



The Benefits of International Standards for Research Data Management

Sven Lieber

Linked Data and International Standards for Cultural Heritage
Brussels, 14 September 2022

Large amount of unconnected data

What is on these two images, and are they connected somehow?



“24 hours in photos”, 2011 from Erik Kessels

350k printed images uploaded to Flickr in a single day

What is Research Data Management (RDM)?

The benefits of International Standards for RDM at KBR

Data Quality in the context of international identifiers

Conclusion

Take home message:

**Use identifiers and
standard values
where possible!**

What is Research Data Management (RDM)?

The benefits of International Standards for RDM at KBR

Data Quality in the context of international identifiers

Conclusion

What is RDM?



“Research Data Management (RDM) entails all actions needed to ensure that data are secure, **easy to find, understand, and (re)use**, not only during a research project, but also in the longer term.” - UGent

“It includes activities such as **planning, collecting and organizing, documenting, storing and backing up, preserving, and sharing** research data.” - UGent

Increasing awareness of RDM

Universities



“Research Data Management (RDM) entails all actions needed to ensure that data are secure, **easy to find, understand, and (re)use**, not only during a research project, but also in the longer term.” - UGent

“It includes activities such as **planning, collecting and organizing, documenting, storing and backing up, preserving, and sharing** research data.” - UGent

Funding agencies



“Research Data Management (RDM) refers to the way research data is managed throughout a research project. **RDM is an essential skill for researchers from all disciplines and in all settings.**” - Science Europe

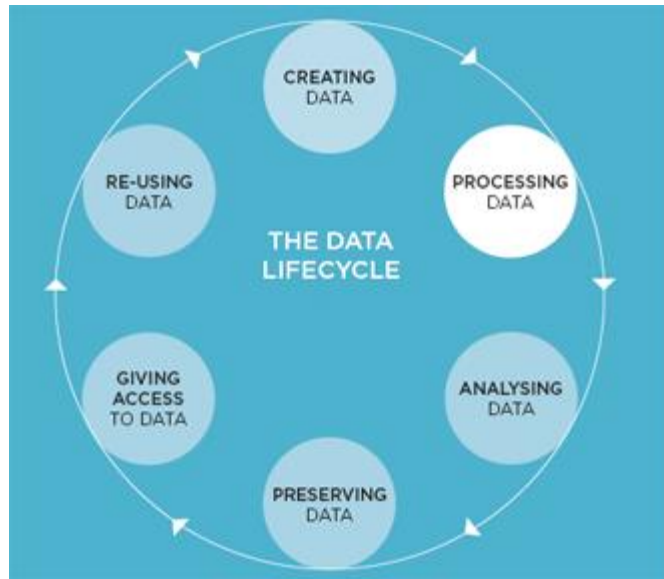
States



“In the **German National Research Data Infrastructure (NFDI)**, valuable data from science and research are **systematically accessed**, networked and made usable in a sustainable and qualitative manner for the **entire German science system**. [...] Together, the Association and the NFDI consortia are **shaping the future of research data management** in Germany.”

Thinking about data and plan throughout its lifecycle

Data Lifecycle



© UGent data stewards

Data Management Plan



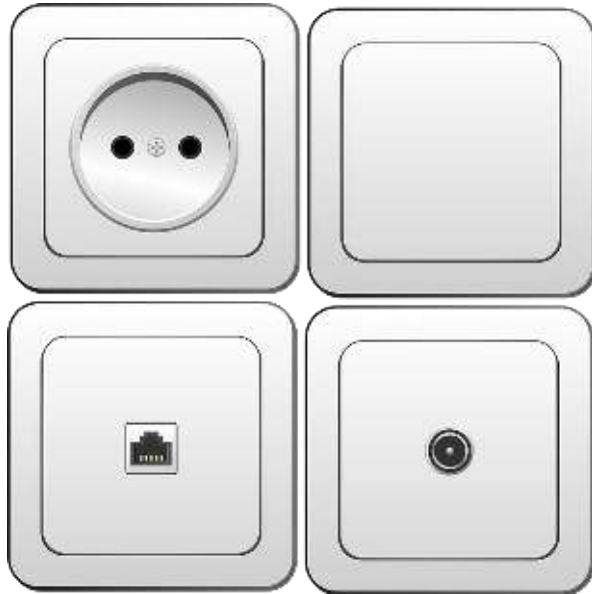
What is Research Data Management (RDM)?

The benefits of International Standards for RDM at KBR

Data Quality in the context of international identifiers

Conclusion

Standards: a blessing and a curse



BELTRANS - Intra-Belgian book translations 1970-2020



Financed by



BRAIN.be 2.0 programme
Pillar 2 “Heritage science”
National thematic project

Studying intra-Belgian translation flows and their context

4 Years project (2021-2025)

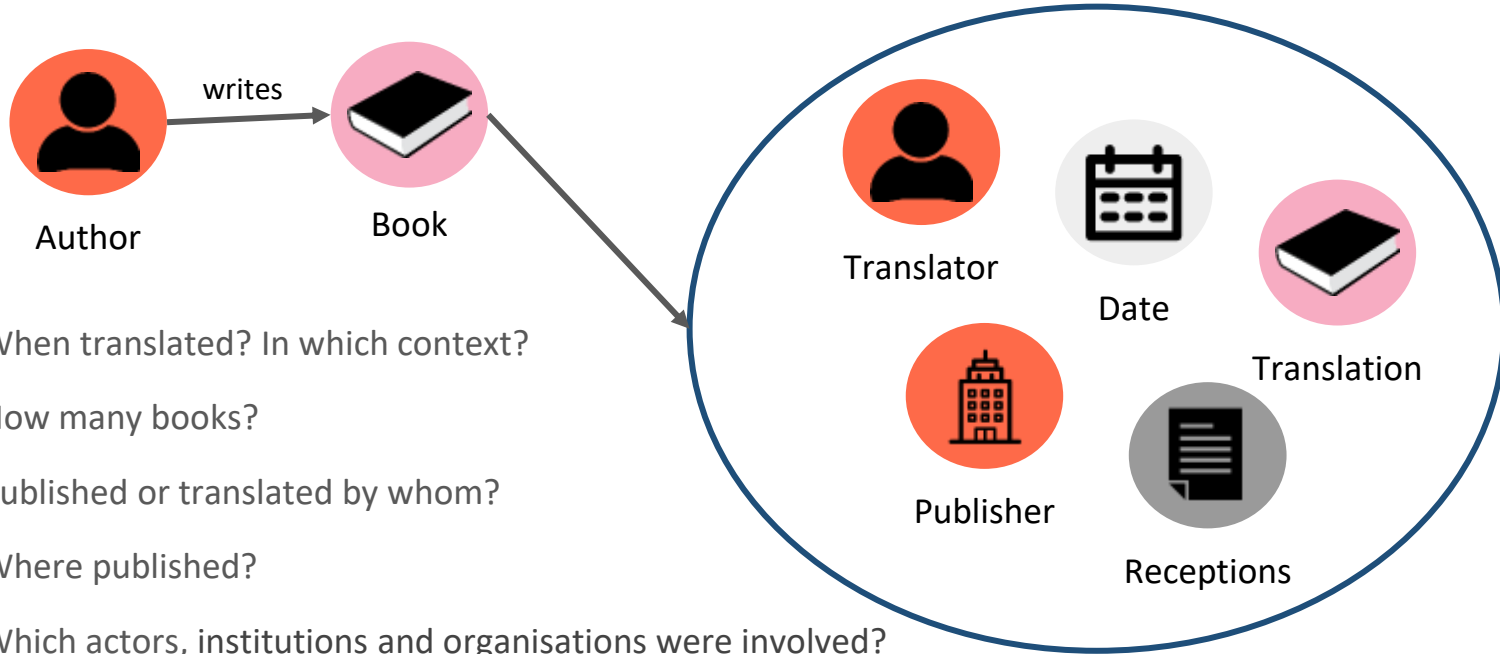
Cooperation with KU Leuven and UCLouvain

3 PhD students working on it

KBR responsible for data management

<https://www.kbr.be/en/projects/beltrans/>

BELTRANS is about *translation flows*



When translated? In which context?

How many books?

Published or translated by whom?

Where published?

Which actors, institutions and organisations were involved?

How were the translations received in the other language community?

...

Use case BELTRANS project: book translations



intra-Belgian



author / illustrator / scenarist / publishing director

location



published anywhere in the world

literary



literary genres (novel, youth literature, comics, poetry)
+ literary non-fiction (mainly

history books)
translations



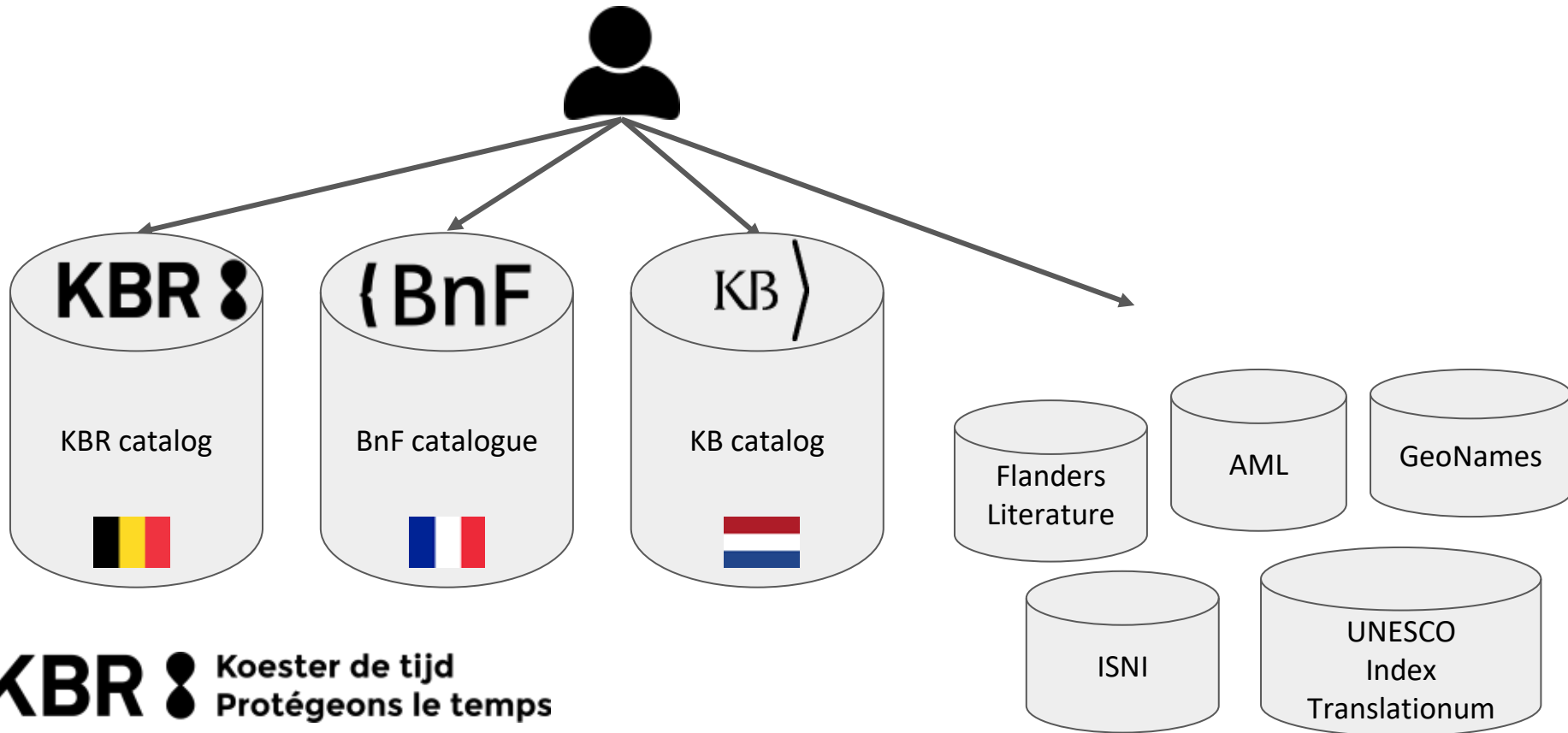
FR-NL / NL-FR

time-period



1970-2020, since the creation of the Communities

Different data *sources* in different *formats* of different *quality*





Relevant standards for BELTRANS

intra-Belgian



MARC relator codes, e.g. 'aut' for author

location



geonames identifiers, coordinates

literary



Belgian Bibliographie, RAMEAU, KBCode,
Universal decimal classification, dewey

translations



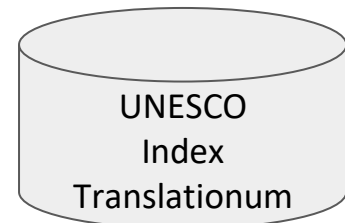
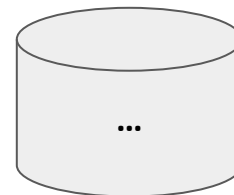
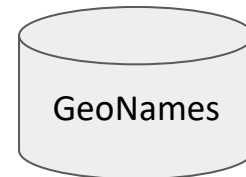
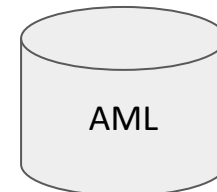
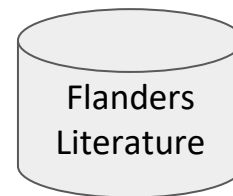
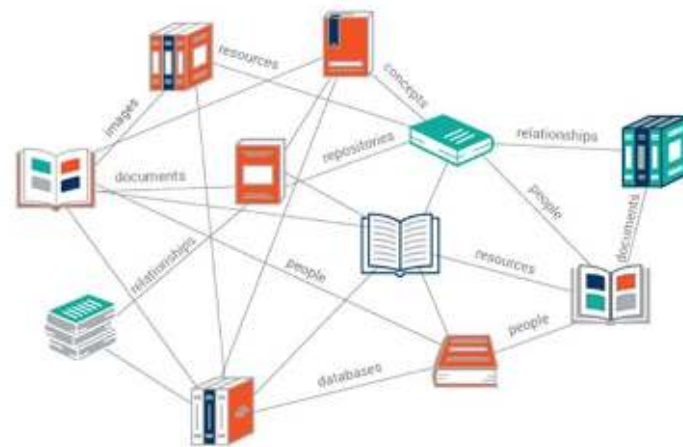
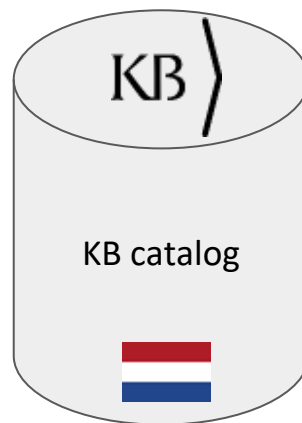
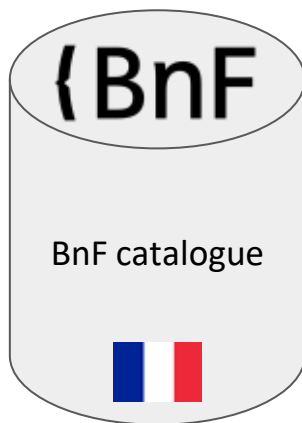
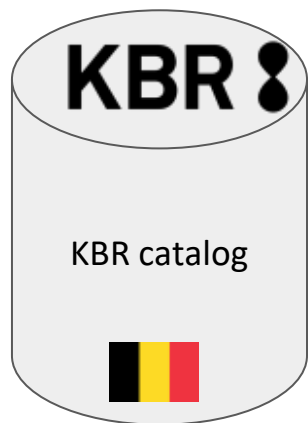
MARC language codes (also old-French, old-Dutch)

time-period



ISO-compliant date formatting

We developed a pipeline with *different processing per data source* to create integrated data



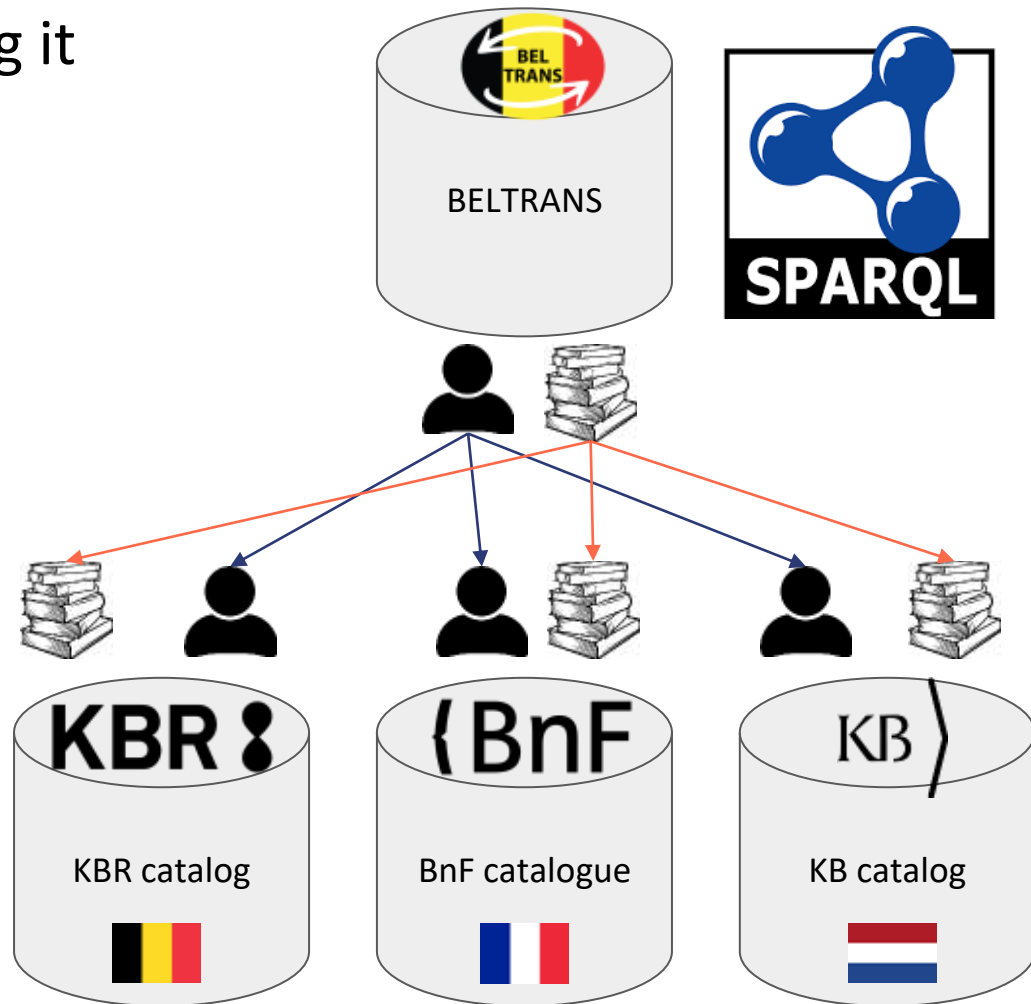
Integrate data by interlinking it
with *standard identifiers*



schema:sameAs



schema:sameAs



Postprocess the data and create a lightweight CSV file

ID	title	year Of Publication KBR	Year Of Publication BnF	Year Of Publication KB	...
1	Book1	2020	2020	2019	
2	Book2	2019		2019	



ID	title	year Of Publication	...
1	Book1	2019 or 2020	
2	Book2	2019	

**Merge data and
report inconsistencies**

Benefits of international standards for BELTRANS and KBR

RDM for BELTRANS: reproducible steps to build the BELTRANS corpus

Abstract: <https://zenodo.org/record/6521682#.YxHG3XaxWUk>

Presentation:

<https://zenodo.org/record/6592382#.YxHHE3axWUk>

GitHub: <https://github.com/kbrbe/beltrans-data-integration>



Besides the project: improve data in the KBR catalogue

Find data quality issues by comparison with other data sources



Upcoming presentation at SWIB (Nov 28 to Dec 2) librarian-in-the-loop Linked Data workflow

What is Research Data Management (RDM)?

The benefits of International Standards for RDM at KBR

Data Quality in the context of international identifiers

Conclusion

Data quality is use case specific

Who is the publisher?

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

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

Data quality is use case specific

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

Where is the publisher from?

In this view: information only in the text

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

Data quality is use case specific

Where is the publisher from?

```
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
```

```
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
```

Data quality is use case specific

From which region is the publisher?

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

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

Identifier links to an authority record

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

Publisher name only as text!
Limited location information

Data quality is use case specific

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

Where is the publisher from?

Text is ambiguous, instead of Brugge it could be Bruges

More ambiguities because of text:
is the publisher from Belgium or from France?

Search in geonames.org

Name	Country
 Bruges  Brige, Briuge, Briugé, Briz, Brizh, Briž, Brjuge, Brjugge, Broegge, Brudje, Bruegge, Brugae, Brugeh, Bruges, Brugg...	Belgium , Flanders West Flanders Province > Arrondissement of Bruges > Brugge
 Bruges Bruges	France , Nouvelle- Aquitaine Gironde > Bordeaux > Bruges

Data quality is use case specific

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

Structured format
+
Standard identifiers
=
Reusability beside initial use case

Search in geonames.org

Name	Country
 Bruges  Brige, Briuge, Briugé, Briz, Brizh, Briž, Brjuge, Brjugge, Broegge, Brudje, Bruegge, Brugae, Brugeh, Bruges, Brugg...	Belgium , Flanders West Flanders Province > Arrondissement of Bruges > Brugge
 Bruges Bruges	France , Nouvelle- Aquitaine Gironde > Bordeaux > Bruges

What is Research Data Management (RDM)?

The benefits of International Standards for RDM at KBR

Data Quality in the context of international identifiers

Conclusion

Reflection on international standards

Standards for structure without standard values do not help much

Better <https://www.geonames.org/2800931>
instead of ambiguous text value “Bruges”

Standards should be open

Ideally in a machine-readable format, e.g. a SKOS taxonomy
instead of data in closed software

Be aware that standards have limitations and biases

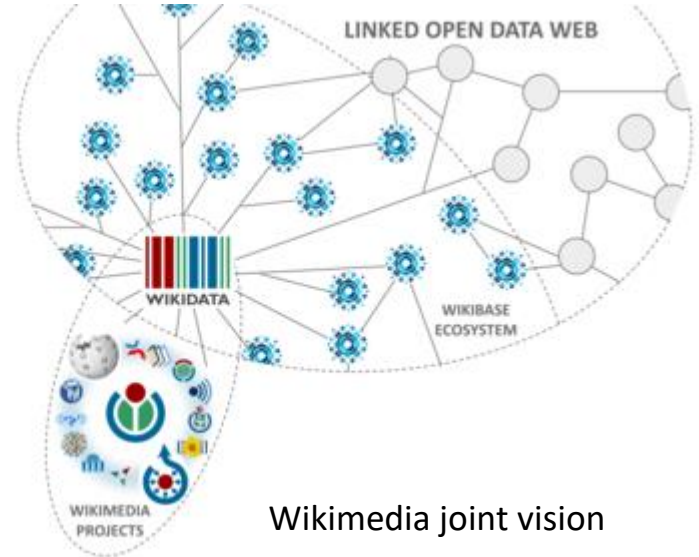
E.g. MARC language codes for French and old-French, but not Walloon

Use standards: make your data reusable for different use cases and ease RDM for yourself and others

Text-based data, limited by space and supporting only a few use cases



Structured data linked via identifiers



Wikimedia joint vision

**Use identifiers and
standard values
where possible!**

Sven Lieber (Sven.Lieber@kbr.be, sven-lieber.org)

Data manager BELTRANS project