Adopting DC Metadata for Union Serial System of KERIS: It's Design and Implementation

Yong Soon KIM, Ji Won LEE, Hyeong Yong PARK
Korea Education & Research Information Service(KERIS), Arirang Tower, 1467-80, Seocho-Dong, Seoch-gu, Seoul, 137-070, Korea
{ysoonkim, jwlee, hypark}@keris.or.kr

Abstract

This paper describes an overview of design and implementation of Union Serial System using Dublin Core(DC) Metadata. The project of developing Union Serial System was for making a better research information service environment leaded by Korea & Education Research Information Service(KERIS). In this project, we defined the qualifiers and additional data elements for locally specific application based on DC Metadata. All the records in Union Serial System were originated from university library as MARC, which was converted into DC Metadata. In this paper, we will introduce the process of designing Union Serial System; data model with Metadata element set, qualifiers, mapping MARC into DC Metadata, system architecture, and tribulations in this project.

Keywords: DC Metadata, Union Serial System, Digital Library, MARC, Qualifier

1. Introduction

The project of developing Union Serial System was proceeded from October 1999 to July 2000 for the purpose of providing better research information service to Korean researchers on the web by the KERIS.

In this project, 37 universities attended to provide their serial catalogs with the following goals; providing a Interlibrary Loan (ILL) service to one another through this Union Serial System and making available for researchers to access research resources and their findings easier and faster way. Now Union Serial System is running on the web page http://www.riss4u.net.

In this project, we adopt the DC Metadata in designing data model rather than making use of MARC format.

The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies and commercial organizations. The key characteristics of the Dublin Core are simplicity, semantic interoperability, international consensus, extensibility, and modularity. The Dublin Core is positioned as a simple information resource description format. However, it also aims to provide a basis for semantic interoperability between other, often more complicated, formats.

Implementations in many countries and languages, and in many disciplines testify to the widely perceived need for such a metadata element set, and the Dublin Core[1][2] is the leading candidate for achieving the goal of simple resource description for Internet resources. If we use the same database scheme and data structure, it will become much easier to share the information among the heterogeneous database system and search engine[3].

One of the major reasons we adopted the DC Metadata is the international acceptance as a *de facto* standard and the future interoperability such as cross-search or meta-search among the heterogeneous resource systems.

This paper explores brief overview of procedural process of design and implementation of Union Serial System; status of cataloging serial data, data model, mapping MARC into DC Metadata, system architecture, and difficulties in implementing this system based on our practical experience.

https://doi.org/10.23106/dcmi.952106693

2 Union Serial System

2.1 Data Elements Set and Qualifiers

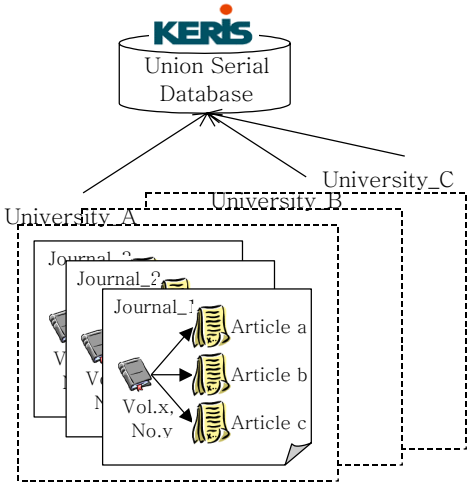

Figure 1. Serial data structure in a real world.

Figure 1 shows the simplified description of a real world of Journal data structure in each university library. Those data are merged into Union Serial Database in Union Serial System of KERIS.

KERIS journal and article index of journal service was designed to implement providing the researchers which journal is kept in which library. Researchers or finder can get proper findings for their research activity by retrieving Union Serial Database. Union Serial Database includes holding information that encodes which library keeps what journal. This holding information gives a clue to users which library keeps their findings and wanted research outcomes. The Union Serial System was designed to implement ILL service based on this holding information.

In the process of building this service system, 37 university libraries attended to share their serial catalogs through KERIS Union Serial System. Bibliography of serial catalog is most recorded as USMARC and KORMARC by catalogers of each university library. USMARC is habitually used for cataloging foreign book and KORMARC for books published in Korea. The way of cataloging volume, number, and article index depends on library automation system library employs. According to the survey in January 2001 by KERIS, 73% of 4 years university, 134 universities, regulate volume and number information of serial. 59 libraries employ holding MARC and 74 libraries employ bibliography MARC to catalog volume, 108 libraries adopt the check-in library system, which treats volume and number as text base attributes as a part of DBMS. In aspect of article index, 96 universities out of 160 regulate table of contents data. 14 university libraries were reported they do not have any systems for cataloging article index according to this survey.

KERIS collects serial catalog and convert it into metadata based on DC. The main data of serial catalog information in KERIS is mainly academic journals excluding general popular magazine.

		Journal	Article
Appli-cation Data Set	**DCMES**	Title	Title
		Creator	Creator
		Subject	Subject
		Description	Description
		Publisher	Publisher
		Contributor	Contributor
		Type	Type
		Date	Date
		Format	Format
		Identifier	Identifier
		Source	Source
		Language	Language
		-	-
		-	-
		-	-
	Locally needed specific elements	Volume_holding ILL_YesNo…	ILL_YesNo…

ILL: Inter-Library Loan

Table 1. Data sets for Union Serial System

Table 1 simplifies the data structuring aspect of serial service system application.

The data set of application is classified into two parts, Dublin Core Metadata Element Set(DCMES) and locally required element data. DC Metadata is necessary factors for representing attributes of Journal and articles such as title, author name, description language, publisher and so on. The other elements contain additional information for data manipulation and retrieval aspect of specific application such as ILL service

https://doi.org/10.23106/dcmi.952106693

based on Metadata attributes of Journal and Article. We draw a distinction between the attributes of metadata and functions of metadata in conceptualizing real world data.

DC Metadata contains 15 elements with qualifiers that describe and specify the meaning of each element in decoding local application. For example, the real object of book can be described with elements of title, author, publishing date, and so on. Each element can be refined and specified by defining refinements and qualifiers to decode the exact meaning of elements for local application. Element title can be translated into originally author named the resource or any form of title or alternative name of resource such as translated title in foreign language. These elements and qualifiers are optional and repetitive.

In Union Serial System, holding information of each serial should be managed to implement ILL feasibility. Locally specific needed elements are represented as another entity and keep relations with main entity of DC Metadata in ER model. This information is required for data manipulation based on DC Metadata and qualifiers.

The results of process conceptualizing real object of serial draw specific local refinements of each DCMES and elements set for ILL service. Local encoding way of DCMES is represented in table 2. We listed items that can express the attributes of Journal and Article itself. In determining the makeup of these qualifiers, we considered some of items which can be picked out from the MARC records and which notation can be automatically acknowledged by the computer in converting MARC into DC Metadata notation automatically.

In table 2, not only qualifiers but also specific definition with schemes that aid the interpretation of qualifier's value are indicated. These schemes include formal notation such as a term from a classification system or a string formatted in accordance with a formal notation (e.g., 2000-01-01 as the standard expression of date). Totally we defined the 65 qualifiers in this project and we decided to hire 46 qualifiers in expressing serial and 44 qualifiers in expressing each article index in journal.

Qualifiers	Brief Definition(Scheme)	Adoption	Usage
DC.TITLE	The name given to the resource, usually by the Creator or Publisher, e.g. Title in Korean	S, A	M
DC.TITLE.PARALLEL	Any form of the title used as a substitute or alternative to the formal title of the resource, e.g. Title in English	S, A	O
DC.TITLE.SUBTITLE	Subtitle of resource e.g. in another language	S, A	O
DC.CREATOR.PERSONALNAME	The person or organization primarily responsible for creating the intellectual content of the resource.	S, A	O
DC.CREATOR.PERSONALNAME.ALTERNATIVE	Any form of name used as a substitute or alternative to the author, name in English	S, A	O
DC.CREATOR.PERSONALNAME.EMAIL	Email address of author	S, A	O
DC.CREATOR.PERSONALNAME.AFFILIATION	Organization which author belongs to	S, A	O
DC.CREATOR.PERSONALNAME.HOMEPAGE	Homepage address of author	S, A	O
DC.CREATOR.CORPORATENAME	Author name of a group or coauthor	S, A	O
DC.CREATOR.CORPORATENAME.ALTERNATIVE	Any form of name of author of a group or coauthor	S, A	O
DC.CREATOR.CORPORATENAME.EMAIL	Email address of the author of a group	S, A	O
DC.CREATOR.CORPORATENAME.AFFILIATION	Author's affiliated institution	S, A	O
DC.CREATOR.CORPORATENAME.HOMEPAGE	Homepage address of the author	S, A	O
DC.SUBJECT	The topic of the content of the resource (LCSH,MeSH,DDC,LCC,DC)	S, A	O

https://doi.org/10.23106/dcmi.952106693

DC.DESCRIPTION	Keywords or phrases describing the subject	S, A	O
DC.DESCRIPTION.ABSTRACT	Abstract of resource	A	O
DC.PUBLISHER	Entity responsible for making this work available in its present form	S, A	M
DC.PUBLISHER.ALTERNATIVE	Alternative name of publisher	S, A	O
DC.PUBLISHER.EMAIL	Publisher's email	S, A	O
DC.PUBLISHER.HOMEPAGE	Homepage of publisher's	S, A	O
DC.PUBLISHER.COUNTRY	Publisher's country	S, A	O
DC.CONTRIBUTOR.PERSONALNAME	Any form of name and title of contributor	S, A	O
DC.CONTRIBUTOR.PERSONALNAME.ALTERNATIVE	Any form of subtitle and name of contributor	S, A	O
DC.CONTRIBUTOR.PERSONALNAME.EMAIL	Contributor's email address	S, A	O
DC.CONTRIBUTOR.PERSONALNAME.AFFILIATION	Organization which contributor belongs to	S, A	O
DC.CONTRIBUTOR.PERSONALNAME.ROLE	Role of contributor	S, A	O
DC.CONTRIBUTOR.PERSONALNAME.HOMEPAGE	Contributor's homepage	S, A	O
DC.CONTRIBUTOR.CORPORATENAME	Group contributor's name	S, A	O
DC.CONTRIBUTOR.CORPORATENAME.ALTERNATIVE	Group contributor's alternative name	S, A	O
DC.CONTRIBUTOR.CORPORATENAME.EMAIL	Group contributor's email address	S, A	O
DC.CONTRIBUTOR.CORPORATENAME.AFFILIATION	Organization which group contributor belongs to	S, A	O
DC.CONTRIBUTOR.CORPORATENAME.ROLE	Role of group contributor	S, A	O
DC.CONTRIBUTOR.CORPORATENAME.HOMEPAGE	Group contributor's homepage	S, A	O
DC.DATE.CREATED	Year of publication, Responsible for making the resource available, e.g. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format	S, A	M
DC.DATE.DISCONTINUED	Year publishing ended	S	O
DC.DATE.Republication	Year of republishing	S	O
DC.DATE.METADATACREATED	Date on which the resource metadata was created	S, A	M
DC.DATE.METADATAMODIFIED	Date on which the resource was changed	S, A	O
DC.TYPE	Category of resource	S, A	M
DC.TYPE.INTERVAL	Any form of Interval period of publishing resource	S	O
DC.FORMAT	Format may include the media-type or dimensions of the resource.	S, A	O
DC.FORMAT.PAGE	Start page number and final page number based on printed copy of resource	A	O
DC.IDENTIFIER	Unique ID of the resource, e.g. URL or URN	S, A	O
DC.SOURCE	A Reference to a resource from which the present resource is derived	S, A	M(A) O(S)
DC.SOURCE.TYPE	Any form of type of referenced resource	S, A	O

https://doi.org/10.23106/dcmi.952106693

DC.SOURCE.VOLUME	Any form of volume and number information of resource	S, A	M(A) O(S)
DC.SOURCE.INTERVAL	Any form of interval period of resource publishing interval	S	O
DC.LANGUAGE	Description Language of resource (ISO 639-2)	S, A	O

S: Serial, A: Article, M: Mandatory, O: Optional

Table 2. Defining DC Qualifiers for Serial and Article Service System.

2.2 ER Model

Figure 2 describes main entities, their attributes, and relation among the entities for Union Serial System. This evolved from figure 1 into Entity Relation(ER) model.

This ER model is composed of four parts; DC element set and qualifiers part, serial data structure, article data structure, and volume holding information part.

15 DC elements set, qualifiers, and their schemes are transformed into entities 'Meta_Element', 'Meta_Type', and 'Meta_Scheme'.

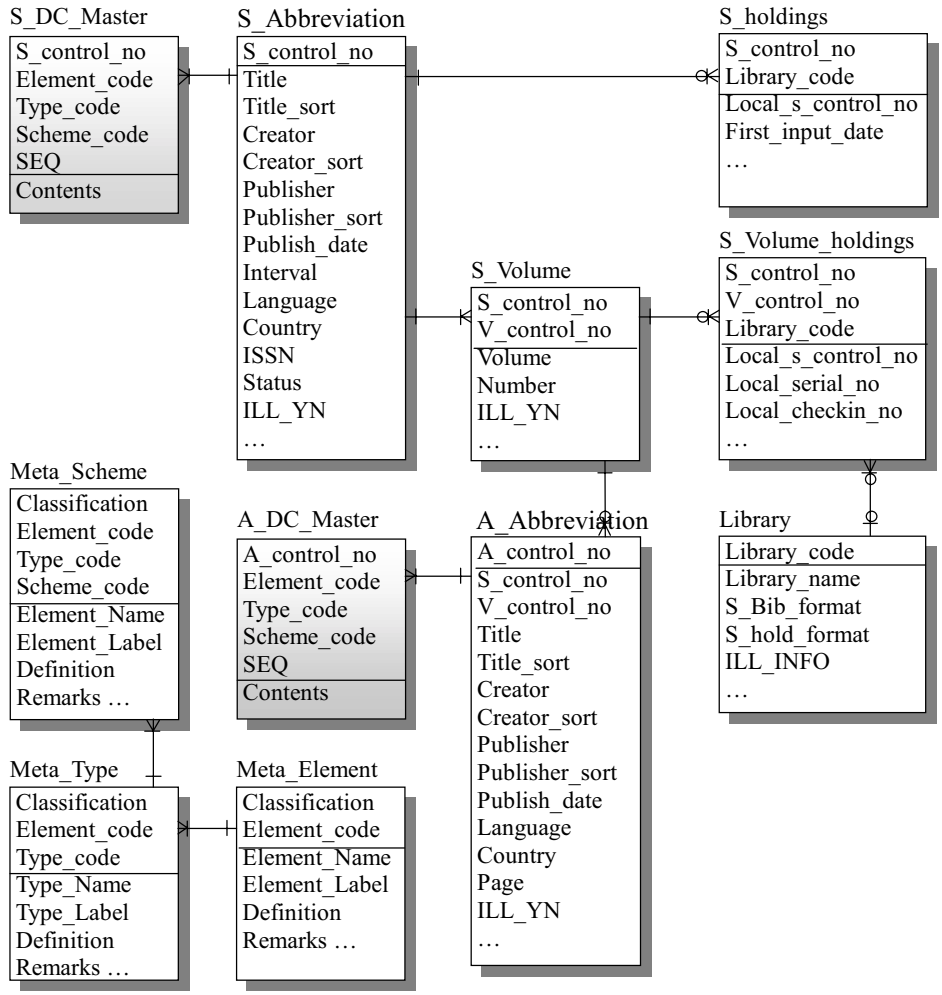

Figure 2. ER Model for Union Serial System.

https://doi.org/10.23106/dcmi.952106693

In data model, symbol 'S_' stands for 'Serial' and 'A_' stands for 'Article'. Entity 'S_DC_Master' and 'S_Abbreviation' contain serial information. One serial can be described as one record in entity 'S_Abbreviation' and several records in 'S_DC_Master'.

Entity 'A_DC_Master' and 'A_Abbreviation' treat table of contents of each serial. Volume holding information, which include which library keep which serial, is special information needed in local application in order to implement ILL service among the libraries. Each article includes volume and number information of journal in expressing DC element set but volume and number data is handled S_volume entity in expressing serial into DC.

The entity ending with '_Abbreviation' contains application search specific information about serial and article such as Interlibrary loan feasibility. The entity ending with '_DC_Master' contains details DC qualifiers.

The process of data retrieving responding user's query requires JOIN process among the entity in DBMS. The entity '_Abbreviation' make it easier to view main information for user's browsing without requiring JOIN process in DBMS. Viewing simple search result does not require the JOIN function among the entity in this application. 15 DC Metadata Elements are stored in an entity, ending name '_DC_Master'. If 15 DC Element was designed to store corresponding each 15 entities, it requires join at least 3 entities and at most 16 ones. For this reason, we designed the '_Abbreviation' entity that includes enough data to list simple search result. Each record of brief result kept a link with detailed information referring '_DC_Master' entity.

As early mentioned in section 2.1, 27% of 4 years university libraries do not catalog volume and number of serial. In order to accept this practical situation, we design the 'S_holdings' entity. A debate on linkage between 'S_holdings' and 'A_Abbreviation' entity should be needed more.

2.3 Mapping MARC into DC

Table 3 and 4 show qualifiers with corresponding mapping information between MARC and DC qualifiers. In Korea, KORMARC is habitually used for recording book and serials published inside of Korea and USMARC is used for foreign book. In the process of defining mapping between MARC and DC Metadata, we found out MARC and DC are not fully reciprocal translocation.

Qualifiers	KORMARC	USMARC
DC.TITLE	245 $a+ 245 $p	245 $a+ 245 $p
DC.TITLE.PARALLEL	245 $x	245 $x
DC.TITLE.SUBTITLE	245 $b, 245 $c	245 $b
DC.CREATOR.PERSONALNAME	100 $a, 700 $a	100 $a, 700 $a
DC.CREATOR.CORPORATENAME	110 $a+110 $b, 710 $a+710 $b	110 $a+110 $b, 710 $a+710 $b
DC.SUBJECT	653 $a, 650 $a, 056 $a, 082 $a, 080 $a	653 $a, 650 $a, 082 $a, 080 $a, 050 $a
DC.PUBLISHER	260 $b	260 $b
DC.PUBLISHER.COUNTRY	81503	81503
DC.DATE.CREATED	80704	80704
DC.DATE.DISCONTINUED	81104	81104
DC.TYPE	0082101=p&0082901=0, 0082101=&0082901=0, 0082101=*&0082901=1	0082101=p&0082901=0, 0082101=&0082901=0, 0082101=*&0082901=1
DC.TYPE.INTERVAL	81801	81801
DC.IDENTIFIER	022 $a, 010 $a	022 $a, 010 $a
DC.LANGUAGE	0083503, 041 $a	0083503, 041 $a

+: concatenation of each value

Table 3. Mapping table DC qualifiers with corresponding MARC in Journal.

DC Qualifiers	KORMARC	USMARC
DC.TITLE	245 $a	245 $a
DC.TITLE.PARALLEL	245 $x	245 $x
DC.TITLE.SUBTITLE	245 $c	
DC.CREATOR.PERSONALNAME	100 $a	100 $a, 700 $a
DC.CREATOR.PERSONALNAME.ALTERNATIVE	700 $a	700 $a
DC.CREATOR.PERSONALNAME.AFFILIATION	502 $b	502 $b

https://doi.org/10.23106/dcmi.952106693

DC.SUBJECT	653 $a	653 $a
DC.DESCRIPTION	510 $a	510 $a
DC.DESCRIPTION.ABSTRACT	520 $b	520 $b
DC.PUBLISHER.COUNTRY	81503	81503
DC.DATE.CREATED	80704	81503
DC.FORMAT.PAGE	300 $a	300 $a
DC.LANGUAGE	0083503, 041 $a	0083503, 041 $a

Table 4. Mapping table DC qualifiers with corresponding MARC in Article.

The level of recording of MARC depends on cataloger in each library. In case of detailed MARC record, it looses some data when it converted into DC Metadata. Even detailed MARC data does not include author subsidiary information such as alternative name and affiliation except name

2.4 System Architecture

The conceptual structure of Union Serial System is summarized in Figure 3. User can access and retrieve Union Serial Database through web. File system 'Article Index' and 'Serial Index' are index data created by search engine for retrieving. Search results have two levels of browsing; first level is amount to simple result referring to entity '_Abbreviation'. Second level forms each result linked detailed information referring to '_DC_Master' entity.

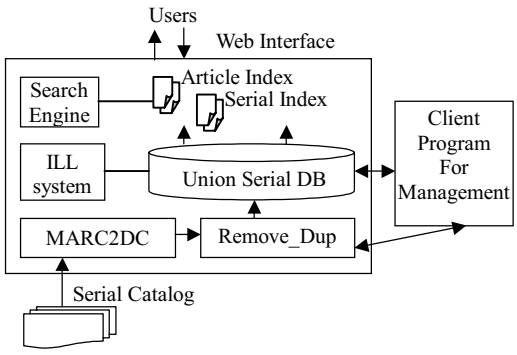

Figure 3. System architecture of Union Serial System.

Serial catalog originate in each university library as MARC. KERIS collects serial catalog and maps it into DC Metadata pattern through 'MARC2DC' program. When a new record of DC appends Union Serial DB, 'Remove_Dup' server program checks existence of identical record in Union Serial DB. If the identical record already exists, only the holding information appended, otherwise the new record and holding information are inserted into Union Serial DB. This procedural steps can be processed both automatic and manually by the person. Holding information is valuable for providing ILL service to the users.

Figure 4 and 5 shows an example of client program interface for manager. Client programs are built for the purpose of easy of monitoring this system for administrator. Interface of web for researcher is different from that of client program. Client program is implemented by the Power Builder software tool.

Figure 4 shows an example of 'S_Abbreviation' and 'S_DC_Master' of figure 2 ER model. Each record of 'S_Abbreviation' is connected with 'S_DC_Master'. Mapping key attribute is 's_control_no'. Left side of screen is each record of Journal and right side of screen shows the detailed contents of DC qualifiers of record in 'S_Abbreviation', reverse colored third record.

Figure 5 shows information of volume and holding information. Left side of screen shows each journal in 'S_Abbreviation' and reversed color record, third record' was chosen and it's volume information was appeared in right side of screen. This information comes from 'S_Volume' entity in ER model of Figure 2. Each records of 'S_Volume' is connected to holding information. In figure 5, Journal 'Korea Informaion Processing…' is composed of 21 lists of volume and number covering from 1997 to 2000. among the journals, publishing year 1997, volume 4 number 2, is kept by the two libraries, the central library of Yonsei University and Seowon university library.

Real Object	Name of Entity In Figure 2	The Number of Records
Serial	S_Abbreviation	75,940
	S_DC_Master	970,332
Vol & Num	S_Volume	2,645,277
Holding data	S_Volume_holdings	3,856,896

Vol & Num : Volume and Number

Table 5. Summary of data records.

https://doi.org/10.23106/dcmi.952106693

An overview of available data in Union Serial System is summarized in table 5. As mentioned in section 2.1, we defined 46 qualifiers. Only 28% of data was filled up according to table 5.

Figure 4.Client PG interface of 'S_Abbreviation' and 'S_DC_Master' table.

Figure 5.Client PG interface of volume and holding data.

3. Conclusions

The Union Serial System is the first running system adopting DC Metadata in expressing academic journal and articles in Korea. This system is accessible on the Web[4] as a part of a digital library so called Research Information Service System(RISS)[5].

A full discussion and evaluation on this system is needed to proceed in various aspects in the near future. One of the obstacles in doing this project is that there is not yet a defined Cataloging standard about serial and volume number, and article index. Developing union serial catalog system requires first standards and actual acceptance of such format in each library. The quality of data should improve more and a full discussion and debate on this matter is necessary along with continuous refinement process on data for better full-fledged service.

References

[1] Stuart Weibel, (1999) "The State of the Dublin Core Metadata Initiative," D-Lib Magazine, Volume 5 Number 4 ISSN 1082-9873 on http://www.dlib.org/dlib/april99/04weibel.html

[2] Stuart Weibel and Eric Miller, (2000) Dublin Core Metadata, on http://purl.org/metadata /dublin_core

[3] James Powell, Edward A. Fox, (1998) Multilingual Federated Searching Across Heterogeneous Collections, D-Lib Magazine

[4] http://riss4u.net/int_search/jlistkdam_fm1_p.html.

[5] Yong Soon KIM, Jung Sun Oh, Kim Min Ki, Digital Library of Theses and Dissertations; improving Graduate Education and University Information Infrastructure, The 6th Joint Conference on Educational technology, Vol. 2, pp. 875~876, 2000, Japan

https://doi.org/10.23106/dcmi.952106693