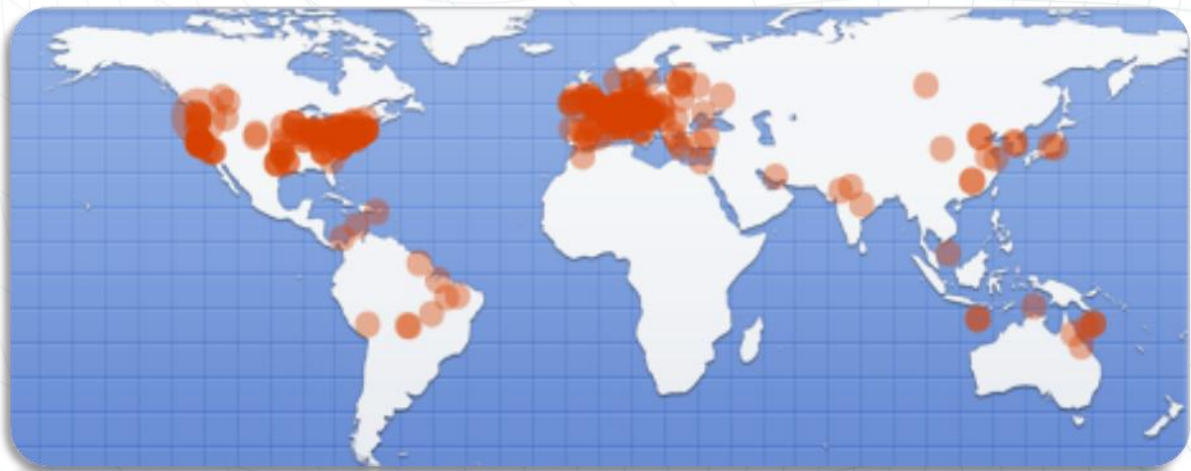


Survey results

Do controlled vocabularies matter?

Florian Kondert, Thomas Schandl, Andreas Blumauer
Vienna, June, 2011



This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit creativecommons.org or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Table of Contents

Summary.....	3
Survey results.....	4
1) Do you or your organization already make use of controlled vocabularies like taxonomies or thesauri?.....	4
2) How long have you been using controlled vocabularies?.....	4
3) Which kind of controlled vocabulary do you use or plan to use in your applications?	6
4) Which kind of controlled vocabulary do you plan to use in your	7
applications?.....	7
5) What are the main application areas of controlled vocabularies from your perspective?	8
6) How important is the usage of standards like SKOS for controlled vocabularies?.....	10
7) Do you think enterprises and other organizations can significantly benefit from using Linked Data?	11
8) Do you think that search engines which utilize thesauri to improve results will become main-stream?	12
Results regarding PoolParty	13
9) Have you heard about the PoolParty Thesaurus Management system before?.....	13
10) In which areas do you think PoolParty Thesaurus Management can be useful for you?	14
Facts about participants.....	15
11) In which of the following sectors is your organization primarily active?.....	15
12) In which area do you work in your organization?	15
13) What size is the organization you work for?	16
Conclusio & words of thanks.....	16

Summary

We are happy to present our survey results about the relevance of controlled vocabularies for the purpose of enterprise information management.

We ran our survey with the online-service [surveygizmo](http://surveygizmo.com). The survey was open from 5th of April, 2011 until May 18. 158 participants from 27 countries (see the map on the front cover of this report which represents the geographic distribution of all participants) followed our call to participate. Most of the attendees are working in IT, Science, Public Sector and Education.

Some words about the motivation:

The target of our investigation was to learn more about some aspects that are related to controlled vocabularies. Since we haven't found any clear information concerning our questions we decided to start this explorative study, knowing that the results most probably wouldn't meet scientific quality standards.

Our interest was focusing on:

- time of using controlled vocabularies
- preferred knowledge models
- main application areas
- importance of standards
- benefit of Linked (Open) Data
- trends in branches
- trends in organizations sizes

Although the data is neither statistically reliable nor representative there are some valuable insights:

- most of our participants do have a clear awareness about controlled vocabularies and 85,4% are using them in their organization
- the bigger the organization the longer controlled vocabularies are used
- taxonomies and ontologies seem to be the preferred knowledge models
- semantic search, data integration and structure for content navigation are the main application areas for controlled vocabularies
- application areas like recommender systems, autocomplete suggestions and support for multilingual search are not seen as very relevant
- linked data is valued very positively as a future topic
- thesauri will support search engines in the near future to improve search results
- experience with controlled vocabularies is varying considerably among the branches
- there is a high awareness for standards like SKOS; the web-paradigm has entered the world of controlled vocabularies
- controlled vocabularies are no more locked in academic frames, they have also arrived in enterprise areas

Survey results

1) Do you or your organization already make use of controlled vocabularies like taxonomies or thesauri?

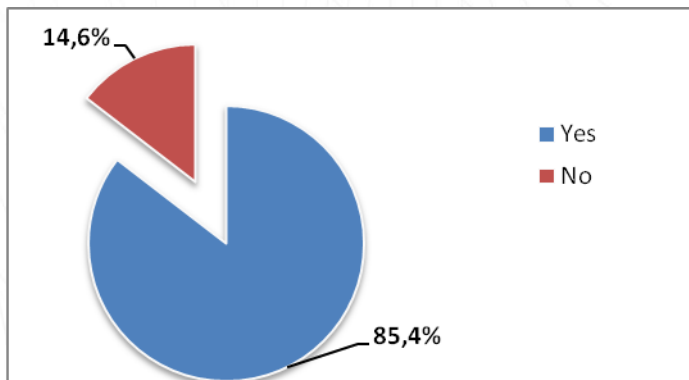


Figure 1: Use of controlled vocabularies

Value	Count	Percent
Yes	135	85,40%
No	23	14,60%
Total	158	

To make sure that the following questions would meet a qualified audience we needed to know if our participants are already using controlled vocabularies.

2) How long have you been using controlled vocabularies?

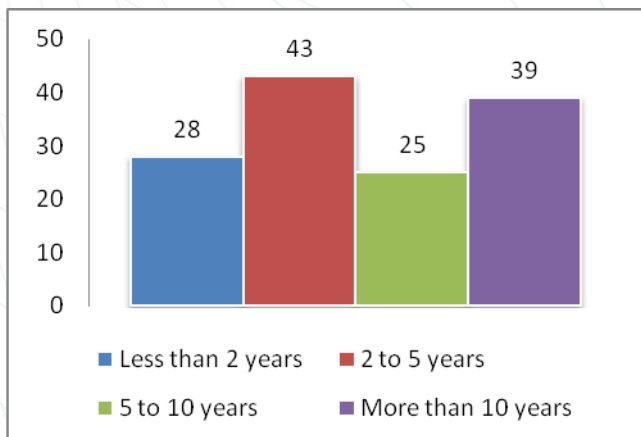


Figure 2: Time using controlled vocabularies

Value	Count	Percent
Less than 2 years	28	20.7%
2 to 5 years	43	31.9%
5 to 10 years	25	18.5%
More than 10 years	39	28.9%
Total	135	

Nearly half of the participants (47,4%) are using controlled vocabularies since at least 5 years. In average controlled vocabularies are used since 3.1 years. Even if we had an idea, we wanted to know if there is any interaction with the size of organizations as heavy users.

The results are what could be expected: parallel to the size of an organization the need of an advanced way of managing information is evident:

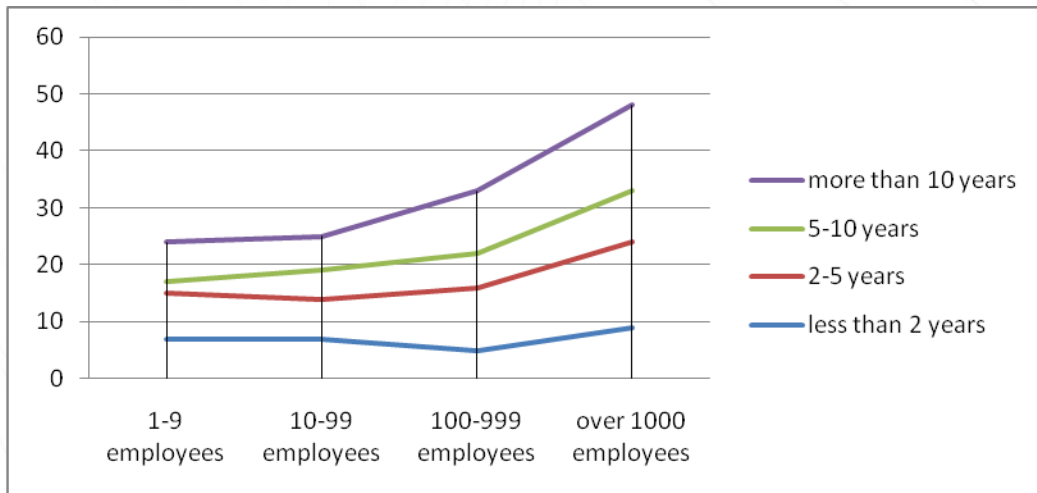


Figure 3: Organizations size and time of use (mentions)

Another question comes up within this context. **Is there a significant interaction between time of using controlled vocabularies and specific sectors?**

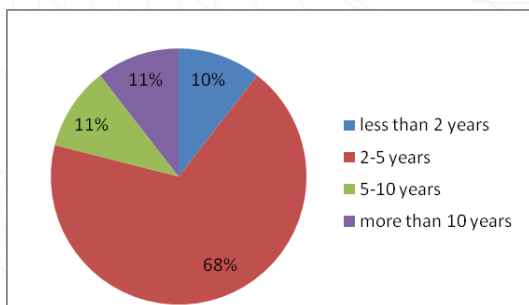


Figure 4: Science using controlled vocabularies

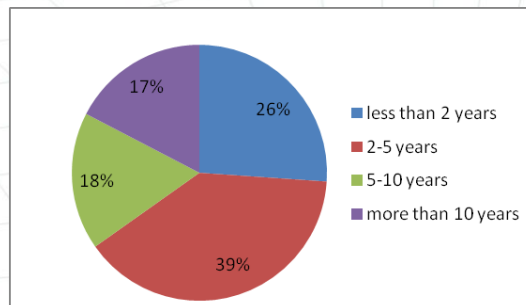


Figure 5: IT using controlled vocabularies

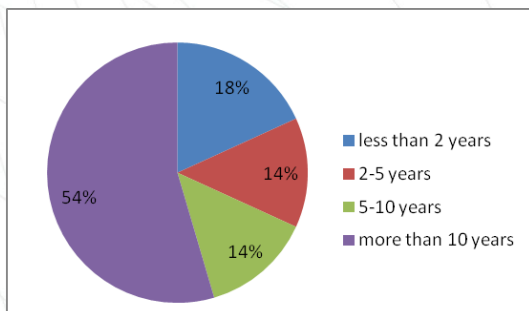
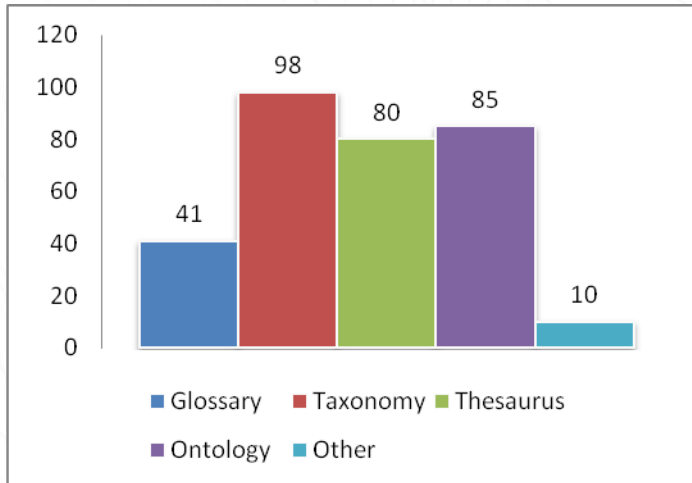


Figure 6: Public sector using controlled vocabularies

Our data reveals that both Science (68%) and Information Technologies (39%) have been using controlled vocabularies between 2-5 years. Whereas the Public sector definitely keeps the pole position with 54% mentions on using controlled vocabularies for more than 10 years.

3) Which kind of controlled vocabulary do you use or plan to use in your applications?



Value	Count	Percent
Glossary	41	30,4%
Taxonomy	98	72,6%
Thesaurus	80	59,3%
Ontology	85	63,0%
Other	10	7,4%
Total	135	

Figure 7: Preferred kind of controlled vocabularies

The figure shows that multiple answers were allowed. What seems to be really interesting is that organizations of all sizes are using / or planning to use controlled vocabularies. It is notable that glossaries are far behind the more complex models that offer more flexible vocabulary management.

Again there comes up the question concerning the relation between sector of a company and the preferred vocabulary-form:

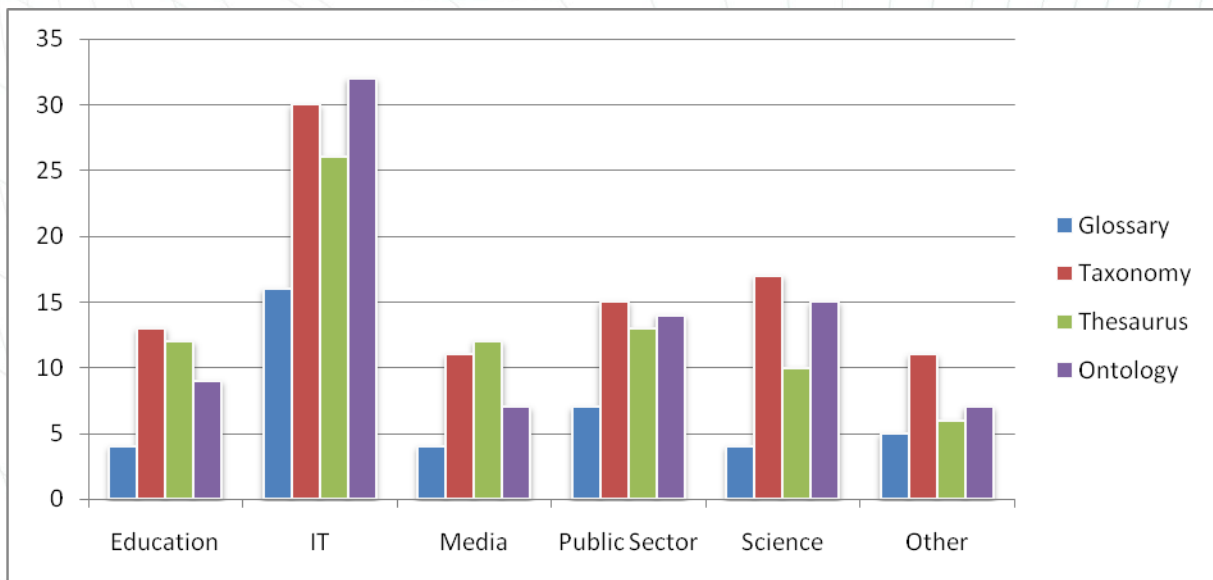


Figure 8: Organization's sector and preferred vocabulary models

The picture fits quite well to figure 7: Whereas taxonomies rule in education, public sector and science, ontologies are at the pole position in the IT sector. Thesauri rule at the media sector and are rated higher in the education sector than ontologies. Overall there are indicators for varying preferences in semantic expressivity of knowledge models.

Probably now you have the same question in your mind as we have: **What about the relation between the preferred vocabulary models and the size of an organization?**

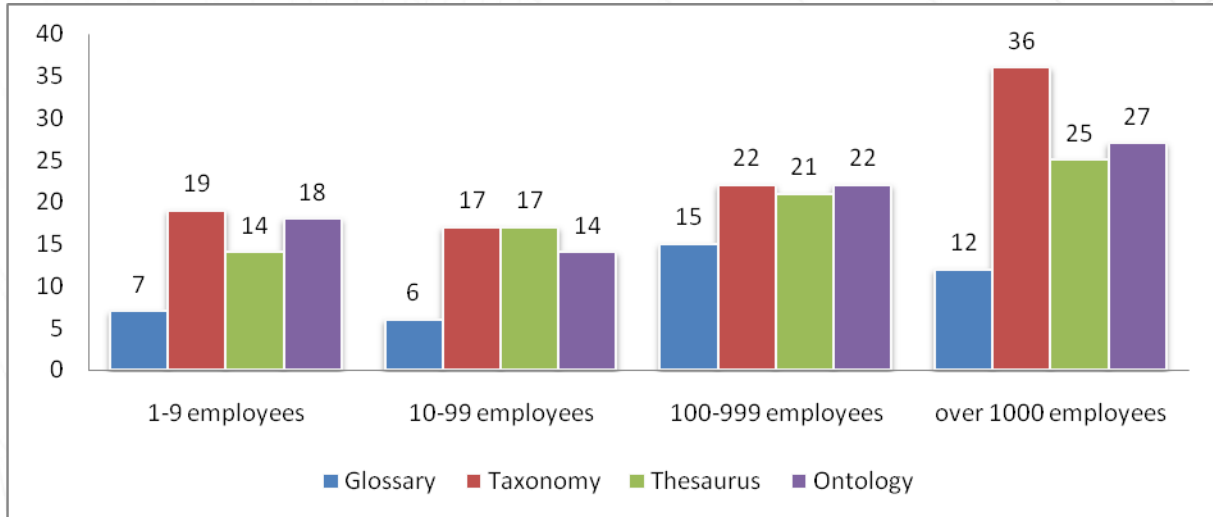


Figure 9: Organizations size related to preferred vocabulary models

Combined with the size of an organization it is notable, that glossaries and thesauri are more preferred by little and mid-sized organizations whereas taxonomies and ontologies are also in the focus of bigger organizations.

4) Which kind of controlled vocabulary do you plan to use in your applications?

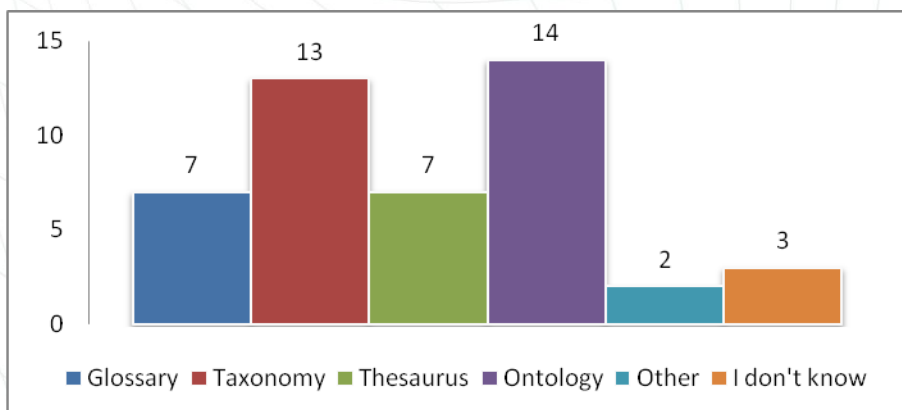


Figure 10: Controlled vocabularies for the future

When it comes to future plans, 46 participants told us what kind of controlled vocabulary models they will integrate in their environment. The two winners are Taxonomies and Ontologies.

5) What are the main application areas of controlled vocabularies from your perspective?

	Very important	Relevant	Somewhat relevant	Not relevant	I don't know	Total
Annotation and Tag Recommendation	36.6%	35.9%	20.3%	5.9%	1.3%	100%
	56	55	31	9	2	153
Semantic Search	52.6%	31.8%	12.3%	1.9%	1.3%	100%
	81	49	19	3	2	154
Data Integration	52.3%	32.0%	11.8%	2.0%	2.0%	100%
	80	49	18	3	3	153
Recommender Systems	16.3%	26.5%	33.3%	16.3%	7.5%	100%
	24	39	49	24	11	147
Support of multilingual applications	19.5%	27.3%	30.5%	18.8%	3.9%	100%
	30	42	47	29	6	154
Autocomplete suggestions	17.1%	38.2%	29.6%	12.5%	2.6%	100%
	26	58	45	19	4	152
(Linked) Open Data Publishing	38.4%	27.2%	17.9%	8.6%	7.9%	100%
	58	41	27	13	12	151
Content authoring and interlinking	31.1%	38.4%	18.5%	7.3%	4.6%	100%
	47	58	28	11	7	151
Structure for content navigation	41.2%	29.4%	22.2%	3.3%	3.9%	100%
	63	45	34	5	6	153

Table 1: Main application areas of controlled vocabularies

We wanted to know, what kind of application areas should especially be supported by controlled vocabularies. As the following graphs will show, the topics semantic search, data integration, structure for content navigation, annotation & tagging and Linked (Open) Data publishing are the main areas of interest.

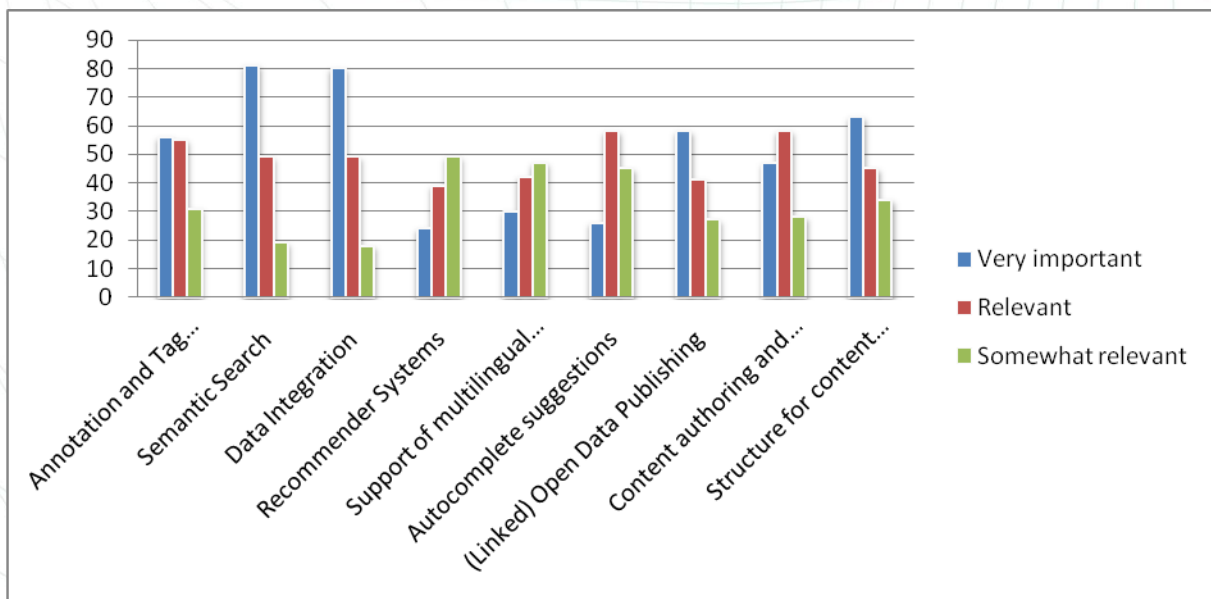


Figure 11: Main application areas of controlled vocabularies

On the other hand recommender systems, autocomplete suggestions and support for multi-lingual search are not found favored to be supported by controlled vocabularies.

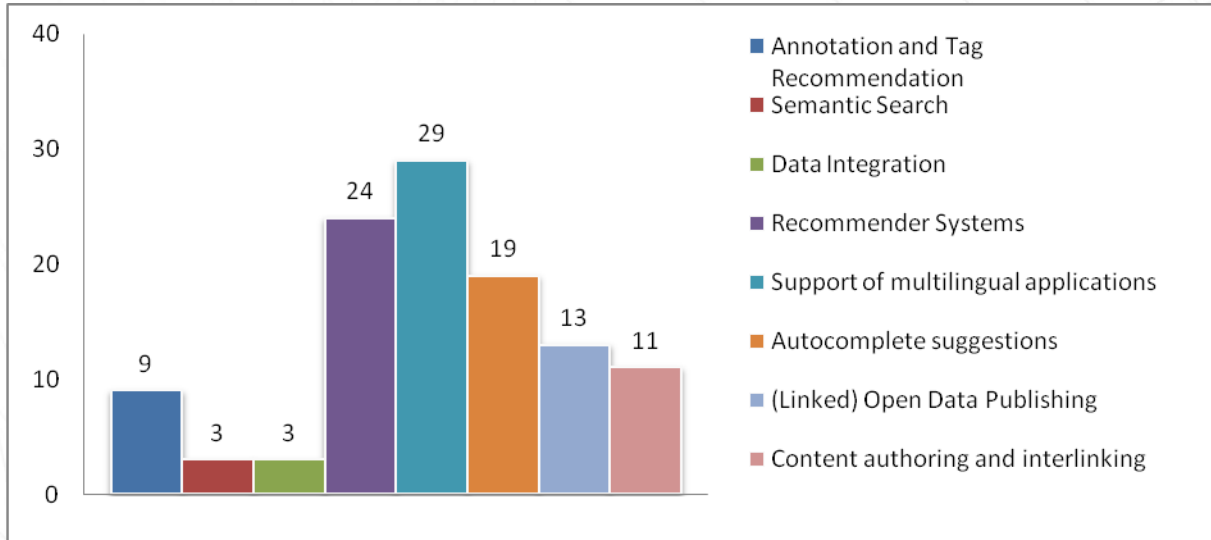


Figure 12: Not favored application areas for controlled vocabularies

We also asked about any other specific application software that should be extended to be used with controlled vocabularies. We decided to merge the nominations while the biggest part is to be found within “Others”. There we found answers like “the web itself”, “Light-room”, “Inhouse systems” and many others.

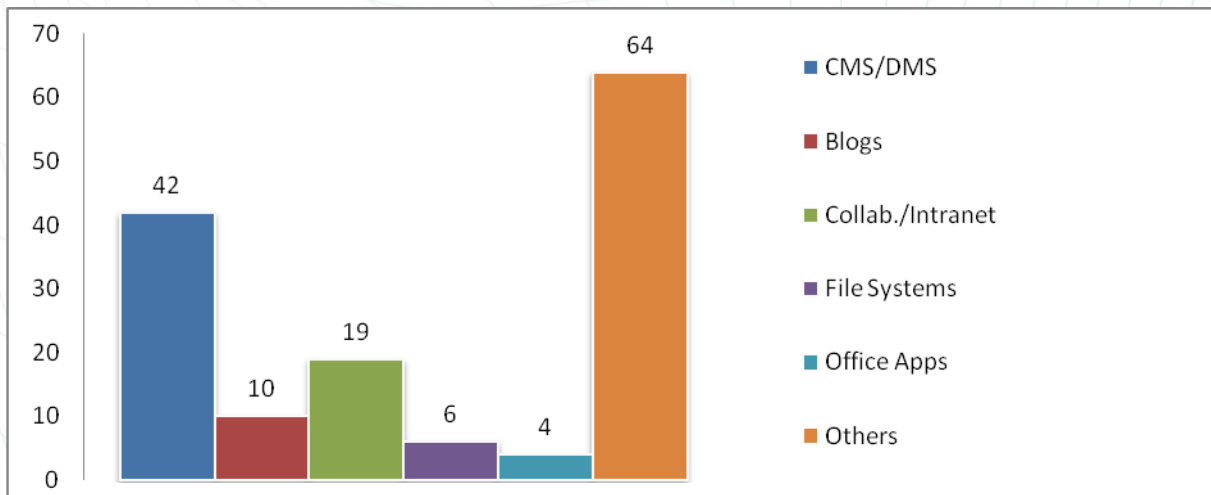


Figure 13: Applications to be extended with controlled vocabularies

6) How important is the usage of standards like SKOS for controlled vocabularies?

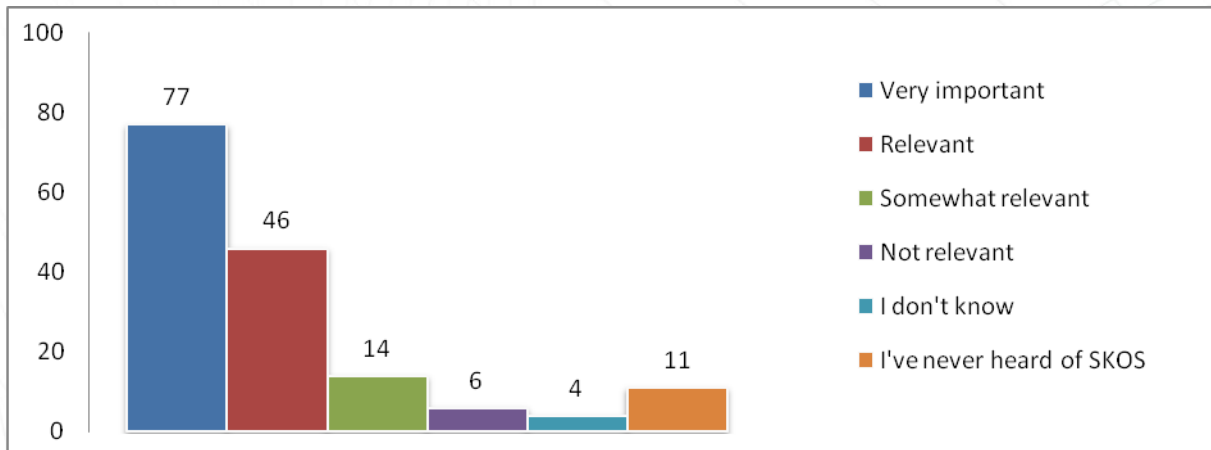


Figure 14: The use of SKOS for controlled vocabularies

In August 2009 the W3C consortium announced the new SKOS standard - developed by the [SWD WG](#) - for bridging between the world of knowledge organization systems and the linked data community. Now, nearly two years after, it looks like this standard has well arrived. 48.7% stated that standards like SKOS are very important and 29.1% voted for “relevant”.

We also asked for **other standards that are important for the participant’s daily work**. From 250 nominations we decided to focus on those that had a score of at least 7. The big three are OWL, RDF and Dublin Core. Others were a variety of ISO terminologies, RSS, SPIN. Remarkable in this context is that SPARQL made it just to 5 mentions.

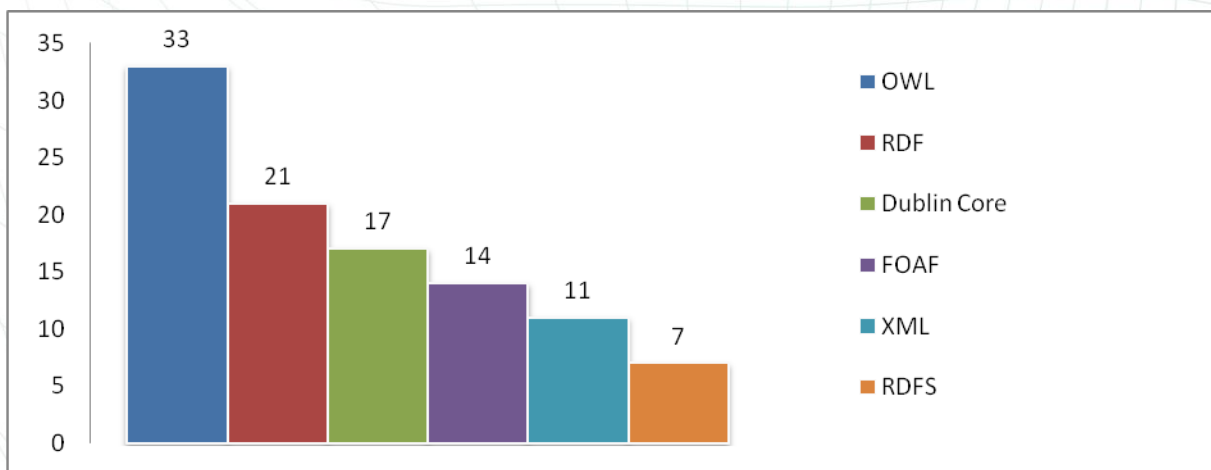


Figure 15: Other important standards

7) Do you think enterprises and other organizations can significantly benefit from using Linked Data?

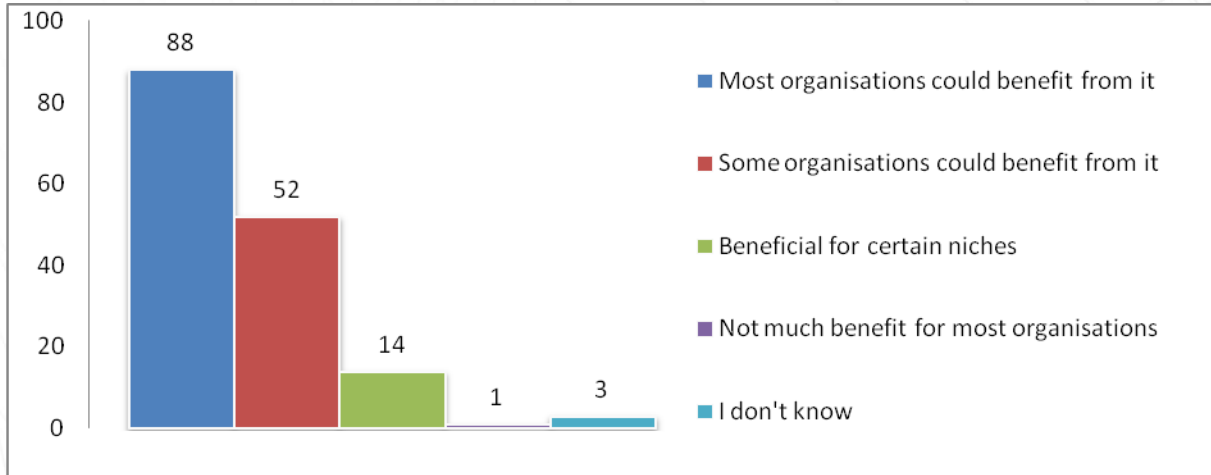


Figure 16: Can organizations benefit from using Linked Data

The results show clearly that Linked Data is seen as a useful approach for most organizations. But we want to mention here, that a big part of the participants of this survey are close to this topic, since we generated many participants from our contact-lists and also through being supported by community-platforms that are handling Linked Data topics.

Interesting at this point again, if organizations of different sizes are thinking the same way, and yes - the results show relatively constant figures.

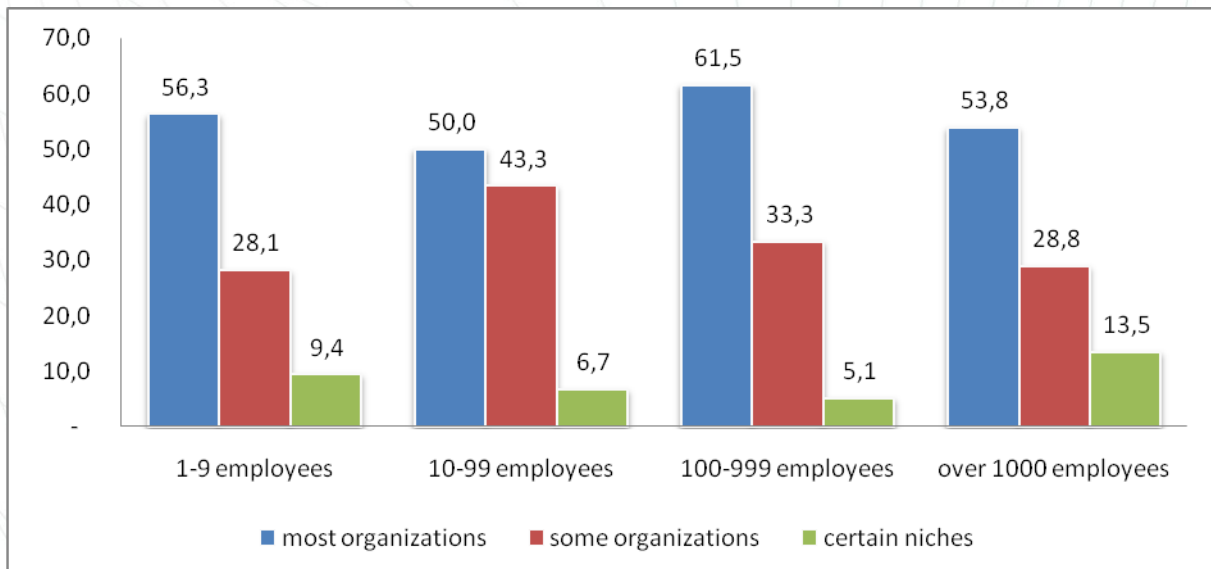


Figure 17: Use of Linked Data by organization size (in %)

8) Do you think that search engines which utilize thesauri to improve results will become main-stream?

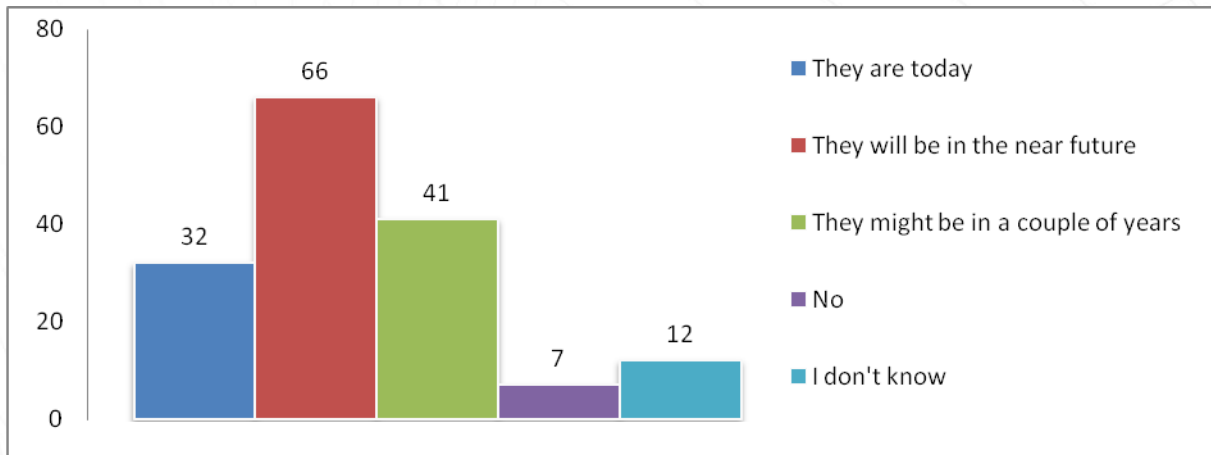


Figure 18: Will using thesauri to improve search engines become mainstream?

Yes, there was a reason why we were asking especially about thesauri, since this is our preferred model for controlled vocabularies. The really positive results look nice, even if we would say (while stating ourselves as experts in this domain), that the 32 answers on “They are today” is a bit too positive.

Let's see, if there are some special changes, when we combine this question concerning the size of an organization:

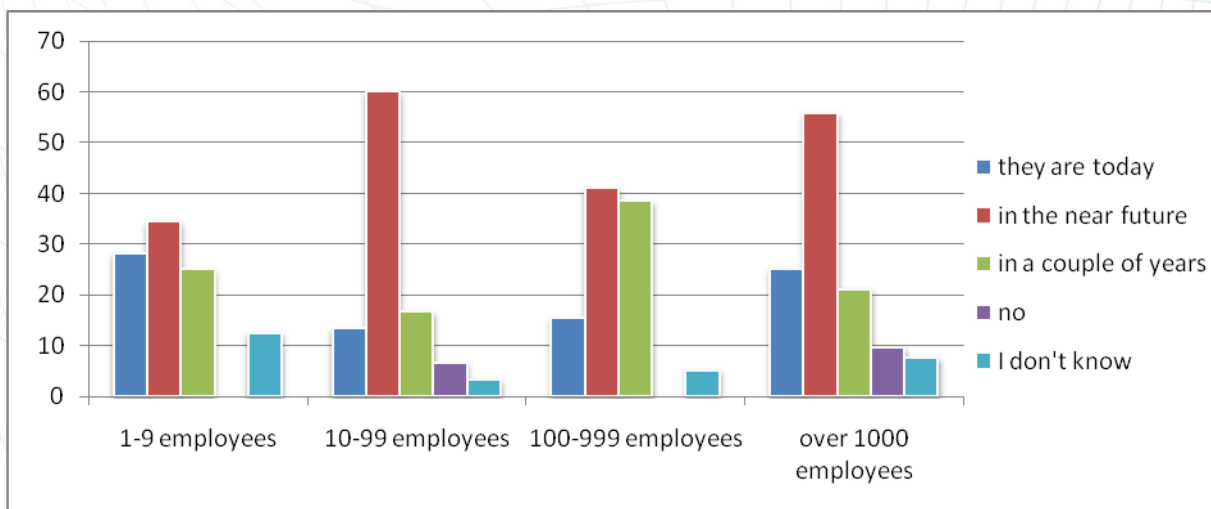


Figure 19: Will using thesauri to improve search engines become mainstream? By size (in %)

Small and big organizations have very positive expectations towards the status quo. Whereas both the second and the fourth group are very confident that thesauri will have a big time in the area of search improvements soon.

Results regarding PoolParty

9) Have you heard about the PoolParty Thesaurus Management system before?

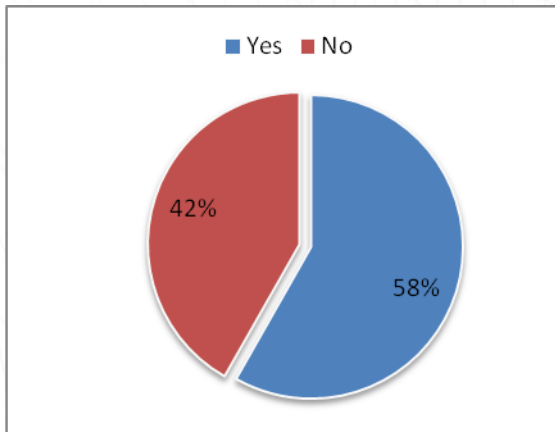


Figure 20: Heard about PoolParty before?

Yes, we are glad about this result consisting of 158 answers. Again, we should be aware that many of our participants are coming from our contact lists.

Besides PoolParty there are some more tools for managing thesauri or taxonomies. The following ones are better known by the participants (from 174 answers in total):

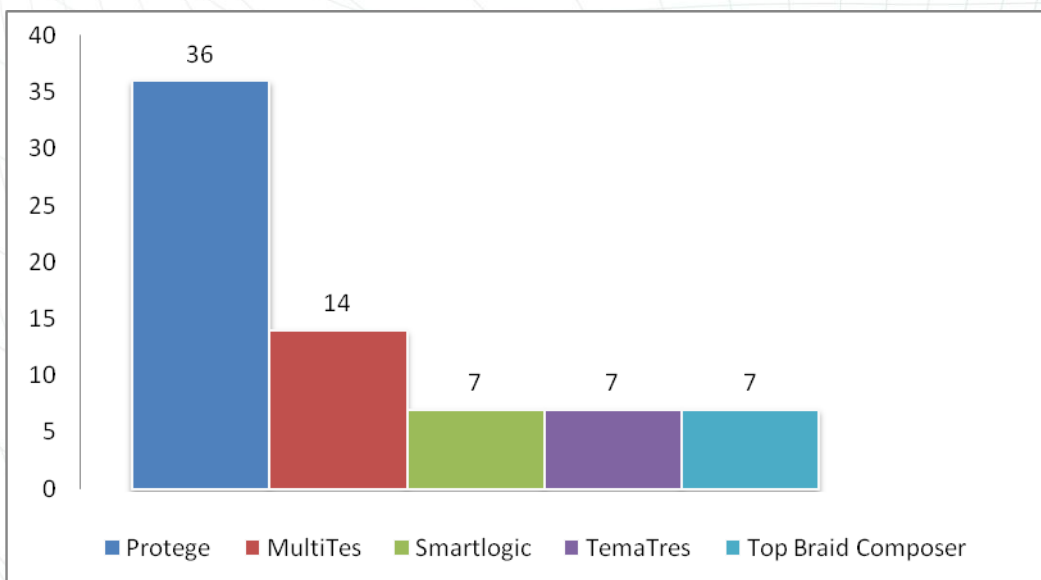


Figure 21: Known tools for managing thesauri and taxonomies

10) In which areas do you think PoolParty Thesaurus Management can be useful for you?

As we provide all those solutions with PoolParty mentioned as answering options in this question, we wanted to know if those opportunities are also known by our survey participants.

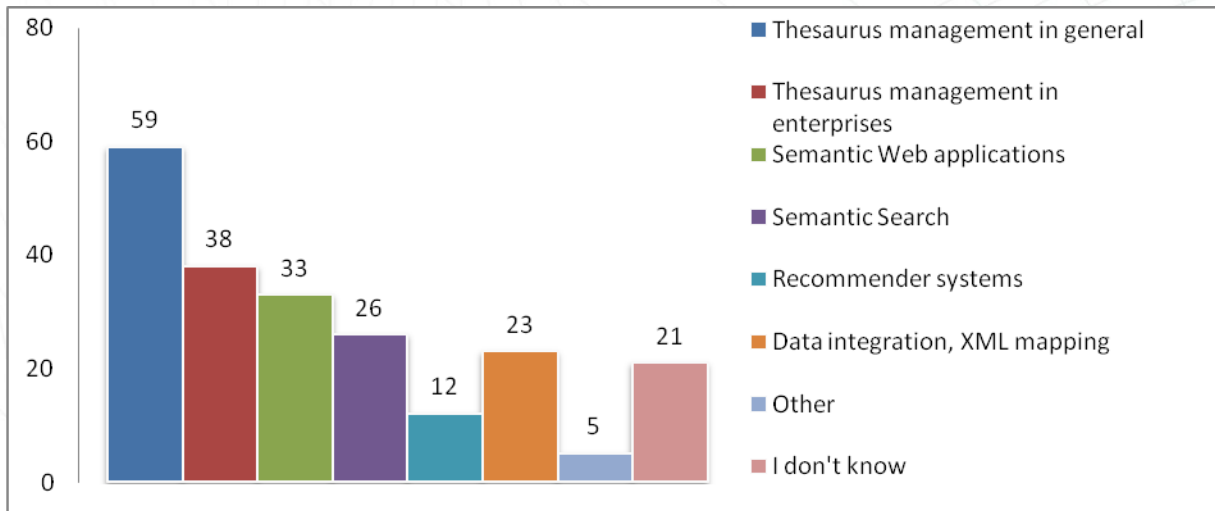


Figure 22: Areas that could be supported by PoolParty

Facts about participants

11) In which of the following sectors is your organization primarily active?

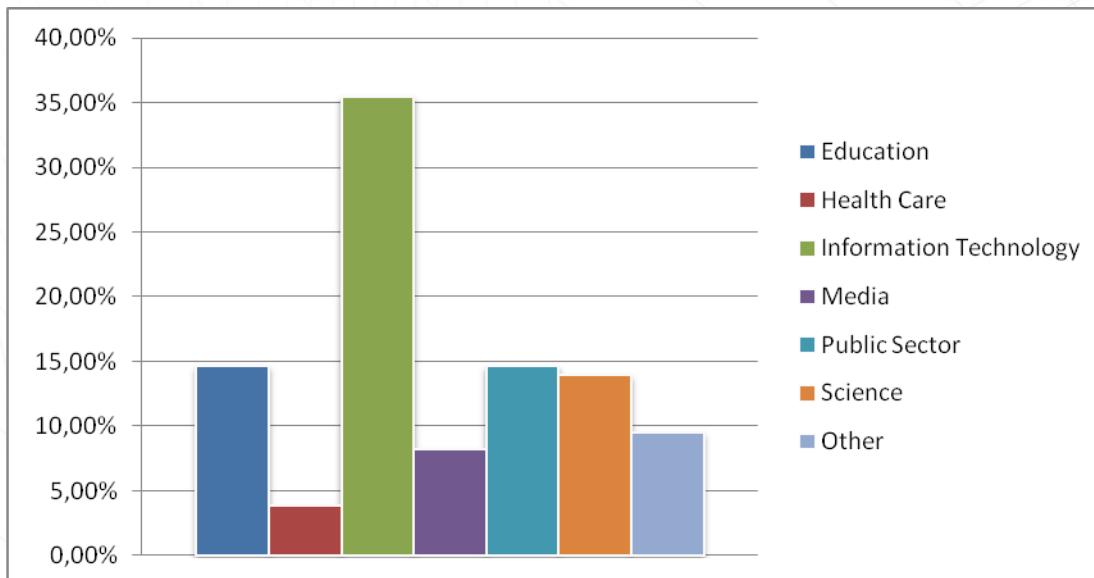


Figure 23: Participants sector

12) In which area do you work in your organization?

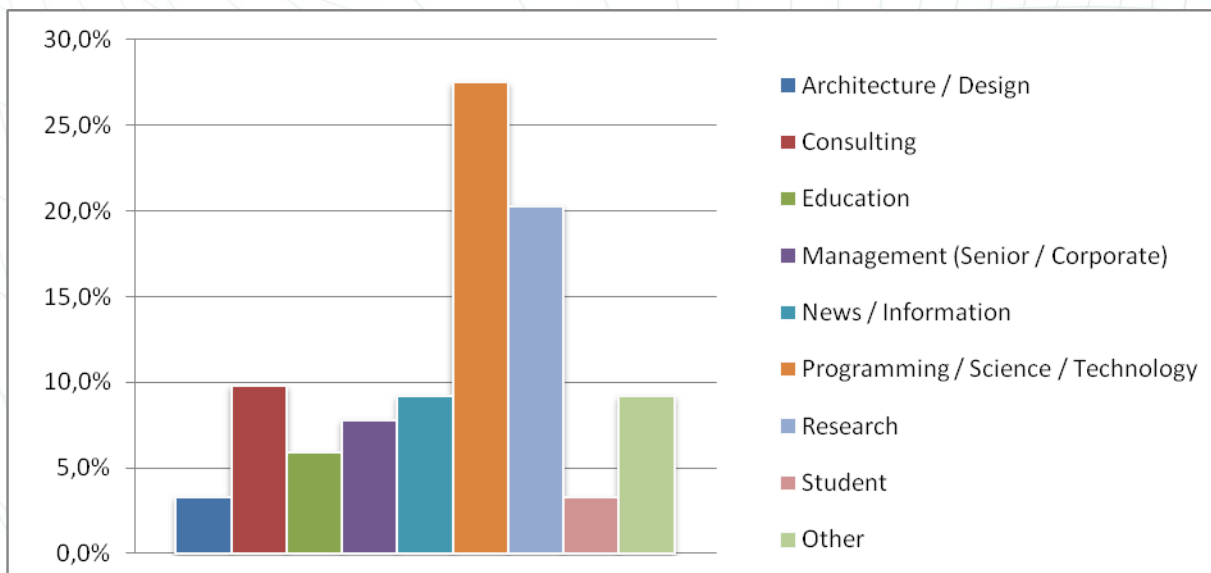


Figure 24: Participants working area

13) What size is the organization you work for?

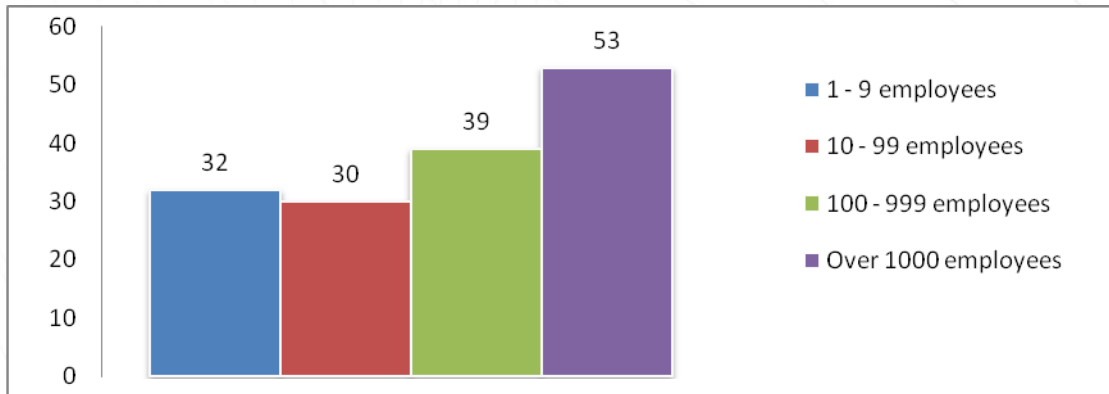


Figure 25: Organization size

Conclusio & words of thanks

We would like to thank all participants of this survey who build the basis for these results. We hope to contribute with this aggregated information in a way which makes the benefit of controlled vocabularies in today's enterprise information management more transparent and obvious. Not only providers of tools but especially (potential) users of taxonomies, thesauri, ontologies or whatever kind of knowledge model they have voted for should have gained better insight into potential use cases and the value of using standards like SKOS.

For more information about this survey or about controlled vocabularies please contact us:

office@semantic-web.at or info@poolparty.biz