Autoevaluation of FAO Program Entity 3HP02

Thomas Baker

June 9, 2010

Contents

Scope and method

This report is the result of an autoevaluation of the Programme Entity (PE) 3HP02, “Standards, norms and procedures for knowledge management and information management,” as defined in the Programme of Work and Budget (PWB) 2008–2009 and managed by the WAICENT Knowledge Exchange Facilitation Branch (KCEW) of the Knowledge Exchange and Capacity Building Division (KCE) of the Food and Agriculture Organization of the United Nations (FAO). PE 3HP02 was, in turn, based on an earlier programme entity, PE 222P7, “Standards, norms and procedures for improved access to agricultural information,” which came into being with PWB 2006–2007 in the context of FAO’s Medium-Term Plan 2006–2011. PE 222P7, in turn, built on prior work that began in 2000.

The report focuses specifically on the impact and future prospects of information-management standards for the “organisation, classification and cataloguing of information in FAO’s areas of expertise” (3HP02 Major Output 003). These standards have been published on a Web portal, Agricultural Information Management Standards (AIMS)[1], in order to support the adoption and implementation of the standards in FAO member countries (3HP02 Major Output 006).

Work on PEs 222P7 and 3HP02 was performed under the leadership of branch chief Stephen Katz and information systems officer Johannes Keizer by roughly eight professional KCEW staff members and three service employees, typically on a part-time basis in conjunction with duties towards other programme entities; overall, roughly two-thirds of their effort was oriented to the maintenance of more traditional, legacy information systems under the higher-level entity 3H.

Specific tasks for 222P7 and 3HP02 such as thesaurus enhancement and software

[1] http://aims.fao.org/

development were sub-contracted to external consultants and to partner organizations, notably in India and Thailand. Participation in external research projects — in particular the EU Integrated Project IST-2005-027595, "NeOn: Lifecycle Support for Networked Ontologies" — funded extra capacity for research in standards-related applications and methodologies for four years starting in March 2006.

This report refers to the group of staff members working on 3HP02-related standards as the "AIMS team" and to the standards on which they worked as the "AIMS standards."

In the course of researching this autoevaluation, the reviewer spoke with more than thirty people by email, Skype, and face-to-face, often with follow-up. People interviewed included members of the AIMS team, colleagues in other departments at FAO, partner organizations in member countries, and other consultants. The reviewer visited FAO headquarters in September 2009 for initial contacts and returned in December to fill in gaps and discuss preliminary results with the AIMS team. His desk research encompassed dozens of published articles, internal budget documents, Powerpoint presentations, and back-to-office reports from as far back as 2000 but with a focus on the four years starting in 2006, as well as documentation for the AIMS Standards themselves. The report takes into account results of an April 2010 workshop held at MIMOS Berhad in Kuala Lumpur. In order to evaluate the ontologies, the reviewer focused on understanding the meaning of the statements ("triples") expressed in the data itself.

Chapter 1

Overview

1.1 AIMS achievements

In the early 2000s, a series of workshops with experts and international partners encouraged FAO to work with Member Countries to become “a key enabler and catalyst to establish a new model of agricultural information management in the 21st century” based on decentralized information management and using “Web-enabled” standards for interoperable data exchange. The guiding theme was provided by Tim Berners-Lee’s seminal keynote at XML2000, which outlined his vision of a Semantic Web based on “ontologies.” Under the banner “Agricultural Ontology Server” (AOS), the AIMS team developed a program with three main components:

- The use of simple descriptive metadata for integrating access to agricultural information in both developed and developing countries and, to a lesser extent, in FAO’s own technical departments.
- The development and maintenance of thesauri and ontologies as descriptors for structuring access to agricultural information and as “building blocks” for application-specific ontologies.
- Networking, capacity development, and outreach aimed at promoting the use of these standards by FAO information providers and partner organizations.

1.2 Course corrections

As an early adopter of Semantic Web technology, the AIMS team has been years ahead of the curve in porting its legacy information management standards from the print world into Web formats and is well-positioned to benefit from current technological trends. In some areas, however, the team is paying a price for having been a bit too far ahead of the curve. This chapter summarizes the work done, lessons learned, and outlines some course corrections already being undertaken:

- The simple but rigid metadata record formats it has defined, such as the AGRIS application profile, have allowed the AIMS team (and others) to merge information from diverse sources into central databases but now need to be loosened to accommodate input that is either simpler (where resources are scarce) or more complex (where requirements are more comprehensive) — something which more flexible technological approaches now support.

- The AIMS staff has been productive far beyond its size by mobilizing voluntary and subcontracted work by partner organizations and by securing substantial outside funding. Their remarkable achievement in porting legacy standards from the print world into the Web environment have given them high visibility in the field and at international conferences. These extraordinary results, however, have come at an unsustainably high cost to its overtaxed staff.

- The metamodel custom-designed in-house for upgrading AGROVOC and other AIMS thesauri into Web-enabled ontologies, while novel and innovative in 2006, has now been superseded by a W3C Recommendation — Simple Knowledge Organization System (SKOS) [8] — that serves the same purpose. The AIMS team can increase the quality and efficiency of its work by aligning with new standards.

- The AIMS team has suffered from the dependence of its strategy on a software development project managed primarily at Kasetsart University, the AGROVOC Concept Server Workbench. This project is, in turn, dependent on software — the Stanford Protégé triple store and its OWL API — which has proven to be difficult to work with. The graphical interface has obfuscated problems with the underlying model, causing an explosion of redundant triples and slowing performance. At the time of finalizing this autoevaluation, extricating AGROVOC from this dependence is proving to be a tricky operation. The best way forward lies in opening the project to input from other groups interested in solving the same problems using SKOS.

1.3 A strategy for Linked Open Data

The Semantic Web vision outlined in 2000 achieved its breakthrough when Tim Berners-Lee radically redefined the message in 2006 around the notion of Linked Data.[1] The term Linked Data refers to a style of publishing structured data on the Web in which all elements of an ontology (properties, classes, and value vocabularies), as well as many of the things described by the ontology (publications, events, people), are identified by Uniform Resource Identifiers (URIs), and data sources are extensively cross-referenced ("linked") among themselves using generic data "statements" ("triples").

The vision of Linked Data is succeeding where Semantic Web did not because it conveys a simple message that can be understood in very concrete terms. People can see that it has to do with how things relate to each other, about making such links resolvable on the Web for practical purposes such as structured browsing and data integration.

Linked Data is strongly associated with an architectural style, Representation State Transfer (REST), [2] which uses HTTP URIs to identify resources, distinguishes between resources and "representations" of resources in multiple data formats, and relies on making those URIs resolve to information resources by using response codes and media types to make the data "self-descriptive" so that its correct interpretation does not depend on knowing "out-of-band" context. The REST style uses URIs as the sole pathways for leading a user agent between information sources ("hypermedia").

In Linked Data terms, an ontology is a conceptual structure represented as data — data over which services can be built. Using HTTP URIs and resolving those URIs to useful information that people can look up replicates the function of a dictionary. By promoting use of the URIs of AIMS standards in tagging (annotating) Web content worldwide, AIMS can empower resource providers to bypass centralized aggregators and search engines, which seek to position themselves as gatekeepers, and connect their resources directly to a growing

[1] http://www.w3.org/DesignIssues/LinkedData.html
[2] http://en.wikipedia.org/wiki/Representation_State_Transfer

Linked Data cloud.

As the technological approach which AIMS helped pioneer now matures, AIMS will be able to benefit from generic software tools developed in the commercial world and open-source communities. With mainstream search engines and applications adopting the Linked Data approach, AIMS can transition from the role of technological innovator to that of developing capacity to help information providers in member countries benefit from the Web revolution.

The sections which follow review technical achievements, user feedback, and planned course corrections with respect to:

- Metadata based on application profiles that use open, Semantic Web vocabularies to describe documents and other objects of interest, such as events, people, and learning materials.
- Thesauri such as AGROVOC, upgraded for publication and use in a networked environment as Web ontologies, and their alignment with specialized vocabularies in domains such as fisheries.
- Collaboration among partner organizations in the creation, maintenance, and deployment of standards for sharing knowledge related to food and agriculture, notably in the context of an umbrella initiative, Coherence in Information for Agricultural Research for Development (CIARD).[3]

All of the standards and projects discussed below are documented or linked on the AIMS Website.[4]

1.4 Recommendations

1. Publish all AIMS concept schemes and namespaces as Linked Data.
2. Before promoting URIs for use in linked data, take strategic decisions about the form of URIs, policies for their long-term maintenance, and the method for their publication (see Appendix A).
3. Promote the use of AIMS URIs for annotating (tagging) resources.

[3] http://www.ciard.net/
[4] http://aims.fao.org/

Chapter 2

Integrating access using metadata

2.1 Achievements in promoting metadata standards

Work on the standards that now fall under the banner of AIMS began under an Agricultural Metadata Standards Initiative (AgStandards) in 2000. Inspired in part by the Dublin Core Metadata Initiative, then five years old, the AgStandards Initiative took the fifteen elements of the Dublin Core Metadata Element Set (DCMES) — basic elements such as Title, Subject, and Date — as a starting point and defined itself as an umbrella under which additional elements could be created. A new namespace for describing document-like resources relevant to agriculture, Agricultural Metadata Element Set (AgMES), was published in 2005 as the first output of the initiative.

The flagship implementation of AgMES is the International Information System for the Agricultural Sciences and Technology (AGRIS), FAO's database of bibliographic references to literature produced by agricultural research centers around the world. From its beginnings in 1969 — the name "AGRIS" dates from 1975 — through the late 1990s, AGRIS was maintained by FAO as a centralized database with its own unique database structure, exchange formats, and software.

With the rise of the World Wide Web and its new paradigm of distributed information management, the AGRIS database was by 2000 looking old-fashioned and unsustainably centralized. Between 2000 and 2003, a series of workshops with experts and international partners encouraged FAO to diversify institutional participation in AGRIS through capacity development, which aimed at empowering local and regional AGRIS centers to improve information management in their own institutions. The workshops endorsed the role of FAO in supporting common standards and protocols for achieving this goal.

The renewed AGRIS effort focused on the use of a simple application profile based on Dublin Core — the AGRIS Application Profile — as the basis for conversions from a wide range of local database formats into a common XML format (Document Type Definition, or DTD). To facilitate the adoption of the AGRIS profile by AIMS partners such as the Global Forestry Information Service (GFIS) and the research centers of the Consultative Group on International Agricultural Research (CGIAR), the AGRIS team defined mappings from legacy data formats and developed simple data input tools ("WebAGRIS" and "MetaMaker").

By 2005, the AGRIS team had converted the entire repository of three million records from its legacy library-catalog-based "AGRIN3" format into XML records based on the AGRIS profile. Over the years, data has accumulated in AGRIS from over two hundred institutes, and of today's one hundred AGRIS providers, roughly sixty remain "very active." Some AGRIS data is carried from very remote locations on thumb drives. Institutions have been encouraged to configure their databases to generate AGRIS-conformant XML data for harvesting and transformation by the central AGRIS team. In practice, it has fallen to the staff in Rome to correct and validate much of the data semi-manually. The introduction of the AGRIS AP as a common exchange format dramatically reduced the need for editing and cleaning incoming data, which before 2000 had been done by a team of more than ten people at the AGRIS processing unit in Vienna.

The AIMS team followed up its publication of the AGRIS profile by developing or promoting profiles for other types of information – e.g., for News (using the standard RSS news format) and Events (a simple profile with starting and ending dates, location, type, and organizer). These were used for an alert service, AgriFeeds[1], which was launched in 2007. The team also created a profile for brief descriptions of organizations which, when published on their own Websites in XML, can be harvested for automatic compilation into lists.

In 2006, work began on a profile for providing structured access to learning resources in a Capacity and Institution Building Portal.[2] This profile uses results from an ongoing effort by DCMI and the Institute of Electrical and Electronics

[1] http://www.agrifeeds.org/
[2] ftp://ftp.fao.org/docrep/fao/010/ai154e/ai154e00.pdf

Engineers (IEEE) to harmonize the simpler approach of Dublin Core metadata with the more comprehensive and complex specification of the IEEE Learning Object Metadata (LOM) standard on the basis of a Linked-Data-compatible representation.

2.2 Feedback from application profile users

The renewal of the legacy AGRIS database as a Web repository is generally seen as a big success, and the AgriFeeds service is widely used. The repository has exposed local research results to a global audience. The AGRIS center in South Korea, for example, has been delighted at the surge in requests for its publications, especially since AGRIS has been picked up by Google.

The AGRIS Application Profile 1.1 of July 2005[3], however, prints out at eighty-one pages, and as various users attest, the profile is widely considered "heavy" to implement:

> We do not use the AgMES application profile. Not that we reject it, but we see that such applications are too heavy-duty for people in developing countries. They do not have the staff to do detailed things, and we do not want to push them to adopt anything. At our home office we have even less capacity for adding metadata or mapping.

> The Application Profile is very comprehensive and it does have substantial advantages over simple Dublin Core. But in practice, it is quite cumbersome to fill in the complete profile.

> Author strings can be constructed many different ways. You are happy there is at least something. But the AGRIS profile is a bit strict. Probably they compromise in reality, but if you just go by the guidelines it requires a higher level of control than we can afford.

An AIMS partner confirms that even the task of mapping from existing formats presents a significant barrier:

---

[3]http://www.fao.org/docrep/008/ae909e/ae909e00.htm

> Today we have over 170 information provider partners from around world, but only half have created RSS feeds links to us — and only because we could show that it did not take much working time. We have had even less success in getting partners to create AGRIS data from their native records — it is a bigger job for them to understand the records and make the mapping.

Quite a few users suggest that AGRIS lower the bar by promoting simpler, lighter alternatives:

> In order to justify the working time, our information providers want to see how this will help them get more users, like offering a simple search tool. Maybe FAO could make the profile simpler and more flexible. Start with something very simple, like RSS, before introducing more comprehensive metadata solutions.

> We would like to submit data to AGRIS. The problem is that the data is very dirty — it is collected from different sources. The funder collects things they no longer fund, and you have to accept everything and get very dirty metadata. We require something a bit lighter than the AGRIS application profile.

A minority of users, on the other hand, see the problem less as one of complexity than of excessive simplicity and lack of flexibility:

> The AGRIS application profile is really useful. The weakness is that it is difficult to revise to meet local needs. For example, the AGRIS profile does not have an “affiliation” element.

Work on an application profile for describing projects ran up against the limits of simple, flat (and therefore more easily interoperable) descriptions with the need to provide contact information for project coordinators and recipient institutions — information that requires descriptions about additional entities, such as people and organizations, to be embedded in records about publications.

Some suggest that basic Dublin Core metadata would suffice:

> The additional complexity of the AGRIS application profile over Dublin Core is not really needed — we could just use Dublin Core.

> An AGRIS Profile Lite would be nice — something like Simple Dublin Core.

> For describing documents, plain Dublin Core is good enough.

Indeed, AGRIS is perceived as competing with Google Scholar because the latter ingests OAIster, which uses Dublin Core:

> FAO should try to work with big search engines, which have a philanthropic side. Putting our data into OAIster gave it alot of visibility, as it was polled by Google Scholar. That is good! But we have no time left over to create AGRIS metadata. It means competing with other international organizations with a special standard, and we cannot compete with Google for visibility so should work with them.

AGRIS staff point out in response that "the AGRIS profile is perceived as complicated because people see the fifty or sixty fields but do not realize that only five or six of those fields are mandatory." The AGRIS team does in fact accept data in whatever granularity it is provided. Many descriptions provide just a minimum, with Title, Subject (typically with an AGROVOC value), Date, Availability (location), Language, and often Conference Name. This message, however, has clearly not been widely understood:

> The requirements of the AGRIS profile are not actually very demanding, but this message should be clearer. The guidelines on the Web site could put in more examples with just title, subject, and one or two other elements.

This would, of course, require the AIMS team to revisit the AGRIS profile, formulate new guidelines, and write new documentation. Having completed its initial push to create a series of AIMS profiles, the AIMS team would need to decide how much internal effort to invest in such ongoing maintenance, moving forward, as opposed to pushing the profiles out to broader maintenance communities:

> Nobody is currently maintaining the AGRIS profile. Maybe a new version could come from outside FAO, for example from DCMI. As it stands, there is not really a way of updating the standard and making it a community thing and not a FAO thing.

AGRIS staff observe that the role of metadata is shifting in ways which de-emphasize the importance of information about the location of a resource. In the Web world resources are, in practice, often moved around or replicated on multiple servers. Google, on the other hand, excels at finding "known entities" — resources for which an exact title, authors, or other bibliographic information is known, if not the location. In the new division of labor between search engines and curated collections, bibliographic databases can help users discover that a resource exists, then Google can help them find and retrieve the resource, wherever it may be.

One user suggests that, if nothing else, tagging resources by subject would by itself be a big win:

> Focus less on application profiles than on using AGROVOC well. If people could pull elements from AGROVOC just to tag their things, it would be fantastic.

Another user cautions, however, that even minimal requirements can be hard to meet:

> When we started, I thought an RSS feed would be simple — just a title, description, date, and link to the source. But our experience, even the publication date, which should be simple, is full of errors.

The underlying problem, according to many of the users, is the lack of basic knowledge and skills in information management methods:

> In our experience with RSS and the AGRIS profile, the main problem is not with the specifications themselves. The biggest problem is that organizations which maintain and create information on the Web do not have knowledge or skills to maintain metadata. They have old-fashioned Web sites — hand-made, not dynamically generated.

> Behind those Web pages, some developers have learned to maintain Web pages, but the structure as a whole is not well prepared. Only a few providers know how to create RSS or AGRIS XML data, upload to the Website, and link to our service.

> Most of our users do not know what XML is.

> With the AGRIS profile, people are sometimes intimidated by the big words, even if it is just their own data fields that are getting mapped.

> We have guidelines, but people do not read them. Instead, they ask us! Some people are simply too busy to read today.

The solution, expressed in many ways by the people interviewed, lies in capacity-developing measures for bringing users up to speed with the technology:

> Ninety percent of our users are in developing countries. The key is capacity building. It is one thing to publish a specification, but to get uptake in twenty institutions, you need to hold face-to-face meetings, identify champions, and train the trainers.

2.3 Metadata enrichment and conversion to Linked Data

The AIMS team is currently exploring ways to leverage AGRIS in the Web environment by publishing the entire repository in the form of RDF "triples," — the fundamental unit of Linked Data. The process involves "metadata enrichment" — the progressive enhancement of descriptions, where possible, with explicit links (URIs). This turns each AGRIS record into an entry point to a web of authors, institutions, and topics — a "hub" for drawing together a global collection of information and, by extension, the community of authors.

The new role of URIs in weaving the Web changes the role of metadata itself by de-emphasizing its function for finding information, for which people often turn to Google. Rather, metadata functions increasingly as a bundle of links that embed a given resource in a web of relationships, thereby giving that resource a context.

With help from the information management company Talis and a team from the Okkam Project at the University of Trento, the AGRIS team is testing the "triplification" of AGRIS XML records. Talis is testing the conversion of string values for Creator, Publisher, Language, and Type into URIs from authority files for authors, journals, languages, and resource types. The Okkam Project is testing algorithms for disambiguating between authors, given inconsistently entered names, by using contextual information such as affiliation, co-authorship, or country. Subject, arguably the most important field in AGRIS descriptions because it links resources to FAO's areas of interest, is also one of the "cleanest" in the dataset because it was populated largely using tools which copied subject strings directly from AGROVOC online.

Before the conversion of strings into URIs, data must often first be cleaned by normalizing variant strings to the "termspell" (normalized string) of a target vocabulary. The process of cleaning, normalizing, and enriching cannot be fully automated — people need to control the results at every step — and the procedure is intended to be a one-way migration, not something that is carried out repeatedly and on-the-fly. It greatly helps that the XML data files of AGRIS are already partitioned according to year and month of ingest, country, and institution because the quality of records has improved over the years as AGRIS centers have acquired data-entry tools.

Moving forward, the AGRIS team aims at facilitating the use of URIs by increasing tool support. AIMS partners are developing small utilities and plug-ins, for example, to tag content with AGROVOC descriptors ("AgroTagger"), enhance string-based record fields with URIs in DSpace repositories, and identify concepts in texts for annotation with URIs in Drupal content management systems ("AgroDrupal"). As one AGRIS manager explained, the AGRIS profile can be taken as a foundation and, starting with a minimal record, tools can be used to enrich the data, automatically, with information extracted from the content of the resource or inferred from its context.

2.4 Accepting "whatever you can get"

For many years, the dominant paradigm for the interoperability of digital information has been syntactic conformance with specific data formats encoded as XML DTDs or XML Schemas. AIMS application profiles were based on a set of well-defined data elements semantically compatible with RDF properties and classes. Transforming AGRIS partner data into the AGRIS XML format was a process of mapping local data elements of AGRIS data providers to common target elements. As the concept of Linked Data had not been developed in 2005, and most AGRIS partners lacked and continue to lack the experience for publishing their data directly in an RDF representation syntax, the AGRIS DTD has served as a transitional aid for creating data that is conceptually and semantically (though not syntactically) interoperable with RDF.

The emerging paradigm of Linked Data, in contrast, explicitly avoids requiring that information providers expose identical formats. RDF provides an abstract model for data that can be serialized in one of several interchangeable syntaxes for representing data as generic "statements" (RDF "triples") that can be joined automatically on the basis of shared global identifiers (URIs). The "Open World Assumption" underlying Linked Data avoids assuming that any one source provides complete and exhaustive information about a given resource and anticipates that information sources may only partially overlap. Whereas formats such as DTDs can be "broken" by omitting data, triples constitute a language in which "missing is not broken" [1]. By anticipating the future integration of new sources even if they are not completely aligned, the architecture of Linked Data is more resilient to imperfections and diversity, while the syntax-independent model of triples makes data more "future-proof."

In the new paradigm, interoperability is an unbroken continuum that depends on the "coherence" of merged triples. Coherence is provided best by shared URIs — URIs identifying resources described, URIs for the properties used to characterize the relationships between resources, and URIs for the classes used to characterize the type of a resource.

String values — sequences of alphanumeric characters such as names, dates, and publication abstracts — are inherently less precise as a basis for merging data due to natural variations in spelling or punctuating subject headings and titles, representing names, or formatting dates. To improve their value for Linked Data, it is important that string values be "qualified," when possible, with descriptive context. Date strings, for example, can be expressed as RDF "datatypes" (in Dublin Core terminology, Syntax Encoding Schemes) by providing a URI identifying the ISO or W3C standard that specifies pattern used to form the sequence of months, days, and years.

Value vocabularies are most effective for use in Linked Data when their individual terms are identified using URIs, as with AGROVOC. However, a URI identifying a Vocabulary Encoding Scheme, or VES (in Dublin Core terminology) can be used to put a string value into the context of a controlled vocabulary. Using a VES URI together with a string is not as precise as using a URI for a specific term, but for controlled vocabularies that have not yet been "Webified," it is much better than providing no context at all. As an example, the string "Agriculture—Biography" is more useful if contextualized with the Vocabulary Encoding Scheme URI http://purl.org/dc/terms/LCSH, which says that the concept represented by the string is a member of the Library of Congress Subject Headings. Since 2009, however, individual subject headings have been assigned URIs by the Library of Congress itself, so this particular heading can be referenced more precisely by using its own individual URI, http://id.loc.gov/authorities/sh88005148#concept.

Shifting the emphasis from shared data formats to the coherence of underlying triples will allow the AGRIS team to relax the requirements for data ingest and more flexibly accommodate data from a growing diversity of providers. Providers using RDFa to embed structured descriptions "invisibly" into normal Web pages, for example, will be able to use tools such as Yahoo SearchMonkey to extract the underlying triples for ingesting into AGRIS. This shift redefines the function of the AGRIS DTD, moving forward, from that of ensuring interoperability through uniformity of format to that of providing a validatable format that is cleanly convertible into RDF triples. If the AGRIS DTD continues to be used, an extensible stylesheet language transform should be maintained to automate this conversion.

Promoting RDF triples as an acceptable input format will give information providers more flexibility to use application profiles that transcend the limits of flat descriptions of a single resource — for example, by adding information about authors, such as affiliation — without sacrificing interoperability.

2.5 Recommendations

1. Correct the perception that AIMS application profiles are technically rigid by redefing their function from that of ensuring interoperability through uniformity of format to that of providing validatable formats that are cleanly convertible into Linked Data (i.e., triples).

2. Correct the perception that the AIMS application profiles are onerous by actively promoting minimal descriptions, such as descriptions consisting of just five elements or, in the extreme case, of just one element relating a

resource to an AGROVOC concept URI.

3. Where URIs are not available, promote the “qualification” of string values as RDF datatypes or Vocabulary Encoding Schemes.
4. Work with research partners to “triplify” datasets such as AGRIS converting value strings, where possible, into URIs (“metadata enrichment”).
5. Promote a value proposition for metadata as providing a bundle of links that embed a given resource in a web of relationships, providing the resource with a well-defined context and increasing its value for finding related resources.

Chapter 3

Thesauri and ontologies

3.1 Achievements in promoting thesauri and ontologies

AGROVOC, a multilingual thesaurus of agricultural topics, was created by FAO and the Commission of the European Communities in the early 1980s. It consists of "terms" (natural-language phrases) in multiple languages cross-referenced with other broader, narrower, and related terms. The thesaurus standardizes term codes and "termspells" (spelling and punctuation) in order to improve the quality of indexing and search.

From 8,660 descriptors (preferred terms) in 1982, AGROVOC grew to 16,607 descriptors by 2000 and has roughly 32,000 descriptors today. Initially available in English, French, and Spanish, AGROVOC is now available in nineteen languages, with additional translations in the works. Periodic releases of AGROVOC can be freely downloaded in its native relational database format or in alternative formats such as Microsoft Access, and the latest version can be accessed by applications via Web services for looking up terms or expanding queries. AGROVOC terms have been mapped to terms in the Chinese Agricultural Thesaurus (CAT), the Schlagwortnormdatei (SWD) Thesaurus of the German National Library, the US National Agricultural Library Thesaurus (NAL), the GEneral Multilingual Environmental Thesaurus (GEMET) of the European Environment Information and Observation Network, and the CAB Thesaurus of the UK-based technical agency CAB International.

In 2001, the planned Agricultural Ontology Server (AOS) was envisioned as "a reference tool that structures and standardises agricultural terminology in multiple languages," providing modules of terms that can serve as "building blocks" for developing more specific domain ontologies. From 2004 through 2006, the AGROVOC project team formulated a conceptual model with "the necessary structure to create precise semantics to facilitate the transition from traditional thesauri to ontologies" [10] — in effect a "metamodel" for thesauri — which modeled thesaurus terms as lexicalizations of underlying concepts. These underlying concepts were represented in the ontology as OWL classes (see Appendix B).

Starting in 2005, the AIMS team focused on "refining" AGROVOC's standard thesaurus relationships ("Broader Term," "Narrower Term," "Related Term," and "Used For") into semantically more specific relationships such as "hasIngredient" or "growsIn."[1] The refinement of thesaurus relationships was undertaken with the implicit assumption that a more precisely engineered ontology would support more intelligent queries — for example, to determine whether a specific farming method has been used in a dryland area for a given crop and to find any relevant research reports in whatever language they may be available. Most of the refinements have been defined by experts at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in Patancheru, India.

Converting the metamodel of AGROVOC into a class-based ontology and refining the relationships among its concepts, however, was only part of the AIMS vision. Equally important was the notion of enabling AGROVOC to evolve dynamically, in response to technical innovation, scientific advances, regional specialization, and linguistic evolution. Just as member institutions were empowered to submit bibliographic data directly to AGRIS, decreasing dependence on the central team in Rome, there was a strong push to enable expert users in AGROVOC's twenty-some language areas to maintain the ontology directly online. Aside from relieving the central AGROVOC team of the cumbersome and relentless task of processing change requests — a frustrating bottleneck both for the team and for its users — the idea of moving maintenance to the Web addressed what Martin Hepp refers to as the trade-off between "ontology engineering lag versus conceptual dynamics" [4] — the insight that knowledge itself is continually evolving, that the process of ontology development is necessarily iterative and dynamic, and that for semantic applications, the most important concepts are frequently also the newest.

In 2005, requirements were developed for a Web-based platform — the AGROVOC Concept Server Workbench — to allow experts in many countries to

[1] http://agrovoc.icrisat.ac.in/agrovoc/relationstree.php

add or translate concepts in their specific areas of interest. The Workbench was conceived as a distributed, Web-based maintenance environment that would enable participants in multiple countries to edit parts of the central AGROVOC ontology simultaneously — adding term translations, adding or refining relationships between terms, or performing batch modifications on the basis of pattern matching. The Workbench was also seen as a platform for plug-in tools that could proactively populate AGROVOC with new concepts extracted by corpus analysis from breaking news stories ("ontology learning"). The move to a distributed architecture was seen as a way to loosen the dependence of AGROVOC on terms entered canonically in English, then "translated" into other languages, towards an environment in which users could create new locally-specific terms in any language.

The system was intended to support levels of authorization ranging from Guest Users through Term Editors, Ontology Editors, Validators, and Publishers, to System Administrators. It was designed to support the extraction and export of sub-sets of concepts for personal use and the upload of entire ontologies for sharing with others. It was conceived of as a generic tool in principle adaptable to other domains, such as health care and medicine. Part of the vision was eventually to provide add-on services such as automatic or semi-automatic translation, ontological reasoning, guided search, and concept disambiguation.

In 2006, having formulated Workbench requirements and finalized the OWL-class-based ontology model, the AIMS team, finding no software capable of fully implementing this vision off-the-shelf, undertook the development of a customized interface to a backend ontology database, Protégé. This software development project has been led since 2006 by Kasetsart University in Thailand with input from implementation testers in Rome and Patancheru. An alpha version of the Workbench was released in June 2008, and development has accelerated in 2010 with the involvement of a development team at MIMOS Berhad in Malaysia. AGROVOC has in the meantime been maintained in the original thesaurus database, with snapshots periodically exported to the Workbench for testing. After a final migration, the original thesaurus database will be retired and maintenance of AGROVOC will continue on a production basis in the Workbench.

In the meantime, AGROVOC term codes and "termspells" have been widely used in agricultural portals and repositories worldwide. At FAO itself, AGROVOC terms have been used in AGRIS; in an International Portal on Food Safety, Animal and Plant Health; in an Emergency Prevention System for Transboundary Animal and Plant Pests and Diseases; in Geonetwork, a repository of geospatial information; and in the Electronic Information Management System, a workflow database used at FAO to track publications.

Although AGROVOC has not yet been used in its "ontological" form for production databases, it has been extensively used for research, most notably in the NeOn Project[2], an EU-funded project of 14.7 million Euros involving fourteen partners in seven countries for four years starting in March 2006. The NeOn Project aimed at providing "lifecycle support for networked ontologies" in large-scale, distributed applications. FAO's role in the project — carried out by the AIMS team in cooperation with FAO's fisheries department — was to implement a prototype Fish Stock Depletion Alert System in support of the long-term goal of sustainable fisheries.

The role of the AIMS team in implementing the alert system was to integrate a diversity of data sources into a decision support system — sources ranging from land and fishing areas (identified using geographical coordinates), to biological entities (including family and species), fisheries commodities (using global statistical codes), fishing vessels (types and sizes), fishing gear (using a global classification scheme), and images from a variety of Websites. Related concepts needed to be aligned; water areas needed to be related to neighboring land areas. The objective was to federate the independent ontologies under a common queryable data infrastructure.

In 2003, a previous project in-house at FAO had attempted to build a comprehensive monolithic fishery ontology as a central focus for mappings from stand-alone databases, but work had bogged down with modeling issues, and the resulting construct was impractical and unwieldly. The NeOn approach, in contrast, was that of a "network of ontologies." It assumed that datasets would continue to evolve within specialized communities of practice, each of which in turn comprised the diverse perspectives of managers, biologists, IT systems administrators, and thesaurus maintainers.

3.2 User experience of AGROVOC and AIMS ontologies

AGROVOC is found useful by its users. Overall, in terms of coverage, it appears to address the needs of its target audience well:

[2]http://aims.fao.org/website/NeON/sub2

> We use AGROVOC thesaurus, primarily for cataloging and geographic information of our products, in conjunction with Library of Congress and find it very useful because AGROVOC covers more the agricultural point of view than LC, which is more general.

The AGROVOC Thesaurus was a loose, sprawling collection of terms added over of the course of thirty years by innumerable unnamed contributors and encompassing common and scientific names for bacteria, viruses, fungi, plants, and animals, as well as geographic names, acronyms, and chemicals. The terms all have something to do with agriculture or nutrition in a broad sense, but the thesaurus does not reflect any particular context, viewpoint, or application requirements. “Petroleum,” for example, is narrower than “mineral resource” and related to “fuels”; the related term “oil spills” is narrower than “pollution,” and “pollution” is narrower than “natural phenomena.”

One important achievement of the re-engineering process of the past few years has been to “clean” the ontology by consolidating hundreds of top terms, linking hundreds of “orphaned” concepts, and correcting thousands of other inconsistencies.

The process of refining semantic relations, described above, has added more precise relationships, but the process has not been guided by an overarching framework — e.g., viewing the entities from a particular scientific, commercial, farming, or environmental standpoint. The semantic multivalence of the terms is augmented further by the subtle differences of perspective and interpretation introduced by their translation into nineteen languages. Advanced ontological reasoning, however, presupposes a commitment to an ontologically well-defined point of view. One user finds the effort to refine relationships useful in principle but hard to exploit in practice:

> For our resource-discovery purposes, we cannot really apply the more refined relationships. I do not see how they can work — at least we do not have the technology to use them for resource discovery. You need an inference engine that can use them. Without an inference engine and a purpose, it is not clear what to do with them.

Another believes the effort to refine relationships has been useful but explains that their particular application requires relationships to be refined *differently*.

Extracting a sub-set of AGROVOC concepts as a starting point, they have refined the terms into an ontology in their own particular way.

A recurring theme in user feedback is the case in which developers set out to create expert systems, using well-engineered ontologies, such as for text mining or decision support, and ended up falling back on less sophisticated uses for the ontology such as simple query expansion and structured browsing. One FAO partner recounts the challenge of building a sophisticated ontology application with domain experts in the field:

> A group of extension officers in plant protection first tried to make a sophisticated portal on pesticides — a resource that extension officers could consult to help farmers diagnose plant diseases. They tried some complex solution and at some point, they completely gave up. They know the reality, they know their plants and all the relationships — the reality they know is so complex — but they couldn't use it to build an information system. They lacked the knowledge for creating a search assistant with an inference engine. The lesson we learned was that getting the various experts together, identifying the relevant material, and submitting it to the system, was actually more important than the highly codified system that resulted. In the end, we're talking here about references to just 1,000 research reports — and that is quite a lot for a specialized field! Once we identified those 1,000 reports, we did not need overly refined discovery methods.

One FAO technical officer with experience in ontology projects feels the requirements for reasoning functionality were never properly clarified:

> The few ontologies in FAO are not exploited fully in terms of reasoning capability, and there are no real specific requirements for reasoning. The real requirements, like language independence and collaborative maintenance, do not require rules and reasoning. Maybe we should investigate whether we really want to have a basis for full-fledged ontologies. Maybe researchers were pushing for more functionality than really required.

Other users confirm that their needs are quite simple — better navigation, search refinement, or ranking hits:

> We have used ontologies in vertical portals to index or classify things. We use OWL formats, but more like thesauri. With mappings, we can continue using legacy thesauri. We find we get better navigation; they help in ranking hits and refining searches.

Another reported a preference for a legacy classification scheme that still forms part of the AIMS offering but has been de-emphasized:

> What we have used from the FAO standards is the AGRIS/CARIS Classification Scheme, though that is not what excites [the ontologists] the most. We use it for news feeds. However, this taxonomy is old-fashioned — it is based on agricultural production, but out-of-date on genomics and environmental aspects. There is no procedure for feeding terms back to AGRIS/CARIS, because this standard is no longer maintained. The classification scheme was inherited from the printed version of the index in the early 1980s.

A colleague in a FAO technical department would like to use AGROVOC to tag reports and publications:

> Increasingly we have stuff to tag: meeting reports, publications, duty travels, case studies. Much mundane, day-to-day stuff. If we had it "in AGROVOC," we could do interesting things. "Where are meetings duty travel reports, institutions, and Web pages we have done about, say, fungus?"

Fishery experts in the NeOn Project express enthusiasm about the potential of ontologies to guide decision-making but recognize that the methods may take a few years to mature. For the AIMS team, the project confirms that the maintenance of alignments within a "network of ontologies" is time-consuming and error-prone, especially between ontologies based on different underlying models (e.g., class- versus instance-based) and between ontologies that are independently evolving towards new versions. Recognized bottlenecks are the lack of tools for automating such tasks and the lack of reliable corpi with which to test automatic alignment methods.

3.3 AGROVOC as a "quarry" of terms

The goal articulated for the Agricultural Ontology Server in 2001 was that of providing "building blocks" for application-specific ontologies. Feedback from users strongly confirms that this is indeed how AGROVOC is being used, only not for the sophisticated applications originally envisioned. In practice, AGROVOC serves as a quarry of conceptual blocks to extract as a starting point for customized vocabularies:

> We need specific vocabularies in many areas. Making derivative products from AGROVOC — terms relevant for a particular area — is what people want to have: go one level down, slice up the pie with very specific terms in a particular area.

Sets of AGROVOC terms often provide a starting point for creating specialized portals about topics like "crop pests" or "bananas." The Organic Edunet[3] used AGROVOC as a starting point for their own set of categories, mapping to AGROVOC wherever possible and inventing the rest. It is simply more efficient to re-use an existing vocabulary than to try to invent one from scratch:

> We need something between Yahoo and Dewey and more specific. It would take alot of discussion to come up with our own. We use taxonomies both for indexing and for creating the structure of Web pages. For each entry in the browsing structure, we want to have a query to the database using subject headings.

In its entirety, however, AGROVOC is simply too big:

> Using all of AGROVOC is cumbersome — putting whole thing into peoples' hands is too much. We want to make a sub-vocabulary. We are moving towards full-text indexing and need vocabularies for very specific portals.

Given the wide range of audiences for which AGROVOC is used, however, the semantic multivalence of its terms is actually desirable. One project in India needs

[3] http://www.organic-edunet.eu/

to customize browsing structures for users ranging from scientists to agricultural extension works and semi-literate farmers. Another user reports:

> We have customers who produce portals for regional development — specific birds, sheep, things in meadows, how to manage meadows in specific ways. We need taxonomies to create a browsing structure for our portals, and not just from a scholarly perspective.

Many users see an inherent tension between centralizing quality control over the maintenance and expansion of AGROVOC with experts in the AIMS team as opposed to devolving control to user groups and language communities with their own local requirements:

> I see a need for lots of country-specific AGROVOCs — for India, Brazil, etc. Everyone has very specific terminology. It is not doable to capture all of these variants in the central AGROVOC ontology. We need distributed vocabularies.

Decentralizing maintenance control, however, implies capacity development — instruction about ontological principles and training in the use of specific tools and procedures:

> AGROVOC is understaffed for the task of maintaining AGROVOC, allowing new concepts without duplicating or creating a mess. One always has to check and think before entering a term — it is not a mechanical job for a clerk but involves brainware. KCEW could explain tagging as a capacity-building effort. This could be useful but would conflict with the maintenance task. There is possibly a built-in friction between the two roles.

Users see this as a crucial role for a United Nations team:

> FAO provides AGROVOC to download and use, but just as important have been the people who provide support. This is extremely helpful! They bring new ideas. As a UN organization, FAO should have this role — to help solve problems.

> In countries such as Thai, Laos, China, and Japan, data needs to be entered in the local language together with English. It is meaningless to create a database only for an international audience, which cannot serve local needs or people in their own country who cannot read. FAO should focus its efforts on strengthening collaboration and on supporting localization of agricultural information to better serve the needs of local users.

> There was talk of using AGROVOC in order to classify news items about FAO. But it seems less useful for KCEW to focus on activities within FAO itself than to use AGROVOC in communities served in the outside world. It should focus on providing stable environments for products such as AGROVOC, on bringing people together to exchange opinions, developing initiatives and services such as linked data — working more as a facilitator than as an implementer — and funding the participation of people from developing countries to participate.

> FAO should work on the normative layer — standards and norms. FAO has a normative role. Let others then use the standards in their work. FAO's role is that of paving the path for smooth operations.

Indeed, the centralized maintenance of AGROVOC is widely seen as a bottleneck:

> It can take three months to get a new term added to AGROVOC. At the time we propose them, we are using them to pull together some material. When they do not get approved quickly, we create a workaround, and once the workaround has been made, we continue using that.

Users feel that decentralizing maintenance would free the vocabulary to grow more quickly:

> AGROVOC is very strong, especially in geographic areas — we like it — but it evolves too slowly to keep pace with emerging research terms. Maybe we need vocabularies in a wiki or blog thing, like

> Wikipedia, where people can quickly post these things and start to adopt terms quickly — where terms can be proposed and used immediately.

Providing an environment in which a large number of users could participate in the maintenance of AGROVOC, as a community undertaking and in multiple languages was was, of course, the goal of the Workbench activity. More than three years after its inception, however, this effort has yet to produce a maintenance tool efficient enough to allow AGROVOC to be transformed definitively from a thesaurus into an ontology and from the legacy relational database environment into an ontology editor.

The more sophisticated Semantic Web uses for ontologies imagined in the early 2000s have not materialized in the AIMS user community. To some extent, this has been both a barrier to understanding and a source of tension between visionaries and practitioners. Ontologies have been seen as bleeding-edge research — a noble undertaking but impractically complicated for the average implementer. The simpler and straightforward goals of today's Linked Data movement, however, are seen by many users as a crucial way forward. It would seem that the goal of honing the precision of well-engineered ontologies stands at cross purposes with the goal of accommodating a broad diversity of language communities and user perspectives.

3.4 Correcting the model for *less* precision

Since the 2006 finalization of a metamodel for expressing a term-based thesaurus (i.e., AGROVOC) as an ontology of Concepts linked to Lexicalizations, the World Wide Web Consortium has finalized a W3C Recommendation for precisely this purpose: Simple Knowledge Organization System (SKOS) [8]. Indeed, a computer scientist from the AIMS team participated in the W3C Semantic Web Deployment Working Group which developed SKOS, and AGROVOC provided a key use case for the requirement that Labels (Lexicalizations) be defined as first-class resources [6]. It is fortunate that the AIMS team has not yet finalized the conversion of AGROVOC from thesaurus to ontology or promoted the URIs of its concepts, modeled as OWL classes, for use in Linked Data, because the shift to a SKOS metamodel can still be undertaken without breaking existing applications.

We have seen above that in practice, concepts are often extracted from

AGROVOC, like building blocks from a quarry, for often quite basic uses. Erring on the side of under-specifying concepts avoids imposing inappropriate ontological commitments and reduces the risk of their being reused incorrectly. Users of SKOS concepts in applications downstream do not inherit the transitivity and entailments of OWL sub-classing.

This ontologically more flexible approach to concept schemes also addresses a difficulty that has emerged in AIMS capacity-development activities. AIMS team members holding seminars at FAO partner institutions report that words like "ontology" and "concept server" are perceived as "confusing," even "scary," and that the finer points of ontologies, such as the distinction between classes and instances, are lost on many audiences. The distinctions are, of course, hard to teach in part because they really are hard to nail down or justify in practice. SKOS should be easier to teach, and with the rapid uptake of SKOS, AIMS trainers should benefit from the growing availability of tutorial materials.

The effort to refine AGROVOC concept relationships has underlined a need to standardize some frequently used properties such as "hasAcronym." The popularity of lightly defined concepts suggests, however, that the push to refine AGROVOC as a whole be given lower priority, moving forward, than the gradual extension of the concept set into new languages and subject areas. A colleague, Mark van Assem, finds vocabulary maintainers generally reluctant to complexify their vocabularies ontologically, as it is not always clear how refinements improve performance and user support, and suggests that vocabulary developers follow the adage "no innovations without clear applications."[4]

The AIMS namespace for AGROVOC currently defines 198 refined relationships, two-thirds of which constitute a "long tail" of properties used less than twenty times, or even just once or twice, as with "isAfflictedBy" or "hasBreedingMethod." The AIMS team may publish these properties as Linked Data, enabling their re-use in other projects, but the AIMS team will not have the resources to pursue their standardization in the global arena. Ideally, this task should be undertaken in the context of a standards organization, perhaps with the goal of starting with a manageable core of, say, fifteen popular and well-understood properties — a "Dublin Core" of thesaurus refinements. In the meantime, specifying all of the existing refinements as sub-properties of the original thesaurus relationships (Broader, Narrower, and Related) would allow an application to "dumb down" the refined relationships for simple purposes such as query expansion.

[4] Personal communication.

Guus Schreiber points out that due to the diversity of their perspectives, vocabularies cannot simply be "merged." Rather, the best one can realistically hope for is to make the vocabularies usable jointly by defining a limited set of mappings in a process of "vocabulary alignment." Published as Linked Data either as a part of AGROVOC or as a separate module, mapping assertions effectively increase the reach of AGROVOC concepts, allowing queries to be expanded to resources indexed with terms from related agricultural vocabularies such as CAT, SWD, NAL, GEMET, and CAB Thesaurus (see above) or more general vocabularies such as WordNet or the Library of Congress Subject Headings. Facilitating the creation of such alignments has been identified as a new priority for the Workbench project.

The impact of AIMS standardization activities has traditionally been measured by indicators counting the number of people and organizations engaged in defining, translating, downloading, and viewing the standards, as in an internal FAO assessment of 3HP02 activities in 2007. [**?**] New types of RDF aggregators and search engines, such Swoogle[5], have the potential to generate potentially vastly more explicit information about how vocabularies are being deployed, combined, and consumed. Using time-series statistics, it should become possible to count not only the number and location of resources referenced using AGROVOC URIs, but the number of documents linked indirectly — through concepts in vocabularies which, like the Chinese Agricultural Thesaurus, have been aligned with AGROVOC through mappings. By revealing trends now hidden from view, these methods will help demonstrate to information providers the practical advantages of tagging their materials using AIMS URIs.

3.5 Recommendations

1. Before promoting the use of their URIs in linked data, convert AGROVOC and other AIMS ontologies from the 2006 OWL-class-based metamodel into SKOS concept schemes (see Appendix B). Rename the AGROVOC Concept *Server* as the AGROVOC Concept *Scheme*, and the development environment as the AGROVOC Concept Scheme Workbench.

2. Complete the migration of AGROVOC to a Workbench environment, based on SKOS, as soon as possible.

3. Re-affirm the role of AGROVOC as a "quarry" of "building blocks" for

[5]http://swoogle.umbc.edu

applications that may be quite simple, such as query expansion and structured browsing.

4. Publish the relationship “refinements” coined for AGROVOC as linked data but de-emphasize the creation of further properties. Consider the AIMS refinements as a provisional “stake in the ground” for properties that may eventually be standardized by global bodies that are better suited for this considerable effort than AIMS. If thesaurus refinements are standardized internationally, AIMS properties might be aligned with the new standard and their use gradually de-emphasized.

5. Emphasize the alignment of AGROVOC with related vocabularies and support the creation of alignment assertions in the Workbench. Aside from aligning with popular vocabularies such as WordNet, there may be interesting opportunities in-house, e.g., with FAOSTAT.

6. As maintenance control over AGROVOC and related concept schemes devolves to the community in the context of the Workbench, re-orient the AIMS team towards capacity development — instruction about ontological principles and training in the use of specific tools and procedures.

7. Medium-term, explore the use of RDF aggregators and search engines such as Swoogle to statistically demonstrate the value of tagging resources with URIs.

Chapter 4

Networking, capacity development, and outreach

4.1 Achievements in outreach and capacity development

A significant part of the AIMS initiative falls under the heading “capacity development” — the development of partnership among international colleagues through distributed teamwork, workshops, and training seminars in member countries or at headquarters. Capacity-development efforts typically focus on the formation of information managers, local champions, and educators at regional universities and research centers (“training the trainers”), often with an effort to involve agricultural extension workers or reach out to farmers directly. Capacity development may involve on-site training sessions by FAO staff or research sojourns by visitors in Rome. Teaching materials have been developed to support these activities, such as the Information Management Resource Kit (IMARK)[1], a series of computer-based distance learning modules available over the Web or on CD-ROMs.

The AIMS team has helped build or provided training for regional initiatives such as the following:

- Red Peruana de Intercambio de Información Agraria (AGRORED), a network of public and private institutions for supporting agricultural science

[1] http://www.imarkgroup.org/modulesintro_en.asp

and innovation in Peru with an emphasis on technical exchange and information management standards.

- The Kenya Agricultural Information Network (KAINet), a three-year national project funded by the UK Department for International Development (DFID), which among other things provided training in the use of metadata to participate in AGRIS.

- The Thai National AGRIS Center (TAC), established in 1980 as part of the Kasetsart University Central Library, which was an early adopter of the AGRIS application profile as the basis for merging content from twenty national research institutes and making it freely available on the Web. TAC has translated AGROVOC concepts into Thai and added concepts specific to the Thai context. As a major provider of specialized agricultural terminology in Thai, AGROVOC ranks high in Thai-language searches on Google for topics related to agriculture.

- The National Agricultural Research Information Management System (NARIMS) in Egypt, a bilingual Arabic-English Web portal for information about research in Egypt related to agriculture, which was developed in cooperation with FAO staff and using FAO tools and standards, notably an Arabic version of the AGRIS application profile. Starting in 2010, NARIMS data will be harvested by Near East Agricultural Knowledge and Information Network (NERAKIN), a platform for agricultural research organizations in the wider Near East region (Egypt, Iran, Jordan, Lebanon, Morocco, Qatar, Oman, Sudan, and Yemen) and, from there, ingested into the central AGRIS database.

- The Global Forest Information Service (GFIS)[2], a portal for information sources related to forestry, from maps and datasets to grey literature and journal articles. The GFIS Consortium aims at making information about the forestry resources of national and regional member initiatives more easily findable by scientists, planners, business people, educators, and private citizens through a single point of entry. GFIS worked closely with FAO on the design of their service and put their Website online in 2005. GFIS functions similarly to AGRIS inasmuch GFIS information providers submit metadata through Service Centers for conversion into a Dublin-Core-based application profile.

[2]http://www.gfis.net

The story of several related projects in India exemplifies the role that the AIMS team can play in developing capacity on several levels. Starting in 2002, the Indian Institute of Technology in Kanpur experimented with using the Web to help semi-literate farmers bypass intermediaries to sell their commodities online. The initial idea of promoting digital commerce failed for lack of uptake, but the project did confirm a need to transfer knowledge about crops (such as dal and sugar), farming methods (sericulture and pest control), and agrarian legislation from India's 11,000 or so PhD-level agronomists to its 100 million farmers to address issues such as crop rationalization, declining soil fertility, the after-effects of chemical use, and pest pathologies.

The initiative enlisted the collaboration of village-level agricultural extension workers in bridging this gap and aimed at disseminating information in broadly consumable forms such as radio broadcasts, comic books, and SMS alerts, written or spoken in the rural vernacular. One strategy for making research outputs accessible to a broader range of participants was to tag available materials with familiar concepts, so parts of the AGROVOC Thesaurus were translated into Hindi and Telugu.

A larger-scale National Agriculture Innovation Project, "Agropedia,"[3] was launched in January 2009 to empower farmers and extension workers with crop- and region-specific information and "accelerate technology-led, pro-poor growth and diffusion of new technologies for improving agricultural yield and rural livelihood." A brainstorming workshop with seventy participants of diverse background generated knowledge models reflecting scientific, clinical, and practical perspectives on the management of key crops such as rice, pigeon peas, and sorghum.

Taking AGROVOC concepts as a starting point, the participants used simple open-source software to define entities and relationships. Experienced ontologists from FAO helped apply standard naming conventions and map the emerging relationships to existing properties in AGROVOC. The workshop served both as a capacity- and a community-developing experience. The resulting knowledge models link local terminology to standardized, language-independent concepts usable for tagging research outputs and learning materials, whether by manual metadata creation or automated keyword extraction, and to access those materials from a variety of perspectives.

[3] http://agropedia.iitk.ac.in

4.2 Fishing in a Sea of Agrovoc?

In 2004, an autoevaluation with focus groups at FAO identified the need for "a prolonged effort to monitor the departmental sites, put a coherent layer of metadata over the different information systems (building on already existing metadata), and do some quality assurance in order to bring some order to the FAO site and better index it." The evaluator reported that previous efforts to put order to the proliferating departmental sites "was never a pretty process; a lot of tension was involved between divergent departments. Everybody is so busy with service/divisional work that coordination is viewed as a burden." Another consultant's report from the following year made a similar observation:

> It was generally recognized and agreed that interoperability across different information object types and applications is necessary and methodologies have to be developed and applied to facilitate this interoperability. Important lessons have been derived from the experiences of different standard-setting communities: Standards should be kept simple, to facilitate their adoption by data owners.

There have been a few cases of successful cooperation between the AIMS team and technical departments within FAO, notably with Fisheries (in the NeOn Project) and Forestry, involving primarily the use of AGROVOC for indexing, Agrifeeds for disseminating information about events, and the use metadata for describing departmental outputs. Overall, however, the observations made in 2004 appear still to apply five years later.

One technical colleague at FAO, however, offers a compelling metaphor for what might possibly be achieved in such a diverse institution:

> There is absolutely a need for more communication between departments at FAO. Everything we do can be seen from multiple angles: capacity building, research, women and development, democracy. If we were swimming in a Sea of AGROVOC, and we were to cast our hook for 'Climate Change,' what things might we pull up?

The same colleague argues that such an approach is essential for preserving and transmitting institutional knowledge in a faster and more mobile age:

> There is quicker turnover now. With quicker staff turnover, institutional memory becomes a bigger problem. I used to be the youngest person in my department, but in the past three or four years, there have been more retirements. Who can tell me what meetings were held?
>
> Looking at shared network drives is terrifying: gigabytes and gigabytes of stuff. Nobody knows which information is useful, and which parts are private; nobody will ever sort this out. If a colleague dies, how can we tell which data is important? Maybe we are ditching librarians too quickly.

The environment in which the organization works also has become more crowded and competitive:

> We need to make the products of our organization more findable. Now we are competing for attention because of information overload. We want to make them findable when people do a search from anywhere, like Google.

Various colleagues offer suggestions for how such a project might be undertaken and where it would start:

> To succeed at FAO, more should be done to convince content owners to use the standards. They stand a better chance of success if they can be used behind the scenes, with automated processes. Perhaps clever ways to configure content management systems to assign default keywords. What would work is clear examples of how the standards increase the visibility of their content.

> To get buy-in, they would need a demonstrator that could be scaled up. This means starting by getting something to work — for example, getting two departments to combine feeds. If we knew a metadata standard for meetings we could just use, chances are better we would just use it.

> There are a handful of information types that account for most needs, day-to-day: meetings — not only in-house, but meetings in which

FAO Is involved — projects, contacts, organizations, and references (annotated bibliographies).

How might such a vision be achieved in practice? One well-developed model is offered by the VIVO service, managed since 2003 by the Cornell University Library as a structured view of information about people and academic resources at Cornell University.[4] The sample of VIVO suggests the following advice:

- Start small, with a few common content types — people, departments, courses, publications — and extend the supported types organically, based on growing relationships to people, activities, and organizations.
- Work with departments and administrators to promote a more uniform approach to self-reporting and demonstrating Return On Investment in the form of improved data consistency and higher public visibility.
- Ingest data from departments and databases with as little manual intervention as possible, adapting automated ingest procedures to specific local data structures and using simple inferencing to enrich data records with information not explicitly encoded in the source databases (e.g., "member of life science field") and, where possible, enriching or replacing text values with URIs.
- Convert data into an open and consistent format, using explicit semantic relationships, and publish the data according to accepted Linked Data principles, avoiding a requirement that any one tool be globally accepted and anticipating instead the future availability of innovative alternatives.
- Present users with a clean, Google-like search box in recognition of the fact that people typically submit queries of just one or two words.
- Take the user from a single-word query to a page that assembles links clustered by type — people, events, publications, institutions, and topics — efficiently exposing the searcher to response sets of high quality and providing a structured browsing experience based on semantic relationships.

[4] http://vivo.cornell.edu

4.3 The global "coherence" of information about food

The AIMS initiative sees itself as part of a broader movement for improving the management of, and access to, agricultural information. FAO is part of an initiative that has coalesced under the banner of Coherence in Information for Agricultural Research for Development (CIARD). The CIARD initiative was the result of expert consultations held in 2005 and 2007 under the name International Information Systems for Agricultural Science and Technology (IISAST). FAO is one of the fifteen core members of CIARD.

CIARD presents a broader context in which AIMS can be effective. Where AIMS focuses on information standards, especially the AGROVOC thesaurus and AgMES-based application profiles, with AGRIS as a key implementer, CIARD represents a broader community, institutional base, and scope of action, with Task Forces on Advocacy, Capacity Building, and Content Management. The CIARD Content Management Task Force advocates the use of common standards for enabling the integration of information across institutions. The CIARD Pathways to Research Uptake offer concrete advice on broader issues, such as licensing and open access, techniques for retrospective digitization, policies for sustainable repositories, digital preservation, the exchange of information about news and events, and effective Website management (Web 2.0, search engine optimization, social media, and the use of Web analytics).[5]

The notion of "coherence" fits beautifully with the message of Linked Data. We live in a diverse and rapidly evolving world in which it is unrealistic to expect that interoperability can be tightly coordinated on the basis of mandatory data formats and specific technical solutions, whether by "lock-step" agreement among big institutions or by the de-facto dominance of specific software platforms. RDF provides an open-ended data model that explicitly avoids requiring that providers information in identical formats — a goal which can only remain, in the best of circumstances, elusive.

Rather, the watchwords of this more loosely-coupled vision of interoperability are "alignment," "harmonization," and "partial understanding." The best we can hope for is "coherence" in the underlying data itself — to ensure that the data can be expressed as, or translated into, RDF triples that can be coherently merged on the basis of shared descriptive properties, shared value vocabularies, and shared resource identifiers.

[5] http://www.ciard.net/index.php?id=607

History shows that all technology is transitional. Most of the applications and data formats we use today will become obsolete in the coming decade. RDF triples represent knowledge in the form of a simple sentence grammar, using noun-like classes and verb-like properties to make statements about things in the world — statements that are expressible in, and freely convertible among, multiple concrete syntaxes.

As of 2010, there are no other compatible models for representing knowledge with the uptake and traction of RDF. For the foreseeable future, RDF offers our best hope for "future-proofing" our cultural and scientific memory. As our applications and formats inevitably lapse into obsolescence, we can only hope to retain the ability to interpret what remains. We must ensure that our data is expressed in a form that we can flexibily re-use today and pass to the next generation tomorrow, especially as it relates to nutrition, agriculture, and the sustainable use of the Earth's natural resources.

4.4 Recommendations

1. Consider advocating a strategy for integrating access to the outputs of FAO technical departments on the basis of Linked Data — with emphasis on tagging materials using AGROVOC URIs (the "Sea of AGROVOC" metaphor) — using successful projects such as VIVO as models.

2. Promote the idea of Linked Data in the CIARD community; the messages of "coherence," "alignment," and "partial understanding" are good fits to the CIARD message and could constitute an additional point in Group 3, Making Content Widely Accessible on the Web, of the CIARD Pathways to Research Uptake[6].

3. Either define the acronym "AOS" (by describing its historical background) or consider dropping it. If AOS is kept for reasons of brand recognition, its "server" aspect should be played down in favor of a more RESTful, data-centric message.

4. In the medium term, consider replacing the Glossary, FAQ, and Registry of Tools — the scopes of which are unsustainably comprehensive in their current form — with short documents more tightly focused on AIMS (see Appendix C).

[6] http://www.ciard.net/pathways

Bibliography

[1] Brickley, Dan. 2003. Missing isn't broken: data validation and freedom on the Semantic Web. FOAF Project Blog, http://blog.foaf-project.org/2003/07/missing-isnt-broken-data-validation-and-freedom-on-the-semantic-web/.

[2] Caracciolo, Caterina. 2009. D7.2.3. Initial Network of Fisheries Ontologies. NeOn Project. http://www.neon-project.org/web-content/images/Publications/neon_2009_d723.pdf

[3] Gruber, Thomas. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* 43(5–6]: 907–928.

[4] Hepp, Martin. 2007. Possible Ontologies: How reality constrains the development of relevant ontologies, Martin Hepp. *IEEE Internet Computing* 11(1): 90-96.

[5] Independent External Evaluation of FAO. 2007. Rome: FAO. ftp://ftp.fao.org/docrep/fao/meeting/012/k0827e02.pdf.

[6] Isaac, Antoine, Jon Phipps, Daniel Rubin. 2009. SKOS Use Cases and Requirements. [W3C Working Group Note, 18 August 2009]. http://www.w3.org/TR/skos-ucr/#UC-Aims.

[7] McGuiness, Deborah, Frank van Harmelen, eds. 2004. OWL Web Ontology Language Overview. [W3C Recommendation 10 February 2004]. http://www.w3.org/TR/owl-features/.

[8] Miles, Alistair, Sean Bechhofer, eds. 2009. SKOS Simple Knowledge Organization System Reference. [W3C Recommendation, 18 August 2009]. http://www.w3.org/TR/skos-reference/.

[9] Sauermann, Leo, Richard Cyganiak. 2008. Cool URIs for the Semantic Web [W3C Interest Group Note 03 December 2008]. http://www.w3.org/TR/cooluris/.

[10] Soergel, Dagobert, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer, and Stephen Katz. 2004. Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information* 4(4). http://journals.tdl.org/jodi/article/view/112/111.

Appendix A: Promoting AIMS URIs

With the rapid success of Linked Data, FAO is uniquely well positioned as a source of trusted URIs. A key aspect of FAO's technology strategy should be to ground all elements of its vocabularies in URIs. As pointed out above, the fact that AGROVOC URIs have not yet been widely promoted offers an opportunity to correct its thesaurus metamodel without breaking applications. Currently, the URIs are only used behind the scenes, for processes such as query expansion. Rather, the idea is to empower resource providers to tag their own content in a way that links it to other resources in the Linked Data cloud without the intervention of a central aggregator.

Before promoting URIs, however, several issues will need to be sorted out:

- **Clarifying how AIMS URIs are composed**. This includes questions such as whether URIs should be minted under one's own domain or under redirect services such as purl.org; decide which terms are most appropriately identified with numbers (e.g. "c_3870") as opposed to word phrases (e.g., "hasAcronym"); whether URI strings should contain versioning information or date stamps; and whether base URIs should end in a hash sign a slash. Until recently, the document "Cool URIs for the Semantic Web" [9] provided the most up-to-date overview of options, but with the standardization of methods for embedding structured data in normal Web pages, such as RDFa, best practice in this area continues to evolve. The Pedantic Web Group[7] currently provides one of the most lively forums for discussion.

- **Clarifying a URI maintenance commitment ("namespace policy")**. AIMS should publish a document describing the institution's commitment to the long-term maintenance of its URIs as persistent identifiers and to ensuring that the URIs resolve to the latest version of their documentation, even if AGROVOC itself should cease to be actively maintained, or that redirects will be provided if retirement is necessary.

- **Publish vocabularies as Linked Data**. The AIMS team will need to decide between alternative ways to publish vocabularies as linked data. Until recently, the favored method was to use content negotiation to resolve a URI to a representation in HTML or RDF depending on the browser preferences transmitted with the request. [9] However, this approach

[7] http://pedantic-web.org/

involves using HTTP response codes and customizing server settings correctly — a complex process that is prone to error.

In the meantime, approaches based on embedding RDF representations in HTML tags using RDFa are gaining favor. Serving one HTML page with two representations — visible text and invisible RDF data — becomes easier when the creation of embedded metadata is supported by content management systems such as Drupal, and complex server configuration becomes unnecessary.

It should also be possible simply to download AGROVOC (and other ontologies) directly, in RDF/XML or N-Triples. The presentation of the Library of Congress Subject Headings in the id.loc.gov service[8] nicely illustrates all of these possibilities (i.e., direct download, resolution of URIs by content negotiation, and structured representations embedded in Web pages).

[8] http://id.loc.gov/authorities/search/

Appendix B: The AGROVOC metamodel

In an article for the Journal of Digital Information in 2004, the AGROVOC project team proposed “a conceptual model that provides the necessary structure to create precise semantics to facilitate the transition from traditional thesauri to ontologies” — in effect a “metamodel” for thesauri — that distinguished three levels: a Concept, a Lexicalization (or Term) designating the Concept, and a String manifesting the Lexicalization [10]. Each level of this model was seen as a first-class entity — e.g., one Lexicalization could could have a formal relationship to another Lexicalization.

The appendix to the 2004 article noted the need for a generalized specification for expressing Knowledge Organization Systems in RDF/XML that transcended the limitations of term-based standards such as ISO 2788. At that time, precisely such a standard, Simple Knowledge Organization System (SKOS), was in the early stages of development, transitioning from European project deliverable to editor’s draft of a W3C working group. Crucially, however, SKOS was at that time limited to Concepts described by literal (string) labels, falling short of the requirements for expressing the multilingual AGROVOC thesaurus.

Translating the 2004 AGROVOC metamodel into RDF/OWL, however, introduced changes and carried the model ontologically further (see Fig. 1):

- The distinction between Lexicalization and String was quietly dropped, helpfully simplifying the model. (It is worth noting that no requirement for distinguishing between Lexicalizations and Strings was later identified for SKOS [6].)
- The natural-language Terms of the AGROVOC Thesaurus were re-conceptualized as Lexicalizations (Labels) for underlying Concepts. Lexicalizations included preferred and alternative labels, synonyms, spelling variants, and translations in multiple languages. Descriptors were conceptualized as “preferred” Lexicalizations.
- Concepts were modeled as OWL Classes (i.e., as sets of things). [7]
- Each Concept-Class was associated with one Instance of that Class as a means of relating a Concept to its Lexicalizations. (This was done to meet a perceived need for description-logic-based computability, as declaring one Class to be an Instance of another Class sacrifices conformance with “OWL DL,” a constrained, description-logic-conformant sub-set of the more expressive but computationally intractable variant “OWL Full.”)

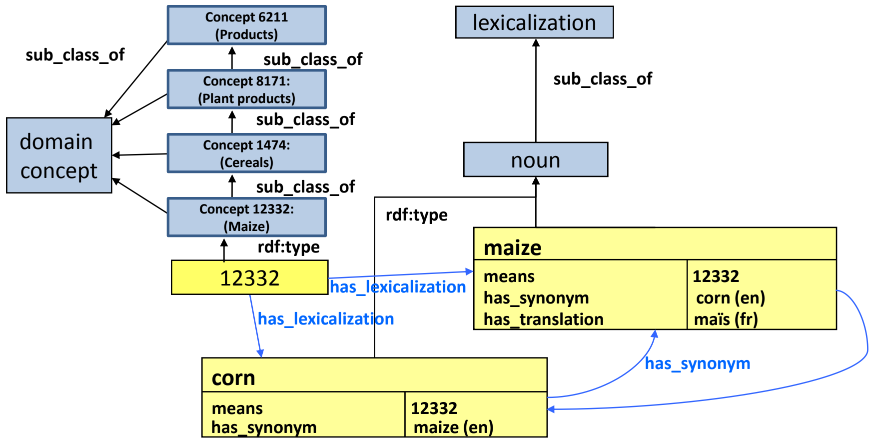

Figure 1: Metamodel for the AGROVOC Ontology, 2006–2010 (simplified)

- Relationships could also be specified between Concepts (such as “isUsedIn” or “causes”) or between Lexicalizations (such as “hasAcronym”). In 2006, this was considered a significant and innovative feature of the metamodel.

This model, which is currently reflected in the RDF/OWL representation of AGROVOC, has several problems:

- Lexicalizations are related to Concepts only by means of a parallel and artificially redundant set of Instances, which is both conceptually problematic and poses practical difficulties for software developers designing queries, mappings, and display interfaces.
- A Concept such as *Maize*, modeled as a Class, is declared to be a Sub-Class of the Class *Cereals* as well as a Sub-Class of the Class *Domain Concept*, whereas conceptually, *Maize* may more properly be seen as an Instance of a Domain Concept.
- Interpreting the Broader Term and Narrower Term relationships used between Concepts in the original AGROVOC Thesaurus as Sub-Class relationships between OWL Classes arguably constitutes “ontological overcommitment” (see below).

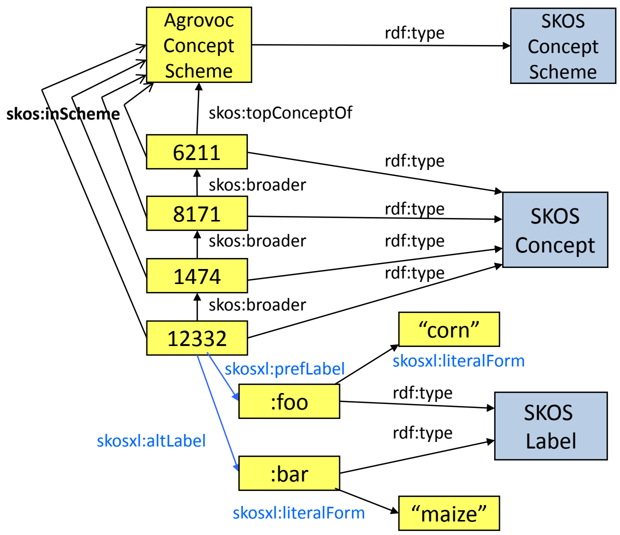

Figure 2: AGROVOC modeled as a SKOS Concept Scheme

The AIMS Website currently offers a representation of AGROVOC based the earlier, superseded version of SKOS which did not support the modeling of Labels as resources ("SKOS 2005") [**?**]. SKOS has in the meantime matured into a stable W3C Recommendation that supports the requirements identified for representing AGROVOC ("SKOS 2009") [8]. Indeed, AGROVOC had a significant influence on the design of SKOS itself by providing a well-articulated use case and requirements for crucial new features, such as Labels as first-class resources [6]. Since its finalization as a W3C Recommendation in August 2009, SKOS has become the de-facto standard for expressing Knowledge Organization Systems in a Linked Data context.

Figure 2 shows how AGROVOC can currently be expressed in SKOS: AGROVOC Lexicalizations (Terms) are modeled as instances of the class SKOS Label, AGROVOC Concepts as instances of the class SKOS Concept, and the AGROVOC Concept Scheme itself as an instance of the class SKOS Concept Scheme (see Fig. 2). With a subtle shift of wording, the AGROVOC Concept *Server* can be renamed AGROVOC Concept *Scheme*, and the development environment can be re-launched as the AGROVOC Concept Scheme Workbench.

This shift solves several problems with the 2006 AGROVOC metamodel, most crucially because SKOS provides a vocabulary for expressing the legacy thesaurus relationships between concepts not as ontologically strong sub-class relationships, but as ontologically weaker “broader” and “narrower” relationships. This is more appropriate for AGROVOC because the mechanical translation of thesaurus terms into OWL classes violates the design principle of *minimal ontological commitment*. As explained by Thomas Gruber [3]:

> An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities. An ontology should make as few claims as possible about the world being modeled, allowing the parties committed to the ontology freedom to specialize and instantiate the ontology as needed. Since ontological commitment is based on consistent use of vocabulary, ontological commitment can be minimized by specifying the weakest theory (allowing the most models) and defining only those terms that are essential to the communication of knowledge consistent with that theory.

SKOS concepts make a minimal ontological commitment to the nature of concepts and of relationships between concepts. Constructs consisting of SKOS concepts do not support the sort of advanced reasoning possible with tightly defined and constrained OWL classes, but they more faithfully reflect the flexible way that people actually think. SKOS concepts, by default lightly specified, prevent modelers from introducing false precision into their models, and they prevent inferencers from drawing unwarranted conclusions.

We have seen above that in practice, concepts are often extracted from AGROVOC, like building blocks from a quarry, for uses that more often than not are quite basic. Erring on the side of under-specifying concepts avoids imposing inappropriate ontological commitments and reduces the risk of their being reused incorrectly. Users of SKOS concepts in applications downstream do not inherit the transitivity and entailments of OWL sub-classing.

Declaring AGROVOC concepts as SKOS Concepts, on the other hand, does not *preclude* the use of OWL properties for defining relationships between concepts (properties) with more precision than plain-vanilla SKOS, e.g., as transitive, inverse, or symmetric. When appropriate, SKOS concepts may also be upgraded to OWL classes, with additional constraints, for use in local ontologies. (It is worth noting that the likewise lightly specified Dublin Core Metadata Terms are often upgraded locally from RDF into OWL properties, then more tightly constrained to support reasoning, and as there are endlessly different ways to do this, the minimal commitment of the Dublin Core specifications in this regard is widely considered a basis of their success.) Defining AGROVOC in SKOS does not, in other words, impede the development of applications that use reasoning.

Putting the Workbench onto a SKOS basis means that its developers will be able to benefit from software libraries and interfaces being developed for what is already the most widely deployed standard for Linked Data vocabularies. This will, in turn, make the Workbench more attractive for contributors from the open-source development community. Users will be able to process the RDF representation of AGROVOC, or an extract thereof, not just with the Workbench but with any SKOS-enabled software. Use of the Workbench will not depend on support for a metamodel unique to AGROVOC.

The conversion into SKOS will also resolve another issue that has emerged as a problem for AGROVOC — the presence of "classes" that should arguably be conceptualized as "instances." Examples include living species, chemicals, languages, and geographic place names, such as AGROVOC Concept 3253 ("Ghana"). AIMS team members holding seminars at FAO partner institutions report that the distinction between classes and instances is lost on many audiences. The distinctions are, of course, hard to teach in part because they really are hard to pin ail down or justify both in theory and in practice.

In SKOS, every Concept is by definition an instance of the class SKOS Concept — in other words, every concept is by definition an instance, and the only question is whether there is a meaningful difference between "concept-like" instances and other, "non-concept-like" instances. Although it has been suggested that SKOS Concepts be reserved for "concepts" instead of "real-world" things — or for "universals" rather than "particulars" — such distinctions are not understood widely enough to provide a basis for consistent distinctions. By design, at any rate, nothing in the SKOS data model prevents AGROVOC Concept "Ghana" from being considered a SKOS Concept. This should help make SKOS easier to teach than more advanced ontology engineering, and with the rapid uptake of SKOS, AIMS trainers should benefit from the growing availability of tutorial materials.

Forcing a distinction between classes and instances may, in fact, force ontological overcommitment. In order to map AGROVOC to an ontology for Aquatic Sciences and Fisheries Abstracts (AFSA), for example, the NeOn Project had to make AFSA comparable to AGROVOC by mechanically converting it into an ontology of OWL classes. On the other hand, while it seemed logical to the NeOn team that a species of fish be considered a class, and that actual fish be considered instances of that class, they found that when mapping to statistical time series, they needed needed to map species as instances. Indeed, the project team concluded "that the domain of interpretation of fisheries can contain entities as well as types of entities, and distinguishing them in a logically sound way would require a huge amount of fishery experts time, and only after they are organized in a team sided by ontology designers and are taught design tools adequately" [2] — a helpful warning against undertaking such a task lightly. Thanks to their ontologically light specification, in other words, SKOS vocabularies can more safely and easily be mapped.

It has been suggested that the maintenance of AGROVOC be rationalized by splitting the concept scheme into separately maintained modules and to phase out the maintenance of some modules in favor of pointing to vocabularies maintained by other organizations, especially for terms outside the core area of agriculture and for parts consisting of instances. The AGROVOC maintenance community is not well-suited to maintain identifiers for things like country names and species. While there may be good reasons to modularize AGROVOC and sharpen its focus content-wise, expressing AGROVOC in SKOS would remove the *ontological* need to distinguish instances and classes. Moreover, this issue need not be handled by deleting, deprecating, or devolving ownership of AGROVOC URIs. Rather, the same effect can be achieved by formally aligning its terms with other, actively maintained vocabularies and promoting the use of those other vocabularies, in specific areas, over the use of AGROVOC.

Appendix C: AIMS messaging and Website

Many of the benefits of ontologies can be seen as "services," but this is potentially a source of confusion. In early presentations and publications, the refinement of AGROVOC was depicted as the "conversion from a traditional thesaurus (AGROVOC) to a new system, the Agricultural Ontology Service Concept Server," described as a "multilingual repository of concepts" (2006) — in other words, from an *information structure* to a *service*. One presentation described an ontology as "a semantic system that contains terms, the definitions of those terms, and the specification of relationships among those terms" — i.e., as a conceptual construct — but added, somewhat ambiguously, that "such a semantic system can be referred to as an *Ontology Service*."

The acronym AOS dates back to a concept paper in 2001, if not before, and has meant either "Agricultural Ontology Server" or, more recently, "Agricultural Ontology Service" (hence AOS Concept Server). Usage has been, and remains, inconsistent. Over the years, AOS has been variously referred to as a "project," an "initiative," a "consortium of information providers," and a "clearinghouse for semantic standards."

Aside from its use today in an ongoing workshop series, the acronym AOS appears by now largely to have receded into the background. A link on the AIMS home page to "AOS Registries" points to a page which simply names the Agricultural Ontology Service without saying anything more about what AOS currently means. The legacy URL http://www.fao.org/agris/aos now simply redirects to http://agris.fao.org. The link to AOS Concept Server cited on the current Wikipedia page for AOS[9], http://aims.fao.org/cs.htm, points to a page that no longer exists. The AOS Concept Server Workbench (2006) is now called AGROVOC Concept Server Workbench.

That the "AOS" has largely disappeared has its advantages:

- **The word "ontology."** The word "ontology" is arguably associated by the general public with complicated and expensive research ventures.
- **The words "server" (or "service").** These are problematic because they evoke a Web of Application Programming Interfaces (APIs) running purpose-built Web services, which is arguably the dominant paradigm for Web applications today but locks data into dependence on software. An

[9] http://en.wikipedia.org/wiki/Agricultural_Ontology_Service

emphasis on services is difficult to reconcile with an emphasis on self-descriptive data and on plain-vanilla, HTTP- and hypermedia-driven ("RESTful") applications, which the reviewer believes to be far more "future-proof."

That said, International Business Machines is still called "IBM," "AOS" is a recognized brand, and there are advantages to continuity, especially for acronyms that are conveniently short for use in URIs. If kept, the Wikipedia page should be corrected and a short historical note should be posted to the AIMS Website.

The reviewer knows from experience with the Website of the Dublin Core Metadata Initiative, which has a similar purpose and scope to the AIMS Website, how difficult it can be to keep material fresh. There are some direct parallels in the Glossary, FAQ, and Registry of Tools, all of which began several years ago as centralized efforts maintained by hand:

- Analogously to the legacy DCMI Glossary, the AIMS Glossary tries to cover a broad range of concepts and acronyms of relevance to AIMS Standards, such as "IEEE," "URI," "Hypertext," and even "FAQ." The reviewer suspects that the AIMS team, like DCMI, will find it impractical to maintain this level of detail in such a rapidly evolving metadata scene. Like DCMI, it may conclude that it would be more practical to redefine the FAQ as a three-page, readable summary of just a dozen or two concepts and acronyms that are specific to the AIMS community — a page that could be downloaded and printed out for purposes of capacity development. For example, the Glossary might be a good place to define "AOS" as a legacy handle for today's activities. For definitions of things like "Hypertext" and "URI," people now turn first to Wikipedia.
- For the Registry of Tools, it may make sense to join forces (or simply point to) other efforts, such as the DCMI Tools Community, which maintains a Tools and Software page[10] that overlaps with the AIMS Registry, and the rdfa.info community, which maintains a page for RDFa Implementations page.[11] The rapid growth of Linked Data is translating into a rapid proliferation of new tools in this area for embedding RDFa attributes in Web pages, maintaining SKOS concept schemes, generating metadata as an integral part of content management systems, or automatically assigning

[10] http://dublincore.org/tools/
[11] http://rdfa.info/rdfa-implementations/

metadata on the basis of content analysis. Realistically, the AIMS Registry of Tools should perhaps point off to other lists, where known, and limit its focus to tools used in AIMS own capacity-developing activities.

- Analogously to the Glossary, the FAQ should perhaps be limited to three printable and readable pages of questions directly related to the development of AIMS standards. (The DCMI FAQ was similarly ambitious in scope and is being rewritten as a short document addressing DCMI-specific issues.) As with the Glossary, the FAQ should be concise and engaging enough to view as a single Web page or simply download and print. Here too, it might make sense to join forces with, or point to, DCMI or other related organizations regarding frequently-asked questions about metadata generally.