

Capturing Research Output in the Field of Anthropology: Metadata Design and Lessons Learned

Sittisak Rungcharoensuksri
Office of Academic Affairs and Information The
Princess Maha Chakri Sirindhorn Anthropology
Centre, Thailand
sittisak.r@sac.ac.th

Wachiraporn Klungthanaboon
Department of Library Science
Faculty of Arts, Chulalongkorn University,
Thailand
Wachiraporn.K@chula.ac.th

Abstract

To advocate open science and knowledge development, the Princess Maha Chakri Sirindhorn Anthropology Centre (SAC) recognizes the significance of collocation of scattered research outputs funded by the SAC for the public use. The SAC Research Database (<http://www.sac.or.th/databases/sac-research/index.php>) was developed and launched in March 2019 to provide free access to digital full-text research outputs under the creative common license (CC-BY-NC-ND 3.0). The ease of use and interoperability are taken into consideration when selecting the metadata scheme. The Dublin Core Metadata Element Set was chosen with some modified elements for the SAC Research Database. This paper presents the lesson learned from the development of this database.

Keywords: research output; metadata; digital curation; digital collection

1. Introduction

This paper describes the development of the SAC Research Database mainly emphasizing on the metadata model and some lessons learned. In section 2, we present the background of the SAC, its roles and the reasons why this database was developed. Section 3 shows the process of organizing digital research output focusing on metadata design. Lastly, the last section describes the challenges and lessons learned from the project and the conclusion.

2. Background

2.1. The Princess Maha Chakri Sirindhorn Anthropology Centre (SAC) and its Roles

The SAC, a leading research centre in Thailand in the disciplines of anthropology, archaeology-history, ethnology and socio-cultural studies, realizes the significance of information at hand for further research and development. As a result, after the establishment of the SAC in 1991, the centre has been creating academic contents and supporting grants to their academic networks for nearly 30 years. Furthermore, the SAC has been disseminating their knowledge through a series of searchable digital databases, such as, Ethnic Identity Research Database, Local Museums Database, Anthropological Archives Database, Potteries in Thailand Database, Inscriptions in Thailand Database, etc. Currently, there are 27 databases in the SAC digital repository which can be accessed via <https://www.sac.or.th/main/th/database>.

2.2. The SAC Research Database

The SAC has been carrying out several research projects and providing research grants to researchers in anthropology and related fields for more than 20 years. At the end of the SAC-funded research projects, it is required that the hard copies and soft files of research reports must be submitted to the Centre. So ideally all SAC-funded research reports should be acquired and available at the SAC library. However, prior to the development of the database, some research reports were kept by those responsible for the projects. In some cases, some hard copies are available at the SAC library whereas some digital files could not be found. Additionally, most research reports lack adequate information about their creation. This data loss affects the metadata enrichment, discoverability, and other data-driven services. The collection of SAC-funded research outputs would have been comprehensive and available if clear policies on managing research

reports had been developed and introduced to the SAC staff. However, it is never too late to initiate the development of the SAC Research Database.

In the mid of 2018, the SAC started the SAC Research Database project. This project aims to collect, manage, and disseminate the SAC publications and research outputs for free online access. This would contribute to the advancement of education and research in anthropology and related disciplines. The database covers the SAC publications and research outputs published between 1991 to 2018. So far, the working group has found 482 research titles from all the SAC projects. However, due to the unclear copyrights status, research title overlaps, and the disappearance of some printed versions of research or digital files, only 192 research titles were chosen as the source for database development (Rungcharoensuksri, Process, 2018).

3. Metadata Design

According to the minute of SAC Research Database working group (2018), three requirements are taken into consideration for metadata design of this database. Firstly, the data structure and data content of this database should rely on an international standard metadata schema. Secondly, the content of this database should be searchable through the single search system of the SAC digital repository. Thirdly, due to limited background knowledge of the SAC researchers, research assistants, and administrators who will be responsible for this database, the metadata description and the usability of this database should be easy for them.

To comply with the above-mentioned requirements, the working group was chosen the "Dublin Core Metadata Element Set" (DCMI, 2005) as the standard data structure for the SAC Research Database. There are two reasons why this standard was selected. As for the first reason, due to different educational backgrounds of online users and the SAC staff, many of them are more familiar with Dublin Core Metadata Element Set than other metadata sets because this metadata has especially designed for social science and humanity contents. Furthermore, some elements of this metadata set are flexible and appropriate for the SAC's administrative management. For instance, the academic users are likely to focus on description, subject, and coverage in research outputs, while the SAC staffs tend to use contents and intellectual property metadata for the planning and administration of their projects. As for the second reason, due to the variety of academic content of the SAC digital repository, the project managers (researchers) of each database have been allowed to use a standard metadata set which is particularly suitable for their content¹ (Rungcharoensuksri, Guideline for SAC Digital Curation, 2016). However, all the contents from 27 databases can be searched through the single search system of the SAC digital repository, because each database has created the metadata crosswalk to search system via the Dublin Core Metadata Element Set.

Consequently, to comply with the user's requirements and standard metadata policies of the SAC, the Dublin Core Metadata Element Set has been chosen in order to design the database. However, during the work process, it is found that even though the core elements of this metadata set may be suitable for the overview description of the SAC research outputs, more details of the resources should be explained. For example, the locations of the field site should be identified to inform researchers where the research data have been collected, and to present those field sites on the mapping search system. Furthermore, some element encoding schemes of the Dublin Core Metadata Element Set are too complicated for research assistants and database administrators.

Therefore, the working group has work towards these three solutions. Firstly, the "SAC qualifier" is designed as specific description for the SAC research outputs. According to data structure design (Rungcharoensuksri, Data Structure, 2019), this database has been designed in imitation of the Dublin Core Metadata Element Set, and some elements of the Dublin Core have been adapted and extended the namespace "sac" taken from the name of the Centre for the SAC qualifier. For example, "LC Call No." <sac:identifier.lc> identifies the Library of Congress Call No. of the researches which are kept in the SAC library, Contract Form No. <sac:right.contractFormNo> identified the contract form number of the funding research contracts which are kept by the

¹ Currently, there are seven types of data structure in the SAC digital repository, namely, Home Made Metadata (based on academic but non-international standard content), Dublin Core Metadata Element Set, International Standard Archival Description (ISAD(G)), Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF), Categories for the Description of Works of Art (CDWA), and MARC21.

procurement staffs, and “DMS Location” <sac:coverage.spatial.dmsLocation> identifies the places where the researchers did the fieldwork. Secondly, to make it easier for the administrators to fill in the content and make it easier for the automated citation system², some elements are separated or created as the choice type. For instance, for the “Type” element, we have created it as the choice type for administrators to choose books, journals, articles, dissertations, researches, reports, others. For the “Creator” element, we have separated it into three elements, namely, First Name of Creator <sac:creator.firstName>, Middle Name of Creator <sac:creator.middleName>, and Last Name of Creator <sac:creator.lastName>. Lastly, to develop a clear understanding between the working groups during the work process, we have decided to compile a “data dictionary” of this database. This document is a communication tool between the researchers who have designed data structure of the SAC Research Database at the first place and the programmers who create the database in the later phase. Furthermore, the database administrators and research assistants also use this document as the guideline to fill in the content in the backend of the database.

TABLE 1: Metadata element set of the SAC Research Database

	Labels	DC Elements	SAC qualifier	Type
1. Content	1.1 Title			
	Title	<dc:title>	-	Text
	Alternative Title	<dc:title.alternative>	-	Text
	1.2 Description			
	Abstract	<dc:description.abstract>	-	Text
	Table of Contents	<dc:description.tableOfContents>	-	Text
	Provenance	<dc:description.provenance>	-	Text
	Statement of Responsibility	<dc:description.statementOfResponsibility>	-	Text
	Budget	-	<sac:description.budget>	Number
	1.3 Subject			
	SAC Thesaurus	-	<sac:subject.sacThesaurus>	Choice type
	Free Term Subject	-	<sac:subject.freeTerm>	Text
	1.4 Relation			
	Relation	<dc:relation>	-	Text
	1.5 Coverage			
	Spatial	<dc:coverage.spatial>	-	Text
	DMS Location	-	<sac:coverage.spatial.dmsLocation>	Text
	DD Location	-	<sac:coverage.spatial.ddLocation>	Text
	Temporal	<dc:coverage.temporal>	-	Text
2	2.1 Identifier			
	LC Call No.	-	<sac:identifier.lc>	Text

² To facilitate and promote online users to refer to SAC research outputs as the source for their work. The working group decided to create the prototype of automated citation system for SAC Research Database. This system is designed by mapping and rearranging Dublin Core Metadata Element Set to bibliography system, namely, APA, MLA, and Harvard. For example, the “Creator” element is mapped to “Author” in APA, MLA, and Harvard system. For more details, go to <https://www.sac.or.th/databases/sac-research/citation.php> (In Thai)

1	Identifier	<dc:identifier>	-	Date: Time stamp
	2.2 Type			
	Type	<dc:type>	-	Choice type
	2.3 Publisher			
	Publication Place	-	<sac:publisher.publicati onPlace>	Text
	Publisher	-	<sac:publisher>	Text
	Publication Year	-	<sac:publisher.publicati onYear>	Date: YYYY
	Publication Month	-	<sac:publisher.publicati onMonth	Text
	Edition Statement	-	<sac:publisher.edition>	Number
	Volume	-	<sac:publisher.volume>	Number
	Issue	-	<sac:publisher.issue>	Number
	Page	-	<sac:publisher.page>	Number
	2.4 Format			
	Total Pages		<sac:format.totalPages >	Number
	Medium	<dc:format.medium>	-	Choice type
	Extent	<dc:format.extent>	-	Text
	2.5 Language			
	Language	<dc:language>	-	Choice type
	Original Language	-	<sac:language.original>	Choice type
	2.6 Date			
	Description Created Date	-	<sac:date.descriptionCr eated>	Date: YYYY- MM-DD
	Submitted Date	<dc:date.submitted>	-	Date: YYYY- MM-DD
	Issued Date	<dc:date.issued>	-	Date: YYYY- MM-DD
	Modified Date	<dc:date.modified>	-	Date: YYYY- MM-DD
	Maintenance Notes	-	<sac:date.maitenanceN otes>	Date: YYYY- MM-DD
2	3.1 Creator			
	First Name of Creator	-	<sac:creator.firstName>	Text
	Middle Name of Creator	-	<sac:creator.middleNa me>	Text
	Last Name of Creator	-	<sac:creator.lastName>	Text
	3.2 Contributor			
3	First Name of Contributor	-	<sac:contributor.firstNa me>	Text
	Middle Name of Contributor	-	<sac:contributor.middle Name>	Text

Last Name of Contributor	-	<sac:contributor.lastName>	Text
Role of Contributor	-	<sac:contributor.role>	Text
3.3 Rights			
Rights	<dc:rights>	-	Text
Type of License	-	<sac:rights.license>	Text
Contract Form No.	-	<sac:right.contractFormNo>	Text
Digital File of Contract Form	-	<sac:right.digitalFileContractForm>	Text
3.4 Administrator			
Cataloger Name	-	<sac:contributor.catalogerName>	Text
Digital Editor	-	<sac:contributor.digitalEditor>	Text
Database Administrator	-	<sac:contributor.databaseAdm>	Text

ประเพณีสิบสองเดือน : พิธีกรรมที่เปลี่ยนไป

นักวิจัย : ศรัศกร วัลลิโภดม, ปี : 2546

คำอธิบาย / บทคัดย่อ

การศึกษาประเพณีสิบสองเดือน ได้รับรวมหลักฐานและรายละเอียดเรื่องราวเกี่ยวกับประเพณีในแต่ละภูมิภาคว่ามีลักษณะอย่างไร มีความเหมือนหรือต่างกันอย่างไร และมีความหมายอย่างไรต่อการดำรงอยู่ของคนในท้องถิ่น รวมทั้งศึกษาบริบทแวดล้อมทั้งทางสังคม เศรษฐกิจ การเมืองและวัฒนธรรมที่มีการเปลี่ยนแปลงจากเดิมหรือเพิ่มเติมการปรุงแต่งขึ้นใหม่

ประเภท	รายงานงานวิจัย
สถานที่พิมพ์	กรุงเทพฯ
สำนักพิมพ์	ศูนย์มานุษยวิทยาสิรินธร
สิทธิ์	ศูนย์มานุษยวิทยาสิรินธร (องค์การมหาชน)
รูปแบบลิขสิทธิ์	Attribution-NonCommercial (CC BY-NC)

+ Additional Details

ระบบอ้างอิงอัตโนมัติฐานข้อมูลผลงานวิชาการ ศมส.

+ APA

ศรัศกร วัลลิโภดม. (2546). ประเพณีสิบสองเดือน : พิธีกรรมที่เปลี่ยนไป. กรุงเทพฯ: ศูนย์มานุษยวิทยาสิรินธร. ค้นคืนเมื่อ 17 พฤษภาคม 2562. จาก ฐานข้อมูลงานวิจัย ศมส.: http://www.sac.or.th/databases/sac-research/research-item.php?ob_id=14

+ MLA

+ HARVARD

File Download

File Name	Size	Download
ปก	332.56 KB	Download
สารบัญ, คำนำ	1.09 MB	Download

FIG. 1. Display page of the SAC Research Database (In Thai)
(http://www.sac.or.th/databases/sac-research/research-item.php?ob_id=14)

4. Challenges and Lessons Learned

The SAC has attempted to collect, organize, and disseminate information on anthropology, archaeology-history, ethnology, socio-cultural studies, and other related fields to the public for better understanding among people in the society. The SAC Research Database aims to make our funded research outputs discoverable and accessible 24/7. However, we faced difficulties to collocate digital files of research reports due to disparate sources and poor digital curation process. Developing the SAC Research Database asked us to rethink about digital curation and information management of the centre. It would be better if we had a clear policy in place at the beginning of the research funding process to collect and manage our research outputs for online access. This enhances the possibility to receive information context for enriched metadata. Working in the midst of unorganized and scattered research outputs provides us with challenges on metadata description which is important for enhancing the discoverability and accessibility of the collection.

Regarding the quality of metadata, the working group described digital research outputs difficultly due to the scattered research outputs and insufficient metadata. Therefore, in some cases, metadata is entirely absent. From this problem, we drafted a guideline for managing the SAC-funded research outputs for the SAC staffs. Besides, we asked for the policy support from the SAC administrative board. Additionally, we are discussing whether it should be a mandatory requirement for the researchers to embed metadata when submitting digital files of their research reports. This will enhance the collection development of the SAC Research Database effectively and the quality of metadata. Significant metadata categories for digital collections are descriptive metadata, structural metadata, rights management metadata, and technical and administrative metadata.

Next challenge is metadata mapping. The SAC has developed and maintained 27 databases which have various metadata schemes. However, we provide an in-house single search tool "SAC Portal" to facilitate the users to do a search across all databases. Therefore, it is important to have a robust metadata structure supporting interoperability and searchability across the SAC-owned databases. In designing and creating a metadata schema for the SAC Research Database project, we consider the interoperability and flexibility of metadata standard as the most important criteria. Moreover, it should be easy to use for non-library background users. Consequently, the Dublin Core Metadata Element Set was selected. However, we modified metadata elements and qualifiers to meet the characteristics of research outputs and researchers' information behavior in anthropology and related fields. We anticipate various types of research outputs and research data and other value-added research support services built upon the quality of metadata. Therefore, we have added some qualifiers to support their search.

In short, the quality and standardization of metadata influence the discoverability, accessibility, and utilization of digital information. However, challenges associated with metadata have been usually apparent to information professionals. The SAC also faces the same challenges as other institutions. This probably takes time and effort to deal with challenges. However, it is worth to invest our time, resources, and effort. Looking at the issue of the administration, metadata will enhance us to create new data-driven reports showing the visibility of the SAC-funded research outputs and the research strength of the centre.

Acknowledgements

This project may not be successful without the assistance of the SAC Research Database working group, namely, Sutasinee Boonkerd (Assistant researcher and Database administrator), Tanapha Eungthanapha (Programmer), Rattanaporn Chiangkham (Programmer), Wisoot Singhso (Programmer), and Kanyanart Siripanya (Coordinator and English prove reader)

References

- Alam, A. W. (2014). *Dublin Core Metadata for Research Data - Lessons Learned in a Real-World Scenario with datorium*. Retrieved August 2019, from DCMI International Conference on Dublin Core and Metadata Applications:
<https://dcpapers.dublincore.org/pubs/article/view/3703>

Database working group. (2018). *Minutes of Working Group of SAC's Research Database meeting 15 August 2018*. BKK: Princess Maha Chakri Sirindhorn Anthropology Centre. (InThai)

DCMI. (2005, November 7). *Using Dublin Core - Dublin Core Qualifiers*. Retrieved August 8, 2018, from Dublin Core Metadata Initiative:
<http://www.dublincore.org/specifications/dublin-core/usageguide/qualifiers/>

Rungcharoensuksri, S. (2016). *Guideline for SAC Digital Curation*. BKK: Princess Maha Chakri Sirindhorn Anthropology Centre. (InThai)

Rungcharoensuksri, S. (2018, December). *Process*. Retrieved May 15, 2019, from SAC's Research Database: <http://www.sac.or.th/databases/sac-research/process.php> (InThai)

Rungcharoensuksri, S. (2019, January). *Data Structure*. Retrieved May 15, 2019, from SAC's Research Database: <http://www.sac.or.th/databases/sac-research/structure.php> (InThai)