Position Paper for the International Semantic Web Workshop (Infrastructure and Applications for the Semantic Web)

Title: Creating the Semantic Web: the Role of an Agricultural Ontology Server (AOS)

Author: Fisseha, Frehiwot; Hagedorn, Kat; Keizer, Johannes and Katz, Stephen (FAO, Library and Documentation Systems Division)

"Knowledge management is vital for effective decision-making.....It is therefore essential to maintain and improve the coverage, quantity, utility, timeliness and accessibility of the information collected and disseminated." (citation from the Strategic Framework for FAO).

FAO is a huge content provider for the World Wide Web. The FAO website has more than 6 gigabytes of information that contains knowledge created by more than 4000 FAO staff working in the world to combat hunger and to help people create a better life.

The semantic web is based on knowledge representation systems. Creating infrastructure for the semantic web is not only an encoding problem. Topic maps need underlying ontologies, as do any RDF description of web sites. Ontologies are emerging as a key aspect of information management in many areas, from the interchange of engineering data to corporate knowledge management.

To create knowledge representation systems (ontologies), knowledge about the represented domains is needed. **This knowledge is not with the developers of encoding systems and software but with the producers and providers of content.**

There is no realistic hope of globally classifying all concepts, terms and relationships; we need to be able to manage and interrelate knowledge representation systems (ontologies) project by project, domain by domain, so that scalability is achieved without either runaway complexity or over-simplification.

FAO, together with partners and other stakeholders in the area, has been developing and maintaining knowledge representation systems in the basic form as represented in the AGRIS/CARIS subject categories and the AGROVOC thesaurus for nearly two decades. The advent of the internet and the World Wide Web gives us the possibility to extend the concepts behind these systems.

We are planning to develop an Agricultural Ontology Server (AOS):

- To allow domain knowledge to be defined and described
- To communicate among domains without semantic ambiguity
- To enable reuse of domain knowledge
- To share the structure and meaning of agricultural information among users and tools
- To provide foundation to build other specific ontologies

And last, but not least:

- To provide more effective dissemination and access to knowledge for users

Briefly defined, the Agricultural Ontology Server (AOS) functions as a central common reference tool for serving ontologies. An *ontology* is a system that contains terms and 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. It is designed to serve as a central focal point for the vocabulary of a particular domain, and to codify and standardise the knowledge within that domain. It enables better communication within and across domains, and structures meaning contained in the domain.

In essence, 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 subsets of the knowledge domain will use in building and maintaining their own ontologies. For instance, in this case, the AOS provides the reference for all the terminology of the agricultural domain. Knowledge domain subsets, including forestry, fishery, plant biology, sustainable development, organic agriculture and nutrition, will use this reference tool to build their own ontologies. Once these ontologies are created, they can be used to inform knowledge bases—and can be re-used and enhanced by other knowledge bases. This is an iterative process that grows and maintains the ontologies.

The existence of a common ontology server guarantees that common concepts are clearly defined by unique identifiers, and basic relations are used throughout the domains.

A thesaurus has 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 those 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.

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—creating rules for developing specific relationships—and thereby provide a means for better knowledge sharing. However, this would have to be balanced with the need for compatibility with existing systems and future interoperability.

Since the server will be the central reference resource for vocabulary control and relationship structure of agricultural terminology, we will need to utilize multiple different sources to build it. The main source will be the AGROVOC thesaurus, which already has the appropriate scope and basic relationships to serve as a base for the AOS. Other sources 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 categorizations of knowledge domains

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 should collect and coordinate terminology, definitions and relationships in the five official languages of the FAO—English, French, Spanish, Arabic and Chinese. Additional languages can be added if necessary by those developing ontologies, if working in the mother tongue of the country is beneficial.

We will need to develop a suite of ontology tools to be used in 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 new relationships
- coding—storage of terms, definitions and relationships in a standard, interoperable format
- indexing—using ontologies to index resources
- discovery—searching and browsing by users in the AOS or in an ontology
- maintenance—ontology collection, storage, dissemination and evaluation by managers

We will need to incorporate current state-of-the-art standards (RDF and XTM) in the encoding of the ontology server for the KOSs to communicate with each other effectively.

We believe that other stakeholders in the area of agriculture, forestry, fishery and nutrition and the like, who have or need Knowledge Organization Systems for their information ensembles would benefit widely from the AOS. We believe also that the development of tools and software could be inspired by the existence of a large structured system of knowledge representation.