

The Development of Application Profile for OAK Institutional Repository *Poster*

Mihwa Lee
Kongju National University,
South Korea
leemh@kongju.ac.kr

Jee-Hyun Rho
Pusan National University,
South Korea
jhrho@pusan.ac.kr

Eun-Ju Lee
Donggeui University,
South Korea
ejulee@deu.ac.kr

Keywords: institutional repository; metadata; OAK; application profile; Dublin Core; DSpace

Abstract

OAK (Open Access Korea) hosted by National Library of Korea is the national portal for integrated search of IRs, participating universities, public institutions, researches, and businesses. OAK had collected metadata from member IR systems, and has accumulated them according to the OAK metadata scheme by mapping IRs metadata into OAK metadata scheme. OAK has developed OAK metadata scheme based on DSpace to build OAK portal in 2009, but the initial OAK metadata schemes could not accommodate all the elements that participant IRs wanted. Elements of member IR metadata were redundant, disorganized and inappropriate. As a result, it brought into missing metadata in harvesting and mapping the metadata from member IR systems, and into searching without the elaborate elements. For solving these problems, this study is to suggest the OAK application profile through developing new OAK metadata scheme by accommodating member IRs metadata and comparing elements between OAK metadata and major IRs such as DSpace, Eprints, Bepress, ETD-db, and dCollection sophisticatedly. As the research methods, it is used to analyze the 17 representative IRs' metadata schemes and compare the major IRs' metadata schemes for conforming standard metadata scheme.

Analyzing the Metadata Elements of 17 IRs

The metadata schemes of 17 among 33 OAK IRs were analyzed in aspect of 15 main elements of Dublin Core.

In title, various titles were used as sub-elements such as translated title, original title, subtitle, part title, and part number. Subject element was qualified into sub-elements for differentiating local classification schemes and various subject headings vocabularies. In description, claim (request for a patent), version (peer-reviewed), and provenance (owner of resource) were used as the element refinement. In contributor, various contributors were used with sub-element such as editor, illustrator, examiner, googleAuthor, college, department as well as author, advisor. Date was divided into element refinements such as created, available, issued, submitted, accessioned, updated, valid, and modified. In type, various type encoding schemes were used including MARC type, DSpace type, and so on. So, the mapping table between resource type schemes should be made to retrieve the resources by resource type. In identifier, it was not affordable to accept all the identifiers such as uri, govdoc, isbn, ismn, issn, sici, patentno, pmid (pubmed identifier), scopusid (scopus identifier). In citation, citation description methods were various in IRs -- one IR subdivided citation into several element qualifiers such as title, volume, date, genre, identifier, issueno, and page, other IR described the citation information under the relation element, and another IR moved the citation to the main element not as the sub-element of identifier. The citation was the element in need to uniformity.

There are some difficulties in retrieving the resources in national portal level because each IR has used its own elements and sub-elements. Therefore, new OAK metadata scheme should be developed to consolidate all the IRs metadata elements.

Comparing the major IRs' metadata

The representative IRs' metadata schemes in the world were selected for comparison of metadata elements such as DSpace, Eprints, BEPress, ETD-db, and dCollection (IR for college and university libraries in Korea, host by KERIS). <TABLE 1> shows the comparison of only title and description sections of 5 IR metadata schemes.

TABLE 1: Comparison table of title and description sections of 5 IR metadata schemes (in part).

DSpace		EPrints	BEPress	ETD-db		dCollection	
element	qualifier			element	qualifier	element	qualifier
title		title	title, article title	title		title	
	alternative	alternative title			alternative		alternative
							subTitle
							translated
description		description		description		description	
	tableofcontents						tableofcontents
	abstract		abstract		abstract		abstract
	provenance	provenance					provenance
	sponsorship	funder					sponsorship
		grant number					
		parent project					
	statementofresponsibility						
	uri						descriptionURI
			embargo period				
			versions		release		
			peer reviewed				indexed
		publication status	publication status				
		data collection method					
		contact					
		administrative note			note		
		additional information					
				degree	name		
		divisions*	disciplines		discipline		
					grantor		
					level		localRemark
			comments				

This table was based on Chung, Yeon-Kyoung, Na-Nee Lee and Mihaw Lee (2007), KERIS (2005), BEPress Home Page, DSpace Home Page, Ensom, Tom and Alexis Wolton (2012), Jones, Richard (2004), Digital Commons@Otterbein.

EPrints includes special elements and sub-elements to reflect the project related information such as funder, grant number, and parent project. BEPress has specialty in describing journal article, for example, version (preprint, postprint, published), publication status, peer reviewed, and embargo period. ETD-db develops thesis related element such as degree, name, level and discipline. dCollection includes alternative, subtitle, and translated in title, and indexed as the sub-elements of description. Also, dCollection has the various sub-elements in identifier.citation to build the citation information - citationTitle, citationIdentifier, citationGenre, citationIssueNo, citationVolume, citationNumber, citationTissueNo, citationPages, citationStartPage, citationEndPage, citationConferenceNumber, citationConferencePlace, citationConferenceDate,

citationEdition, citationAuthor, place, duration, publisher, uriType, uriEntity, ISBN, ISSN, SICI, KERISIdentifier.

It is to find that OAK metadata scheme should have been developed with the various elements and sub-elements through analyzing the world wide IRs metadata schemes. These elements and sub-element should be applied to OAK metadata schemes.

Development of New OAK Application Profiles

The objective of this study was to suggest new OAK application profiles by analyzing the elements and sub-element of member IRs metadata and comparing major international IRs. The previous OAK metadata scheme was changed -- new elements and sub-elements were added, and elements with similar meaning were integrated into one element or sub-element. The new OAK application profile's features are as follows:

- The different elements which have same meaning are unified to one element or sub-element, for example, in description, abstract, summary as sub-elements → summary.
- Subject, degree, eprintVersion, contributor, nameIdentifier, and identifier are changed to subjectType, degreeType, eprintType, contributorType, nameIdentifier, and identifierType to select type from controlled vocabularies to accommodate various elements of member IRs and to differentiate the value. For example, it is possible to differentiate various contributor such as author, advisor, editor, translator, illustrator, examiner, department, reviewer by inputting type from controlled vocabularies of contributorType.
- New elements and sub-elements are added such as subject.keyword (keyword written by author), and description.degree (degree type).
- According to Dublin Core Metadata Initiative Citation Working Group (2005), bibliographicCitation is to capture the bibliographic citation information for a resource, but, bibliographicCitation is not enough to construct and describe the citation information in uniform. Therefore, in OAK, the citation element was changed as the main and administrative element and was subdivided into citation.title, citation.volume, citation.number, citation.date, citation.startPage, citation.endPage, citation.conferenceName, citation.conferenceNumber, citation.conferenceDate, citation.author, citation.author, citation.edition, citation.place, citation.publisher to embrace all kinds of resource's citation information. These citation related elements are used only as administrative element to get the data from input interface, and identifier.bibliographicCitation was used as element in displaying the citation information by collecting data in citation such as Table 2.

TABLE 2: Citation description example

Property	Value String	notes
OAK:title	A critique of the FRBR user task and their modifications	
OAK:contributor (value=author)	Hider, Philip	
OAK:publisher	Taylor & Francis	
OAK:identifier.bibliographic Citation	Cataloging and Classification Quarterly, 55(2), 55-74.	This data is displayed for the citation information.
OAK:citation.title	Cataloging and Classification Quarterly	This data is used only for getting the data from user.
OAK:citation.volume	55	
OAK:citation.number	2	
OAK:citation.date	2016	
OAK:citation.startPage	55	
OAK:citation.endPage	74	

Future work

This study is to make OAK AP to accommodate all member IR metadata according to international standard such as DC, DSpace et al. This OAK AP would contribute to uniformity and consistency of portal metadata. After system development using new OAK metadata scheme, user survey and evaluation must be completed as the future works.

Acknowledgements

This paper was funded by National Library of Korea.

References

- BEPress Home Page. Retrieved, September 20, 2016, from <http://www.bepress.com>.
- Chung, Yeon-Kyoung, Na-Nee Lee and Mihaw Lee. (2007). A study of an extension of metadata for institutional repository. *Journal of the Korean Society for Library and Information Science*, 41(1), 323-344.
- Digital Commons@Otterbein. Retrieved, September 20, 2016, from http://digitalcommons.otterbein.edu/metadata_elements_master_list.pdf.
- DSpace Home Page. Retrieved, September 20, 2016, from <http://www.dspace.org>.
- Dublin Core Metadata Initiative Citation Working Group. (2005). Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata. Retrieved, September 20, 2016, from <http://dublincore.org/documents/dc-citation-guidelines>.
- Ensom, Tom and Alexis Wolton. (2012). RDE Metadata Profile for EPrints. Retrieved, September 20, 2016, from http://www.data-archive.ac.uk/media/375386/rde_eprints_metadataprofile.pdf.
- Jones, Richard. (2004). DSpace VS. ETD-db: Choosing software to manage electronic theses and dissertations. *Ariadne*, 38. Retrieved, September 20, 2016, from <http://www.ariadne.ac.uk/issue38/jones>.
- KERIS. (2005). dCollection Manual for System Managers. Seoul: KERIS.