

Producing and Consuming Linked Open Data on Art with a Local Community

Fuyuko Matsumura¹, Iwao Kobayashi², Fumihiro Kato³, Tetsuro Kamura^{4,5},
Ikki Ohmukai^{1,4}, and Hideaki Takeda^{1,4}

¹ National Institute of Informatics,
fuyuko@nii.ac.jp, iwao@scholex.com, takeda@nii.ac.jp

² Open Community Data Initiative
iwao@scholex.com

³ Research Organization of Information and Systems
fumi@nii.ac.jp

⁴ The Graduate University for Advanced Studies

⁵ Tokyo University of the Arts
kamura.tetsuro@noc.geidai.ac.jp

Abstract. The importance of Linked Open Data (LOD) has not yet been recognized by ordinary people. As the effort to involve them into LOD activity, we have developed *Yokohama Art Spot*, a mash-up application on local museum information to demonstrate the value of information sharing and reuse where data from different communities are used together. Thanks to the flexibility of LOD, it provides useful views for local museum information and uses three different data sources, i.e., LODAC Museum which is a museum collection dataset, Yokohama Art LOD which is an event dataset maintained by a local community, and PinQA which is a Q&A dataset maintained by a company. In particular, Yokohama Art LOD was the result of communication among researchers, local volunteers, and people of the local government. Yokohama Art Spot works as a good example of how such efforts by local people can be rewarded by flexible use of the provided data.

1 Introduction

Open government is a move to publish government information to enhance government transparency and encourage citizens to participate in politics. Several western countries and cities such as the United States¹, the United Kingdom², Paris³, and Berlin⁴ have websites to provide data to citizens and some of them have already started to publish data in the *Linked Data*[1] format. Publishing Linked Data accelerates the move toward *open data* and enables data sharing, not only within the same field, but also with other fields. Indeed, in the United Kingdom, citizens and companies can develop and provide web application or smartphone applications using Linked Data provided by the government.

¹ <http://data.gov/>

² <http://data.gov.uk/>

³ <http://opendata.paris.fr/>

⁴ <http://daten.berlin.de/>

In Japan, the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry lead in the effort toward open government and publish their data in various formats. However, most of these are published in PDF or Excel file formats, and it is difficult for computers to automatically recognize and extract a required data from such files in unstructured formats.

Under these circumstances, dissemination of the idea of Linked Open Data (LOD) is important to the future of the Internet but it has not yet been recognized by ordinary people. The Linked Open Data for ACademia (LODAC) project⁵ has started to publish LOD of museum, biodiversity, and geographic information aiming to advance sharing and reuse of academic data in Japan. In particular, the integrated museum collection data is called *LODAC Museum*[2].

Such actions for open data are becoming increasingly common in local communities that publish local information with social tools. In the city of Yokohama, which is the most populous city in Japan, there are many portal sites that distribute information about that area, however, a lot of overlaps occurs between data used in each site, e.g., profiles of the same concert hall or the same restaurant. Several citizens and organizations in Yokohama realized the importance of open data for the efficient use of data on local information about Yokohama. They established *Yokohama LOD project*⁶ for individual persons or groups to accelerate information publishing by the sharing and reuse of local information in the LOD format. As a first step, the project has started to publish LOD of information on art events as *Yokohama Art LOD* and generated typed links to LODAC Museum based on the discussion between researchers in LODAC projects, local volunteer developers, and people of the local government.

In order to demonstrate the value of information sharing and reuse in the LOD format, *Yokohama Art Spot*⁷, a web application that provides information on art in Yokohama, was developed. It uses PinQA, which is a local Q&A dataset in addition to LODAC Museum and Yokohama Art LOD. This paper thus introduces the best practice of collaboration with a local community toward the acceleration of information distribution in the LOD format.

2 Linked Open Data (LOD)

Linked Open Data (LOD) represents datasets that give a Uniform Resource Identifier (URI) to each thing as a global ID on the web and express relations between those as links. LOD datasets connect to each other and construct a global data space on the web called *web of data*[3], according to the following principles: 1) Use URIs to identify things, 2) Use HTTP URIs so that these things can be referred to and looked up (*dereferenced*) by people and user agents, 3) Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML, and 4) Include links to other, related URIs in the exposed data to improve discovery of other related information on the web.

⁵ <http://lod.ac/>

⁶ <http://ocdi.jp/>

⁷ <http://lod.ac/apps/yas>

The advantages of LOD can be divided into the following three points: 1) encourage open data, 2) realize data sharing within the same field, and 3) realize data sharing with other fields. On the web, open data is published in various data formats, such as HTML, PDF, and Excel. People can easily read data expressed in these formats; however, it is difficult for programs to automatically recognize and extract a required data from such files, whereas, each thing is given a URI to be identified as a resource in LOD. Compared to data such as HTML documents, this makes it easier to access and reuse open data when the data is published as LOD. Moreover, it is also effective for data sharing within the same field or with other fields to identify unique items with URIs when several sites have data indicating the same thing in common.

3 LODAC Museum: Integrating Museum Collection Data in Japan

3.1 Overview of LODAC Museum

LODAC Museum is an LOD of museum information constructed by the LODAC project at the Transdisciplinary Research Integration Center in the Research Organization of Information and Systems. In Japan, museums publish their collection information on their own websites, except for some web services that integrate collection information, such as *Cultural Heritage Online*⁸. This means that currently it is impossible, for example, to extract a list of museums that own the art works of *Taikan Yokoyama*, a famous Japanese-style painter.

The LODAC project is thus attempting to build an LOD of museum collection information utilizing the advantages of LOD that advance data sharing within the same field, as mentioned in section 2. Essentially, it is desired that each museum publish their collection data in the LOD format and coordinate them with semantic links. However, currently, the LODAC project obtains collection data from various websites and publishes such obtained data as a single LOD.

LODAC Museum consists of over 40 million triples extracted from 114 museums and research institutes in Japan. This information is integrated according to the *Database of Japan Arts Thesaurus* which is a reliable thesaurus of art in Japan[4]. The *database of government-designated cultural property*⁹, Cultural Heritage Online, *DBpedia Japanese Lite* and *DBpedia Japanese*¹⁰ are also utilized to populate LODAC Museum.

3.2 Expression and implementation of data

Fig. 1 shows an RDF triple of a part of LODAC Museum. The main resource types of LODAC Museum are works, creators, and institutions. Each resource includes links to other resources, and all the data are described as graphs consisting of resources and links. For instance, the fact “The artist who painted *Spring*

⁸ <http://bunka.nii.ac.jp/>

⁹ <http://kunishitei.bunka.go.jp/bsys/>

¹⁰ <http://ja.dbpedia.org>

Sea is Taikan Yokoyama” is described as a triple like $\langle \text{http://lod.ac/id/20428} \rangle$ $\langle \text{http://purl.org/dc/terms/creator} \rangle$ $\langle \text{http://lod.ac/id/1767} \rangle$ as shown in Fig. 1.

To integrate information of a certain thing distributed to several data sources, a resource representing the thing is given a URI as an *ID* resource and includes links to *reference* resources consisting of information about the thing included in each data source. For example, if information on Taikan Yokoyama can be obtained from museums A and B, an ID resource would be created to express the identity of Taikan, and two reference resources are generated to describe information obtained from museums A and B, respectively.

Furthermore, LODAC Museum uses the OWLIM-SE¹¹ as our RDF store and provides a SPARQL endpoint and a browsing interface with keyword search.

4 Yokohama Art LOD: Providing Information on Art Events in Yokohama

4.1 The circumstances on publication of local information in Yokohama

Nowadays, there are more and more events in local communities actively planned and managed by citizens. At the same time, online tools such as Twitter, Facebook, and blogs that enable ordinary people to publish their news or opinions have become popular. In the city of Yokohama, there are many people or groups who provide local information on Yokohama using such online tools. However, it is difficult to make people reach the information published by each citizen or each group because of too many information floods. Thus, the local government and several organizations provide several portal sites to aggregate and deliver such event information. For instance, the Yokohama Arts Foundation distributes event information and news on art in Yokohama at their website *Yokohama Art Navi*¹². On the other hand, the Yokohama Convention & Visitors Bureau lists local sightseeing spots, events, and restaurants on the website of the *Yokohama visitor's guide*¹³. *Hamaspo.com* provides local sports information and *Ototokebin* delivers event information in each ward of Yokohama.

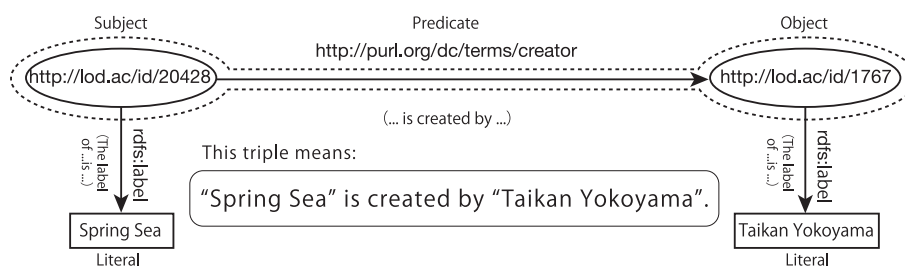


Fig. 1. A triple that expresses Taikan Yokoyama.

¹¹ <http://www.ontotext.com/owlim>

¹² <http://yan.yafjp.org/>

¹³ <http://www.welcome.city.yokohama.jp/eng/travel/>

4.2 Yokohama LOD Project and information on art events in the LOD format

Nevertheless, sometimes overlaps of information, i.e., address, genre, and opening hours of institutions, occur between such portal sites because these are common in the same area, Yokohama. Thus, some local volunteers came up with an idea that open data realizes efficient utilization of local information. Then, they have established the Yokohama LOD project to integrate those data in the LOD format to encourage publishing, sharing, and reusing local information in Yokohama. It would enable the reuse of data created by others according to the web standard and the reduction of cost for such data creations.

As a necessary first step, Yokohama Arts Foundation has allowed them to use data of all art events provided at Yokohama Art Navi website, and they have started to construct Yokohama Art LOD from those data. Researchers at LODAC projects have also discussed how to design their data and how to choose vocabularies with them. The number of triples of Yokohama Art LOD is 627,257 at present.

Currently, Yokohama Arts Foundation aggregates event information from each museum or each gallery and manually inputs these into the appropriate format for publication by the staff. The event information is written in text format; however, if the information of events, institutions, artists, and nearest station is expressed as resources and the relation between them are written as semantic links, it becomes possible to complement the missing information of a certain resource with other accumulated information via links.

In fact, event resources of Yokohama Art LOD have semantic links to the resources of LODAC Museum that describe the institutions which are the venues for those events, and institutions resources have also links to the resources of DBpedia Japanese which indicate same venues. Identification of those venues in two different datasets were processed based on text matching. It would be capable to obtain more detailed information on art via those links and demonstrate the value of the information sharing and reuse.

Providing information in the LOD format is suitable for the publication of local information, which contains groups or communities of various scales, because sharing and linking resources realize broader information distribution in local communities without making the effort to collect all the data separately.

In the United States and the United Kingdom, the government and the public administration lead the effort on open data[5][6], whereas in Yokohama, their work should be noted because their LOD activities for open data is being attempted from the grass-root level.

5 Yokohama Art Spot

This section introduces Yokohama Art Spot, a web application developed as a result of the collaboration between researchers, a local community and a company to make ordinary people notice the value of sharing and reuse of the data provided by others. It uses following three different datasets as data sources; LODAC Museum, Yokohama Art LOD and PinQA.

PinQA is a social geographical Q&A service provided by NTT Resonant, Inc. that obtains and shares users Q&As corresponding to the specified geographical regions in a map. They have also started to publish a small LOD dataset consisting of 777 triples generated from aggregated Q&A data, but it has not been linked to LODAC Museum and Yokohama Art LOD yet. It would help users to imagine their visit to the event, for instance, they can plan to take a rest at a coffee shop near the venue according to the past question about restaurants around there made by other users. Thus, typed links only exist between the resources of institutions in LODAC Museum and the resources of events in Yokohama Art LOD in this system. Links represented by dotted line are planned to connect in the near future.

Therefore, Yokohama Art Spot is provided conventional and novel local information on art in Yokohama based on a museum collection dataset, an art events dataset, and a local Q&A dataset.

5.1 Overview of the developed system

Fig. 2 shows an overview of this application. First, the people who will use this system are divided into two types: a user who visits the application on the web and a staff who manually aggregates and maintains the data. A user uses this application to obtain desired information and a staff manages updates and addition of LOD datasets used in this system.

The interface is developed with HTML and JavaScript. If a user takes an action, such as opening a new page that describes details about an institution, this is queried to the SPARQL endpoints of each LOD. The application then presents new information to the user using the data returned from the SPARQL endpoints.

5.2 Main features of the developed system

Yokohama Art Spot consists of 1) map view, 2) institution view, and 3) administration view. Both 1) and 2) provide information on art in Yokohama by embedding data obtained with SPARQL into templates of web pages.

Map view When a user accesses the application, a map view is presented, as shown in Fig. 3. This view gives users information on the kind of art they can experience by displaying institutions. The locations of the current day's events and Q&As related to the displayed area are also presented on the map of Yokohama. In this map view, the name, address, and geographical coordinates of institutions to be presented are extracted from LODAC Museum with SPARQL using the latitudes and longitudes of the displayed area. As for the event information, the application obtains data for the events that would be held at those institutions included in the displayed area from Yokohama Art LOD using SPARQL.

Institution view The institution view appears when a user clicks on the marker of the institution on the map view as shown in Fig. 4. This page shows basic information on the institution, event information and collection information. To

display basic information on the institution, the name, access options, hours of operation, and the geographical coordinates are extracted from LODAC Museum with SPARQL according to the URI of the institution. Moreover, events that have links indicating a venue for the institution are extracted by SPARQL, and those event data are displayed in a timeline interface as shown in Fig. 4. Collection data are also acquired from LODAC Museum by tracing art works linked to the institution with SPARQL.

Administration view Event data of Yokohama Art LOD are frequently updated and added, whereas information on collections and institutions is fairly stable. Currently, the same data is published in the art calendar on the Yokohama Art Navi and is manually updated by the staff of Yokohama Arts Foundation. However, it is difficult for staff who are not familiar with RDF to input data in the LOD format. This application, however, creates forms to make it easier to

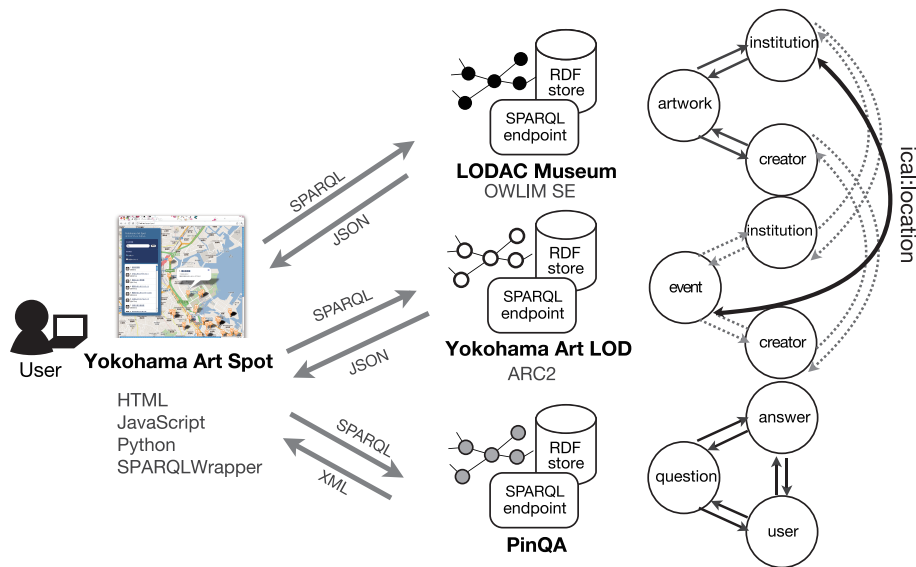


Fig. 2. The system overview of Yokohama Art Spot.

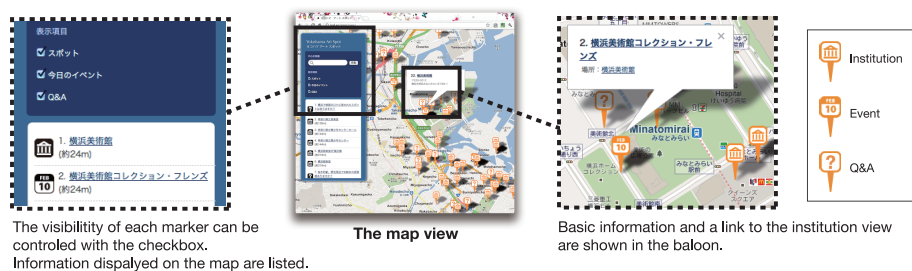
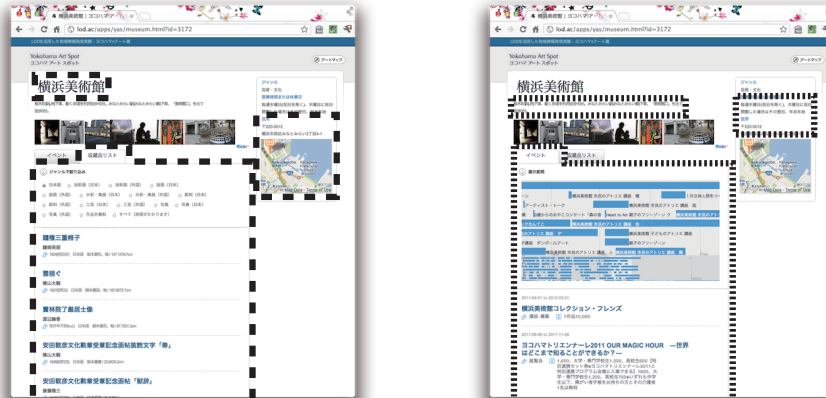


Fig. 3. The map view of Yokohama Art Spot



is obtained from LODAC Museum.
 (name, address, latitude, longitude, collection)

is obtained from Yokohama Art LOD.
 (genre, opening hours and event information)

Fig. 4. The institution view of the Yokohama Museum of Art.

properties (corresponding to appropriate vocabularies)

Fig. 5. The input form for institution information generated by Semantic MediaWiki.

input event or artist information based on *Semantic MediaWiki*¹⁴[7]. The staff would thus be able to generate RDF of Yokohama Art LOD by inputting required data into the form as illustrated in Fig. 5. This would enable addition or correction of data and realize data circulation of LOD consisting of creation, sharing, aggregation and usage.

6 Discussion

Yokohama Art Spot was developed by using three different LOD datasets aiming to make ordinary people notice the importance of sharing and reuse of information, and involve them into the LOD activities.

First, the effects of interdisciplinary use of three datasets are discussed. The developed system provides a couple of views such as map view and institution

¹⁴ <http://semantic-mediawiki.org/>

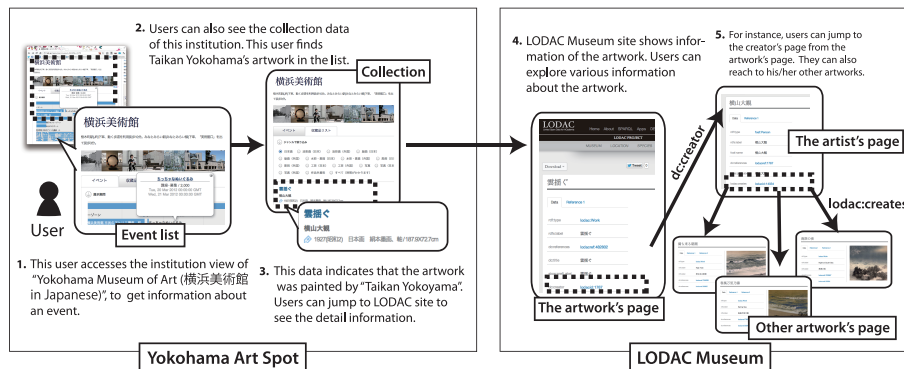


Fig. 6. An example of a scenario that guides users to discover new knowledge.

view even though the link connects only an institution data and an event data in LODAC Museum and Yokohama Art LOD, respectively. As shown in Fig. 6, for example, when users access the institution view of a specific institution to view the events to be held there, they may have a chance to see the collection data of that institution. If they find their favorite or famous paintings, they can explore the information through links to LODAC Museum. It is also possible to reach the same artist's paintings owned by other institutions. Therefore, more extensive information distribution would be realized by linking three different types of LOD datasets such as collection data containing ancient artworks and the data of current art events.

Second, the benefits of the compliance with the web standard are discussed. Even if data from other applications were available via an API, some modifications would be required to use such data because each API is different. Thus, when an application uses data from several data sources, this costs more compared to using a single data source. In contrast, in LOD, data are described using RDF, the standard data model, and they can be queried with SPARQL, the standard query language. Even if the number of LOD used in the application increases, any modifications to the application would be small. Because of this, the event data in the LOD format are easily presented on the institution view and it also acts as a reference of the institution when it has no webpage by simply providing accumulated data in the LOD format. In addition, developers of PinQA have also released their data in the LOD format and the map view presents word of mouth from PinQA on the displayed area without any major modification.

Finally, Yokohama Art Spot has been awarded the prize at Linked Open Data Challenge Japan 2011 for the contribution on the collaborative use of LOD datasets. In response, Yokohama LOD project has been interviewed as a leading-edge example of open data in the LOD format by the people from the government who work for open data. There is no doubt that the developed system acts an important role to get people to notice the importance of information sharing and they would lead the open data with other ordinary people in Japan.

7 Related Work

Several related projects have attempted to provide users the chance to access cultural heritage information.

The Google Art Project¹⁵ is an ambitious project that allows viewing of artworks with a brush stroke-level detail for greater appreciation as well as a virtual tour of 151 museums across 40 countries. They prepared this application to provide an environment to enjoy the fine art in famous museums around the world. The developed application, Yokohama Art Spot, aims to introduce a wide variety of modern art that are still being created and exhibited in the local area in addition to the fine art.

The following projects structured their data in the LOD format and enable a multifaceted search based on the properties of resources. Europeana[8] aggregates collection data from museums in Europe and structure the data based on the Europeana Data Model[9]. They organized some hackathons to encourage the use of Europeana via the Europeana API, and some applications have been developed. Although these applications use Europeana only via API, possible relations between Europeana and other LOD datasets have not been defined and utilized yet.

CultureSampo[10] provides Finland's national cultural heritage content from various perspectives generated by integrated multiple metadata schemas and cross-domain content. Yokohama Art Spot is similar to this system because it affords diverse semantic visualizations based on metadata: map view, timeline view, and collection view[11][12]. The developed system also utilizes several types of content, such as collection data and events, just as CultureSampo does. The MultimediaN E-Culture project[13] also merged partial collection data of the Rijksmuseum Amsterdam and the Rijksmuseum voor Volkenkunde into a single LOD dataset. Yokohama Art Spot is different from these projects because they manually integrate data aggregated from several data sources into a single dataset, whereas the developed system separately uses three independent datasets.

The CHIP Project[14] and the British Museum¹⁶ generate their own LOD datasets from collection data of the Rijksmuseum Amsterdam and the British Museum, respectively. The former tries to enrich a user's experience with a personalized interactive tour guide using recommendations based on user models[15][16]. These projects, however, focus only on data and users inside each museum, in contrast to the developed system.

The novelty of our research activities is the collaborative use of three different types of LOD: LODAC Museum, Yokohama Art LOD, and PinQA based on the actual collaboration between researchers, the local volunteers, and the people from the local government. We also succeeded to involve ordinary people to open data or the LOD activities by the demonstration of the benefit to share data and reuse them each other.

¹⁵ <http://www.googleartproject.com/>

¹⁶ <http://www.britishmuseum.org/>

8 Conclusion

In this paper, Yokohama Art Spot, a web application for the art guide in Yokohama, has been developed and it uses three different types of LOD datasets, i.e., LODAC Museum which is a collection dataset crawled from museum web sites, Yokohama Art LOD which is an art event dataset maintained by a local community, and PinQA which is a local Q&A dataset maintained by a company based on the collaboration between researchers, local volunteers, and the people of the local government.

LOD datasets such as collection data containing ancient artworks and the data of current art events or Q&As complement each other, and this is expected to enable broader information distribution. The addition of a new data source has become easy by the standard format and technologies such as RDF and SPARQL. As a result, Yokohama Art Spot works as a good example of how the efforts by local volunteers can be rewarded by flexible use of the LOD dataset provided by them.

Currently, it is required to input an institution's LODAC Museum URI to link with the event resource of Yokohama Art LOD, and thus the usability of updates will be improved by suggesting candidates according to the keywords input for the actual operation. In addition, ATR Creative Co., Ltd. has added new functionality to display Points Of Interest (POIs) included in LODAC Museum on the map on an iPhone application. Thus, our project plans to develop a smartphone application that will allow a user to search the contents close to the current position and annotate his/her own personal perspective to resources. Finally, it is desired to advance the use and the creation of museum collection LOD datasets and local information LOD datasets for future work by developing a web application package like the developed system to apply to other local communities in Japan.

Acknowledgements

We would like to thank Yokohama Arts Foundation and Hiroyuki Sato from NTT Resonant, Inc. for providing their datasets on art events and Q&As, respectively. We also gratefully appreciate the contribution of Toru Takahashi and Hiroshi Ueda from ATR Creative Co., Ltd. for the discussion on the developed application.

References

1. Bizer, C., Heath, T. and Berners-Lee, T.: Linked Data - The Story So Far, International Journal on Semantic Web and Information Systems (IJSWIS), Vol.5, No.3, pp.1-22 (2009)
2. Kamura, T., Takeda, H., Ohmukai, I., Kato, F., Takahashi, T., and Ueda, H.: Study Support and Integration of Cultural Information Resources with Linked Data, in Proc. of 2nd International Conference on Culture and Computing, pp.177-178 (2011)

3. Berners-Lee, T. and Fischetti, M.: Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor, HarperOne (1999)
4. Fukuda, H. and Omuka, T.: Constructing Web-based Art Thesaurus, Focusing Around Data Conversion (*in Japanese*), The Bulletin of Japan Art Documentation Society, No.14, pp.56–66 (2007)
5. Coglianese, C.: The Transparency President? The Obama Administration and Open Government, *Governance*, Vol.22, No.4, pp.529–544, 2009.
6. Shadbolt, N., O'Hara, K., Berners-lee, T., Gibbins, N., Glaser, H., Hall, W. and Schraefel, M. C.: Linked Open Government Data: Lessons from Data.gov.uk, *Intelligent Systems, IEEE*, Vol.27, No.3, pp.16–24, 2012.
7. Krötzsch, M., Vrandečić, D. and Völkel, M.: Semantic MediaWiki, *Lecture Notes in Computer Science*, 4273/2006, pp.935–942 (2006)
8. Haslhofer, B. and Isaac, A.: data.europeana.eu - The Europeana Linked Open Data Pilot, *Proceedings of the International Conference on Dublin Core and Metadata Applications(DC2011)*, pp.21–23 (2011)
9. Hennicke, S., Olensky, M., de Boer, V., Isaac, A. and Wielemaker, J.: A data model for cross-domain data representation: The Europeana Data Model in the case of archival and museum data, In *Proceedings of the 12th International Symposium on Information Science (ISI 2011)* (2011)
10. Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., and Kuittinen, H., et al.: CultureSampo:A National Publication System of Cultural Heritage on the Semantic Web 2 . 0 1 A Basis for Cultural Heritage on the Semantic Web, *THE SEMANTIC WEB: RESEARCH AND APPLICATIONS, Lecture Notes in Computer Science*, Vol.5554/2009, pp.851–856 (2009)
11. Kauppinen, T., Paakkari, P., Mäkelä, E., Kuittinen, H., Väättäinen, J., Hyvönen, E., Lytras, A. M., et al.: Geospatio-temporal Semantic Web for Cultural Heritage. *Digital Culture and E-Tourism: Technologies, Applications and Management Approaches*, Vol. 2010, pp.48–64 (2010)
12. Hyvönen, E., Lindquist, T., Törnroos, J., and Mäkelä, E.: History on the Semantic Web as Linked Data - An Event Gazetteer and Timeline for the World War I, *CIDOC2012* (2012)
13. Schreiber, G., Amin, A., van Assem, M., de Boer, Viktor et al: MultimediaN E-Culture Demonstrator, In *Proceedings of the 5th International Semantic Web Conference (ISWC2006)*, pp.951–958 (2006)
14. Roes, I., Stash, N., Wang, Y. and Aroyo, L.: A Personalized Walk through the Museum : The CHIP Interactive Tour Guide, In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA 2009)*, pp.3317–3322 (2009)
15. Wang, Y., Aroyo, L., Stash, N. and Rutledge, L.: Interactive User Modeling for Personalized Access to Museum Collections : The Rijksmuseum Case Study, In *Proceedings of the 11th international conference on User Modeling (UM 2007)*, LNAI 4511, pp.385–389 (2007)
16. Wang, Y., Wang, S., Stash, N., Aroyo, L. and Schreiber, G.: Enhancing Content-Based Recommendation with the Task Model of Classification, In *Proceedings of the 17th international conference on Knowledge engineering and management by the masses (EKAW 2010)*, LNAI 6317, pp.431–440 (2010)