The Agricultural Ontology Server (AOS)

A Tool for Facilitating Access to Knowledge

Contents

Food and Agriculture Organization of the United Nations (FAO)
Library and Documentation Systems Division
AGRIS/CARIS and Documentation Group
July 2001

1.0 Executive Summary

At FAO (Food and Agriculture Organization) of the United Nations, 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 the community of people doing this very work. The wide recognition of FAO as a neutral international centre of excellence for agriculture positions it perfectly to **lead in the growth and improvement of knowledge representation systems in the agricultural domain**, and to **lead in developing more functionality for users** looking for information.

We need to develop an Agricultural Ontology Server (AOS) that will function as a **reference tool that structures and standardises agricultural terminology in multiple languages for use by any number of different systems**. The AOS will serve the following purposes:

- increasing the efficiency and consistency of describing and relating multilingual agricultural resources
- decreasing the random nature and increasing the functionality for accessing these resources
- enabling sharing of common descriptions, definitions and relations within the agricultural community

Or, stated more simply, the purpose of the AOS is to achieve:

- **better indexing** of resources,
- **better retrieval** of resources, and
- **increased interaction** within the agricultural community.

The goals of the AOS are realised by assisting community partners in **building 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, more specific and more powerful relationships. An ontology captures and structures the knowledge in a domain, and by doing so captures the meaning of concepts that are specific to that domain. This meaning is then extended to end-users through the use of tools (e.g., indexing, retrieval) that apply the ontologies.

Our plan is to:

1. **Strategise about participation of agricultural community partners**.
2. **Utilise all possible knowledge organization systems within the agricultural domain**.
3. **Utilise current and developing state-of-the-art interoperability standards**.
4. **Develop formal ontological relationships among topics**.
5. **Build in functionality for describing and finding multilingual resources**.
6. **Create an agriculturally focused biological species micro-ontology.**
7. **Develop and pilot test development, storage, management and retrieval tools**.
8. **Test functionality of the server with end-users**.

Thus, the AOS will **provide terms, definitions and relationship components that can be shared** among associated partners, thereby enhancing communication and interaction within the community. Use of these components **increases the functionality for indexing and retrieving resources** by providing a standard source for terminology and offering richer, more powerful ontological relationships.

2.0 Introduction

2.1 Purpose and Need

At FAO (Food and Agriculture Organization) of the United Nations, 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 the community of people doing this very work. The wide recognition of FAO as a neutral international centre of excellence for agriculture positions it perfectly to **lead in the growth and improvement of knowledge representation systems in the agricultural domain**, and to **lead in developing more functionality for users** looking for information.

Users are finding it **harder and harder to retrieve information**—particularly on the web, where growth has been exponential and unmanaged. Users are unsure where to go to retrieve the resources they need and how to retrieve them once they get there. There are **numerous, independently 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.[i] 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.

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. But, not all systems use the same kind of structure. We need to be able to **communicate different information structures among a community of systems**.

How can we develop a structure that is controlled, yet allows differences among systems?

We need to develop an Agricultural Ontology Server (AOS) that will function as a **reference tool that structures and standardises agricultural terminology in multiple languages for use by any number of different systems**. The AOS will serve the following purposes:

- increasing the efficiency and consistency of describing and relating multilingual agricultural resources
- decreasing the random nature and increasing the functionality for accessing these resources
- enabling sharing of common descriptions, definitions and relations within the agricultural community

Or, stated more simply, the purpose of the AOS is to achieve:

- **better indexing** of resources,
- **better retrieval** of resources, and
- **increased interaction** within the agricultural community.

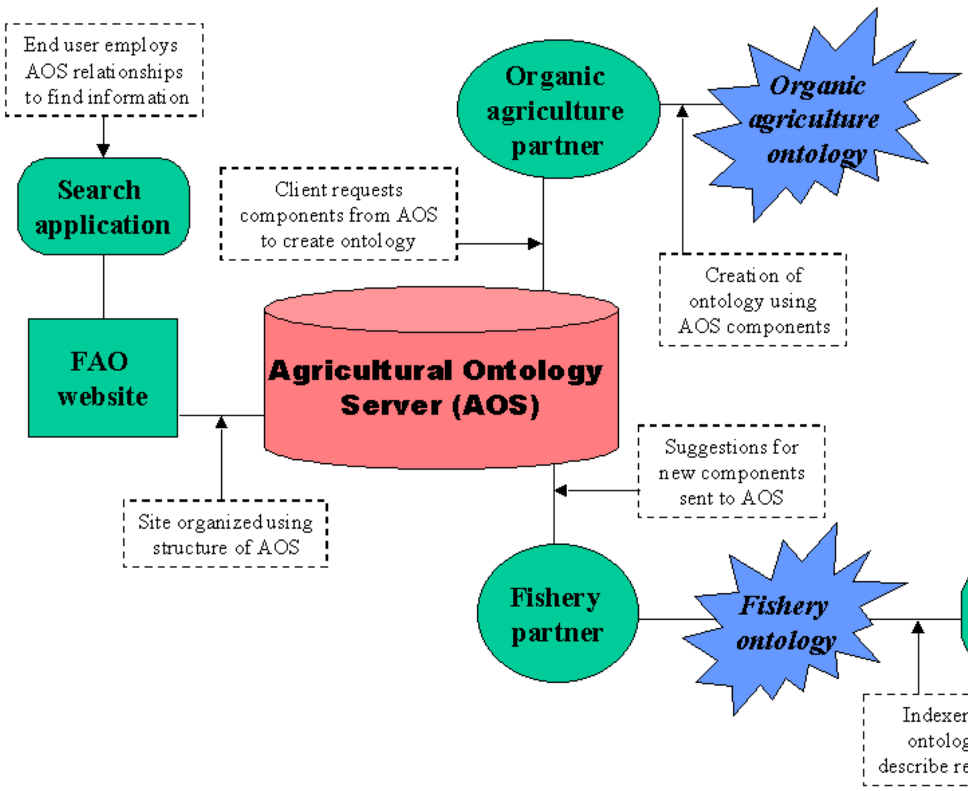
2.2 Features and Functionality

The goals of the AOS are realised by assisting community partners in **building 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, more specific and more powerful relationships. An ontology captures and structures the knowledge in a domain, and by doing so captures the meaning of concepts that are specific to that domain. This meaning is then extended to end-users through the use of tools (e.g., indexing, retrieval) that apply the ontologies.

Using the services provided by the AOS, community partners can increase the functionality of their current or planned knowledge representation systems by:

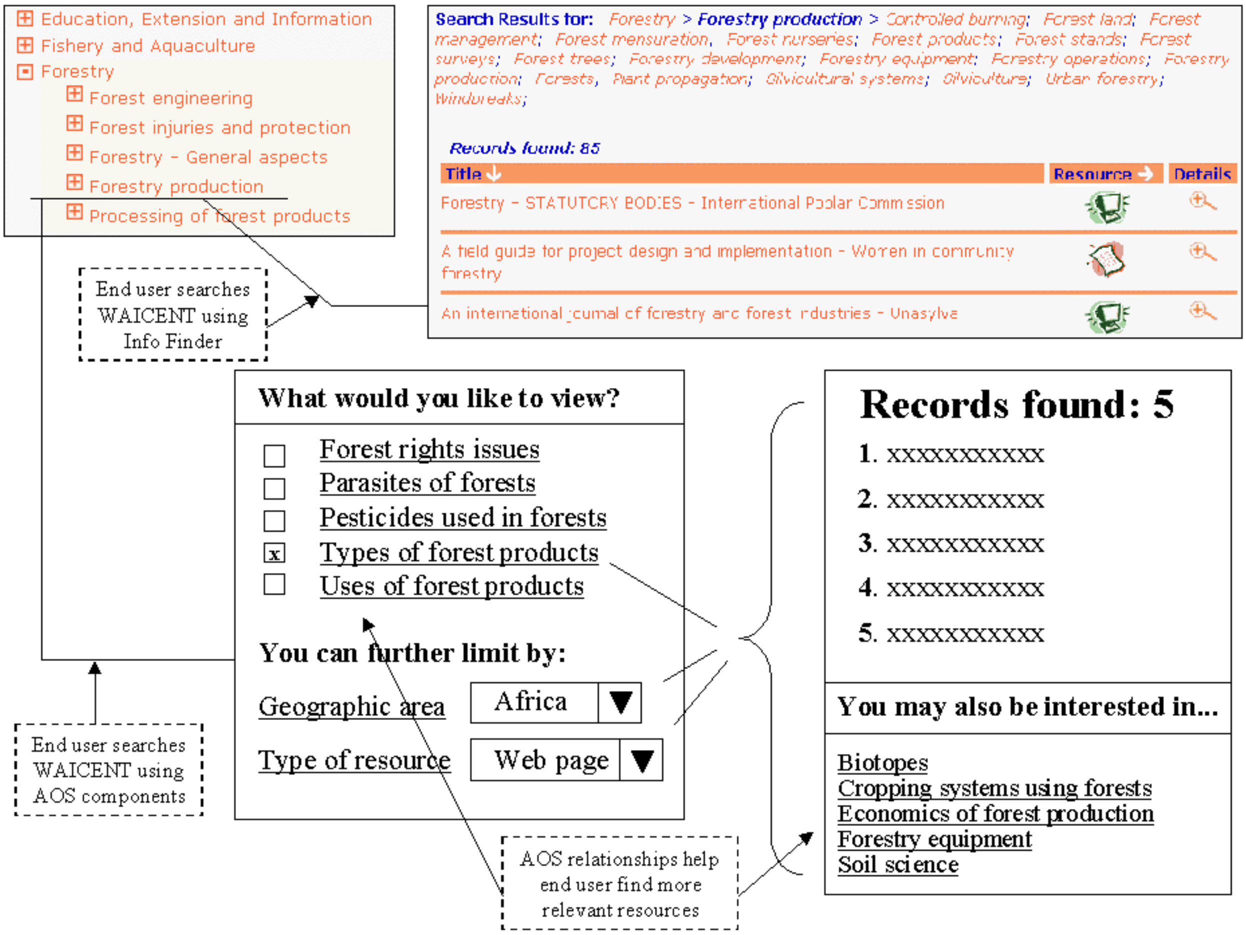
- **building their own** subject-specific ontologies for indexing and retrieval
- **integrating** their ontologies with the AOS, thereby increasing the knowledge of the server
- making their ontologies available to others for **re-use** of components in building other ontologies

The following diagram illustrates how community partners can use the services of the AOS. They can retrieve, check, map, suggest, re-use and maintain their ontologies using components of the AOS. Several of these mechanisms are shown in the diagram. The diagram also shows how tools, such as an indexing application or a searching application for an end user, take advantage of the AOS by utilising the ontologies created using AOS components.



Thus, the AOS will **provide terms, definitions and relationship components that can be shared** among associated partners, thereby enhancing communication and interaction within the community. Use of these components **increases the functionality for indexing and retrieving resources** by providing a standard source for terminology and offering richer, more powerful ontological relationships.

The following diagram illustrates the current use of relationships and the potential use of ontological relationships, using the current WAICENT Information Finder[ii] system as an example.



In the current system, the Information Finder retrieves all resources indexed with the topical category "forestry production." In the proposed system, the Information Finder will be used to retrieve resources indexed using AOS ontological relationships that result in topical categories that are more finely described, e.g., "types of forest products," "pesticides used in forests." As a consequence, the user of the proposed system is **able to retrieve a more granular and more relevant set of resources**. In addition, along with the result set, the user can be shown other possibly relevant categories, again developed using ontological relationships, that they might want to use to retrieve other relevant resources.

The AOS should be **built using AGROVOC[iii]—the FAO Multilingual Agricultural Thesaurus—as its platform**, since this thesaurus was developed at FAO and has the appropriate scope and basic relationships to serve as the underlying structure for the AOS.

Because the AOS 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, we **will need to build associations with community partners** for its development. For instance, in the fisheries area, the AOS could partner with oneFish[iv], FIGIS[v], ASFA[vi] and SIFAR[vii], among others.

Also, there should be **no reason to work in a vacuum**. Other organizations have developed knowledge servers—for instance, Cycorp[viii] and UMLS[ix]—and it is important to learn from the development of these other servers, both in terms of process and issues.

Therefore, to achieve the stated goals of the AOS, we need a server that includes the following:

- a **standardised multilingual agricultural controlled vocabulary with rich relationships**
- **partners with interest and capabilities** for development of ontologies
- use of **state-of-the-art interoperability standards** for effective communication of components
- development of **tools for development, storage, management and retrieval**

This document discusses the steps to be taken in achieving these goals.

3.0 The Plan

The AOS should provide an avenue for effective resource description and retrieval, and efficient knowledge system development and integration. The proposed project to develop the AOS would:

3.1. **Strategise about participation of agricultural community partners**.

3.2. **Utilise all possible knowledge organization systems within the agricultural domain**.

3.3. **Utilise current and developing state-of-the-art interoperability standards**.

3.4. **Develop formal ontological relationships among topics**.

3.5. **Build in functionality for describing and finding multilingual resources**.

3.6. **Create an agriculturally focused biological species micro-ontology**.

3.7. **Develop and pilot test development, storage, management and retrieval tools**.

3.8. **Test functionality of the server with end-users**.

Each of these steps is discussed below.

3.1 Participation

An **inventory of possible participants** will need to be made before any project planning can start. It is important that we develop a plan that **ensures wide involvement of the agricultural community**. As part of this inventory, we will need to **determine the methods of partner participation**. For example, what is the partner's level of participation and what resources can they offer? After taking the inventory, we will need to **form a working group of participants**. Invitations to a meeting of the working group at FAO should occur in early fall 2001.

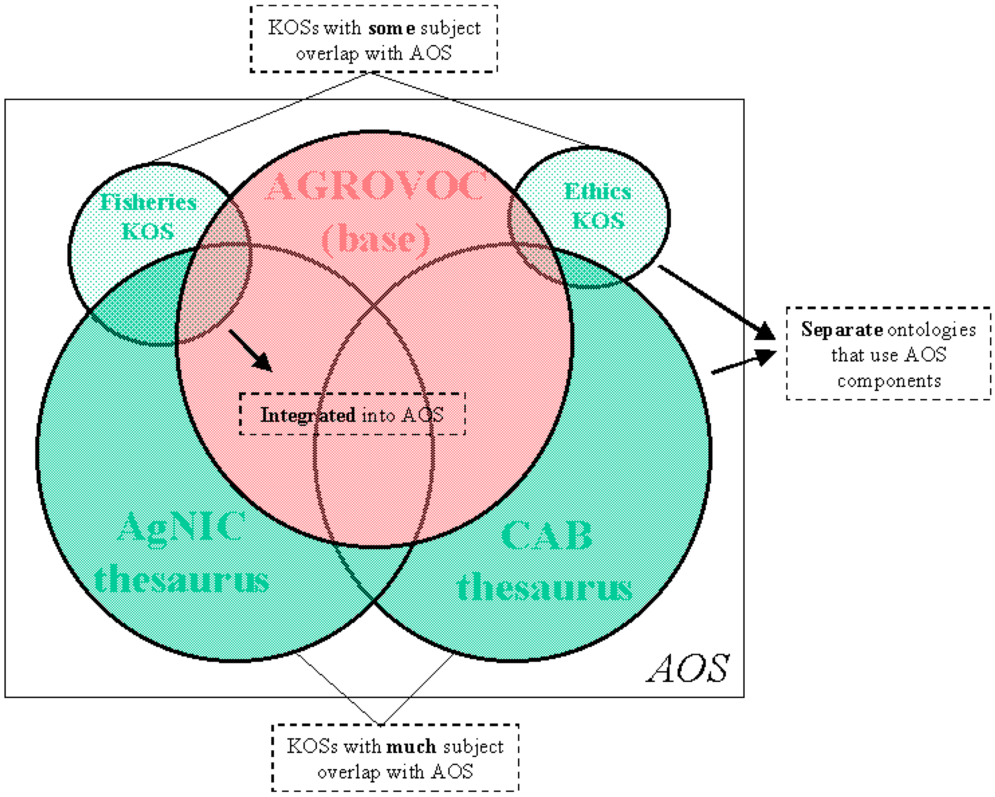
3.2. Knowledge Organization Systems (KOSs)

Since the server will be the reference for terminology and relationships, we should inform the development of this reference with as many knowledge organization systems (KOSs) in the agricultural domain as possible. A KOS is **any system that attempts to classify the information within its boundaries**, whether that is a web site, a domain, a network, or any other type of environment. To start, we will need to **take a complete inventory of potential KOSs**. In our inventory, we should determine the degree of overlap of the terminology and relationships in the KOS with AGROVOC, the platform of the AOS. This will give us an idea of:

- the amount of effort required to analyse and use the systems

- the benefits that can be obtained by both the KOS and the AOS, e.g., sharing of terms and relationships
- the process for mapping terms, definitions and relationships between the AOS and the KOS

This will **result in the development of ontologies**. Some KOSs, such as those developed by such organizations as CABI (the CAB thesaurus[x]) and NAL (the AgNIC thesaurus[xi]), have a high degree of subject overlap with AGROVOC, yet will want to remain separate entities. These systems may add the richer functionality of the AOS components to their KOS, creating a separate, newly created ontology. Other KOSs may choose to be fully integrated into the AOS and use the AOS as their primary point of departure for agricultural terminology and relationships. The following graphic illustrates this development.



In summary, the use of KOSs will be two-fold, by **helping grow the AOS** and by **assisting in the re-deploying of KOSs as micro-ontologies**.

3.3. Standards

There are now opportunities to use and share controlled vocabularies in web environments because of new standards that offer more power and flexibility. The advent of **XML**[xii] offers a common method for sharing knowledge across different tools. The **RDF**[xiii] standard allows storage and sharing of metadata (the data that describes resources) across systems. These new standards are part of **the Semantic Web** that allows machines to share information the way humans currently share information on the web. The syntax and schema behind the Semantic Web gives us **interoperability**—the **means to share resource descriptions among systems**. The AOS is dependent on this ability, since it will need to share information between the AOS and the micro-ontologies. The conjunction of these standards will enable the sharing of machine-readable URIs (Universal Resource Identifiers) among a variety of different systems. For the AOS, this communication will **allow components of ontologies**—their terms, definitions and relationships—to be **shared, evaluated and maintained using the central AOS**.

3.4. Ontological Relationships

A thesaurus has equivalence (USE/UF), broader term (BT), narrower term (NT) and related term (RT) relationships. These **relationships provide the scope and structure** for the thesaurus. We can use **an extended set of relationships** to perform more granular and more consistent indexing, and to enable more effective searching and browsing for users. We need to **formalise rules for their development** and **implement processes for using them in indexing and retrieval**.

For example, for the topic "plant production" we can describe the associations the topic has with other topics. In the table below, for instance, "raw product" is formally associated with the topic "plant production." In practice, an indexer would see all the appropriate topics and relationship types when describing a resource—a resource about "cotton balls" might receive the topic "plant production" and the relationship type "raw product." A searcher requesting the topic "plant production" could be presented with the option to limit his search to particular kind(s) of relationships, e.g., "Would you like to see raw products?". The prospect for **retrieving more relevant resources is greatly increased**.

Relationship Type	*Examples*
Primary Activity	Seasonal Cropping
Type of Plant	Cotton
Cultivar	"U-name-it" Cotton Cultivar
Location	Southern States (USA)
Production System	Agropastoral Systems
Raw Product	Cotton Balls Cottonseed Oil
Derived Product	Protective Clothing
By-Product	Mattresses Oilseed Cakes
Derived Activity	Handicrafts Milling
Resources	Cotton Gin Machines
Environmental Impact	Soil Degradation Fertilizers Insecticides
Infecting Agent	Cotton Boll Weevil
Infection	Early Blight (Alternaria)
Limiting Factor	Drought
Conflicting Activity	Re-forestation

To effectively develop this functionality, a list of potential relationships needs to be compiled, in conjunction with the inventory of KOSs. These relationships should be compared among the KOSs, with the **end result of a common set of relationships**. These common relationships will be stored in the AOS for future re-use by other ontologies. As a result of continuous sharing, development and maintenance, a refined collection of commonly used ontological relationships will be available.

3.5. Multilinguality

A key aspect of the AOS is that it will be multilingual. For any user in any country who needs access to resources, we should 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—at least. Additional languages should be added by those partners developing their own ontology, as needed. There are two large issues surrounding multilinguality: **the collection of terms in different languages** and the **mapping of terms and concepts from different languages**. For the AOS to be a functional multilingual reference, it **needs to be**

comprehensive. We will need to keep in close communication with partners to be aware of changing needs, and write thorough policies to be followed by those managing the AOS.

3.6. Biological Species Micro-ontology

A biological species taxonomy should be part of the AOS. A taxonomy is **an ordered classification of living things detailing the relationships among them.** Good web-based biological taxonomies exist (e.g., ITIS[xiv]), however we need to develop a biological taxonomy that is appropriate for use specifically by the agricultural community. We should use current taxonomies to inform development. We will want to add functionality to these taxonomies by developing new relationships and including additional terms (e.g., common names) that will **result in a biological species micro-ontology**.

3.7. Tool Development and Testing

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 include **development, storage, management and retrieval tools**, which can be further defined as:

- **mapping**—discovery of overlap in terminology and mapping of common terms and definitions
- **relationship building**—creation of ontologies using common AOS relationships and building ontology-specific relationships
- **indexing**—using ontologies to automatically and manually index resources
- **encoding**—storage of terms, definitions and relationships in a standard, interoperable format
- **maintenance**—quality assurance and upkeep of ontologies by managers
- **resource discovery**—searching and browsing resources using an ontology

As a proof-of-concept, we need to **test AOS functionality at an early stage**. An application of the Community Directory Server (CDS)—the oneFish site spoken of earlier—will provide a good opportunity to do this. oneFish uses CDS to perform many of the tools listed above, however, it does not use a controlled knowledge representation system as a reference for terminology and relationships. Many of the same steps will need to be taken by the oneFish team and the AOS team. Working closely with the oneFish team will make it possible to test a large part of the functionality of the AOS.

3.8. User Testing

It wouldn't make sense to develop these tools without **developing interfaces that address the needs of the end-users**. We need to ask a number of questions, among them if only partner end-users will be using the tools and how expert the end-users are with the subject and with the types of tools. To appropriately answer these, we need to strategise and we **need to do usability testing**. Effective usability testing involves meeting the users and can include such methods as focus groups, one-on-one evaluation, observation and task analysis.

The results of this testing will make it **possible to update and enhance AOS functionality**. Testing also **informs the methodology for training end-users** in how to operate the tools, which will need to be initiated before the end of the project. Testing should be done iteratively throughout the lifecycle of the project to ensure that the AOS continues to be suitable for the needs of end-users.

4.0 Key Issues to Consider

The development of this plan involves the discussion of a number of important, strategic issues. These include (not an exhaustive list at present):

- Which participants should be involved fully and which should be involved peripherally?
- Which KOSs in which sub-domains will we use initially to grow the server?
- What encoding standards should be used—RDF, a newer standard like XTM[xv], others?
- What process should we develop to map terminology among the KOSs and the AOS?
- What process should we develop to define ontological relationships and develop a common set?
- Who are the users of the AOS?
- What are possible other uses of the AOS besides its function as a reference for indexing and retrieval software?
- What kinds of tools are needed for these other uses?
- Are there off-the-shelf products available or what type of development is needed?

It is important to discuss and resolve our approach to these issues before the plan is instigated. This can be achieved in a number of venues (e.g., conferences, workshops, listservs) but should include all participants, who need to be at least aware of the issues whether they are involved in the resolutions or not.

5.0 References

The following documents generally informed the development of this paper:

- All articles in JODI (Journal of Digital Information) Volume 1, Issue 8 – Networked Knowledge Information Systems. http://jodi.ecs.soton.ac.uk/?vol=1&iss=8
- The DESIRE (Development of a European Service for Information on Research and Education) project, in particular the toolkit and handbook. http://www.desire.org/

The following are specific references made in this paper:

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

ii. World Agricultural Information Centre (WAICENT) Information Finder. http://www.fao.org/waicent/search/default.asp

iii. AGROVOC—FAO Multilingual Agricultural Thesaurus. http://www.fao.org/agrovoc/

iv. oneFish Community Knowledge Directory. http://www.onefish.org/

v. Fisheries Global Information System (FIGIS). http://www.fao.org/fi/figis/index.jsp

vi. Aquatic Sciences and Fisheries Abstracts (ASFA). http://www4.fao.org/asfa/asfa.htm

vii. Support Unit for International Fisheries and Aquatic Research (SIFAR). http://www.sifar.org/

viii. Cycorp: Makers of the Cyc Knowledge Server for Artificial Intelligence-based Common Sense. http://www.cyc.com/overview.html

ix. Unified Medical Language System (UMLS) Knowledge Source Server. http://umlsks.nlm.nih.gov/ Registration required. Please contact Johannes Keizer (johannes.keizer@fao.org) for access information.

x. Commonwealth Agricultural Bureaux (CAB) International Information for Development (IFD). http://www.cabi.org/IFD/index.asp

xi. Agriculture Network Information Center (AgNIC). http://www.cornell.agnic.org/agnic/about/index.html

xii. eXtensible Markup Language (XML). http://www.w3.org/XML/

xiii. Resource Description Framework (RDF). http://www.w3.org/RDF/

xiv. Interagency Taxonomic Information System (ITIS). http://www.itis.usda.gov/access.html

xv. XML Topic Maps (XTM). http://www.topicmaps.org/xtm/1.0/