

From CegeSoma spreadsheets to Linked Data: a Wikibase journey

Linking the Past, KBR, 22.11.2019

Anne Chardonnens



@annechardo

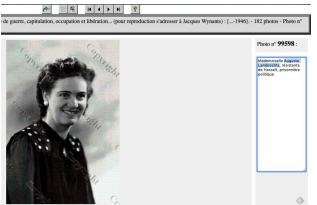




Andrée de Jongh
Andrée De Jongh
Andrée DeJongh
De Jongh, Andrée Eugénie Adrienne
Countess Andree 'Dedee' de Jongh
Dédée
Cyclone DD
Petit Cyclone
The Postman

Bruxelles

Le Bruxellois



```
<russea>
<row num="1">
<type>1,3</type>
<naam>Abeel S</naam>
<voornaam>Joseph</voornaam>
<geb_datum>1916</geb_datum>
<geb_plaats>Seraing</geb_plaats>
<nationaliteit>Belg</nationaliteit>
<regist_nr> 275</regist_nr>
</row>
<row num="2">
<type>1,1</type>
<naam>Abels C</naam>
<voornaam>Christian</voornaam>
<geb_datum>1927</geb_datum>
<geb_plaats>Anvers</geb_plaats>
<nationaliteit>Belg</nationaliteit>
<regist_nr> 110</regist_nr>
</row>
<row num="3">
<type>3,1</type>
<naam>Abem Lis</naam>
<voornaam>Robert</voornaam>
<geb_datum>1926</geb_datum>
<geb_plaats>Aarsimont</geb_plaats>
<nationaliteit>Belg</nationaliteit>
<regist_nr> 276</regist_nr>
<dossier_nr>881</dossier_nr>
<archief_nr>278125</archief_nr>
</row>
```

714-1053. « Béleg zur Verwaltungsbach ». Preuves de versements en numéraire ou à d'opérations financières allemandes (Reichsbahndirektion, Deutsche Reichsbahn, William F. du Diamondbank d'Amsterdam) à partir de comptes ou de valeurs appartenant à des particuliers. 1940-1945. 1 volume, 293 chemins et 83 pièces
714-717. « 1940-1941 », juillet 1944 - avril 1945 et s.d. 2 lisses et 2 chemins
714. Juillet - décembre 1946. 1 lisse
715. Janvier - juin 1941. 1 lisse
716. Juillet - décembre 1941. 1 lisse

718-1153. < 1942 a. juil 1940 - juillet 1943 et s.d. 291 chemises et 83 pièces
Par dossier nominal, en fonction des numéros de dossiers (ordre numélique croissant)

718. Richard Ruben (Glückstadt)
juli 1943 - oktober 1942.

719. Trichtbauer I.

729. Assenheimer D. (Antwerp).

221 Mandel Moses Leib. mei 1941 - april 1942.

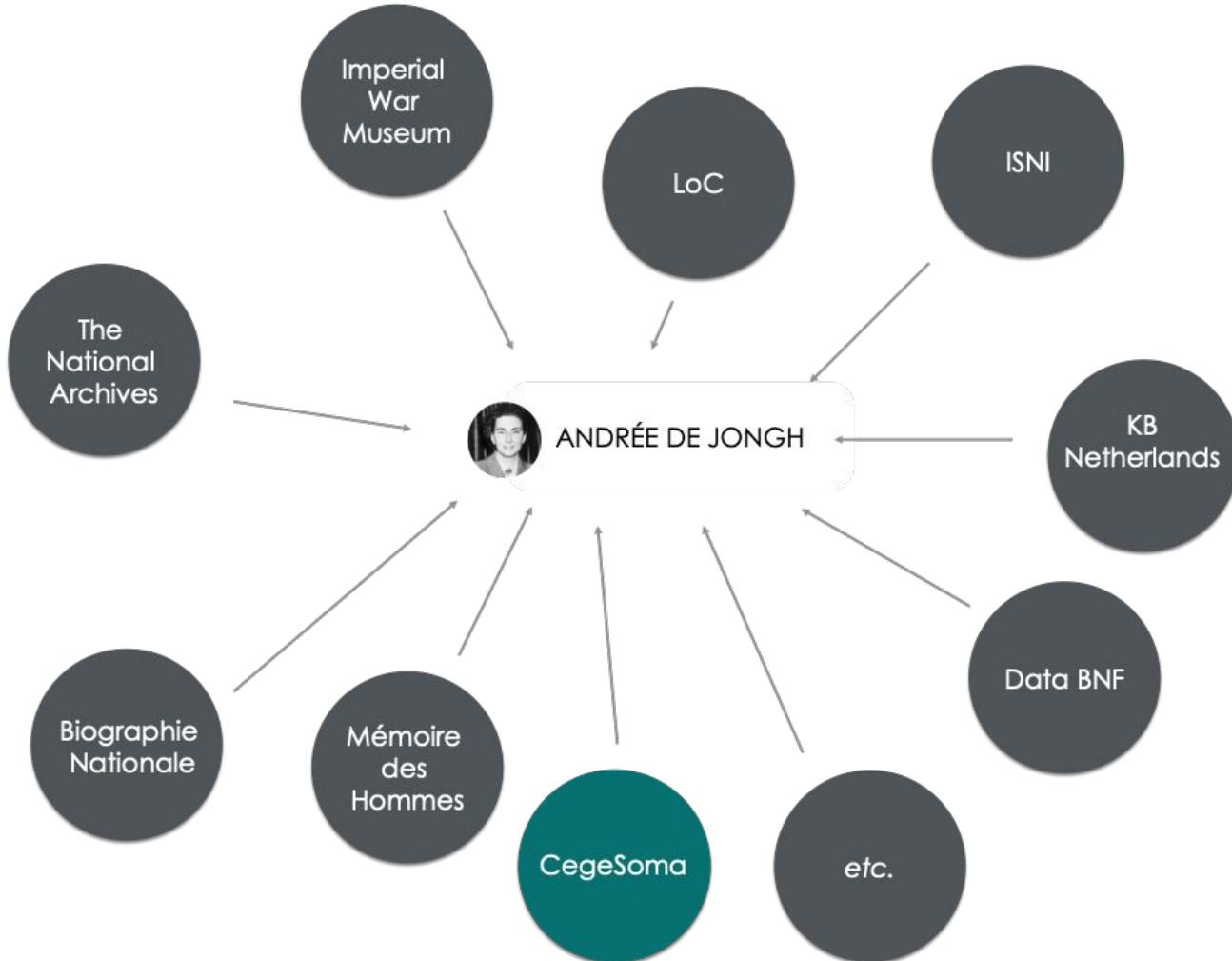
L	V	F	G	B
De Heidt	Albert Benedict Leonard	Artbergen	28.03.1910 Antwerpen	S81410040
De Klerk	Edmund	Antwerpen	28.03.1910 Antwerpen	S81410040
Beur	Geoffrey Alfred Dick	Verwerven	28.03.1910 Antwerpen	S81410037
De Bruyn	John	Antwerpen	28.03.1910 Antwerpen	S81410037
Schmitz	François Joseph Hubert	Gemert	28.03.1910 Antwerpen-Houthalen	S81410037
Bethune	Frank	Euden	13.12.2000 Antwerpen	S81410037
Hughes	Frank	Euden	04.04.2010 Antwerpen	S81410037
Hughes	Agnes	Euden	04.04.2010 Antwerpen	S81410037
De Groot	Frederick Hendrikus	Verwerven	04.02.2010 Antwerpen	S81410037
Schmitz	Hans Michaelina Francisca	Verwerven	04.02.2010 Antwerpen	S81410037
Breuk	John	Hoornbeek-Hout	21.12.2000 in Calais	S81410037
Geddes	Henry Adelinde	Anvers	17.03.2010 Antwerp-Haringey	S81410037
Lutz	Camil Ferdinand Leonidas	Leuven	17.03.2010 Antwerp	S81410037
Borrel	Ignace Hubert	Brugge	17.03.2010 Brussel	S81410037
De Groot	Gerard	Antwerpen	05.12.2000 Antwerpen	S81410037
Pora	Albert Ernest Gullane	Elsene (Bel.Cong.)	01.01.2010 St.Gillis	S81410037
Verhae	Henriette Maria	Verwerven	02.04.2010 Antwerpen	S81410037
Verhae	Hendrik Cornelis	Ortendaal	02.04.2010 Antwerpen	S81410037
Hegeman	Hegemarie Elisabeth	Oostende	24.10.2010 Oostende	S81410037
Verhae	Albert	Verwerven	02.04.2010 Antwerpen	S81410037



1 PERSON = 1



```
<russ>
<row num="1">
  <type>1.3</type>
  <naam>Abeel</naam>
  <voornaam>Joseph</voornaam>
  <geb_datum>1916</geb_datum>
  <geb_plaats>Seraing</geb_plaats>
  <national>Belge</national>
  <regist_nr>275</regist_nr>
</row>
<row num="2">
  <type>1.1</type>
  <naam>Abels</naam>
  <voornaam>Christian</voornaam>
  <geb_datum>1927</geb_datum>
  <geb_plaats>Anvers</geb_plaats>
  <national>Belge</national>
  <regist_nr>110</regist_nr>
</row>
<row num="3">
  <type>3.1</type>
  <naam>Abemlis</naam>
  <voornaam>Robert</voornaam>
  <geb_datum>1926</geb_datum>
  <geb_plaats>Arimont</geb_plaats>
  <national>Belge</national>
  <regist_nr>276</regist_nr>
  <dossier_nr>881</dossier_nr>
  <archief_nr>278125</archief_nr>
</row>
```





Andrée De Jongh (Q461027)***

résistante belge

► Recoin: Propriétés manquantes les plus pertinentes

► Plus de langues

Conquerir

Langue	Libellé	Description	Alias
français	Andrée De Jongh	résistante belge	
anglais	Andrée de Jongh	Belgian resistance member	Andree de Jongh
néerlandais	Andrée de Jongh	Belgisch verzetstrijdster in de Tweede Wereldoorlog	Andree de Jongh
allemand	Andrée de Jongh	belgische Krankenschwester und Fluchthelferin	Andree de Jongh

Toutes les langues saisies

Déclarations

nature de l'élément	être humain ***	modifier
	► 1 référence	
+ ajouter une valeur		

image		modifier
Civilian Bravery Awards during the Second World War HU55451.jpg 583 × 800 ; 84 Kio		

modifier

Wikipédia (6 entrées) modifier

fr Andrée De Jongh [ref]

de Andrée de Jongh [ref]

en Andrée de Jongh [ref]

es Andrée de Jongh [ref]

nl Andrée de Jongh [ref]

pl Andrée de Jongh [ref]

Wikilivres (0 entrée) modifier

Wikinews (0 entrée) modifier

Wikiquote (0 entrée) modifier

Wikisource (0 entrée) modifier

Wikiversité (0 entrée) modifier

Wikivoyage (0 entrée) modifier

Wiktionary (0 entrée) modifier

Autres sites (1 entrée) modifier

commons Category:Andrée De Jongh



Using Wikibase as a Platform to Develop a Semantic Biodiversity Standard

David Fichtmueller

Botanic Garden and Botanical Museum Berlin



WikidataCon 2019, 25-26 October, Berlin

CONTENTdm User Group Meeting, August 1-2, 2018, Columbus Metropolitan Library

Using Wikibase to Manage Linked Data for Digital Collections

Jean Godby

Research Scientist

OCLC Membership & Research

Jeff Mixter

Software Engineer

OCLC Membership & Research



Main page Discussion Read View source View history Search FactGrid Q

Welcome!

— on FactGrid, the Wikibase instance for historians, Q1749 on the global Wikibase Registry.

The project is still under construction. You might like to check our blog to learn more about the aims. Visit the mailing list for ongoing work.

If you do not have an account yet, it will be the first thing you should offer (but read our guidelines for Service before you complete the technical questionnaire).

The present installation is by default in English. Anyone with an account can log in and change that to any other language in the Preferences section. We have guest accounts for visitors who just want to switch the interface language — see the different languages offered on top. The software is multi-lingual but it needs translation. German and English are so far the two languages it speaks with greater precision. It also offers a multilingual (ReST API) interface, so that even users without an account can switch languages to their preferences.

Dr. Olaf Simons
Forschungszentrum Gotha der Universität Erfurt
Am Schlossberg 2
99867 Gotha
<http://factgrid.org>

FactGrid
a database for historians

Rechercher dans Lingua Lit Rechercher ☰

Lingua Libre PARTICIPE SONOTHÈQUE

Hibou anglais (Q208463)

You can now exclude automatically words you have already recorded! [More information here.](#)

De Lingua Libre

audio record - fra - ERIC LEWIN (ERIC LEWIN) 499as

- Plus de langues	Catégorie	Libellé	Description	Alias
français	Pas de libellé défini	Aucune description fournie		
anglais	Hibou	audio record - fra - ERIC LEWIN (ERIC LEWIN)		

Déclarations

nature de l'élément : enregistrement

Assessing Wikibase as the core for the French National Entities File (FNE)

Benjamin Bober



Anila Angjeli (BnF)

WikidataCon 2019, 25-26 October, Berlin

DEUTSCHE
NATIONALE
BIBLIOTHEK

3 | 7 | GND meets Wikibase | WikidataCon 2019

Why we go for it
GND meets Wikibase





ADOCHS

Item Discussion

Andrée de Jongh (Q5)

résistant belge
petit cyclone I Dédeé

+ Plus de langues Configurer

Langue	Libellé	Description	Alias
français	Andrée de Jongh	résistant belge	
anglais	Andrée de Jongh	Belgisch verzetsstrijder	petit cyclone Dédeé
néerlandais	Andrée de Jongh	Belgische resistance member	
allemand	Andrée de Jongh	Aucune description fournie	Little Cyclone

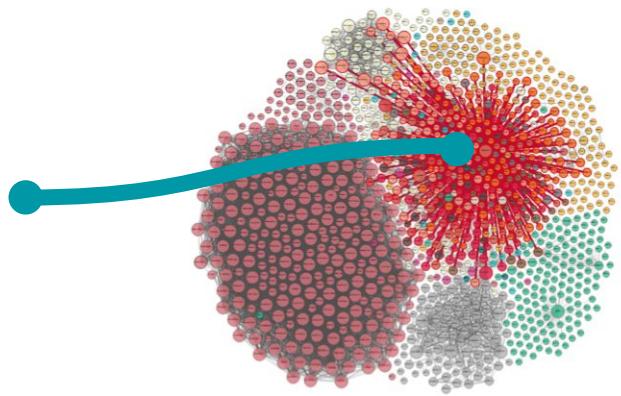
Déclarations

occupation

infirmier (Q186360)
professionnel de santé, dont la profession est de délivrer des soins infirmiers

> 1 référence

+ ajouter une valeur



Item [Discussion](#)

Lire

[Voir l'historique](#)

Plus ▾

[Rechercher dans DataCegeSom](#)

Andrée de Jongh (Q5)

résistant belge

 [modifier](#)

petit cyclone | Dédée

▼ Plus de langues [Configurer](#)

Langue	Libellé	Description	Alias
français	Andrée de Jongh	résistant belge	petit cyclone Dédée
anglais	Andrée de Jongh	Belgisch verzetsstrijder	Little Cyclone
néerlandais	Andrée de Jongh	Belgiance resistance member	
allemand	Andrée de Jongh	Aucune description fournie	

Déclarations

occupation



infirmier (Q186360)

 [modifier](#)

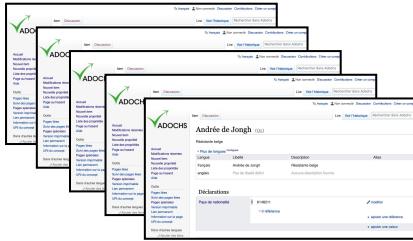
professionnel de santé, dont la profession est de délivrer des soins infirmiers

► 1 référence

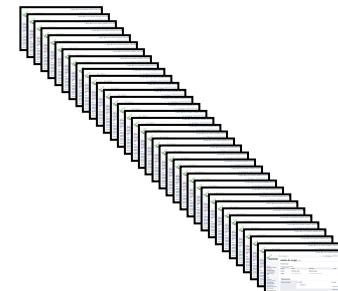
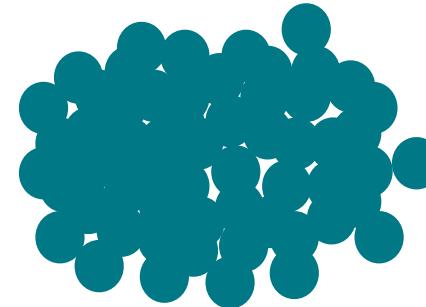
+ ajouter une valeur

Two processes run in parallel...

0. The data



Working on a sample



Working on the whole dataset

1. Cleaning and deduplicating the data

A	B	C	D	E	F	G	H
1	Nom	Prénom	Profession	Domicile	Région	Date naissa	Date naissa
2	Ackermann	Jean					
3	Aerden	Henri	directeur école aviation	Anvers	Flandre	1897-01-20	1897-01-20 marié
4	Albert	Philippe		Beauraing			
5	Albert	Romain		Bièvre	Wallonie	1921-02-15	1921-02-15
6	Altenhoff	Richard	ingénieur	Woluwe-St-Lamb	Bruxelles	1913-06-26	1913-06-26 cé.
7	Aneye	Julien	électricien	Schelle	Flandre	1901-08-28	1901-08-28 marié
8	Ancion	Jacques		Boncelles	Wallonie	1919-11-15	1919-11-15



```
[7]: # Import necessary libraries
from recordlinkage.datasets import load_fabrl1
dfA = load_fabrl1()
# Initialize step
Indexer = recordlinkage.Indexer()
Indexer.set_index_column('id')
candidate_links = Indexer.index(dfA)

# Comparison step
comparator = recordlinkage.Comparator()
comparator.set_index('id', name='given_name', label='given_name')
comparator.set_index('surname', name='surname', method='jaccard', threshold=0.85, label='surname')
comparator.set_index('date_of_birth', name='date_of_birth', label='date_of_birth')
comparator.set_index('state', name='state', label='state')
comparator.set_index('address', name='address', label='address')

Features = comparator.compute(candidate_links, dfA)

# Classification step
matches = Features.features.sum(axis=1) > 3
print(matches)

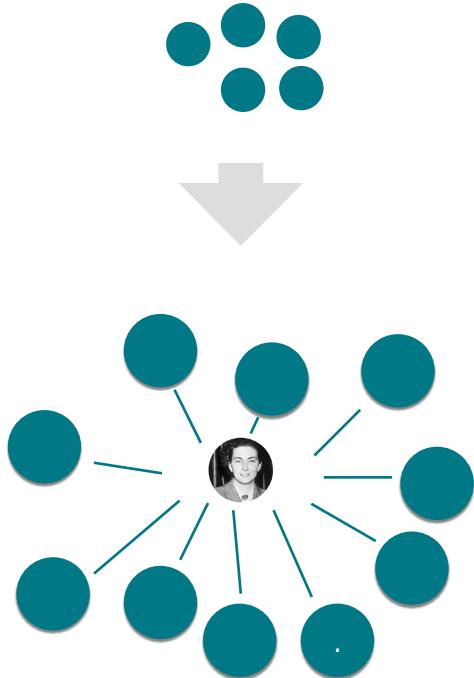
[18]:
```

given_name	surname	date_of_birth	state	address
rec_149-dap-0	rec_149-dap-0	1	1.0	1 1 1 1 1.0
rec_322-dap-0	rec_322-dap-0	1	1.0	1 1 1 1 1.0
rec_248-dap-0	rec_248-dap-0	1	1.0	1 1 1 1 1.0
rec_373-dap-0	rec_373-dap-0	1	1.0	1 1 1 1 1.0
rec_319-dap-0	rec_319-dap-0	1	1.0	1 1 1 1 1.0
rec_220-dap-0	rec_220-dap-0	1	1.0	1 1 1 1 1.0
rec_397-dap-0	rec_397-dap-0	1	1.0	1 1 1 1 1.0
rec_472-dap-0	rec_472-dap-0	1	1.0	1 1 1 1 1.0
rec_510-dap-0	rec_510-dap-0	1	1.0	0 1 1 1 1.0

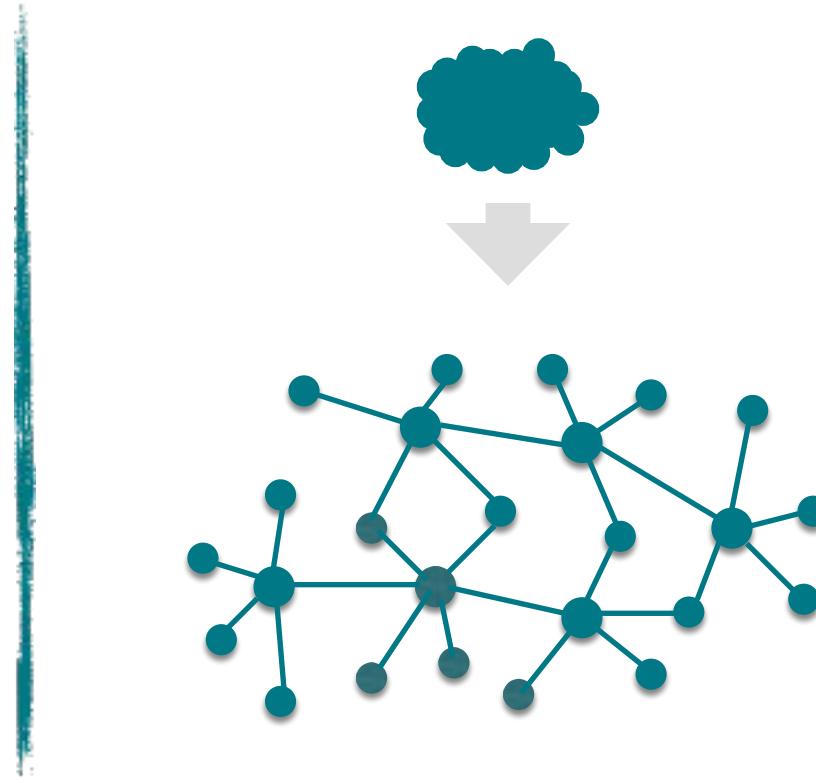
Manual process based on deductions

OpenRefine &
The Python Record Linkage toolkit

2. Reconciling names with external resources

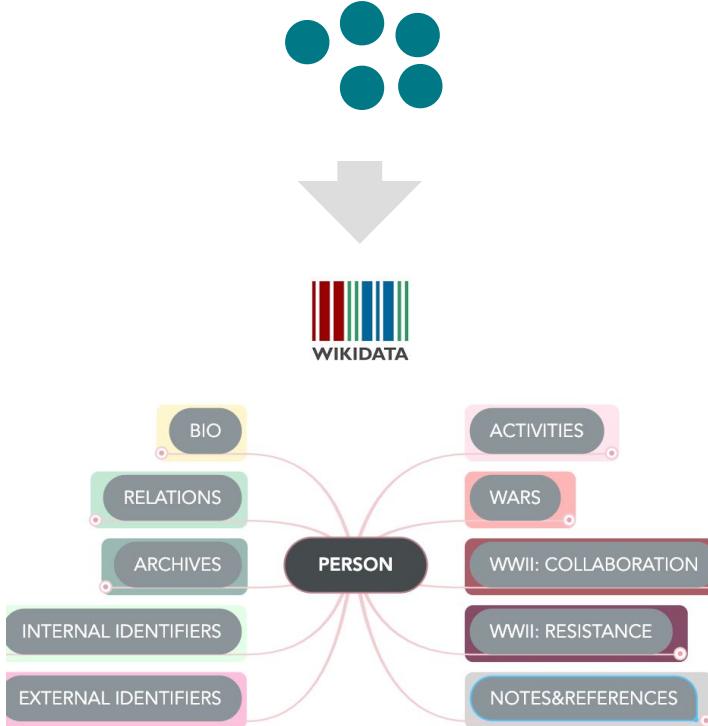


Custom research

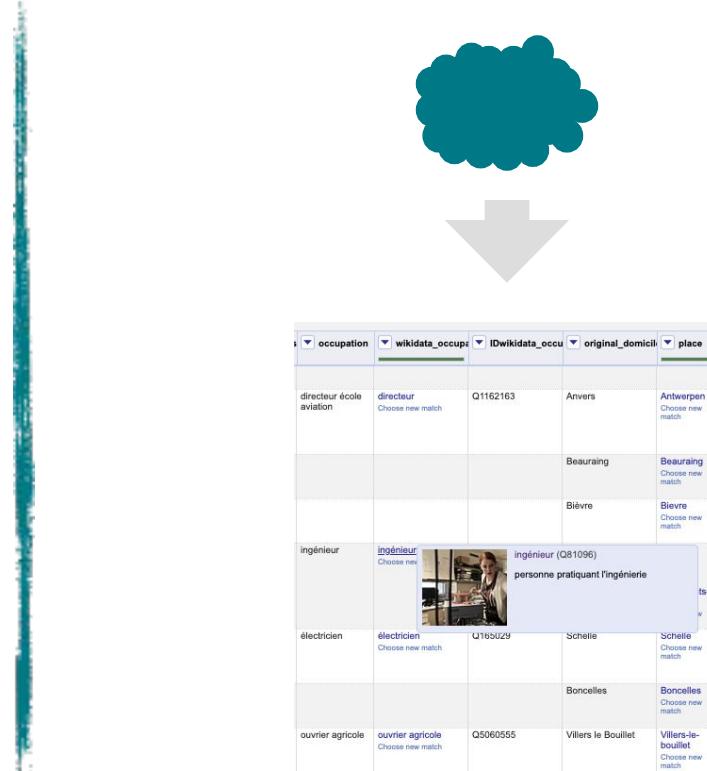


*OpenRefine reconciliation services
(Wikidata, Viaf, ULAN etc.)*

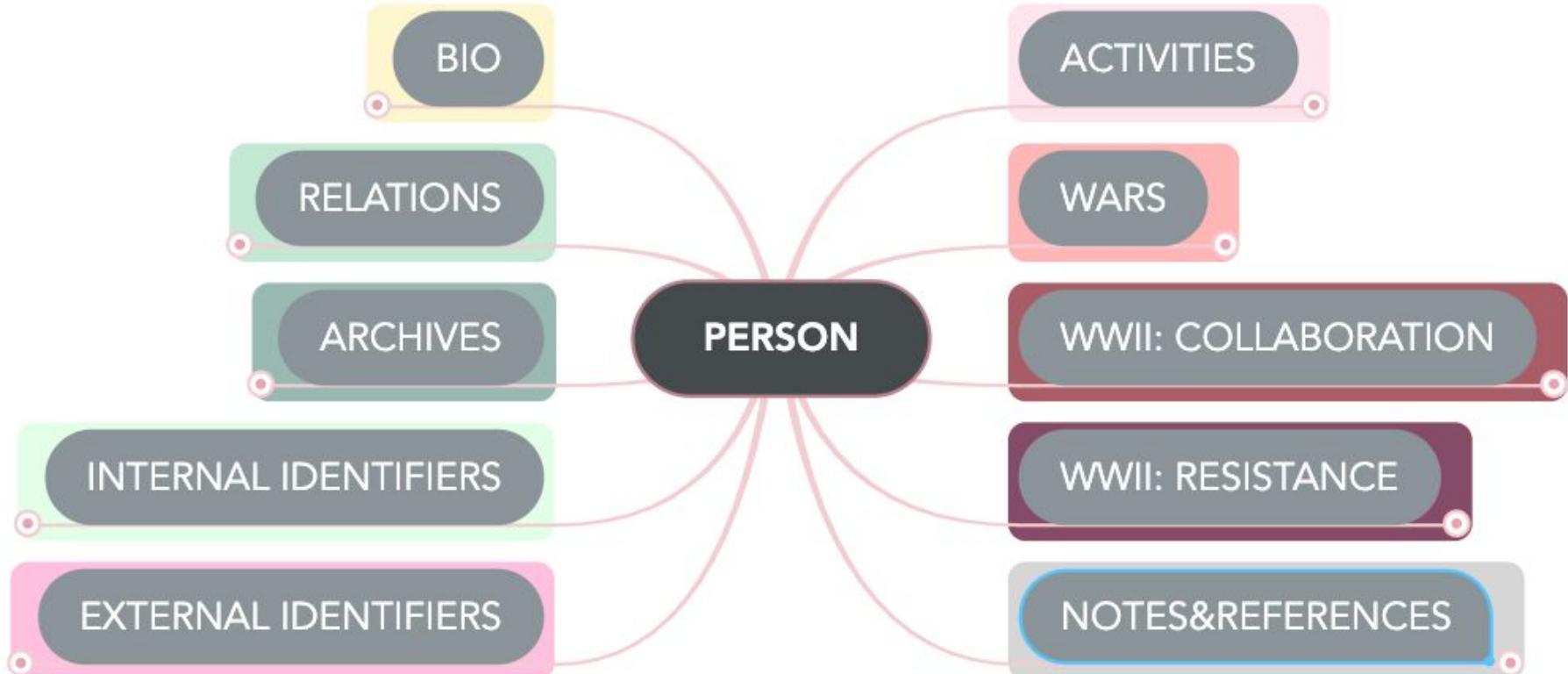
3. Modeling the data



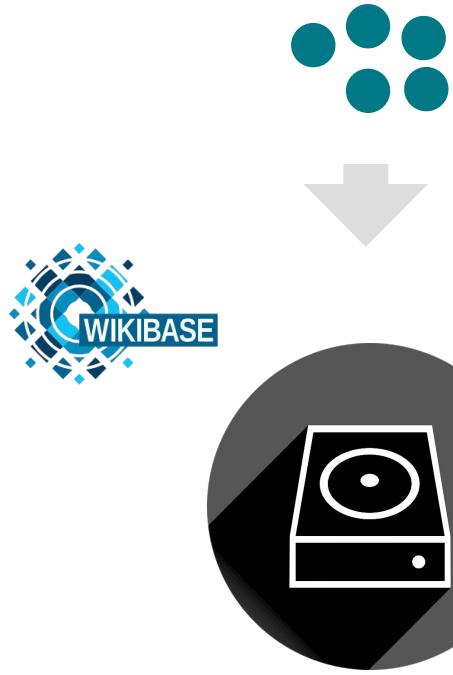
Data modeling based on a representative sample



From strings to things (URLs)



4. Running a Wikibase

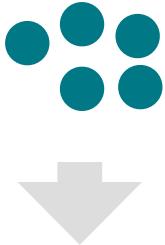


Locally



AGR server

5. Importing data



Andrée de Jongh (Q5)

Résistante belge

[modifier](#)

▼ Plus de langues [Configurer](#)

Langue	Libellé	Description	Alias
français	Andrée de Jongh	Résistante belge	
anglais	Andrée de Jongh	Aucune description fournie	
français canadien	Pas de libellé défini	Aucune description fournie	
néerlandais	Pas de libellé défini	Aucune description fournie	

Déclarations

occupation [infirmier \(Q186360\)](#) [modifier](#)

professionnel de santé, dont la profession est de délivrer des soins infirmiers

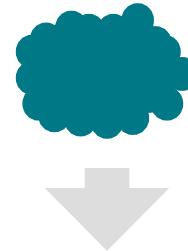
▼ 0 référence [+ ajouter une référence](#) [+ ajouter une valeur](#)

pays [État dont la personne possède la nationalité](#) [modifier](#)

État dont la personne possède la nationalité [enregistrer](#) [annuler](#) [\[?\]](#)

▼ 0 référence [+ ajouter un qualificatif](#)

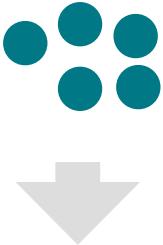
MediaWiki Graphical User Interface



Wikibase-edit
QuickStatements
Wikidata Integrator
Pywikibot
Heard Library Python script

Several solutions for
(Semi-)automatic data ingestion

6. Querying data



andrée

Andrée de Jongh
Résistante belge

contenant...

andrée

Résultat 1 parmi 1

Simple search box



Wikidata Query Service Examples Help More tools

```
1 # Instances of "lattice tower"
2 SELECT ?item ?itemLabel ?lat ?lon
3 WHERE {
4   ?item wdt:P31 wd:Q1440476; wdt:P625 ?pt.
5   ?item p:P625?loc.
6   ?loc psv:P625?node.
7   ?node wikibase:geoLatitude?lat.
8   ?node wikibase:geoLongitude?lon.
9   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
10 }
11 }
```

SPARQL Query Service





Item Discussion Lire Voir l'historique Plus Rechercher dans DataCage/Some

DOCHS

Henry Carton de Wiart (Q770)

écrivain et homme politique belge [modifier](#)

actions récentes Item le propriétai re propriétai rs utilisés

plus récentes Item le propriétai re propriétai rs utilisés

jeux

les pages liées spéciales imprimerable récents

lien sur la page concept

'autres langues ajouter des liens

environ et homme politique belge

+ Plus de langues [Continuer](#)

Langue	Libellé	Description	Alias
français	Henry Carton de Wiart	écrivain et homme politique belge	
anglais	Henri Carton de Wiart	Belgian former prime minister	
néerlandais	Henri Carton de Wiart	Belgisch politicus	
allemand	Henri Carton de Wiart	belgischer Politiker und Premierminister	

Déclaration

nature de l'élément [être humain](#) [modifier](#)

+ 0 référence

+ ajouter une référence

+ ajouter une valeur

pays de nationalité [Belgique](#) [modifier](#)

+ 0 référence

+ ajouter une référence

+ ajouter une valeur