belgium

Royal Museums of Fine Arts

Royal Institute for Cultural Heritage

Royal Museums of Art and History
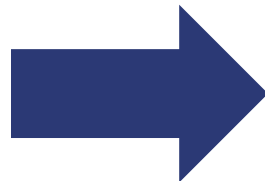
(2) Manual data curation supported by reconciliation for lower quality data

KBR
Koester de tijd
Protégeons le temps

Data gap - problem and envisioned solution

Data model - RDF and Wikibase

Data integration

Data governance

Data access

KBR ⦂ **Koester de tijd**
Protégeons le temps

Data governance for MetaBelgica

After the initial data integration, how to keep the data in sync?

Which data fields should be common?

How to deal with changes?

How to respect/align the different cataloging guidelines?

KBR **Koester de tijd**
Protégeons le temps

Organizational structures: Also link people, not just data

Synthesize different **guidelines**

  Part of desk research and workshops when creating a common data model

Integrate the operation of MetaBelgica in **daily curations tasks** of each institution

  Create documentation of usage in NL/FR

  Organize internal trainings

Investigate the inception of a **governance and editorial board**

  Formalize collaboration and promote platform

KBR
Koester de tijd
Protégeons le temps

Support governance with technical components

Wikibase as back-end to curate common data
  Having Wikibase in the first place helps, because non-technical personnel
  can curate and integrate data

Semi-automatic means to keep local databases and MetaBelgica in sync
  E.g. weekly report about difference between local and shared
  which can be used to keep the data clean and aligned

KBR ⦁ **Koester de tijd**
Protégeons le temps

Data gap - problem and envisioned solution

Data model - RDF and Wikibase

Data integration

Data governance

Data access

KBR ⦙ **Koester de tijd**
Protégeons le temps

In general: different possibilities of access

Data dumps in common formats such as CSV or different RDF serializations

Application Programming Interfaces (APIs) or query interfaces such as SPARQL or Linked Data Fragments

Web platform with user interface

KBR **Koester de tijd**
Protégeons le temps

Which data to access by whom?

We will survey possible stakeholders
   Identify the **needs of our users** and building up a user base in the first place

Which **data fields** to **share** among partners and which to make public
   Part of data governance that touches upon policies like the GDPR

KBR
Koester de tijd
Protégeons le temps

Data gap	Fill the Belgian gap with FAIR data shared by the partners
Data model	RDF for interoperability + Wikibase as a platform
Data integration	Different integration strategies: semi-automatic + manual
Data governance	Organizational and technical
Data access	As open as possible, as closed as necessary

KBR

IRPA KIK Royal Institute for Cultural Heritage

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE

From Silos to FAIR: Strengthening Belgian Cultural Heritage with a Wikibase-driven Research Data Infrastructure

Sven Lieber (Sven.Lieber@kbr.be, https://sven-lieber.org)
Data manager, research and innovation department

https://www.kbr.be/en/projects/metabelgica/

KBR

IRPA KIK Royal Institute for **Cultural Heritage**

Royal Museums of Fine Arts of Belgium

ROYAL MUSEUMS OF ART AND HISTORY
KONINKLIJKE MUSEA VOOR KUNST EN GESCHIEDENIS
MUSÉES ROYAUX D'ART ET D'HISTOIRE