

The Creation of a Linked Data-based Application Service at the National Library of Korea

Wonhong Jang, Sungkyunkwan University
Sangeun Han, Sungkyunkwan University
Sam Oh, Sungkyunkwan University

Abstract

Since the advent of Linked Data (LD) in circa 2006, the participation of many institutions in publishing their data online has led to rapid growth of a Linked Open Data (LOD) cloud. In the library field, major libraries from countries around the world have become aware of the value of LOD and are making an effort to publish the quality metadata that they hold as LOD. The National Library of Korea (NLK) joined this movement in 2012. With the quantitative increase in LOD, it has become necessary to build LOD-based application services both to enable more institutions to participate in publishing LOD and to provide real benefits to users. This paper describes the LOD data model of the National Library of Korea, the core of its developing LOD-based application service and planned interface. The paper evaluates and discusses the anticipated effects of the new system for its target audience.

Keywords: National Library of Korea (NLK); Linked Data; Linked Open Data; LD Data Modeling; LD Application Services)

Citation: Jang, W., Han, S., Oh, S. (2015). The Creation of a Linked Data-based Application Service at the National Library of Korea. In *iConference 2015 Proceedings*.

Copyright: Copyright is held by the authors.

Acknowledgements: This project is funded by the National Library of Korea.

Contact: jangwonhong@gmail.com, silver86eun@gmail.com, samoh@skku.edu

1. Aim of Study

The publication and sharing of bibliographic data as Linked Open Data (LOD) were studied in detail as of 2011 by the W3C Library Linked Data Incubator Group (Baker et al., 2011). Institutions such as the national libraries of England, France, and Germany, and OCLC (US), have actively published Linked Data (LD) so that quality bibliographic information could be used in search engine indexing. Examples such as Europeana, the EU's long-term digital library project, and the DPLA (Digital Public Library of America), a collaborate effort of major US libraries and archives, are seen as successful endeavors to effectively publish and expose diverse information (including digital content in the possession of libraries and similar institutions) as well as connect it with other information (Haslhofer & Isaac, 2011).

In accordance with such efforts, the NLK began a project to publish its bibliographic metadata as LD in 2012. In 2014, NLK provided initial results through a SPARQL Endpoint so that anyone could access it. In order to attract more institutions to take part in publishing LD, enhance the quality of bibliographic data through interlinking, and on this basis to provide information services with which users will be satisfied, a responsive interface and application service are urgently needed in order to demonstrate the potential impact of NLK LOD. Therefore the aim of this study is to expand the interlinking of LOD published by NLK, to build an interface and application service which can meet the needs of users, and evaluate its effectiveness.

2. Current Situation of the NLK's LD Publication

The status of NLK's LD publication from 2012 to 2013 is shown in the table below. Targeting bibliographic data about 88 million triples were created, but after interlinking with outside LD further progress was limited to only partial data such as books and authors.

Category	Number of Records	Mapping property	Triples
Bibliography (Book)	3,961,034	-	87,301,188
Bibliography (Periodicals)	69,502	-	1,188,286
NLK to Overseas Libraries	219,360	owl:sameAs	219,360
NLK Bib to LCSH	849,512	dct:subject	2,079,454
NLK Subject to LCSH	10,331	skos:closeMatch	10,331
NLK Bib to National Library of	255,540	dct:subject	456,505

Japan Subject			
---------------	--	--	--

Table 1. Number of published NLK LD

Typical methods of publishing LD are: 1) after creating RDF file, to insert it into a Web page; 2) to apply mapping rules for the publication of existing RDBMS; and 3) if the old OpenAPI is available, making a Wrapper and publishing it (Bizer, Cyganiak & Heath, 2007). In the case of MARC data, it has been suggested that the conversion method of the Library of Congress's BIBFRAME be used (Kroeger, 2013). In the case of NLK, elements of KORMARC were analyzed, an ontology based on KORMARC was modeled, a mapping algorithm was created, and KORMARC data was thereby converted into LD and published. In 2014, works are in progress to build an application service based on this LD and reinforce interlinking to improve the quality of data.

3. Advanced Researches on developing LD-Based Application Services

Europeana, DPLA, and the BBC are prominent examples of LD-based application services. Europeana has converted existing metadata to fit an LD environment based on the EDM (Europeana Data Model). Through this Europeana semantic searches were made possible, seamlessly linking data from libraries and museums across Europe (Europeana, 2013). Data platform services on certain topics such as World War 2 are eliciting active user participation. DPLA has referenced Europeana's EDM and built a model of its own (Guthro, 2013). DPLA LD triggers users' interest with visualizations such as timelines, which promotes increased use.

The BBC published their data as LD in order to provide users with seamless access and use of BBC programming. Through this it became possible to identify certain persons, programs, and other things of interest using an RDF triple data. Information on the World Cup 2010 was provided as an LD platform, which BBC users could continue to explore within the search results on the BBC website (Georgi, 2009). 'ResearchSpace,' which through faceted search and semantic description provides an environment in which studies on history and culture can be properly managed, and 'Open Pharmacology' which provides a semantic research environment for the study of pharmacology, are also successful examples of LD application services (Simperl et al., 2013).

4. Reinforcing LD with Interlinking and NLK LD Data Modeling

One rule of LD publication calls for inclusion of 'links to other related things (using their URLs) when publishing data on the Web.' The effectiveness of LD can be maximized when data is linked between different data sources. This interlinking has not been properly implemented at NLK to date, so it is difficult for the user to access external data through NLK data. To make for a more enhanced service and interlinking, relationship links have been reinforced. Through this improvement, search results have been enhanced with additional relationship information. Furthermore, through identity links such as ISNI (International Standard Name Identifier) and DBpedia, identical entities have been connected across data sources. In order to effectively display LD combined from different institutions, an appropriate data model is essential. In order to meet such needs, Europeana and DPLA have proposed separate data models (Doerr et al., 2010; Matienzo & Rudersdorf, 2014). The diagram below Figure 1 is a DPLA metadata model, which has referenced the Europeana Data Model (EDM).

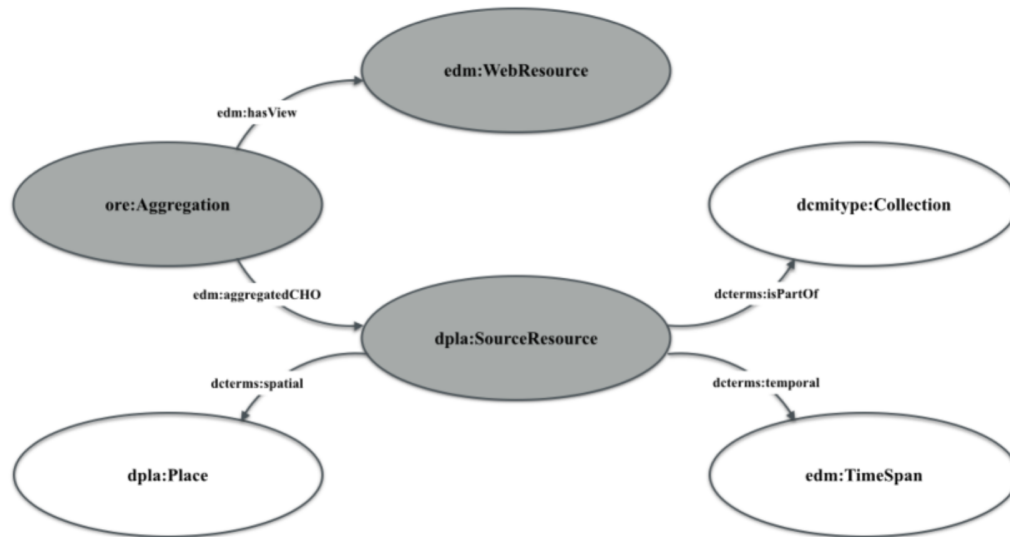


Figure 1. Model of DPLA's Key Classes and Properties

The Figure 1 above is a data model schematized through interlinking; based on detailed information about the author and bibliography, major LDs such as DBpedia, Freebase, and Europeana are connected.

NLK manages not only books, but also many types of resources. If NLK data model is compatible with Europeana and DPLA, linking data internally and externally will be made easier. However, the prototype system developed for this study was not able to fully implement those data models. In terms of the data modeling for this study, a new data model was constructed to increase the rate of inter-linking by incorporating well-established two data models. The Figure 2 below is a sample model showing how, based on the NLK LD model (which is an expansion of EDM), major LD data on authors and bibliography, DBpedia, Freebase, and Europeana can be connected.

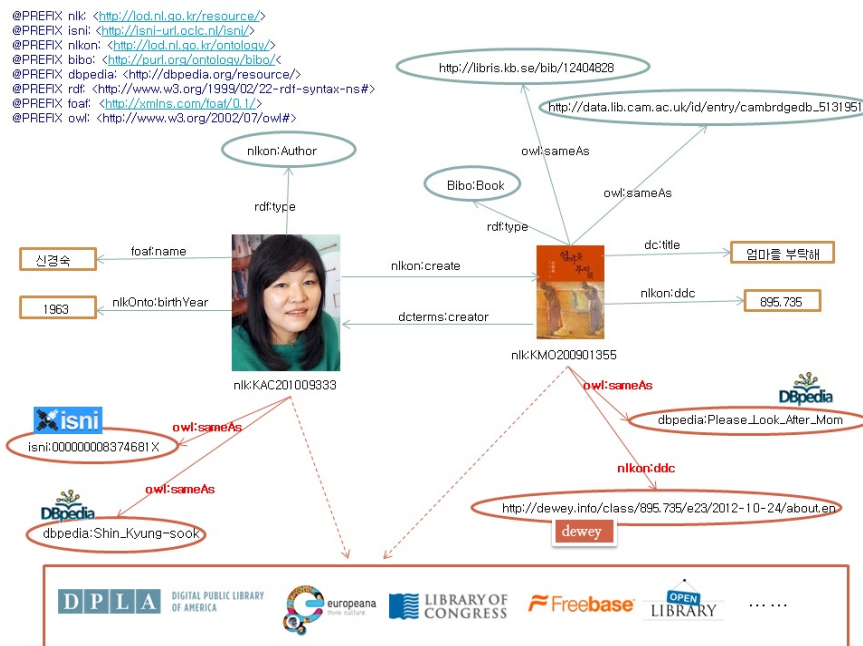


Figure 2. NLK LD Model

5. Developing LD-based application services at NLK

The key elements needed to develop an LD-based application service are largely divided into two areas, Server Side and Client Side. The key elements on the Server Side are a Triple Store that can effectively save large quantities of LD; a management module for the input and output of data saved in Triple Store; an LD search module; and a SPARQL Endpoint. Currently various publicly available frameworks can be used to develop LD-based application services, such as LMF (Linked Media Framework), Callimachus, Synth, and Apache Jena (Schaffert et al., 2012; Battle et al., 2012; de Souza Bomfim & Schwabe, 2011; Simperl et al., 2013). These frameworks usually include the aforementioned key elements for service creation and therefore can greatly reduce the time and effort needed to build application services.

After reviewing the options for framework development, installation and prototyping, we chose Linked Media Framework (LMF) as the basis for the NLK LD-based application service. LMF, a LD framework implemented based on Java script, includes all the elements needed for implementing application services, i.e. a Triple Store linkable with RDBMS, the search engine Apache Solr, and Apache Stanbol and Reasoner, which are needed for implementation of LD-based faceted searching (Kurz & Schaffert & Burger, 2011). The comparative strength of LMF lies in its organically combined search function and LD Cache. The search engine Solr, which is included in LMF, is noteworthy for its multilingual searches and faceted searches; the completeness and quality of its functions have been verified in various projects and studies. By defining simple search rules, LMF helps to easily and quickly create an index optimized for LD faceted searching. Another of its strengths is that with its Cache function it can quickly process SPARQL results from interlinked LOD repositories. Figure 3 shows the NLK LD-based application service model based on LMF.

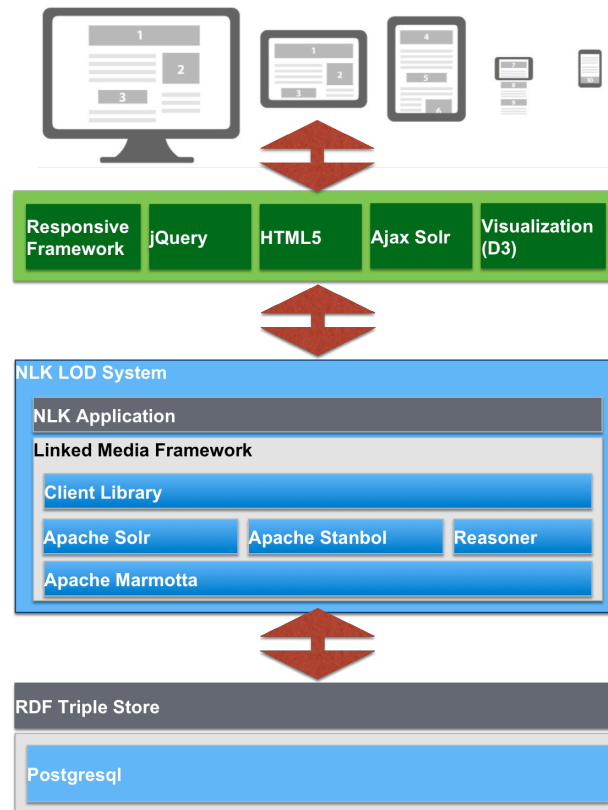


Figure 3. NLK LD-based Application Service Model

In order to enhance the input/output speed of data we plan to save LD in the PostgreSQL database, and through LMF to leverage existing links to import an enhanced form of LD. Based on search rules we create indexes for faceted searching. Using Ajax Solr, a representative Solr client, we provide a faceted search interface so that users can do searches through a Web browser; for this we also used and implemented specific categories of data and the client library provided by LMF. The most prominent characteristic of this application service is that we applied a responsive Web framework on the client

class, thereby allowing users to use this application service as an optimized UI on devices with a variety of resolutions—PCs, tablets, smartphones, etc. In order to enable users to explore LD content intuitively, we plan to use the framework D3 to visualize LD elements such as location, publication dates, and authors' birth date. Figure 4 shows a screenshot of the LMF management page on the NLK LD service server, with search settings and LD inhibition and modification functions.

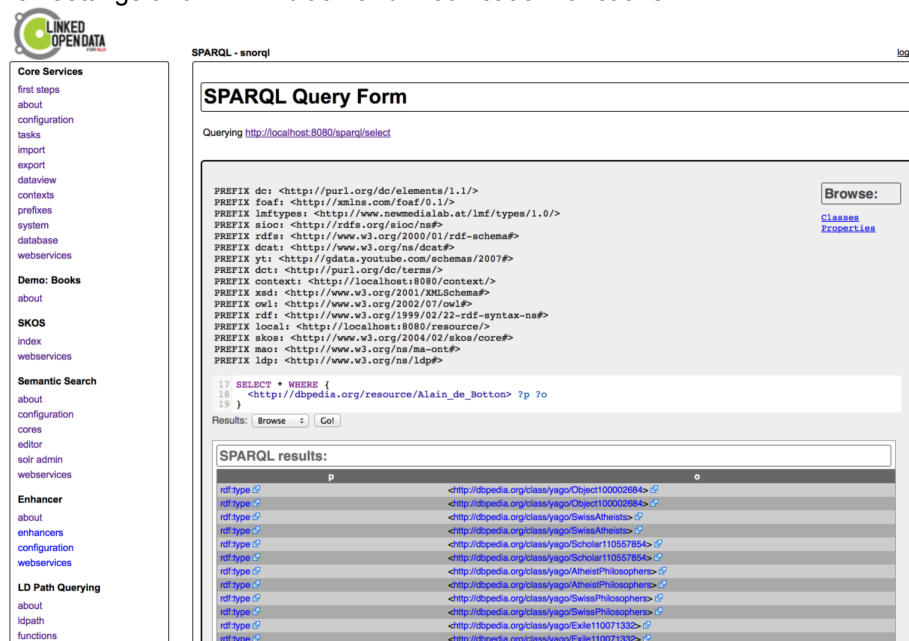


Figure 4. NLK LD-based Application Service Management Page

6. Evaluating the LD-based system and anticipated affects of application services

The implemented system will be evaluated two ways. To see what impacts the proposed system for the NLK managers, we will ask NLK system managers about the conveniences and inconveniences they experienced as they used the LD-based platform. To test the performance of the proposed system, given the diverse characteristics of users of the NLK LD-based system, we intend to perform an objective and subjective evaluation of the differences between the previous system and the LD-based system. For the second evaluation, we will employ the evaluation criteria developed by DeLone and McLean(1992). The table 2 and 3 describe the major criteria for evaluation to be performed.

Point of Evaluation	Description	Method of Recording	Unit of Recording
Retrieval Time	How much time was spent before arriving at the final retrieval results?	Recorded Timing	Individual queries
Number of Retrievals	How many keywords were inputted before arriving at the final retrieval results?	Recorded number of retrievals	Individual queries
Page Views	How many pages were viewed before arriving at the final retrieval results?	Recorded page views	Individual queries

Table 2. Objective Evaluation Items of the Information System

Point of Evaluation	Description	Survey Content	Unit of Recording
Accuracy	Was the user able to obtain the desired results from the	I was able to obtain accurate answers to the	Query type

	system?	given tasks by using this system.	
	Were the results for each keyword accurate?	I was able to obtain accurate results to the keywords inputted into the search engine.	Query type
Usability	Does the system provide a variety of different search paths?	This system provides a variety of different search paths.	Query type
	Is the system easy to navigate?	This system is easy to navigate.	Query type
Usefulness	Was the system useful in obtaining the final results?	I was able to obtain a quick answer to the given tasks by using this system.	Query type
	Did the system enable the user to obtain additional information?	I was able to obtain additional information while searching for answers to the given tasks.	Query type
User Satisfaction	Is the user satisfied with the results of the retrieval?	I am satisfied with the retrieval results obtained by using this system.	Individual systems
	Is the user satisfied with the system in general?	I am satisfied with this system overall.	Individual systems

Table 3. Subjective Evaluation Items of the Information System

To fully test what is proposed in this study, query templates will be provided to the subjects. The templates are composed of three levels of difficulty. The following illustrates the templates with examples.

Query Type	Query Template	Example Query
Simple Query	Search for the year of birth and year of death of (Person).	Search for the year of birth and year of death of Charles Darwin.
	Search for information on the members of (Person)'s family.	Search for information on the members of Charles Darwin's family.
Composite Query	Search for the title and publication year of the book written by Author) in (Year).	Search for the title and publication year of the book written by George Orwell in 1948.
	Search for the company which published (Book Title) in (Year). Then search for a list of the company's representative publications.	Search for the company which published Bill Clinton's autobiography My Life in 2004. Then search for a list of the company's representative publications.

Table 4. Query Types

The implementation of this application service is expected to have the following effects:

- To provide information about relationships: The strength of LD is that related information can be closely linked on the basis of shared identifiers. Through queries to external LD repositories, the service can provide users not only with local data but with information in external LD repositories.

- To reduce service implementation time and cost: The strength of LD-based application service implementation is the availability of mature tools as open source. That the use of such tools leads to reduction of implementation time and cost has been proven itself in this project.
- To provide an LD-based service implementation guide – a report that describes in detail the whole process of implementation that can be used as a guide for the implementation of other LD-based application services that use open source technology.
- To provide a platform service basis: The strength of LD is that flexible expansion of data and services are possible. Based on the results of the development of this application service, a convenient basis for creation of new LD and additional implementation of Web applications will be provided, and in the long run this can be used as a basis for the NLK LD platform service.

7. Conclusion and Further Plans

This paper has introduced the plans of the National Library of Korea for the implementation of LD-based application services, plans for system evaluation, and anticipated effects. In order to enhance the quality of Linked Data and its interlinking with external LD sources we have applied faceted searching and visualization technology to implement an application service based on Linked Media Framework (LMF) and other proven open source technologies.

As of August 2014, projects to reinforce LD and the management of LD are in progress. Implementation of Web applications that use responsive Web technology and implementation of visualization are scheduled to continue through the end of 2014. After the services go online, Web analytics tools such as Google Analytics will be used to collect and analyze quantitative data about users. Through surveys and interviews of user groups, qualitative data of users' satisfaction level will be collected to evaluate the impact of LD-based application services.

8. References

- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., & Zeng, M. (2011). Library Linked Data Incubator Group Final Report. *October, available at: www.w3.org/2005/Incubator/ld/XGR-ld-20111025/(accessed August 25, 2014).*
- Battle, S., Wood, D., Leigh, J., & Ruth, L. (2012). The Callimachus Project: RDFa as a Web Template Language. In *COLD*.
- Bizer, C., Cyganiak, R., & Heath, T. (2007). How to publish linked data on the web. July, available at: www4.wiiss.fu-berlin.de/bizer/pub/linkdatatutorial/(accessed August 25, 2014).
- de Souza Bomfim, M. H., & Schwabe, D. (2011). Design and implementation of linked data applications using SHDM and synth. In *Web Engineering* (pp. 121-136). Springer Berlin Heidelberg.
- Delone, W. H. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.
- Guthro, C. (2013). Digital Public Library of America. *Maine Policy Review*, 22(1), 126-129.
- Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & van de Sompel, H. (2010, August). The europeana data model (edm). In *World Library and Information Congress: 76th IFLA general conference and assembly* (pp. 10-15).
- Haslhofer, B., & Isaac, A. (2011, September). data. europeana. eu: The europeana linked open data pilot. In *International Conference on Dublin Core and Metadata Applications* (pp. 94-104).
- Kroeger, A. (2013). The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC Future. *Cataloging & Classification Quarterly*, 51(8), 873-890.

Kurz, T., Schaffert, S., & Bürger, T. (2011, September). LMF-A Framework for Linked Media. In *Proceedings of the 2011 Workshop on Multimedia on the Web* (pp. 16-20). IEEE Computer Society.

Matienzo, M. A., & Rudersdorf, A. (2014). The Digital Public Library of America Ingestion Ecosystem: Lessons Learned After One Year of Large-Scale Collaborative Metadata Aggregation. *arXiv preprint arXiv:1408.1713*.

Schaffert, S., Bauer, C., Kurz, T., Dorschel, F., Glachs, D., & Fernandez, M. (2012, September). The linked media framework: Integrating and interlinking enterprise media content and data. In *Proceedings of the 8th International Conference on Semantic Systems* (pp. 25-32). ACM.

Simperl, E., Norton, B., Acosta, M., Maleshkova, M., Domingue, J., Mikroyannidis, A., ... & Power, R. (2013). Using Linked Data Effectively.

9. Table of Figures

Figure 1. Model of DPLA's Key Classes and Properties	3
Figure 2. NLK LD Model	3
Figure 3. NLK LD-based Application Service Model.....	4
Figure 4. NLK LD-based Application Service Management Page	5

10. Table of Tables

Table 1. Number of published NLK LD	2
Table 2. Objective Evaluation Items of the Information System	5
Table 3. Subjective Evaluation Items of the Information System.....	6
Table 4. Query Types	6