

---

## AGRIS Application Profile : Methodology

### 1. Where was the demand?

---

The first **COAIM** (Consultation on Agricultural Information Management), held in June 2000 recognized that AGRIS should become “a key enabler and catalyst to establish a new model of agricultural information management”. Furthermore, it was agreed that FAO should develop AGRIS into a capacity building initiative as well as an information system. The AGRIS network now has 201 Resource Centres, which vary in their resources and level of participation in the network, and opportunities exist to improve the effectiveness of the initiative through enhanced collaboration.

A proposal for a new XML based metadata standard for AGRIS reference was presented at COAIM 2002. This document explains the methodology followed for creating the “*AGRIS: Guidelines for the Description of Information Objects in the International Information System on Agricultural Sciences and Technology*”.

- The standard itself is available from:  
<ftp://ext-ftp.fao.org/agris/agmes/AGRISAP-UserGuide.pdf>
- The accompanying DTD to validate the XML inputs is available from:  
<http://www.purl.org/agmes/agrisap/dtd/> (used in XML DTD declaration)  
[http://www.fao.org/agris/agmes/Documents/20040421\\_agris\\_xml.dtd.txt](http://www.fao.org/agris/agmes/Documents/20040421_agris_xml.dtd.txt) (text file)
- Sample XML document is available at:  
[http://www.fao.org/agris/agmes/Documents/XML\\_input\\_sample.xml](http://www.fao.org/agris/agmes/Documents/XML_input_sample.xml)

### 2. What is the AGRIS Application Profile?

---

The AGRIS Application Profile (AGRIS AP) is a high quality metadata format for exchange of agricultural information resources within the International Information System for the Agricultural Sciences and Technology (AGRIS) and is directed at improving accessibility of agricultural information materials via the web. The AGRIS AP is a data-model created by taking elements *already* specified in the Agricultural Metadata Standards Element Set<sup>1</sup> (AgMES), the Dublin Core Metadata Element Set<sup>2</sup> (DCMES) and the Australian Government Locator Service<sup>3</sup> (AGLS).

An application profile has the following properties. It

- draws on one or more existing namespaces,
- introduces no new elements or refinements,
- specifies the permitted schemes and values, and
- refines, when necessary, standard definitions of elements or refinements (definitions provided in the namespace).

The objectives of the AGRIS AP are as follows:

- To serve as a flexible, platform independent, information exchange format that adheres to current metadata standards while being interoperable with different information systems for exchange of different types of information within the AGRIS Network.

---

<sup>1</sup> Agricultural Metadata Element Set <http://www.fao.org/agris/agmes/>

<sup>2</sup> Dublin Core Metadata Initiative <http://www.dublincore.org/>

<sup>3</sup> Australian Government Locator Service [http://www.naa.gov.au/recordkeeping/gov\\_online/agls/cim/cim\\_manual.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/cim/cim_manual.html).

- To facilitate exchange of more types of agricultural information resources that are not presently covered by the old exchange format, consequently enriching the information base for the participating members.
- Encourage participation by new members to the AGRIS Network.

### 3. Why do we need an application profile?

The available information on the internet, in most domains, is growing exponentially. To create quality metadata for this information is a colossal task that we will never be able to complete even with the help of automatic indexing tools. The real promise of metadata is not only its ability to help the user find relevant resources but also to facilitate sharing across proprietary systems as well as organizational boundaries.

Within a loosely-knit community or domain, the use of an agreed standard such as an application profile for sharing information is cost-effective, albeit with small amounts of data or information loss. The cost of moving towards complete homogeneity of information or metadata is as high as the cost of creating metadata from scratch.

This social system, of sharing using a common exchange layer, provides the platform on which other services can be based. For example, a portal that brings together all the country specific regulations on “Animal Nutrition”. In this case, the information is shared (by the different databases), merged (by normalizing and removing duplicates) and presented to the user (for browsing, searching etc.).

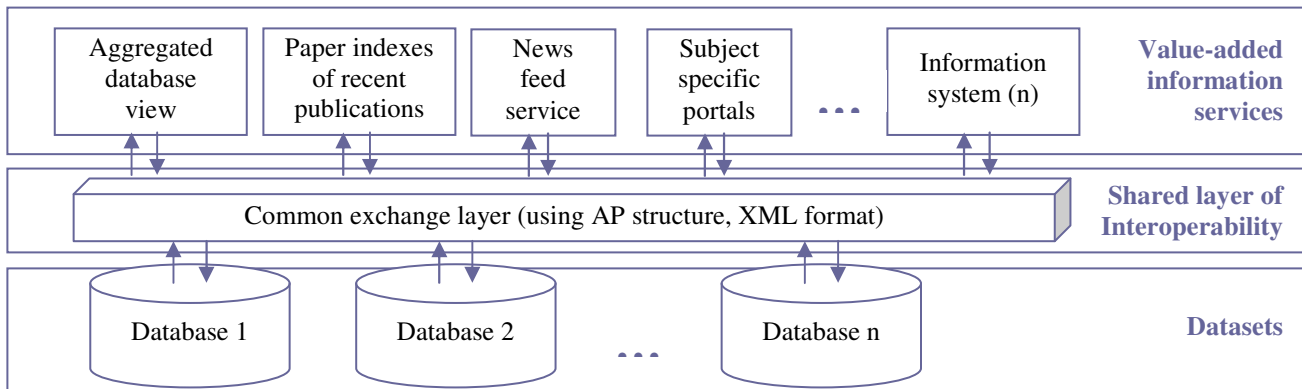


Figure 1: Interoperability between datasets allow for creation of value-added services and systems

Thus, the databases that are exposing their data, are not only becoming interoperable with the Portal itself but also with each other.

An application profile, or in database terms a data model, prescribes the vocabulary, content and structure rules that can be used to share information between heterogeneous datasets without requiring any change to the local system. The terms of this AP need to be ‘negotiated’ to represent the community and the domain in which it is being applied. Then, with the possibility of using tools such as XSLT, the information extraction and conversion becomes a simple yet extremely important task towards facilitating interoperability. The fact that the resource itself is not required to be attached to the metadata makes it easy to control access rights on it.

---

In the Agricultural community, the process of sharing data with AGRIS and between each other is illustrated below.

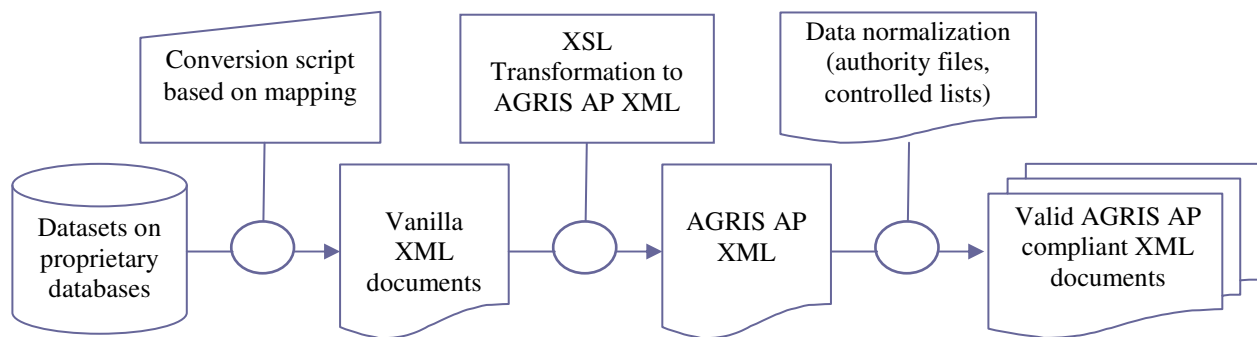


Figure 2: Converting from one dataset to AGRIS AP

The first step involves mapping the local data-model to the Application Profile. This is followed by either creation of a conversion script or direct export of data into XML (as some databases allow this). Then XSL Transformation is applied to convert the simple XML or Tag Text document into the AGRIS AP-compliant XML format. This is followed by data normalization step which makes sure not only the XML file validates correctly against the DTD but also that all the requirements, such as use of controlled vocabularies or declaration of used scheme, are met. The resulting XML documents, from this conversion step, represent the common exchange layer shown in Figure 1.

This platform independent sharing of metadata, with or without sharing the actual resource, allows communities to provide relevant results to their users.

#### 4. Some commonly used terms and their definitions

---

The following definitions explain the concepts used in this document.

##### Elements

An **element** is described as a unit of data or metadata. The element allows us to give more information about the described information.

##### Refinements

An **element refinement or qualifier** makes the meaning of an element either narrower or more specific. Additionally, element refinement shares the meaning of the unqualified element, but with a more restricted scope. When a client does not understand the element refinement, it can be ignored and the value used as content of the unqualified element.

##### Encoding Scheme

An encoding scheme aids in the interpretation of the value of an element. Encoding schemes may either be controlled vocabularies or formal notations. A value drawn from an encoding scheme can be taken from a controlled list of vocabulary (e.g. a term from a classification such as ASC (AGRIS Subject Categories) or a term from a thesaurus such as AGROVOC). When a client does not understand the encoding scheme, it can be still useful for human readers. To see some standardized schemes, see the following:

- Three letter codes for the Representation of Names of Languages  
<http://www.loc.gov/standards/iso639-2/langcodes.html>

- 
- Codes for the Representation of Countries  
<http://www.iso.ch/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>

### **Namespaces**

An **XML namespace** is a collection of names, identified by a URI reference, which are used in XML documents as element types and attribute names. XML namespaces differ from the "namespaces" conventionally used in computing disciplines in that the XML version has an internal structure and is not, mathematically speaking, a set. To learn more about XML namespaces, see the following:

- Namespaces in XML <http://www.w3.org/TR/REC-xml-names/>
- XML <http://www.w3.org/XML/>

### **Application Profile**

An **application profile** is a type of metadata schema, which consists of data elements drawn from one or more namespaces, combined together by implementers, and optimised for a particular local application. The AGRIS AP has been created by taking elements and refinements that are already in existence, such as those declared by organizations like *Dublin Core* and *Australian Government Locator Service Metadata* and those declared by the *Agricultural Metadata Standards Initiative*. This endorses the current trend for promoting reuse instead of reinvention. To understand more the concept of AP, see the following:

- Baker, Dekkers, Heery, Patel and Salokhe (2001) "What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives". JoDI, Vol 2., Issue 2. <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/>
- Heery, Rachel and Manjula Patel (2000) "Application profiles: mixing and matching metadata schemas". Ariadne, No. 25, September. <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>

### **Term**

A term is an element, a refinement or a scheme declared in a namespace. Each of these is defined and has a unique URI. In this document, elements, refinements and schemes are referred to as "term" when they do not need to be mentioned specifically.

## **5. Methodology adopted**

---

The creation of AGRIS AP involved several phases. The duration of each phase depended on the tasks involved to complete the phase.

### **Phase 0 : Define the project, its goal and scope**

The first and foremost task before starting any such project is to establish the clear need for it. The project must be well-defined in what it wants to achieve, what are its short-term and long term objectives. The scope of the project will also have to be defined as this helps define the ground and helps to keep the work within its boundaries. In the case of AGRIS system, there were two major goals:

1. Provide a platform independent exchange format and
2. Update the rules of AGRIS to match current levels of digital information management standards and tools.

---

The solution for the first goal was clearly to provide a format, such as XML, that would not bind resource centres to any AGRIS specific databases yet allow them to share their data, regardless of the information system they were using. As for the second goal, the task was much more time-consuming. The AGRIS rules were written in the early 1970s and reflected the standards and technology of the time, sometimes made specifically for paper based inputs. The data structure was also mainly flat and not immediately applicable or extendable to a more popular and practical relational model.

The scope of our project was, of course, the existing resource types of the AGRIS System. It was, however, extended to include other document-like information objects<sup>4</sup> (DLIOs): databases, Web pages, national portals on scientific and technical information on Agriculture.

### **Phase 1 : Assessment of the information objects**

To create the AGRIS application profile, a group was set up at the FAO headquarter. The group consisted of well trained cataloguers, information specialists and a reference librarian who provided the user's perspective.

The first task was to collect different types of Document-like information objects (IOs) that were already existing in the AGRIS System and the new forms such as databases, websites, etc. Each of these IOs had to be analyzed and a list of properties that were needed to describe them was prepared. This helped to set up the requirements for the final application profile. For example, to describe a book, it is important to be able to indicate its title, its identifier which usually is expressed as an ISBN. Similarly, to describe a journal article, it is important to be able to give not only the title of the article and its ISSN but also the title of the serial.

### **Phase 2 : Assessment of the existing metadata standards and creation of the AP**

The Dublin Core Metadata Initiative<sup>5</sup> (DCMI) has come up with a set of widely accepted elements, refinements and schemes which help in resource description and discovery. It is clearly defined yet limited by its shallow scope. This is mainly due to the aim to be simple; its simplicity supports cross-domain description, discovery and retrieval of information objects. Another important trait of the DC is its extensibility: there is no limit on extending the set with additional elements, refinements or schemes. However, this extensibility has to be well controlled, without creating problems with interoperability by creating hoards of local extensions.

The Agricultural Metadata Element set<sup>6</sup> (AgMES) initiative aims to encompass issues of semantic standards in the domain of agriculture with respect to description, resource discovery, interoperability and data exchange for different types of information resources. AgMES acts as an umbrella under which namespaces can be defined for newly declared elements that are deemed necessary and are used for different resources (DLIOs, projects, fishing gear presentations, images, technologies, practices, maps etc.) in all areas relevant to food production, nutrition and rural development.

---

<sup>4</sup> A DLIO is a unit that is comparable to a paper document. The term is used to indicate resources such as websites, power-point files, photos etc. but may not cover, for example, organizations or projects.

<sup>5</sup> Dublin Core Metadata Initiative : <http://www.dublincore.org/>

<sup>6</sup> Agricultural Metadata Element Set : <http://www.fao.org/agris/agmes/>

---

The list of properties that were needed to describe the different DLIOs available in the AGRIS System were looked at iteratively, following the process mentioned below. This helped to decide if we really needed to have a particular property to describe the DLIO.

Is there a real requirement for the elements/refinement/scheme for :

- supporting resource description?
- supporting resource discovery?
- supporting interoperability?

In not, then could it be implemented in some other way, for example, putting the content in dc:description field instead of creating a new one?

Then, with the “sieved” list of necessary elements (for description and searching), the following iterative process was conducted<sup>7</sup>.

- Determine if there is a real need for the property? Does it really describe the resource or is it describing a related resource? If there is a need, then
- Can the need be solved with a scheme value for an existing DC element? If yes, then create an AgMES scheme for an existing DC element, or else
- Can the need be solved with a refinement for an existing DC element? If yes, then create an AgMES refinement for an existing DC element, or else
- Can the need be solved by a qualifier from an existing non-DC set? If yes, then use that as a qualifier for the DC element, or else
- Can the need be solved by an element from an existing non-DC set? If so, then use that element, or else
- Create a new AgMES element (and if necessary a scheme).

This task of trying to match each property to an existing term made sure that we were not reinventing the wheel. Thus, all the declared elements, refinements and schemes in AgMES look like a hotchpotch and need to be seen along with their DC parent<sup>8</sup>.

During this process, we not only looked at Dublin Core but also other element sets. In the case of the “availability” element, we found this already declared in the Australian Government Locator Service element set. Thus, instead of reinventing, we took the element and furnished it in AgMES with qualifiers, namely, availabilityLocation and availabilityNumber.

### **Step 2.1 : Provide the ISO11179 metadata for each term in the AgMES namespace**

Once the element, refinements and schemes were given entry into AgMES, they were then carefully described using the ISO/IEC 11179 standard for the description of metadata elements. The use of the ISO/IEC 11179 helps to improve consistency with other communities, and augments the scope, consistency and transparency of the AgMES.

---

<sup>7</sup> Adopted from original guidelines from Stuart Sutton <http://www.schemas-forum.org/workshops/ws2/ws2-presentations/DC-Ed.ppt>

<sup>8</sup> List of all the elements endorsed by AgMES : <http://www.fao.org/agris/agmes/Documents/Elements.html>

<b>Attribute Name</b>	<b>Definition</b>
Name	The unique identifier assigned to the data element.
Label	Label assigned to the data element.
Version	The version of the data element.
Registration Authority	The entity authorized to register the data element.
Language	The language in which the data element is specified
Definition	A statement that clearly represents the concept and essential nature of the data element
Obligation	Indicates if the data element is always or only sometimes required (mandatory, optional, conditional)
Data type	Indicates the type of data that can be represented in the value of the data element
Maximum Occurrence	Indicates any limit to the repeatability of the data element.
Comment	A remark concerning the application of the data element.

The terms **Name** and **Label** are not as they are in ISO/IEC 11179 and were modified to adhere with the terminology currently being used in the XML community. This approach was taken to facilitate the easy assimilation of this set into the XML and RDF communities.

Additionally, the following two attributes were also used.

<b>Attribute Name</b>	<b>Definition</b>
Element Refined	The name(s) of element(s) refined.
Scheme	The applicable schemes for encoding the values of the term.

The full description of all the AgMES terms is available at: [ftp://ext-ftp.fao.org/agris/agmes/AGMESNS-DLIO\\_en.doc](ftp://ext-ftp.fao.org/agris/agmes/AGMESNS-DLIO_en.doc)

### **Step 2.2 : Create the data-model of the application profile**

The next step involved taking each of the terms and defining them in the context of AGRIS. Remember that APs allow us to provide application specific definitions as long as they do not change the concept itself. For each element, for example Title, we wrote its definition, provided cardinality and data type information. This was followed by providing some examples of best practice guidelines. The best practices guidelines tried to cover as many scenarios as possible but could not be exhaustive for obvious reasons. These guidelines sometimes suggested the use of a scheme, for example, ISO639-2 scheme should be used to indicate the language of the resource.

This was done for all the elements and refinements.

### **Phase 3 : Create XML DTD or schema**

The guidelines then were converted into an XML DTD which would be used to validate all the XML based inputs to the AGRIS System.

The XML DTD allows you to provide the following:

---

Logical structure of the record	Which element should follow which element or how the elements should be nested.
Obligation	If a term is mandatory or optional
Cardinality	How often can this term appear in one record (0, 1 or more times)

---

The DTD is available at : <http://www.purl.org/agmes/agrisap/dtd/>

#### **Phase 4 : Test the real data with XML DTD**

The application profile was then made available as both, as a document and also as an XML DTD, necessary for validating XML inputs. The guidelines were applied by the Dutch Library, who helped to refine the document and thus the DTD. Sample of the Dutch input is available at : [http://www.fao.org/agris/agmes/Documents/XML\\_input\\_sample.xml](http://www.fao.org/agris/agmes/Documents/XML_input_sample.xml)

#### **Phase 5 : Supporting the institutes/databases which implement the Application profile**

Once the Standard was established, with repeated testing and updating, it was re-published. The technical implementers will be provided with a Technical Guide (coming soon). However, before that, each implementer was given one-to-one feedback to help them to successfully implement the exchange standard. In some cases, this involved creation of crosswalks or mapping between the local systems and the AGRIS AP.

## **6. The remuneration**

---

To date, records in AGRIS AP compliant format have been submitted by The Netherlands and Finland. The AGRIS AP compliant XML format has also been implemented in the free tool, WebAGRIS, provided by FAO to its AGRIS Centres allowing all those who use it to bypass much of the conversion process mentioned in Figure 2.

As mentioned before, the data/information loss during the mapping and conversion process takes places at the cost of increased interoperability.