

Centre for
Data Analytics

Insight 

Using Semantic Technologies to Create Virtual Families from Historical Vital Records

Christophe Debruyne^{1,2}, Oya Beyan¹, Stefan Decker¹ and Sandra Collins²

¹Insight @ NUI Galway

²Digital Repository of Ireland

2014-09-25 @ EUON 2014

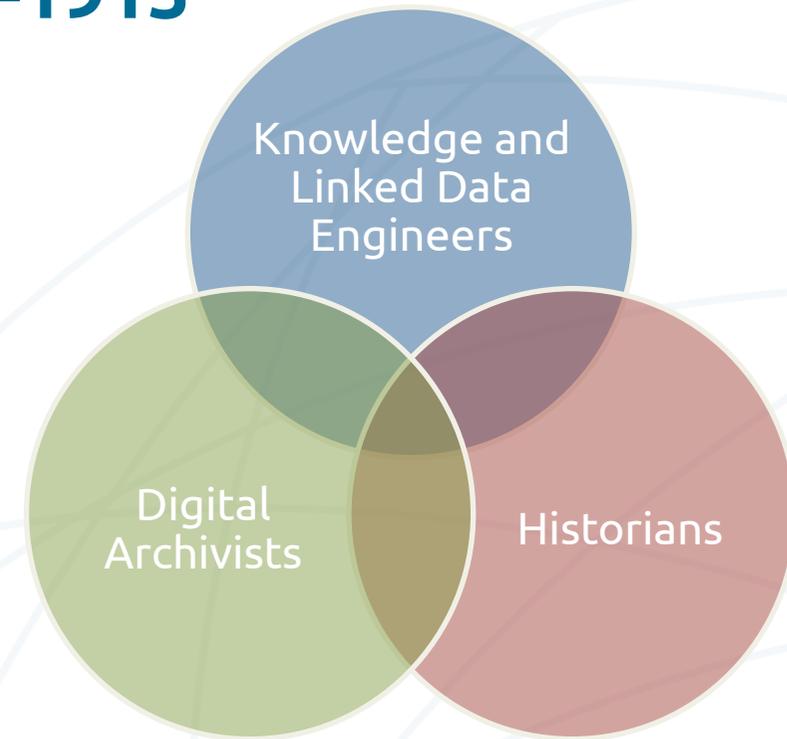


Irish Record Linkage, 1864-1913

Developing a platform applying semantic technologies to historical birth-, death and marriage certificates.

Answering questions such as: “How accurate are historic maternal mortality rates (MMR) and infant mortality rates (IMR) for Dublin?”

Team consists of researchers (historians), digital archivists, and knowledge engineers.



General Records Office

- Vital registration data
 - Birth-certificates
 - Death-certificates
 - Marriage records
- Digitised TIFF images of hardcopy indexes and registers.
- 2 TB of data
- Database describing the digitised records allowing searches on some fields.

First Page. [Please note that all Copies made on this Page should be certified at foot.] 04556605

Superintendent Registrar's District Bunmahon Registrar's District Bunmahon

1916 DEATHS Registered in the District of Bunmahon in the Union of Bunmahon
in the County of Wexford

No. (1)	Date and Time of Death (2)	Name and Surname (3)	Sex (4)	Condition (5)	Age last Birthday (6)	Rank, Profession, or Occupation (7)	Period of Incubation (8)	Signature, Qualification and Rank of Informant (9)	Date Registered (10)	Signature of Registrar (11)
55	9/10	John Spittle	Male	10	Chief of Police	3 Days	John Spittle	John Spittle	1916	John Spittle
62	9/10	John Spittle	Male	58	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
65	9/10	John Spittle	Male	80	Farmer	3 Days	John Spittle	John Spittle	1916	John Spittle
66	9/10	John Spittle	Male	87	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
67	9/10	John Spittle	Male	35	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
68	9/10	John Spittle	Male	71	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
69	9/10	John Spittle	Male	59	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
70	9/10	John Spittle	Male	62	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
71	9/10	John Spittle	Male	71	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle
72	9/10	John Spittle	Male	40	Wife of John Spittle	3 Days	John Spittle	John Spittle	1916	John Spittle

I, John Spittle, Registrar, do hereby certify that this is a copy of the original Book of Deaths within the said District, from the Book of Deaths of Bunmahon District, Wexford County, as kept by me, and as a true and correct copy of the same.

Witness my hand, this 10th day of September, 1916.

John Spittle
Superintendent Registrar.

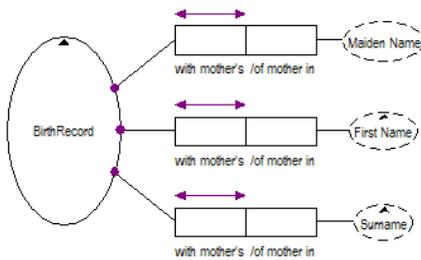
Challenges

- **With respect to requirements**
 - Identifying certified causes of death that can be attributed to maternal death.
 - Death certificates with no corresponding birth certificate
 - Terminology used pre-1900.
 - Capturing the socio-economical status of the families via, for instance, the professions, ranks of fathers.
 - ...
- **With respect to the platform**
 - Data protection
 - Records vs. Knowledge
 - Provenance vs. Interpretation

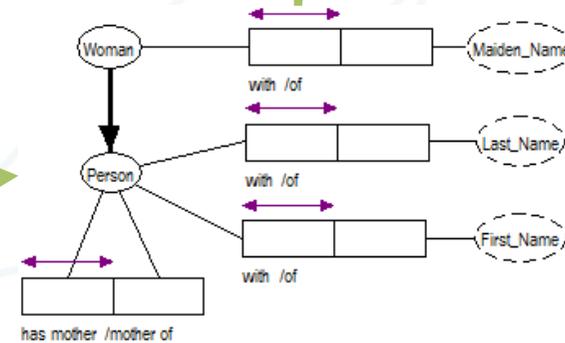
ISBN	Author	PubAddress
03-4472822	Random House	123 4th Street, New York
04-7733903	Wiley and Sons	45 Lincoln Blvd, Chicago
03-4858223	O'Reilly Press	77 Boston Ave, Cambridge
02-3920868	City Lights Books	99 Market, San Francisco

ISBN	Author	PubAddress	PubDate
345-28-2928	Haile Selassie		14-Aug-92
392-48-9965	Joe Blow		14-Mar-15
454-22-4012	Sally Hemmings		12-Sep-70
663-59-1254	Hannah Arendt		12-Mar-06

ISBN	Author	PubDate	Title
1-34532-482-1	345-28-2928	03-4472822	1990 Cold Fusion for Dummies
1-38482-995-1	392-48-9965	04-7733903	1985 Macrame and Straw Tying
2-35921-489-4	454-22-4012	03-4858223	1952 Fluid Dynamics of Aqueducts
1-39278-293-4	663-59-1254	02-3920868	1907 Beads, Baskets & Revolution



SEPARATION OF CONCERNS



Obviously, due to the sensitive nature of the data, data protection is key.

Transformation from one model to another

- SPIN – SPARQL Inference
- SWRL / RuleML
- SPARQL Construct
- ...

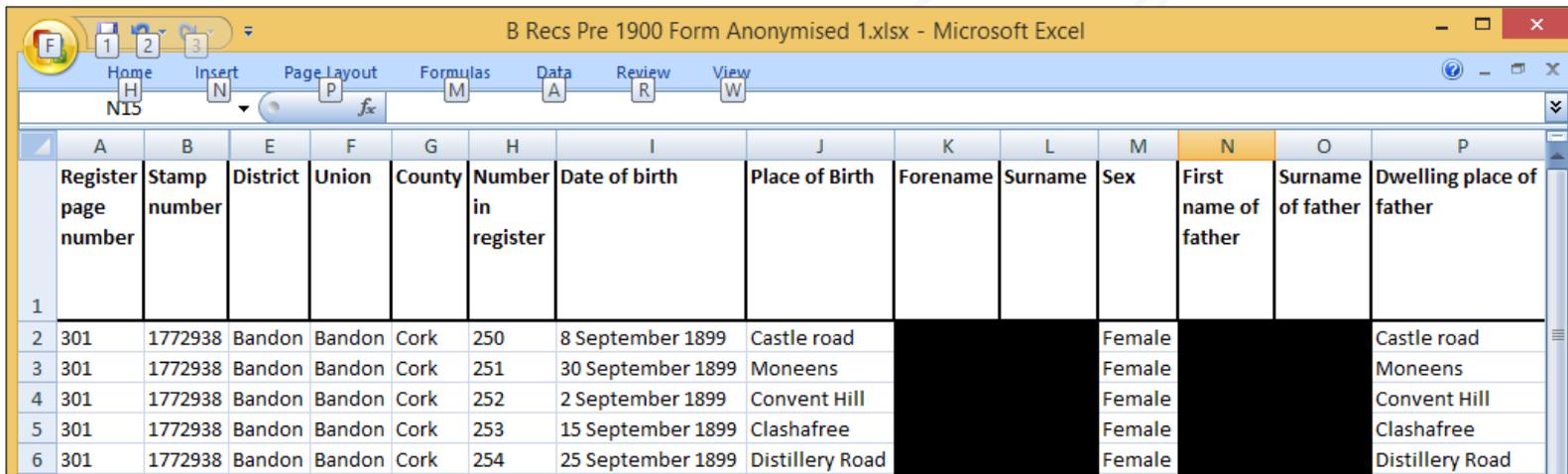
Development of 2 ontologies

- 2 ontologies were developed – separation of concerns
- First ontology for describing the contents of records
 - OWL 2 shallow, “flat ontology”
 - Created by “lifting” the structure of the vital records
 - (Marriage) Record, (Birth|Death) Certificate, Return
- Second ontology for data analysis
 - OWL 2 + Rules to capture background and domain knowledge
 - Created by means of Competency Questions (Grüninger and Fox)
 - Person, Birth, Marriage, Death, withChild, motherOf, ...

Grüninger, M., Fox, M.S.: The role of competency questions in enterprise engineering. In: Benchmarking Theory and Practice, pp. 22-31. Springer (1995)

Tool for the Digital Archivist

- Records are encoded using spreadsheets – a tool the digital archivist is familiar with



The screenshot shows a Microsoft Excel spreadsheet titled "B Recs Pre 1900 Form Anonymised 1.xlsx". The spreadsheet has 14 columns labeled A through P. The data is as follows:

	A	B	E	F	G	H	I	J	K	L	M	N	O	P
	Register page number	Stamp number	District	Union	County	Number in register	Date of birth	Place of Birth	Forename	Surname	Sex	First name of father	Surname of father	Dwelling place of father
1														
2	301	1772938	Bandon	Bandon	Cork	250	8 September 1899	Castle road			Female			Castle road
3	301	1772938	Bandon	Bandon	Cork	251	30 September 1899	Moneens			Female			Moneens
4	301	1772938	Bandon	Bandon	Cork	252	2 September 1899	Convent Hill			Female			Convent Hill
5	301	1772938	Bandon	Bandon	Cork	253	15 September 1899	Clashafree			Female			Clashafree
6	301	1772938	Bandon	Bandon	Cork	254	25 September 1899	Distillery Road			Female			Distillery Road

- RDF-to-RDF mapping files were defined to generate RDF from the in-memory databases created for each spreadsheet.

Next steps

- Encoding a significant amount of vital records in the excel files
 - To create the first triplestore; and
 - To obtain a dataset for validating the transformations; and
 - By consequence, validating the second ontology.
- To investigate proper interaction with the data for the historians.
- Linking the data with additional context; i.e., Linked Logainm
 - <http://data.logainm.ie/>
 - Nuno Lopes, Rebecca Grant, Brian Ó Raghallaigh, Eoghan Ó Carragáin, Sandra Collins, Stefan Decker: Linked Logainm: Enhancing Library Metadata Using Linked Data of Irish Place Names. TPDFL Workshops 2013: 65-76

More information

- @IRL_Project
- Project website <http://irishrecordlinkage.wordpress.com/>
- In partnership with



UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

dri 
Digital Repository of Ireland
Taisclann Dhigiteach na hÉireann