

Improving Archival Records and Service of Traditional Korean Performing Arts in a Semantic Web Environment

Ziyoung Park^{1†}, Hosin Lee¹, Seungchon Kim², Sungjae Park¹

¹Department of Library & Information Science, Hansung University, Republic of Korea ²Department of IT Convergence Engineering, Hansung University, Republic of Korea

Abstract

Purpose: This research project aims to organize the archival information of traditional Korean performing arts in a semantic web environment. Key requirements, which the archival records manager should consider for publishing and distribution of gugak performing archival information in a semantic web environment, are presented in the perspective of linked data.

Design/methodology/approach: This study analyzes the metadata provided by the National Gugak Center's Gugak Archive, the search and browse menus of Gugak Archive's website and K-PAAN, the performing arts portal site.

Findings: The importance of consistency, continuity, and systematicity—crucial qualities in traditional record management practices—is undiminished in a semantic web environment. However, a semantic web environment also requires new tools such as web identifiers (URIs), data models (RDF), and link information (interlinking).

Research limitations: The scope of this study does not include practical implementation strategies for the archival records management system and website services. The suggestions also do not discuss issues related to copyright or policy coordination between related organizations.

Practical implications: The findings of this study can assist records managers in converting a traditional performing arts information archive into a semantic web environment-based online archival service and system. This can also be useful for collaboration with record managers who are unfamiliar with relational or triple database system.

Originality/value: This study analyzed the metadata of the Gugak Archive and its online services to present practical requirements for managing and disseminating gugak performing arts information in a semantic web environment. In the application of the semantic web services' principles and methods to an Gugak Archive, this study can contribute to the improvement of information organization and services in the field of Korean traditional music.

DOI: 10.2478/jdis-2020-0006 Received: Jan. 17, 2020 Revised: Mar. 17, 2020 Accepted: Mar. 25, 2020



Citation: Park, Ziyoung, Hosin Lee, Seungchon Kim, and Sungjae Park. "Improving archival records and service of traditional Korean performing arts in a semantic Web environment." *Journal of Data and Information Science*, vol.5, no.1, 2020, pp. 68–80.

[†] Corresponding author: Ziyoung Park (E-mail: zgpark@hansung.ac.kr).

Keywords Gugak archive; Korean traditional music; Performing arts archive; Linked semantic data; K-PAAN

1 Introduction

This research project aims to organize the archival information of traditional Korean performing arts in a semantic web environment. Key requirements, which the archival records manager should consider for publishing and distribution of gugak performing archival information in a semantic web environment, are presented in the perspective of linked data. Specifically we analyze the characteristics of the archival records in the Gugak Archive of the National Gugak Center with the cooperation of the records managers in the archives. Based on the analysis, we make suggestions for improving the organization of the archival information in terms of identifiers, data schema, and authority data. Technical aspects of the semantic web and linked data were also examined for the web service of the performing arts archival records. This is because the records manager should to understand and accept the proposals of study. We hope that the records manager could extend the legacy archival records management process and actively request for the improvement of their informationa organization practices.

In Section 2, the scope of research is defined, and in Section 3, the methodology is presented based on the semantic web and linked data principles. Section 4 describes the characteristics of gugak, traditional Korean music, and the Gugak Archive's performing arts archives. Section 5 discusses the use of metadata and funcations of web services for searching and browsing of the performing arts information on the Gugak Archive website. Lastly, in Section 6, implications are derived from the analysis results.

2 Scope of analysis: Gugak archival records and online finding aids

This study analyzes the metadata provided by the Gugak Archive, the search and browse menus of Gugak Archive's website and K-PAAN, the performing arts portal site from the perspective of the semantic web and linked data. Traditionally, an archivist analyzed the bibliographic attributes and subject materials of the archival records and produced their metadata. He/she also designs the tools for providing metadata search and the user interface. Under the traditional finding aids, the metadata structure produced by an archivist was virtually identical to the search results provided to users. The assignment of headings (access points) to enable searching or the ordering of shelf arrangment was also at the archivist discretion.

However, with the widespread use of online catalogues published through websites, it is becoming increasingly difficult for an archivist to personally design



the search or browse functions and services to re-organize search results of online catalogue. Unlike in a traditional environment, with online search tools, the metadata structure stored in a server is not identical to what's on the users' screens, including the search results. Depending on the system, functions and interfaces—which are distinct from the original data structure's intended search and browse functions—may be added. When organizing archival information, it is important to consider how information is displayed on the user's web interface. Accordingly, this study extended the analysis of the Gugak Archive's metadata structure to include how data are provided in the online environment on both the Gugak Archive's own website and the Korea Performing Art Archives Network (K-PAAN), performing arts search portal.

3 Framework for analysis: Semantic web and linked data principle

We share large amounts of information with others through web, and the senders and receivers of information vary depending on the type of information. Now, the most commonly used unit of information is the web page, which contains text, images, and videos. Every time we type a keyword into a search box, the search engine compares the query term with strings that were pre-extracted from web documents. A web search engine does not reveal the exact internal processes the search word goes through for thousands and even millions of hits to come up. The semantic web differs from the legacy worldwide web in many ways, mainly in the following characteristics (Antoniou et al., 2012):

- (1) make structured and semi-structured data available in standardized formats on the web;
- (2) make not just the datasets, but also the individual data-elements and their relations accessible on the web;
- (3) describe the intended semantics of such data in a formalism so that this intended semantics can be processed by machines.

The semantic web and linked data complement each other in terms of their purpose and usage. In the semantic web, for a machine (or robot) to access, understand, and analyze data, several technologies are required. These include three interconnected technologies: a resource description framework (RDF) data model to describe resources in triples, uniform resource identifiers (URI) for web resource identification, and an ontology for linking and inference of data. When datasets compliant with these technologies are created, data can be searched or arranged and then restructured using SPARQL. In other words, the semantic web publishes data in a way that the computer understands web documents, and when necessary, new documents can be produced.



Data built using semantic web technologies can be directly used as linked data without further processing. "Linked data" refers both to technologies for publishing certain web data and the data that are published using these technologies. In a technical context, "linked data" is a concept associated with semantic web technologies, which also include resource identifiers, such as URIs and IRIs, and RDF—the data structure. It is necessary to build an ontology, which provides the basis for data linking and inference, to increase the linked data's usability. Equally important is interlinking, which expands data within the web space by interlinking URIs. The linked data principles outlined by Berners-Lee make it clear that they include the semantic web's basic principles (Berners-Lee, 2006; Bizer, Heath, & Berners-Lee, 2009).

- (1) Use URIs as names for things
- (2) Use HTTP URIs so that people can look up those names
- (3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- (4) Include links to other URIs, so that they can discover more things

Brunetti et al. (2012) stated that most data exposed on the web are difficult to analyze or use as they are often in raw tabular form and argued that the use of linked data could improve data's structure and semantics. They, however, added that it is not easy for end-users to intuitively understand the structure of a dataset that is published as linked data. As a matter of fact, to make a SPARQL query on exposed linked data, a user needs to understand the overall data structure, such as classes, attributes, and even values. Accordingly, for our Gugak Archive data and web services analysis, we used the four following criteria:

- (1) Whether URIs are assigned in such a way that archival records can be continuously identified and accessed on the web:
- (2) Whether the archival records' metadata can be converted into an RDF format in a consistant way;
- (3) Whether the Gugak Archive website's information can be provided with information about related performances or similar performances from other organizations; and
- (4) Any differences between Gugak Archive information provided through the global search portal and Gugak Archive's own website.



4 Gugak Archive and performing arts information management

"Gugak" is a term that refers to traditional Korean music. The National Gugak Center (NGC) was established in 1951 for the preservation and legacy of traditional.

Korean performing arts. The NGC started its archiving project in 2007 and has emphasized the need to expand the Gugak Archive. The Gugak Archive began collecting and cataloging materials related to Korean classical performance, and in 2008, it began to develop a metadata and classification scheme for traditional arts archive. The Gugak Archive analyzes, collects, provides, and preserves Korean musical performances, ranging from traditional music, dance, plays, and Contemporary Korean music. At present, the Gugak Archive has produced and collected over 380,000 records, including videos, sounds, images, and textual materials. The NGC aims to provide Gugak-related information to the government and public sector, as well as foreign institutions (Kweon, 2016; NGC, 2019).

An important characteristic of performing arts is that they are not tied to specific media and are complete by and of themselves, by virtue of their expression. For this reason, while it is impossible to preserve performances as such, ironically, they tend to promote preservation through the two most common methods of recording—filming and photography. Moreover, the media environment's recent advancements have opened more possibilities in using recorded performances (Reason, 2006; Lee, 2018).

However, searching for information about performing arts—a field where aesthetic qualities are important—using keywords is still difficult because they are often in image or video form rather than text. For the same reason, browsing through a performing arts archive by text like a book is not possible because the performance posters, flyers, and program brochures are often in image files with special design elements. Therefore, searching and browsing, in this case, requires information with detailed metadata. Rather than conducting known-item searches, users may prefer versatility in browsing records by inputting their artistic preferences, genres, or themes, which the keyword-based search function or vertical hierarchical classification cannot adequately support.

Linked data can address these limitations in searching and browsing performing arts records. Waitelonis and Sack (2009) tested whether using linked data in multimedia search can provide unexpected yet interesting results for users. Suppose a user's query about a certain dancer's videos had no results, but through the linked data's link information, the website can suggest alternative videos featuring dancers with a similar style. This test making use of linked data's key characteristics demonstrates that linked data produces greater results when there are abundant relationships between the data—which are defined in detail—than when the data exists in an isolated storage (like islands) or when there is a limited network of hierarchical or equal relationships between the data (like a tree's structure).



5 Issues of using Gugak Archival information

5.1 Resource identification

When a new record is added into a library or an archive, it is assigned a shelf number or a call number. Likewise, when a new record is uploaded onto the web, it must be assigned a number according to a widely used and understood identification system in the web environment. Furthermore, these record numbers must be consistently managed so that they remain unchanged even when a website is redesigned or reorganized. URI-type numbers that are shown to users must be continuously valid to ensure uninterrupted user access to records. If the record numbers in a offline system can identify the collection to which a record belongs or describes a record's attributes, URIs provided on the web must have a similar or equivalent system.

Currently, information on traditional Korean performing arts and related performing arts archival records is provided not only in the NGC and Gugak Archive but also on APIs of the Cultural Data Plaza and integrated metasearch sites, such as K-PAAN. In this system, the Gugak Archive and the other sites create a link using a webpage URL, not permanent URIs. However, permanent URIs for the performing arts and related archival materials are needed. It is difficult to clearly identify and link the performing arts record on the web with a URL that only shows the page address. Also, in case of website renewal, the URL link could be broken. In order to improve the utilization of the archival information through the website, it is vital to give the URIs access to the archive's own data and provide the URI-granted data to external organizations.

Depending on policy or copyright management of the performing arts institution, there is a chance that the external connection to the archival information via direct URL will be blocked and provide encrypted URIs (do include). This must be improved because, if not, the data on the website may be isolated and users would be severely inconvenienced. It is also worth noting that when assigning a URI, a large number of performing arts archival materials are produced in a single performance. It is important to maintain the relationships between archival materials and the performance itself, too. Therefore, using permanent URIs to identify and link archival information among various web services is needed.

At present, we can estimate the data format of archival records on the web site through URL. Gugak metarials are managed by item and clip unit in Gugak Archive. If the data format is video, audio, or image, describe the video, audio, and the image as a whole in the item level, and describe the individual performance recordings as a clip. Unlike live performance recordings, items such as leaflets, pamphlets, banners or programs are not expanded by clip. The URL of the Gugak archive



materials contains the clip number, so we can guess the data format by URLs. In this URLs, the item number is "V013358", which contains the character of the data format, and the clip number is "20939", which is the serial number excluding the data format designation.

[Example of Item URL]

- Korean Court Music: Royal Ancestral Ceremonial Music-Chon Pye Hee Mun (eng)
- "한국의 궁중음악: 의식음악-종묘제례악 중 전폐희문" (kor) [1966]
- URL for this item "V3358": http://archive.gugak.go.kr/portal/detail/searchVideoAiDetail?clipid=V013358 &system_id=AI&recording_type_code=V (or http://archive.gugak.go.kr/portal/detail/searchVideoAiDetail?clipid=V013358)

[Example of Clip URL]

- Korean Court Music: Royal Ancestral Ceremonial Music-Chon Pye Hee Mun - 01. Korean Court Music-Royal Ancestral Ceremonial Music 'Chon Pye Hee Mun' (eng)
- "한국의 궁중음악: 의식음악-종묘제례악 중 전폐희문 [1966 추정] 01. 한국의 궁중음악-종묘제례악, 전폐희문"(kor)
- URL for this clip "20939" of the item "V3358": http://archive.gugak.go.kr/portal/detail/searchVideoDetail?clipid=20939&sys tem id=AV&recording type code=V

In addition, within the Gugak Archive, items are grouped once more by folder numbers for management purposes. Until now, they are not reflected in browsing or searching on the website of Gugak Archive.

5.2 Data schema based on ontology

The Gugak Archive designed its own classification and metadata scheme to categorize and organize archival records (Noh, 2017), the "Traditional Arts Archive Classification Scheme," which is linked to the Korean Decimal Classification (KDC). The KDC is widely used in Korean library classification, with the entry "679 Korean Traditional Music" used to link two classification schemes. The Gugak Archive also has its own descriptive metadata standard, "Traditional Arts Archive Metadata," which is based on International Standard Archival Description (ISAD (G)) and Dublin Core Metadata. It may not be scalable in the semantic web environment, even though a descriptive standard, such as ISAD (G), can guarantee interoperability in the records management community. Therefore, upper ontology



is needed to exchange and share data with various fields and institutions. ISAD (G) was also developed as an XML-based encoding schema, Encoded Archival Description (EAD) and considered as a basis for archival linked data (Tillman, 2018). There is also a research project that builds linked data based on the EAD such as Locah project (LOCAH Linked Archives Hub (2013)).

Physically and Logically organized performing arts records have been linked with other units of related records through vertical and horizontal relationships. Link information between records, which need to be maintained for searching and browsing through performing arts records, can be given various forms such as a network structure established through RDF-type expressions and linking methods. Moreover, link information can be created between the Gugak Archive's records and other institutions' performing arts records using linked data techniques. Linked data makes it possible for archival information to be more systematically and effectively provided.

In Archival Science, a study has been carried out to map the Encoded Archival Description, an XML version of ISAD (G), to CIDOC CRM (Bountouri & Gergatsoulis, 2011; Park, 2018). For example, a part of the current descriptive elements can be mapped into the CIDOC CRM class and attributed as follows.

Descriptive Elements Matching to Upper Ontology

Title → E19 Physical Object (P102 has title) E35 Title

(The name of an archival resource)

Time → E19 Physical Object (P4 has time-span) E2 Time-Span

(The time when an archival resource is created)

Place → E19 Physical Object (P55 has current location) E53 Place

(The name of a place where an archival resource is created)

Type of Records → E19 Physical Object E55 Type {Fonds, Series, Files, Items}

(The name of a type of material)

5.3 Interlinked performing arts information network design

In the semantic web environment, the link between resources with URIs, or those that interlink, increases data scalability and quality. To increase interlinked data, the value of descriptive elements must have URIs, not strings. This requires authority-controlled vocabularies. In the Gugak Archive, the vocabulary and meaning of Korean traditional music can be controlled using a Gugak dictionary and thesaurus. However, at present, mainly uncontrolled keywords are listed in the archival description of the Gugak Archive.

At present, the Gugak Archives provide detailed classification scheme and keywords (tags) on Korean traditional performances. The Korean traditional music



classification scheme, based on KDC, is marked with headings instead of class numbers. The gugak thesaurus is also built and provided on the Gugak archive website. Thesaurus has synonym clusters centered on descriptors, and provides BT, NT, and RT relationship information. The Korean traditional thesaurus consists of key terms that describe Korean traditional music in detail.

However, in the Gugak Archive website, the classficiation scheme and the thesaurus does not used sufficiently for navigation or reciprocal linking of Gugak performance records. The below is the screen of the Gugak thesaurus for the term, "Jongmyo rite music". The term cluster is formed around the descriptor.



Figure 1. Gugak Thesaurus term "Jongmyo jeryeak" https://archive.gugak.go.kr/portal/division/thesaurusMain.

However, the term relationships in the thesaurus are not yet applied to the Gugak Archive website when searching and browsing gugak performing records. On the screen for the individual records, classification information and keywords are displayed only. We cannot navigate the gugak term network by clicking the headings in the classification information. The figure below is an example of gugak record display. There is no link in the classification headings. If you click on a keyword, only the same keyword can be searched or not. In linked data, individual headings or index terms can be semantically linked and searched in conjunction with URIs.

In particular, performance works, performances, and performing arts records need to be interlinked to structure performing arts information. Because most works are performed several times, it is important to link them with the higher-level unit "contatiner work" as well, rather than just interlinking individual performances and performance records. It is also necessary to clearly distinguish performance works from individual works contained within a certain work and manage them separately.





Figure 2. Gugak archival records "Jongmyo jeryeak" http://archive.gugak.go.kr/portal/detail/searchVideo Detail?clipid=20939>.

Instead of simply entering search keywords related to individual works into the system, the data must be modeled in a way that treats them as separate works by also assigning a unique URI so that they may be linked to various authority data.

5.4 Browsing and searching Gugak archival information

The Gugak Archive's information is not only distributed through its own website but also through K-PAAN—a performing arts search portal. Besides Gugak Archive, K-PAAN also includes information from the Museum of Performing Arts, Arko Arts Archive, and the National Intangible Heritage Center, making it unrealistic to redesign the portal according to the Gugak Archive's requirements. Moreover, K-PAAN does not have its own identification system for the performing arts item as it indexes these organizations' data and performs a parallel search to display the results by organization successively. When a user clicks on a K-PAAN search result,



they are redirected to the source website because the portal does not have or manage its own link information on these four organizations' information. Going forward, for K-PAAN to provide LOD search services, it will need to implement required elements such as URIs, RDF, and interlinking information.

In addition, the classification and keyword information provided on the website of Gugak Archives are not equally reflected in K-PAAN. There is a difference in the unit of classification of records among the institutions that make up K-PAAN service, which may not be distinguished on the portal performing arts archive website. To improve this, the records manager and the website architeture should cooperate concretely the building criteria for the structure of records provided to K-PAAN. In most cases, the records managers of individual institutions have less authority over data structure or services of the web site in the case of portal sites than their own websites. It is also convenient for users to use the portal site as a starting point for searching. Therefore, the quality control of the portal site should also be considered.



Figure 3. Individual archival records in Gugak Archive website http://archive.gugak.go.kr/portal/detail/searchVideoDetail?clipid=20939.



Figure 4. Individual archival records in K-PAAN website. https://www.iha.go.kr/k-paan/main.



Discussion

The importance of consistency, continuity, and systematicity—crucial qualities in traditional record management practices in a library setting—is undiminished in a semantic web environment. However, a semantic web environment also requires new tools such as web identifiers, data models, and link information. We therefore propose the following suggestions to improve the openness and quality of the traditional Korean performing arts record in the semantic web environment:

- Identify traditional Korean performing arts and related archival records with URIs permanently and consistently.
- Design data schema for archival records based on the formal ontological point of view for data linking and sharing among various institutions.
- Refine name and subject authority data to strengthen the linkages between archival data, as well as interlinking with external web data.
- The records manager should be aware of the difference between providing archival services on individual websites and providing services on performing arts search portal, and design and support the user's search and browsing experience.

Acknowledgments

We wish to give our special thanks to Dasom Jung, Seunghee Son, and Yoonwhan Kim for their help and support throughout this project. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016S1A5A2A03927725).

Author contributions

Ziyoung Park (zgpark@hansung.ac.kr) led the analysis of gugak archival records and web services, and write a draft of the paper. Hosin Lee (leehs@hansung.ac.kr) was responsible for the analysis of the performing arts information and review the paper. Seungcheon Kim (kimsc@hansung.ac.kr) was in charge of open data system requirements and review the paper. Sung Jae Park (sungjae.p@gmail.com) was in charge of the quality control of linked data and review the paper.

References

Antoniou, G., Groth, P., Harmelen, F.V., & Hoekstra, R. (2012). A Semantic Web Primer, 3rd edition. Cambridge/London: The MIT Press.

BelliniPaolo, P., & Nesi, P. (2013). A linked open data service for performing arts. In Nesi, P., & Santucci, R. (Eds.) Information Technologies for Performing Arts, Media Access, and Entertainment: proceedings of Second International Conference, ECLAP 2013, Porto, Portugal, April 8–10, 2013, Revised Selected Papers (pp. 13–25). Berlin, Heidelberg: Springer.



- Berners-Lee, T. (2006). Linked Data—Design Issues. Retrieved from: https://www.w3.org/DesignIssues/LinkedData.html.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. International Journal on Semantic Web and Information System, 5(3), 1–22.
- Bountouri, Lina & Gergatsoulis, Manolis. (2011). Mapping encoded archival description to CIDOC CRM. Advances in Water Resources—ADV WATER RESOUR.
- Brunetti, J.M., Rosa, G., Gimeno, J.M., & Garcia, R. (2012). Improved linked data interaction through an automatic information architecture. International Journal of Software Engineering and Knowledge Engineering, 22, 325–343.
- Kweon, Hyekyung. (2016). Current status and future of Gugak archive for collecting and providing Gugak information resources. In proceedings of the Seminar on Gugak Archive: Role and Value Creation in the Digital Age, Performing Arts Archive, Seoul, Korea.
- Lee, H. (2018). Performing arts videos: Beyond the record. Journal of National Gugak Center, 40, 59–79. http://doi.org/10.29028/JNGC.2018.37.059
- LOCAH Linked Archives Hub. (2013). Retrieved from http://data.archiveshub.ac.uk/
- National Gugak Center. (2019). Introducing Gugak Archive. https://archive.gugak.go.kr/portal/main/introduction
- Noh Yea-ri. (2017). The preservation and utilization of Korean traditional music records—Focusing on the records housed at Gugak Archive. The Society for Korean Historico-Musicology, 59(0), 43–82.
- Park, Z. (2018). An exploratory study on linking ISAD(G) and CIDOC CRM using KARMA. Korean Society of Archives and Records Management, 18(2), 189–214.
- Park, Z. (2019). Performing arts information: User-friendly knowledge publication and service design on the Web: With focus on linked semantic data. In proceedings of 2019 Gugak Archive Academic Seminar. Seoul: National Gugak Center.
- Park, Z., Lee, H., Kim, S., Park, S., Jung, D., Son, S., & Kim, Y. (2019). Improving archival records of traditional Korean performing arts in a semantic web environment. Presentation at the DCMI2019. September 23–26, 2019. Seoul, Korea.
- Park, Z., Lee, H., Kim, S., Park, S., Jung, D., Son, S., & Kim, Y. (2020). Organizing performing arts records of Korean traditional music in a semantic web environment—Focusing on the case of Gugak archive. In proceedings of 16th ISKO Conference, Aalborg, July 2020. Baden-Baden: Ergon. [In progress].
- Reason, M. (2006). Documentation, disappearance and the representation of live performance. New York: Palgrave Macmillan.
- Tillman, R.K. (2018). Opportunities for encoding EAD for linked data extraction and publication. Journal of Archival Organization, 19, 19–36.
- Waitelonis, J., & Sack, H. (2009). Towards exploratory video search using linked data. In proceedings of 11th IEEE International Symposium on Multimedia (pp. 540–545). San Diego, CA, USA.
- Wikipedia. (2019). Music of Korea. https://en.wikipedia.org/wiki/Music of Korea



Journal of Data and Information Science

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).