

Content Management Task Force Progress Report

2nd Expert Consultation

**International Information Systems
for Agricultural Science and Technology**
23-24 September 2007
FAO Headquarters, Rome



International Information Systems in Agricultural Science and Technology

Content Management Task Force

Progress Report: September 2007

1. TASK FORCE MEMBERSHIP

The Content Management Task Force (hereafter referred to as CMTF) members comprise representatives of key organizations participating in the Expert Consultation¹ in October 2005, and others co-opted from major constituencies in the initiative that were not represented ([Annex 1](#)).

2. ACTIVITIES CARRIED OUT SINCE OCTOBER 2005

The CMTF was established to address priority actions identified by the 2005 Expert Consultation in the area of 'content management'. Terms of reference² (TOR) were later developed for the CMTF covering issues which it was felt the group should address, as well as the mode of operation and the timeline for each of these activities. These TORs identified three main areas of work:

1. Open Access document repositories in the field of agriculture (in a broad sense);
2. subject-specific vocabularies; and
3. exchange of newsfeeds.

The rest of this section provides a summary of all the activities undertaken by the CMTF members since the first Consultation.

2.1 Open Access document repositories in the field of agriculture

2.1.1 Implementation of the AGRIS Application Profile (AGRIS AP)

The AGRIS application Profile was developed to exchange metadata on "Document-like Information Objects" (DLIOs). These include books, journal articles, thesis, and refer to any resource which can be described like a document. The AGRIS AP has been created specifically to improve the quality and interoperability of metadata in this domain. Currently, the AGRIS AP is used by all centres in the AGRIS Network. The standard has also been adopted, for exchange of DLIO metadata, by:

- Global Forest Information Service (GFIS)
- NARIMS project in Egypt
- Kenya Agricultural Information Network (KAINET) in Kenya
- AGRORED project in Peru.

The AGRIS AP has been implemented as output formats in various software products such as

- WebAGRIS system, widely used in the AGRIS network
- NewGenLib, an integrated Library Management System
- NARIMS System, currently used for the NARIMS partners
- FAO Electronic Information Management Systems.

¹ International Information Systems for Agricultural Science and Technology – Review of Progress and Prospects

http://www.fao.org/gi/gil/consultations/consult_infosys_en.asp

² Terms of Reference of the Agriscontent TF were published in July 2006 after a first meeting in Nairobi in May 2006. The TOR are accessible at <http://agriscontent.pbwiki.com>

Efforts are on way to make output from DSpace compliant with the AGRIS AP.

AGRIS AP: <http://www.fao.org/docrep/008/ae909e/ae909e00.htm>.

AGRIS AP XML Guidelines:
<http://www.fao.org/docrep/008/ae908e/ae908e00.htm>.

2.1.2 Open archives architecture paper

A paper has been authored by FAO about a possible architecture of networks of open archives in Agricultural Science and Technology. This paper has been discussed in the Content Management Taskforce and an earlier version of this paper has been presented at the International Conference for Digital Libraries and the Semantic Web in Bangalore, February 2007.

Conference paper: <ftp://ftp.fao.org/docrep/fao/009/ah766e/ah766e00.pdf>

2.1.3 Development of Search Engine

The bibliographic references (about 2.5 Million) in the AGRIS database, hosted by FAO, have been converted into XML format compliant with the AGRIS AP. This conversion coupled with advances in search technologies have given the possibility to provide improved search services to the AGRIS community. In the last two years, gradually improving versions of the AGRIS Search Engine have been released:

- v1.0 (June 2006): The open source Lucene technology has been chosen to power the new search engine application. The challenge has been to query the entire AGRIS XML repository, exploiting the great performance of the Lucene full-text search engine technology on structured content encoded in XML tags. On the front end, a "Google-like" simple search has been introduced. On top of that a user-friendly "Search Assistant" was put in place where the user could search in specific metadata fields represented in the structure of the AGRIS AP XML records.
- v1.3 (July 2007): This release has mainly addressed bug-fixes. Two main additions have rendered the search more user-friendly: the "Search Assistant" has been enhanced with some new functionality and a "Preferences" section has been introduced to allow the user to define a customized profile by selecting, for example, the preferred "type of publication" or the "language of the resource".
- v1.5 (under development): The AGRIS AP promotes the use of AGROVOC or other agricultural thesauri and classification schemes for describing the subject of a DLIO. In this version, the semantics inherent in the subject keywords shall be further exploited, using the relationships of AGROVOC keywords in order to provide semantic search query analysis and help the user retrieve the best possible results for the search query.

Search engine: <http://www.fao.org/agris/search/search.do>.

2.1.4 OAI compliance of WebAgris

WebAGRIS is a Web based CDS/ISIS application used by many AGRIS Centres to store and retrieve bibliographical metadata and associated links from metadata to the full text of the publications online. It is a tool which ensures high quality content, based on usage of built-in authority files (AGROVOC, AGRIS Subject categories, etc.), cross-checking for accuracy through validations and common standards (DC, AGRIS AP compliant metadata exported in many different formats: XML, HTML, tag/delimited etc). This year FAO in collaboration with Associazione per la documentazione le biblioteche e gli archivi (DBA) worked together on ensuring OAI-PMH compliancy for Web based CDS/ISIS applications. An open source software layer (plug in) was developed, that permits a harvester to pull metadata from internet accessible general structure CDS/ISIS databases using OAI-PMH

protocol. The plug-in allows the harvesting of metadata in not only simple Dublin Core format but also AGRIS AP, which provides a semantically rich description. This is an important step not only for achieving interoperability and improving visibility and access to the metadata and full text agricultural resources from CDS/ISIS Web based applications but also a step towards creation of new value-added services based on the shared information, compliant with the common standards.

Poster: <http://oai5.web.cern.ch/oai5/posters.html#Poster30>.

2.2 Subject Specific Vocabularies (Application Profiles for Data Exchange)

2.2.1 Organizations Application Profile (Ag-Org AP)

Agricultural sciences and technology boasts a large number of organizations, both in the developing and developed countries, mainly because agriculture is the primary industry in nearly all the countries. Knowledge of these organization has become important in the sharing and exchange of agricultural related information as they provide contacts to individual institutional activities, information on planned and current development programs and projects of major funding and implementing to mention a few. As any information type, exchanging information on organizations has its limitations especially due to the lack of standards for the description and exchange. Studies, on the available systems that contain organization information, indicate that most of them have been created to meet their individual needs. The goal of this project is, therefore, to define a standard exchange format for "basic" metadata about an organization. Metadata about an organization are a means to help identify regional, national and international organizations specializing in different agriculture-related domains.

Ag-Org AP:
<ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/organizationap.pdf>.

2.2.2 Events Application Profile (Ag-Event AP) and Exchange

A standard to interchange information about events is urgently required. There are many aspects of an event, ranging from simple announcement to detailed description with session breakdowns. The main objective of this project is to allow users to 'know' about an upcoming event and guide them to the event web site which provides further detailed information. The information communicated will thus remain simple yet interoperable across domains and organizations. This document is designed to help create valid web-feeds (in RSS or Atom) for events.

Ag-Event AP: <ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/eventap.pdf>.

The Ag Event AP has been implemented in the Agrifeeds Application. The AGRIS Secretariat at FAO has produced the Agrifeeds Application to help aggregate news and event feeds. The Application then allows the users to disaggregate feeds based on filter criteria such as subject, country of location and language. The overall functionalities of the current service are:

- Registry of sources (Web feeds exposing data on news and events) with metadata about each feed;
- Harvesting of items from all registered feeds and creation of one (single) feed;
- Filtering of feeds (language, country, region, subject) and access to the custom feeds at any time through a URL;
- iCalendar functionality for users;
- Guidelines and tools for creating and embedding feeds into websites.

More information is available from: <http://www.agrifeeds.org/>.

2.2.3 Projects Application Profile (Ag-Project AP)/Collaboration between FAO and Wageningen International

FAO has been collaborating with Wageningen International to improve the possibilities of sharing project information management. The collaboration concentrated on publishing CARIS information using WISARD Platform and creation of a standard (Ag-Project AP) to share Project information between Project meta-databases. In light of this, the status of the activities is as follows:

- CARIS legacy data, and the most recent contributions, was successfully transferred to the WISARD system.
- The new CARIS-WISARD portal (which provides search functionalities) was published for testing and evaluation. The new portal provides simple search, advanced search by geographic terms (continents, regions, countries), and a combined simultaneous search for research projects in InfoSys+, R4D, WISARD. The final release of the portal is expected towards the end of 2007.
- Work for a draft of a project application profile is ongoing and a DTD has been made available for comments.

CARIS-WISARD Portal: <http://www.carisprojects.org/home/>.

2.2.4 Learning Resources Application Profile (Ag-LR AP)

Capacity and institution building is a core function of FAO, and the Organization has recently started the "Capacity and Institution Building Portal" to provide structured access to information on FAO's capacity and institution building services and learning resources. To ensure that the Portal can be searched by users and to enable interoperability with other recognized educational repositories, an application profile was created conforming to available and commonly used standards, to describe agricultural learning resources. This article presents the application profile, provides an example of an FAO learning resource described and displayed using FAO Learning Resource Application Profile, and presents the lessons learned.

Ag-LR AP: <ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/learnap.pdf>.

2.2.5 Job announcements

In this case, the schema currently used by GFIS will be used. More information is available from: <http://www.gfis.net/gfis/exchange.faces#jobs>.

2.2.6 Experts

This will be developed if required by partners. Some basic evaluation of existing schemas such as "Friend-of-a-Friend" has already been carried out.

2.2.7 Fellowships

A Fellowships Application Profile is currently under development at FAO.

2.3 Subject Specific Vocabularies (Thesauri and Ontologies)

2.3.1 Localization of AGROVOC (Mango Project)

As a result of the Expert meeting held in FAO in October 2005, there was the need of bringing together local knowledge and be able to visualize it in order to clearly identify concepts and terms in multiple languages. The working group on subject vocabularies

proposed to establish a prototype project in order to bring together the knowledge represented in several languages in the area of Mango. However, due to a lack of resources no international efforts could be undertaken. However, a test on visualization of mango related knowledge in Hindi was done in India by the Indian Institute of Technology in Kanpur.

2.3.2 AGROVOC Content Development

In the past two years, FAO and its partners have contributed to the development of AGROVOC with inclusion of new terms (over 550), inclusion of language versions (Czech, German, Lao, Italian, Polish, Farsi and Hindi), revision of language versions (Hungarian), and improvement of the structure. AGROVOC is available in following formats for free downloading (non-commercial use): MySQL, MS Access, SKOS, Postgres, OWL, TagText, ISO2709, and other formats which may be requested for specific applications. AGROVOC is downloaded on average 35 times a month for mainly research purpose, offline use, as a translation aid and for use in applications. AGROVOC is also available for real-time access and use via web services.

More on AGROVOC: http://www.fao.org/aims/ag_intro.htm

2.3.3 AGROVOC Concept Server

In the past two years, FAO has developed a thesaurus model for AGROVOC using the web Ontology language (OWL). Based on this model an AGROVOC concept server and workbench has been developed together with the Thai AGRIS centre and Kasetsart University (Thailand). The AGROVOC concept server workbench will allow collective and distributed maintenance of AGROVOC.

An online E-Conference was held, between May-July 2007, to understand the AGROVOC user community and their requirements. A total of 101 participants registered for the E-Conference from 43 countries, providing an extremely interesting opportunity to collaborate on the future development of AGROVOC.

E-Conference report:
ftp://ftp.fao.org/gi/gil/gilws/aims/publications/papers/AGROVOC_E-Conference_Full_report.pdf

2.3.4 Domain Ontologies

Out of AGROVOC and other knowledge organization systems and based on the requirements from various projects, domain ontologies have been developed for specific systems

- Ontology for Country based Information Systems
- Crop Wild Relatives Ontology : <http://www.fao.org/aims/aos/cwr.owl>
- Food & Nutrition Ontology :
<http://www.fao.org/ag/agn/publications/fna/index.jsp?lang=en>
- Language Code Ontology: <http://www.fao.org/aims/aos/languagecode.owl>
- Fishery Ontologies: <http://www.fao.org/aims/aos/fi/>

2.3.5 NeOn

Starting March 2006, FAO has been a partner in a 12 Million Euro project funded by the European Commission to develop Lifecycle Management Systems for networked ontologies (NeOn). The participation in this project has given the possibility to develop various domain ontologies, especially those related to fishery (vessels, species, water

bodies and others). FAO will test the tools developed in this project through a Use Case on ontology-based assessment of fish stock depletion. The NeOn project will deliver in 3 years time the tools to manage the Agricultural Ontology Service.

3. CONCLUSION

The CMTF has worked on all the areas that were defined as being important during the 2005 Expert Consultation. In nearly all areas considerable progress has been made in establishing standards and methodologies and providing tools. In the next two years, it will be necessary to:

- Conduct technical meetings to fine tune and complete the developed tools and standards;
- Improve the community platform and collaboration mechanisms;
- Promote the work done by the CMTF through appropriate advocacy channels.

Membership Content Management Task Force

CGIAR:	Enrica Porcari
CLAES:	Ahmed Rafea
CTA:	Kevin Painting; Koda Traore
FAO:	Gauri Salokhe; Johannes Keizer
GFAR:	Ajit Maru; Valeria Pesce
GFIS/IUFRO:	Eero Mikkola
IAALD:	Peter Ballantyne
IICA:	Federico Sancho
WIS International:	Joost Lieshout; Koen Beelen
ZADI:	Marc Bernard
Moderator:	Hugo Besemer
Observer:	Chris Addison

Useful Materials/Links³

1. Content Management Task Force Blog: <http://agriscontent.wordpress.com/>
2. Content Management Task Force Wiki: <https://agriscontent.pbwiki.com/>
3. Metadata Flyer:
ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/metadata_en.pdf
4. Ontology Flyer:
ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/ontologies_en.pdf
5. AGROVOC Flyer:
ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/agrovoc_en.pdf
6. Event Application Profile:
<ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/eventap.pdf>
7. Event Aggregator: <http://www.agrifeeds.org/>
8. Organization Application Profile:
<ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/organizationap.pdf>
9. Organization Registry Concept Paper:
<http://www.egfar.org/egfar/website/opensite/collabwebsite?contentId=1599§ionId=1463&folderId=1463>
10. Learning Resources Application Profile:
<ftp://ftp.fao.org/gi/gil/gilws/aims/metadata/docs/learnap.pdf>
11. CARIS Project: <http://www.carisprojects.org/home/>
12. AGROVOC Language versions: <http://www.fao.org/aims/>
13. NeOn Project Homepage: <http://www.neon-project.org/>
14. "Agricultural Information and Knowledge Management Papers" published by FAO and partners:
http://www.fao.org/documents/advanced_s_result.asp?FORM_C=AND&SERIES=339

³ All resources last accessed on: 10 September 2007

List of acronyms and abbreviations

AGMES	Agricultural Metadata Element Set
AGRIS	International Information System for the Agricultural Sciences and Technology
AGROVOC	Multilingual Agricultural Thesaurus
AIMS	Agricultural Information Management Standards Initiative
AOS	Agricultural Ontology Service
ARD	Agricultural Research for Development
ASARECA	Association for Agricultural Research in Eastern and Central Africa
CABI	CAB International
CGIAR	Consultative Group on International Agricultural Research
CIRAD	Centre de coopération internationale en recherche agronomique pour le développement
CLAES	Central Laboratory for Agricultural Expert Systems
CTA	Technical Centre for Agricultural and Rural Cooperation
GFAR	Global Forum on Agricultural Research
GFIS	Global Forestry Information System
IAALD	International Association of Agricultural Information Specialists
ICT	Information and Communication Technology
IICA	Inter-American Institute for Cooperation on Agriculture
IUFRO	International Union of Forest Research Organizations
NAL	National Agricultural Library
NARIMS	National Agricultural Research Information Management System Project
NARS	National Agricultural Research System
OAI	Open Access Initiative
RDF	Resource Description Framework
RSS	RDF Site Summary
ZADI	Zentralstelle für Agrardokumentation und -information (Agricultural Information and Documentation Centre)

Report from the Content Management Taskforce Meeting in Wageningen