The ISIS Facilities Ontology and OntoMaintainer

Louisa Casely-Hayford, Shoaib Sufi

CCLRC e-Science Centre, CCLRC Daresbury Laboratory
Warrington WA4 4AD,UK

Abstract

This paper presents an ISIS facilities ontology based on keywords in the ISIS Metadata catalog (ICAT) and OntoMaintainer, a web application to facilitate the collaborative development of ontologies. The ISIS facilities ontology aims to organize and make keywords in the ISIS Metadata Catalogue more explicit which will improve the search and navigation of data by category. The Ontology Maintainer allows users to view current versions of the ontology, and send feedback on concepts modeled back to the maintainers of the ontology. To ensure ease of use, this service has been made available to the user community via the World Wide Web.

1. Introduction

ISIS, the worlds leading pulsed neutron and muon source, is one of the major large scale facilities operated by CCLRC. The ISIS Metadata Catalogue (ICAT) is a twenty year back catalogue of experiments conducted at ISIS. Currently the ISIS facility produces about 700GB of combined Neutron and Muon data each year, and the addition of a second target station will expand ISIS to twice its current size, hence the volume of data generated is set to rise [1]. Consequently, the efficient storage, retrieval and management of data is vital for the full value of these data resources to be realised [2]. To address this problem, e-science is developing numerous software solutions and ontologies are seen one of these useful approaches. Ontologies provide central controlled vocabularies that can be integrated into catalogues, databases, web publications and knowledge management applications. They are becoming increasingly popular amongst members of the scientific community, because they offer a powerful means to formally express the nature of a domain.

At present over 10,000 keywords describing experiments are housed in ICAT many of which are synonyms. These keywords are used to index experimental studies, however this is seen as a limited method as these free text keywords have no context, much implicit meaning and are hard to map by non-experts to terms used by facilities in the same domain and harder still to those outside. For example, a particular keyword 'HRPD' designates an instrument which is a 'powder diffractometer'; there are situations in which other collaborating Neutron facilities (e.g. SNS at ORNL in the US) understand the meaning of a powder diffractometer but do not understand that the cryptic 'HRPD' refers to such an instrument. The creation of ontologies aids in the mapping of concrete manifestations of familiar terms in one domain as well as related concepts in different domains. Thus the creation of such ontologies at CCLRC could aid in the cross facility search of related scientific data from the various science facilities housed at CCLRC e.g. ISIS, Central Laser Facility (CLF) and Diamond Light Source (DLS). An ISIS facility ontology was built using the web ontology language (OWL) within the java based editing environment Protégé, to address the need for the keywords in ICAT to be organized, classified and formally defined.

Ontologies are not static and continually evolve, therefore with time they will need to be updated and extended. A single ontology is usually built co-operatively by a group of people in different geographical locations. This is because the knowledge contained within an ontology represents a common view shared within a community. The construction and maintenance of an ontology is a difficult task, and dialogue is essential to reach a consensus and distribute information. This is a very important activity within communal design, however there are few tools that address this problem. For these reasons the Ontology Maintainer was developed, as it aims to facilitate the collective task of designing, building, extending and updating ontologies. This paper describes the creation of the ISIS Facilities ontology and Ontology Maintainer. In the following sections, several aspects of ontologies, ontology building, and the structure of the ISIS facilities ontology and the OntoMaintainer will be examined. At the end of the paper, some directions for future work in this area will be mentioned.

2. Background

2.1 What is an Ontology?

The word ontology stems from Philosophy, where it means a branch of metaphysics that investigates and explains the nature, essential properties and relations of all beings [2]. To date there are several existing definitions of the word ontology. One of the most quoted definitions in literature by the ontology community is by Tom Gruber. He defined an ontology as “an explicit specification of a conceptualization” [3]. In 1997 Borst slightly altered Gruber’s definition saying that; “Ontologies are defined as a formal, specification of a shared conceptualization” [4]. In Gruber’s definition 'Conceptualization' refers to an abstract model of some phenomena in the world by having identified the relevant concepts of those phenomena. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' indicates that the ontology should be machine readable, which excludes natural language. 'Shared' reflects the belief that an ontology captures consensual knowledge meaning that it is not private to some individual, but accepted by a group [5].

In all, ontologies are explicit specifications of the concepts in a given field, and of the relationships between those concepts [6]. The required minimum for an ontology is to include concepts, definitions of concepts, and defined relationships between the concepts. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, which promotes knowledge reuse and integration [3]. Ontologies are widely used for different purposes (natural language processing, e-commerce, knowledge management, intelligent information integration and the semantic web) and by different communities (knowledge engineering, bioinformatics, databases and software engineering) [7].

An example of an ontology widely used by biologists to search for associations of genes and gene products is GO. The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The backbone of GO is made up of three ontologies, that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner [8].

2.2 Building an Ontology

Ontology building is more of a craft than an engineering process. Several research groups propose various methods more commonly known as methodologies for building ontologies. There is no consensus between these groups and each employs its own methodology. During the process of building an ontology several questions arise. Some of these include:

- What domain will the ontology cover?
- What will the ontology be used for?
- What types of questions should the information in the ontology provide answers for (competency questions)?

The construction of ontologies is necessarily an iterative process, and answers to these questions may change during the life cycle of the ontology. The four basic aspects to consider when creating an ontology are content (the content of the ontology), the language in which it is implemented, the methodology which has been followed to develop it, and the software tools used to build and edit the ontology. Furthermore the type of language and development environment chosen is dependent on the type of ontology being built. It is important to know what the ontology is going to be used for, and how detailed the ontology should be, so that the possibility of over modelling (e.g. attempting to model the “whole world”) is lessened.

2.3 Ontology Languages

A formal language is used to encode the ontology and different languages provide different facilities [9]. The most recent development in standard ontology languages is the web ontology language (OWL) from the World Wide Web Consortium (W3C) [10]. OWL is a language based on description logics, and can be used explicitly to represent the meaning of terms in vocabularies, and the relationships between those terms [10]. The Web Ontology language (OWL), is a revision of DAML+OIL web language and has more facilities for expressing meaning and semantics than XML, RDF and RDF-S.

OWL ontologies consist of Individuals, Properties and Classes or Concepts (terms relevant to the domain). Classes (concepts) are a concrete representation of concepts, and in OWL classes are interpreted as sets that contain individuals [9]. A concept can be anything about which something can be said, for example human, John or pet. Classes within the ontology

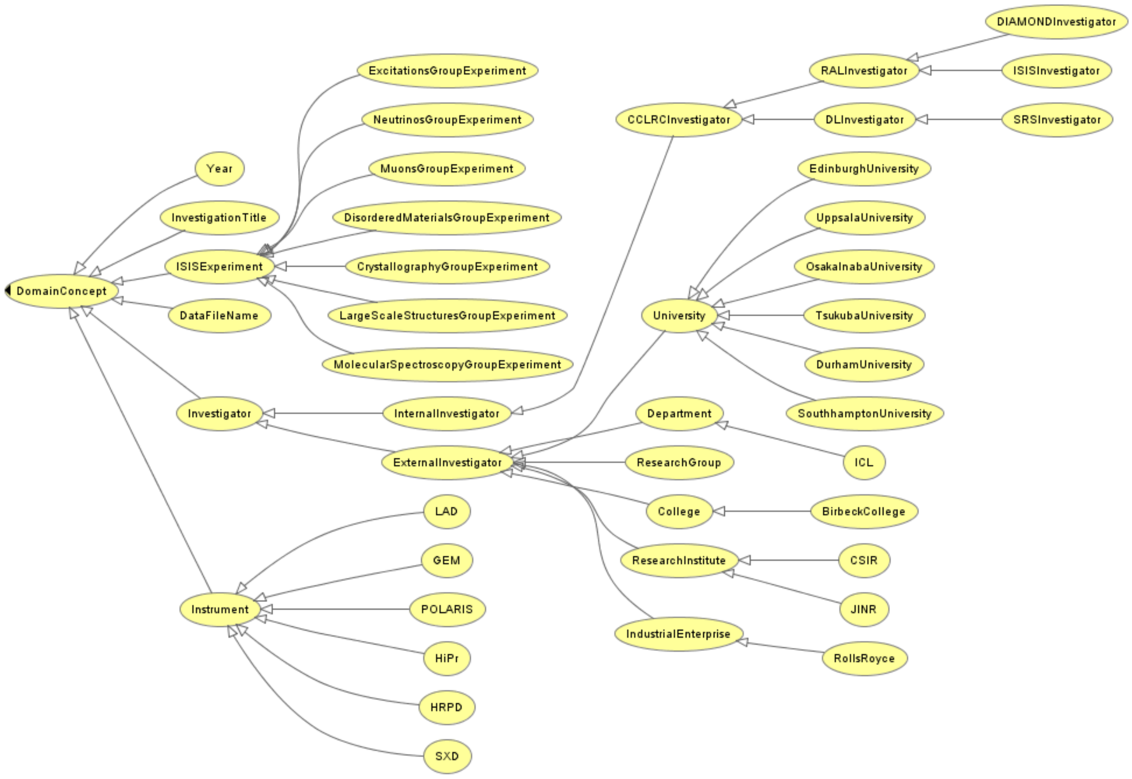


Figure 1 Diagram of Overall Hierarchy of ISIS Facilities Ontology

are normally organised in taxonomies with a subsumption ("subclass") hierarchy. OWL classes can be specified as logical combinations (intersections, unions, or complements) of other classes, or as enumerations of specified objects. OWL can also declare properties, organize these properties into a "sub-property" hierarchy, and provide domains and ranges for these properties [10]. Properties allow general facts about the members of classes and specific facts about individuals to be asserted and also represent relationships between two individuals. For example the property hasSibling might link the individual Matthew to the individual Gemma. There are two main types of properties, object properties and data type properties. Data type properties are relations between instances of classes and RDF literals and XML Schema data types. Object properties are relations between instances of two classes. Properties may have a domain and a range specified. The domains of OWL properties are OWL classes, and ranges can be either classes or externally-defined datatypes such as string or integer. Properties link individuals from the domain to individuals from the range. Individuals, represent objects in our domain of interest [10]. For example Italy, England China and France are individuals in the class Country.

OWL provides three sublanguages, namely OWL Lite, OWL DL and OWL Full [10]. OWL Lite supports those users primarily needing a classification hierarchy and simple constraints. OWL DL supports those users who want the maximum expressiveness while retaining computational completeness for reasoning systems. OWL Full is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees [10]. OWL's logical model allows the use of a reasoner which can check whether or not all of the statements and definitions in the ontology are mutually consistent, and can also recognise which concepts fit under which definitions. The reasoner can therefore help to maintain the hierarchy correctly. This is particularly useful when dealing with cases where classes can have more than one parent [11].

2.4 Ontology Editing Environments

Ontology development or engineering tools include suites and environments that can be used to build a new ontology from scratch or by reusing existing ontologies [12]. Since the mid-nineties there has been an exponential increase in the development of technological platforms related with ontologies [7]. Protégé is one of the most widely used ontology building tools and has been developed by the Stanford Medical Informatics (SMI) Group at Stanford University [13]. Protégé instances, slots and classes roughly corresponds to OWL Individuals, Properties and Classes (concepts).

3. ISIS Facilities Ontology

As mentioned earlier, an ISIS facilities Ontology was built to compile a comprehensive structured vocabulary of terms representing the keywords describing experiments in ICAT. This ontology will be used as a means of grouping data across studies, disambiguating keywords and improving the search and navigation of data by category. The Protégé OWL plug-in was selected to create the ISIS facilities ontology. The web ontology language (OWL) was chosen because of its ability to provide formal semantics, built-in reasoning support and additional features such as metadata annotation and ontology versioning. OWL DL was used as it supports maximum expressiveness while allowing the use of a reasoner [1]. After close examination of the keywords and their definitions, it was decided that they would be grouped into five categories namely: datafile, instrument, investigation_title, investigator and year. This is because all of the keywords in ICAT fell into these five groups.

Examples of keywords in these five categories are:

1. HRP00145.RAW which is a datafile name.
2. HRPD which is a High Resolution Powder Diffractometer one of the many instruments used in experiments at the ISIS facility
3. Hydrazinium which is an investigation title, chemical names and compounds were used as investigation titles of experiments in ICAT
4. 1986 which is the year in which a particular experiment was conducted
5. JINR (Joint Institute for Nuclear Research in Russia) which is the name of an investigator.

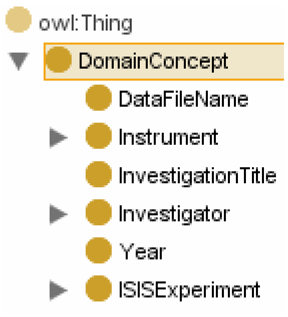


Figure 2 Snapshot of general ontology hierarchy in Protégé

The field of ontological engineering is still young and is not as developed as that of knowledge engineering. Although there is some cooperative experience of building and using ontologies, up to now there are no standardised methodologies for building ontologies [14]. One of the most recognized ontology construction guidelines, developed by Thomas Gruber, was followed to build the ISIS facility ontology [15]. A top-down modelling approach was used to build the ontology with an overview of the system being formulated first, afterwhich each part of the system was then refined by designing it in more detail. The overall hierarchy of the ontology can be viewed in Figure 1. The classes DataFileName, Instrument, InvestigationTitle, Investigator and Year formed the core of the ontology (Figure 2). A class ISISExperiment was added to represent Experiments carried out at ISIS by different scientific groups (Figure 4). More subclasses were added to Instrument, ISISExperiment and Investigator to represent the different kinds of investigator, instrument and ISIS experiments (Figure 1).

All ISIS Instruments (e.g. HRPD, POLARIS, LOQ, IRIS Figure 1) were added as subclasses of the class Instrument to define the different kinds of instruments used in experiments at ISIS (Figure 3). A subsumption relationship exists between classes and subclasses in OWL, therefore for example the intstrument HRPD (the High Resolution Powder Diffractometer) is a subclass of the superclass Instrument. This subsumption relationship indicates that HRPD is a kind of Instrument and an Instrument is HRPD. Therefore if M is an instance of class HRPD, and HRPD is a subclass of Instrument, then we can infer that M is an instance of Instrument (Figure 3).

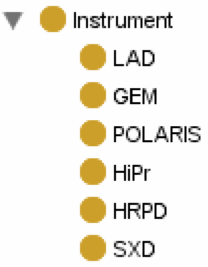
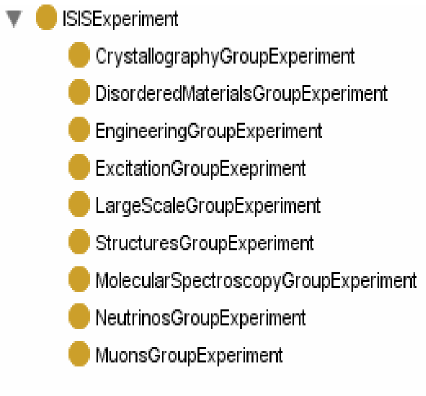


Figure 3 Snapshot of subclass Hierarchy of Instrument Class

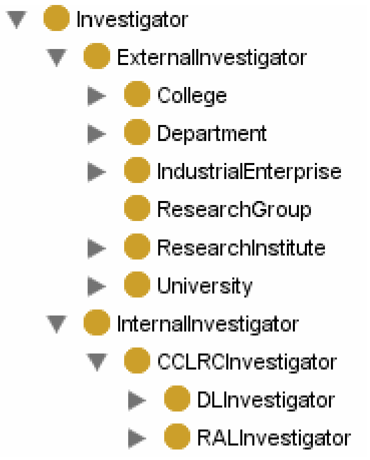
Subclasses representing experiments carried out by different scientific groups (e.g. CrystallographyGroupExperiment), at ISIS were added to the class ISISExperiment (Figure 4). This is to facilitate common instrument requirements related to different research groups and to provide a focus for user interaction.



ISISExperiment
CrystallographyGroupExperiment
DisorderedMaterialsGroupExperiment
EngineeringGroupExperiment
ExcitationGroupExepriment
LargeScaleGroupExperiment
StructuresGroupExperiment
MolecularSpectroscopyGroupExperiment
NeutrinosGroupExperiment
MuonsGroupExperiment

Figure 4 Snapshot of subclass Hierarchy of ISIS Experiment Class

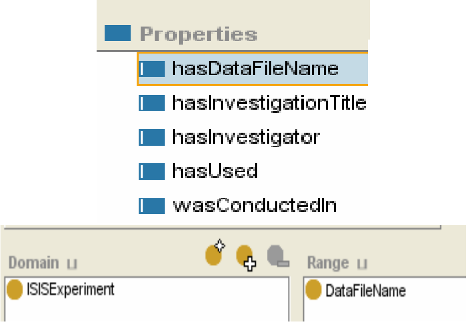
The Class Investigator was subclassified into internal and external investigator. Next external investigators were subclassed into College, Department, ReaserchGroup, ReasearchInstitute and University (Figure 5). Furthermore internal investigators were subclasified into DL (Daresbury Laboratory) and RAL (Rutherford Apppleton) investigators (Figure 5).



Investigator
ExternalInvestigator
College
Department
IndustrialEnterprise
ResearchGroup
ResearchInstitute
University
InternalInvestigator
CCLRCInvestigator
DLInvestigator
RALInvestigator

Figure 5 Snapshot of subclass Hierarchy of Investigator Class

Properties were created to provide a link or relationship between individuals in the class ISISExperiment and individuals in the other five main classes. The properties hasDataFileName, hasInvestigationTitle, hasInvestigator, hasUsed and wasConductedIn were created and domain and range constraints set in Protégé. All five properties have ISISExperiment as their domain and DataFileName, InvestigationTitle, Investigator, Instrument, and Year as their range respectively (Figure 6).



Properties
hasDataFileName
hasInvestigationTitle
hasInvestigator
hasUsed
wasConductedIn
Domain
ISISExperiment
Range
DataFileName

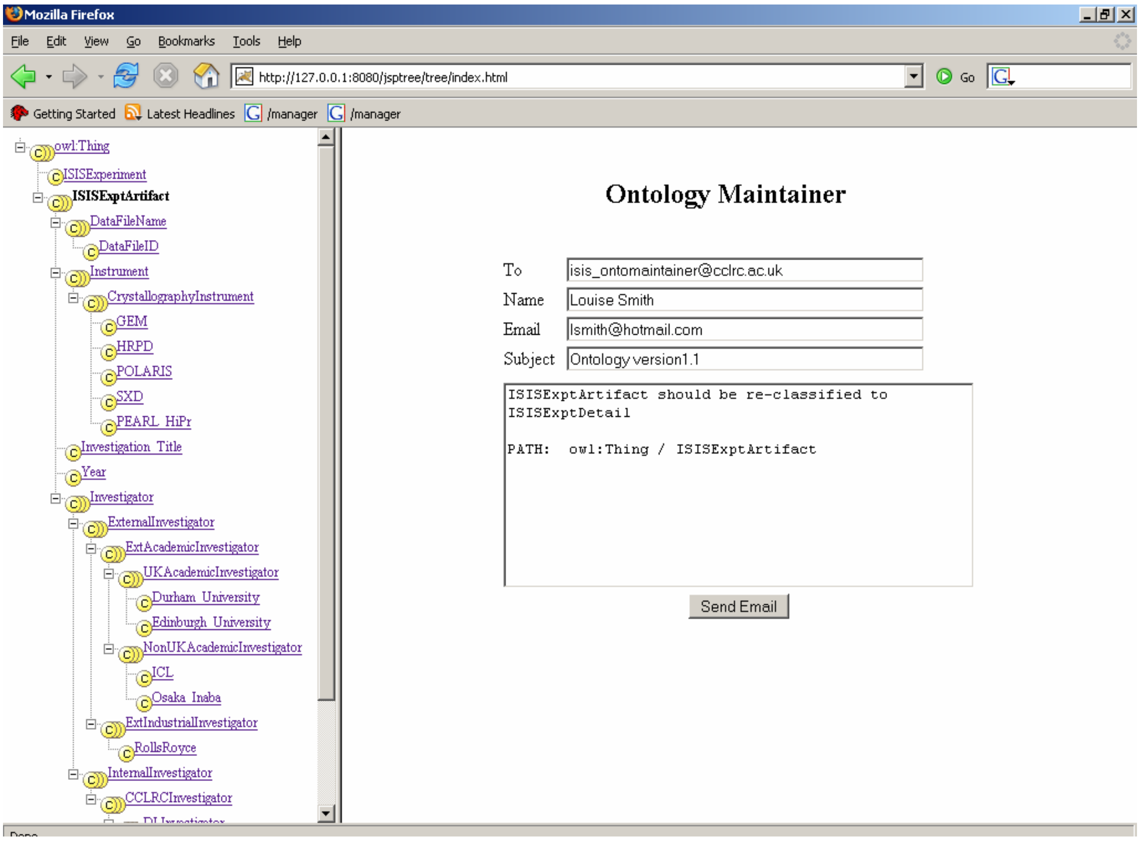
Figure 6 Snapshot of Properties in Properties Tab

4. Ontology Maintainer

Consensus on Concepts modelled in the ISIS Facilities ontology, was achieved through a series of interviews with domain experts. During the design and creation process, there was a difficulty in sharing current versions of the ontology with our collaborators at ISIS.

This is because to view the hierarchical structure of the ontology, scientists would have to download and install Protégé locally which is a time consuming and complicated process. Although there are a few web-based ontology library systems and editing environments for communal building, many of these tools are: complicated to use, not easily accessible, do not provide a user-friendly interface, and do not provide a graphical view of the hierarchy of the ontology. For these reasons the Ontology Maintainer was developed to aid the community to remotely view current versions of the ontology. It is purely a visual tool aimed at the social engineering (collective effort) aspect of ontology design and is not for the editing or building of ontologies.

The Ontology Maintainer is a web application with a client server-based architecture that was developed using Java Server Pages, servlets, CSS (Cascading Style Sheets) and Tomcat. The



Ontology Maintainer

To	isis_ontomaintainer@cclrc.ac.uk
Name	Louise Smith
Email	lsmith@hotmail.com
Subject	Ontology version1.1

```
ISISExptArtifact should be re-classified to
ISISExptDetail

PATH:  owl:Thing / ISISExptArtifact
```

Figure 7 Screenshot of the OntoMaintainer

major advantage of this tool is that it provides an easily accessible, user friendly method of displaying the hierarchy of the ontology being built. In addition to visualisation, it enables users to submit feedback through a simple web form to maintainers of the ontology.

The web-page is divided into two panels one displaying the tree hierarchy and the other showing a web form. The first panel shows a tree hierarchy with class names displayed alongside each of the tree nodes, users can expand the tree fully by clicking on nodes or classes in the tree. The web application is modular and the look and feel of the tree can be changed easily. To send comments users can click on the class they wish to comment on in the first panel and this will generate a path of that particular class through the ontology, which is added to the text area of the comment form. Next users can enter their full name, email address, subject and comments in the second panel, as shown in Figure 7. Parameters entered in the form will be sent directly to the maintainer of the ontology.

Ontologies are constantly being changed due to technological advances, errors in earlier versions of the ontology, or the release of a novel method of modelling a domain. The process of updating an ontology is not an easy task, as there needs to be a general consensus amongst members of the community. The Ontology Maintainer will provide a huge impact on this process, as it will allow individuals to easily access current versions of the ontology through the World Wide Web.

5. Future work

The current ISIS facilities ontology serves as a base for future ontologies, and in the near future it will be expanded with more concepts, properties and restrictions. An ISIS Online Proposal System has been developed to automate the capture of metadata [1]. This system enables users of the ISIS facility to electronically create, submit and collaborate on applications for beam-time [1]. This new electronic proposal system provides access to a rich source of metadata which can be fed directly into the experimental set. Creation of the ISIS facilities ontology has generated interest by ISIS in the use of ontologies to mark up proposals submitted through the online proposal system. Further ontologies for sample, experiment and investigator are being developed to mark up submitted proposals. These separate modular ontologies will be developed and fed back into the Metadata catalog. Once feedback is submitted and a consensus is achieved, a person (the designated Facility ontology maintainer) will update the ontology, which will automatically serve these new terms to reflect a shared understanding. Ontologies will facilitate searching of data by category and grouping of data into keywords across studies.

The ultimate aim of these ontologies is to support or promote cross-facility searching. We can envisage an ontology at each of CCLRC's large scale facilities ISIS, CLF (Central Laser Facility) and DLS (Diamond Light Source), with mappings between them such that users can search data from their perspective and receive relevant hits from other domains. The User Interface of the Ontology Maintainer will be improved through the addition of properties. Properties will be displayed in the user interface to enable relationships between individuals in classes to be shown. Additionally a graphical view of the ontology will be generated with the click of a button. Also feedback entered by users will be stored in a file system or database. Finally the tree hierarchy in the ontology maintainer will be made more dynamic through automatic updating of classes for example based on records in a database derived from the master ontology designed in an ontology engineering tool e.g. Protégé.

6. Conclusion

The ISIS facilities Ontology serves as a stepping stone to the creation of further ontologies, which will be used to help maximise the value of data collected at ISIS. Ontologies will function to improve the access, navigation and reuse of these large scale data resources. Throughout this process, collaboration between scientists at ISIS and ontology builders will be made easier through the use of the Ontology Maintainer.

7. References

[1] Damian Flannery
ISIS Data, Metadata, Proposals and the CCLRC Data Portal
http://lns00.psi.ch/nobugs2004/papers/paper00147.pdf

[2] Diana Marcela Sánchez, José María Cavero, Esperanza Marcos
On models and ontologies
http://kybele.escet.urjc.es/PHISE05/papers/sesionIV/SanchezCaveroMarcos.pdf

[3] T.R. Gruber
A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993
http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html

[4] A.J. Duineveld, R.Stoter., M.R. Weiden, B. Kenepa, and V.R. Benjamins. (1999) Wondertools? A comparative study of ontological engineering tools.
http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Duineveld1/wondertools.pdf

[5] R. Studer, V.R.Benjamins., D. Fensel (1998) Knowledge engineering: principles and methods. Data and Knowledge Engineering 25, 161–197.

[6] M. Fernaandez-Loopez, A. Goomez-Peerez, A. Pazos-Sierra, J. Pazos-Sierra, Building a chemical ontology using METHONTOLOGY and the ontology design environment, IEEE Intelligent Systems & their applications 4 (1) (1999) 37–46
http://www.aegean.gr/culturaltec/Kavakli/MIS/papers/Corcho_2003.pdf

[7] Oscar Corcho, Mariano Fernandez-Lopex, Asuncion Gomez-Perez. Data &Knowledge Engineering 46(2003) 41-64. Methodologies, tools and languages for building ontologies. Where is their meeting point?
http://www.aegean.gr/culturaltec/Kavakli/MIS/papers/Corcho_2003.pdf

[8]M Ashburner et al, Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25 - 29 (2000
http://www.nature.com/ng/journal/v25/n1/pdf/ng0500_25.pdf

[9] M. Horridge, H. Knublauch, A. Rector, R. Stevens, C. Wroe (2004) A Practical Guide To Building OWL Ontologies Using The. Protege-OWL Plugin and CO-ODE Tools. Edition 1.0.
http://www.coode.org/resources/tutorials/ProtegeOWLTutorial.pdf

[10] Mike Dean, Dan Connolly, Frank van Harmelen, James Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. OWL web ontology language reference. W3C Working Draft, 31 March 2003. http://www.w3.org/TR/2003/WD-owl-ref-20030331

[11] R.Stevens, C.Goble, S. Bechhofer (2001) Ontology-based Knowledge Representation for Bioinformatics. University of Manchester http://www.cs.man.ac.uk/~stevensr/papers/briefings-ontology.pdf

[12]Lambrix P, Habbouche M, Perez M. Evaluation of ontology development tools for bioinformatics.Department of Computer and Information Science, Linkopings universitet, 581 83 Linkoping, Sweden. patla@ida.liu.se http://bioinformatics.oxfordjournals.org/cgi/reprint/19/12/1564

[13] N.F. Noy, R.W.Fergerson, M.A. Musen (2000) The knowledge model of protege-2000: combining interoperability and flexibility. In: 12th International Conference in Knowledge Engineering and Knowledge Management (EKAWÕ00), Lecture Notes in Artificial Intelligence (Springer, ed.), pp. 72, Berlin.

[14] Oscar Corcho , Mariano Fernandez-Lopez , Asuncion Gomez-Perez. Methodologies, tools and languages for building ontologies. Where is their meeting point? Facultad de Informatica, Universidad Politecnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte, Madrid 28660,Spain (2001) http://www.aegean.gr/culturaltec/Kavakli/MIS/papers/Corcho_2003.pdf

[15] Thomas R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing.Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University. http://www.cise.ufl.edu/~jhammer/classes/6930/XML-FA02/papers/gruber93ontology.pdf