

# Study support and integration of cultural information resources with Linked Data

Tetsuro KAMURA

Graduate University for Advanced Studies (SOKENDAI)  
Tokyo, JAPAN  
kamura@nii.ac.jp

Ikki OHMUKAI

National Institute of Informatics  
Tokyo, JAPAN

Toru TAKAHASHI

ATR Media Information Science Laboratories  
Kyoto, JAPAN

Hideaki TAKEDA

National Institute of Informatics  
Tokyo, JAPAN

Fumihiko KATO

National Institute of Informatics  
Tokyo, JAPAN

Hiroshi UEDA

ATR-Promotions Inc.  
Kyoto, JAPAN

**Abstract**—A museum collection search system called **Linked Open Data for Academia (LODAC) Museum** has been developed that uses **Linked Data**. The **LODAC Museum** identifies and associates artists, artworks, and museum information from some different museums to provide integrated data that are published as **Linked Data** with the **SPARQL** endpoint. This project's purpose is to provide an information distribution system that can share and publish a wide range of data as **Linked Data**, especially in the artistic and cultural fields in Japan. Different types of data are currently being integrated, and new approaches and support for studying these fields are being investigated.

*Museum; Linked Data; Semantic Web; RDF; SPARQL;*

## I. INTRODUCTION

In this paper, we introduce a prototype system called “LODAC Museum” to integrate museum information across multiple resources. We identified and associated artists and artwork information from some museum collections with different types of information to provide integrated views of them. Then we investigated the possibility of new approaches and support for studying arts and culture fields.

## II. PURPOSE

Valuable information should be used. To do so, we should establish a cycle of information, i.e., *Publish, Share, Collect, Use, and Create*. This is crucial in the creative fields, such as the arts and culture. For this purpose, information for re-use needs to be published. **Linked Data** exactly meets these needs since its purpose is to share data openly by using a re-useable format. Japanese museums maintain and publish information with the individual metadata schema. This leads to difficulty in crossover searching. Therefore, we only obtain fragments of information during a search and need to integrate information from several sources by using **Linked Data**. In addition, we suppose that new knowledge can be found and new methods discovered by using not only museum collection data, but also different sources, for example, libraries, thesauruses, terminology, and GIS.

## III. APPROACH

### A. Method

The **LODAC Museum** is an integrated metadata database of Japanese museum collections. It provides metadata for artworks, creators, and relevant museum information in various **RDF** formats. The data is now ca. 130,000 from 15 museums, **DBpedia Lite Japanese**, and **GIS** data from the National and Regional Planning Bureau. The procedure in the **LODAC Museum** is as follows:

- 1) *Scraping from Web pages*: Collect data from web pages in different sources, identify and extract metadata from each page, and store data with the identified metadata schema.
- 2) *Mapping Vocabularies*: Map from the individual metadata schema to the single common schema with the essential elements.
- 3) *Integrating unique items*: Identify the same items (artwork, creator, museum) across museum collections and associate them to single identifiers.
- 4) *Publishing*: Publish data as **RDF** with permalinks that work as identifiers for people, artworks, and museum locations, accessible through a **SPARQL** Endpoint. In this way, a user can use information with string and link data from other sites.

### B. Development

- 1) *Canonical Data*: This data will enable integration of data from different sources. We adopt the “**Japan Art Thesaurus**” as the canonical data that contain a lot of types of objects, like the creators, artwork titles, museum locations, books, etc.
- 2) *Vocabularies*: We do not describe detailed vocabularies for context. Since our purpose is to integrate information from different resources, we provide metadata schema only with the essential elements. These include people’s names, titles,

genres, and...etc., which are mostly general among metadata schema for museum sites.

3) *Problem of People's Names*: Kana pronunciation is a peculiarly Japanese problem. Therefore, we decided to use language tags to solve it (see Table 1). This also happens in other metadata vocabularies such as artwork titles.

TABLE I. KANA PRONUNCIATION

Person Reference	Property
Creator (Popular Name)	foaf:name / skos:prefLabel
Creator (pronunciation)	foaf:name @ja-hrkt / skos:altLabel
Peculiar title	foaf:nick
Peculiar title (pronunciation)	foaf:nick @ja-hrkt
Creator (English)	foaf:name @en / skos:altLabel

4) *Data Integration*: We provide two kind of resources. One is "ref-resource", which is taken from each museum's website and is assigned a unique identifier corresponding to the original information source. In contrast, the other resource is "id-resource", which is minimum metadata with a "LODAC identifier" that is different for all information sources. It also holds links "dc:references" to a single or multiple "ref-resource" identifier, which works as a hub to integrate information from different sites, while a "ref-resource" identifier maps between LODAC and museum website objects.

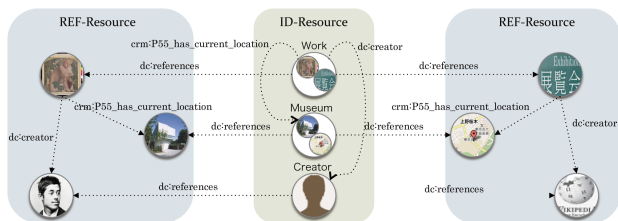


Figure 1. Data Integration

5) *Results of Integration*: We integrate museums, artworks, and people from multiple sources. For example, museum location are matched between "the Japanese Art Thesaurus (648 triples)" and "Cultural Heritage Online (915 triples)". As a result, we obtained 77 integrated resources of facilities for museum information. For further information see the work of Kamura et al. [2].

#### IV. RELATED WORK

Several studies have reported that cultural heritage information model CHoWDer is use as CIDOC-CRM and RDF. This model is described about a simple approach to interlinking cultural heritage information of different institutions [3]. Besides that, the CULTURE SAMPO is use a semantic web technology portal application for semantic annotation and search in huge information of Finnish cultural objects with original vocabularies and ontologies. They offer map-based interfaces for a user to large cultural objects and points of interest (POI) in Finland [4]. In a similar study by Cornejo et al, the eCultura project. They present semantic web applications to access and integrate diverse web-based contents of the cultural fields [5].

#### V. RESULT AND FUTURE WORK

The SPARQL Search is wide-ranging and flexibly used for searching for RDF data. For example, search for a work by entering Japanese painter's name and year the work was created across 15 museums' collection data, SPARQL returns integrated four museums and 20 works information. If we prepare more integrated data, we will be able to help the study of arts and culture by using a statistically based analysis [6].

This prototype system indicates the great capability of Linked Data to integrate artistic and cultural fields that are naturally separated. In particular, generating collections for individual artists from different sources is practically useful both for museum staff and visitors. However, there are still some issues to be solved:

- A. *Multiple creators names*: We need to reference a good artist thesaurus.
- B. *Updating resource*: Since LODAC Museum resources are taken from each museums web sites, we need find out how to give them regular or on-demand updates.
- C. *Links to other sites*: It is useful to have links to other data sites such as National Diet Library (Japan) and Europeana.
- D. *Easy participation*: We need to provide easier ways to participate in Linked Data such as importing collection data from CSV format data. This is particularly important for small museums.

We believe that further studying cultural information resources with Linked Open Data will greatly benefit all the artistic and cultural fields.

#### ACKNOWLEDGMENT

We thank the collaborators including the LODAC project and Kazunori YASHIRO (SOKENDAI) for valuable discussions.

#### REFERENCES

- [1] Hideaki Takeda, Ikki Ohmukai, Fumihiro Kato, Toru Takahashi, Hiroshi Ueda, Tetsuro Kamura. "Challenges for Linked Data in Japan", The 25th Annual Conference of the Japanese Society for Artificial Intelligence, 2011.
- [2] Tetsuro Kamura, Hideaki Takeda, Ikki Ohmukai, Fumihiro Kato, Toru Takahashi, Hiroshi Ueda: "Building Linked Data For Cultural Information Resources In Japan", Demonstation at Museum and the Web 2011, April, 2011
- [3] Jacek Jankowski, Yolanda Cobos, Michael Hausenblas and Stefan Decker: "Accessing Cultural Heritage using the Web of Data", 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2009)
- [4] Tomi Kauppinen, Panu Paakkari, Eetu Mäkelä, Heini Kuittinen, Jari Väätäinen and Eero Hyvönen: Geospatio-temporal Semantic Web for Cultural Heritage. Digital Culture and E-Tourism: Technologies, Applications and Management Approaches, 2010.
- [5] Cornejo, C.M.; Ruiz-Rube, I.; Dodero, J.M.; Comput. Languages & Syst. Dept., Univ. of Cadiz, Cadiz, Spain. "eCultura, a semantically-enriched web-based approach to manage cultural contents", 2010 IEEE International Conference on Information Reuse and Integration (IRI-2010)
- [6] Kazunori YASHIRO. "An Assessment Approach of Museum Collection Metadata on the Web", IPSJ-CH, May 17, 2011