

Report on the Linked Logainm project

Rebecca Grant
Nuno Lopes
Catherine Ryan





Linked Logainm steering committee

Nuno Lopes – Digital Enterprise Research Institute; Sandra Collins, Rebecca Grant – Digital Repository of Ireland; Catherine Ryan – Digital Repository of Ireland/National Library of Ireland; Úna Bhreathnach, Mairéad Nic Lochlainn, Caoilfhionn Nic Pháidín, Brian Ó Raghallaigh– Fiontar, Dublin City University; Eoghan Ó Carragáin, Ger Wilson – National Library of Ireland; Dónall Mac Giolla Easpaig, Pádraig Ó Cearbhaill, Micheál Ó Conaire – Placenames Branch, Department of Arts, Heritage and the Gaeltacht.

Linked Logainm project team

Úna Bhreathnach, Rebecca Grant, Nuno Lopes, Mairéad Nic Lochlainn, Eoghan Ó Carragáin, Brian Ó Raghallaigh, Catherine Ryan.

We gratefully acknowledge the help of James Wogan (Decipher Project), and Sarah Gearty and Jennifer Moore (Irish Historic Towns Atlas, Royal Irish Academy).



This work is licensed under a Creative Commons Attribution 3.0 Ireland License.

When citing or attributing this work, please use the following: Grant, R., Lopes, N., Ryan, C. (2013), 'Report on the Linked Logainm project'. Dublin: Royal Irish Academy and National Library of Ireland; Galway: NUI Galway.

DOI: 10.3318/DRI.LODer.2013.1
Digital Repository of Ireland series, no. 3

Executive summary

Linked Logainm is a collaborative project undertaken by the Digital Repository of Ireland (DRI), the Digital Enterprise Research Institute (DERI), Fiontar at Dublin City University (DCU), the National Library of Ireland (NLI) and the Placenames Branch of the Department of Arts, Heritage and the Gaeltacht.

The Linked Logainm project has created a Linked Data version of the authoritative bilingual database of Irish place names, logainm.ie, which is being developed by Fiontar in collaboration with the Placenames Branch. The new Linked Logainm dataset provides Irish place name data in a structured, computer-readable format, which allows its value to be fully exploited by collaborators, web developers, computer scientists, the heritage community and information professionals. It also encourages the international dissemination of Irish language place names, which have been given legal status by the Department of Arts, Heritage and the Gaeltacht.

As Linked Data, the place names included in logainm.ie can be linked to external geographic data on the Web, such as place name information from wikipedia.org. In practice, this means that heritage institutions like NLI can integrate Linked Logainm into their catalogues, and pull in additional contextual information of interest to researchers such as climate and population statistics or governance information for a particular place. It can also be used to link together digitised heritage collections of Irish interest from multiple sources such as europeana.org.

The Location LODer demonstrator website gives an interactive introduction to the potential of the Linked Logainm concept, using logainm.ie data and a map interface to allow users to explore content from sources including Europeana, NLI and Wikipedia.

The Digital Repository of Ireland and the National Library of Ireland will support the dissemination of Linked Logainm by engaging with the Irish heritage community and by providing a detailed introduction and guidelines to the implementation of Linked Data in cataloguing workflows.

This is a wonderful example of a self-started project by leaders in the field, to demonstrate what can be achieved when cutting-edge ICT technology is applied to trusted Irish digital content. All the partners have dedicated their time and expertise to making this trusted high-quality Irish data accessible for the public good through state-of-the-art technology, and the widespread open adoption of Irish place names via best practice in digital platforms.

Linked Logainm

Linked Logainm is a collaborative project led by the Digital Repository of Ireland (DRI), the Digital Enterprise Research Institute (DERI), Fiontar at Dublin City University (DCU), and the National Library of Ireland (NLI).

The project aim was to create a Linked Data version of the place name data held by logainm.ie, which is the authoritative database of Irish place names, created by Fiontar in collaboration with the Placenames Branch of the Department of Arts, Heritage and the Gaeltacht.

The Linked Logainm project demonstrates the significance and potential applications of a Linked Data controlled vocabulary of Irish place names, available in a structured format, which allows its value to be fully exploited by collaborators, developers, the heritage community, information professionals and others; and ensures the dissemination of accurate, authorised versions of Irish place names internationally.

In addition, the demonstrator website, *Location LODer*, gives an accessible introduction to the potential applications of linked place name data, and the accompanying guidelines provide an overview of developments in linked data and cataloguing for heritage professionals interested in using the logainm.ie data in their own institutions.

Project participants

The **Digital Enterprise Research Institute** is a Centre for Science, Engineering and Technology established in 2003 with funding from Science Foundation Ireland at the National University of Ireland, Galway (NUI Galway). The mission of DERI is to exploit semantics for people, organisations and systems to enable collaboration and interoperability on a global scale. After more than nine years of operation, DERI has become an internationally leading institute in semantic web research, education and technology transfer, which directly contributes to the Irish government's plan to transform Ireland into a competitive knowledge economy. The technology developed by DERI is driving tens of thousands of websites (for example in the Content Management System, Drupal) and is installed on countless desktops (due to the fact that technology by DERI is part of any Linux installation). International Web standards have been initiated, shaped and led by DERI (for example, the data catalogue metadata formats VoiD and DCAT or World Wide Web Consortium (W3C) query language standards like SPARQL). This world-leading expertise has been created and developed in Ireland to propel the Irish ICT landscape forward.

The **Digital Repository of Ireland** is an interactive, trusted digital repository for social and cultural content held by Irish institutions. By providing a central Internet access point and interactive multimedia tools, DRI facilitates engagement with contemporary and historical data, allowing the public, students and scholars to research Ireland's cultural heritage and social life in ways never before possible. As a national digital infrastructure, DRI is working with a wide range of institutional stakeholders to link together and preserve Ireland's rich and varied humanities and social science data.

DRI was launched in 2011, when it received funding from the Irish government's Programme for Research in Third-Level Institutions cycle 5 for €5.2 million over four years. The Royal Irish Academy is the lead institution in the DRI consortium, which is also composed of the following partners: National University of Ireland, Maynooth; Trinity College Dublin; Dublin Institute of Technology; National University of Ireland, Galway; and the National College of Art and Design. The DRI Research Consortium is currently collaborating with a network of cultural, social, academic and industry partners, including the National Library of Ireland, the National Archives of Ireland and RTÉ.

Fiontar, at Dublin City University, is conducting cutting-edge research in the area of electronic resources for the Irish language, especially with regards to terminology, place names and Irish-language biographies; as seen on focal.ie, logainm.ie and ainm.ie. These research projects develop and use digital resources such as databases, Internet tools and data corpora, to provide terminological and place name material to researchers and to the global Irish-speaking community. These resources are developed in collaboration with national and international institutions with national and European funding. A team of researchers work on these projects in disciplines such as terminology (editing, development of specialised dictionaries, etc.) and technology (developing Internet tools, search methods, databases, mapping, etc.). Fiontar also develops Irish-language terminology for European Union (EU) institutions. The focal.ie database currently contains over 300,000 terms and c. 1 million searches are performed each month on the website. There are 100,000 place names as well as sound recordings and images in logainm.ie.

The mission of the **National Library of Ireland** is to collect, preserve, promote and make accessible the documentary and intellectual record of life of Ireland and to contribute to the provision of access to the larger universe of recorded knowledge. The NLI offers an exciting programme of exhibitions, events and learning opportunities for people of all ages and interests. It is driving a forward-looking programme of digitisation, digital preservation, and innovative access, visualisation and engagement tools for Ireland's rich cultural collections. For example, the NLI is one of the main contributors to VuFind, an open-source discovery interface, which is used by hundreds of libraries around the world to enhance access to research materials. The NLI also has considerable experience in

converting and enhancing metadata for cultural heritage items, including working with Linked Data resources like Freebase, VIAF and DBpedia.

The primary function of the **Placenames Branch** in the Department of Arts, Heritage and the Gaeltacht is to undertake research into the place names of Ireland to provide authoritative Irish-language versions of those place names for official and public use. The Irish versions determined by the Placenames Branch are given legal status by means of a place names order made by the minister of state at the Department for Arts, Heritage and the Gaeltacht, following public consultation and upon consideration of the advice of the Placenames Commission. Fourteen place names orders have been made to date since the enactment of the Official Languages Act, 2003.

Even though the branch's research is focused primarily on administrative names – the names of counties, baronies, civil parishes, townlands and electoral divisions – Irish-language versions have been established for other types of names also, such as towns and other centres of population, street names, rivers and various geographic features. Names within the various Gaeltacht (or Irish-speaking) areas have also been researched, both administrative place names and other types of names whose English versions are found on Ordnance Survey Ireland maps.

Introduction to the Linked Data concept

Linked Data refers to data published on the Web following a set of principles designed to promote linking between Web entities. An essential requirement to enable this linking is that each entity (for example a place name or personal name) is given a unique identifier, generally in the form of a Uniform Resource Identifier (URI). Having determined these URIs, Linked Data reuses other data models such as the Resource Description Framework (RDF)¹ to specify the links, and their type, between two URIs. Geographic Linked Data forms a substantial portion of the current Linked Data ecosystem.

The creation of authoritative URIs as subdomains of logainm.ie establishes logainm.ie as the recognised entity responsible for Irish place name Linked Data identifiers. The structured data published on the Web enables developers to reuse logainm.ie data to build applications, taking advantage of query languages, such as SPARQL, that allow the user to go beyond string matching when searching for place names. For example, by using SPARQL one can retrieve only entities of a specified type, with specific values for any property, or simply count the number of entities in a dataset.

¹ RDF is a data model developed by W3C for exchanging information on the Web. RDF makes statements, called 'triples', and they take the form <subject> <predicate> <object>, where subject is the entity or resource, object is another resource or value, and predicate is the relationship between them.

Project overview

- Logainm.ie is an online database containing over 100,000 Irish geographical names, including authoritative Irish-language versions for approximately 80,000 of these. The data is generated by the Placenames Branch of the Department of Arts, Heritage and the Gaeltacht, and the database was created and is being developed in collaboration with Fiontar, DCU.
- The logainm.ie data is intended as a resource for researchers, including educators, students and genealogists. As a bilingual authoritative list of place names, it is also used as the basis for cataloguing and key wording collections from heritage institutions, including museums, archives and libraries nationally and internationally. While the data had previously been made available by Fiontar on a request basis, the generation of a Linked Data version of the site allows immediate access to structured data dumps, which can be used by cataloguers, and computer scientists and application developers.
- As part of this project, the National Library of Ireland provided a use case for the newly linked place name data, with the enhancement of the catalogue for its Longfield map collection to allow greater searchability of the content.
- In order to further demonstrate the potential application of linked place name data, the Location LODer demonstrator website was created, using Linked Logainm place name URIs as its focal point, drawing in content from a range of sources including Europeana, DBpedia, the National Library of Ireland's Longfield map collection, and digitised archival content from logainm.ie.

Project methodology

Initially, Fiontar supplied an XML dump of the logainm.ie database to the DERI/DRI team, including all English and Irish place names, their type, and if available, their geolocation in Irish Grid Reference format. This data and its hierarchical structure formed the basis of the new RDF dataset.

Creating RDF data

The translation of the logainm.ie database into Linked Data was performed using XSPARQL, by developing a query that transforms the input XML into the target RDF. The RDF follows the NeoGeo² vocabulary, in which every place is designated as being of type 'Feature'. The URIs for identifying each place name from logainm.ie were generated based on the place name's unique identifier in the logainm.ie database and thus refer to the original record in that database. The target RDF representation also contains the geolocation of the place using the World Geodetic System coordinates, the reference coordinate system used by the Global Positioning System (GPS). These coordinates were translated from the Irish Grid Reference coordinates and used to aid the determination of links between the logainm.ie dataset and other Linked Data sources on the Web.

This process of determining links between place names in the different datasets was previously tested and applied in the RDF version of OpenStreetMap, LinkedGeoData, where they used similar approaches to determine the links to other datasets like DBpedia.

Linking the datasets

In order to link the newly created RDF dataset to other external datasets, the project team identified relevant sources of geographical Linked Data, and selected those which could be linked to logainm.ie: DBpedia,³ LinkedGeoData⁴ and Geonames.⁵ DBpedia consists of a partial export of data from Wikipedia into RDF and Linked Data. Similarly, data from OpenStreetMap is made available as RDF by the LinkedGeoData project. GeoNames is another geographical database accessible on the Web that also exposes its data as Linked Data.

To determine links between the logainm.ie RDF data and the other datasets, the Silk Link Discovery Framework was used.⁶ This framework compares entities from the different datasets according to a predefined set of rules and assigns a value to the similarity between entities. The entities with the highest similarity value are considered to be equivalent and this is recorded in the RDF data.

² <http://geovocab.org/doc/neogeo.html> (22 October 2013).

³ www.dbpedia.org (22 October 2013).

⁴ www.linkedgeo.org (22 October 2013).

⁵ <http://geonames.org> (22 October 2013).

⁶ <http://wifo5-03.informatik.uni-mannheim.de/bizer/silk> (22 October 2013).

Link evaluation

The rules used to determine the links between the datasets were based on the place name, its geographical coordinates (if available), and how it related to the hierarchy of other place names. For example, Blackrock in Dublin and Blackrock in Galway were differentiated within each dataset based on their coordinates and their counties. The process generated around 16,000 links between the datasets and based on the provided rules, the software assigned a confidence level to the links with a cut-off threshold of 95%. The generated links were then separated into two sets based on this cut-off value, those with a confidence level above 95% (assumed to be correct but still needed to be checked for incorrect links) and another set of links with a confidence level lower than 95% (still might have contained correct links).

In order to check whether the links between the datasets were created as expected, a process of manual link-checking was undertaken by the project team at Fiontar and DRI/NLI. This involved saving the suggested links in a .CSV file, opening each one in a web browser, checking if the suggested matching of places was correct or incorrect and making a note of whether they were linked as expected. By checking whether a selection of links that the software considered to be correct were indeed correct, and whether those that were considered incorrect were really incorrect, a set of 'guide links' was generated. This allowed the team to estimate the overall correctness of the entire set of links. The confidence level for the generated links currently stands at 97%.

To enhance the Linked Dataset further, the National Library of Ireland suggested that links could be generated to the Library of Congress FAST (Faceted Application of Subject Terminology) dataset, which is commonly used for library cataloguing. Approximately 1,000 links were generated between logainm.ie and FAST, and these were all checked manually to create a set of links with a 98.8% confidence level.

Following the link checking and the acceptance of the confidence levels of the generated links by Fiontar and the Placenames Branch, the linked RDF dataset was migrated to the logainm.ie server. Although the user interface of the logainm.ie website has not changed, the data can now be downloaded as RDF, JSON, RDF/XML, Turtle and N-Triples, and queried via a SPARQL interface at <http://data.logainm.ie/sparql/>

Updating the National Library of Ireland's catalogue

Once the URIs were created for the place names in logainm.ie, they could be appended to the National Library of Ireland's MARCXML catalogue records for its Longfield map collection. By matching the place names found in the records to the logainm.ie URIs, 1,570 records were updated automatically. Over 300 of the URIs added to the records were checked by DRI/NLI, and were found to match completely. Because this pilot set of links were 100% correct, the remaining links were assumed to be correct also.

The application of URIs to the MARC records allows the National Library of Ireland to exploit the Linked Data in a number of ways. By enhancing the catalogue's Web interface to capitalise on the presence of the Linked Data, the searchability of the collection can be augmented. For example, place names in the catalogue could be searched both in Irish and English, including any variants of the Irish language place name. The library could also choose to display data pulled in via the Linked Data sets, for example information taken from the DBpedia data to contextualise the historical maps.

Creating the demonstrator

The demonstrator website Location LODer was designed as a pilot application providing an accessible introduction to the applications of Linked Data for heritage institutions. The links generated between logainm.ie and the DBpedia, LinkedGeoData and Geonames datasets allow the additional place name data held in the external datasets to be pulled in and contextualised in a single interface. To enhance the site further, digitised objects were pulled in from Europeana, and digital collections added from the National Library of Ireland's Longfield map collection, the Placenames Branch's digitised archival records, and data and maps from the Royal Irish Academy's Irish Historical Towns Atlas' Dublin volumes.

The interface is based on an interactive map of Ireland, built on Google maps, which allows the user to explore places that have both logainm.ie data, and data from external datasets. As the example below shows, a pin icon marks locations on the map for which data is available, and the user can click through to view the range of content. The Irish and English language place name and its type comes from the Logainm data. There is an infobox from DBpedia, including a description of the place, which partially replicates the infobox present on Wikipedia's pages. Additionally, places on the map may have associated Longfield maps, digitised archival documents from the Placenames Branch or Europeana, or detailed street-level information from the Irish Historic Towns Atlas.



Figure 1 - Screenshot of the Location LODer interface

Guide to using Linked Data

In order to support the use of Linked Data for cataloguing in Irish institutions, an introduction to Linked Data for information professionals was prepared by DRI and the National Library of Ireland.

The document gives an accessible technical introduction to the concept of Linked open data, and a survey of its current application in the library, archives and museum domains. It discusses the advantages of using controlled vocabularies and ontologies such as the logainm.ie dataset for cataloguing, and the additional advantages of using a Linked Data vocabulary. It also gives an overview of how Linked Data can be incorporated into cataloguing and systems workflows, and the many ways in which Linked Data can be exploited using the Linked Logainm dataset to illustrate this potential.

Future work

- By providing guidelines and use cases demonstrating the value of Linked Data for heritage professionals, the project team hopes to encourage its widespread adoption in Ireland. The Digital Repository of Ireland platform will support Linked Data cataloguing, and its Raidió na Gaeltachta demonstrator project at National University of Ireland, Galway intends to enhance its catalogue records using the Linked Logainm dataset.
- The adoption of the Linked Logainm data as a controlled vocabulary for Irish place names will encourage consistency in the creation of geographic keywords for cataloguing, while disseminating the Irish and English versions approved by the Placenames Branch.
- From the collaboration with DERI, a world-leading Linked Data research institute, DRI can also avail of new techniques and software applications that aim to improve and enhance Linked Data, and specifically the logainm.ie data.