PROJECT PROPOSAL: *Reference:*

The Agricultural Ontology Server: A Tool for Knowledge Organisation and Integration

Food and Agriculture Organization of the United Nations (GILW)
Rome
June 2001

Contents

Executive Summary

At FAO, we are committed to helping combat and eradicate world hunger. Information dissemination is an important and necessary tool in furthering this cause—we need to provide consistent, usable access to information for users in places doing this very work. And, the wide recognition of FAO as a neutral international centre of excellence for agriculture positions it perfectly to lead in the development of system specific agricultural ontologies.

The Agricultural Ontology Server (AOS) will be instrumental in this effort by structuring agricultural terminology, thus making describing, defining and relating this information manageable for distributed facilities, and by standardising agricultural terminology, thus making resource access and discovery more efficient.

The AOS will function as a central common reference tool for serving ontologies. Itself an ontology using the AGROVOC thesaurus as its core, it will contain and serve terms, definitions of those terms and the relationships among those terms. It is designed to serve as a focal point for the vocabulary of the agricultural domain, and to codify and standardise the knowledge within this domain. It will serve common core terms and relationships, as well as richer relationships that designate it as an ontology.

The AOS will provide the “building blocks” needed to develop and maintain other ontologies. Knowledge systems will be able to interact directly with the AOS using a suite of ontological tools to perform resource description, relationship building, indexing and discovery functions, among others. Once ontologies are created for these knowledge systems, they can be used, enhanced and re-used by other knowledge systems. In effect, the AOS will provide modules of terms, definitions and relationships that can be shared among knowledge systems. The use of current standards and tools developed for this purpose will ensure efficient communication among these systems.

The proposed project to develop the AOS would:

1. Determine the types of users and tools that will use the server.
2. Utilise all the important knowledge organisation systems presently available and being developed.
3. Utilise current state-of-the-art encoding and interoperability standards.
4. Map and enhance relationships among terms to create ontologies.
5. Build in functionality to describe and find multilingual resources.
6. Create an agriculturally focused biological taxonomy to become an integral part of the server.
7. Develop and license authoring and maintenance tools and interfaces for use of the server.
8. Pilot test the server using specific user and application types, and further develop based on the results.

The existence of a common ontology server guarantees that common concepts are clearly defined by unique identifiers, basic relationships are developed, and that these elements are used and shared throughout the agricultural domain. The AOS is a huge step in the direction of sharing contextualised information. It will achieve this by defining and structuring the meaning of terms centrally, and providing access and re-use of this central core for ontology building in systems throughout the agricultural domain.

Introduction

Background and Definition

Because of the "information explosion," a currently trite phrase that fits our purposes here, users have found it harder and harder to retrieve information—particularly on the Web, where growth has been exponential and unmanaged. Within the agricultural community, the proliferation of documents and other resources has led to chaos. Users are unsure where to go to retrieve the resources they need and how to retrieve them once they get there. There are numerous, separately created knowledge systems that contain large numbers of resources, each with their own method for describing, defining and relating these resources.

The information science community has long known that terminology that is controlled improves information retrieval by standardising terminology and by providing a structure for languages.[1] Controlled terminology, known as *controlled vocabularies*, improves the finding of information by increasing recall—helping users find resources that use different terminology for the same concept—and by increasing precision—defining the structure of the terminology so users can understand the scope of resources to be found. A fully structured type of controlled vocabulary is a *thesaurus*, which includes hierarchical and associative relationships that add to the definition of scope.

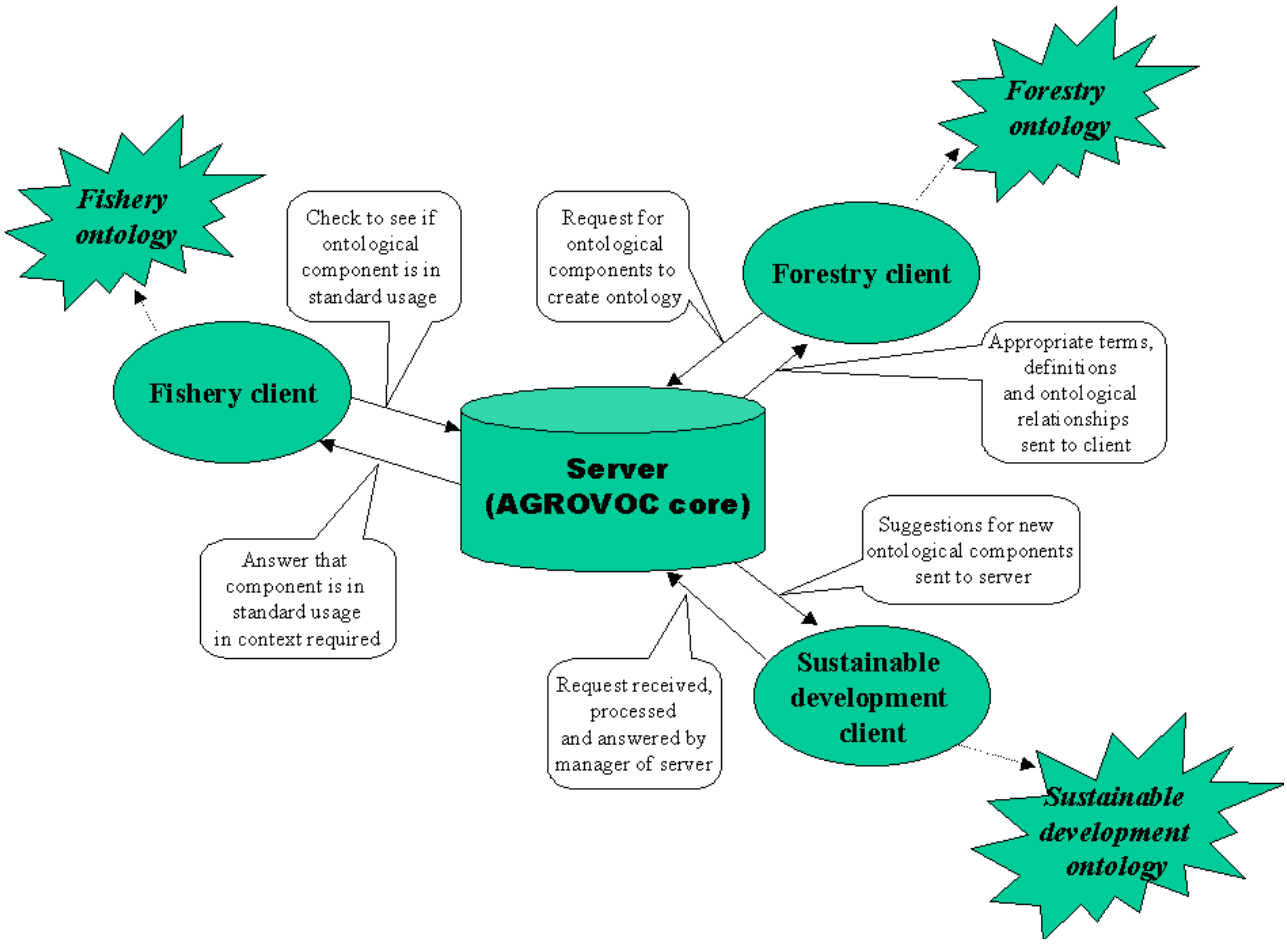
There are now opportunities to use these controlled vocabularies on the Web because of new standards that offer more power and flexibility. The advent of XML (eXtensible Markup Language) provides the ability to share knowledge across different tools, using a standard schema. The RDF (Resource Description Framework) standard allows storage and sharing of *metadata* (data about resources) across systems. The topic mapping language, XTM (XML Topic Maps), currently in development, may provide even stronger functionality for the use of metadata. These new standards allow us to leverage controlled vocabularies in the development of common methods for describing, defining and relating resources.

Briefly defined, the Agricultural Ontology Server (AOS) will function as a central common reference tool for serving ontologies. An *ontology* is a system that contains terms, the definitions of those terms, and the specification of relationships among those terms. It can be thought of as an enhanced thesaurus—it provides all the basic relationships inherent in a thesaurus, plus it defines and enables the creation of more formal and more specific relationships. The AOS, using the AGROVOC thesaurus as its core, is designed to serve as a central focal point for the vocabulary of the agricultural domain, and to codify and standardise the knowledge within this domain. It enables better communication within and across systems, and structures the meaning contained within systems.

The AOS provides the "building blocks" that assist in developing and maintaining other ontologies. It will contain the core vocabulary and definitions (multilingual) and the core relationships (including common richer relationships) which knowledge systems within the agricultural domain will use in building and maintaining their own ontologies. Knowledge systems, including forestry, fishery, plant biology, sustainable development and organic agriculture, among others, will use this reference tool to build their own ontologies. Once ontologies are created for these knowledge systems, they can be used, enhanced and re-used by other knowledge systems. In effect, the AOS will provide modules of terms, definitions and relationships that can be shared among knowledge systems.

The following illustration shows how the server can interact with these knowledge systems:

[1] Guidelines for the Construction, Format, and Management of Monolingual Thesauri, ANSI/NISO Z39.19-1993, p. 1.



The illustration shows how the AOS can assist clients in building their ontologies by:

- Mapping ontological components between existing knowledge systems and the AOS to determine overlap and, as a result, the amount of work required to build an ontology.
- Creating new ontologies, as in the “Forestry” illustration.
- Maintaining existing ontologies, by having updates of terms, definitions and relationships requested and sent from the server to the client on a regular basis.
- Checking on ontological components to ensure standard usage, as in the “Fishery” illustration.
- Suggesting new ontological components, as in the “Sustainable Development” illustration.
- Using ontological components created and suggested by other ontologies.

The Need

At FAO, we are committed to helping combat and eradicate world hunger. Information dissemination is an important and necessary tool in furthering this cause—we need to provide consistent, usable access to information for users in places doing this very work. And, the wide recognition of FAO as a neutral international centre of excellence for agriculture positions it perfectly to lead in the development of system specific agricultural ontologies.

The AOS will be instrumental in this effort by structuring agricultural terminology, thus making describing, defining and relating this information manageable for distributed facilities, and by standardising agricultural terminology, thus making resource access and discovery more efficient. Although there are organisations

who have developed ontology or taxonomy servers—for instance, Cycorp[2] and UMLS[3]—there are none for the agricultural community being developed by a content provider.
At the fundamental core of the server is the history of the explosion of information. As the amount of information has increased, we have needed more and better tools to help us manage, access and share it. Once, we had to use the same hardware in order to do this. Later, we were able to use different hardware, provided we used the same software. Now we can use different software, provided we structure the information exactly the same way. The next step is the development of structures that convey the different contexts of the information, but that can be communicated among systems. How can this be achieved?

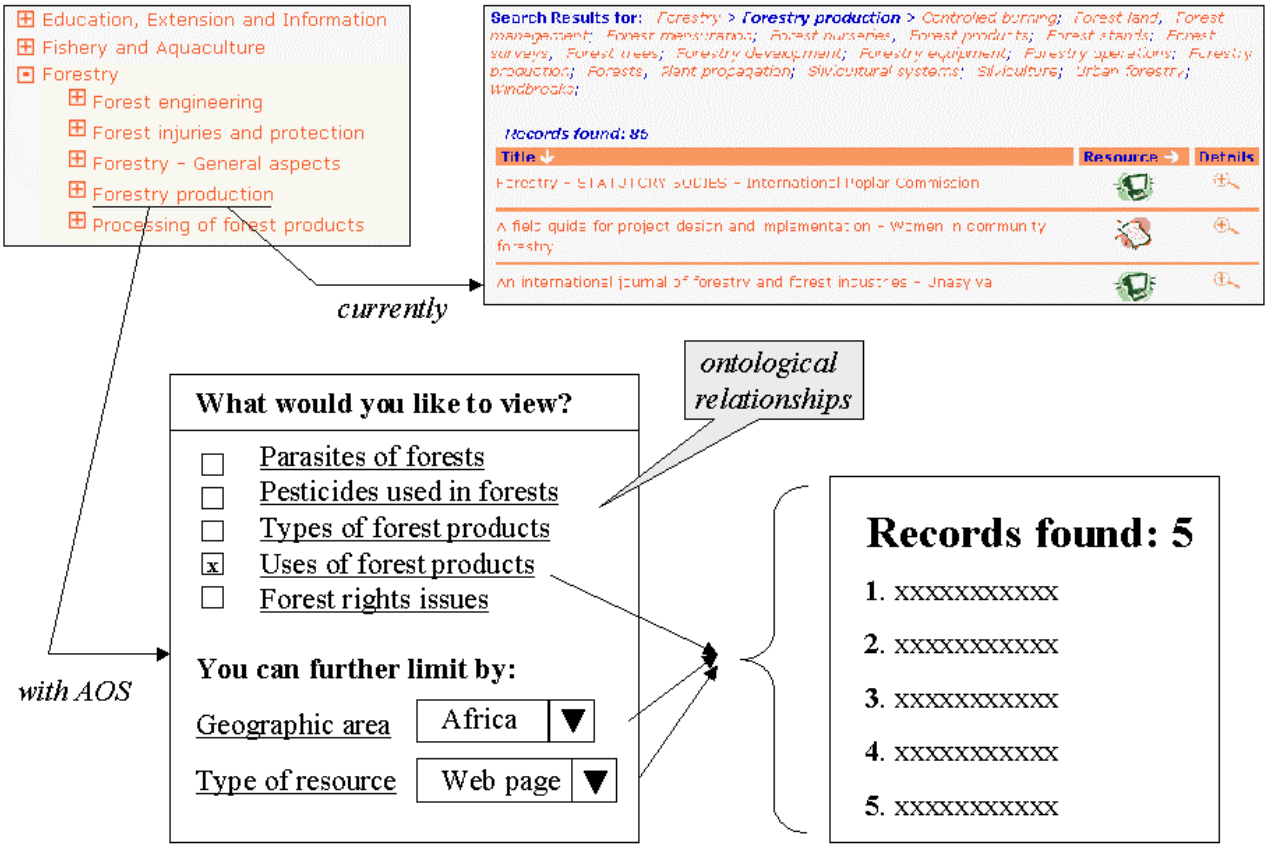
The AOS is a huge step in the direction of sharing contextualised information. It will achieve this by defining and structuring the meaning of terms centrally, and providing access and re-use of this central core for ontology building in systems throughout the agricultural domain.

Therefore, we need an AOS for the following purposes:
- To allow domain knowledge to be centrally defined and described
- To guarantee that common concepts and rich relationships are clearly defined
- To communicate among systems without semantic ambiguity
- To share the structure and meaning of agricultural information among users and tools
- To provide foundations to build other ontologies
- To enable re-use of domain knowledge
- To save time and effort re-deploying similar knowledge systems

And last, but certainly not least:
- To provide more effective dissemination and access to resources for users

The following illustration shows how the AOS will help these users find information more effectively than is achieved in the current system:

[2] http://www.cyc.com/overview.html
[3] http://umlsks.nlm.nih.gov/



In the current system, the WAICENT Information Finder retrieves for users all records related to a broad AGRIS/CARIS category. The proposed system uses new ontological relationships to filter the records by topics that are more useful to the user, and consequently retrieves a smaller, more relevant set of records.

The Proposal

The sum total of the AOS is to provide an effective method for knowledge system description, development, integration, sharing and dissemination. The proposed project to develop the AOS would:

1. Determine the types of users and tools that will use the server.
2. Utilise all the important knowledge organisation systems presently available and being developed.
3. Utilise current state-of-the-art encoding and interoperability standards.
4. Map and enhance relationships among terms to create ontologies.
5. Build in functionality to describe and find multilingual resources.
6. Create an agriculturally focused biological taxonomy to become an integral part of the server.
7. Develop and license authoring and maintenance tools and interfaces for use of the server.
8. Pilot test the server using specific user and application types, and further develop based on the results.

Each of these project points is discussed below.

1. Types of Users and Tools

We need to know up front who and what will be using the system.

By discovering who the users are, we will be able to develop the appropriate type of access. For instance, we need to understand if we are serving users at the local level as well as at the international level, whether experts and/or novices will be accessing the system, the level of detail to be sought in terminology, and what users need from the ontologies they will be creating.

It is also imperative to understand what types of tools will be interfacing with the AOS to assist users in building their own ontologies. Our server needs to be interoperable with a wide variety of tools—e.g., indexing, searching, metadata collection, vocabulary management—and the needs of these should inform the development of the server. To achieve this, we need to utilise appropriate interoperable standards, discussed below in project point number three.

2. Knowledge Organisation Systems (KOSs)

Since the server will be the central reference resource for vocabulary control and relationship structure of agricultural terminology, we will need to use a variety of knowledge organisation systems (KOSs) in developing the AOS to serve systems. Thus, the use of KOSs is two-fold: we will use KOSs to initially develop the server, and we will serve these and other KOSs in developing ontologies after initial development of the server. Some KOSs may eventually become integrated into the AOS because of overlap in terminology, while others will stand apart as separate ontologies.

Firstly, we will need to determine the relationships of these KOSs with the server. We can categorise the relationships as follows:

- core—e.g., fisheries, forestry
- adjacent—e.g., organic agriculture, sustainable development
- associated—e.g., biotechnology, ethics in agricultural research

Determining the relationship of the KOSs with the server allows us to understand the overlap in knowledge and the amount of effort required to work with the systems. Core systems are our first priority for utilisation in development of the server.

The primary KOS used will be the AGROVOC thesaurus, which already has the appropriate scope and basic relationships to serve as a base for the AOS. Other types of systems will include:

- classifications—lists of terms often using hierarchical relationships
- controlled vocabularies—controlled lists of preferred and variant terms based on concepts
- thesauri—controlled vocabularies containing hierarchical relationships
- authority files—controlled lists of preferred and variant names
- glossaries—lists of terms with definitions
- gazetteers—dictionaries of place names
- subject headings lists—broad categorisations of subject areas

A number of these KOSs are already available to us, although this is by no means a complete list:

Examples of KOSs	*Type of KOS*	*Subject Area*
AGROVOC	thesaurus	agriculture
AGRIS/CARIS Categories	classification	agriculture
WAICENT Information Finder	classification	agriculture
FAOTERM	authority list	agriculture
CGIAR	thesauri	agriculture
FIGIS	classification	fishery
oneFish	classification	fishery
ASFA	thesaurus	fishery
IUFRO SilvaVoc	thesaurus	forestry

...	...	...

Other possible sources include external KOSs such as CABI and NAL, and the development of partnerships with these bodies will be instrumental in the development of the AOS. Naturally, discussions with stakeholders of the KOSs are an integral part of the use of them.

To begin development, there will need to be a complete inventory of potential KOSs. Such an inventory should include an evaluation of each source to assess overlap in terminology and to determine the process for the mapping of terms, definitions and relationships between the core and the KOS.

3. Standards

The Semantic Web will allow machines to share information the way humans currently share information on the Web. The basic elements of the Semantic Web are the XML and RDF standards, discussed earlier. The syntax and schema behind these standards allows resource description and interoperability of these descriptions among systems.

We will need to incorporate current state-of-the-art standards to achieve two important tasks for the server to work effectively:

- availability of common terms, definitions and relationships to ensure standard use across systems
- opportunity for interoperability of those elements to enable sharing and communication among systems

To ensure the first task, the elements of the AOS will need to be encoded within the RDF framework. Common terms and definitions and their associated relationships from the core of the AOS will be identified by Universal Resource Identifiers (URIs) and stored in this common framework. (XTM is a parallel standard in development that may provide richer associations for better encoding.)

To enable the second task, the AOS will use XML language to communicate among systems for the exchange of the URIs to build ontologies. The systems interested in utilising the AOS will need to use this language to be capable of interoperability.

The conjunction of these standards will enable the communication of machine-readable commonly used URIs among a variety of different tools. In the case of the AOS, this type of communication will allow ontologies created by multiple tools—their terms, definitions and relationships—to be shared, evaluated and maintained using the central AOS.

4. Ontological Relationships

The relationships in a thesaurus are equivalence (USE/UF), broader term (BT), narrower term (NT) and related term (RT) relationships. These relationships provide structure for the terms. For instance, knowing that a broader term for “cereals” is “plant products” and that narrower terms are “maize” and “rye” provides a structure that defines the scope of information represented by these terms.

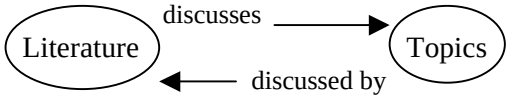
Recently, there has been considerable discussion relating to extending this core set of relationships. In the late 1990s, the American Library Association Subcommittee on Subject Relationships/Reference Structures examined over 165 relationships within the English language alone and from these produced a checklist of twenty candidate subject relationships for information retrieval.[4]

We can use a richer set of relationships to develop tools that provide more granular and more consistent indexing, and more effective searching and browsing for users. With ontologies we can more fully define these relationships—by creating rules for the development of these specific relationships—and thereby provide a means for better knowledge sharing.

[4] http://www.ala.org/alcts/organization/ccs/sac/rpt97rev.html

Each KOS that isn't integrated into the AOS will remain separate. Through communication with the AOS, they will build their own ontological relationships specific to their needs, which will result in the creation of their own ontologies. As a result of continuous sharing, development and maintenance, a variety of ontological relationships will be available for future ontological development.

An example of an ontological relationship is:



The following is a scenario describing how an ontological relationship is more powerful than combining terms using a Boolean AND:

> A teacher wants to inform her class about ticks and the diseases they bear. If she searches "ticks AND diseases" she retrieves all resources that contain those terms. She may end up retrieving resources that discuss the diseases that kill ticks. If instead, she searches "ticks" using the ontological relationship "bears_diseases" she limits her retrieval set to resources about the diseases that ticks bear.

The following is another scenario describing how an ontological relationship eliminates the need to do multiple searches:

> A researcher is interested in finding resources about the types of infestations of tomatoes. Instead of doing multiple searches for each type of infestation (e.g., "tomatoes AND tomato mosaic tobamovirus," "tomatoes AND fungal wilt") he uses the ontological relationship "infected_by." This relationship has been used in indexing each tomato infestation resource in his system. By using it, he saves himself work by doing a single search that ensures he will find all the resources he needs.

To effectively build this functionality, a list of potential relationships needs to be compiled (hopefully in conjunction with the inventory of KOSs). These relationships should be compared among KOSs, with the end result of a common set of relationships that can be used among KOSs. A subset of these will form the core relationships stored in the AOS. (The compilation of relationships will need to be compared with those being considered by NISO and ISO for inclusion in a future standard for electronic thesauri.[5])

5. Multilinguality

A key aspect of the AOS is that it will be multilingual. For users in all countries who need access to resources, we need to provide the ability to index and find information in any language needed. The AOS needs to collect and co-ordinate terminology, definitions and relationships in the five official languages of the FAO—English, French, Spanish, Arabic and Chinese. Additional languages should be added if necessary by those developing the ontology, if working in the mother tongue of the country is beneficial.

One of the biggest issues surrounding multilingual thesauri is the relating of terms and concepts from different languages. Often there is no one-to-one mapping of terms. The same will be true for the development of relationships. Concept-based mapping and cross-concordance (the fusing of clusters of similar concepts from different social groups into a new concept) will need to be incorporated into the AOS.

6. Biological Taxonomy

There needs to be a biological *taxonomy* included in the AOS, one that is appropriate for agricultural use. In this sense, a taxonomy is an ordered classification of living things detailing the relationships among them.

---

[5] http://www.niso.org/thes99rprt.html

This type of taxonomy should include both Latin and common names of plants and animals, structured appropriately, along with other richer ontological relationships. For example, creating the relationships "species_in" and "genuses_in" should provide more powerful searching for resources like those detailed in the scenarios above (e.g., using "lepidoptera" and "genuses_in" together).

The development of a biological taxonomy that is an ontology would be welcomed by the agricultural community for use in describing and indexing resources. Currently, there is no one organisation that has developed such an ontology, although several organisations have been developing standards for biological nomenclature (e.g., NBII[6], FGDC[7]).

7. Ontology Tools

We will need to develop a suite of ontology tools to be used for accessing the AOS and its set of ontologies. This suite should contain tools that allow:

- description—discovery of overlap in terminology and mapping of common terms and definitions
- relationship building—creation of ontologies using common relationships and building richer relationships
- encoding—storage of terms, definitions and relationships in a standard, interoperable format
- indexing—using ontologies to automatically and manually index resources
- discovery—searching and browsing by users in the AOS or in an ontology
- maintenance—ontology collection, storage, dissemination and evaluation by managers

For some of these tools, off-the-shelf products may be sufficient, in which case a set will need to be tested and evaluated for its possible use with the AOS. In certain cases, it may be possible to tweak the product to fit our needs, depending on the availability of staff to assist in this. Otherwise, if there are no off-the-shelf products, we will need to develop our own (preferably Web-based) tools using staff knowledge and skills.

8. Testing and Further Development

In addition to the tools themselves, appropriate interfaces need to be developed that address the needs of the users and systems that will be using the AOS tools. To build good interfaces, they need to be tested with users to find out if they suit their needs. Testing is best achieved through focus groups, one-on-one evaluation, observation, task analysis, and other commonly used methods. User profiles will need to be developed so the most relevant results of the testing are obtained—e.g., demographics, use of tools, level of subject knowledge.

It wouldn't be feasible to test the entire AOS initially. One of the KOSs should be used as a concrete example (e.g., the oneFish classification), mapping terms and developing relationships using the AOS, and then testing the result with users of that KOS. The results of this testing will make it possible to update and enhance the AOS functionality. Testing should be done iteratively throughout the lifecycle of the project to ensure that the AOS continues to be suitable for the needs of users.

---

[6] http://www.nbii.gov/datainfo/metadata/

[7] http://www.fgdc.gov/standards/status/sub5_8.html

Appendix A: Activity List

0. Initiate project plan.

- Build complete picture of purposes and uses of the AOS for description, sharing, integration, dissemination, discovery.
- Determine the specific functionality of the AOS to realise goals.
- Identify partners and stakeholders for the AOS system and their roles in relation to the system. Establish contacts.
- Sensitise stakeholders on reasons for use of AOS and its components. Develop initial marketing plan for use during implementation.
- Establish electronic communication (e.g., mailing lists) for exchanging ideas among partners involved in the development of the AOS.
- Ensure funding for development, testing and implementation.

1. Determine the types of users and tools that will use the server.

- Identify types of users who will be using the AOS (e.g., their level of expertise, their language skills) and their needs.
- Identify types of KOS tools that will be accessing the AOS (e.g., indexing, searching, metadata collection, vocabulary management) and their needs.
- Develop user and tool profiles. Use these profiles to prioritise development of the AOS.
- Summarise user goals and system related tasks in the new system. Develop use cases.

2. Utilise all the important KOSs presently available and being developed.

- Conduct a survey among core, adjacent and associated knowledge domains to determine what is available and being developed.
- As a result, inventory the existing and emerging KOSs.
- Determine mechanisms needed for communicating between the AOS and the KOSs.
- Decide which KOSs should be integrated into the AOS and which should stand alone to be developed as ontologies. (*In conjunction with step 4*.)
- Describe technical requirements needed to use the AOS. (*In conjunction with step 7*.)
- Analyse the AGROVOC thesaurus for necessary modifications.

3. Utilise current state-of-the-art encoding and interoperability standards.

- Conduct survey of suitable encoding standards.
- Determine the suitable set of standards for use in the AOS.
- Gather information on tools that use these standards. (*In conjunction with step 7*.)

4. Map and enhance relationships among terms to create ontologies.

- Evaluate potential KOSs for overlap in terminology and relationships.
- Determine what constitutes the core vocabulary (AGROVOC plus integrated KOSs).
- Determine the common set of relationships that are best served in the AOS.
- Develop a process for mapping common terms, definitions and relationships between the KOSs and the AOS.
- Discuss with stakeholders the types of new relationships that need to be developed.
- Identify and develop the new core ontological relationships that should be included in the AOS.
- Determine methods for maintaining the AOS (e.g., updating, suggestions handling). Include handling updates to and from satellite ontologies.
- Based on new core ontological relationships developed, decide on new organisational systems for presenting the AOS to users (e.g., topic mapping, developing new classification for the FAO web site). (*In conjunction with step 7*.)

5. Build in functionality to describe and find multilingual resources.

- Research and identify issues related to multilingual organisation systems (e.g., cross-concordance).
- Identify languages needed to satisfy our multilinguality requirement. (*In conjunction with step 1.*)
- Determine the sources for multilinguality (e.g., AGROVOC). Determine whether feedback to the source is allowed.
- Identify terms and relationships in the AOS that need to be translated.
- Develop a process for mapping multilingual terms from KOSs to the AOS. (*In conjunction with step 4.*)

6. Create an agriculturally focused biological taxonomy to become an integral part of the server.
- Analyse functionality of such a taxonomy for the AOS.
- Research developing standards for biological taxonomies.
- Establish contacts with those organisations responsible for these standards.
- Utilise those taxonomies that can inform development of our taxonomy (e.g., ITIS[8]).
- Develop proof-of-concept and present to contacts. (*In conjunction with step 8.*)
- Build the biological taxonomy.

7. Develop and license authoring and maintenance tools and interfaces for use of the server.
- Identify and evaluate off-the-shelf web based tools that are suitable for the development and maintenance of the AOS.
- Analyse areas that are lacking in the tools to consider customisation of the tools as required.
- Determine information architecture for the user interface(s). (*In conjunction with step 8.*)
- Test the tools with specific subject areas (e.g., fisheries). Re-evaluate options.

8. Pilot test the server using specific user types and application types, and further develop based on the results.
- Generate and assess feasibility of design for the AOS.
- Identify design constraints.
- Build prototype(s) to test as proof-of-concept.
- Decide on methodology for testing (e.g., card sorting, task analysis).
- Determine frequency of user testing, and when initial testing will take place.
- Determine what parts of the AOS (e.g., biological taxonomy, for indexing, with local users) will be tested when.
- Run several tests, using the results of one to inform the next (e.g., contextual inquiries can assist in creating prototypes to be tested with other users).
- Analyse results of user testing, and loop feedback into development of the design of the AOS.
- Implement the AOS when certain portions are ready to be released (e.g., indexing interface) in a reasonable time frame.
- Publicize the AOS.

[8] http://www.itis.usda.gov/access.html