Collaborative Editing and Linking of Astronomy Vocabularies Using Semantic Mediawiki

Stuart Chalmers[1], Norman Gray[2], Iadh Ounis[1], and Alasdair Gray[3]

[1] Computing Science, University of Glasgow, Glasgow, UK
[2] Physics and Astronomy, University of Glasgow, Glasgow, UK
[3] School of Computer Science, Manchester University, Manchester, UK

Abstract. The International Virtual Observatory Alliance (IVOA) comprises 17 Virtual Observatory (VO) projects and facilitates the creation, coordination and collaboration of standards promoting the use and reuse of astronomical data archives. The Semantics working group in the IVOA has repurposed five existing vocabularies (modelled using SKOS), capturing concepts within specific areas of astronomy expertise and applications. A major task however, is to promote the uptake of these semantic representations within the Astronomy community, and further, to let astronomers model (and in turn create links from) their own custom vocabularies to use these existing definitions. In this paper we show how Semantic Mediawiki (SMW) can be used to support expert interaction in the lifecycle of vocabulary creation, linking, and maintenance.

1 Introduction

Astronomy as a discipline incorporates a broad range of topics and data analysis across the wavelength spectrum, from gamma-rays to radio waves, and a wide range of expertise from professional researchers to amateurs. Because of the collaborative nature of astronomy working groups and projects, and a culture where sharing data is the norm, there is a well-established need for consensus definitions describing data (mostly image and object catalogue data). To this end a number of standardised vocabularies have emerged, which are mostly, at present, focused on the search for and retrieval of resources, primarily data and journal articles.

Thus, multiple independent controlled vocabularies have evolved to meet the various terminological needs of these different sub-communities (Table 1). The most widely-known of these is the keyword list maintained jointly by the three main astronomy journals A&A, ApJ and MNRAS (these keywords are used to tag journal articles, so that most astronomers have a familiarity with this set), and the largest is a thesaurus developed by the International Astronomical Union (IAU) (with the IVOA starting work on an update, the IVOAT). Newer than both are the AVM vocabulary – a recent effort intended for use when tagging astronomy outreach images – and the UCD list, in increasingly wide use as a set of standardised database column headings[4]. For further discussion see [1].

[4] `http://www.ivoa.net/Documents/latest/Vocabularies.html`

Vocabulary	Original Publisher	Purpose	Number of Concepts
Journal Keywords	Journal publishers	Tagging articles to aid retrieval	311
Astronomy Visualization Metadata (AVM)	various	Tagging images for dissemination	208
The IAU Thesaurus (IAUT)	IAU	Library cataloguing	2551
The IVOA Thesaurus (IVOAT)	IVOA	Update of the IAU Thesaurus	2890
Universal Content Descriptors (UCD)	IVOA	Labelling data repository column headings	473

Table 1. Astronomy vocabularies

While the IVOA vocabularies have provided a basis for standardisation of experimental terminology, there remain a few problems:

- There are no standardised tools or methodology for creating custom experimental descriptions based on these vocabularies.
- Users may be familiar with specific IVOA vocabularies relating to their subdiscipline, but not others, meaning that their description cannot describe their data as fully as a searching colleague might require.
- Searching of user-defined vocabularies and data is limited to terminology in the IVOA vocabular(ies) used to define them. For instance, a user vocabulary described using the IVOAT thesaurus has no relation to searches using keywords from the IAUT thesaurus.

Recent work in the Explicator project[5] has laid the foundations for a solution to these problems, by representing the main IVOA vocabularies in SKOS, and exploiting SKOS relationships to help domain experts articulate cross-vocabulary links [2].

2 Current Vocabulary Building Tools

The Explicator project has developed a number of tools for the creation and use of SKOS astronomy vocabularies. The main entry point for searching and exploring terminology is the Web Vocabulary Explorer[6], built upon the Terrier Information Retrieval Platform [3] and providing an AJAX frontend for searching and browsing the astronomy vocabularies by entering a simple search string to find matching concepts. Fig. 1 (left) shows the search results for "star". The use of Terrier is important, in order to provide useful ranking of results: this vocabulary contains a large number of labels with common strings, so a naive search for "star" produces more than 600 concepts which have that string somewhere in their label, with the key concept 'Star' appearing uselessly far down the list. Using Terrier's ranking support, however, the appropriate concepts from

[5] `http://explicator.dcs.gla.ac.uk`
[6] `http://explicator.dcs.gla.ac.uk/WebVocabularyExplorer`

the three searched vocabularies appear at the beginning of this list. The explorer allows users to expand results and view details of concepts, such as alternate labels, available definitions and semantic relationships. Related concepts, both within a vocabulary and across vocabularies, can be explored by following links to broader, narrower, related, and equivalent concepts. Searches can be configured by selecting sets of vocabularies and mappings. This service is also available via XML-RPC, so that it can be embedded within other applications.

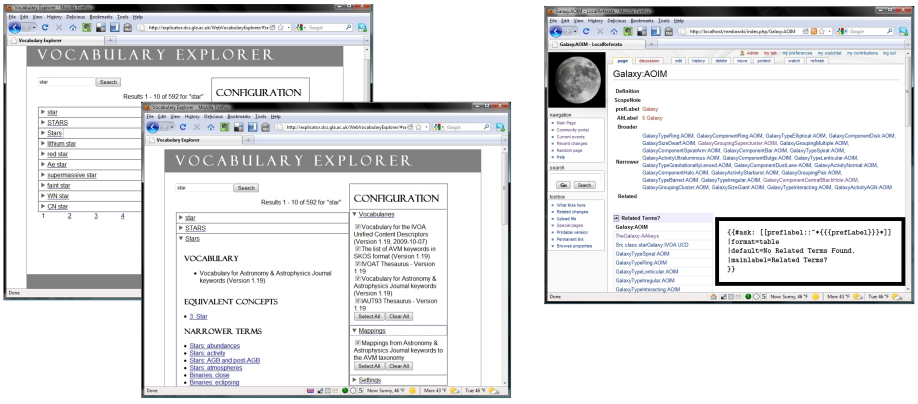

Fig. 1. The Web Vocabulary Explorer interface (left), and the inline search query and its use in the AOIM Galaxy definition (right)

To create links between the main vocabularies in Table 1 we have a Java mapping application providing a GUI interface to declare mappings between vocabularies that can then be integrated into the Web Vocabulary Explorer. The five vocabularies listed here were pre-existing ones, though not published as SKOS, and so were converted from their original formats as part of the process of developing [4]. The tool also allows the inclusion of automatically created RDF representations of databases, created using the D2RQ database to RDF mapping tool[7]. The other important source of ontology information within the VO is the IVOA's resource registry[8], which curates resource metadata using a standardised set of XML Schemas, which we have also converted to RDF Schemas using XSLT transformations.

Part of the point of the tool's search functionality is to help users find relevant concepts in multiple vocabularies, and to support them in articulating inter-vocabulary mappings. However we do not aim to do any automatic vocabulary alignment.

3 Semantic Mediawiki in the Vocabulary Lifecycle

While the astronomy community is in general technically adept, the immediate payoff from adopting the tools described in section 2 and converting to SKOS

[7] `http://www4.wiwiss.fu-berlin.de/bizer/d2rq/`
[8] `http://rofr.ivoa.net`

representations is not obvious (or apparent) enough to users to make this an attractive option (this is a general problem, also discussed in [5]). What is needed is a cohesive, familiar and easily understandable interface that integrates these tools in a way that allows the creation of SKOS-based experiment descriptions and vocabularies (based on and utilising current IVOA standard vocabularies) with minimal expenditure on learning the underlying semantic representations. To this end we have proposed a coherent vocabulary 'lifecycle' methodology (creation, collaborative editing, linking and searching/use) – see Fig 2. This uses SMW as a collaborative vocabulary building tool to create and edit vocabularies (1), *link* (1) these to existing IVOA vocabularies (2) and have them automatically *exported* to (3) and *imported* from (5) their corresponding SKOS representations (4) for use in the Web Vocabulary Explorer.

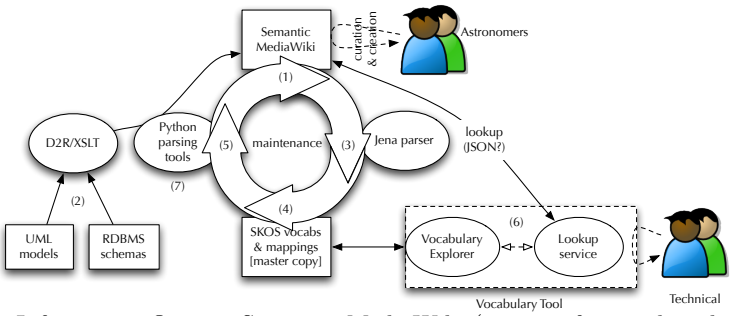

Fig. 2. Information flows in Semantic MediaWiki (see text for numbered notes)

To link SMW to our existing tools (6), we have developed a general set of python scripts (7), using pywikipediabot[9] and the rdflib[10] library to automate the uploading and parsing of our SKOS vocabularies into Wikipedia pages[11]. The SMW pages are based on a simple semantic form/template structure, parsed from the main SKOS vocabularies (4) and uploaded using the python bots. Similarly we use a Jena-based parser to parse the SMW OWL/RDF export (3) for a particular vocabulary and create the corresponding SKOS version for reinclusion in the Web Vocabulary Explorer search.

This linking of the five main IVOA vocabularies into SMW pages means that we now have a base set of terms for users to begin using in their own experimental vocabularies. To help users find related terminology (e.g. for broader, narrower, or related matches in their SKOS terms) we use simple inline queries embedded in the main vocabulary term template to show (on each term's page) the possible related terminology. Fig. 1 (right) shows the inline query used in the main template of the vocabulary wiki pages and an example, the AOIM term 'Galaxy'. This shows the main definition (scopenote, prefLabel, altlabel, broader, narrower and related) and a table of the possible related terms (including TheGalaxy in the AAKeys vocabulary and the src.class.starGalaxy from the

[9] `http://meta.wikimedia.org/wiki/Pywikipedia`
[10] `http://www.rdflib.net/`
[11] We currently host this testbed at `http://vocabularies.referata.com`

UCD vocabulary) that may be linked to by the user as cross-vocabulary related terms.

4 Related and future work

There are other vocabulary development systems in existence, including the NeOn project's ontology editor[12], and its Cicero project, which is also based on SMW, and which supports an elaborate argumentation structure for collaborative ontology development (NeOn deliverable 2.3.1). On a similar theme is LexWiki[13], which is a platform for developing a biomedical vocabulary. The problem we are addressing, however, is *not* that of collaboratively creating a large ontology from scratch, but supporting the collaborative inter-relation of multiple existing vocabularies from various sources, with a community which is made more rather than less comfortable by having some of the underlying technology visible, and repurposable from user-written applications.

At present we are working on a mediawiki extension that will allow us to use the XML-RPC search from the Web Vocabulary Explorer to find related terms. This will use the Terrier search described above, to provide more accurate ranked searches for related terms, than is possible with the existing inline searches.

A key advantage, for us, of using a wiki-based solution is that it provides a good match to the expectations of the domain experts – they feel comfortable and in control when using it. Both the wiki and its embedded functionality must therefore evolve in tune with the user base, and an important strand of our future work on this project is to evaluate the provided functionality in use.

References

1. Gray, A., Gray, N., Ounis, I.: Vocabularies in the VO. In Bohlender, D., et al., eds.: Proc. Astronomical Data Analysis and Software Systems Conference (ADASS XVIII). Volume 411., Astronomical Society of the Pacific (2009) 179–182
2. Gray, A.J.G., Gray, N., Hall, C.W., Ounis, I.: Finding the right term: Retrieving and exploring semantic concepts in astronomical vocabularies. Information Processing and Management (2009). In press.
3. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, C.: Terrier: A high performance and scalable information retrieval platform. In: Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006), Seattle (Washington, USA), ACM (2006)
4. Gray, A.J.G., Gray, N., Hessman, F.V., Preite Martinez, A.: Vocabularies in the virtual observatory. IVOA Recommendation (2009) Available at: `http://www.ivoa.net/Documents/latest/Vocabularies.html`.
5. Gray, N., Linde, T., Andrews, K.: SKUA - retrofitting semantics. In Auer, S., et al., eds.: Proc. 5th Workshop on Scripting and Development for the Semantic Web at ESWC 2009, Heraklion, Greece. Volume 449 of CEUR Workshop Proceedings ISSN 1613-0073. (2009)

[12] `http://www.neon-project.org/`

[13] `http://informatics.mayo.edu/vkcdemo/lexwiki1/`