On a Linked Data Platform for Irish Historical Vital Records

Christophe Debruyne[1,2], Oya Deniz Beyan[2], Rebecca Grant[1], Sandra Collins[1], and Stefan Decker[2]

[1] Digital Repository of Ireland, Royal Irish Academy, Dublin, Ireland
{c.debruyne,r.grant,s.collins}@ria.ie
[2] Insight @ NUIG, National University of Ireland Galway, Galway, Ireland
{firstname.lastname}@insight-centre.org

Abstract. The Irish Record Linkage 1864-1913 is a multi-disciplinary project aiming to create a platform for analyzing events captured in historical birth, marriage and death records by applying semantic technologies for annotating, storing and inferring information from the data contained in those records. This enables researchers to, for instance, investigate to what extent maternal and infant mortality rates were underreported. We report on the semantic architecture, provide motivation for the adoption of RDF and Linked Data principles, and elaborate on the ontology construction process that was influenced by both the requirements of the digital archivists and historians. Concerns of digital archivists include the preservation of the archival record and following best practices in preservation, cataloguing and data protection. The historians in this project wish to discover certain patterns in those vital records. An important aspect of the semantic architecture is the clear separation of concerns that reflects those requirements – the transcription and archival authenticity of the register pages and the interpretation of the transcribed data – that led to the creation of two distinct ontologies and knowledge bases.

Keywords: Historical Vital Records, Cultural Heritage, Linked Data, Ontology Engineering, RDF Graph Transformation

1 Introduction

We report on the semantic architecture and ontology creation of the multi-disciplinary Irish Record Linkage (IRL) 1864-1913 project. The IRL project aims to create a knowledge base containing historical birth-, marriage- and death records translated into RDF and create a Linked Data [6] platform to analyze those events. The project involves the expertise of three disciplines [3]: historians, digital archivists and knowledge engineers. With the help of knowledge engineers creating the ontologies and setting up the platform and the digital archivists who curate, ingest and maintain the RDF, the historians will be able to analyze reconstructed "virtual" families of Dublin in the 19th and early 20th centuries,

allowing them to address questions about the accuracy of officially reported maternal mortality and infant mortality rates. To aid the historians in their data analysis, the knowledge engineers also contribute in linking people across records and the contextualization of the information with other datasets.

2 General Records Office

In Ireland, the General Register Office – GRO for short – is Ireland's civil registry responsible for recording information on births, deaths and marriages. In this project, the Registrar General of Ireland generously offered us records of 6,009,781 births (from 1864 to 1912), 4,314,963 deaths (from 1864 and 1912) and 1,443,110 marriages (from 1845 to 1912) under strict terms and conditions. It became compulsory to report and register births, deaths and marriages in 1864, but *non-Catholic* marriages were already being registered from 1845 onwards.[1] This explains the broader timespan for marriage records. *Records* of these events were captured on *register pages* (up to 10 per page for births and deaths, and up to 4 for marriages) divided by district and sent to the GRO where volumes were then created and an *index* compiled. The information was provided to us as a database dump of the GRO's database with digitized versions of the register pages and indexes.[2]

The information system the GRO has built allowed one to search for vital records concerning a person based on a person's name, geographical area (to the level of district) and year; one of their core services to the public. Not only has the GRO spent resources in the construction of such a service, an enormous amount of effort also went into the digitization of register pages and indexes as accurately as the recording of a subset of the information in a relational database. A rational decision was made to only enter in the database the information sufficient to efficiently find records. While the system developed by the GRO works perfectly for finding historical records, information that is key in answering the IRL historians' questions were not captured by the database (such as the places of death, names of the informant, etc.). As such, we should call on the expertise of digital archivists – trained in processing, transcribing and curating the information – in preparation for the Linked Data platform to be developed.

The vital records and the goals of the IRL project lead to various challenges that need to be taken into account and those challenges reside at different levels: data protection, data transcription, historical evolution (medical knowledge, geographical, etc.) and, of course, the method for answering the historians' research questions. We will highlight some of the pertinent challenges below that will influence the design of the semantic architecture and the transcription workflow.

[1] `http://www.irish-genealogy-toolkit.com/Irish-marriage-records.html`

[2] The terms and conditions of our data sharing agreement do not permit us to make public any data that would identify any individual [3]. One can access the historic records of the GRO at its dedicated research room in Dublin, but it is restricted per diem and there is an associated charge.

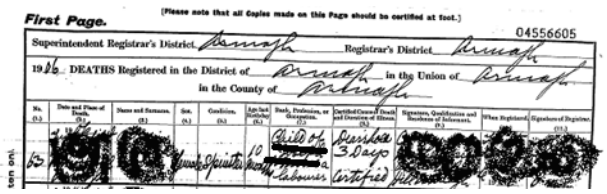

Fig. 1. Part of a register page containing death records (redacted as per our data sharing agreement). Copyright held by the General Register Office and reproduced with permission.

Data security and protection in terms of transfer, storage and use by authorized parties were covered by the data sharing agreement with the GRO. The goal of the IRL project is to build a platform that allows one to analyze the data captured in those records and not to replace the service already built by the GRO, although the new platform would support the queries typically executed by GRO as well. As per our data sharing agreement, the dataset in its entirety (that means data and digitized objects) should only be available to the members of the project team. With the help of the digital archivist, who is familiar with data protection legislation and best practices, we furthermore identified which guidelines to follow.

Records, knowledge and interpretation. Another challenge is the varying levels of detail in the records (seen in, for instance, the causes of death) and the variances in how subject names and places were recorded (initials, short hands, name of a building versus street name, etc.) [3]. These variances might imply something, which we are currently unaware of. Therefore, we should ensure that the transcription of the register pages transcribes exactly what was written down. In other words, the manipulation of the information should be kept to a minimum. This leads to another, yet related challenge, *clearly separate two concerns*: the exact transcription of what has been captured on the register pages as to have an authentic virtual account of historic events; and the interpretation, possibly with background knowledge, of certain aspects based on these interpretations. Examples of how interpretation can differ are the evolution of Ireland's geography (place names changing and streets disappearing, merging and even reappearing), evolution in knowledge (e.g., new insights in medicine) and even the adoptions of different theories (e.g., different classifications of social status).

Provenance and archival authenticity. Archival theory is based on two key principles, *respect de fonds* (original order) and *archival provenance*. Respect de fonds is the principle which guides archivists when exerting intellectual control over a collection, and ensures that the archival record is always described in relation to the context in which it is created as far as possible (for example a letter should only be described in terms of a set of correspondence where it is available). We follow this principle by transcribing not a line of data about an individual, which is meaningless in an archival context, but the entire register page that constitutes an archival record or object. The principle of *respect de fonds* is linked closely to provenance, which forms

the foundation of archival description. Provenance refers to how the archival record relates to its creator, and can only be maintained through the appropriate description of an archival record. These principles are important in the digital sphere, and describing and authenticating records in this way gives meaning through the provision of context.

Other data challenges include the conversion to appropriate data formats as well as cataloguing of the digitized objects so as to ensure compliance with digital preservation best practices. These challenges, however, fall outside of the scope of this paper; work on the ingestion of the digitized objects in a suitable digital long-term preservation platform will be disseminated elsewhere.

3 IRL Semantic Architecture

This paper focuses on the semantic architecture on which the user interfaces for data analysis will be built. These interfaces are currently being developed and investigated, and will be reported elsewhere (see Section 8). The architecture is set up to cope with the requirements defined by the data challenges described in the previous section and the research questions the historians aim to address. Fig. 2 depicts graphically our architecture in which the two aforementioned concerns – exact transcription on the left vs. interpretation on the right – are strictly separated. We will first motivate the adoption of RDF and semantic technologies and discuss some aspects of each concern. Details on the ontologies developed for this platform will be discussed in subsequent sections and build further upon the work reported in [3].

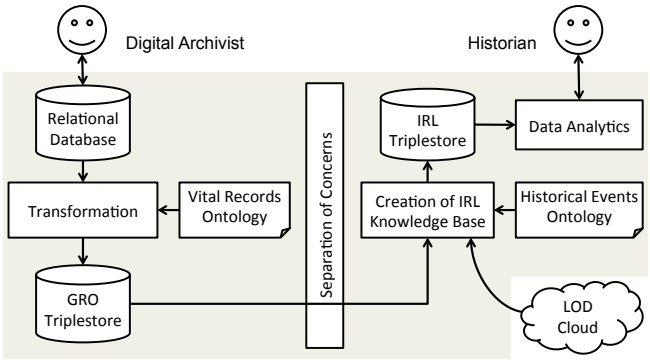

Fig. 2. The conceptual architecture of the IRL Linked Data platform. Transcription of register pages and the interpretation of the data are strictly separated.

RDF and Linked Data principles were adopted for various reasons. RDF allows us to use a simple data model that facilitates the integration of internal and external data by creating links. Using RDF, the management of knowledge is scalable, and data access – for analysis, amongst others – is pushed closer

to the user and application level by adopting the Linked Data principles (e.g., content negotiation) and the W3C SPARQL recommendation.

By reusing the existing HTTP infrastructure on which Linked Data is built, datasets that are behind firewalls can still link to other datasets in the Linked Data cloud. This allowed us to take a conservative approach by setting up our services behind a firewall and create (and exploit) outbound links; we thus benefit from all the Semantic Web technologies and the Linked Data cloud has to offer without violating our data sharing agreement and data protections legislations. Datasets relevant for this project that provide additional context include DBpedia [1] and Linked Logainm [13]. The latter is a Linked Data version of the authoritative bilingual database of Irish place names `logainm.ie`. Linked Logainm also provides links to places in DBpedia and `geonames.org`.

OWL 2 was adopted for the creation of the two ontologies allowing us to infer implicit information and rule languages were adopted to encode domain expert knowledge (historical, medical, etc.) to infer additional information that falls outside the capabilities of OWL.

There are four principles that Linked Data datasets should adhere to [2]: 1) use URIs as names for things; 2) use HTTP URIs so that people can look up those names; 3) provide information with standards (e.g., RDF) when URIs are looked up; and 4) include links to other URIs. Principles 1 to 3 are adhered to by both triplestores. The GRO triplestore provides links to other URIs within the same dataset to avoid interpretation and contextualization. The IRL triplestore links to external datasets to provide that contextualization. Since the datasets are behind a firewall, inbound links are not possible. Outbound links can be followed to discover more information. The authors are aware that the firewall can pose problems if one wishes to execute federated queries (across different datasets), but this has not yet been encountered within the context of this project.

For the platform, we adopted Jena TDB as triplestores and Jena Fuseki to provide the SPARQL endpoints.[3] Pubby is used to create a simple Linked Data frontend via those endpoints.[4] Details on the technologies adopted for the generation of RDF triples from the relational database and the transformation of triples for the interpretation of the data will be provided in the next sections.

4 Transcription of the Register Pages

We reiterate that the existing system the GRO has built took into account the attributes necessary to find records about individuals, thereby leaving out all fields on the register pages that were not relevant for this task. The digital archivists thus have the meticulous and laborious task of transcribing all that was captured on register pages, which is not merely transcribing those records, but also involves undertaking research and controlling the quality of what has been transcribed. Adopting Optical Character Recognition (OCR) was not possible as a very high level of precision in the transcription process was necessary.

[3] `http://jena.apache.org/`
[4] `http://wifo5-03.informatik.uni-mannheim.de/pubby/`

In order to cope with the tension field of transcribing exactly what has been written down and the normalization of the data in some of these fields, a relational database has been set up that can capture in greater detail what can be observed on a register page. On death register pages, for instance, one can find a field "Certified Cause of Death and Duration of Illness". We observed variances in detail, which depended for instance on the registrar or on the informant (practitioner vs. relative). That field was sometimes used to indicate that the cause of death was uncertified. The database thus provided an additional field to indicate whether a death was explicitly certified, explicitly uncertified or neither. The duration of illness can be unknown or not applicable, e.g., in the case of drowning. The field can thus be NULL in case no information was provided.

Notes for each record and register page can be kept to capture anomalies or peculiarities such as signatures with a cross or crossed out information. As the project continues and the digital archivists transcribe register pages, these notes could be used as input for the creation of a controlled vocabulary for anomalies in register pages (see future work).

The database schema was developed in such a way that the data entered adheres to certain integrity constraints, thus effectively preventing certain errors. This relational database is then annotated with the Vital Records Ontology, presented in the next section, using D2RQ [5] and the generated triples are stored in a records triplestore.

5 Vital Records Ontology (VRO)

Births, deaths and marriages were captured per district (within a union, within a county) as single records on register pages. These pages can contain up to 10 records after which such a page is signed off by the registrar and sent to the superintendent registrar for inspection and validation. To create a first version of the Vital Records Ontology (VRO)[5], we "lifted" the information one could see on one such register page to an ontology.

To minimize interpretation, we choose to develop a "flat" ontology, which means that most information that can be found on such a register page was captured as literals. For example, instead of creating a concept `Person` that can have a `forename` and `surname`, we choose to relate the concept of a `Record` to these attributes. For the VRO, we thus defined a few concepts. A `RegisterPage` and a `Record` for representing the different types of records were declared. Each record must belong to a register page and each register page can have zero (which implies a blank pages) or more records. We make a distinction between a `Certificate` and a `MarriageRecord`, both of them being disjoint subclasses of the concept `Record`. The first has as a subject only one person and the latter two. The two concepts are disjoints, which makes that no instance of a certificate can be an instance of a marriage record and vice versa. Finally, we created two disjoint subclasses of the concept `Record`: `BirthRecord` and `DeathRecord`. The

[5] Available via `http://purl.org/net/irish-record-linkage/records`.

only object property, a relation between two concepts, we needed was to relate records to register pages. All other properties are datatype properties. Datatype properties are related to the greatest common denominator. For instance, all records are signed off by a registrar on a certain date. The date of registration as well as information on the registrar are therefore related to the concept of `Record` so that all subtypes of this class inherit this property.

One of the challenges is to capture the domain as well as possible, yet maintain a valid OWL 2 ontology. As explained by Motik and Horrocks in [14], it is difficult to reason about date and time intervals, and therefore only specific points in time (captured by both `xsd:dateTime` and `xsd:dateTimeStamp`) were "amenable for implementation" and those "can be handled by techniques similar to the ones for numbers." Together with the digital archivist, we choose not to capture dates mentioned in records as instances of `xsd:dateTime` as we do not know the exact times and we felt that encoding "default" times would not be in keeping with archival principles. We thus chose to declare the range of these properties as being `rdfs:Literal`, but provided transcription guidelines in which the use of `xsd:date` was to be highly encouraged.

One key requirement for Linked Data platforms in general is adequate identifiers. For our records knowledge base, we need to identify instances of records and register pages. Each register page and record is identified by a URI under the new subdomain `http://irl.dri.ie/`. Register pages are identified by a unique, physically stamped number provided by the GRO while digitizing. We use this stamp number for the creation of URIs identifying register pages. Individual records are identified by the combination of the stamp and entry-number. Fig. 3 depicts the triples from a death record on a register page of a woman who died of paralysis in the year 1890.

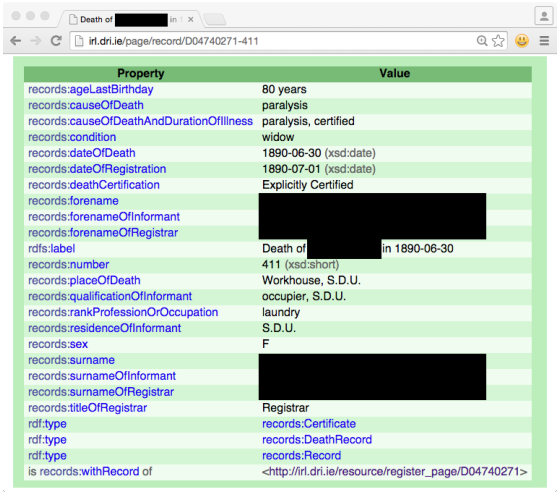

Fig. 3. Example of the triples from a death record in a register page.

6 Interpretation of the Register Pages and Records

We already described the importance of separating the information captured in the register pages and the interpretation thereof. The ontology that needs to support that kind of interpretation of the GRO data is more challenging given that the historians wishing to analyze the content are not necessarily familiar with ontology engineering and the knowledge base needs to support their activities, we adopted – reported in [3] – the approach proposed by Grüninger and Fox of having the stakeholders formulating *competency questions* [12]. The ontology must contain a necessary and sufficient set of axioms to represent and solve these questions [12]. These competency questions are not used to generate an ontology, but rather to evaluate it [11]. Using the types of queries the stakeholders wish to see answered, the knowledge engineers built an ontology, which was specifically tailored for the project, yet aimed to reuse existing, established vocabularies where possible. Competency questions formulated by historians included (paraphrased from [3]): "How many women died within n days after childbirth due to complications related to labor [...]?" and "What is the average sibship interval where the first child did not survive under various socio-economic conditions?" Those questions can be broken down in smaller competency questions such as: "Which infants died within the first 24 hours of their life?" and "What was the cause of death of a person?"

The questions were analyzed to identify the concepts and relations for the ontology, which were validated by the stakeholders. Graphical representations of the developed ontologies were used during discussions, e.g., as shown in Fig. 4. The VRO serves to reflect the historical records. Although it contains information about *events*, *people*, *places*, etc., the VRO does not capture these as *distinct entities*. However, to reconstitute families and analyze, we need distinct representations of events and persons involved. Therefore we developed the *Historical Events Ontology* (HEO) on top of the VRO as to provide a base ontology for answering the competency questions. The choice was made not to declare these concepts in the VRO as they fulfill the requirement of one particular set of tasks. This strict **separation of concerns** would allow for a **greater reuse of the historical records** for different kinds of analyses.

We looked at existing ontologies for reuse and integration as well as the creation of missing concepts and relations for the creation of the HEO. To describe people, we take into account FOAF[6] and the Persona Vocabulary[7]. Both are used to describe people, their activities and their relations to other people and objects. The latter has more relations such as `hasChildren`.

As the project aims to reconstitute families and health histories of people, we also included concepts related to time (events), relations, and reused available domain disease ontologies [7]. The construction of the HEO also included

[6] Friend-of-a-Friend: `http://xmlns.com/foaf/spec/`
[7] `http://wiki.eclipse.org/Persona_vocabulary`

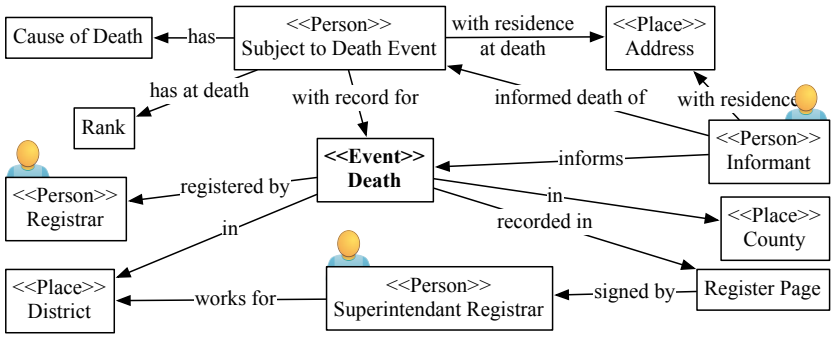

Fig. 4. Concepts and relations in the Historical Events Ontology for deaths.

formalizing information found in classification systems such as the International Statistical Classification of Diseases and Related Health Problems.[8]

Some of the concepts in the HEO are: `Person` for those involved during the event or registration; `Event` to capture the recorded births, deaths and marriages; `Place` for locations related to events or people; `CauseOfDeath` to facilitate reasoning and classifying causes of death; `Rank` for capturing the rank and occupation of involved persons; and `RegisterPage` to assure provenance. In a first instance, the data from the first triplestore is transformed to populate concepts and relations in the HEO by a series of SPARQL CONSTRUCT queries and SWRL rules. For instance, the following query allows us to create instances of the class `foaf:Person` from death records (prefixes omitted):

```
CONSTRUCT { ?new a foaf:Person; rdfs:seeAlso ?r;
                foaf:firstName ?f; foaf:familyName ?s.
}  WHERE { ?r a rec:DeathRecord; rec:forename ?f; rec:surname ?s.
           BIND (URI(CONCAT(STR(?record),"/person")) AS ?new). }
```

Transforming graphs from the first knowledge base into the second leads to the creation of many persons. Matching techniques are adopted to identify the same persons across different vital records to assert `owl:sameAs` statements. This is an important as some names are very common and women adopted the name of their husband after marriage. Other fields (place, time) need to be taken into account to properly identify the same persons across records. When transforming the graphs from the first knowledge base into the second, many instances of persons are created. Another goal of the IRL platform is to add contextual information from other datasets [3]. We adopted Linked Logainm [13] for information on Irish place names and links with DBpedia resources.

7 Discussion and Related Work

On Extracting RDF from Databases. Though D2RQ does not yet fully support the R2RML W3C recommendation[9], it proved to be easy to test the

[8] `http://apps.who.int/classifications/icd10/browse/2010/en`
[9] `http://www.w3.org/TR/r2rml/`

mappings using the built-in Linked Data server. This server also allowed one to access the database's content with SPARQL. D2RQ also comes with means to generate RDF dumps that can be used to populate a triplestore. Tools that support R2RML do exist, such as XSPARQL [4] – which has been extended to support R2RML, see [8] – and RML [9]. Though D2RQ so far accommodates our needs, "porting" the mapping to R2RML and investigate its different implementations will be investigated in the future.

On the Digitized Objects and the Transcriptions. We explained the reason why the Linked Data platform was placed behind a firewall in Section 3. Although not part of this project, one could investigate which subsets of the knowledge bases, and in particular the one containing historical events, do not violate the agreement and could be of benefit to the scientific community. The GRO also digitized the indexes for finding individual records. Indexes are currently not transcribed as they provide no additional information for our data analysis and individual records can be queried with SPARQL.

Important to consider in the future is the long-term preservation of the digitized objects and their RDF transcriptions. The Digital Repository of Ireland (DRI, `http://www.dri.ie`) is the national trusted digital repository for Ireland's social and cultural data. The DRI platform supports the ingestion of digitized objects and metadata, including Qualified Dublin Core (QDC) and Encoded Archival Description (EAD), and the configuration of access policies and licenses for these objects. Each object receives a Digital Object Identifier which will be referred to by the RDF transcription via, for instance, `rdfs:seeAlso` statements. We create an RDF file for each register page and related records by executing SPARQL DESCRIBE queries (an example with prefixes omitted is shown below). Those files are then used as input to create QDC files via an appropriate XSPARQL mapping.

```
DESCRIBE * WHERE { ?page r:stampNumber "4740271"; r:withRecord ?record. }
```

On Ontology Engineering. The digital archivists keep track of any anomalies or peculiarities in the register pages and individual records in a notes field in the database. Examples of anomalies include strikethroughs in fields or the occurrence of crosses where signatures are necessary. The first *could* indicate a correction or removal of information, the latter could indicate an illiterate person. We carefully chose to use the verb "could" as these are historical vital records and we should not give an interpretation to these anomalies when we are not sure. Depending on the nature of these anomalies and their frequency, we could consider using these for the creation of a controlled vocabulary; allowing one to look up these anomalies and decide how to interpret them. This vocabulary, captured as an ontology, would then reside next to the VRO.

8 Conclusions and Future Work

We reported on the creation of the semantic architecture, the ontologies and knowledge bases of the IRL Linked Data platform. Taking into account the

requirements of both the digital archivists (archival authenticity, preservation, cataloguing and data protection) and the historians (answering their research questions), the Linked Data platform is comprised of two distinct knowledge bases, each supported by a different ontology, to separate those two concerns: the Vital Records Ontology for the exact transcription of the historical vital records and register pages, and the Historical Events Ontology for an interpretation of the register pages. The creation of the first was fairly straightforward and primarily the result of a collaboration between the knowledge engineers and digital archivists. The latter also involved the historians who were asked to formulate competency questions to identify concepts and relations. Reasoning provides one motivation for adopting semantic technologies. The second is the creation of links with other datasets providing additional context to interpret the data. As the transcription of register pages is a laborious process, the latter can only be meaningfully evaluated when we have an adequate number of transcriptions.

The lessons learned in this study arise from the value of the separation of concerns. Though digital archivists could have elicited facts from the register pages immediately and solely fit for answering the competency questions in this project, the resulting dataset would have had limited value for reuse and future research questions. We argue that the return in value justified the extra overhead in terms of transcription and platform complexity. Our approach is thus different from, for instance, the Dacura platform [10], which adopts crowdsourcing techniques to elicit facts from datasets such as newspaper articles according to a schema for a particular purpose.

Future work that we will prioritize will be the ingestion of the digitized images and their RDF in a long-term preservation platform according to best practices and standards and the investigation to what extent parts of the knowledge bases can be made available to the public without revealing the details of individuals. Finally, this paper focused on the semantic architecture upon which applications can be built. The user interfaces which will aid the historians in answering their research questions built on top of the semantic architecture are still being investigated and will be reported elsewhere.

Acknowledgements We thank the Registrar General of Ireland for permitting us to use this rich digital content contained in the vital records for the purposes of this research project. This publication has emanated from research conducted within the Irish Record Linkage, 1864-1913 project supported by the RPG2013-3; Irish Research Council Interdisciplinary Research Project Grant, and within the Science Foundation Ireland Funded Insight Research Centre (SFI/12/RC/2289). The Digital Repository of Ireland (formerly NAVR) gratefully acknowledges funding from the Irish HEA PRTLI programme.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.G.: DBpedia: A Nucleus for a Web of Open Data. In: Aberer, K., Choi, K., Noy, N.F., Allemang,

D., Lee, K., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Lecture Notes in Computer Science, vol. 4825, pp. 722–735. Springer (2007)
2. Berners-Lee, T.: Linked Data - Design Issues. Last accessed: June 7th, 2015. (2006), http://www.w3.org/DesignIssues/LinkedData.html
3. Beyan, O., Breathnach, C., Collins, S., Debruyne, C., Decker, S., Grant, D., Grant, R., Gurrin, B.: Towards Linked Vital Registration Data for Reconstituting Families and Creating Longitudinal Health Histories. In: KR4HC Workshop (in conjunction with KR 2014). pp. 181–187 (2014)
4. Bischof, S., Decker, S., Krennwallner, T., Lopes, N., Polleres, A.: Mapping between RDF and XML with XSPARQL. J. Data Semantics 1(3), 147–185 (2012)
5. Bizer, C.: D2R MAP - A database to RDF mapping language. In: King, I., Máray, T. (eds.) Proceedings of the Twelfth International World Wide Web Conference - Posters, WWW 2003, Budapest, Hungary, May 20-24, 2003 (2003)
6. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)
7. Bodenreider, O.: Disease Ontology. In: Dubitzky, W., Wolkenhauer, O., Cho, K., Yokota, H. (eds.) Encyclopedia of Systems Biology, pp. 578–581. Springer New York (2013)
8. Dell'Aglio, D., Polleres, A., Lopes, N., Bischof, S.: Querying the Web of Data with XSPARQL 1.1. In: Verborgh, R., Mannens, E. (eds.) Proceedings of the ISWC Developers Workshop 2014, co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 19, 2014. CEUR Workshop Proceedings, vol. 1268, pp. 113–118. CEUR-WS.org (2014)
9. Dimou, A., Vander Sande, M., Colpaert, P., Verborgh, R., Mannens, E., Van de Walle, R.: RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In: Bizer, C., Heath, T., Auer, S., Berners-Lee, T. (eds.) Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014), Seoul, Korea, April 8, 2014. CEUR Workshop Proceedings, vol. 1184. CEUR-WS.org (2014)
10. Feeney, K.C., O'Sullivan, D., Tai, W., Brennan, R.: Improving curated web-data quality with structured harvesting and assessment. Int. J. Semantic Web Inf. Syst. 10(2), 35–62 (2014), http://dx.doi.org/10.4018/ijswis.2014040103
11. Fox, M.S., Gruninger, M.: Enterprise modeling. AI magazine 19(3), 109–121 (1998)
12. Grüninger, M., Fox, M.S.: The role of competency questions in enterprise engineering. In: BenchmarkingTheory and Practice, pp. 22–31. Springer (1995)
13. Lopes, N., Grant, R., Ó Raghallaigh, B., Ó Carragáin, E., Collins, S., Decker, S.: Linked Logainm: Enhancing Library Metadata Using Linked Data of Irish Place Names. In: Bolikowski, L., Casarosa, V., Goodale, P., Houssos, N., Manghi, P., Schirrwagen, J. (eds.) Theory and Practice of Digital Libraries - TPDL 2013 Selected Workshops - LCPD 2013, SUEDL 2013, DataCur 2013, Held in Valletta, Malta, September 22-26, 2013. Revised Selected Papers. Communications in Computer and Information Science, vol. 416, pp. 65–76. Springer (2014)
14. Motik, B., Horrocks, I.: OWL Datatypes: Design and Implementation. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T.W., Thirunarayan, K. (eds.) The Semantic Web - ISWC 2008, 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008. Proceedings. Lecture Notes in Computer Science, vol. 5318, pp. 307–322. Springer (2008)