Towards an infrastructure for semantic applications: Methodologies for semantic integration of heterogeneous resources

By Anita Liang, Gauri Salokhe, Margherita Sini, Johannes Keizer

SUMMARY:

As with many domains, information retrieval and knowledge management (IR/KM) in agriculture suffers from the problems of semantic heterogeneity, making it difficult for providers to disseminate their services effectively and for users to retrieve the information they need. Based on the analysis of resources in the domain of agriculture, this paper proposes a) application profiles for dealing with the problem of heterogeneity originating from differences in terminologies, domain coverage, and domain modelling, and b) a root ontology based on the application profile which can serve as a basis for extending knowledge of the domain.

KEYWORDS: ***Application profiles, Information integration, Information retrieval, Knowledge management, Metadata, Ontologies, Semantic standards***

1 Introduction

The information resources available from the fora of international agriculture related arena differ in degree of coverage (e.g., some resources, such as AGROVOC, principally concern generic agriculture, food safety, etc.), sub-discipline (e.g., nutrition, animal and plant health), type (e.g., databases, images, news announcements), and content (e.g., journals, institutions, expert information, project descriptions, thesauri). What is immediately evident is the distributed, heterogeneous nature of the resources comprising this domain, and that no single search engine could retrieve a comprehensive set of the resources relevant to a user's needs.

Users who are looking for information on, say, French researchers working on stem cells, have at their disposal search engines that can go through these files (and indeed, the millions of files on the World Wide Web) at the blink of an eye. Yet, there is little guarantee that what the search engines

will find and display will correspond in meaningful ways to the user's query: She may enter the query *stem cell researchers in France* but the resources that might constitute good responses to her query refer to *French researchers* and employ the plural form *stem cells*[1]; or, alternatively, it may be the case that relevant information is included within the results displayed, but that she has to manually sift through a dozen pages of irrelevant results (e.g., pages containing her query in the form of a bag of words, i.e., *stem*, *cell*, *researchers*, and *French*, appearing as separate terms) to get to them; or, the page might contain biased or inaccurate statements about the topic.

To handle the huge quantity and heterogeneity of information published on the Web intelligently and efficiently, the WWW needs to transform itself into a system for disseminating knowledge that can be interpreted not only by humans but also and especially by computers. This implies an evolution to a web that is first and foremost meaning based rather than form based. An intelligent WWW, that is, one using semantic technologies, could then process the query *stem cell researchers in France*. Based on its "understanding" of the query (e.g., through a process of resolving the query terms into concepts and matching those concepts to an ontology over which reasoning can be performed), it could conduct not only a comprehensive search, but also retrieve/suggest related concepts and resources, irrespective of the actual terms and language of the query.

AGRIS[2] is the international information system for the agricultural sciences and technology, created by the Food and Agriculture Organization of the United Nations (FAO) in 1974, to facilitate information exchange and to bring together world literature dealing with all aspects of agriculture. AGRIS is a cooperative system in which participating countries input references to the literature produced within their boundaries and, in return, draw on the information provided by the other participants.

The aim of this paper is to specify the rationale and the methodologies for developing semantic standards in the domain of Agriculture. In particular, we propose, on the one hand, an *AGRIS*

Application Profile (AAP) to address the problem of semantic heterogeneity of exchanging metadata on document-like information objects. On the other hand, from the AAP, we derive a corresponding ontology, the AGRIS *Application Ontology* (AAO). The application ontology upon which the application profile is based makes explicit the semantics that already exists within the application profile, and may be further enriched with additional semantics through the introduction of schemes, thesauri, and other terminologies. Thus, the semantic richness of the application ontology varies according to the extent to which additional concepts and relations have been incorporated into the ontology. We will refer to this AAO alternatively as a root ontology, since it serves as a starting point for further semantic extensions.

Developing and applying standards for resource description is a prerequisite for creating the infrastructure for a network of information services that can alleviate the semantic heterogeneity of the diverse and distributed services providing information resources in the Agricultural domain. Moreover, this emphasis on meaning over form allows for the development of "smart" applications for areas such as content management (e.g., automatic mark-up of documents), knowledge management (e.g., expert locators, concept-based search), and advisors/recommenders (e.g., mediators).

2 Semantic Web

In his vision of the Semantic Web, Tim Berners-Lee (Berners-Lee 2000) outlines an architecture for the Web that is multi-layered and machine processable, as depicted in the much-reproduced image in Figure 1. The layers with which we will principally be concerned are the resource description framework layer and the ontology layer. The XML layer will be touched upon insofar as it addresses the issue of content.

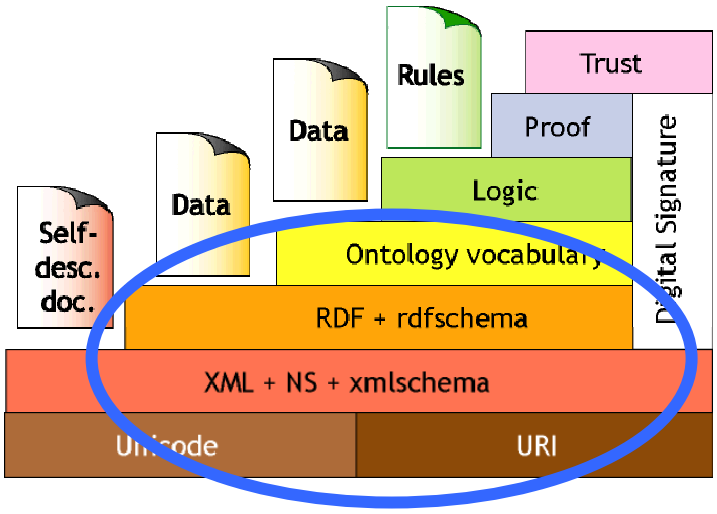

Figure 1. Layers of concern for the development of semantic standards in the agricultural domain

2.1 The XML layer: Content of a resource

The XML layer is concerned with the description of what a document or resource is about. Inasmuch as data that is proprietary to an application has limited use, the XML layer provides for standardized means of describing content in order to free up that content for use by any number of applications.

2.1.1 Granularity

Domain-specific XML tags can be used to mark up the content of a resource at various levels of granularity ranging from the level of the resource itself (i.e., to describe what the resource is about using descriptor terms or abstracts) down to the level of the section or passage within the document (if it is a text), to the sentence-level, to the level of a single term (i.e., to describe what the term means or refers to). In the case of structured data such as databases, the database itself might be described, or the fields of the database might be semantically indexed.

The level of granularity at which data is indexed is directly related to the types of queries the user can ask and the types of results that can be retrieved. If resources are marked up coarsely, such as at the level of the website (or individual pages on a website) or metadata record, then the user's query,

normally in the form of one or more keywords, will retrieve a set of links that either contain or are associated with (via, for example, matching strings in the resource's metadata) the user's keywords. This option limits resources to those that can be identified via a URL or metadata records retrieved via a SELECT SQL query. Contrastively, at finer-levels of marking up information, such as the chapter of a book, or a passage, the results retrieved can be more directly targeted to the user's query, more so than matching keywords against metadata describing a document or URL, which may not contain the information needed by the user. For instance, if the user is looking for information on the health hazards to humans of pesticide use in Africa, she might indeed find a document keyworded with (or having significant frequencies of occurrences of) “pesticides”, “health hazards”, and “Zambia,” but the document might be about the removal of these substances, or the amount sold of those types of substances, or regulations about their use, etc., rather than about the ways in which they are hazardous. This is because there is little or no indication of what the relationship is between the terms used to describe a given resource and the resource itself, or among the terms themselves. When information is described below the level of the resource, retrieval results may match user queries more effectively. So an article containing a section on pesticides, health hazards, and Zambia would be indexed differently, and ranked at a higher position than one containing one section on pesticides and health hazards and another on Zambian culture. At the finest level of mark-up, where individual words are indexed, the system “understands” the meaning of each term in a sentence, as well as its relation to other terms. In such a system, the user can issue a well-formed question as a query, and the result elicited would be in the form of an individual sentence, based on an analysis of the user’s query and a search for the best match among the sentences within the resources. For example, it would be possible for the user to input “What are the health hazards of pesticides used in Zambia?” and for a direct response to be in the form of a sentence drawn from resources, e.g., “Pesticide use in Zambia are associated with the following toxic effects.”

Within the domain of agriculture, content description is initially envisaged at the metadata level, within the RDF layer (cf. next section), using controlled vocabularies. Thus, what can be retrieved are resources such as experts, software, and DLIOs (as opposed to individual answers, discussion threads, or text passages). This is mainly for practical reasons, given that there already exist numerous bibliographical databases that contain descriptions of bibliographic metadata using controlled vocabularies. The database structures can be studied to develop the initial version of the AAP and the AAO, while the controlled vocabularies can serve as the basis of further developing the AAO. However, it must be stressed that this is a starting point, and that more sophisticated systems can be developed once the AAO has been extended using vocabularies containing rich semantics.

2.2 The RDF layer: Metadata of a resource

The RDF layer contains information about a resource, viewed externally, that is, from outside the resource, and includes information such as its title, author, and publisher. This information that describes a resource is called metadata. Standardized XML tags can be used to mark up metadata. For resource description, there already exist standards such as the Dublin Core Metadata Element Set[3] (DCMES). The Agricultural Metadata Element Set[4] (AgMES), which complements the DCMES, has also become a standard commonly used in the domain of Agriculture, with its specific emphasis of agricultural vocabularies and terminologies. What distinguishes, however, the lower XML layer, which merely describes a resource, from the RDF layer (Figure 2) is that the latter is able to express relations between resources.

```
<rdf:Description rdf:about="http://www.fao.org/ag-pub/citation/1">
  <rdf:type rdf:resource="http://purl.org/agmes/1.1/citation"/>
  <ags:citationTitle rdf:resource="http://www.fao.org/ag-pub/citationTitle/1"/>
  <ags:citationNumber rdf:resource="http://www.fao.org/ag-pub/citationNumber/1"/>
  <ags:citationIdentifier rdf:resource="http://www.fao.org/ag-pub/citationIdentifier/1"/>
</rdf:Description>
<!-- Start description of the first resource -->
<rdf:Description rdf:about="http://www.fao.org/ag-pub/XM20053081089">
  <dc:title xml:lang="en">Application of a novel disposable film culture system</dc:title>
  <dc:creator rdf:resource="http://www.fao.org/ag-pub/creator2"/>
  <dc:creator rdf:resource="http://www.fao.org/ag-pub/creator3"/>
  <dc:creator rdf:resource="http://www.fao.org/ag-pub/creator4"/>
  <dc:publisher rdf:resource="http://www.fao.org/ag-pub/pub2"/>
  <dc:date rdf:resource="http://www.fao.org/ag-pub/date2"/>
  <ags:subjectThesaurus rdf:resource="http://www.fao.org/aos/2005/cabt#SQ00095"/>
  <ags:subjectThesaurus rdf:resource="http://www.fao.org/aos/2005/cabt#NT05852"/>
  <dcterms:abstract xml:lang="en">To overcome various disadvantages [...etc]</dcterms:abstract>
  <dc:identifier rdf:resource="http://www.ingentaconnect.com/content/cabi/ivp/"/>
  <dc:type rdf:resource="http://purl.org/dc/elements/1.1/type/JA"/>
  <dc:language xml:lang="en-US">en</dc:language>
  <agls:availability rdf:resource="http://www.fao.org/ag-pub/availability/2"/>
  <ags:citation rdf:resource="http://www.fao.org/ag-pub/citation/1"/>
</rdf:Description>
</rdf:RDF>
```

Figure 2: Section of metadata record expressed in RDF.

In contrast to the WWW, where associations, i.e., hyperlinks between resources are meaningful to the extent that they are interpretable by humans (e.g., while a human could understand why a string *Mahatma Gandhi* would be hyperlinked to an image of the Indian flag, to a computer, such a relation would be indistinguishable from any other text that was hyperlinked to an image). RDF provides a standardized format for uniquely defining resources and a well-defined syntax for making statements about those resources. Figure 3 exemplifies the type of statements that RDF allows about a resource.

Subject	Predicate	Object
http://http//www-smi.stanford.edu/pubs/SMI_Reports/SMI-97-0685.pdf/	http://purl.org/dc/elements/1.1/title	"A Study of Collaboration Among Medical Informatics Research Laboratories"

Figure 3. [resource] --dc:title--> *v*[dc:title]

As mentioned, for developing an integrated information service for the domain of Agriculture,

resources will be described using an *application profile* (AP), metadata elements that are drawn from one or more standardized metadata element sets that may also be extended and customized to the types of resources to be provided by the information service. This will allow resources to be described using standard mark-up languages that are independent of local platforms and applications and can enhance the possibility of semantic interoperability of resources within the domain. In section 3, we specify a methodology to develop an AP.

2.3 The Ontology layer: Modelling the domain

In the RDF layer, resources are defined by virtue of their relationships to other resources. The ontology layer offers, in addition, the possibility of reasoning within the domain through precise specifications of concepts, relations, and rules, thereby creating the possibility of inferring new data from existing data. In other words, an ontology provides a knowledge model of a given domain that can interface with the RDF layer via mappings to its metadata elements. The model is made explicit via a knowledge representation language. Although many such languages exist, we use OWL Web Ontology Language[5], the W3C standard knowledge representation language that offers rich semantics and is native to the Web (i.e., is serialized in XML).

For the domain of Agriculture, we distinguish two levels of knowledge to be represented: (1) One consists of the root ontology, where concepts, relations, and rules corresponding to the resource metadata will be specified for and mapped to the elements comprising the aforementioned application profile. (2) The other consists of all other ontologies derived from knowledge organisation systems such as thesauri and terminologies that can extend the root ontology. These other knowledge organization systems may provide a set of valid metadata values for resource attributes, or they may comprise an entire (sub) ontology in their own right that can extend the root ontology.

3 Methodologies for semantic integration

In the previous section, we outlined the parts of the Semantic Web architecture with which we will be concerned for the development of a semantic integration solution for Agriculture, as we have defined it. In this section, we will define the terminologies we are using and discuss explicitly the rationales and the methodologies for the development of those components.

3.1 AGRIS Application Profile

3.1.1 Definition and rationale

An application profile (AP) is a flexible, platform- and architecture-independent, information exchange format to facilitate the exchange of information resources via the Web for a given project or application. It consists of data elements (i.e., XML tags), drawn from one or more namespaces (i.e., named collection of elements and attributes), combined together and optimised for a given domain. By reusing elements specified in already-existing metadata standards, such as the DCMES, AgMES and the Australian Government Locator Service[6] (AGLS), the AP transcends proprietary systems and organizational boundaries, and thus creates the possibility of improving management of and accessibility to domain-specific information materials.

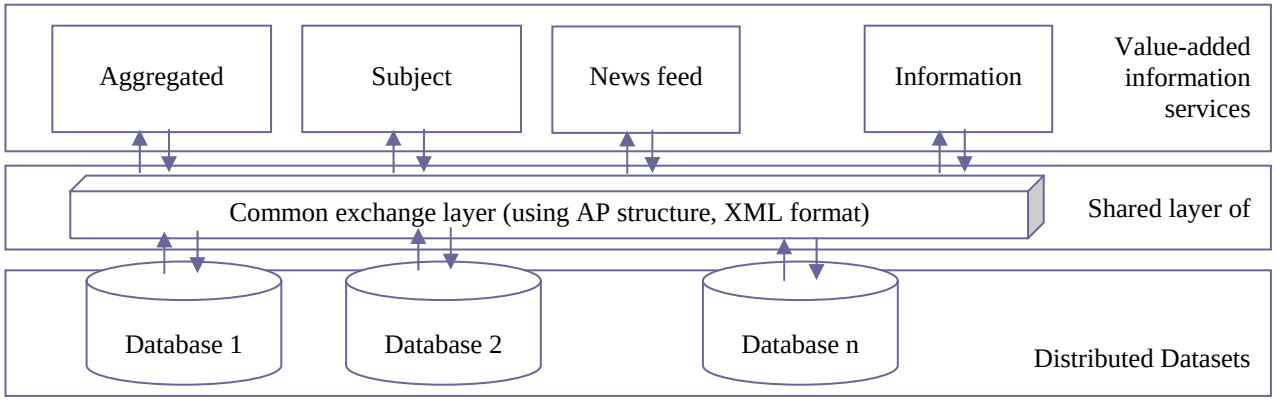

Figure 4: Interoperability between datasets allow for creation of value-added services and systems

Figure 4 shows the use of an AP as a common exchange layer to resolve the heterogeneity among information systems, and as a basis for the development of value-added services.

An AP 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. With the possibility of using tools such as XSL Transformation (XSLT)[7], the information extraction and conversion becomes a simple yet extremely important task towards facilitating interoperability. The fact that the resource itself does not have to be attached to the metadata makes it easy to control access rights on it.

The following steps briefly describe the process (Figure 5) of generating valid AGRIS XML records from proprietary XML-enabled databases:

1. Identify the fields in the catalogue of the local database that will match the AGRIS AP XML DTD elements and schemes. Export the desired fields into well-formed XML documents from the local system.
2. Map, normally with the help of cataloguers or librarians, fields from the local database to the fields of the AGRIS DTD.
3. Create an XSLT stylesheet is then used to encode the mapping document produced by the cataloguers.
4. Convert the well-formed XML documents in step 1 to AGRIS AP XML resources by means of an XSL processor.
5. Validate the generated XML documents against the AGRIS AP XML DTD by means of XML parsers.

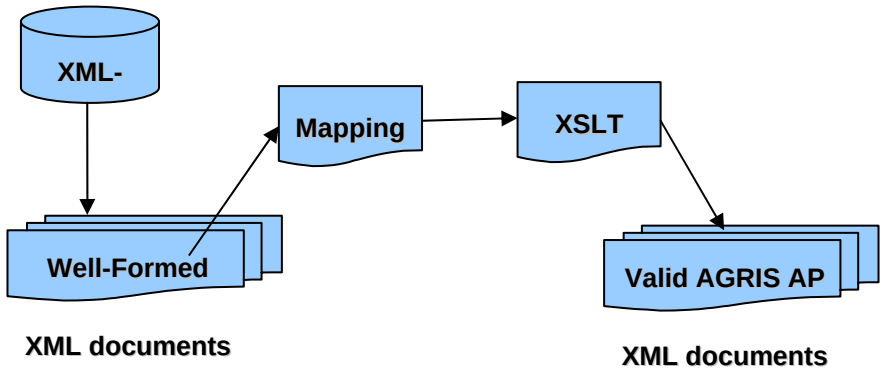

Figure 5: Converting from a proprietary XML enabled dataset to the AP

The next section describes the methodology adapted for developing the AGRIS Application Profile.

3.1.2 Methodology for the development of the AGRIS application profile (AAP)

The creation of the AAP involves several phases.

Phase 0. Definition of the project, its goal, and its scope

The first and foremost task is to specify the goals of the project, especially in terms of its short-term and long-term objectives. This allows the scope of the project to be defined in order to keep the work within its boundaries. In the case of the AGRIS AP project to develop an integrative information system for Agriculture, the following goals were identified:

1. provide a platform independent exchange format that can alleviate the semantic heterogeneity characterizing the resources provided by the 200-some information systems identified thus far;
2. do the groundwork to enable information service providers of agricultural resources to achieve digital information management standards for the next generation Semantic Web.

The solution for the first goal is clearly to provide a format, such as XML, that will not bind resource centres to any specific information system yet allow them to share their data, regardless of the platforms and technologies they are using. These resources can remain distributed and can use either Hypertext Transfer Protocol (HTTP) or Web Services[8] to perform multi-host searches, or they can be centralized to a single database. If they remain distributed, Web Services is recommended both for scalability and for automatic discovery of resources, especially for the future, when the adoption and implementation of semantic technologies (presumably) becomes

more widespread. The second goal, which depends on the first, as well as on the development of the agricultural metadata and subject ontologies (defined below in section 2), provides a means of converting data into machine-processable smart data, a prerequisite for transformation of the WWW into the Semantic Web.

As shown in Phase 1, a review of the resource types identified by the initial surveys of the agricultural information services indicated that in all likelihood, it will need to be extended to include not only document-like information objects[9] (DLIOs) but other entities such as institutions, individuals, and projects, and even services provided by collaboration software and texts generated by means of those services.

Phase 1. Assessment of the information objects

The major objective of this phase was to specify the range of resource types comprising the agriculture domain. Within the domain of Agriculture, the following resources have been identified thus far:

- internet portals, link collections, personal web pages, web pages of institutions and organisations
- databases: institutes, experts, literature, press articles, multimedia files, bibliographic data, projects, events
- publications: journals, newsletters, book excerpts, online texts
- collaboration software: discussion fora, calendars, event notification service, etc.

Once the actual resource types to be accounted for had been determined, for example, through user surveys, web logs, etc., each needed to be analyzed to determine the properties characterizing it. Such analyses established the initial requirements for specifying an application profile for the agricultural domain. Note that different types of resources will be described using different criteria.

For example, part of an adequate description of a book should include information such as its title and its identifier, which usually is expressed as an ISBN. An adequate description of a journal article should include not only the title of the article and its ISSN but also the title of the serial. In contrast, to describe an individual, information such as the employing institution, the individual's title, research areas, and email address, might be deemed crucial.

Phase 2. Assessment of the existing metadata standards and creation of the AP

As shown above, different types of resources exist within the domain of Agriculture: DLIOs; non-DLIOs such as persons, institutions, events, and projects; and services provided by collaboration software. Because many of the resources are in fact document-like resources found in digital library collections, a natural starting point from which to create the AP was the set of elements, refinements, and schemes recommended by the DCMES. It is clearly defined yet shallowly scoped to serve the aim of wide applicability, i.e., cross-domain description, discovery and retrieval of information objects. It is also extensible in that additional elements, refinements or schemes may be added. However, this extensibility has to be controlled as it can be counterproductive to achieving the aim of interoperability.

Identify other entity types for which suitable metadata standards must be found or developed. For instance, to describe persons, the suitability of standards such as vCard and FOAF can be assessed.

Phase 3. Developing new properties

Because currently available metadata standards may not be sufficient to cover all of the needs particular to agricultural resources, rather than extend the current standards beyond recognition, a metadata element set specific to the domain, namely the AgMES, was developed to act as an umbrella namespace under which new elements that are deemed necessary for resource description in the domain of agriculture can be declared.

The initial step was to determine a set of properties needed to describe the different resources available in agriculture independent of any given standard. This process helped to decide if a particular property was really needed to describe the resource. Table 1 outlines the series of questions that were posed for determining the necessary properties.

Is the elements/refinement/scheme really required to support: • resource description? • resource discovery? • interoperability?

Table 1: Determining the necessary properties to describe a given type of resource

Then, with the properties evaluated as necessary (for description and searching), another sequence of questions, shown in Table 2, concerning each of these elements was posed, again in iterative fashion (Adopted from original guidelines from Stuart Sutton)

Once it is determined that the need for a given property exists, 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 EEMES element (and, if necessary, a scheme).

Table 2: Determining the need for a new element

This task of trying to match each property to an existing element, refinement, or scheme was meant to avoid reinventing the wheel. One consequence was that all declared elements, refinements and schemes in AgMES have ended up looking like a hodgepodge. To make sense of them, they need to be seen along with their DC "parent element". Two further steps were necessary for completing this phase.

(a) Provide the ISO/IEC 11179 metadata for each element, refinement and scheme in the AgMES namespace

Once the elements, refinements, and schemes were given entry into the AgMES, they were then described using the ISO/IEC 11179[10] 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.

The following ten attributes were used for defining the elements.

Attribute Name	**Definition**
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 appear in ISO/IEC 11179 and were modified to adhere to the terminology currently being used in the XML community. This approach was taken to facilitate the assimilation of this set into the XML and RDF communities.

Additionally, the following two attributes were also used.

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

(b) Create the data model of the AP

The next step involved taking each of the terms and defining them in the context of Agriculture. APs allow us to provide application specific definitions as long as they do not change the concepts themselves. For each element, we provided definition, cardinality, and data type information by giving some examples of best practice guidelines. These guidelines try to cover as many scenarios

as possible but are not exhaustive for practical reasons and suggest the use of schemes whenever possible; for example, the ISO639-2 scheme to indicate the language, when necessary. This process was applied to all the elements and refinements.

Phase 4. Create an XML DTD or Schema

The guidelines were then converted into an XML DTD which is used to validate all the XML-based inputs to the AGRIS Network. The XML DTD provides the following:

Logical structure of the record	The sequence and/or nesting of elements
Obligation	If a term is mandatory or optional
Cardinality	How often can this term appear in one record (0, 1 or more times)

Phase 5. Test the schema, and the application profile, with real data

The application profile was then made available as both a document and also as an XML DTD, which was necessary for validating XML inputs. The guidelines were then applied by a test information provider for subsequent refinements of both the document and the DTD.

The technical implementers, i.e., those who would be responsible for converting their proprietary databases to the AP format, were provided with documentation on how to handle the conversion. Each implementer was given one-to-one feedback to help them successfully implement the exchange standard.

3.2 AGRIS Application Ontology

3.2.1 Definition and rationale

An ontology is a shared model of a given domain whose basic components consist of a vocabulary of terms, a precise specification of those terms, and the relations between them. Although an ontology has a structure similar to that of a taxonomy, the real power of an ontology comes from the ability to go beyond the information encoded in the structure to generate new information

through inferencing. Using an ontology creates a separate knowledge layer distinct from any local information technology, information architecture, or application. It is more scalable than traditional methods of integration, where fields from separate data sources are mapped to each other. In traditional methods, the addition of a single database to be mapped to *n* databases requires *n* mappings from each field in the new database to each corresponding field(s) in the other *n* databases. Moreover, drawing the correspondences between fields from the new database to those in the others requires an understanding of the semantics of each field in each database. Thus, the task of integrating every new database to the system, or indeed, making a change to any one of the databases, becomes more and more unwieldy, increasing by an order of magnitude the number of mappings to be carried out[11]. However, when the knowledge layer is abstracted away from the details of a specific application, each new system has only to perform a single mapping in order to communicate with the other systems. This facilitates management of and communication among otherwise heterogeneous systems.

3.2.2 Mapping the application profile to the application ontology

By definition, a standardized metadata element set consists of uniquely defined concepts that are in specific relations to each other. Whether explicit, as in the relationship between translations of corresponding resources, e.g., the ags:isTranslationOf element, or implicit, as in the relationship between a resource and its file name, the semantics of those elements can be expressed via an ontology.

The Agriculture Application Ontology (AAO), is the root ontology of the system. The representation of resource metadata elements as an ontology is motivated by the recognition that, as far as a resource metadata is concerned, the normally underexploited semantics existing between extrinsic descriptors of resources could be used to enhance the user's information retrieval/knowledge acquisition experience. For example, nearly all bibliographic metadata contain

the following assertions.

[resource] dc:creator *v*[dc:creator] or [resource] dc:subject *v*[dc:subject]

where *v* represents the *value* of the property.

A simple but useful inference that can be drawn from these assertions is

v[dc:creator] hasWrittenOn/hasPapersAbout dc:subject

An application such as a search engine could make use of such meanings not asserted by the metadata or the resource (e.g., to make suggestions to the user, to enhance the user's learning experience, etc.). Yet, rarely do bibliographical information retrieval systems take advantage of the ability to make these kinds of inference.

Other metadata standards describing other types of resources (e.g., events, experts, etc.) are treated analogously.

Figure 6 depicts the three-tiered organization outlining the relationships between the resource, the metadata elements from, in this case, the AP, and then the application ontology.

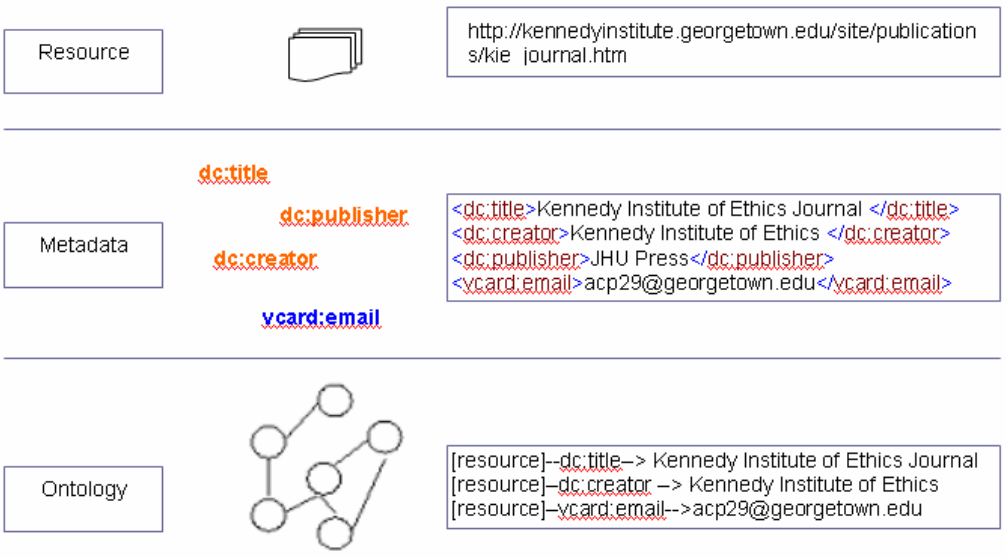

Figure 6: The three-tiered relation between Resource, Metadata, and Ontology

The resource metadata is marked up using the AP. The AP elements and the relations between them correspond to concepts and relations in the application or root ontology that make it possible to generate new information through the extraction of inferences. For instance, a search for a certain journal could also yield, by inference, the email address of the institution responsible for that journal. Indeed, in the future, with the use of the W3C OWL standard, this inference could be made even if the information about the email address and the journal were on different websites.

Query: search for “Food, Nutrition and Agriculture” Journal

Inference: contact email of the “Food, Nutrition and Agriculture” Journal

In the domain of food, nutrition and agriculture, FAO has developed a multilingual metadata ontology containing few concepts (corresponding to the metadata elements), some relationships between them (such as “has_author”, “publication_date”, “has_subject”, etc.), and many instances which correspond to the metadata records of a bibliographical database.

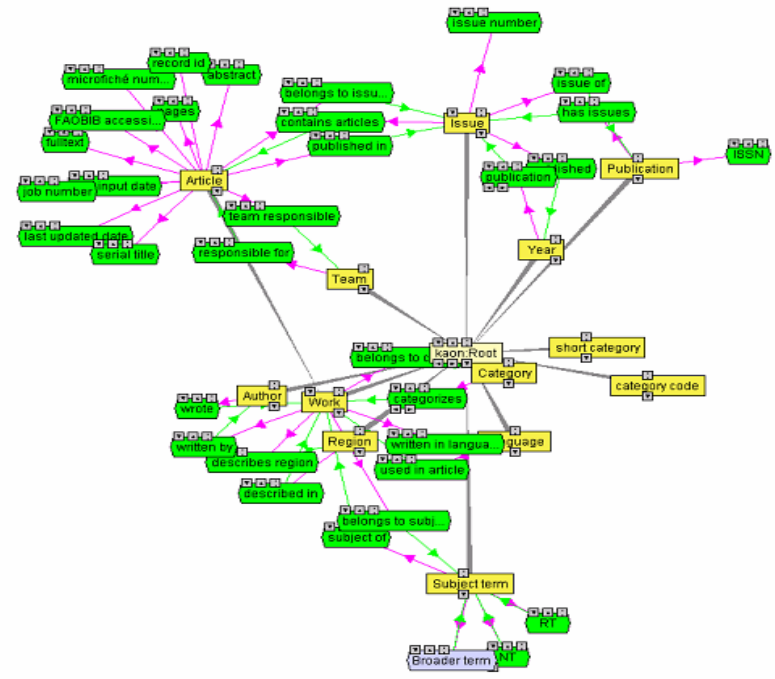

Figure 7: Metadata ontology for the Food, Nutrition and Agriculture online catalogue

3.2.3 Extending the AAO

Using the AAO as a root ontology to express the semantics of metadata descriptors (e.g., title, creator, publisher) considerably enhances the value of the resources that are described using them for the reasons mentioned above. However, the ontology can be further developed. On the one hand, subtypes of concepts already existing in the root ontology can be added. For instance, the dc:title concept subsumes the sub-concept dcterms:alternative. The root ontology can also be extended through the incorporation of controlled vocabularies. These vocabularies may simply consist of a flat list of terms, such as language codes. When the controlled vocabulary has some explicit semantics, as does a thesaurus, it lends itself to realization as a sub-ontology. These extension types are elaborated in the next two sections.

3.2.3.1 Addition of sub-concepts

As mentioned previously, Dublin Core, whose semantics is being used as the basis of the AAO root ontology, was deliberately designed to be shallow. This shallowness allows for flexibility in its applicability to the specific needs of a given domain or application. Within Agriculture, the following sub-concepts have been identified:

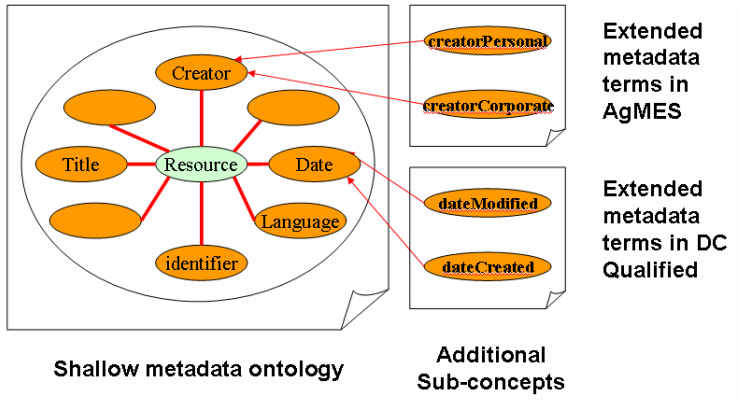

Figure 8: Additional Sub-concepts[12]

3.2.3.2 Controlled vocabularies

Controlled vocabularies are terminologies consisting of a set of terms and associated meanings that have been standardized for describing and searching resources. They often represent the intellectual work of experts and/or standards bodies that can and should be reused to avoid reinventing the wheel and to increase the possibilities for interoperability. This paper is concerned with controlled vocabularies that can be used as valid metadata values and those with rich(er) semantics useful for development of sub-ontologies.

3.2.3.2.1 For valid values

Their main distinguishing features are that they tend to occur as a flat file containing standardized names or symbols. They extend the root ontology inasmuch as they supply a list of valid values for specifying resource attributes. Examples include language codes, identification types for bibliographical resources, etc.

3.2.3.2.2 For sub-ontologies

Vocabularies that contain rich semantics are often accompanied by prose definitions where the semantics is implicit, i.e., interpretable exclusively by humans, as in a glossary or dictionary, or where the relations among terms or concepts are (more) explicit and thus (more) amenable to machine processing, as in a taxonomy. In contrast to our discussions thus far on the AAP and the corresponding AAO, where we have been concerned with the extrinsic properties of resources (e.g., title, author, publication type), these kinds of vocabularies tend to describe the concepts and relations that make up a given domain, that is, those that describe the content of resources. Vocabularies such as thesauri are a good starting point for ontology development because they

already are to some degree of machine readable. With regard to the root ontology, they serve as a sub-ontology attaching to the root via the dc:subject concept.

Several vocabularies may exist that are of relevance to the domain. That is, one provider might use *gmo crop* while another might use *novel food* to refer to the same concept. The subject sub-ontology can act as a mediating structure for multiple thesauri within the same or overlapping domains. Because it is concept- and not string-based, terminologies can map their specific terms to the corresponding concepts within the ontology. Further, with the help of domain experts, relations can be drawn between each uniquely defined concept. Consequently, providers can maintain the use of their terminologies while also being semantically interoperable with other vocabularies by integrating them based on a common semantic structure that can specify both terminological relationships (such as synonymy) and taxonomic and other semantic relationships (such as part-of).

3.2.4 Methodology for ontology-building

3.2.4.1 AGRIS Application Ontology

The AAO is based on elements constituting the AAP. Indeed, an ontology already exists for the Dublin Core Metadata Element Set (cf. Kamel-Boulos *et al.* 2001) that can be used both as a model and as a starting point for the construction of the AAO. Therefore, the construction of this ontology should not create significant problems. Figure 9 shows the correspondences among the concepts derived from resource, the AAP, and the AAO.

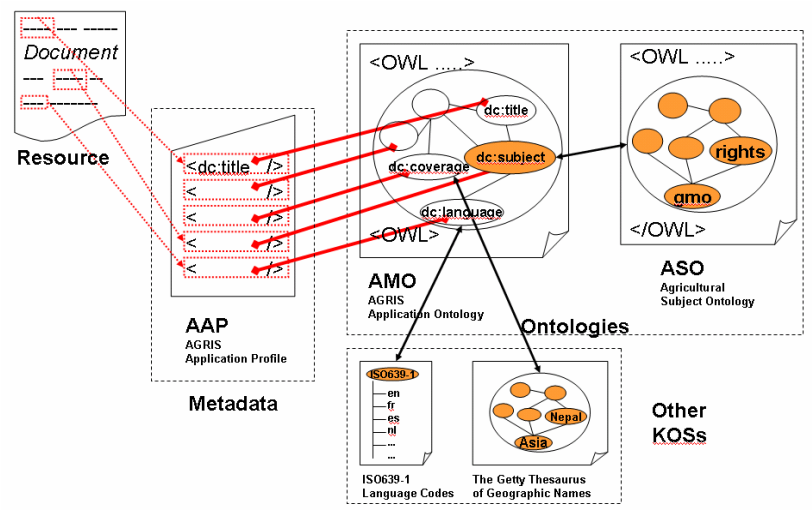

Figure 9: Corresponding concepts in the resource, application profile, and application ontology

3.2.4.2 Sub-ontology: Agricultural Subject Ontology

This involves at least two strata of ontologies: the core domain and component sub-domains.

Phase 1. Gather and characterize existing terminological resources in the domain.

In keeping with the principle of reuse (and in the service of interoperability), the first step is to identify the lexical resources that can furnish the raw materials, i.e., terms and meanings, from which to build the ontology. These lexical resources may involve semantics of varying degrees of explicitness (e.g., a word list only identifies concepts without definitions or relations; a taxonomy has some semantics expressed through terms connected via a hierarchy), that may or may not be machine-interpretable (e.g., a glossary is intended for human interpretation; a database scheme can be "understood" and used by a computer). They include glossaries, wordlists, thesauri, taxonomies, subject classifications, XML DTDs, and database schemes as well as ontologies. Figure 10 shows how these resources fall along a continuum, according to the explicitness of their semantics and their amenability to machine interpretation.

Figure 10: Terminological resources on a continuum of semantic explicitness. (Based on McGuinness 1999.)

The degree to which the resource covers the domain in question, as well as the sub-domains covered, should also be assessed. For instance, a dedicated Aquatic Sciences and Fisheries Thesaurus[13] (ASFA Thesaurus) would obviously be relatively coextensive with the domain whereas a thesaurus such as AGROVOC, which is a general agricultural thesaurus, would contain only parts, scattered throughout the thesaurus that were relevant. Other information that should be determined are:

- Number of concepts/terms
- How concepts and/or equivalence classes are identified
- Semantic relations, hierarchical and associative (e.g., RT)
- Number of top-level terms
- Depth of trees
- Classes v. individuals
- Annotations

Thus far, in agriculture, over 40 terminological resources[14], whose content is of varying degrees of relevance to the domain, have been identified. These include, among others the AGROVOC Thesaurus, the NAL Thesaurus from the National Agricultural Library of United States, the CAB thesaurus.

With the help of subject matter experts (SMEs), those parts that are relevant to the domain and need to be incorporated into the ontology would need to be identified. Based on the findings in phase 1, the next phase can be implemented.

Phase 2. Analyze data models

This phase requires analysis of each of the individual terminologies to establish the correspondences to be made to the data model of the ontology. Thus, for example, for thesauri, the following correspondences hold:

- terms are treated as strings not concepts
- concepts correspond to classes and are not equivalent to subjects, topics, or domains
- BT/NT are converted to superclass/subclass relations
- RT is generalized to top-level conceptual relations (in fact concept-to-concept relations will be represented in a hierarchical manner; see Fig. 9)
- An individual is distinct from and a member of a class
- USE/UF may or may not correspond to synonymy relations

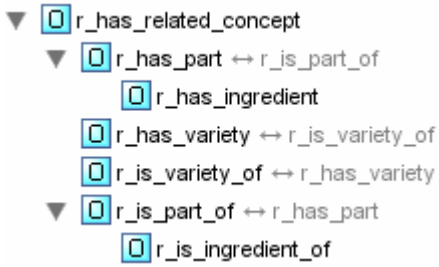

Figure 11: Hierarchical organizations of concept relationships

Note that other terminologies may have other correspondences, e.g., in an XML DTD, elements may correspond to concepts; in a glossary, each term might correspond to a concept while relations to other terms or concepts might be derived from informal definitions.

Phase 3. Convert mapped data models into OWL

This stage involves the transformation of the correspondences made in Phase 2 to a knowledge representation language. Each transformation should retain information about the source terminology. For instance,

AGROVOC Thesaurus: *fires* ==> <owl:class rdf:id="fires@agrovoc">

NAL thesaurus: *fires* ==> <owl:class rdf:id="fires@NALThes">

Phase 4. The core subject ontology: Capturing knowledge from Subject matter experts (SMEs)

In this phase, SMEs are given a set of key questions or use cases to identify fundamental entities, roles, components, functions, and relations for building the ontology. The objective of this stage is to specify the domain-specific concepts and relations at the highest level of abstraction. The ontology that is developed at this stage then can serve as the foundation for the hierarchies identified and extracted in the next phase.

Phase 5. Identify hierarchies within terminologies

Once the resources have been identified, they need to be classified according to the degree of explicit structure contained in the resource. Terminologies containing hierarchical structures can be (re)used to build the structure of the ontology while those with semantics meant for human interpretation such as glossaries can serve to provide synonyms and annotations.

Phase 6. Alignment

These top terms along with their hierarchies are then aligned to the core domain ontology created in Phase 4. If a corresponding class does not exist, and the term is pertinent to the domain, enrich the relevant part of the core domain ontology to create a place for alignment. Figure 12 shows a graphic of how alignment is done.

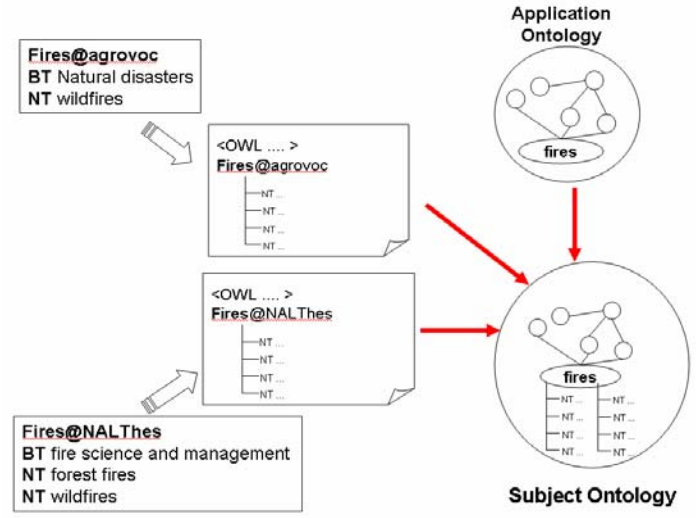

Figure 12: Aligning hierarchies from different terminologies to the core domain ontology

Phase 7. Merging

Merging is the process of integrating corresponding concepts from the source terminologies. For example, the concept {climatic change} is homonymous in the AGROVOC, CAB Thesaurus, and NAL Thesaurus and UNBIS Thesaurus, to the concept {climate change}, i.e.,

AGROVOC[15]: climatic change

CAB Thesaurus[16]: climatic change

NAL Thesaurus[17]: climate change

UNBIS Thesaurus[18]: climate change

Thus, the four sources are realized in the core domain ontology as lexicalizations of the same concept. Tools[19] are available to help SMEs with this process.

In other cases, homonymous terms may refer to different concepts. In AGROVOC, *euthanasia* refers to *putting animals to death*, while in the CAB Thesaurus, the context in which it occurs suggests that it refers to the *putting a human to death*.

AGROVOC:
Euthanasia
USE Destruction of animals

CAB Thesaurus:
Euthanasia
UF: mercy killing
RT: health protection; pain

For synonyms referring to the same concept, e.g., *GMO crop* and *novel food*, a SME is required to

make the determination.

Phase 8. Enrich through annotations

Multilingual labels, synonyms, comments, identifier numbers can be mapped to the concepts.

Phase 9. Post-processing

Once alignment and merging has taken place, the resulting structure should be checked for inconsistencies. The exposure of such inconsistencies is facilitated using tools[20] built for such purposes.

XThes: DisjointClasses (a:female a:male)

XThes: Class (a:Sam partial a:female)

XThes: Class (a:Sam partial a:male)

➔ Inconsistent

4 Scenarios: Relating ontologies and data

Thus far, the emphasis has been placed on the use of the AO for retrieving information from structured metadata repositories. In this section, the discussion is broaden to the relationship between ontologies and data in general. That is, we show that it is possible

(1) to make use of the AO to search structured data that has not been indexed;

(2) to make use of the AGRIS Application ontology to search unstructured data;

(3) to make use of a semantically enriched AO at various stages of complexity.

The continuum depicted in Figure 10 shows the kinds of resources that can be incorporated into the AO to enhance the latter's semantic richness. The scenarios will be described using vocabularies taken from different locations on the continuum.

We first outline the main architectures within which the AP and AO may be contextualized. These vary along two dimensions, namely, the centralization/distributedness of the AO, on the one hand, and of the data repositories described by the AO, on the other. Then we show that while the development of semantic structures is an ongoing process, each stage can be used as the basis for solving many of the problems currently plaguing heterogeneous domain-specific information retrieval and knowledge management systems. The richer the semantics, the more possible it becomes to develop "smart" applications that not only increase the effectiveness of IR/KM but also enhance the user's experience in the process of search and retrieval.

The overall architecture integrating the AP and the root ontology is depicted in Figure 13. However, there are two principal architectures that combine these two components. Owners of distributed databases map records to and exposing content in the AP format. This metadata is sent to a centralized database or it is made available on individual websites. The former is called the mediator approach while the latter is the federated multi-host approach. Both metadata and subject ontologies are made available in a pre-defined location, although in the future, they may be located in a registry of ontologies and accessed via web services.

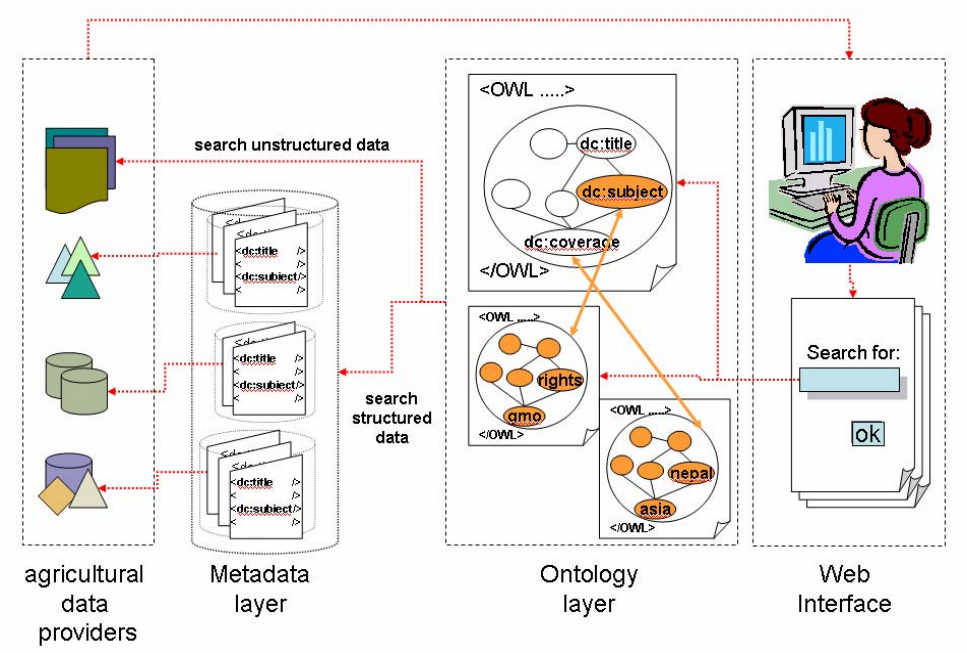

Figure 13: Integrating the ontology layers with the metadata layers

4.1 Principle architectures

In Figure 4, we showed graphically the role of the AP as a layer mediating, on the one hand, between heterogeneous, distributed datasets and, on the other, applications making use of the information contained in those datasets. Once owners of distributed databases have mapped their records to and exposed their content in the AP format, this mediating layer may be realized in one of two ways. Either the resulting metadata is sent to a centralized database or it is made available in individual databases that are accessible to multi-host searching. In either case, the corresponding AO will be made available in a pre-defined location so that the metadata vocabulary can be interpreted (i.e., ascribed meaning). In the future, the AO in its entirety may be stored in a registry of ontologies and accessed dynamically via Web Services technologies. However, the use of OWL to describe the AO will allow distributed storage, maintenance, and enhancement of the ontology. Applications based on the development of this AO will see it as a single ontology.

4.1.1 Mediated

In the mediated approach, all metadata is stored in a single centralized location. Queries are first pre-processed (e.g., parsed, spellchecked, normalized for singular/plural forms) and then interpreted via the AO, where they are resolved to concepts or instances. These concepts or instances are then matched to the relevant fields within the central database. Note that, as mentioned, the AO itself may be centralized or distributed.

4.1.2 Federated

In contrast, in a federated architecture, databases are stored locally and made available for WWW access. As in the mediated approach, queries would undergo pre-processing and interpretation via the AO. But rather than conducting a search on a single database, the interpreted query would be sent via web services to databases hosted on multiple distributed servers and a search executed on each of those databases.

Thus, there are two dimensions to the building of the architecture: the centralization (or not) of the data, i.e., resources, and the centralization (or not) of the domain knowledge that describes those resources.

4.2 Four sample scenarios

4.2.1 The AO and unindexed structured data

In the first scenario, the AO is exclusively based on the AP, and consequently, consists only of concepts describing resources. That is, it is the root ontology without any further extensions. In this case, relatively little analysis is required for the development of the ontological structures involved. The concepts concern only those used to describe the extrinsic properties of resources. Controlled vocabularies or lists may provide values for the attributes of the resource, e.g., language, keyword,

etc., and in such cases, should be specified; however, they are not essential for this scenario.

In and of itself, the AP already resolves, in a substantial way, the issue of semantic heterogeneity and the associated problems of maintaining interoperability among multiple distributed data repositories. However, further explicating the semantics of the AP in the form of an ontology provides the foundation for the development of semantically driven applications. For instance, an information retrieval application based on the AO could have the following features.

- The user could search resources by resource type, author, year, publisher, etc.
- The user could issue complex queries, such as checking to see if a given author wrote any articles written within a given time period.
- The system could present the user with information related to her query, such as titles associated with a particular author, their frequent co-authors, etc.
- The user could learn about resources and their properties, e.g., discovering that the same individual who led a project promoting a certain technology also wrote a paper arguing against it, by traversing the links in the ontology.

4.2.2 The Subject Ontology and unstructured data

The second scenario comprises the use of the agricultural subject ontology (SO) and one or more unstructured data repositories. In this case, the SO is used primarily to help the user form effective queries using the terms and relations in the ontology for query expansion. The actual query expansion that is realized depends on whether the data repository is domain-specific, or if it is a general one such as the WWW. The richer the vocabulary (synonyms and translations), the more effective the search. For instance, if the user is looking for information on BSE within domain-specific repositories, and he issues the query term BSE, the query would be expanded to include all of the synonyms for that concept. This ensures the greater recall of resources.

Domain-specific context
User's Query: BSE
Expansion: BSE OR bovine spongiform encephalopathy OR mad cow disease

Because the search context is limited to the domain, the problem of false positives is reduced. That is, in a repository of information on, say, bioethics, the string BSE is more likely to refer to the *bovine spongiform encephalopathy* concept than the *Bombay Stock Exchange*.

Within general data repositories, the expansion strategy may be slightly different. In this case, the abovementioned query would be much more likely to retrieve false positives. To minimize ambiguity, the query would then be expanded to include the parent concept, with which it would be combined with the Boolean AND.

General context
User's Query: BSE
Expansion: (BSE OR bovine spongiform encephalopathy OR mad cow disease) AND (disease OR syndrome OR disorder)

By expanding the query to include the parent term, which disambiguates the term, false positives arising from the ambiguity of the original query are reduced.

When the contents of a repository are determined (automatically or manually) to occur in a particular language, the terms in the appropriate language can be used to expand the query in the same manner as described above. When the repository is general and multilingual, as is the WWW, a simple OR query consisting of all synonyms and translations corresponding to the user's query could be issued. Thus, in addition to the functionalities such as aided query formulation through query expansion and multi-lingual search, the SO also enables the user to learn the domain vocabulary as well as the domain itself and search related terms.

4.2.3 The AO and the use of one or more thesauri

The third scenario comprises the root ontology with one or more thesauri containing the usual BT/NT/RT/UF relations. In contrast to the previous scenarios where semantic relations exist only between metadata descriptors but not between terms within controlled vocabularies, or only

between concepts specific to the agricultural domain, in this scenario, the controlled vocabularies supplying values to one or more of the concepts, usually **dc:subject**, contain some explicit semantics. Although the terms and relations within thesauri are often ill-defined and cannot be used for reasoning, they nonetheless contain some machine-readable semantics that can be exploited for the development of intelligent applications.

An information retrieval application based on this extended ontology can have, in addition to the features described above in the previous two scenarios, the following:

- Resources containing terms related to those in the user's query, as well as the terms themselves, could be displayed.
- If the user enters a query, she receives a list of results containing the metadata for the articles associated with those keywords. Clicking on an author's name retrieves all the resources containing those same keywords.
- The system can help the user find the information she is searching through a series of questions that filter through the information, e.g., what resource type? (e.g., author), wrote what publication type? (e.g., introductory text), about what? (e.g., keywords k1, k2, and k3), when? (e.g., between the years y1 and y2), in which language? (e.g. language l).

4.2.4 The AO and the use of sub-ontology

The fourth scenario is the most complex. It consists of the root AO containing concepts corresponding to all the resource types. Where a given concept is associated with multiple controlled vocabularies, those vocabularies are integrated. If they consist of flat lists, a list akin to an authority file can be incorporated into the ontology. If they contain a more complex structure, a sub-ontology can be developed that integrates the different terminologies (cf. Section 3.2.4.1.1). Moreover, in this scenario, the knowledge itself may be distributed. Thus, parts of the root and sub-

ontologies may be stored on distributed servers. Access to these distributed parts may be made via URL references or through web services.

All functionalities described in the previous scenarios are possible. Because the semantic structures are truly concept-based, other features such as cross-language information retrieval, terminology brokering across multiple databases using different terminologies, and intelligent query expansion become possible.

In addition, more sophisticated applications can be developed. For instance, a customizable information delivery system can filter information for people needing to monitor and assess large volumes of information. The volume of targeted information is reduced based on its relevance according to the user's "need to know." In a real-time monitoring system consisting of online RSS news feeds, the user could enter parameters of interest (i.e., concepts). A change or update in information that conforms to those parameters (i.e., that contain those or child concepts) could trigger an alert. In a well-designed robust ontology, information can be filtered independently of language or specific terms used.

A robust ontology also serves as the basis for automatic indexing of texts at multiple levels of granularity. Based on the usual statistical analyses of term frequencies, terms can be resolved to corresponding concepts in the AO (and indeed to those in corresponding vocabularies). Documents can then be tagged with those concepts. At the content level, semantic tags can be provided to allow resources to be "better known" by one or more systems so that search, integration, or invocation of other applications becomes more effective. Tags are automatically inserted based on natural language analyses of texts.

5 Conclusion

We have seen how semantic standards promise to be an effective approach to resolving the problems posed by semantic heterogeneity and how they can be the source of applications that help users find and discover information efficiently and effectively. The development of semantic technologies is an ongoing process, whereby any given stage can be the source of application development. The richer the semantics, the greater the possibilities for developing smart applications.

Contrary to popular belief, it is not the case that large expenditures of time and effort are necessary to develop nor to enjoy the advantages of semantic technologies, nor is it the case that structured indexed data are necessary to realize the benefits. As we have shown, even a small investment in the enhancement of relations between vocabularies, both metadata and domain-specific, yields a relatively large return on investment. We have shown several scenarios of varying complexity that enable information providers within the agricultural domain to exploit semantic technologies to provide information effectively and allow their users to access it easily.

REFERENCES

Berners-Lee, T. *Semantic Web on XML*. Keynote presentation for XML 2000. Slides available at: http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html. Reporting available at: http://www.xml.com/pub/a/2000/12/xml2000/timbl.html

Heery, R. and Patel, M. 2000. *Application profiles: mixing and matching metadata schemas*. Ariadne, N. 25, September 2000. Available at: http://www.ariadne.ac.uk/issue25/app-profiles/intro.html

Kamel-Boulos, M. N., Roudsari, A. V. and Carson. 2001. Towards a semantic medical web: HealthCyberMap's Dublin Core ontology in Protégé-2000. In *Fifth International Protégé Workshop, SCHIN, Newcastle, UK, July 2001*.

Lauser, Boris, Wildemann, T., Poulos, A., Fisseha, F., Keizer, J. and Katz, S. 2002. A Comprehensive Framework for Building Multilingual Domain Ontologies: Creating a Prototype Biosecurity. In *Proceedings of International Conference on Dublin Core and Metadata for e-Communities 2002:* 113-123. Firenze University Press. Available at:

http://www.bncf.net/dc2002/program/ft/paper13.pdf

McGuinness, Deborah L. 2002. Ontologies come of age. In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster (eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press.

Onyancha, I. Weinheimer, J. Salokhe, G. Katz, S. and Keizer, J. 2004. Metadata Exchange without pain: the AGRIS AP to harvest and exchange quality metadata, In *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC2004)*

Salokhe, G. and Le Hunte Ward, F. (Eds.), 2004. *Fifth Agricultural Ontology Service (AOS) Workshop Proceedings 27 - 29 April 2004, Beijing, China.*

Salokhe, G. Pastore, A. Richards, B. Weatherley, S. Aubert, A. Keizer, J. Nadeau, A. Katz, S. Rudgard, S. Mangstl A. 2004. FAO's role in Information Management and Dissemination – Challenges, Innovation, Success, Lessons Learned. In *Quarterly Bulletin of the International Association of Agricultural Information Specialists (IAALD)* 1019-9926, v.49 (= no. 3/4) p. 73-83. Available at: http://www.fao.org/docrep/008/af238e/af238e00.htm

Salokhe, G., Weinheimer, J., Bovo, M.G., Agrimi, M. 2003. Structured Metadata for Direct Resource Location: A Case Study. In *Proceedings of International Conference on Dublin Core and Metadata for e-Communities 2002:* 113-123. Firenze University Press. Available at: http://www.siderean.com/dc2003/404_Paper84-color.pdf

Soergel, D. Lauser, B. Liang, A. Fisseha, F. Keizer, J. and Katz, S., 2004. Reengineering Thesauri for New Applications: the AGROVOC Example, *Journal of Digital Information vol.4, n.4.* Available at: http://jodi.tamu.edu/Articles/v04/i04/Soergel/

Volz, R. Studer, R. Maedche, A. Lauser, B. 2003. Pruning-based Identification of Domain Ontologies. In *Proceedings of I-KNOW '03. Graz, Austria, July 2-4, 2003*. Available at: http://i-know.know-center.tugraz.at/previous/i-know03/papers/kc/volz.pdf

---

[1] For example, Google (http://www.google.org/) displays 55.5m results for *stem cells* and 88.6m results for *stem cell* with an overlap of 6 links on the first page of results for each query. [19/April/2006]

[2] The AGRIS Network http://www.fao.org/agris/

[3] Dublin Core Metadata Initiative http://www.dublincore.org/

[4] Agricultural Metadata Element Set http://www.fao.org/aims/agmes_elements.jsp

[5] Cf. the OWL Web Ontology Language Overview (http://www.w3.org/TR/2004/REC-owl-features-20040210).

[6] AGLS: http://www.naa.gov.au/recordkeeping/gov_online/agls/cim/cim_manual.html

[7] XSL Transformations (XSLT) Version 1.0 http://www.w3.org/TR/xslt

[8] Web services http://www.w3.org/2002/ws/

[9] A DLIO is a unit that is comparable to a paper document. The term is used to indicate resources such as websites, presentation files, photos, etc. but may not cover, for example, organizations or projects.

[10] ISO/IEC standard http://metadata-standards.org/11179/

[11]

$$n\_P\_2 = \frac{n!}{(n-2)!}$$

Where n = number of databases that want to share information with each other.

[12] AgMES extensions for DC elements http://www.fao.org/aims/agmes_elements.jsp

[13] Aquatic Sciences and Fisheries Thesaurus http://www4.fao.org/asfa/asfa.htm

[14] Knowledge organizations systems (KOS) currently in use within the Agriculture and related domains http://www.fao.org/aims/kos_list_type.htm

[15] AGROVOC Thesaurus: http://www.fao.org/aims/ag_intro.htm

[16] CAB Thesaurus http://www.cabi-publishing.org/

[17] NAL Thesaurus http://agclass.nal.usda.gov/agt/dne/search.shtml

[18] UNBIS Thesaurus http://unhq-appspub-01.un.org/LIB/DHLUNBISThesaurus.nsf

[19] Multiple Ontology Management Tools: Prompt (http://protege.stanford.edu/plugins/prompt/prompt.html) and Chimaera (http://www.ksl.stanford.edu/software/chimaera)

[20] For instance, the OWL plug-in in Stanford's Protégé tool is able to highlight logical inconsistencies in an ontological structure.