

Metadata Framework for Resource Discovery of Agricultural Information

ECDL 2001 OAI Workshop

to be held in conjunction with:

**5th European Conference on Research and Advanced Technology
for Digital Libraries September 4-9 2001, Darmstadt, Germany**

Irene Onyancha, Fynvola Le Hunte Ward, Frehiwot Fisseha, Kafkas Caprazli, Stefano
Anibaldi, Keizer Johannes, Steve Katz
**Food and Agriculture Organization (FAO) of the United Nations
Library & Documentation Systems Division (GIL)
AGRIS/CARIS & Documentation Group**

1.0 Introduction

Network technologies have helped to lower many of the geographical barriers that impede access to information resources, but other obstacles have appeared in their place: one is the heterogeneous use of resource descriptions; another, more serious, is the lack of resource description at all.

Resource description varies depending on the structure, type and content of resources; it also varies with the interests of the information keepers responsible for the management of these resources. A further consideration in resource discovery is the cross-domain information needs of users who require access to information about relevant resources irrespective of where they are located, how they have been stored or by whom. With the current enabling technology, the more complex needs of users nowadays can be met: querying more than one domain-specific information system in parallel while information managers seek to have a system that enables access to separately managed collections in-house. Example of initiatives that have been developed to encourage timely dissemination of scholarly information is the Open Archive Initiative (OAI)

To meet such information demands, a framework needs to be developed that would allow users to access information regardless of the above-mentioned barriers while giving the managers better control of information management and preservation. One main step forward in the development of such a framework, is the creation of a low barrier metadata format that allows for interoperability between cross-domain information systems. The Dublin Core initiative is a potential example of such a format; the initiative has many positive characteristics that distinguish it as a prime candidate for resource description for the primary goal of electronic resource discovery.

The Open Archives Initiative (OAI) develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The OAI approach of interoperability attempts to combine the best of library and Internet techniques into a new model of accessing resources. It has adopted a low-barrier interoperability solution known as *metadata harvesting*, which allows content providers to expose their metadata via an open interface. The open interface prescribes to *Unqualified Dublin Core Metadata set*.

This report outlines a proposed metadata framework for resource discovery of agricultural resources, and in particular to describe documents and document-like resources in agricultural sciences. The project lies within the framework of a wider and more comprehensive project proposal that promotes a metadata set of core elements and qualifiers that are generic to the description of agricultural resources, mainly project and document resources. The overall work is the result of a collaborative effort between a number of partners in the agricultural community and the World Agricultural Information Centre of FAO. The endeavour is formally referred to as the “Agricultural Metadata Standards Initiative”. It is based upon the elements and qualifiers of the Dublin Core Metadata Initiative (DCMI).

The report first provides the overall context for the metadata framework; why the standard is needed; how the work was done, and then offers thoughts on the way forward from here. Annex 1 provides the elements and qualifiers of the proposed standard presented in a hierarchical structure. The hierarchical structure offers a flexible framework to implement the proposed standard at different levels of granularity, depending on the how rich each metadata source are. In its simplest form, metadata can even be supplied at the most general level of 13 core fields.

2.0 Objectives of the Agricultural Metadata Framework

The overall objective of the agricultural metadata framework is to define a low-barrier interoperability layer using emerging standards that aim to facilitate the efficient dissemination of agricultural content. This document presents an application profile for description of document and document-like agricultural resources. “

The specific goals are to:

1. To define a low-barrier interoperability layer to aid primary resource discovery. This is to support the emerging initiatives that promote “open” exchange of information and in particular the Open Archive Initiative (OAI).
2. To define a richer interoperability layer to aid secondary and tertiary resource discovery.
3. To assist the management of resources by the owners;
4. To enable interoperability between different metadata structures through a simple common format;
5. To develop a metadata framework that is compliant with other standards such as MARC, ISBD and new emerging ones like Dublin Core;
6. Describe the metadata set in XML and RDF. By doing so we aim:
 - to provide better search engine capabilities which will assist retrieval,
 - to allow interoperability between applications
 - to support automated processing of web resources.

3.0 Strategy and Methodology Adopted

With respect to the strategy and methodology adopted to formulate the metadata framework, specific actions were taken to:

- Develop a conceptual map of the different types of information resources used in agriculture.
- Evaluate standards and common resource description practices currently used in the agricultural domain.
- Initially focus on the description of project, document and document-like (bibliographic) resources.
- Identify the pool of elements and qualifiers that apply to project and document/document-like resources, in conformance with the guidelines of the Dublin Core Metadata Initiative.
- Develop a specific Application Profile for description of document and document-like resources.
- Document a full element description for document and document-like resources using the set of attributes recommended by the Dublin Core Metadata Initiative.

As a result of adopting this strategy and methodology in a participatory manner with all partners of the Agricultural Metadata Standards Initiative, a list of 13 elements for agricultural resources description was proposed.

4.0 Implementation aspects compared to generic Dublin Core

Suggestions and comments were received from all partners of the Agricultural Metadata Standards Initiative, as well as from Stuart Weibel, Executive Director of the Dublin Core Metadata Initiative. These led to the following implementation decisions with respect to the generic specification of Dublin Core:

- To merge the DC elements *Creator*, *Contributor* and *Publisher* to one main element called *Creator*;
- To drop the element *Source*, but elaborate the element *Relation* to include information about the source;
- To propose a new element called **TARGET AUDIENCE**;
- A number of new qualifiers and attributes were introduced that are vital to the description and discovery of information in the agricultural domain. They are all described in detail in Annex 1
- In addition to these implementation aspects with respect to generic Dublin Core, *Authority files* were created for elements and qualifiers that have secondary information that is not included in the metadata description of a resource but is relevant for resource discovery.

5.0 Future developments

As mentioned earlier, this paper only represents the first step in the development of tools to aid resource discovery in the agricultural domain. The initiative will be posted and advertised in agricultural forums so as to impact the targeted audience. Work is still in progress and the logical frameworks that have been developed are in the process of being converted into technical frameworks. The proposal will also be presented to the intergovernmental process of FAO for possible endorsement by member countries.

Some of the immediate future developments are as follows:

- To assess the suitable technical framework for describing, storing and processing the metadata. RDF being the foundation for processing metadata, has been considered as a data model to describe, store and process the metadata. The motivation to use RDF for modelling the data lies on the following features of RDF namely:
 1. RDF uses XML which is the future of electronic data processing and information exchange
 2. RDF provides a machine-understandable system for defining schemas for descriptive vocabularies
 3. RDF is a flexible and easily extensible metadata description and data modelling system
 4. RDF is the W3C recommendation

- To initiate a pilot project between FAO and a number of important and successful agricultural gateway services. The project aims to provide a single access point with multi-host searching using the Agricultural Application Profile as the standard for linking common metadata across the different gateway services;
- To develop software tools in support of the proposed standard (e.g. for import, export, validation, query purposes, etc.);
- To register the metadata framework and specific application profile with authoritative metadata registries.
- To develop guidelines for the application profile to assist implementers and users.
- To monitor the impact of the proposed metadata application profile for agricultural resources, making any changes or enhancements based on the results of the impact study, and undertaking outreach work to promote and facilitate the rational and widespread use of metadata.
- Present the project for further discussions at the DC-2001 Conference in Tokyo, Japan in October 22-26, 2001.
- Project to be discussed in a workshop to be held in November with FAO Partners, Members from different agricultural institutions and collaborative Governments.

6.0 Benefits of the application profile to FAO and the agricultural community

Format for describing and maintaining FAO in-house databases

A crosswalk of the Dublin Core, AGRIS (International Information System for the Agricultural Sciences and Technology), CARIS (Current Agricultural Research Information System) and other document repositories at FAO was developed. The crosswalk consists of the proposed core elements as container elements, while the sub-elements that qualify a specific core element are layered under the hierarchy. This mapping gives homogeneity to the different application profile under one set of defined core elements. This provides a working example of how a low level format, enables interoperability between different information systems to allow resources discovery.

Format for a unified interface for searching heterogeneous archives:- The AGRIS MHS

AGRIS Multi-Host Server is a search engine that allows parallel searching across distributed databases that are heterogeneous and have different data structure and metadata information. The search engine is being developed in corporation with ZADI (Zentralstelle für Agrardokumentation) in Germany. It searches distributed bibliographical databases giving a one stop access to them without the need of centralising data. The proposed application profile gives common metadata elements that homogenise search set results.

Format for resources discovery through agricultural subject gateways and information providers

Subject gateways are online services and sites that provide searchable and browseable catalogues of Internet based resources. Subject gateways will typically focus on a related set of academic subject areas. They generally consist of databases of detailed metadata or catalogued records.

Some Examples of agricultural subject gateways such as NOVAGATE, BIOME, AGRIGATE, AGNIC have the following benefits;

- Participation in a global network to bring agricultural and related information to the Web
- Offer users the opportunity to interact and resource share with other national and international agricultural institutions
- Offer opportunity to provide value-added services to constituencies

The proposed metadata frameworks offer a uniform format that could be used as a means of interoperability between these gateways. The framework offers opportunity for both low level and detailed description according to the users needs. The different levels of description indicated in the metadata framework express these.

7.0 References

1. Dublin Core Qualifiers. <http://www.dublincore.org/documents/2000/07/11/dcmes-qualifiers/>
2. Dublin Core Metadata Element Set, Version 1.1, Reference description. <http://www.dublincore.org/documents/1999/07/02/dces/>
3. DMCI Type vocabulary. <http://dublincore.org/documents/dcmi-type-vocabulary/>
4. Food and Agricultural Organization (FAO) Cataloguing and Indexing Manual.
5. Resource Description Framework (RDF) Model and Syntax Specification. <http://www.w3.org/TR/REC-rdf-syntax/>
6. Metadata: standards for Retrieving WWW Documents (and Other Digitized and Non-Digitized Resources). <http://www.stsci.edu/stsci/meetings/lisa3/ruschfejad.html>
7. Interoperability Metadata Standard for Electronic Thesis and Dissertations- Version .03. <http://ndltd.org/standards/metadata/current.html>
8. Using Dublin Core. <http://dublincore.org/documents/2001/04/12/usageguide/>
9. Discovering Online Resources across the Humanities: A Practical Implementation of the Dublin Core. <http://ahds.ac.uk/public/metadata/discovery.html>
10. Dublin Core Metadata and the Cataloguing Rules. American Library Association, Committee on Cataloguing: Task Force on Metadata and the Cataloguing Rules (Final Report). <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei5.html>.
The open Archives Initiative: Building a low-barrier interoperability framework. <http://www.openarchives.org/documents/oai.pdf>

Annex 1

Presentation of a metadata set for the description of agricultural documents and document-like resources

The paper presents qualifiers of the proposed Dublin Core elements for documents and document-like information resources. Preference is given to notation, vocabularies and terms that are currently used in describing agricultural resources.

As indicated below, agricultural information resource can be described from broad to very specific levels. The top level in the hierarchy defines the Dublin Core elements and additional proposed elements while the qualifiers suitable for agricultural resource description are shown down the hierarchy. A more detailed description of all the elements and qualifiers, including information on definitions, rules, and data typing has been done and is available.

Important notice:

1. The element **Creator** has been revised to represent all the agent elements namely, ***Creator, Contributor and Publisher***.
2. Some attributes of elements that have been in the past considered necessary in resource description are not included in this description of a specific resource because this information is currently not considered as primary information that is important for discovery of a particular resource. However to include this information which is also important for resource discovery at a secondary level, ***Authority files*** will be created and linked to the metadata. Element that will have Authority files include: ***Author, Corporate Author, Publisher and Type***, qualifier ***Event*** (Conferences, Workshops, Meetings)

The hierarchical notation presents the different levels of description that is noted by the use of different formats and colours as indicated in the legend below

Legend

Bold:	DC Elements & Proposed elements for agricultural resources
<u>Underlined:</u>	Qualifiers for agricultural resource description
<i>Italic:</i>	Sub-elements or attributes of the qualifiers
Blue:	Attributes of sub-elements

The proposed levels of description are:

- DC & Proposed elements for agricultural resources
- Qualifiers
 1. Element refinement
 2. Encoding schemes
 - Sub elements of qualifiers

Recommended DC qualifiers are noted with prefix (DC) while the new qualifiers are noted with (new)

Proposed core elements and the qualifiers for agricultural resources description

1. Core element: Creator (new)

Qualifiers for Agricultural Resources

- (DC) Corporate author
Name
- (DC) Personal author
Name
- (DC) Publisher
Name
- (DC) Editor
Name
- (DC) Compiler
Name

2. Core element: Rights

Qualifiers for Agricultural Resources

- (DC) Statement
- (new) Terms of use
- (new) Patent

3. Core element: Title

Qualifiers for Agricultural Resources

- (DC) Main title
- (new) Title supplement
- (DC) Alternative title

4. Core element: Relation

Qualifiers for Agricultural Resources

- (DC) Is Part Of
 - part of monograph (AM)*
(Unique identifier of related record. If URI should be PURL)
 - part of monographic series (AMS)*
(Unique identifier of related record. If URI should be PURL)
 - part of series (AS)*
(Unique identifier of related record. If URI should be PURL)

- (DC) Has Part (*analytical*)
(Unique identifier of related record. If URI should be PURL)
- (DC) Edition
Version Of Monograph,
(Unique identifier of other version)
URI
(Unique identifier if URI should be PURL)
- (DC) References
(Unique identifiers of related record. If URI should be PURL)
(Unique identifiers of related language versions)
URI
(If URI should be PURL)

5. Core element: Subject

Qualifiers for Agricultural Resources

- (DC) Subject classification
AGRIS subject Categories
CABI Codes
LCSH
DDC
LCC
UDC
- (DC) Thesaurus
CABI Thesaurus
AGROVOC Thesaurus
NAL Thesaurus
ASFA Thesaurus

6. Core element: Coverage

Qualifiers for Agricultural Resources

- (DC) Thesaurus
AGROVOC Thesaurus
GTN
- (DC) Country codes
ISO 3166 country codes

7. Core element: Description

Qualifiers for Agricultural Resources

- (DC) Abstract
(DC) Table of contents
(new) Notes

(new) Holdings
Location
Address
Online
URI (PURL)

8. Core element: Date

Qualifiers for Agricultural Resources

(DC) Publication date
(DC) Date of Creation
(DC) Date of Modification

9. Core element: Identifier

Qualifiers for Agricultural Resources

URI
ISBN
ISSN
Report no.
Accession no.
Call no.
Patent no.
Job no.

10. Core Element: Language

Qualifiers for Agricultural Resources

(DC) Language of text
ISO 639 language codes

11. Core element: Type

Qualifiers for Agricultural Resources

(DC) Collection
Monographs
Serials
Monographic serials
Web Pages
Analyticals

(new) Document class
Publication
Non conventional

(new) Literary indicator

Grey Literature

Legislation

Standard

Bibliography

Summary

Statistical data

Directory

Thesis

(DC) Event

Conference

Training course

Workshops

Seminars

Consultations

(DC) Images

Photograph

Film

Picture

Map

Slide

Microfiche

Video Cassettes

(DC) Sound

Audio cassettes

Digital Audio

(DC) Interactive Resources

CD-ROM

Web Forms

(DC) Software

(DC) Dataset

(DC) Text

12. Core element: Format

Qualifiers for Agricultural Resources

(DC) Medium

Internet Media Type, (IMT)

[Application](#)

[Audio](#)

[Image](#)

[Text](#)

[Video](#)

Print

Electronic media

CD-ROM

Diskette

DVD

(DC) Extent
Collation
Pagination
Duration

13. Core element: Target Audience (new)

Policy makers
Planners
Researchers
Research Institutions
Educational institutions
Students
Information intermediaries
Media

Note:

The new proposed core element, Target audience will be further developed to have a standardised list. It is currently under review.