



# Representing CIDOC-CRM in Wikibase

for the Luxembourg Shared Authority File

Michelle Pfeiffer

Centre national de recherche archéologique

Jose Emilio Labra Gayo

WESO research group

Slides available at: <https://tinyurl.com/6edaepk>



# What?

The Luxembourg Competency Network on Digital Cultural Heritage is developing a Shared Authority File, to **combine the knowledge from national heritage institutions** and to **increase the impact of the member institutions' digitised collections**.

The ambition of the Luxembourg Shared Authority File (SAF) is to create and maintain an authority file for:

- **persons**,
- corporate bodies,
- conferences and events,
- geographical areas,
- technical terms and work titles.



# Network members

- Ministry of Culture
- National Museum
- National Archives
- National Library
- Audiovisual Archives
- Archaeological institute
- Sites and Monuments
- Institute of Natural History



Musée national  
d'histoire et d'art  
Luxembourg



Bibliothèque nationale du Luxembourg  
Nationalbibliothéik



CNRA Centre national de recherche archéologique





# External partners

Project coordination:

- Maarten Zeinstra (IP Squared)
- Maarten Brinkerink (Digitaal Werktuig)

IP<sup>2</sup>

Information Professional  
Intellectual Property Lawyer

DIGITAAL  
WERKTUIG

External development:

- Jeroen de Dauw (Professional.wiki)
- David Raison (TenTwentyFour)

1024.lu Professional.Wiki

External expertise:

- José Emilio Labra Gayo (WESO research group University of Oviedo, Spain)
- Andra Waagmeester (Micelio)
  - WBStack used to test and iterate



WESO





# Overview

Request for data integration among the different cultural institutions.

Adopt CIDOC-CRM **conceptual model**, starting with person entities.

Use the open source product **Wikibase** as its main data store and internal user interface.

We consider that the mapping between the developed CIDOC-CRM model and the Wikibase model that we present in this paper can be valuable for other cultural heritage institutions working with Wikibase.

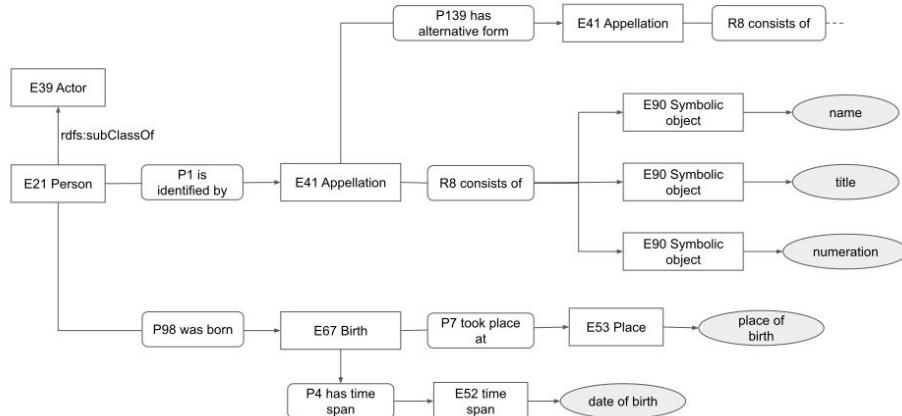
# Technical approach

We started with a data modeling phase by domain experts using diagrams and Excel sheets

Selection of some running examples

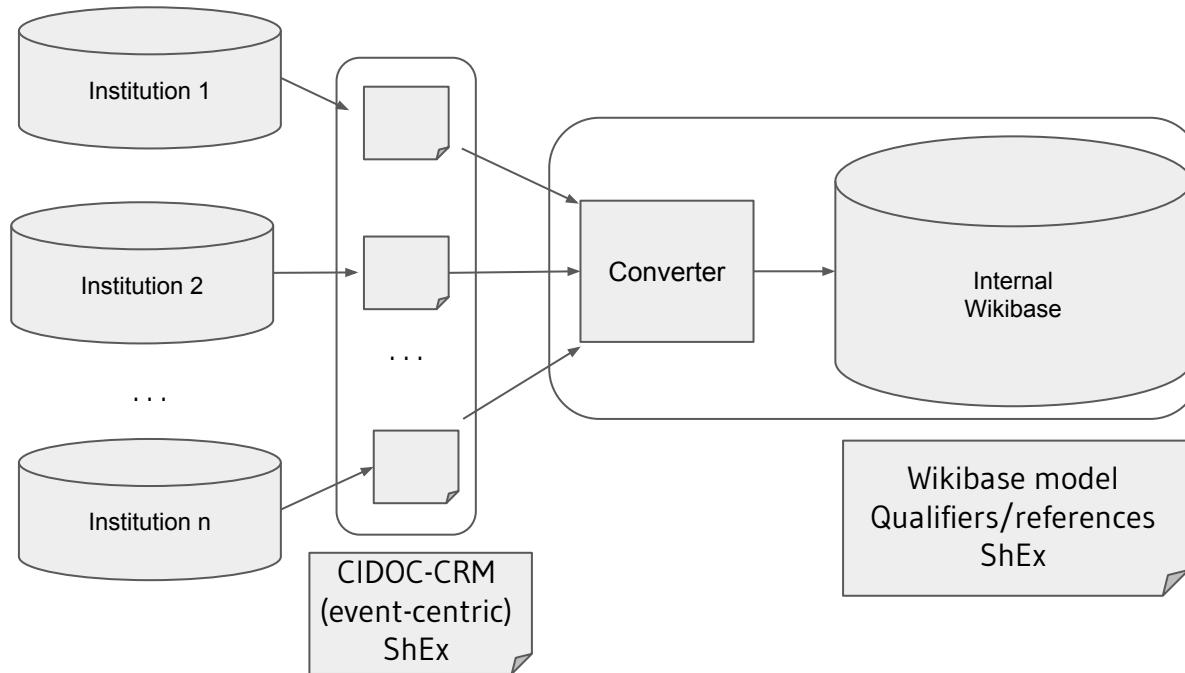
Mapping from CIDOC-CRM event-centric approach to Wikibase data model (qualifiers/references)

Validation using Shape Expressions schemas





# Overview of the approach





# Example: Philip IV, Count of Nassau-Weilburg

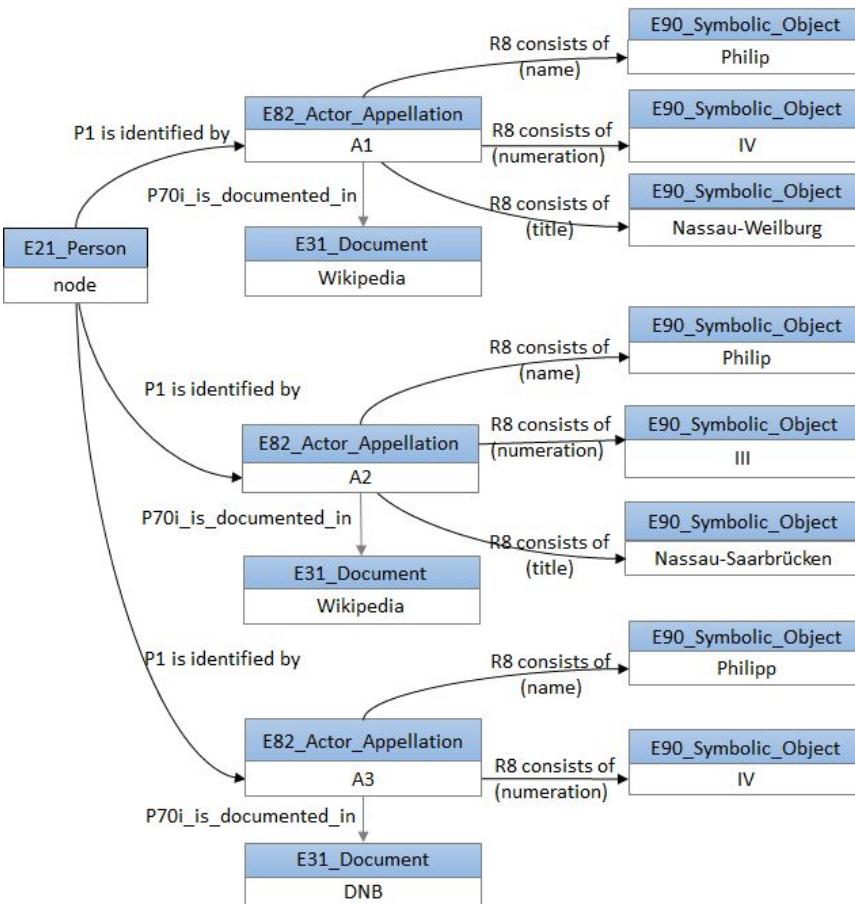
- Basic information:
  - Gender: male
  - Birth place: 1542-10-14, Weilburg
- Name, numeration and title:
  - According to Wikipedia:
    - Philip IV, Count of Nassau-Weilburg and
    - Philip III, Count of Nassau-Saarbrücken
  - According to DNB:
    - Philipp IV, according to DNB



[Philip IV, Count of Nassau-Weilburg](#)



# CIDOC-CRM example





# Requirements of mapping to Wikibase

- **Intuitive:** Represent properties and values directly in the main item of a person, without requiring the user to navigate to different items to view the information.
- **Source of information:** it must be possible to add references that include metadata about the source of information url/text of each statement at the different levels.
- **Consistency:** The values of name/numeration/title/order must form a group which cannot be combined with other values.

For example, (Philip,IV,Nassau-Weilburg) (Philip,III,Nassau-Saarbrücken) = 2 groups

The values of those groups cannot be mixed, forming inconsistencies like (Philip,IV,Nassau-Saarbrücken).



# Mapping option selected

We identified several approaches but selected to follow Wikibase references/qualifiers model

Name represented as a string value statement

Add qualifiers/references to the statement to encode numeration and title.

Advantages

- All information of each name, title, numeration, order for a particular person available in one item
- It is only necessary to edit one item

Disadvantages

- It is not possible to add references to specific qualifiers for the alternative numeration/title values, only to the whole name/numeration/title group



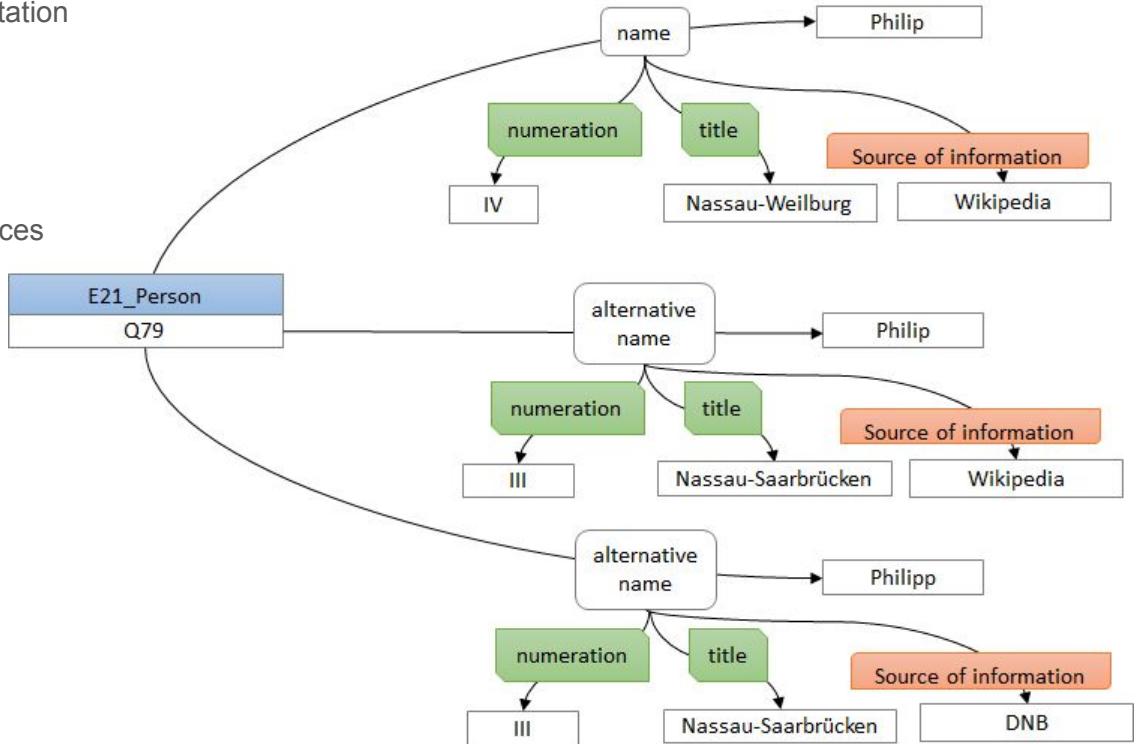
# Wikibase example

Wikibase data model allows an item based representation

All info about a person directly visible

Numeration/title represented as qualifiers

Source of information represented by references



Wikibase data model (blue=item, green=qualifier, orange=reference)

# Wikibase example

We used WBStack as a playground

Easy to create Wikibase examples

Facilitates data modeling and exploration

## Example item

Philip IV with embedded qualifiers (Q79)

[Item](#) [Discussion](#)

No description defined

[In more languages](#) [Configure](#)

Language	Label	Description	Also known as
British English	No label defined	No description defined	
English	Philip IV with embedded qualifiers	No description defined	
Dutch	No label defined	No description defined	

**Statements**

**name**

Philip	numeration	IV
	title	Count of Nassau-Weilburg
<a href="#">▼ 1 reference</a>		
source of information		Wikipedia

**alternative name**

Philip	numeration	III
	title	Count of Nassau-Saarbrücken
<a href="#">▼ 1 reference</a>		
source of information		Wikipedia

Philipp	numeration	IV
	source of information	DNB



# Conclusions

We have created a mapping between a subset of CIDOC-CRM ontological model and Wikibase model

Shape Expressions were helpful for data model

Human readable and machine processable descriptions

Available both for RDF and for Wikibase entity schemas

Production system currently under development

We consider that this approach could be helpful for other cultural heritage institutions



# Thank you!

Presenters:  
Michelle Pfeiffer

Centre national de recherche archéologique, Luxembourg  
[michelle.pfeiffer@cnra.etat.lu](mailto:michelle.pfeiffer@cnra.etat.lu)

Jose Emilio Labra Gayo  
WESO research group, University of Oviedo, Spain  
[labra@uniovi.es](mailto:labra@uniovi.es)



# backup slides



# Example data

Field	Value	Source of information	Group
Name	Philip	Wikipedia	A1
Numeration	IV	Wikipedia	A1
Title	Count of Nassau-Weilburg	Wikipedia	A1
Alternative name	Philip	Wikipedia	A2
Alternative numeration	III	Wikipedia	A2
Alternative title	Count of Nassau-Saarbrücken	Wikipedia	A2
Alternative name	Philipp	DNB	A3
Alternative numeration	IV	DNB	A3

# CIDOC CRM and RDF

CIDOC CRM maintains a logical form which is independent of RDF

CIDOC CRM is an ontology, not a data model

We follow “[Implementing the CIDOC CRM in RDF](#)” recommendations

Classes and direct/inverse properties use `crm:` namespace

`saflux:` represents the SAF Lux namespace

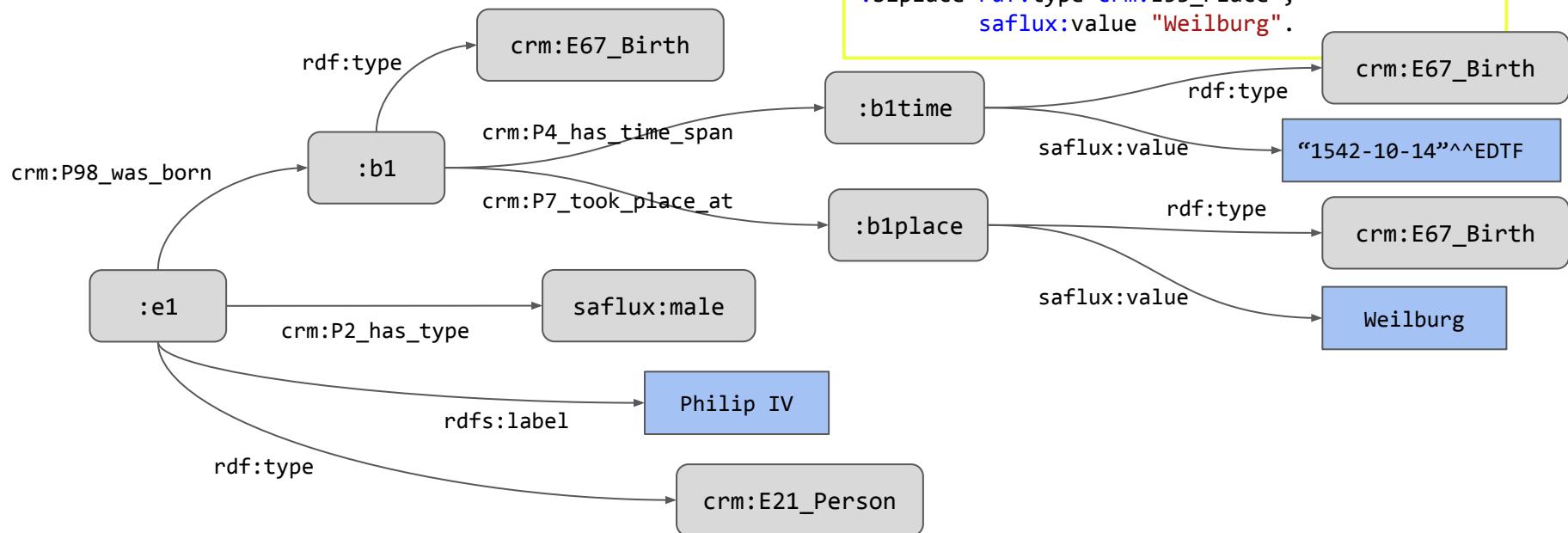
```
prefix :      <http://data.culture.lu/>
prefix saflux: <http://data.culture.lu/ns/saf/>
prefix crm:   <http://www.cidoc-crm.org/cidoc-crm/>
prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

:e1 crm:P2_has_type saflux:male ;
    rdfs:label      "Philip IV" ;
    rdf:type        crm:E21_Person .
```

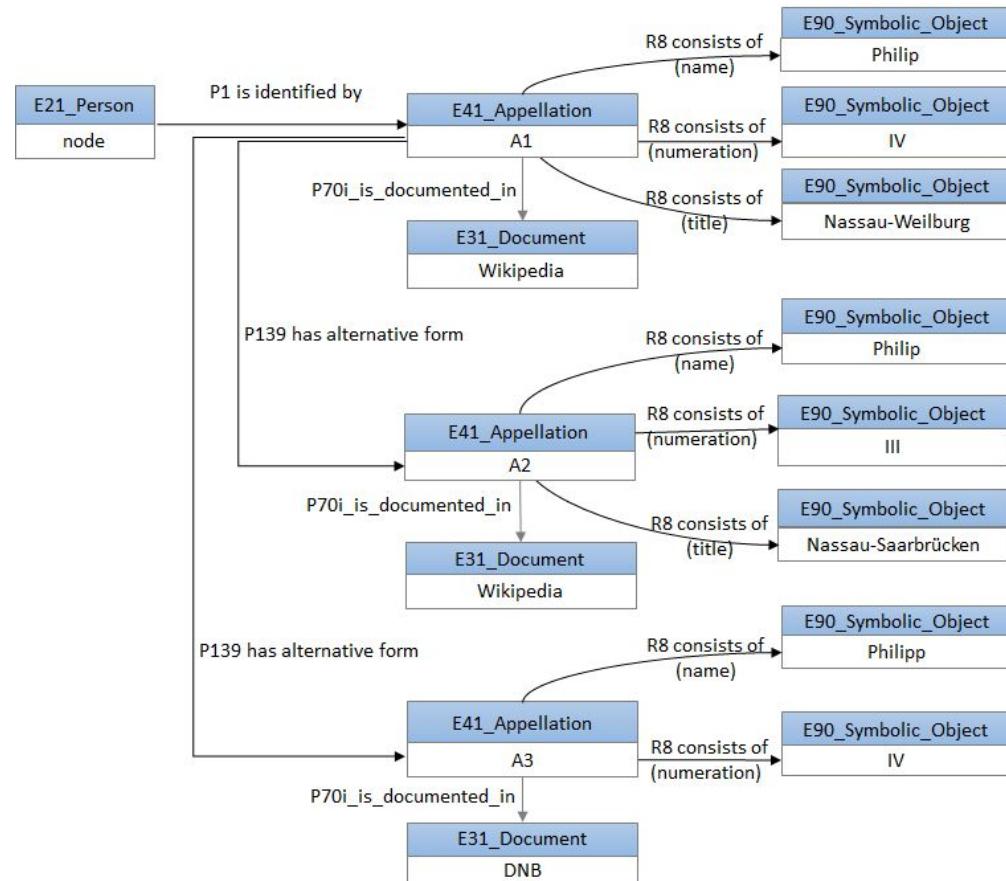
# CIDOC CRM and RDF - events

Events represented as nodes

Dates represented as edtf: datatypes



# CIDOC CRM and RDF - Names



```

:name1 rdf:type crm:E21_Person ;
      crm:P1_is_identified_by :a1 .

:a1 rdf:type crm:E41_Appellation ;
      efr:R8_consists_of :name1, :numeration1, :title1;

:a2 a crm:E41_Appellation ;
      efr:R8_consists_of :name2, :numeration2, :title2;
      crm:P70i_is_documented_in :wikipedia .

:name2 rdf:type crm:E90_Symbolic_Object ;
      saflux:label saflux:alternativeName ;
      crm:P2_has_type saflux:direct .

:a3 rdf:type crm:E41_Appellation ;
      efr:R8_consists_of :name3, :numeration3 ;
      crm:P70i_is_documented_in :dnb .

:name3 rdf:type crm:E90_Symbolic_Object ;
      saflux:label saflux:alternativeName ;
      crm:P2_has_type saflux:direct ;
      saflux:value "Phillip" .

:numeration3 rdf:type crm:E90_Symbolic_Object ;
      saflux:label saflux:alternativeNumeration ;
      saflux:value "IV" .
  
```

# ShEx example

RDF data

```
:e1 crm:P2_has_type saflux:male ;
  rdfs:label "Philip IV" ;
  rdf:type crm:E21_Person .
```

ShEx

```
prefix :      <http://data.culture.lu/>
prefix saflux: <http://data.culture.lu/ns/saf/>
prefix crm:   <http://www.cidoc-crm.org/cidoc-crm/>
prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

saflux:Person {
  rdf:type      [ crm:E21_Person ] ;
  rdfs:label    xsd:string ;
  crm:P2_has_type [ saflux:male saflux:female ]
}
```

Try it: [shex-simple](#), [rdfshape](#)

# ShEx example 2 - events (birth information)

RDF data

```
:e1 rdf:type crm:E21_Person ;
    crm:P98_was_born :b1 .

:b1 rdf:type crm:E67_Birth ;
    crm:P4_has_time_span :b1time ;
    crm:P7_took_place_at :b1place ;
    crm:P70i_is_documented_in :wikipedia .

:b1time rdf:type crm:E52_Time-Span ;
    saflux:value "1542-10-14"^^edtf:EDTF .
:b1place rdf:type crm:E53_Place ;
    saflux:value "Weilburg".
```

ShEx

```
saflux:Person {
    rdf:type [ crm:E21_Person ] ;
    crm:P98_was_born @saflux:BirthInformation
}

saflux:BirthInformation {
    rdf:type [ crm:E67_Birth ] ;
    crm:P4_has_time_span @saflux:TimeSpan ? ;
    crm:P7_took_place_at @saflux:Place ? ;
}

saflux:TimeSpan {
    rdf:type [ crm:E52_Time-Span ] ;
}

saflux:Place {
    rdf:type [ crm:E53_Place ] ;
}
```

Try it: [shex-simple](#), [rdfshape](#)

# ShEx example 3 (Names)

```
:e1 rdf:type crm:E21_Person ;
  crm:P1_is_identified_by :a1 .

:a1 rdf:type crm:E41_Appellation ;
  efr:R8_consists_of :name1, :enumeration1, :title1;
  # ...

:a2 a crm:E41_Appellation ;
  efr:R8_consists_of :name2, :enumeration2, :title2;
  crm:P70i_is_documented_in :wikipedia .

:name2 a crm:E90_Symbolic_Object ;
  :a3 a crm:E41_Appellation ;
  efr:R8_consists_of :name3, :enumeration3 ;
  crm:P70i_is_documented_in :dnb .

:name3 a crm:E90_Symbolic_Object ;
  afl:label afl:alternativeName ;
  crm:P2_has_type saflux:direct ;
  saflux:value "Phillip" .

:enumeration3 a crm:E90_Symbolic_Object ;
  saflux:label saflux:alternativeNumeration ;
  saflux:value "IV" .
```

```
saflux:Person {
  rdf:type [ crm:E21_Person ] ;
  crm:P1_is_identified_by @saflux:NameAppellation ;
  # ...
}

saflux:NameAppellation {
  rdf:type [ crm:E41_Appellation ] ;
  efr:R8_consists_of @saflux:Name ;
  efr:R8_consists_of @saflux:Numeration ? ;
  efr:R8_consists_of @saflux:Title ? ;
  crm:P70i_is_documented_in @saflux:SourceOfInformation * ;
  crm:P139_has_alternative_form @saflux:AltNameAppellation * ;
}

saflux:AltNameAppellation {
  rdf:type [ crm:E41_Appellation ] ;
  efr:R8_consists_of @saflux:AlternativeName ? ;
  efr:R8_consists_of @saflux:AlternativeNumeration ? ;
  efr:R8_consists_of @saflux:AlternativeTitle ? ;
  crm:P70i_is_documented_in @saflux:SourceOfInformation * ;
}

saflux:Name {
  rdf:type [ crm:E90_Symbolic_Object ];
  afl:label @saflux:nameLabel ;
  crm:P2_has_type @saflux:NameType ;
  saflux:value xsd:string
}
```

# Validating wikibase data with ShEx

```
start = @saflux:person

saflux:person {
  wdt:P1 [ wd:Q78 ] ;
  p:P8 @saflux:P8_name ;
  p:P17 @saflux:P17_alternative_name* ;
  p:P19 @saflux:P19_place_of_birth ? ;
  p:P18 @saflux:P18_date_of_birth ? ;
  p:P21 @saflux:P21_place_of_death ? ;
  p:P20 @saflux:P20_date_of_death ? ;
  p:P22 @saflux:P22_gender ;
  p:P23 @saflux:P23_profession ?;
  p:P26 @saflux:P26_activity * ;
  p:P29 @saflux:P29_AFL_identifier ;
  p:P36 @saflux:P36_ISNI ? ;
  p:P37 @saflux:P37_VIAF ?;
  p:P41 @saflux:P41_ARK ?;
}

saflux:P8_name {
  ps:P8 xsd:string ;
  pq:P10 xsd:string ? ;
  pq:P11 xsd:string ? ;
  prov:wasDerivedFrom @saflux:reference +
}
```

validate (ctl-enter)

Query Entities to check

```
PREFIX wd: <http://lux-saf-beta.wiki.opencura.com/entity/>
PREFIX wdt: <http://lux-saf-beta.wiki.opencura.com/prop/direct/>
SELECT ?person WHERE {
  ?person wdt:P1 wd:Q78 .
} limit 5
```

✓saflux:Q351@START  
✓saflux:Q352@START  
✓saflux:Q353@START  
✓saflux:Q354@START  
✓saflux:Q355@START

[Try it](#) (ShEx-simple)

## 2 data models: CIDOC-CRM and Wikibase

CIDOC-CRM model ([rules-1.0.3.shex](#))

“Logical” names like: crm:E21\_Person, crm:P1\_is\_identified\_by

Hierarchical and event based structure

Wikibase model ([wikibase\\_v1.0.3.shex](#))

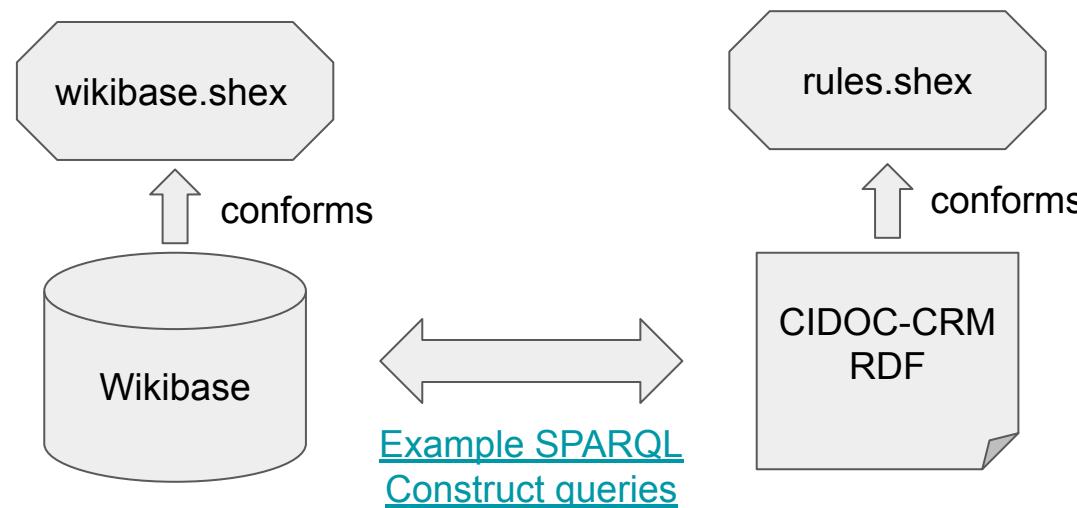
Number-based identifiers like Q78, P8

Flatten structure making use of qualifiers and references

# Bridging the gap between CIDOC-CRM and wikibase

Convert from CIDOC-CRM and Wikibase data

SPARQL construct queries or other techniques



# Factsheet – Luxembourg Shared Authority File

## What

The Luxembourg Competency Network on Digital Cultural Heritage is developing a Shared Authority File, to combine the knowledge from national heritage institutions and to increase the impact of the member institutions' digitised collections.

The ambition of the Luxembourg Shared Authority File (SAF) is to create and maintain an authority file for:

- persons,
- corporate bodies,
- conferences and events,
- geographical areas,
- technical terms and work titles.

The data of the authority file uses a custom developed [CIDOC-CRM](#) RDF/XML data model.

## Who

The Luxembourg Competency Network on Digital Cultural Heritage consists of:

- [The Ministry of Culture](#),
- [Archives nationales de Luxembourg](#) (ANLux),
- [Bibliothèque nationale du Luxembourg](#) (BnL),
- [Centre national de l'audiovisuel](#) (CNA),
- [Centre national de littérature](#) (CNL),
- [Centre national de recherche archéologique](#) (CNRA),
- [Musée national d'histoire et d'art](#) (MNHA),
- [Musée national d'histoire naturelle](#) (MNHN),
- [Services de sites et monuments nationaux](#) (SSMN).

The authority file is developed with the help of [IP Squared](#), [Digitaal Werktuig](#) (Project Coordination), [Mcelcio](#), [WESO](#), [Professional.wiki](#), and [TenTwentyFour](#) (experts)

## Where

The Shared Authority File uses the open source product [Wikibase](#) as its main data store and internal user interface.

It will use [persist.ly](#) as a provider for persistent identifiers (ARKs) for any publicly available data.

Wikibase contains a datetime representation based on [ISO 8601:2004](#). The project extended the Wikibase software that includes [ISO 8601:2019](#), and their extensions. This resulted in two open source products. A [library](#) for the validation and humanisation of EDTF dates in PHP, and an [extension](#) to WikiBase to use this library. Both are available as open source software for all to use and improve. The humanisation is translated in various languages using [Translate.wiki](#).

## When

The first modality (natural persons) will be realised by September 2021. After this our roadmap contains creating a public portal on top of the Authority File, and expanding the modalities that the Authority File will cover.

## More information

Contact [Marianne Backes](#) (Ministry of Culture) for more information about this project.

Factsheet Version 1.1.0 – March 2021

## [Factsheet](#) (live document v1.1.0)