Anderson, K.L. & C. Thiery (eds.). 2006. Information for Responsible Fisheries : Libraries as Mediators : proceedings of the 31st Annual Conference: Rome, Italy, October 10 – 14, 2005. Fort Pierce, FL: International Association of Aquatic and Marine Science Libraries and Information Centers.

KNOWLEDGE RETRIEVAL IN AQUATIC ECOLOGY AND FISHERIES – DO WE NEED (AND CAN WE AFFORD) ONTOLOGIES?

Ian Pettman[1]
Librarian, Freshwater Biological Association,
Ontology Consultant to Freshwater Life Initiative

ABSTRACT:
What do we mean by the term "ontology" in the context of this paper?
Why might we (as information professionals) need ontologies and
where would we use them?
What sort of ontologies might we use and ***how*** could we build them?
Who would need to be involved?
When – what would they cost and when would that cost be justified?
Has IAMSLIC a role as a mediator in ontology building and use?

Introduction

This paper aims to start (or continue) a discussion within IAMSLIC concerning issues relating to terminology tools and controlled vocabulary systems for information and knowledge retrieval. The focus will be on ontologies in the aquatic sciences. Aquatic ecology and fisheries are two domains in the subject area where the author has some information retrieval experience combined with some experience of ontologies – hence the paper's title.

What do we mean by the term "ontology" in the context of this paper?

Terminology issues are central to effective information and knowledge retrieval. The term ontology is now widely used in knowledge retrieval circles but, needless to say, there is often confusion over what different people mean when they use the term. In other words, they are using one word to cover more than one concept!

---

[1] The author has been an active "SLIC" member for over 20 years. In 1996, after 25 years of providing library and information services for the Freshwater Biological Association in the UK, he converted to digital and spent 3 years working on Z39.50 and the inter-operability of distributed library catalogues for a European Union project. From the year 2000 he has worked on various web based information systems – mainly in the fields of fisheries and aquatic ecology (with interesting forays off topic into such fields as Rural Finance, Aquaculture genetics etc). He was one of the instigators of the FAO Fisheries Ontology Service project and worked on the initial stages. Having flunked retirement, he is back with the Freshwater Biological Association using both his library and digital experience. He also acts as Ontology Advisor for the Freshwater*Life* Initiative and collaborates with the Biological Ontology Department of Manchester University.

Some people use ontology in a very wide sense – covering all aspects of controlled vocabularies from simple term lists through taxonomies and thesauri to what are termed "formal ontologies" in the following diagram.

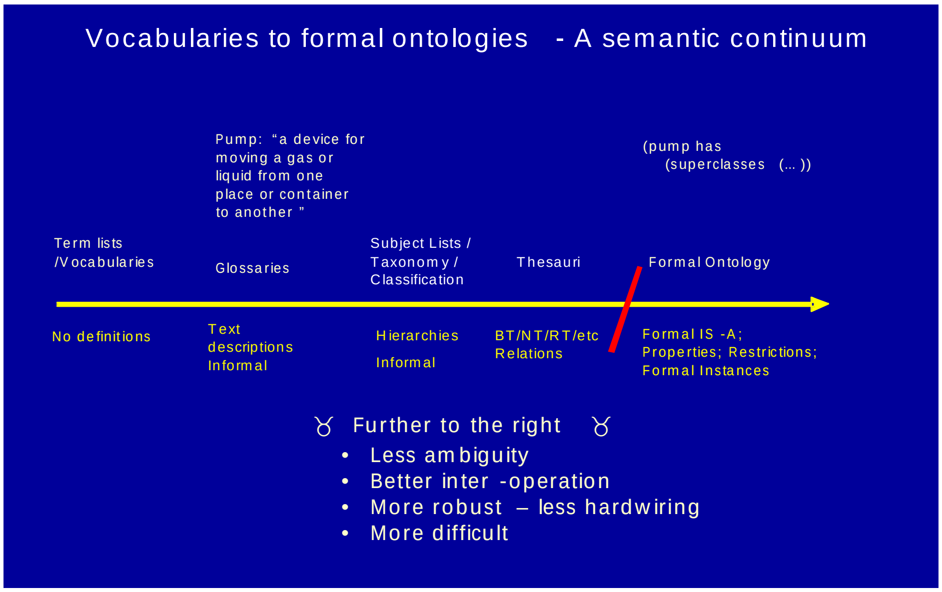

Diagram 1: the controlled vocabulary continuum.

The semantic tools to the left of the diagonal red line in diagram 1 are considered to be standard information retrieval tools that we are already familiar with. In this paper, we are only questioning the uses and financial viability of the "Formal Ontology" section to the right of the red diagonal line.

What are the differences between the "Formal Ontologies" and the other tools? Formal ontologies are only a further extension of the existing and familiar tools. The diagram illustrates, the complexity and hence difficulty in building an ontology increases beyond that of thesaurus building.

Some of the definitions of formal ontologies are concise – "An ontology is an agreed *formal* conceptualization of the world or of a domain".

Concise but not necessarily clear! Maybe we should use different terms (or words) to clarify the concept.

- An ontology = agreed concepts and their meanings for a subject area.
- The concepts are represented by various terms (or words or symbols)
- The terms represent the concepts completely or partially.

It is important to grasp this distinction between concepts and the terms used to refer to the concepts. A simple example might be as below:

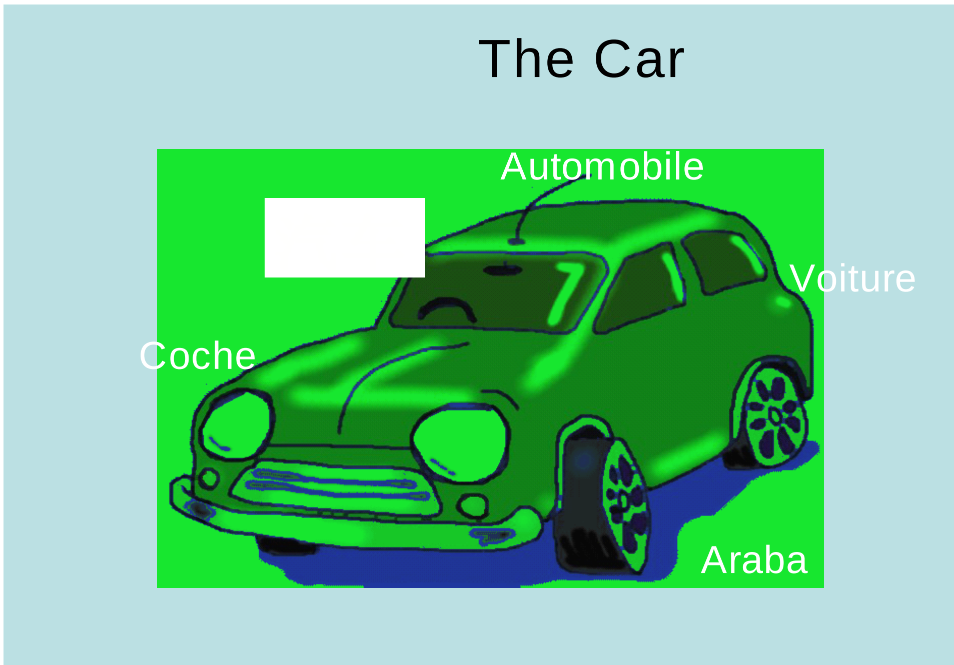

Diagram 2 Concept versus term - Courtesy of Manchester University

Here we have a concept (the green image) which in England we refer to using the term or word “car”. Other people use different terms and symbols to represent the same concept.

This is obvious and fairly simple. However, an ontology is more than this.

An ontology also specifies a set of **constraints**, which declare that which should necessarily hold in the domain. Or, put in clearer terms, it describes the meaning of the concepts and the relationships between the concepts. However, it does this using logic constructs not language sentences (Diagram 3).

Formal description of concept ŅFisheryÓ

```
(defconcept  FISHERY
        :IS-PRIMITIVE (:AND  ACTIVITY
                             (:SOME REGULATED-BY
                                      FISHING-REGULATION)
                             (:SOME MANAGED-BY
                                      MANAGEMENT-METHOD)
                             (:SOME INSTRUMENT
                                      DEVICE)
                             (:SOME TARGET
                                      AQUATIC-ORGANISM))

        :IMPLIES (:AND (:SOME  RESULT
                                      POST-HARVEST-USE)
                             (:SOME RESULT
                                      IMPACT)
                             (:SOME PERFORMED-BY
                                      CREW)
                             (:SOME PARTICIPANT-PLACE
                                      WATER-AREA))
```

Diagram 3 Fragment of the formal description of the concept "Fishery" from the FAO Fisheries Ontology.

The above is *only a small part of* the formal description for the concept "fishery" in the FAO Fishery Ontology Service draft ontology. Here the logical constructs are expressed in English.

What does a formal ontology look like? Visualization of an ontology is different to other controlled vocabulary tools. Ontologies cannot easily be printed out – there are too many relationships, restrictions and values to make this possible. They can be viewed as OWL files in XML but these are not easy for everyone to understand. They can also be viewed as node graphs with interconnections shown but these can get unwieldy as the ontology grows. One of the easiest ways at the time of writing is to view the files in an ontology editor. One of the more widely used of these is Protégé.

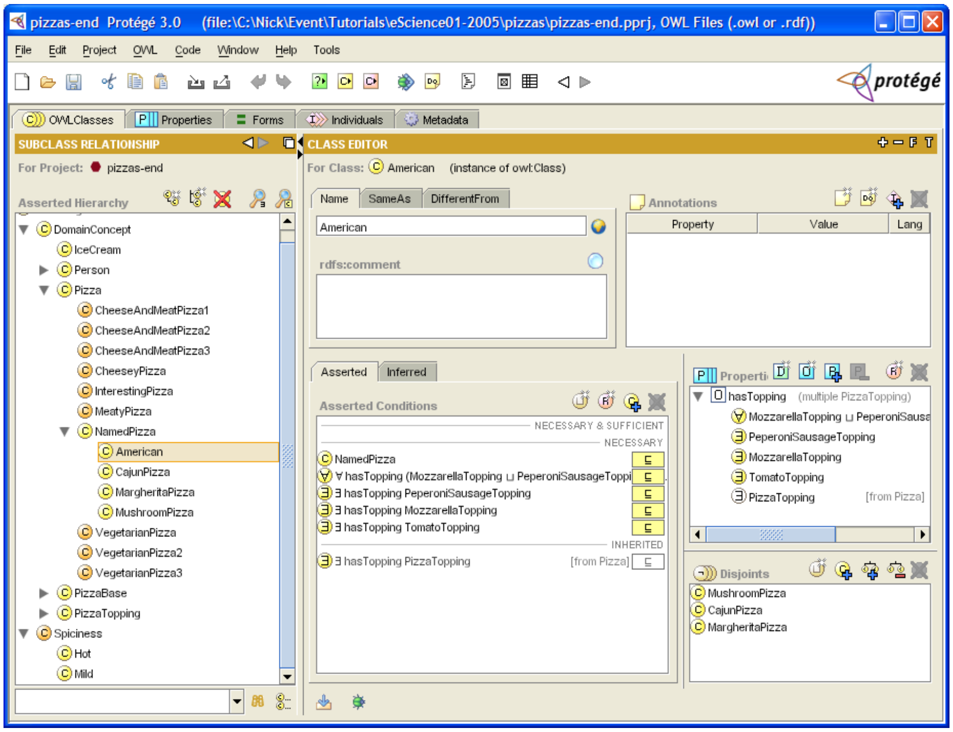

Diagram 4: a Pizza ontology in the Protégé ontology editor software

This is a view of the Classes Editing Screen and, in the left hand column, shows the Asserted Hierarchy of concepts in this ontology. This hierarchy is a strict IS-A hierarchy (e.g. Cajun Pizza is a Named Pizza; Named Pizza is a Pizza). Many thesauri and classification schemes have hierarchies that are mixed or non-IS-A.

This view also shows the Asserted Conditions (or Restrictions) which indicate that this Pizza (American) has three toppings (Pepperoni Sausage, Mozzarella and Tomato); the Properties (has Topping); and in the bottom right box that it is disjoint from the other named Pizzas (i.e. if it is an American Pizza it cannot be a Cajun Pizza etc.).

This brief introduction has indicated just some of the ways that an ontology is different from other controlled vocabulary tools.

Why might we (as information professionals) need ontologies?

This is an area that is wide open for discussion. The most interesting possibilities are beginning to emerge through developments in digital libraries and electronic document repositories. However, a look at some immediate advantages and disadvantages might be one place to start:

- Advantages of an ontology:
  - Extends interactive search and retrieve possibilities:
    - Choice from more than one concept

      For example, in a keyword search, more than one concept matches the word used. Using the ontology, the application can offer the user the choice e.g. Turkey – application offers the choice of "Turkey the Country" OR "Turkey the Animal".
    - Choice by relationship

      If the keyword has a rich complexity of relationships, this can help the user focus the query. For example, if the user searches for "Aquaculture" a range of choices could be offered – by environment (freshwater, brackish water, sea water); by organism (shrimp, tilapia etc); by technique (cage culture, tank culture, etc); by socio-economic factors (management, regulation, economics, etc) and so on.
    - Choice by language

      The application can give choices for the languages of the retrieved information. For example, I only want results in Russian or Japanese please.
  - Improves and extends the possibilities for manual and machine indexing and automatic data mining:
    - Manual indexing

      This can be time consuming and difficult. However, if the indexer has a similar interface to the ontology as the searcher options outlined above, manual indexing becomes quicker and more accurate and enables the above improvements in retrieval.
    - Machine indexing and data mining

      These are the first steps in to the more interesting possibilities. The increase in the number of digital documents now available has led to a lot of work being done in these areas. Ontologies not only play a role in making this possible, but they are essential if the end product is to be accurate.
  - Thesauri might be able to do a lot of the above – however, ontologies extend this by:
    - Using formal structures which enforce more "logical" thinking when choosing concepts and their relationships
    - Allowing the definition of a wider range of relationships between concepts and individuals

- Enabling logical restrictions and values to be applied to concepts so that ***reasoning*** is possible (we will return to reasoning later)

- Disadvantages
  - This is a rapidly developing field and hence some of the challenges are still research areas e.g.:
    - There are different ontology "languages" all of which still have some limitations
    - Many of the tools necessary for building, editing and maintaining ontologies are still being developed
    - Interaction between (and reasoning over) ontology "modules" and versioning of ontologies are still research areas
  - Ontologies can be difficult to build and "get right" – some would say that there is no such thing as a correct ontology?
  - They can be resource heavy to build (this is covered more in the costs and commons sections later).

Where would we use them?

Ontology building could be viewed as a lot of extra effort for present document retrieval requirements and it could be argued that this effort would be better expended improving existing thesauri and the user interfaces for these. So, where would we make this extra effort to build and use ontologies?

- The most likely uses will be in:
  - Digital Libraries
  - Web based Information Systems
    - Particularly Web based systems integrating Distributed Information Systems which include digital documents and data
- Some extra possible applications of ontologies for the above systems:
  - Improve document indexing
    - Automatic hyper linking to sources for identified concepts in documents or electronic resources (e.g. to a glossary definition of the terms, or to another web enabled document, etc)
    - Using "Open Office" documents or the next release of MS Word – tag concepts and relationships within a document using the ontology – this would then enable knowledge retrieval of data from within the documents.
  - Extend complexity of search and retrieve
    - Improved document indexing allows more complex searches over a range of both information and data sources

<u>Example of experimental complex searching over both data and information sources using the prototype FAO Fishery Ontology:</u>

A complex query was constructed:
"tell me what vessels from a nearby country are currently in the marine area 50N060W within the Atlantic Ocean, fishing for Thunnus alalunga stock using allowed techniques“

This query was run over a range of databases available to FAO using the services of the draft fisheries ontology and (at that time) received the response:

→ (Atlantic Enterprise II)

The result lists the name of the only vessel that (at the time of the query) was in the specified marine area (the *50N060W* quadrant in the *Atlantic Ocean*), was registered in a nearby country (its "flag state"), and had equipment able to fish – by legally admitted techniques, which in this case meant "not using *drifting longlines*" – the tunas of type *Thunnus alalunga* that were present in a stock located at the same marine area.

This is a real example but using more data sources than bibliographic sources. However, if electronic documents are tagged with the ontology concepts and relationships, it then becomes possible to retrieve data from within digital documents themselves.

What sort of ontologies might we use?

So, if we think that we have valid uses for ontologies, how do we start? The first choice is what ontology language to use.

This is a topic in itself, the details of which fall outside the scope of this paper. We have said that our use emphasis relates to digital libraries and web-based knowledge retrieval systems. This leads us at the present time to choose OWL as our starting point. What is OWL?

- Web Ontology Language (OWL)
    - The latest standard in ontology languages from the World Wide Web Consortium (W3C).
    - There are three ‘species' of OWL - OWL-Lite; OWL-DL; and OWL-Full.

OWL-Lite is the simplest to implement but is limited in scope and functionality. OWL-Full has the most functionality and allows most things to be done. Unfortunately, this functionality also means that it is “non-computable” i.e. it is not possible to use a reasoner software with the ontology.

- OWL-DL is based on description logics and can be computed.

For ontologies that fall into the scope of OWL-DL, we can use a ***reasoner*** *to infer extra information and further concepts in an ontology. This also opens up more opportunities for innovative knowledge retrieval in the future.*

Reasoners

The following is a simple example of the use of a reasoner. Let us look again at the Pizza ontology.

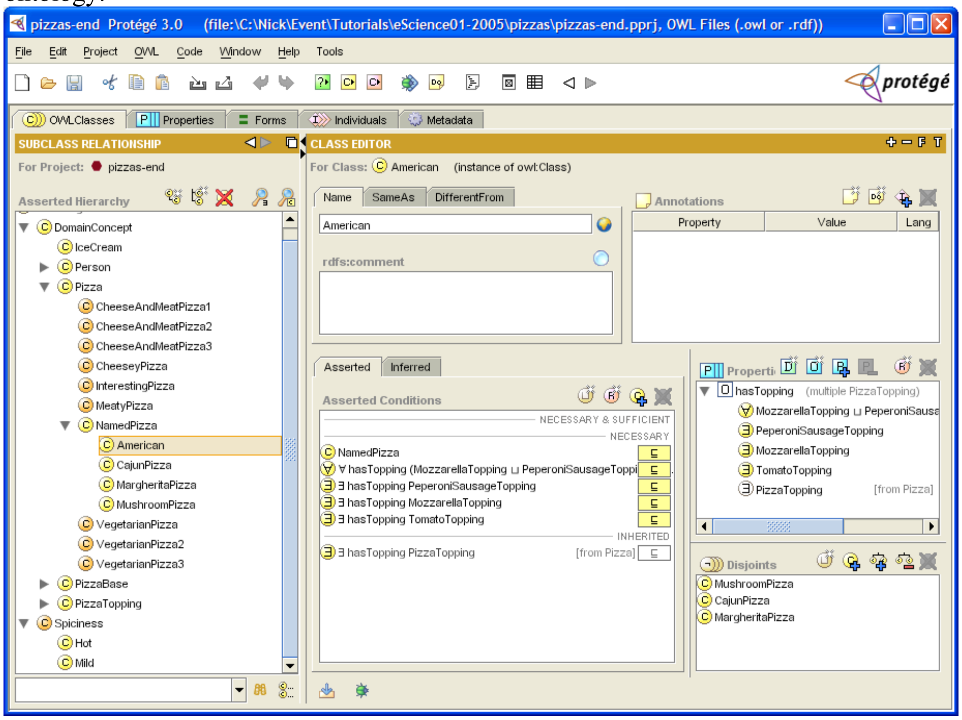

Diagram 5: the Pizza ontology as in Diagram 4.

This is the ontology as we built it. Note that there are no sub concepts under Cheesy Pizza or Interesting Pizza. Note also that the Named Pizza "American" has three toppings in its formal description (Pepperoni Sausage, Mozzarella, and Tomato).

What we cannot see on this screen shot (but is there) is that the formal description of Interesting Pizza says that an Interesting Pizza must have 3 or more toppings.
[Also, the Cheesy Pizza concept has the formal description which says that a Cheesy Pizza must have a topping from one of those listed as Cheesy Toppings under the Class Pizza Topping.]

If we now run the reasoner against the ontology, we get an "inferred Hierarchy" column in our ontology editor. This column shows the new ***extended*** ontology as built by the reasoner using Description Logic algorithms and based on the formal descriptions that we gave to the concepts that we entered.

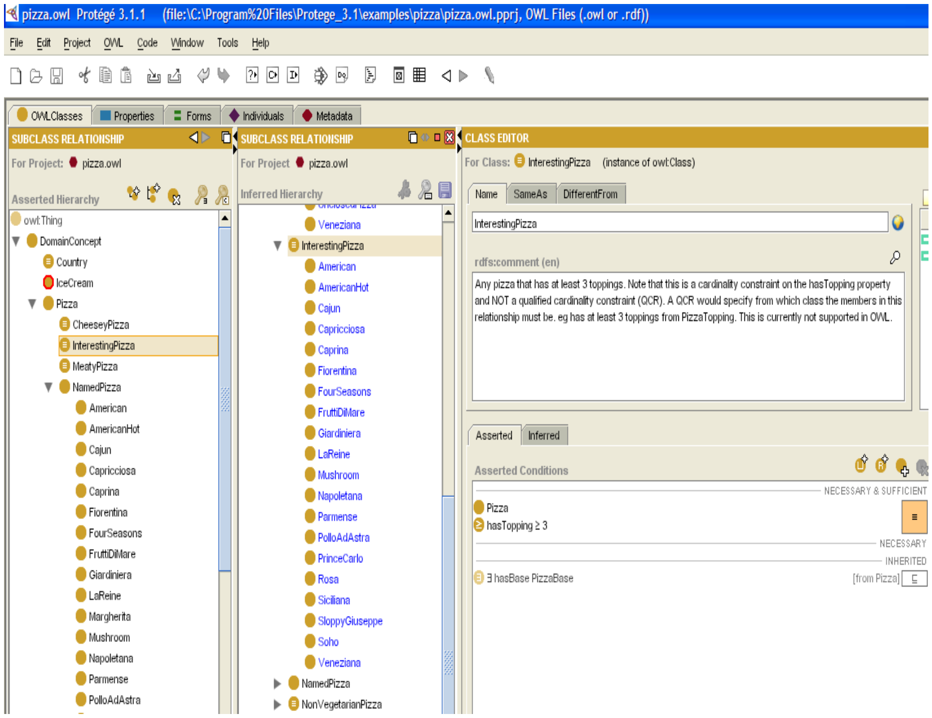

Diagram 6: the Pizza ontology after running the reasoner software

The reasoner has inferred that all the Pizzas listed in blue are Interesting Pizzas. It has inferred this from their formal descriptions. It has done the same for Cheesy Pizzas.

This example shows a first simple but useful uses of a reasoner.

- It simplifies ontology ***building*** by eliminating the necessity for us to add all the concepts to all other relevant concepts in a multiple tree (if we had to add all the Named Pizzas to all the relevant concepts, it would be time consuming and prone to omission errors.
- It simplifies ***maintenance*** of the ontology – if in the future we add another Named Pizza, the reasoner will automatically add this to all the other relevant concepts such as Interesting Pizza.

The possibilities reasoners open up for knowledge retrieval have still to be fully explored and exploited.

How could we build them?

Now we have chosen an ontology language, how do we start to build an ontology? Again, this is a topic for a paper in itself.

An obvious but important point to make is that they need to be built with a purpose in mind – i.e. need to have an application as a focus. In this way there is a chance of building a "correct" ontology for your needs. When there are difficult choices of what to include and what to exclude, on how to model a particular set of concepts, etc (and there will be such difficulties) the answer usually comes from looking at your application focus.

There are two main approaches to the build although many ontologies are a hybrid of the two:

1. Build from existing vocabularies, thesauri etc in the subject area.
    - This results in the need for mapping between vocabularies (not an easy task – and one which may need almost as much work as starting from "scratch")
2. From "scratch" - By extracting knowledge from subject experts.

Tools and techniques are still being developed for both approaches.

Who needs to be involved?

Both people and organizations are required – people for the work, organizations for the sustainable namespaces, web services etc.

The following is a list of the essential participants:

- Domain experts (Subject Specialists)
- Ontology Engineers (Philosophers)
- Ontology building experts (You and I?)
- Sustainable organizations for namespaces
- "Technique" developers
- Tool builders and providers
- Programmers – for application interfaces etc.

You will note that these are all highly specialized and hence "expensive" resources.

What do they cost and when can that cost be justified?

This is not an easy question to answer, there are so many variables. We can make a start by examining some of these.

The main cost factors include:

- Tools (ontology editors, merging tools, reasoners, etc)
    - Many of these are robust open source and free – however, there is a learning curve for each
    - They are still being developed – new releases are frequent – hence the learning curve is an ongoing investment.
- OWL and Description logics
    - These are now fairly stable so although there is a learning curve again – it is more of a one-off investment. Tutorials are available on the Web for both OWL and Description Logics. (Manchester University etc. – some courses becoming available)
- Professional Time – needed for 3 main areas:
    - for building – some tutorials on ontology building are available on the Web (Manchester University is a good starting point)
    - maintaining
    - applications interfacing
- Variables
    - Domains vary
        - Some concept rich, others properties rich, some almost flow charts – this can affect the build time.
    - Techniques vary
        - Merging existing sources, starting from scratch, hybrids (e.g. build a core from scratch and map existing vocabularies to core etc.)
    - Team Composition to build the ontology
        - This will also be domain dependent

          Minimum of 2 ontology builders and 1 domain expert – often more. Access to at least 1 ontology engineer.

A rough guide to estimating the costs for ontology *building* (not maintenance or application interfacing) follows:

- Times and costs
    - Just to extract concepts and the IS-A and Part-of relationships between them – unlikely to manage more than 200 concepts a day (probably less)
    - Agreeing other relationships and building restrictions for concepts will take longer – possibly 10 to 20 concepts a day
    - Might expect to get 3,000 concepts per person year
    - Approximately $20,000 per 1,000 concepts in Western Europe wage terms?
    - End point - $20 per concept.

The above is an example and although it is reasonably realistic it should NOT be taken as anything but a guide to the process! A comparison with the costs of building a thesaurus and the relative cost benefit analysis would be an interesting exercise – but one that has yet to be done.

Ontologies as Commons

Can we reduce these costs by ensuring that ontologies are made freely available as "commons"? There is a strong ethos in the ontology community for open access tools and for re-use of ontologies

- Reducing the expense of the build
  - Work is progressing on the editing tools etc such that ontology building by distributed teams might become more realistic soon [Framework and Rules needed]. This could reduce costs of getting experts together etc.
- Re-Use of Ontologies
  - Once the first version of a domain ontology is built it is generally made available freely for reuse by others
  - Larger ontologies are made available as "services" over the Web e.g. the GO ontology for genetics (it is also envisaged that the FAO Fisheries Ontology should be available in this way).
- Maintenance costs
  - Tools are being developed for web based comment and maintenance of built ontologies – [Framework and Rules needed]
- Gain from funded projects
  - The FAO Fisheries Ontology will gain from further development under a large EU project due to start in March 2006.

Since the FAO Fisheries Ontology Service (FOS) may be of particular interest to this audience, an initial outline is given here. Further details will be available from the FAO Web Site as the project progresses.

Although work is continuing on the merging of concepts in the FOS, this will be augmented when the European Union project commences – this is scheduled for March 2006. From March 2006 to March 2008 further resources will be available to continue improving the existing ontology from the project partners. During this period, work will also be undertaken on a tool for distributed ontology maintenance. The first version of this tool is scheduled for March 2008. This will be tested and improved and a second version is scheduled for March 2009. The final project goal for the FOS is the realization of an alert system for the over-fishing of fish stocks.

Has IAMSLIC a role as a mediator in ontology building and use?

If IAMSLIC members embrace ontologies, then some possible IAMSLIC roles could be:

- An IAMSLIC Ontology Group? To encourage interest in and uses of ontologies for the aquatic sciences.
- On the IAMSLIC Web Site:
  - Links to ontology training materials for ontology construction
  - Links to relevant ontologies available for the aquatic domain
  - Links to relevant project sites e.g. the EU site for the Fishery Ontology
- At a future IAMSLIC Conference – A workshop in ontology building.

Summary

Going back to the title of this paper:

- Do we need Ontologies in Fisheries and Aquatic Sciences?
  - I think we probably will – particularly as we start to exploit the content of digital resources (and data sources) – *BUT this is open for discussion*
- Can we afford them?
  - Maybe – particularly if we cooperate as "Commons" and develop the necessary frameworks and rules – this *is also open for discussion*
- Does IAMSLIC have a role?
  - Several if it wants them – *open for discussion*.

It is my belief that information professionals make good ontology builders. I encourage you to pick up the baton.

REFERENCES

Gangemi, A., Fisseha, F., Pettman, I., Pisanelli, D.M., Taconet, M., Keizer, J. 2002. A Formal Ontological Framework for Semantic Interoperability in the Fishery Domain. In: *Workshop on Ontologies and Semantic Interoperability* (ed. by J. Euzenat, et al), CEUR-WS, vol.64. [Online]. Available: http://sunsite.informatik.rwthaachen.de/Publications/CEUR-WS/Vol-64/

Gangemi, A. 2004. WonderWeb Deliverable D16: Reusing semi-structured terminologies for ontology building: A realistic case study in fishery information Systems. European Union Project Report - IST Project 2001-33052. *WonderWeb: Ontology Infrastructure for the Semantic Web*, pp.31. [Online]. Available: http://wonderweb.semanticweb.org/deliverables/D16.shtml