KNOWLEDGE MEDIA

KMi

INSTITUTE

Ontology Users' Survey – Summary of Results

Technical Report kmi-13-1

June 2013

Paul Warren

paul.warren@cantab.net

The Open University

1 Introduction

This report describes initial findings from the Ontology Users' Survey conducted in early 2013. The next six sections follow the structure of the survey, providing information about the respondents; their ontologies; their ontology tools; the ontology languages and language features they use; the use of ontology patterns; and their final general comments.

These responses were obtained by using a number of contacts and relevant mailing lists. The latter included:

- ontolog-forum (ontolog-forum@ontolog.cim3.net)
- UK Ontology Network (ontology-uk@googlegroups.com)
- two LinkedIn groups: Semantic Web for Life Sciences; Description Logic
- lists maintained by the Open Knowledge Foundation: okfn-en@lists.okfn.org; ok-scotland@lists.okfn.org; okfn-nl@lists.okfn.org
- the internal mailing list within the Knowledge Media Institute at the Open University

In all, there were 118 respondents. A number of respondents only answered a subset of the questions; most questions resulted in several tens of responses. No attempt was made to achieve a representative sampling of all ontology users, and this should be borne in mind when interpreting the results.

2 The respondents

2.1 Sectors

Respondents were asked to categorize themselves into three sectors plus 'other'. 116 people responded to this question, i.e. all but 2 of the total respondents. The percentage breakdown was as follows[1]:

- academic 45%
- from research institutes 25%
- industrial 17%
- other 13%

'Other' included consultancy and freelance; publishing; and government. The relatively large number of respondents classifying themselves as 'other' may be in part a result of the use of the term 'industrial'; 'industrial and commercial' might have been a better term to capture those who work in for-profit organisations. Additionally, it might have been useful to include the category 'government and international organisations'.

[1] These and similar subsequent statistics are rounded to the nearest whole percentage. In some cases, because of rounding errors, the percentages may not add up to 100%.

2.2 Application domains

Respondents were also asked to specify their primary application area. All 118 respondents answered this question, giving the following breakdown:

- biomedical 31%
- business 9%
- engineering 19%
- physical sciences 7%
- social sciences 5%
- other 30%

A number of the respondents in the 'other' category were involved in computer science and information technology; indicating that this would have been a useful category to include in the questionnaire. A smaller number cited 'cultural heritage' and 'linguistics'; indicating that, e.g., 'humanities', might also have been a useful category.

2.3 Length of time working with ontologies

Respondents were also asked how long they had been working with ontologies. The results are shown in figure 2.1.

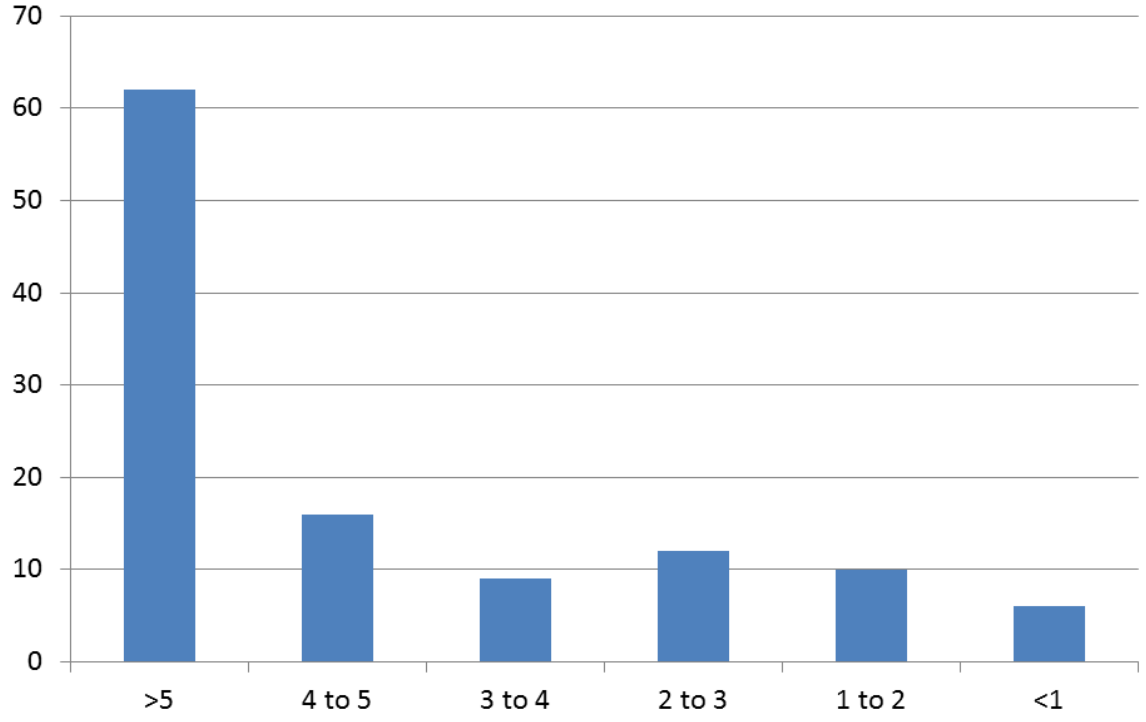


Figure 2.1 Number of years working with ontologies
– showing the number of respondents in each of six categories.

115 people answered this question, i.e. all except three of the total set of respondents. The majority of them had over five years of experience with ontologies. Comparison of this data with the answers to the questions on sector and application area showed no obvious relationships.

3 Ontologies

3.1 Commonly used ontologies

Respondents were asked about the ontologies they used. They were invited to specify their up to five most frequently used ontologies, indicating the order of frequency of use. Of the 69 respondents who answered this question, 32 listed five ontologies and only seven listed just one. Figure 3.1 shows the most popular five ontologies. Ten people indicated that they used their own ontologies, putting 'own' in sixth place, as shown. Note that, as up to five ontologies could be indicated, the total number of ontologies exceeds the number of respondents to the question. The figure also shows how the ontologies were broken down by frequency of use. Thus, for example, the majority of users of FOAF indicated that it was their second most frequently used ontology.

Besides the ontologies shown in the figure there were a number relating to biology, medicine and chemistry; and also some generic ontologies. Amongst the latter were: the W3C provenance ontology; the RDF data cube vocabulary; schema.org; upper level ontologies, e.g., DOLCE; and lexical databases, e.g., WordNet.

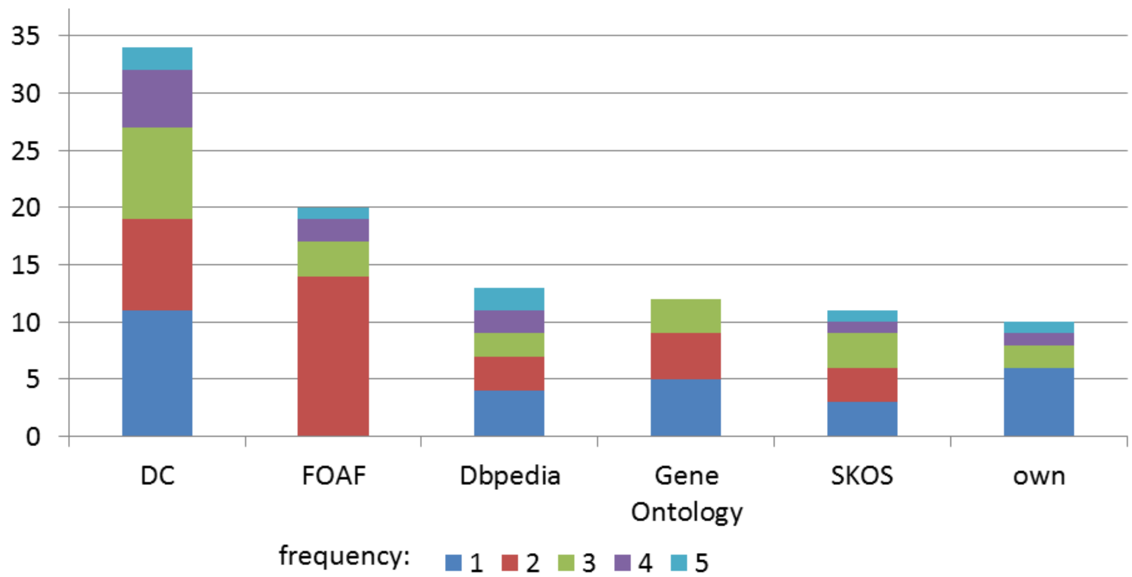


Figure 3.1 Most commonly used ontologies, showing their frequency ranking

3.2 Editing ontologies and knowledgebases

To understand better how respondents were using ontologies they were asked, for each of the up to five ontologies they had listed, to specify whether they:

- edited the ontology and knowledgebase;
- used the ontology (without editing it) and edited the knowledgebase;
- used ontology and knowledgebase (without editing either).

The results from this question are shown in figure 3.2., where the percentage breakdown for each frequency category is shown in columns 2 to 4. The number of people who responded to this question for a particular frequency category is shown in column 5. Thus, for example, the 42% in the column headed 'edit ontology and knowledgebase' and the row headed '1' indicates that, of the 69 people who provided information relating to their most frequently used ontology, 42% edited that ontology.

frequency ranking	edit ontology and knowledgebases	use ontology and edit knowledgebases	use ontology and knowledgebases	number of respondents
1	42%	31%	26%	69
2	10%	55%	33%	59
3	21%	46%	32%	56
4	18%	46%	34%	43
5	18%	40%	40%	32

Figure 3.2 Usage of ontologies related to their frequency ranking

It was thought that the propensity to edit ontologies might be linked to the number of years working with ontologies. To test this, the number of people who edited an ontology was compared with the number who edited only knowledgebases or edited neither ontologies nor knowledgebases. This was done for those with up to five years' experience and those with more than five years. Whilst a higher percentage of the more experienced people did edit ontologies, this result was not statistically significant.

3.3 Size of ontologies

Respondents were asked about the size of their ontologies. They were specifically asked about the size of that part of the ontology with which they worked, i.e., not necessarily the total ontology if that was not appropriate. There were five questions relating to size:

- number of classes
- number of individuals
- number of properties
- number of top-level classes
- depth of the hierarchy

In each case respondents were asked to select from a number of ranges. 40 respondents completed this question. They were invited to provide the information for the up to 5 ontologies previously specified. As a result there were 125 total responses; these are the basis for figures 3.3 to 3.7. These figures show the responses to questions relating to number of classes, number of instances, number of properties, number of top-level classes and depth of hierarchy. They illustrate the wide range of sizes of ontologies being used.

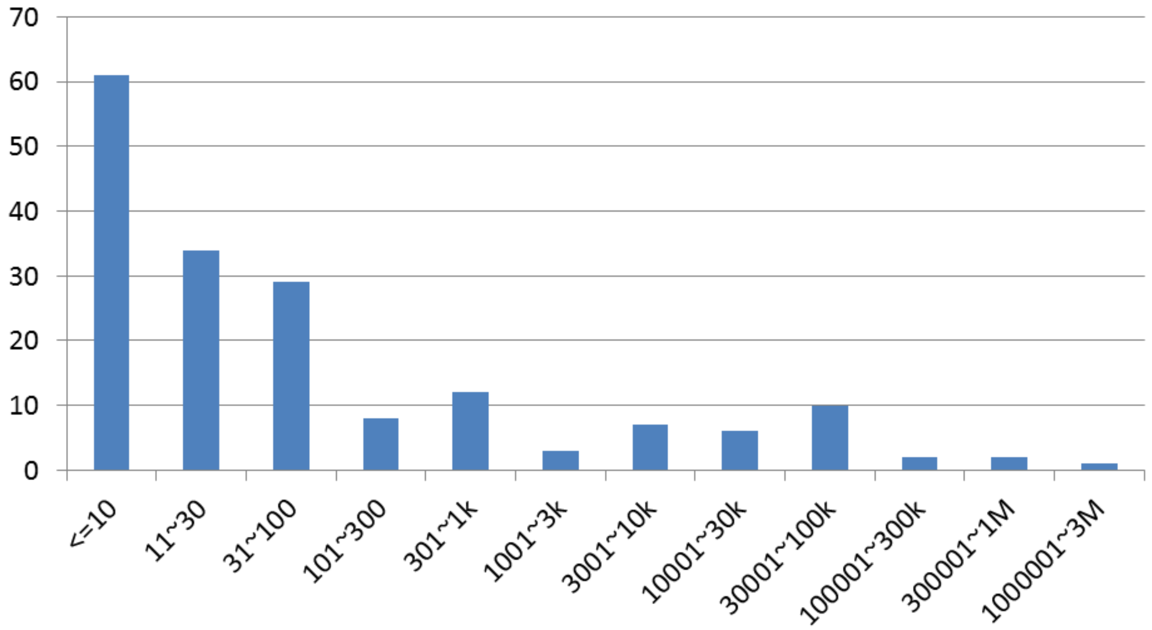


Figure 3.3 Number of classes

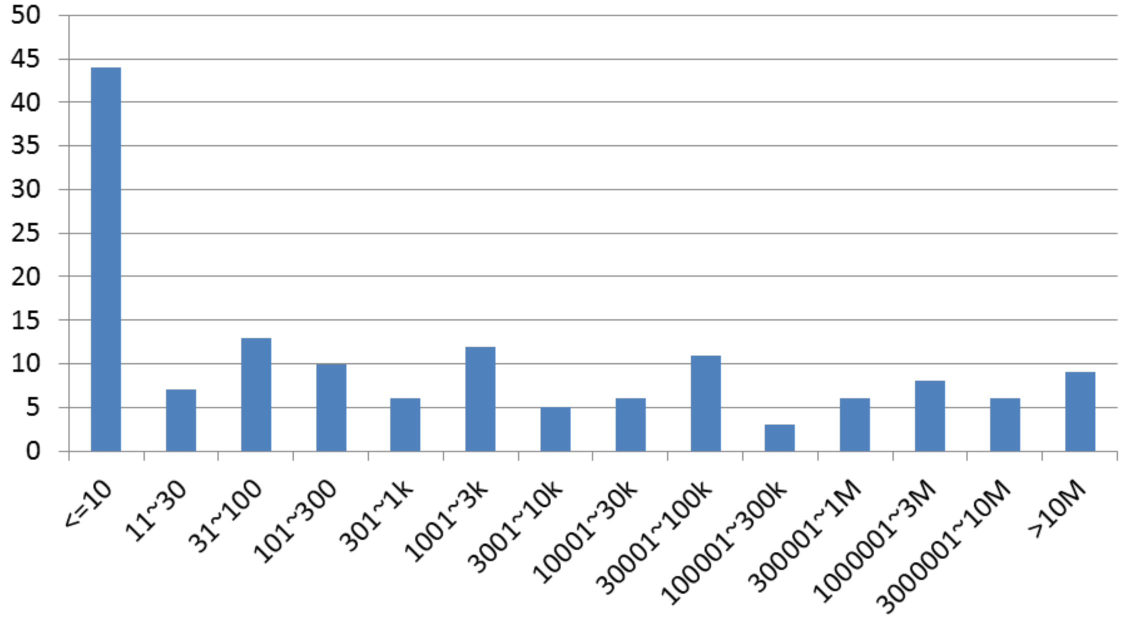


Figure 3.4 Number of instances

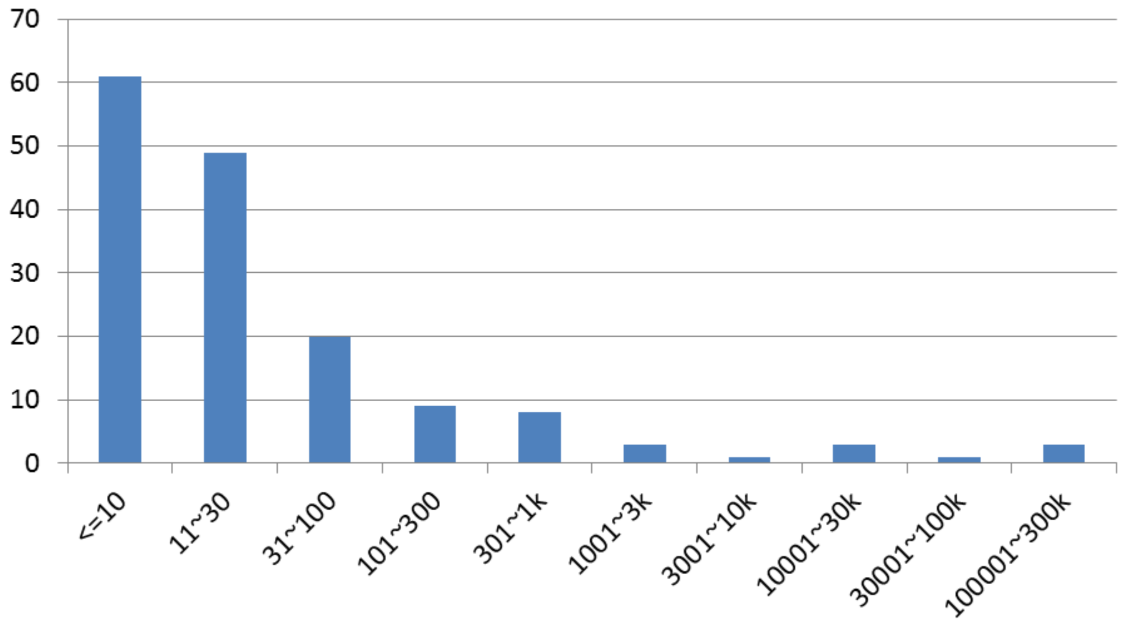


Figure 3.5 Number of properties

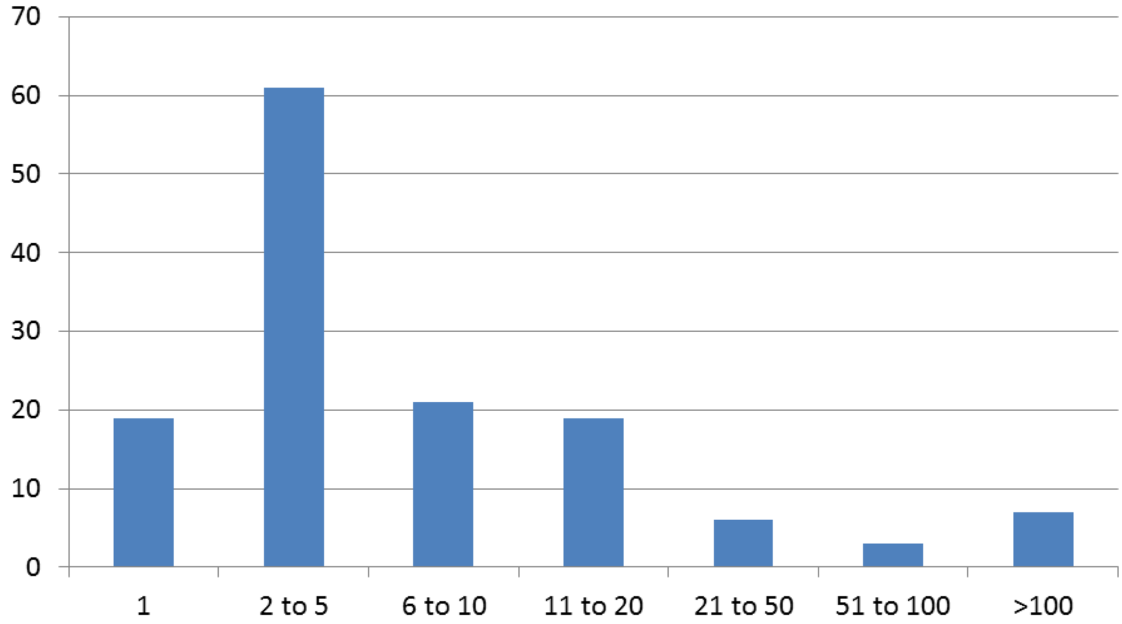


Figure 3.6 Number of top-level classes

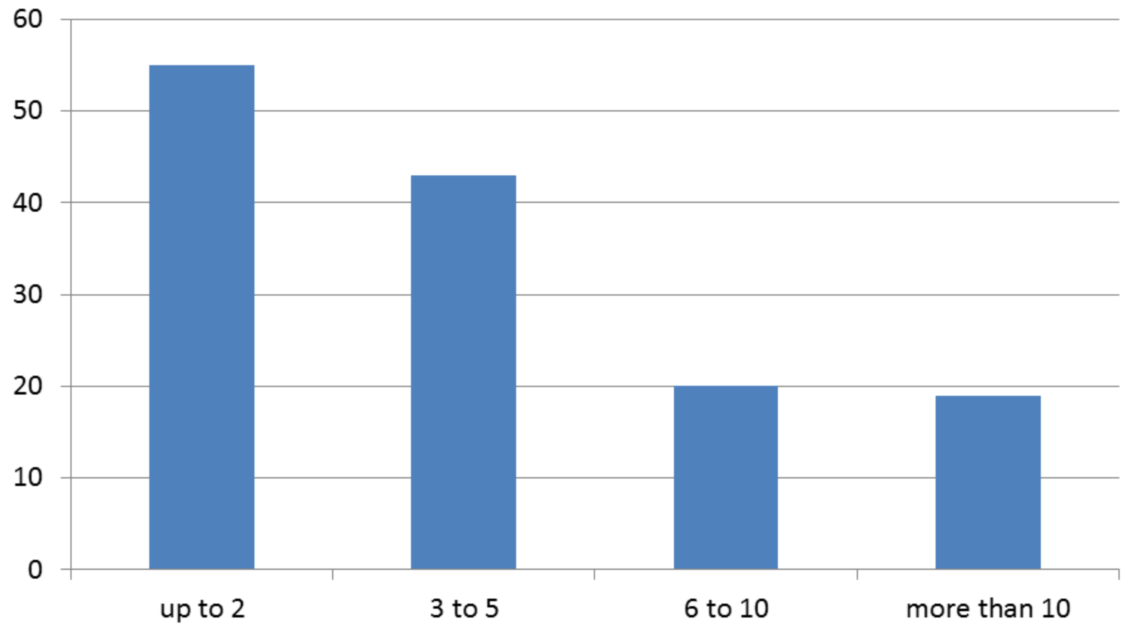


Figure 3.7 Depth of hierarchy

Many ontologies with a very large number of classes were in the biomedical domain. For example, the response in the range 1,000,001 to 3 million represented the Open Biomedical Ontologies which are really a set of ontologies, including the Gene Ontology. SNOMED CT was one of the responses in the range 300,001 to 1 million. Many of the ontologies of depth greater than 10 were also in the biomedical domain. They include, for example, the OBO ontologies. Outside the biomedical domain, CYC was an example of an ontology with depth more than 10.

3.4 Purposes for using ontologies

Respondents were asked for which purposes they used ontologies. There were eight options from which to choose, plus 'other'; multiple responses were permitted. There were in all 341 responses, including 'other', from 73 respondents. The description of each of the predefined options, as provided to the respondents, is shown in figure 3.8.

Figure 3.9 shows the responses. The figure also shows the application area of the respondents; there is no obvious relationship between application area and choice of purposes. A number of the 'other' responses could be seen as particular cases of one of the options, e.g., "formalising a domain to infer knowledge" is a specific case of conceptual modelling. Excluding 'other' there were 332 responses from 72 respondents, i.e., an average of 4.6 responses out of 8 possibilities.

- Conceptual modelling
  - e.g., formally defining a domain
- Data integration
  - i.e., merging a number of databases
- Defining knowledgebase schemas
  - e.g., as a means of storing and retrieving information
- Knowledge sharing
  - e.g., between individuals in an organisation
- Linked data integration
  - e.g., linking data from different public knowledgebases
- Ontology-based search
  - i.e., using ontologies to refine search
- Providing common access to heterogeneous data
  - i.e., providing a common schema for data access
- Supporting natural language processing
  - e.g., analysing text to identify instances of concepts in the ontology

Figure 3.8 List of 'purposes' provided to respondents

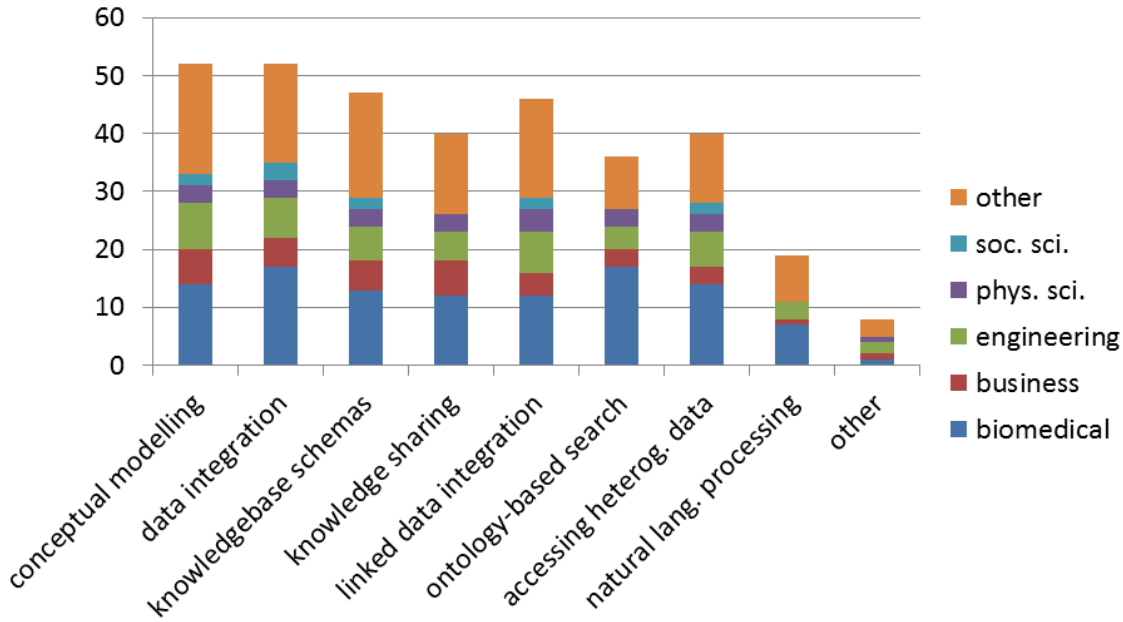


Figure 3.9 Responses to question 'for which purposes do you use ontologies'

4 Tools and visualization

4.1 Ontology editors

Respondents were asked which ontology editors they used. They were given a choice of 12 editors, and there was also an 'other' option. Multiple responses were permitted. 65 respondents replied to the question. Figure 4.1 shows all the tools for which there was more than one response. All the tools indicated were amongst the predefined options, except for OBO-Edit and Neurolex. The predefined categories also comprised the following tools for which there was no response: Internet Business Logic, OilEd, Ontolingua, OntoTrack, WebODE.

Figure 4.1 also shows the breakdown of responses amongst the application area. Apart from the two specifically biomedical editors (OBO-Edit and Neurolex), there appears to be no particular relationship between editor and application; except perhaps the high usage of Protégé 4 within the biomedical community. What is striking is the dominance of OWL editors; the four most common are all in this category, as is SWOOP. In all, 50 of the 65 respondents had experience of OWL editors; some having used more than one.

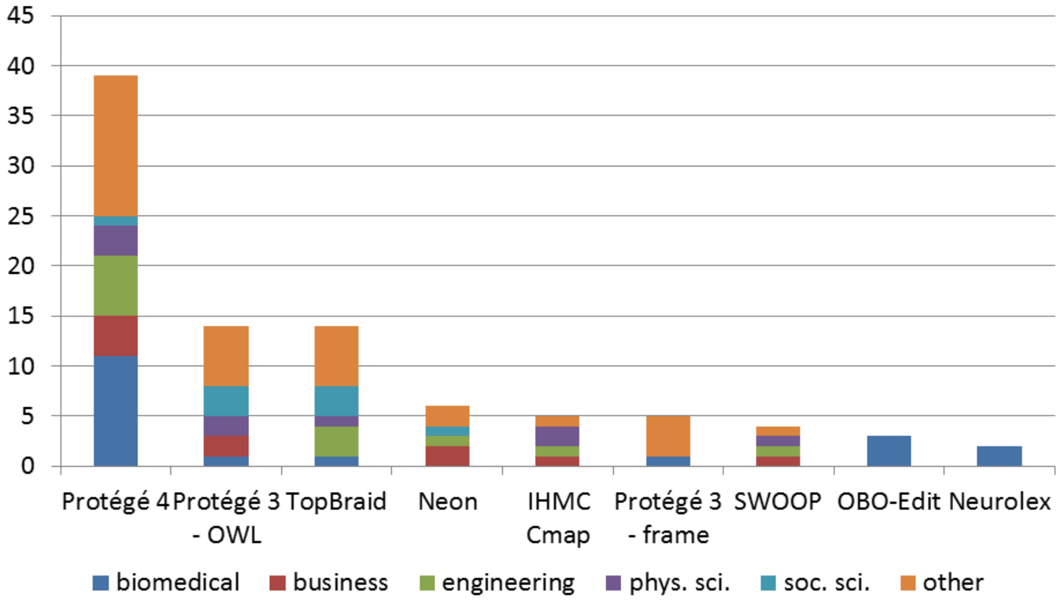


Figure 4.1 Usage of ontology editors

Respondents made a number of comments about the strengths and weaknesses of ontology editors. A brief summary is given here, under a number of headings.

Conceptual One respondent noted the need for two kinds of tools: for domain experts and for the ontology developer. Related to this, another respondent commented: "No tool I know is able to abstract from the technical details and allows non-experts to model useful and correct ontologies". Another complained that the approach based on triplets created a limitation to "property-based entity databases rather than proper graphs".

Functionality Some respondents wanted more of the kind of functionality which one would normally find in other systems development tools, e.g., auto-completion, version control, and distributed development features; one commented that "no good collaborative ontology editor exists".

Usability A few respondents commented on usability. Two commented specifically on the Protégé user interface; one saying that it needs simplifying for new users; the other that it was not suitable for working with domain experts. A comment about ontology tools in general was the lack of overview documentation.

Performance There were comments in particular about problems when working at large scale.

4.2 Visualization

Respondents were asked about their use of visualization tools. They were given a choice of 10 tools plus an ‘other’ option. Figure 4.2 shows the visualization tools for which there was more than one response. TopBraid and OBO-Edit were amongst the ‘other’ option. Amongst the predefined options, GrOWL only received one response. The other predefined options, which received no responses, were CropCircles, GoBar, OntoSphere and TGVizTab.

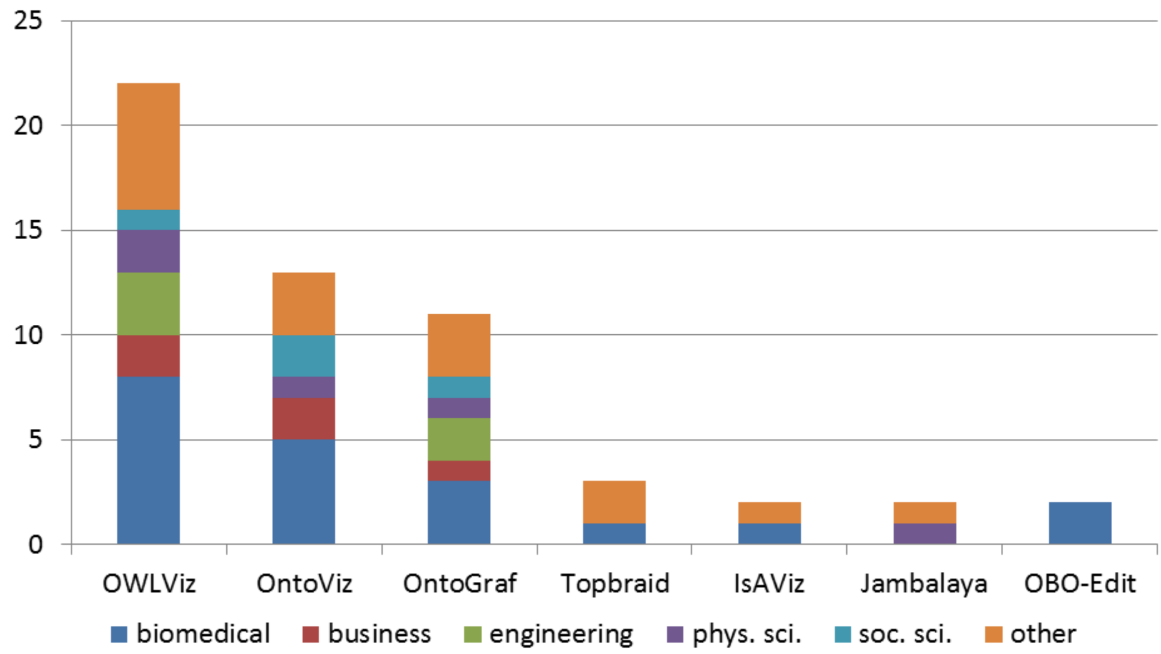


Figure 4.2 Usage of visualization tools

Figure 4.3 shows the response to the question: ‘how useful do you find visualization”. There were 56 responses. Again, these are shown broken down by the application area of the respondent; although there is no obvious relationship. It is clear that there are a range of views, with the majority being in the ‘useful to a small extent’ and ‘quite useful’ categories. More work is required here to understand how people are using visualization and how visualization is and is not proving useful.

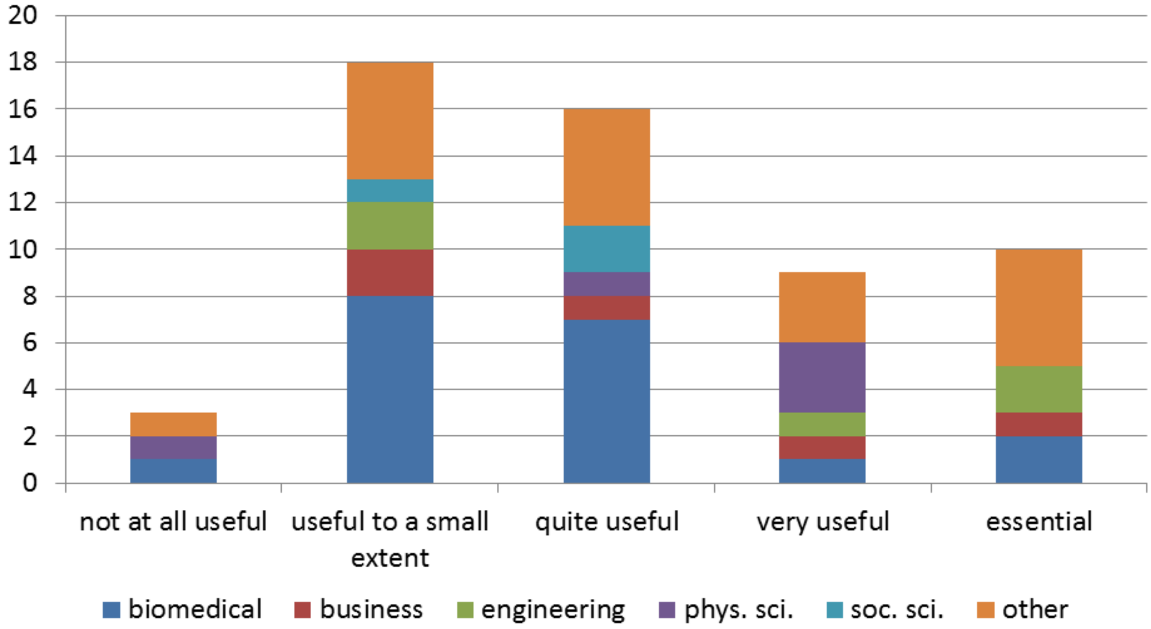


Figure 4.3 Usefulness of visualization

One respondent commented that "visualization is, especially for end-user, really hard and not task-specific". Another wanted "proper visual editing features" and to bring together "the conceptual schema with huge amounts of instance data in order to analyze the effects of changes in real time". In general, it is probable that visualization works best when it is specific to an application; this is likely also to be the case for ontology-based applications. It may be that domain experts are best served by domain specific tools incorporating domain specific visualization.

5 Ontology languages

5.1 OWL and OWL profiles

Respondents were asked about which ontology languages they used. 58 out of the 65 respondents used OWL. This is consistent with the previous information about tools, see section 4.1. Respondents who use OWL were asked which profiles, or fragments, they used. They were given the choice between the variants of OWL and OWL 2, and multiple responses were permitted. There were 54 respondents and a total of 133 responses. Figure 5.1 shows the results.

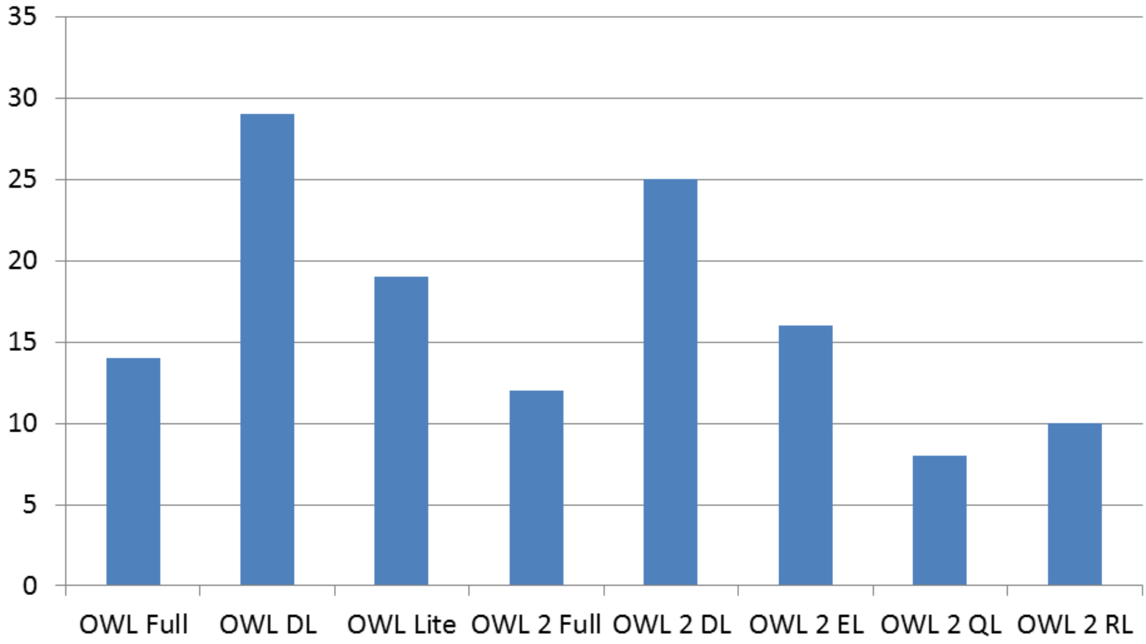


Figure 5.1 Use of OWL profiles

The large number of multiple responses seems surprising. It was thought possible that this might have been caused by the migration from OWL variants to OWL 2 variants, or the simultaneous use of both generations. That is to say, it was thought there might have been a large number of respondents indicating the use of two variants, one in each generation. However, this proved not to be the case. The number of OWL profiles indicated by each respondent is shown in figure 5.2. As can be seen, by adding the number of respondents in the columns headed 3 to 8, 19 people indicated the use of more than two variants. Moreover, of the 13 respondents who indicated use of two profiles, only 3 indicated one profile from each generation, i.e. the other 10 indicated two profiles in the same generation. Of the two people who indicated the use of all 8 profiles, one was anonymous and the other was a researcher in the area of the semantic web.

number of OWL profiles	1	2	3	4	5	6	7	8
number of respondents	22	13	4	8	5	0	0	2

Figure 5.2 Number of OWL profiles indicated by respondents;
shows considerable use of multiple profiles

The survey also asked how people constrained themselves to a particular fragment; multiple responses were permitted. Of the 32 respondents to this question, 21 indicated ‘by conscious effort’ and 14 indicated ‘constrained by editing tool’. Amongst the ‘other’ options were use of a script and use of the validator at http://www.mygrid.org.uk/OWL/Validator.

5.2 Description Logic features

Respondents were asked which of 23 description logic features they used. The results are shown in figures 5.3, 5.4, 5.5 and 5.6, relating to class features, restrictions, object properties and datatype properties respectively. These four groupings were not identified in the questionnaire, although the questions did follow this order. There were 47 respondents to the question. The average number of responses per respondent was just under 13.

One striking feature is the number of respondents who indicated use of the less common features. The least commonly used feature was the specification of object properties as irreflexive, and even for this there were 8 responses. One hypothesis is that it is only a small number of respondents who are using the less common features. However, this was not borne out by inspection of the data. For example, there were 40 respondents out of the 47 who used at least one of the 8 least common features and 45 respondents who used at least one of the 12 least common features. These numbers do need to be interpreted carefully. It is likely that the 47 respondents to this question were the more sophisticated, in terms of usage of Description Logics, and hence not representative of all users of OWL.

Figure 5.3 Usage of class features

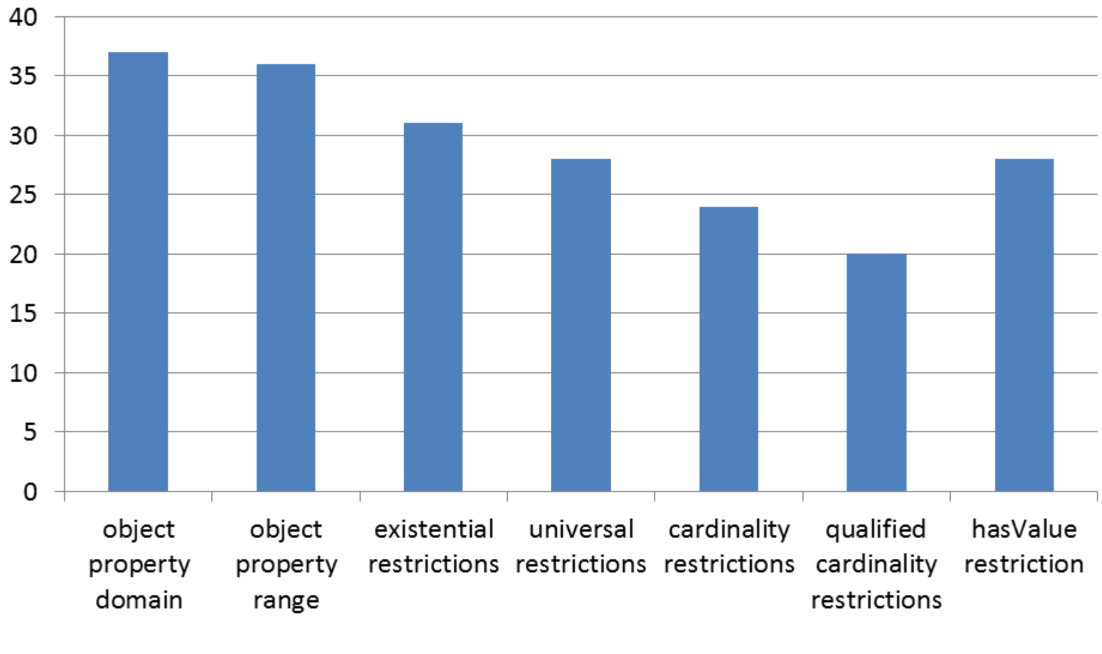


Figure 5.4 Usage of restrictions

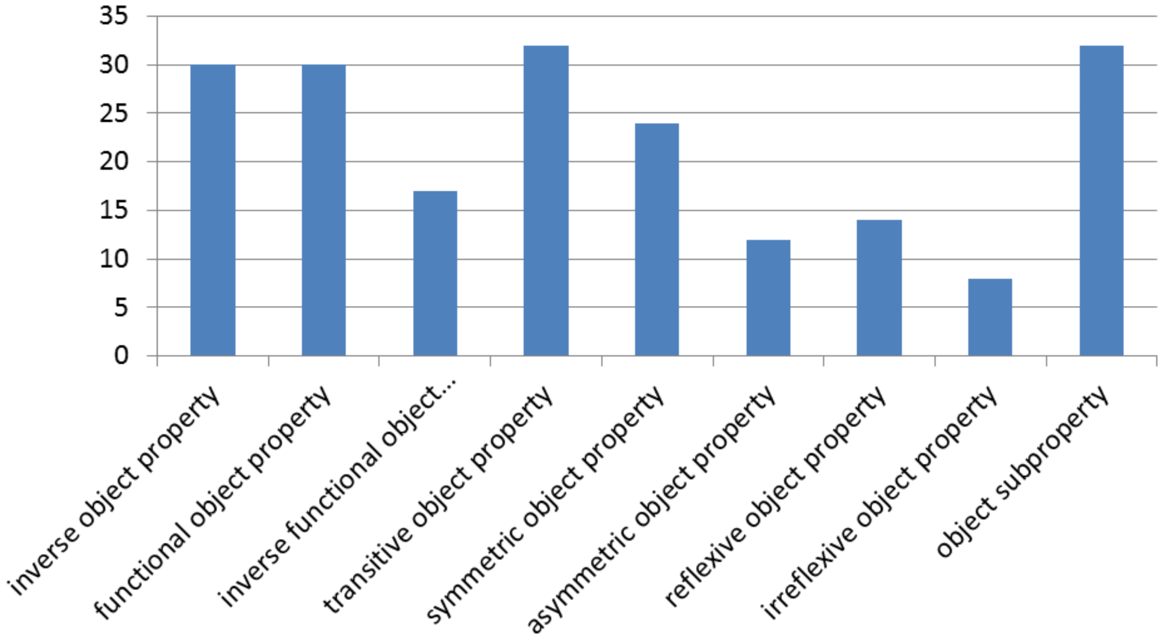


Figure 5.5 Usage of object property characteristics

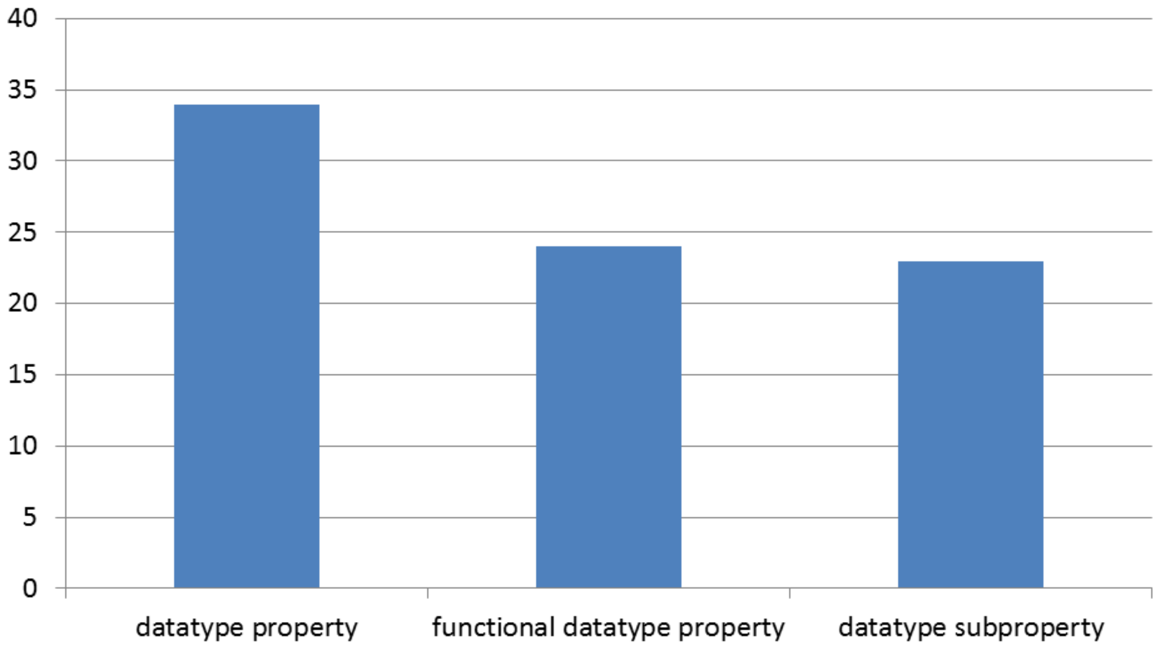


Figure 5.6 Usage of datatype properties

5.3 Comments on ontology languages

Respondents were asked for comments about their experiences with using ontology languages. Key aspects of the responses are discussed here, under a number of headings.

Conceptual One respondent wanted languages to be extended to include "acceptable" punning and also the qualification pattern, where a class is used to enable a relation to be qualified. The same respondent commented on the difficulty of classes vs. individuals and classes vs. properties design decisions. Another respondent wanted lightweight approaches to overcome the difficulty of characterising all information in a strongly semantic fashion; an example cited was where 'temperature' was interpreted as a physical phenomenon or a measurement depending on the system. Two respondents commented on the open world assumption; one noted the difficulty of grasping the implications of open world reasoning for those used to closed world reasoning; the other asking for partial support for closed world reasoning.

Complexity There were a number of comments about the complexity of ontological modelling. One respondent commented "meta models … may not suit what you're doing … The complexity of it all is way beyond what we can hope to hold in our minds at any given time, but I have yet to use a tool that makes this complexity easily understood, or even easily workable with". Another comment was "The rigor of the languages exceeds the rigor of the typical user by a wide margin".

Functionality There were requests for additional functionality, e.g., modalities, time, arithmetic functionality, more sophisticated property chain axioms. At the same time, two of the respondents making suggestions admitted that these led to problems of scalability. One respondent asked for "easier ways to integrate and reuse vocabularies". Another noted the lack of discussion, in the survey, of "the issues of overlapping ontologies", quoting the OBO Foundry ontologies as a collaborative set "that are open and can share terms".

6 Ontology patterns

6.1 Sources of patterns

The questionnaire contained a section specifically for those who used patterns. They were first asked from where they obtained their patterns. The available options and the responses are shown in figure 6.1, broken down by application domain of the respondents. There were 35 respondents, with multiple responses permitted. The ODP public catalogue is at http://www.gong.manchester.ac.uk/odp/html/index.html and contains 16 patterns. OntologyDesignPatterns.org contains a rather larger collection of patterns. The question wording relating to 'own mental models' specifically stated that these should not be written down.

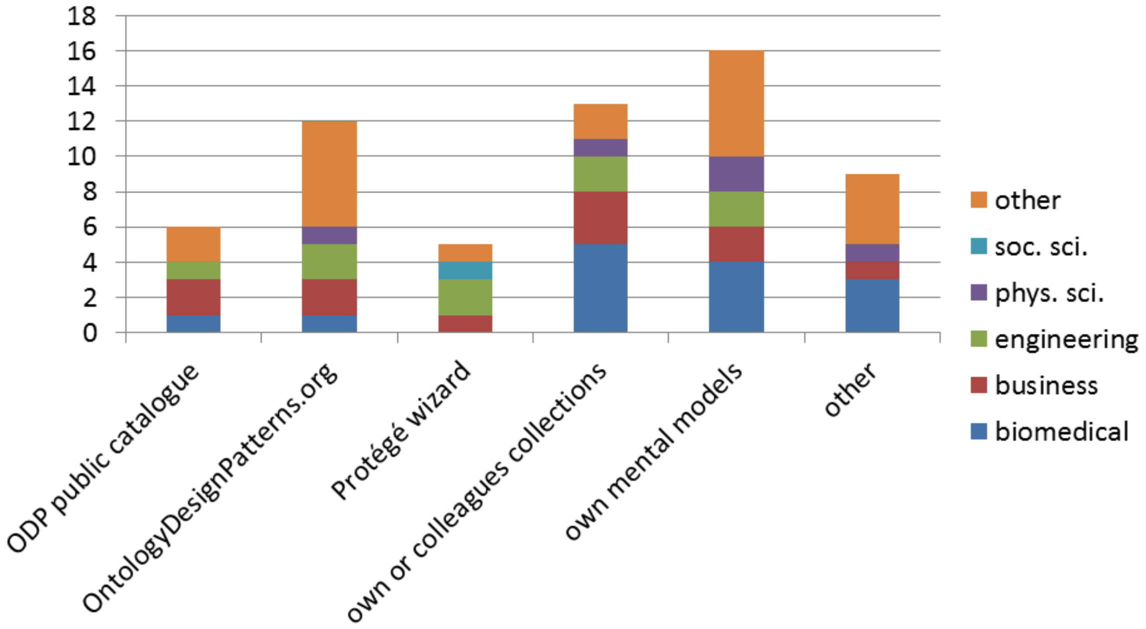


Figure 6.1 Sources of patterns

6.2 *Reasons for using patterns*

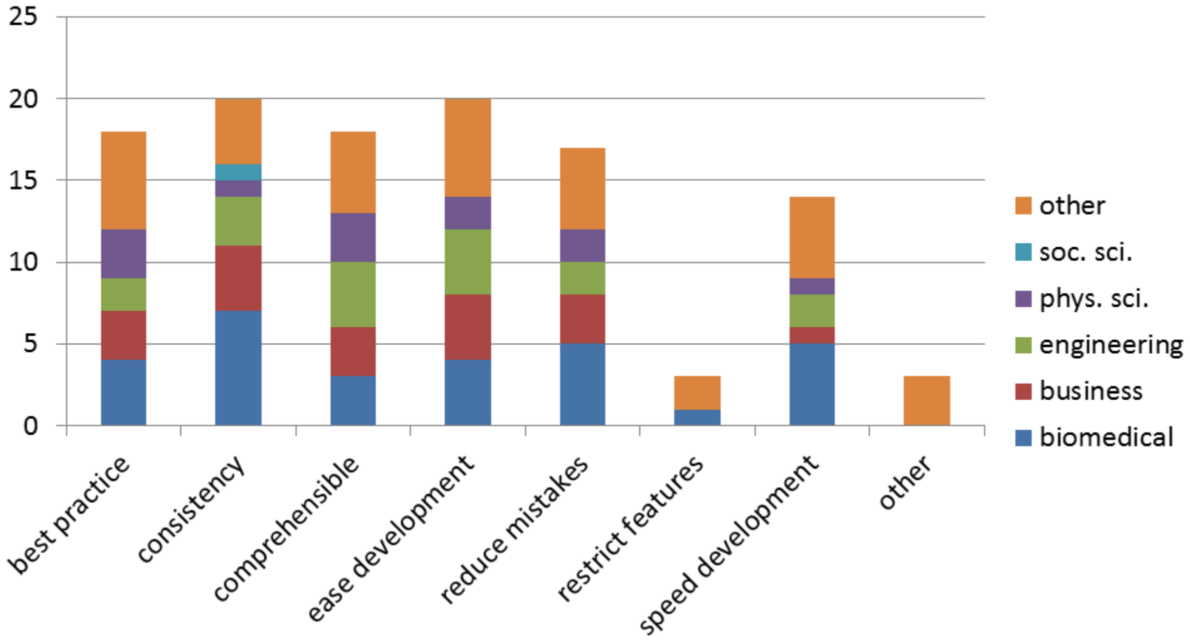
Respondents were asked why they used patterns, with seven possible answers plus 'other' and multiple responses being allowed. The results are shown in figure 6.2. Only 3 people specified 'other'. One of these could be regarded as a rewording of one of the responses provided; another was a specialization of a response option; the other was to restrict the "context and time of … roles and statuses".

6.3 *Method of use*

Respondents were also asked how they used patterns; specifically whether they imported pre-created patterns or whether they used the patterns as examples and then recreated them. The results are shown in figure 6.3. There were 32 respondents; multiple responses were permitted and the total number of responses was 38. The majority of people (24) recreated patterns; only 9 people imported patterns; there were 4 who did both. Among the 'other' responses were "fully integrated into the tool" and "sub-model fragments grouped as templates for further extensions".

6.4 *Experience with using patterns*

Respondents were asked for general comments on their experience with using patterns. One respondent commented that the "best patterns are rather simple, not very complex, basic". A researcher in the biomedical domain expressed the view that there are "seldom some available patterns out there for us to use". This may be because the required patterns are frequently domain-specific rather than generic. Another respondent called for better tool support, stating that "tools should suggest suitable patterns". One comment was about the difficulty of understanding patterns: "initially hard to learn, but provide required functionalities"; this suggests the need for better ways of representing patterns in human-readable form.



encourage best practice	enforce consistency	make ontologies more comprehensible and usable	make ontology development easier
reduce modelling mistakes	restrict language features (e.g., to OWL profiles or fragments)		speed-up ontology development process

Figure 6.2 shows the response to the question 'why do you use patterns?'
- table at bottom shows full text of each option

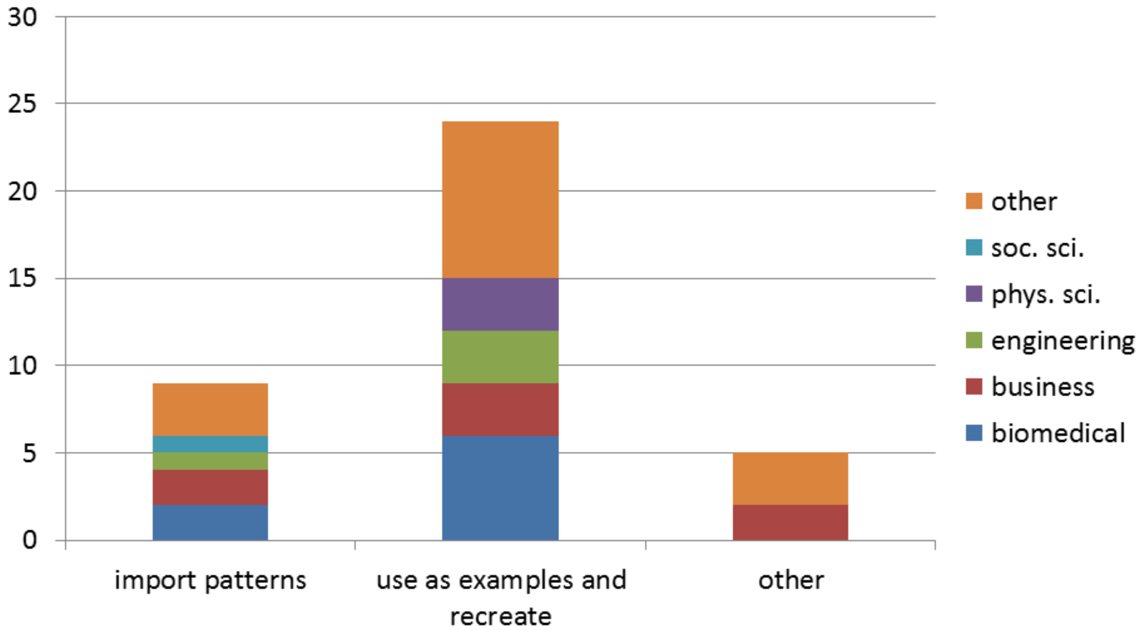


Figure 6.3 Methods of using patterns

7 Additional comments on using ontologies

Respondents were asked for any final comments on their experience with using ontologies; some of the most significant are listed here.

7.1 Modelling ontologies

A number of interesting comments related to the difficulty of designing ontologies. One related to the difficulty of defining classes, which had taken the respondent "many years of learning". The respondent also noted that domain scientists are "not too comfortable with ontology relationships" and have difficulty "building term representations".

Another referred specifically to the difficulty of modelling: "(light-weight) ontologies are great to model domain knowledge and to interchange this with machines AND colleagues. Unfortunately, their modelling isn't an easy task".

A respondent noted that in healthcare and lifesciences they used a pragmatic approach and they "need to not be as atomistic as many 'academic' systems try to be".

One comment seemed to refer to the need for stability in the domain in order for ontologies to be useful: "Perhaps useful in the domain of life sciences were Mother Nature has been organizing things for millions of years, but I'm largely in financial services where fashion rules". This suggests a need for tools to easily and rapidly adapt ontologies.

There was a call, from a respondent in the biomedical domain, for a more mature discipline of ontology design: "Ontologies should be built towards use cases and answering biological questions and this is not always the case. Engineering practices in the domain are rarely applied and immature."

7.2 Implementation

There were a few comments relating to practical implementation issues. One respondent, complaining that the ontology research community is "far out of touch with the practical applications of ontology", wanted "libraries to facilitate the integration of ontologies with day-to-day development practices" and also wanted to "leverage the best practices of UIs, catalogs, and linked open data to produce a seamless user experience". This might be seen as a call to introduce best practices from software engineering into ontology design. In a similar vein, a respondent noted that the integration of ontologies "in applications or new services is still far from being close to developers." There was also a call for more tool support, specifically for "more support for ontology mapping so that any ontology can be used."

Another seemed to echo a comment already made about the split between ontology experts and domain specialists: "The biggest problem is the world outside of your ontology editor - defining stuff in Protégé is fine, and doing domain analysis to model something is also fine. But then what? Tool support for non-experts working with ontologies/knowledgebases is generally poor."

7.3 General

There were some very general comments. One respondent was looking forward to the further evolution and application of ontology science: “Ontologists will one day wonder how they managed with the present incomplete conception of ontologies. So will IS developers”

Inevitably comments often concentrate on difficulties and what needs to be improved, but there were also some very positive comments. A respondent from the cultural sector, whilst admitting experience of the “pitfalls”, made the comment: “All in all we see an enormous advantages for the cultural heritage sector and we see that using a "pure" back-end line with RDF/OWL/SPARQL for such houses can actually deliver and be stable.”

Another respondent simply said “I couldn't build what I build without them”.