KNOWLEDGE MEDIA

KMi

INSTITUTE

Ontology patterns

A survey into their use

Technical Report kmi-14-02
March 2014

Paul Warren

paul.warren@cantab.net

The Open University

Abstract

This report describes the results of a survey into the use of ontology patterns. Responses came from industry, research institutes and academia and covered a range of different application areas. The survey covered: pattern characteristics and sources; pattern usage; and pattern identification, creation and storage.

Pattern size and domain specificity appear to be correlated. There was a significant rank correlation between specificity and number of classes in the pattern, and a close to significant rank correlation between specificity and number of individuals. Whilst some patterns were sourced from pattern catalogues, a number were from collections created by the respondent or the respondent's colleagues, or were simply the respondent's mental models, i.e. not written down. Patterns were sometimes imported. However, they were also often used as examples and recreated, either with or without modification.

Respondents described their difficulties using patterns. Two commented on the need for documentation and examples. There were comments about the need for improvements to tools, in the areas of finding patterns, pattern generation and integration, and visualization when using several patterns. There was also an observation that generally ontologies do not include information about the patterns used in their creation.

Only about a quarter of the respondents had a systematic approach to identifying the need for a pattern. A variety of techniques were used for creating and storing patterns, including the use of an ontology editor and diagrams. The latter approach was used by over a half of the respondents. Response to a question about the use of antipatterns produced a number of generic comments plus a number of specific examples.

There were only two respondents from the biomedical domain. However, their responses differed from most of the other responses in a number of ways. They did not use generic pattern libraries; one used discipline-specific patterns from own or colleagues' pattern collections; the other used mental models specific to the person's work and not written down. They did not import patterns; one created patterns with a pattern creation tool; the other used patterns as examples and modified them. The patterns they used were at the large end of the scale and they were unique in using formal languages to create and store patterns.

Keywords: ontology design patterns; survey; content ontology patterns; logical ontology patterns

Contents

1 Introduction

This report provides the results of a survey into the use of ontology patterns, conducted during the second half of 2013. This work follows on from a more general survey into the use of ontologies, see Warren (2013). Section 6 of that report provided information on:

- the sources of patterns, e.g. pattern libraries, or own or colleagues collections;
- the reasons for using patterns, e.g. encourage best practice, or reduce modelling mistakes;
- the methods of use, i.e. whether they were imported or used as examples and recreated;
- experience with using patterns.

The follow-on survey looked in more detail at some of these questions besides investigating some additional issues relating to patterns.

There are a variety of pattern categories. The OntologyDesignPatterns.org[1] library provides a classification into structural, correspondence, content, reasoning, presentation and lexico-syntactic. Three of these categories are further subdivided, see figure 1.1. Falbo et al. (2013) discuss the classification of patterns, describing the classification at OntologyDesignPatterns.org and also an alternative classification approach related to the ontology development phase at which the pattern is to be used. The introduction to the survey defined ontology patterns as "reusable solutions for use in the design of ontologies". The survey did not provide a classification of patterns. However, a number of the questions were oriented towards logical and content patterns.

Logical patterns are intended to overcome any limitations of expressivity in the language being used. In the terminology of OntologyDesignPatterns.org they are a subset of structural patterns. An example given there is the n-ary relation, which "cannot be directly expressed in OWL". With minor exceptions, such as Thing ($\top$), they do not contain entities, e.g. classes and properties. They can often be expressed as a macro, which may generate entities specific to the target ontology. Content patterns do contain entities, and can be regarded as mini-ontologies. An example of a content pattern is the PartOf pattern[2], which is concerned with part-whole relations.

The following sections report in the order of the survey. The next section provides some information on the respondents. Section 3 then provides information about the patterns being used. Section 4 talks about how patterns are used, including the difficulties of using patterns. Section 5 provides information on how patterns are identified, created and stored, including the tools used. This section also includes some information about the use of antipatterns. Section 6 describes some general feedback from respondents. Finally, section 7 draws some conclusions.

[1] http://ontologydesignpatterns.org
[2] http://ontologydesignpatterns.org/wiki/Submissions:PartOf

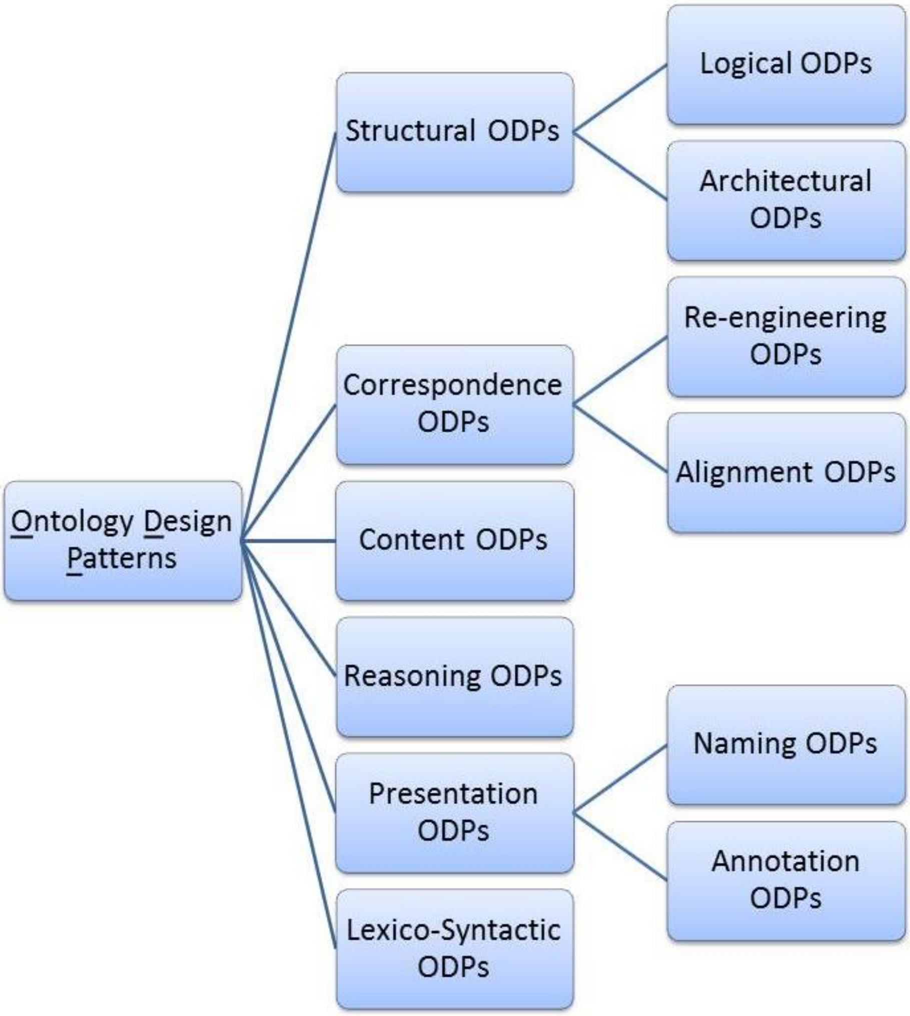


Figure 1.1 Classification of ontology design patterns, from OntologyDesignPatterns.org. Survey questions were largely oriented towards logical and content ODPs.

2 The respondents

The respondents were obtained using a number of relevant mailing lists, including:

- ontolog-forum (ontolog-forum@ontolog.cim3.net)
- UK Ontology Network (ontology-uk@googlegroups.com)
- three LinkedIn groups: Semantic Web Research, Semantic Web for Life Sciences, and Description Logic.

In addition, certain respondents to the original survey who expressed an interest in ontology patterns were re-contacted. 13 respondents provided information. Their breakdown by application area is shown in Figure 2.1. The figure also shows the split between academic, research institute and industrial.

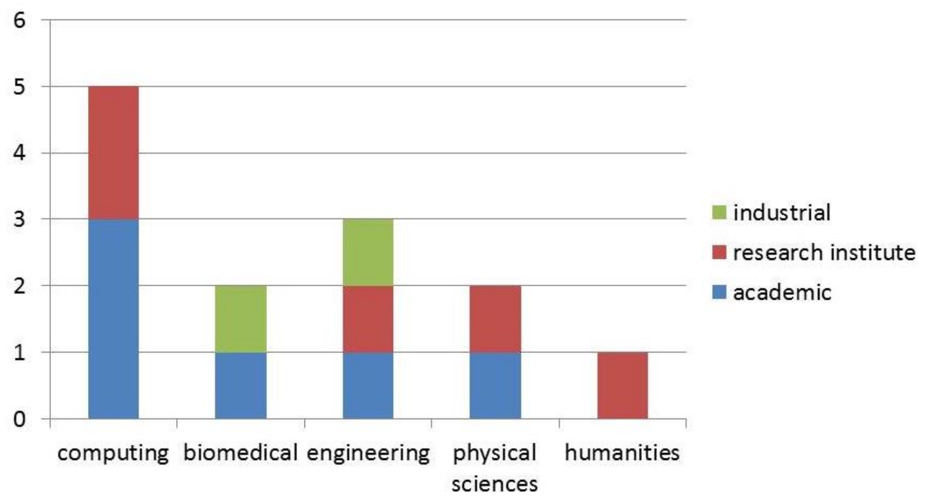


Figure 2.1 Breakdown of respondents; no information was received relating to the two other category options: ‘business’ and ‘social sciences’

3 The patterns

3.1 Pattern characteristics

Respondents were asked about the types of patterns they used. Specifically they were asked:

- the number of classes;
- the number of individuals;
- the number of properties;
- the domain specificity;
- the number of patterns of this type that they used.

Respondents were allowed to specify up to five different pattern types. In fact, no one specified more than four. Figure 3.1 shows the range of responses permitted in each case.

number of classes, individuals and properties
0
1-2
3-5
6-10
11-30
31-100
greater than 100

specificity
generic
suited to a broad discipline, e.g. anatomy
suited to a specific discipline, e.g. human anatomy
specific to my work

number of patterns of this type
1-2
3-5
6-10
11-30
31-100
greater than 100

Figure 3.1 Permitted responses to specify type of patterns

In all 11 respondents provided 17 patterns. In each case respondents provided all five statistics requested, as in the bullet points above, except that four submissions did not include the number of individuals. In these cases this number was assumed to be zero. Figure 3.2 shows the distribution of the number of classes and number of individuals, broken down by specificity. Figure 3.3 shows the distribution of the number of properties.

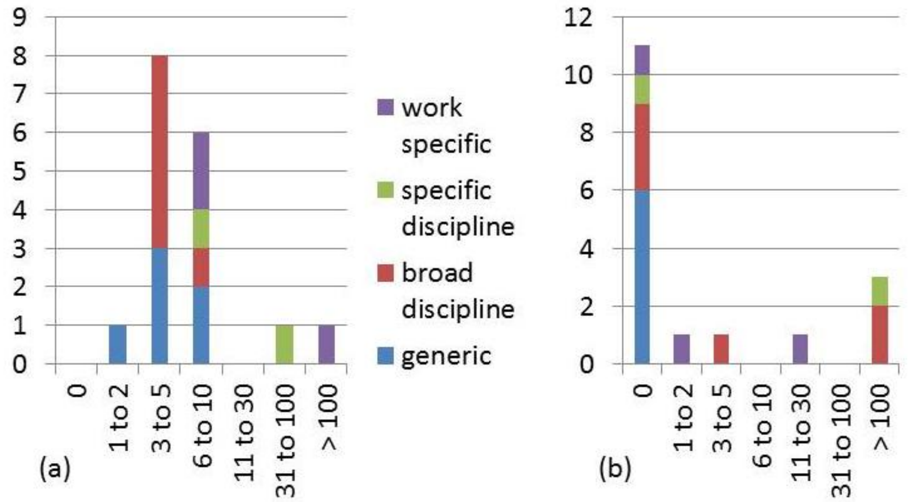


Figure 3.2 (a) Distribution of number of classes; (b) distribution of number of individuals; in both cases broken down by specificity

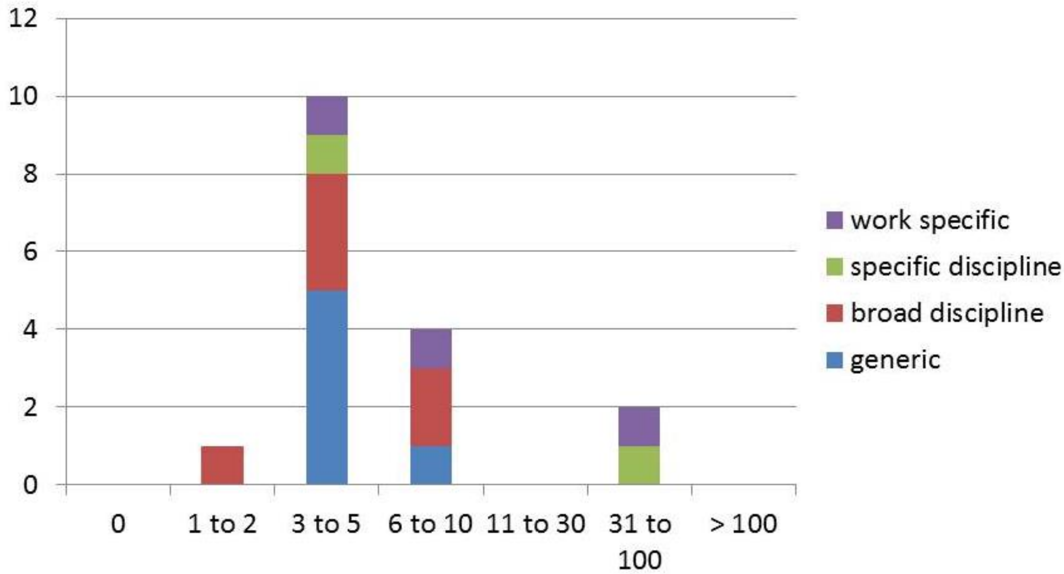


Figure 3.3 Distribution of number of properties, broken down by specificity

Figure 3.4 shows all 17 submissions, ranked by number of classes, number of properties, number of individuals, specificity and number of patterns of that type. Each of the 17 pattern types differed in at least one of number of classes, number of people, number of individuals, or specificity, with the exception of the two submissions at the top right of the figure.

classes	props	indivs	spec	numb
1-2	3-5	0	gen	1-2
3-5	1-2	0	br. dis	6-10
	3-5	0	gen	3-5
			gen	31-100
			br. dis	>100
		>100	br. dis	1-2
	6-10	0	gen	11-30
			br. dis	3-5
		3-5	br. dis	1-2

classes	props	indivs	spec	numb
6-10	3-5	0	gen	1-2
				3-5
		11-30	work	6-10
		>100	br. dis	1-2
			sp. dis	3-5
	6-10	0	work	>100
31-100	31-100	0	sp. dis	1-2
>100	31-100	1-2	work	3-5

Figure 3.4 17 pattern submissions, all differentiated by size or specificity, except for pattern at top right which is differentiated only by number of patterns used

In this sample of 17 patterns, all the generic patterns are small, i.e. no more than 3 classes and 10 properties, and no individuals. Patterns in the 'broad discipline' category were also small, with the exception of two with more than 100 individuals, both related to environmental data. There is an apparent tendency for pattern size to increase with increasing specificity. Spearman's rank correlation showed that the only significant rank correlation between any pair of the five dimensions was between specificity and number of classes ($p = 0.015$). The rank correlation between specificity and number of individuals was close to significance ($p = 0.055$). The rank correlation between specificity and number of properties was not significant ($p = 0.136$) and all other p-values were greater than this.

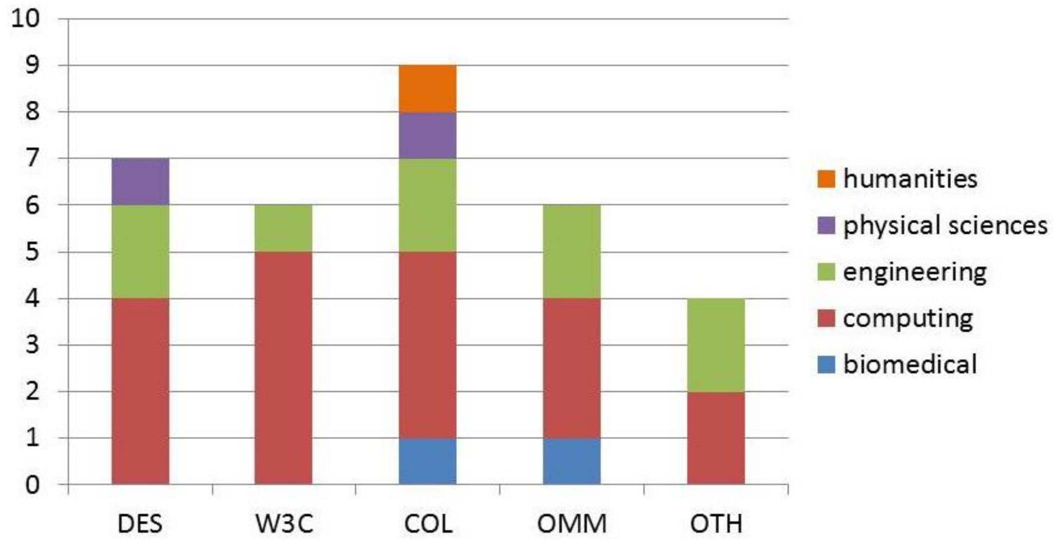
There were two pattern types where the respondent used more than 100 such patterns, and another pattern type where the respondent used 31 to 100 patterns. In fact, all three of these submissions were from the same respondent who worked in the domain of computing. Another submission, also in the computing domain indicated the use of 11 to 30 patterns. Apart from that, no other respondents indicated more than 10 patterns for any one pattern type.

3.2 Pattern sources

13 respondents provided information about the sources of their patterns. The distribution of responses is shown in figure 3.5. Apart from the categories shown in the figure, the following options were provided and received nil responses:

- the ODP public catalogue (http://www.gong.manchester.ac.uk/odp/html/index.html);
- the OBO foundry (http://www.obofoundry.org/);
- the Protégé patterns wizard.

Amongst the 'other' category, two respondents cited Linked Open Vocabularies (http://lov.okfn.org/dataset/lov/).



DES: OntologyDesignPatterns.org
W3C: World Wide Web Consortium
COL: Pattern collections created by yourself or colleagues
OMM: Your own frequently used mental models (i.e. not written down)
OTH: Other

Figure 3.5 Distribution of pattern sources; data from 13 respondents. Where a respondent has cited the same pattern source for two different pattern types, this is counted once.

Whilst it is not possible to draw any firm conclusions from a small sample, it is worth noting that the two disciplines most removed from the mathematical and physical sciences, i.e. biomedical and humanities, use only patterns from their own discipline. This implies, that for these respondents at least, the more generic patterns developed in computer science are not relevant; or perhaps they are simply unaware of their existence. The two respondents in the biomedical domain also reported the use of large patterns, specifically the two at the bottom right of figure 3.4.

4 Using patterns

4.1 Modes of usage

Respondents were also asked how they use patterns. The object of this question was to identify, for example, whether the respondent uses patterns as is, or whether the respondent modifies them. Figure 4.1 shows the response to this question, and also the wording of the question options. An additional option, 'Generate in ontology editor, e.g. using Protege patterns wizard' received a nil response. Note that the option on the extreme left, 'import patterns' is equivalent to what Falbo et al. (2013) call 'reuse by extension'. The two central options, 'recreate unmodified' or 'recreate modified' are equivalent to 'reuse by analogy'. The rightmost option, 'generate with a tool' represents either the original pattern creation phase, or alternatively 'reuse by analogy'.

Again, the sample is too small to permit definitive conclusions. However, it is relevant that the respondents in the biomedical and humanities domains did not use existing patterns but either modified patterns or created their own patterns with pattern tools.

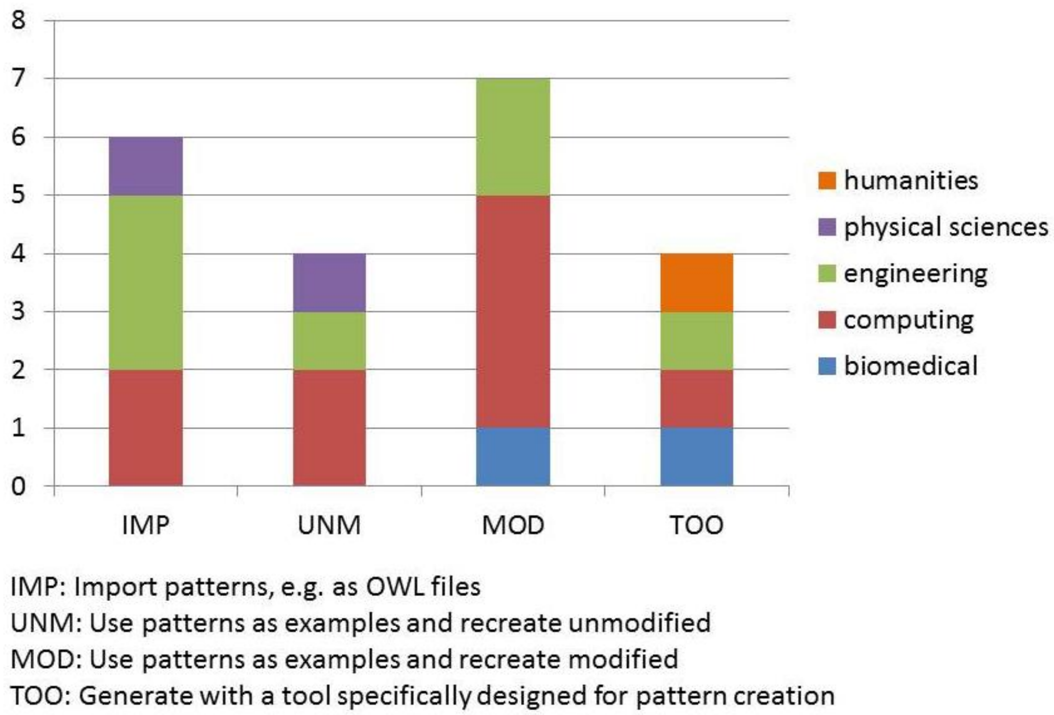


Figure 4.1 Distribution of pattern usage modes; data from 13 respondents

There were four distinct responses in the 'other' category:

- Extreme Design (XD) Tools, see Presutti et al. (2009) and Blomqvist et al. (2010).
- Tawny OWL, see Lord (2013). This uses a programming paradigm, in contrast to the approach of, e.g. Protégé.
- Excel and XSLT.
- Own unpublished tools.

4.2 Difficulties in using patterns

Respondents were asked whether they experienced any problems in using patterns. Two commented on the need for documentation and examples. One of these noted that "the documentation should be extensive and rich of examples". The same respondent also commented on the complexity of visualization when several patterns are used simultaneously. The respondent regarded this as a problem of the editor. There were two other comments relating to tools. One was concerned with the difficulty of finding the right pattern; the other with the difficulty of pattern generation and integration into existing ontologies.

One respondent noted that when ontologies are imported, information about patterns is not available. This respondent's strategy was to draw the ontology "until the pattern appears". This may have been a reference to architectural patterns rather than content or logical patterns, see the discussion in section 1. In any case, the same general point could be made about other types of patterns. Taken together with the comment in the paragraph above about the need to visualize several patterns being used simultaneously, this suggests that an ontology editor could usefully identify patterns, e.g. through colour coding, and keep track of their provenance.

Another respondent found that some patterns had "too many constraints" to be fit for that respondent's purpose. To an extent this may be inevitable; patterns are often designed for a particular context and may not be appropriate for another. More generic design of patterns should mitigate this problem. There is a need for guidelines to help people create patterns suitable for reuse by others.

4.3 Example patterns

Respondents were asked to give examples of patterns which they used, and to indicate from where the patterns were obtained. No one used English or Manchester Owl Syntax (Horridge et al., 2006) to describe a pattern, as respondents were invited to do. Most of the eleven responses cited patterns in libraries. Three respondents referred to the OntologyDesignPatterns.org library, already mentioned ni section 1. The other libraries were:

- The *Some More Individual* webspace at http://smiy.sourceforge.net/. The webspace refers to itself as providing ontologies, although they could be regarded as large content patterns.
- Linked Data Patterns, a pattern catalogue in the form of a book, at http://patterns.dataincubator.org/book/ .
- A set of ontologies and patterns at http://www.essepuntato.it/ . The particular pattern cited was the *time-indexed value in context* pattern at http://www.essepuntato.it/2012/04/tvc .

References to specific patterns were chiefly to logical or content patterns. There were two references to the use of domain specific patterns for environmental data. One respondent mentioned the use of annotation patterns. Another mentioned "display formats", which can be regarded as another form of pattern. The same respondent mentioned "process organizations, e.g. for ETL, giving reflective runtime control".

5 Identifying and creating patterns

5.1 Identifying the need for a pattern

Respondents were asked how they identified the need for a pattern. The results are shown in figure 5.1. The intention of the question was to identify how many respondents used a systematic approach rather than relying on serendipity. As can be seen, only three respondents out of thirteen undertook a systematic analysis to identify potential patterns.

Response	number of respondents
By noticing repeated use of identical or similar structures	10
By systematically analysing the ontologies I work with, e.g. using a tool	2
"Both of the above" (in 'other' category)	1

Figure 5.1 Responses to the question "How do you identify the need for a pattern?"
Only one response was permitted.

5.2 Creating and storing patterns

Figure 5.2 shows the response to the question "How do you create and store patterns?" There were thirteen respondents, with multiple responses permitted. The most striking feature is that over half make use of diagrams. It is also interesting that the two respondents from the biomedical domain both used formal languages, but no other technique. Conversely, none of the other eleven respondents used formal languages. Whilst this is again too small a sample to draw definitive conclusions, it is another potential point of differentiation between the biomedical and other domains.

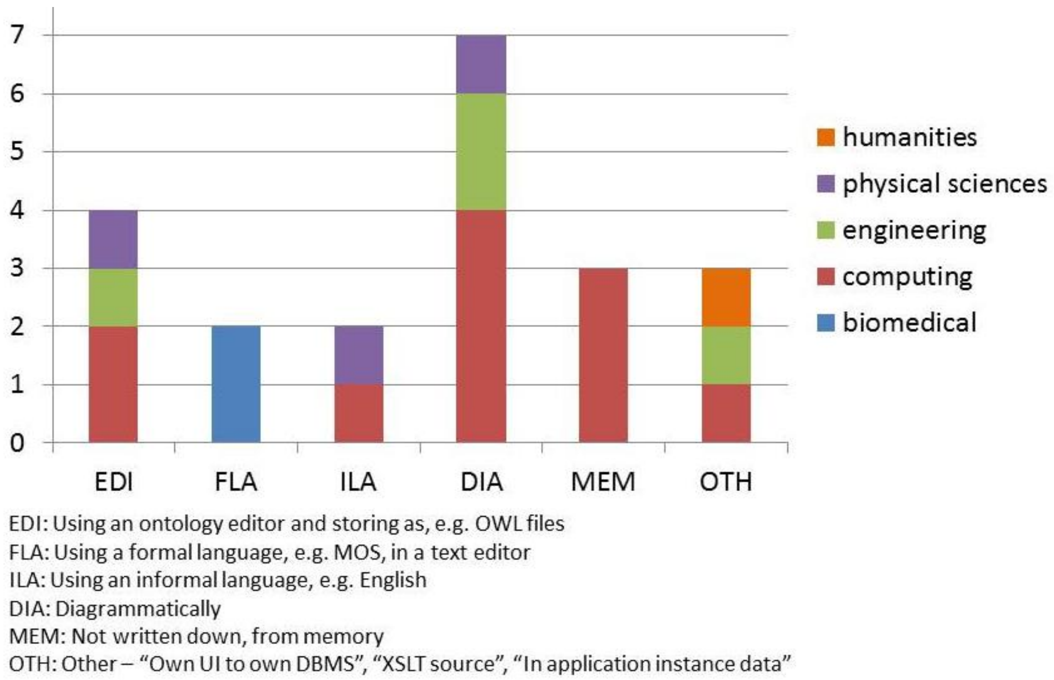


Figure 5.2 Creating and storing patterns; data from 13 respondents.

5.3 Antipatterns

Respondents were also asked whether they identified “erroneous or ineffective patterns to avoid (sometimes called antipatterns)”. There were a number of generic responses to the question and also some specific examples of antipatterns. The latter are shown in figure 5.3. Amongst the generic responses, one referred to “the classic ETL runtime horror stories”, another cited the *only* constructor as leading to “bad modelling structure”. One respondent used OOPS, the Ontology Pitfall Scanner[3], which scans an ontology for common pitfalls. It is based on a pitfall catalogue which can be viewed, although only some of the pitfalls can be detected by OOPS.

Antipatterns	
“AND and OR abuse”	“lack of connection between nary projections”
“part as membership”	“properties that have the same name as a class”
“cause as entailment”	“modelling roles as subclasses of people”
“broader as subsumption”	“use the class hierarchy as a topic hierarchy”

Figure 5.3 Antipatterns reported by respondents

6 Final comments

Respondents were asked if they had any final comments on their experiences with using patterns. One comment was that “many of them are just ontologies that somebody might want to reuse”, which indeed is true for content patterns, “and examples of really generic and abstract patterns are quite rare”. It is certainly true that the OntologyDesignPatterns.org library contains many more content patterns than logical patterns. This is not surprising, given the opportunities to create domain-specific

[3] http://oeg-lia3.dia.fi.upm.es/oops/index-content.jsp

content patterns. Another respondent viewed abstract patterns as being the more useful: “the more abstract the pattern is – the more it helps framing out this or that part/context of an ontology”.

A very general observation was about the need to build a bridge between the community which is concerned with patterns and the community which is concerned with linked open data.

7 Conclusions

Despite the relatively small number of respondents, some tentative conclusions can be drawn. Turning first to the nature of the patterns being used, there is a correlation between domain specificity and the number of classes and properties, but not the number of individuals, in the pattern. Related to this, the small sample of respondents from the biomedical domain used large specific patterns created within their community. Generic patterns were only used by computer scientists and engineers.

A number of areas for improvement were identified. Better documentation and more examples are required. Visualization could also be improved, perhaps including visualizing patterns within ontologies. Related to the need for better visualization, it is worth pointing out that half of the respondents used diagrams to create or store patterns. Tools for finding and generating patterns, and integrating them into ontologies, could also be improved. Finally, few respondents had a systematic approach to identifying patterns, and this is an area which requires further work.

References

Blomqvist, E., Presutti, V., Daga, E., & Gangemi, A. (2010). Experimenting with eXtreme design. In *Knowledge Engineering and Management by the Masses* (pp. 120–134). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-16438-5_9

Falbo, R., Guizzardi, G., Gangemi, A., & Presutti, V. (2013). Ontology patterns: clarifying conceptsa and terminology. Presented at the Workshop on Ontology and Semantic Web Patterns - WOP2013, Sydney, Australia. Retrieved from http://ontologydesignpatterns.org/wiki/WOP:2013

Horridge, M., Drummond, N., Goodwin, J., Rector, A., Stevens, R., & Wang, H. H. (2006). The manchester owl syntax. *OWL: Experiences and Directions*, 10–11.

Lord, P. (2013). The Semantic Web takes Wing: Programming Ontologies with Tawny-OWL. *arXiv Preprint arXiv:1303.0213*. Retrieved from http://arxiv.org/abs/1303.0213

Presutti, V., Daga, E., Gangemi, A., & Blomqvist, E. (2009). eXtreme design with content ontology design patterns. In *Proc. Workshop on Ontology Patterns, Washington, DC, USA*. Retrieved from http://ceur-ws.org/Vol-516/WOP2009_Proceedings.pdf#page=90

Warren, P. (2013). *Ontology Users' Survey - Summary of Results* (KMi Tech Report No. kmi-13-01). Retrieved from http://kmi.open.ac.uk/publications/pdf/kmi-13-01.pdf