http://latc-project.eu

D3.2 Evaluation Report of P&C Library

Project GA No.	FP7-256975
Project acronym	LATC
Start date of project	2010-09-01
Document due date	2011-31-08
Actual date of delivery	2011-31-08
Lead Partner	INFAI
Reply to	Michael Martin, martin@informatik.uni-leipzig.de
Document status	FINAL

Project GA No.	FP7-256975
Project acronym	LATC
Project full title	Linking Open Data Around The Clock
Dissemination level	PU
Number of Pages	43
Task Responsible	INFAI
Other contributors	none
Authors	Michael Martin, Philipp Frischmuth, Jens Lehmann
EC Project Officer	Stefano Bertolo
Keywords	survey, tool library, user evaluation

Contents

1. Executive Summary **4**

2. Introduction and Goals **4**

3. Overview of the Publication & Consumption Tool Library **5**

4. Structure of the Survey **6**
4.1. Survey Section 1: Questions About You and Your Use Cases 6
4.2. Survey Section 2: Tool Evaluation 7
4.3. Survey Section 3: Other tools 10

5. Execution of the Survey **11**

6. Survey Analysis and Discussion **12**
6.1. Survey Section 1 12
6.2. Survey Section 2 14
6.2.1. Neologism 14
6.2.2. RDF Extension for Google Refine 14
6.2.3. Pubby 14
6.2.4. LDSpider 14
6.2.5. Sindice 14
6.2.6. any23 14
6.2.7. SILK Link Discovery Framework 15
6.3. Survey Section 3 15

7. Possible Extensions of the P&C Tool Library **16**

A. Survey Result Export: single choice answers **17**

1. Executive Summary

This deliverable is part of Work Package 3 of the LATC project, which involves the creation of a Linked Data publication and consumption tool library. In a first phase of this work package, an initial library was created. Herein, we conducted a survey on this library in order to improve the library and the presented tools further. We describe which questions were asked in the survey and why, describe the execution of the survey and present the obtained results.

2. Introduction and Goals

The LATC tool library serves to purposes:

- support data set owners in publishing their data on the web
- support data consumers in using published data

The library aims at covering all stages of the publication and consumption process.

In a first stage, in LATC Deliverable 3.1.1, an initial version of the library was created. The aim of Task 3.2, which is presented in this deliverable, is to:

- get feedback from target users (consumers and publishers) about the tool library, for instance what kinds of tools are missing, what parts of the consumption and publication process may not be covered and other weaknesses of the library
- collect information from target users about the tools themselves, i.e. regarding their compatibility, ease of configuration and deployment, standard conformance, completeness of documentation etc.

Ultimately, the feedback obtained in Task 3.2 feeds into Task 3.3 – the final version of the LATC tools & consumption library.

To achieve these goals, we created a survey containing a series of questions, which was then send to the LATC advisory committee[1]. The core of this deliverable describes the structure, execution and results of this survey.

Overall, the deliverable is structured as follows: Section 3 briefly describes the existing publication & consumption tool library. In the following section, we present the structure of the survey and how it is aligned to the above goals. Section 5 describes how we executed the survey. Finally, the main results of the survey will be presented and discussed in Section 6.

[1] `http://latc-project.eu/about/ac`

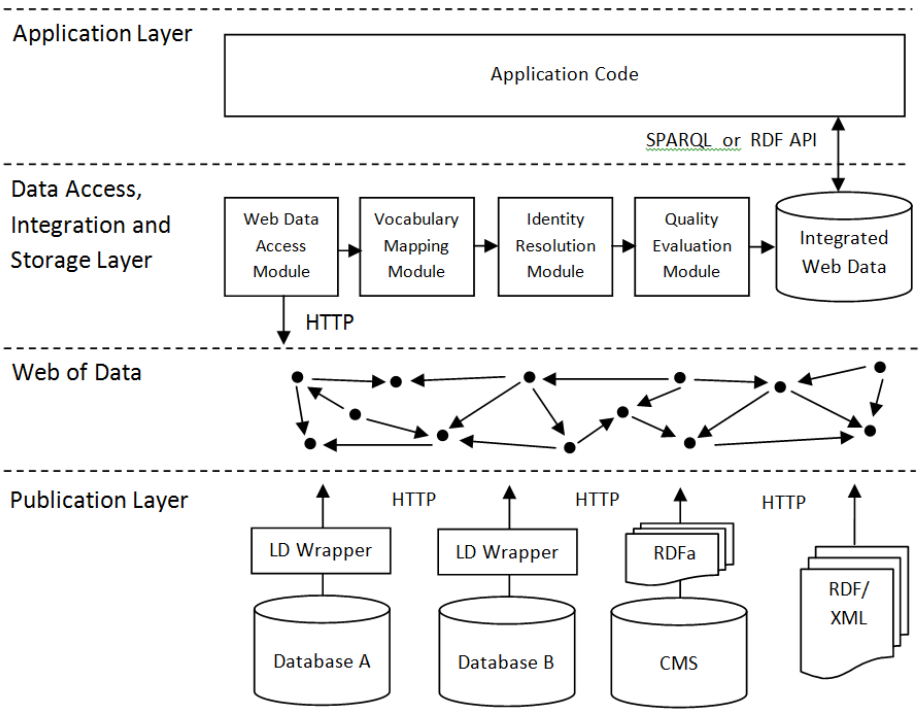


Figure 1: Schematic Architecture of Linked Data Applications

3. Overview of the Publication & Consumption Tool Library

The initial version of the library is published online at `http://www4.wiwiss.fu-berlin.de/latc/toollibrary/`. The described tools were grouped into different layers as shown in Figure 1. An explanation of this figure, in particular the different modules in the data access, integration and storage layer, can be found in Deliverable 3.1.1. The tools library uses this schema to organise their components. One aim of the library is to give users guidance for (open source) tools for each of the layers. This helps users not to get lost in the vast amount of available, but not necessarily well-documented and high quality, tools and scripts on the web.

Overall, the library contains the following tools (category in brackets):

- Neologism (Modelling)
- D2R Server (Publishing)
- RDF Extension for Google Refine (Publishing)
- Pubby (Publishing)
- Any23 (Data Conversion)

- LDSpider (Discovery / Crawling)
- Sindice (Crawling / Searching)
- R2R Framework (Vocabulary Mapping)
- Silk Link Discovery Framework (Identity Resolution)
- Sig.ma (Application / Exploration)
- RelFinder (Application / Exploration)
- OntoWiki (Application / Exploration)
- SPARQL Views (Application / Exploration)

In our result discussion in Section 6, we will often mention those tools, but we refer to the tool library itself for their description.

4. Structure of the Survey

Regarding the goals of the survey we, divided it into the following three parts:

- Survey section "**Questions about you and your use cases**" was created to get basic information about the survey participant.
- Survey section "**Tool Evaluation**" is the core of the survey and allowed to answer questions for one up to three tools from the tool library. The results of this section will be used to improve the presentation of each of the tools in the tool library as well as serve as feedback for tool maintainers.
- Survey section "**Other tools**" was needed to obtain information about tools that are not yet listed in the tool library. Such information can be used to increase the coverage of the tool library by including popular software.

In the following, we present the structure and question of the survey more in detail.

4.1. Survey Section 1: Questions About You and Your Use Cases

This survey section aims to evaluate the experience level of the survey participant.

Question 1.1 The first question the participant has to answer was about her actual field of work, the use cases and application domains of Linked Data. It is a free text question without special options.

Listing 1: Question 1.1

```
Please give us a short summary about the use cases you work
on in which the tool library is relevant for you?
```

Question 1.2 This question is about the application domain in which the participant works at the moment. Its purpose is to related tools to use cases. It is a free text question without special options.

Listing 2: Question 1.2

```
What topic/domain are you applying Linked Data in?
```

Question 1.3 This question tries to figure out why linked data technologies are needed in use cases of the participant. Answers to it may help to increase the coverage of the workflows described in the tool library. It is a free text question without special options.

Listing 3: Question 1.3

```
Why do you use or plan to use Linked Data?
```

Question 1.4 We contined by letting the participant judge whether she is an expert in the field or a novice. Naturally, judging experience is difficult. Herein, we mainly used the experience of the participant as criterion. We are aware that this is not the only criterion upon which expertise of person in a field can be determined, but we refrained from making the survey more complex than necessary.

Listing 4: Question 1.4

```
Please indicate your level of experience in the field of
Semantic Web / Linked Data?
```

This question is of type single-choice and has the following options.

- Option 1 **No Experience**
- Option 2 **I occasionally use Semantic Web technologies**
- Option 3 **Up to three years of professional experience (academic or industry)**
- Option 4 **3+ years of professional experience**

4.2. Survey Section 2: Tool Evaluation

The second section of the survey was made to receive information about the listed tools in the tool library. Participants had the option to select up to three tools from the library (minimal one tool) and to provide some information/impressions/insights about each selected tool.

The questions listed in the following were those asked for each single tool. After answering questions about a tool, the survey provided the option to continue with another tool from the library.

Listing 5: Question 2.1

```
Name of the first selected Tool
```

This question is of type single-choice and has the following options.

- Option 1 **Neologism**
- Option 2 **D2R Server**
- Option 3 **RDF Extension for Google Refine**
- Option 4 **Pubby**
- Option 5 **Any23**
- Option 6 **LDSpider**
- Option 7 **Sindice**
- Option 8 **R2R Framework**
- Option 9 **Silk Link Discovery Framework**
- Option 10 **Sig.ma**
- Option 11 **RelFinder**
- Option 12 **OntoWiki**
- Option 13 **SPARQL Views**

Question 2.2 It is a free text question without special options.

Listing 6: Question 2.2

```
Please summarize the specific task you used (or want to use)
the tool for. We would appreciate if you tell us something
about your (intended) usage of the tool. Which input/output
formats do you expect and what kind of input sources/output
targets do you need?
```

Question 2.3 It is a free text question without special options.

Listing 7: Question 2.3

```
Did you use the tool's online documentation? If so, did you
find it clear or confusing? Any parts that you found
particularly lacking? Did you use any other information
sources?
```

Listing 8: Question 2.4

```
Is the tool easy to configure and deploy
(only applicable if deployment is needed)?
```

Question 2.4 is of type free-text and has no special options.

Listing 9: Question 2.5

```
Did you install and run the tool?
```

Question 2.5 is of type single-choice and has the following options:

- Option 1 **yes**
- Option 2 **tried but did not get it working (within the time I had)**
- Option 3 **no**

Listing 10: Question 2.6

```
Do you consider the tool easy to use?
```

Question 2.6 is of type single-choice and has the following options:

- Option 1 **very easy to use**
- Option 2 **easy to use**
- Option 3 **moderately easy to use**
- Option 4 **difficult to use**
- Option 5 **has several usability issues**

Question 2.7 Participants have the possibility to give information about possible usability problems if exists when answering question 2.7. It is a free text question without special options.

Listing 11: Question 2.7

```
If you found usability issues, can you name specific
problems?
```

Listing 12: Question 2.8

```
How do you rate the standards compliance and compatibility
of the tool?
```

Question 2.8 is of type single-choice and have the following options:

- Option 1 **strictly uses and adheres to standards (e.g. W3C standards)**
- Option 2 **derivations from standards, but you consider those derivations necessary (please explain)**
- Option 3 **does not adhere to standards**

Question 2.9 Question 2.9 has a wide scope and allows the participants to provide additional thoughts on the selected tool. This way, we allowed them to give more information without being restricted to the particular previously covered aspects of a tool. It is a free text question without special options.

Listing 13: Question 2.9

```
Please feel free to add more comments to this tool,
especially in regard to one ore more of the following
questions:
1. What features were lacking from the tool, or would
   have been valuable in your project?
2. Did you use the tool in combination with other
   off-the-shelf tools?
3. If so, which?
4. Did you find it easy to integrate this tool into
   your workflow or overall system?
```

4.3. Survey Section 3: Other tools

One of the goals of the survey was to find out whether we cover the most important aspects of Linked Data publication and consumption in the LATC tool library and have indeed included popular tools. Therefore, we provided an additional free text question in the third part of the survey to suggest additional tools, which are in use for Linked Data publication and consumption. This information will feed into the next library release in LATC Task 3.3.

Question 3.1 It is a free text question without special options.

Listing 14: Question 3.1

```
What other tools do you often use in your Linked Data
or Semantic Web projects?
```

5. Execution of the Survey

In this section, we briefly describe the evaluation of existing tools suitable to execute the developed survey. Furthermore, we present overall statistics about received responses.

Survey tool selection Before executing the survey, described in Section 4, we made a short evaluation of suitable survey tools. In particular, we wanted to execute the survey using a tool with the following characteristics:

- It must be executable in a Web environment.
- It must be deployable on own servers (and reachable via a LATC specific domain).
- It should be Free- and/or OpenSource Software to minimise costs.
- It must be adaptable to LATC style sheets.
- The designed survey itself should of course be representable in the tool.
- The survey result statistics should be visualizable (tables, graphs).
- The survey results shoud be exportabe in suitable formats.

As a result of the evaluation, we decided to use *LimeSurvey*[2] which fits to all of the listed criterias. We installed the used instance[3] of LimeSurvey on an AKSW server.

Overall Statistics The survey is available for execution since the 19. of July 2011 and is still active (until end of September 2011). At the moment there are 19 responses (only 6 of them are completed). We disseminated the LATC survey in the LATC advisory committee and the Linked Data mailing list. We did expect to get more feedback (completed surveys), in particular from the advisory committee.

The following Table 1 visualizes how much time participants used to complete each survey section. The statistics only contain information about complete responses. As part of these 6 survey responses we get 11 responses about tools (3 participants responded about 1 tool, 1 participant responded about 2 tools and 2 participants responded about 3 tools). Which tools and how often thy were selected and commented is listed the following:

[2] `http://limesurvey.org/`
[3] `http://latc.aksw.org/survey/`

Table 1: Time statistics about completed responses (m = minutes, s=seconds).

Overall	Sec 1	Sec 2: Tool 1	Sec 2: Tool 2	Sec 2: Tool 3	Sec 3: Other
12 m 26 s	1 m 23 s	1 m 39 s	2 m 49 s	2 m 30 s	4 m 02 s
25 m 24 s	10 m 47 s	11 m 04 s	44 s	9 s	2 m 39 s
31 m 58 s	2 m 17 s	8 m 47 s	3 m 37 s	15 m 21 s	1 m 53 s
9 m 02 s	5 m 26 s	1 m 37 s	10 s	5 s	1 m 42 s
46 m 16 s	2 m 41 s	4 m 43 s	31 m 48 s	1 m 6 s	5 m 55 s
4 m 18 s	2 m 04 s	1 m 20 s	12 s	9 s	30 s

- A1 **Neologism** (1x)
- A3 **RDF Extension for Google Refine** (1x)
- A4 **Pubby** (1x)
- A5 **LDSpider** (2x)
- A6 **Sindice** (2x)
- A7 **any23** (2x)
- A9 **SILK Link Discovery Framework** (2x)

6. Survey Analysis and Discussion

In this section, we present the analysis created on the basis of the received results. Raw data responses are listed in the appendix for future reference.

6.1. Survey Section 1

We first give an overview about the answers we received for survey section about general participant information.

Question 1.1: Relevance of the Tool Library for Use Cases The following list contains the use cases which are relevant for the tool library (we extracted this from the free text responses):

- Data-Integration
- Data-Publication as RDF
- Creating software platform capable of aggregating, preserving, managing and disseminating blogs
- End to End Applications building in many domains

- Consulting about publishing consuming Linked Data
- Large scale linked open data processing

Question 1.2: Topic/Domain for applying Linked Data The participants work in the following listed topics / domains (again, extracted from free text responses):

- Cross-domain
- Environment-statistics
- Blogosphere
- Research information
- Cultural heritage
- Libraries
- Government data
- Media
- Law
- Corporate intranets

Question 1.3: Usage of Linked Data Reasons for using Linked Data, benefits and efforts of the participants are listed below (extracted from free text):

- Eliminate manual labour in downloading, converting and joining datasets
- Enhancing existing offered services
- Decrease costs and risks better than conventional technologies (e.g. integration problems)
- Integrate heterogeneous data from multiple sources

Question 1.4: Experience Level As visualized in Appendix A most of the participants have more than 3 years of professional experience in the field of the Semantic Web and / or Linked Data.

6.2. Survey Section 2

Overall, the received answers show that the most selected tools are deployable, easy to use, strictly adhere standards and are compatible with other tools at least for people with a high experience level. This is the main result of the survey as it indicates that a reasonable selection was made for the tool library. Naturally, more responses to the survey would allow to be more confident on this outcome.

In the following subsection we will present some of the free-text hints we received regarding each of the selectable tools.

6.2.1. Neologism

This tool is being used to quickly generate vocabularies. It is intuitive, but the documentation could be enhanced – at least for inexperienced users.

6.2.2. RDF Extension for Google Refine

This tool was used to investigate the possibility of analysing the output of files generated by LDSpider. It is easy to use and the documentation is very clear.

6.2.3. Pubby

Because that tool is providing dereferencable URIs, received from underlying data stores, it was part of the deployment environment of participants. The documentation is short but useful.

6.2.4. LDSpider

This tool was used in an experimental environment in order to use links to FOAF files, Twitter and blog resources as seeds in order to gather data from the data web (with limited success). The documentation seems to contain inconsistencies. It would be useful to have more specific and diverse case studies as part of the documentation. To acquire data, information about methods and maybe also further tools could be helpful.

6.2.5. Sindice

This tool is used for searching Linked Data and/or to feed into sameas.org for creating projections between natural language and URIs. For basic usages, the interface was intuitve enough. The documentation about Sindice was used and clear enough even for advanced use.

6.2.6. any23

This tool was used to develop a spider and to provide parsing of some MarkUps. The documentation was described as clearly understandable.

6.2.7. SILK Link Discovery Framework

This tool was used to create non-trivial links between datasets. It was also used to match a list of species with DBpedia. SILK is easy to deploy but difficult to configure. The documentation was helpful – at least for easy tasks – but sometimes confusing and could be enhanced with more details and examples. Specifically, the correct syntax for functions used to transform and compare could be enhanced / integrated. Also an explanatory tutorial would be helpful for one participant. The workbench web application was not intuitive and involved a lot of trial and error to learn its correct usage. Furthermore, more logging and output functionality seems to be helpful. The task of debugging link specs – especially for beginners – is perceived as challenging. The execution time is sometimes very long and users do not receive useful messages about the reason for negative interlinking results.

6.3. Survey Section 3

In this survey section, participants were able to give us hints about semantic web tools / classes of tools which give benefits in their use cases. Obviously, each of these tools will be checked carefully whether it is mature, popular and useful enough to be included in the next tool library release.

- Vapour
- Semantic web client library semradar
- Elda
- Tabulator
- Rapper (`http://librdf.org/raptor/rapper.html`)
- ARC2 (`http://prefix.cchttps://github.com/semsol/arc2/`)
- Puelia-php (`http://code.google.com/p/puelia-php/`)
- Virtuoso
- Sameas.org
- Rkbexplorer.com
- Libraries for rdf/sparql (jean, rdflib, sparqlwrapper ...)
- Libraries for parsing SPARQL XML-results

Furthermore, we have to check if it is possible to provide some more information on existing tools as part of the tool library. For instance, participants mentioned the performance of tools as a critical criterion, e.g. one participant was interested in fast parsers or APIs for Linked Data.

7. Possible Extensions of the P&C Tool Library

As a result of Section 6, we recommend to extend the existing state of the P&C Tool Library as follows.

- **Neologism**
  1. Check & Enhancement of the documentation for inexperienced users
- **LDSpider**
  1. Updating the documentation
  2. Adding tutorials and method recommendations to the documentation
- **LDSpider**
  1. Updating the documentation
  2. Adding tutorials and method recommendations to the documentation
- **SILK Link Discovery Framework**
  1. Enhancing documentation to address configuration possibilities (logging and debugging options)
  2. Enhancing documentation with more details and examples(e.g. syntax of functions) to better support complex tasks (also for the workbench)
  3. Adding an example work flow (tutorial) to the documentation
  4. Enhancing the logging/output functionalities of SILK
- **New tools:** We recommend to carefully examine the suggetsions in Section 6.3 for inclusion the tool library.

A. Survey Result Export: single choice answers

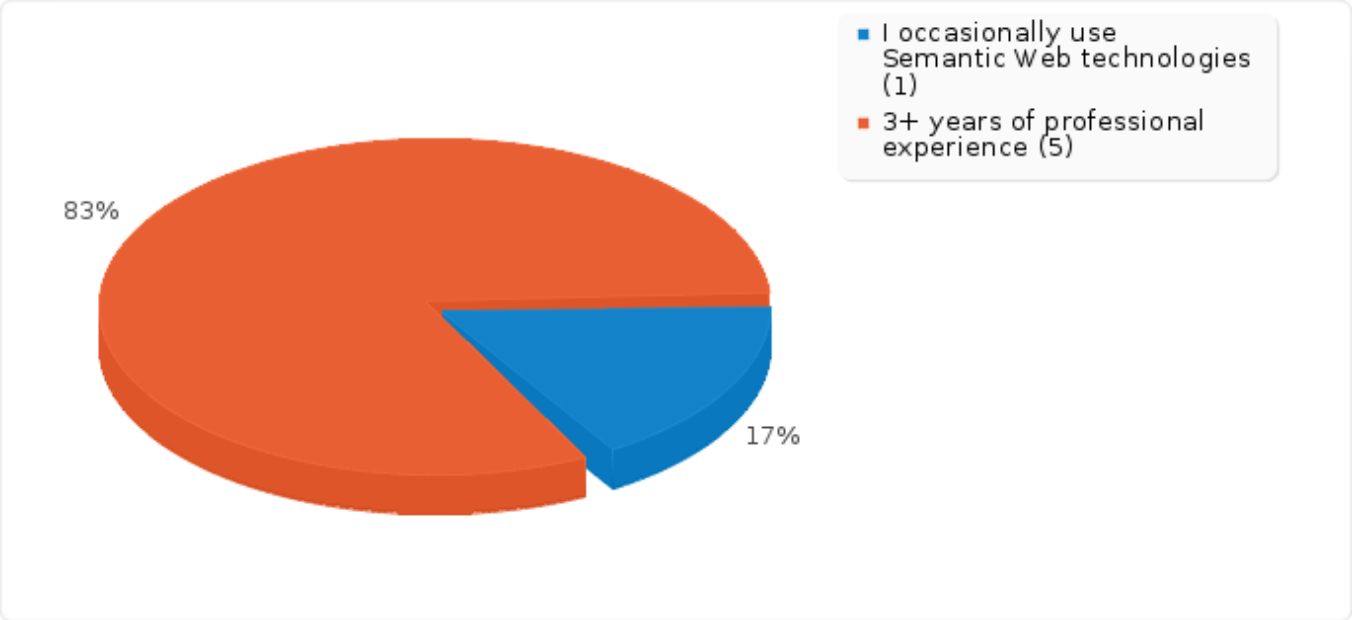
Field summary for 1.4

Please indicate your level of experience in the field of Semantic Web / Linked Data?

Answer	Count	Percentage
No experience (A1)	0	0.00%
I occasionally use Semantic Web technologies (A2)	1	16.67%
Up to 3 years of professional experience (academic or industry) (A3)	0	0.00%
3+ years of professional experience (A4)	5	83.33%
No answer	0	0.00%

Field summary for 1.4

Please indicate your level of experience in the field of Semantic Web / Linked Data?



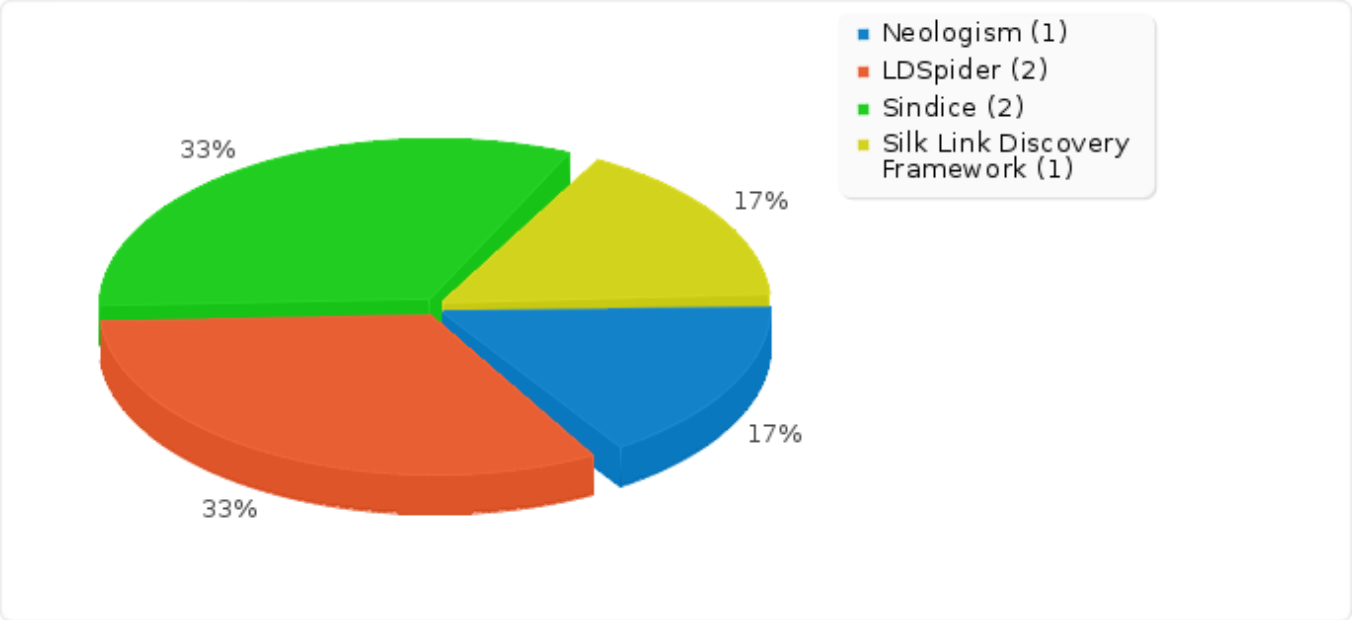
Field summary for 3.1.0

Name of the first selected Tool

Answer	Count	Percentage
Neologism (A1)	1	16.67%
D2R Server (A2)	0	0.00%
RDF Extension for Google Refine (A3)	0	0.00%
Pubby (A4)	0	0.00%
LDSpider (A5)	2	33.33%
Sindice (A6)	2	33.33%
any23 (A7)	0	0.00%
R2R Framework (A8)	0	0.00%
Silk Link Discovery Framework (A9)	1	16.67%
Sig.ma (10)	0	0.00%
RelFinder (11)	0	0.00%
OntoWiki (12)	0	0.00%
SPARQL Views (13)	0	0.00%
No answer	0	0.00%

Field summary for 3.1.0

Name of the first selected Tool



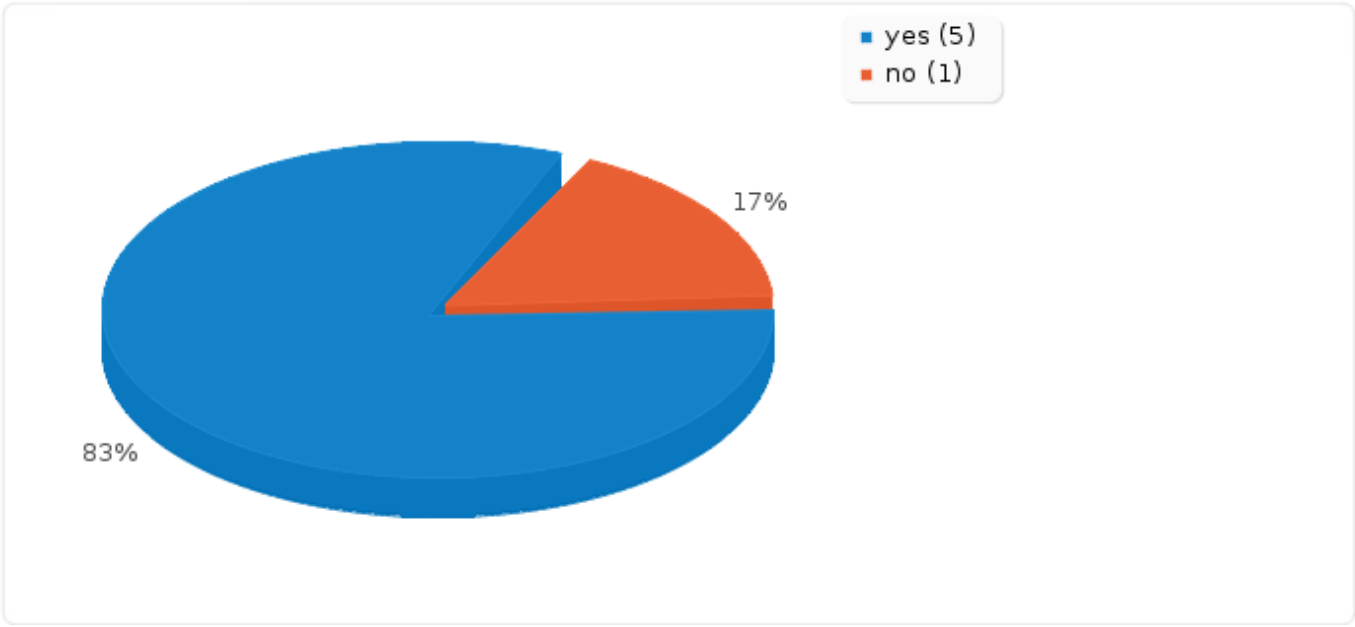
Field summary for 3.1.4

Did you install and run the tool?

Answer	Count	Percentage
yes (A1)	5	83.33%
tried, but did not get it working (within the time I had) (A2)	0	0.00%
no (A3)	1	16.67%
No answer	0	0.00%

Field summary for 3.1.4

Did you install and run the tool?



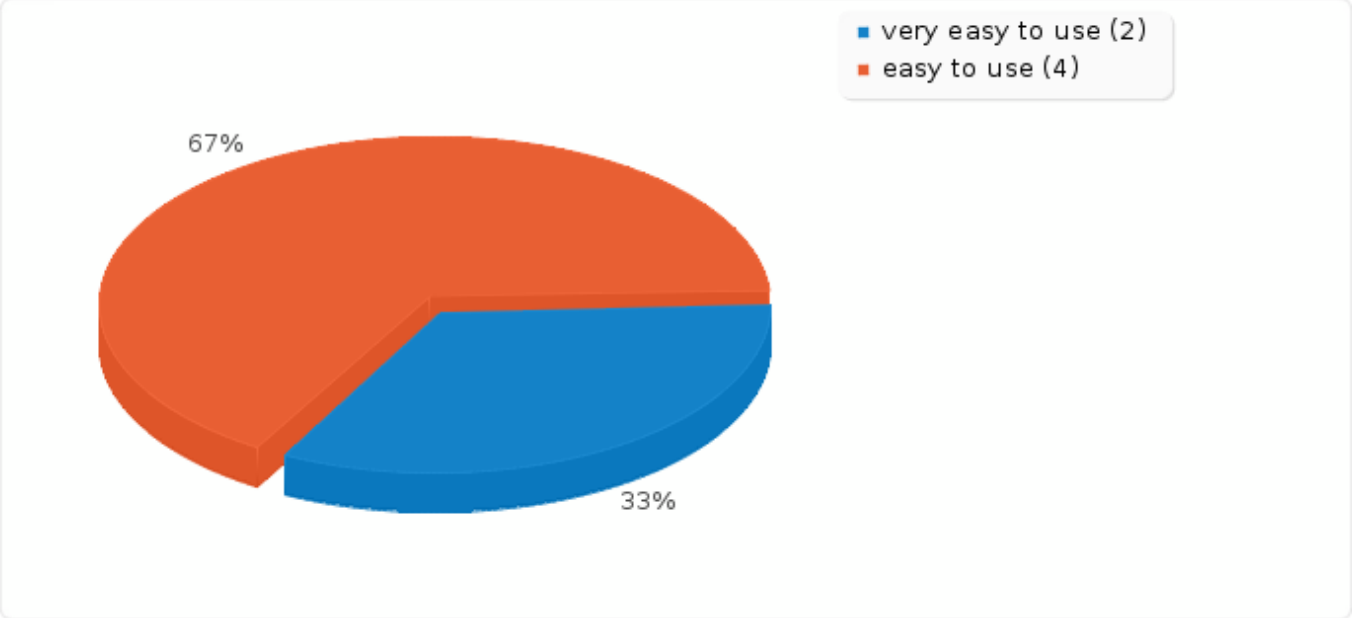
Field summary for 3.1.5

Do you consider the tool easy to use?

Answer	Count	Percentage
very easy to use (A1)	2	33.33%
easy to use (A2)	4	66.67%
moderately easy to use (A3)	0	0.00%
difficult to use (A4)	0	0.00%
has severe usability issues (A5)	0	0.00%
No answer	0	0.00%

Field summary for 3.1.5

Do you consider the tool easy to use?



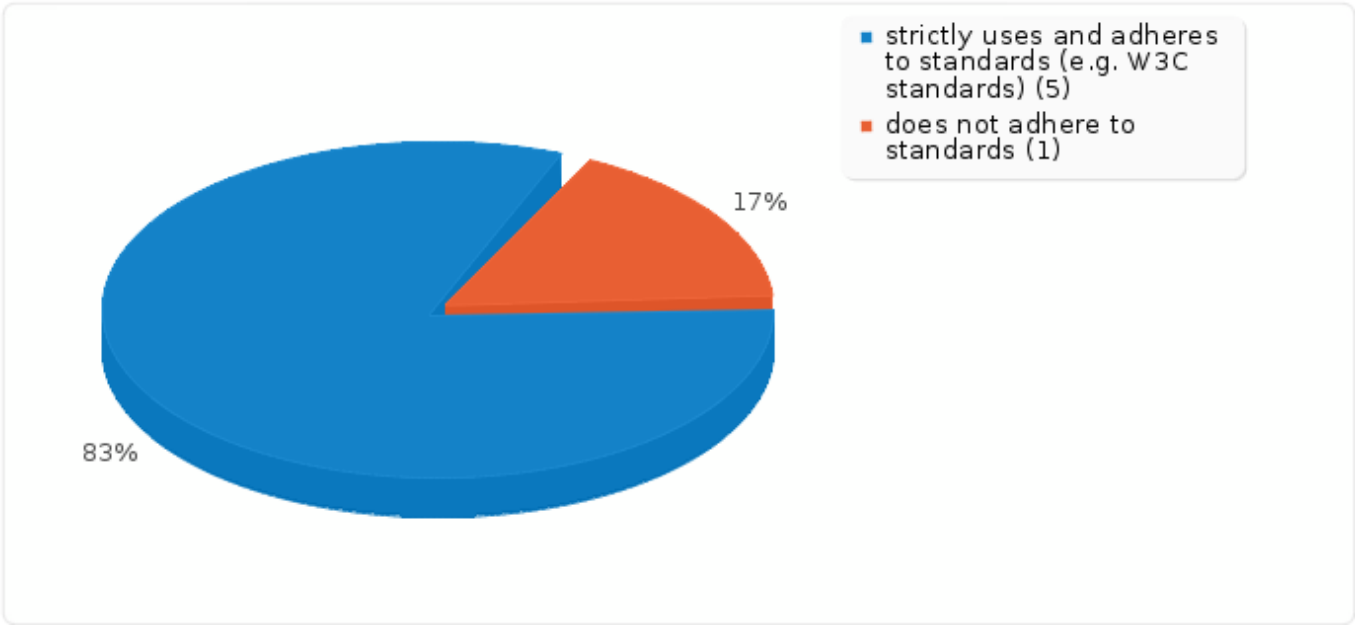
Field summary for 3.1.7

How do you rate the standards compliance and compatibility of the tool?

Answer	Count	Percentage
strictly uses and adheres to standards (e.g. W3C standards) (A1)	5	83.33%
deviates from standards, but you consider those deviations necessary (please explain) (A2)	0	0.00%
does not adhere to standards (A3)	1	16.67%
No answer	0	0.00%

Field summary for 3.1.7

How do you rate the standards compliance and compatibility of the tool?



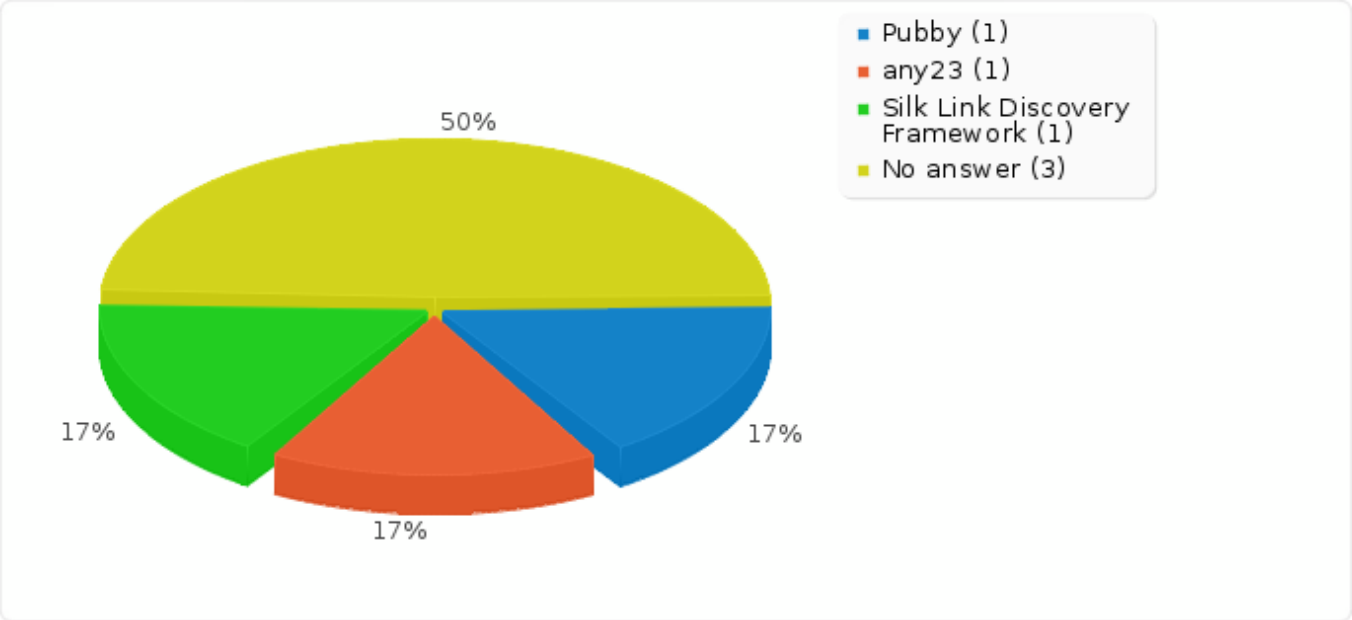
Field summary for 4.1.0

Name of the second selected Tool

Answer	Count	Percentage
Neologism (A1)	0	0.00%
D2R Server (A2)	0	0.00%
RDF Extension for Google Refine (A3)	0	0.00%
Pubby (A4)	1	16.67%
LDSpider (A5)	0	0.00%
Sindice (A6)	0	0.00%
any23 (A7)	1	16.67%
R2R Framework (A8)	0	0.00%
Silk Link Discovery Framework (A9)	1	16.67%
Sig.ma (A10)	0	0.00%
RelFinder (A11)	0	0.00%
OntoWiki (A12)	0	0.00%
SPARQL Views (A13)	0	0.00%
No answer	3	50.00%

Field summary for 4.1.0

Name of the second selected Tool



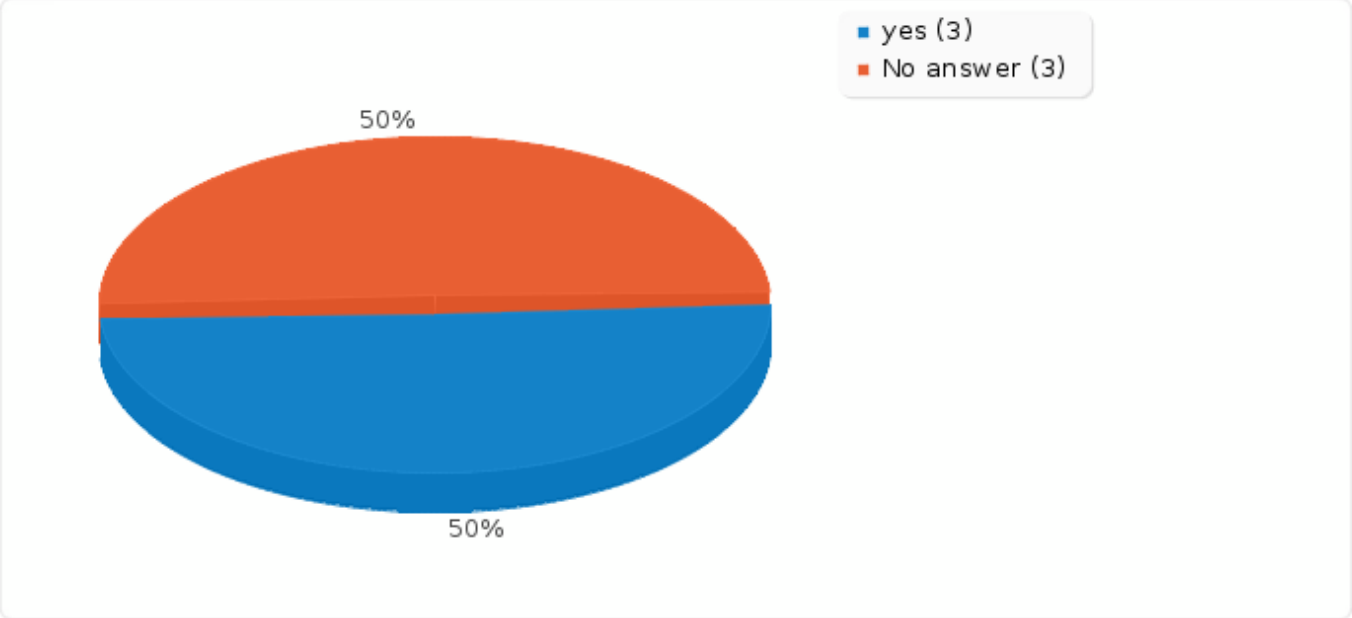
Field summary for 4.1.4

Did you install and run the tool?

Answer	Count	Percentage
yes (A1)	3	50.00%
tried, but did not get it working (within the time I had) (A2)	0	0.00%
no (A3)	0	0.00%
No answer	3	50.00%

Field summary for 4.1.4

Did you install and run the tool?



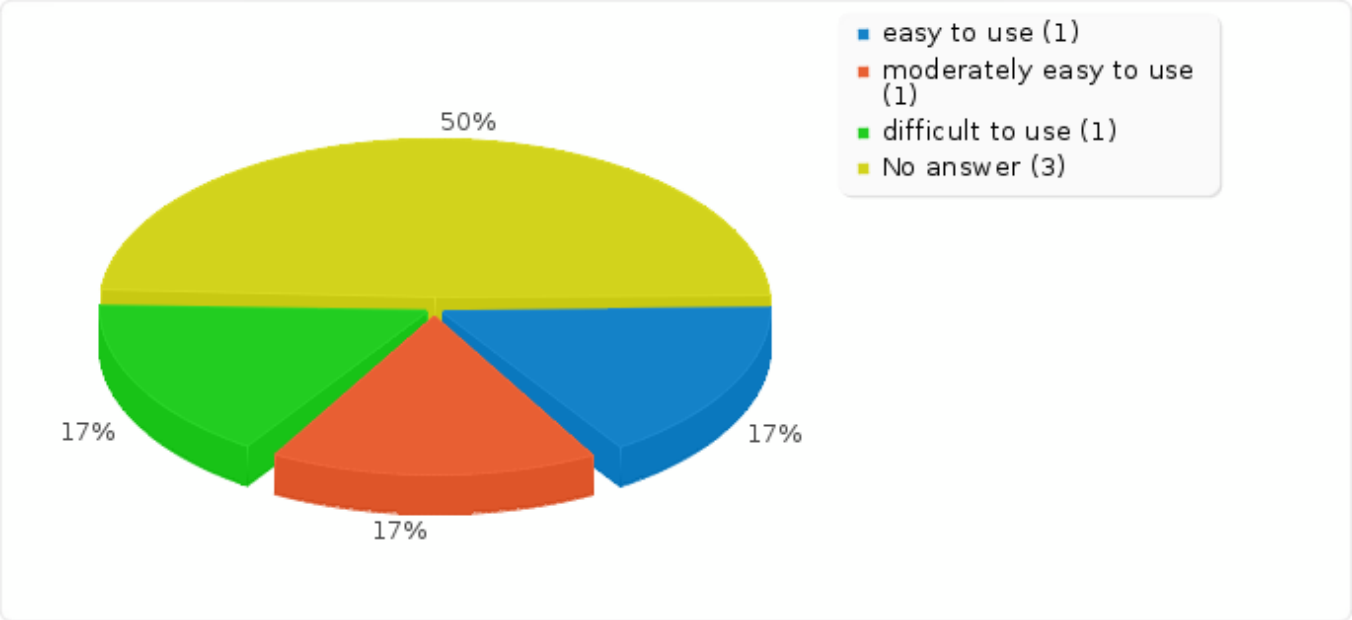
Field summary for 4.1.5

Do you consider the tool easy to use?

Answer	Count	Percentage
very easy to use (A1)	0	0.00%
easy to use (A2)	1	16.67%
moderately easy to use (A3)	1	16.67%
difficult to use (A4)	1	16.67%
has severe usability issues (A5)	0	0.00%
No answer	3	50.00%

Field summary for 4.1.5

Do you consider the tool easy to use?



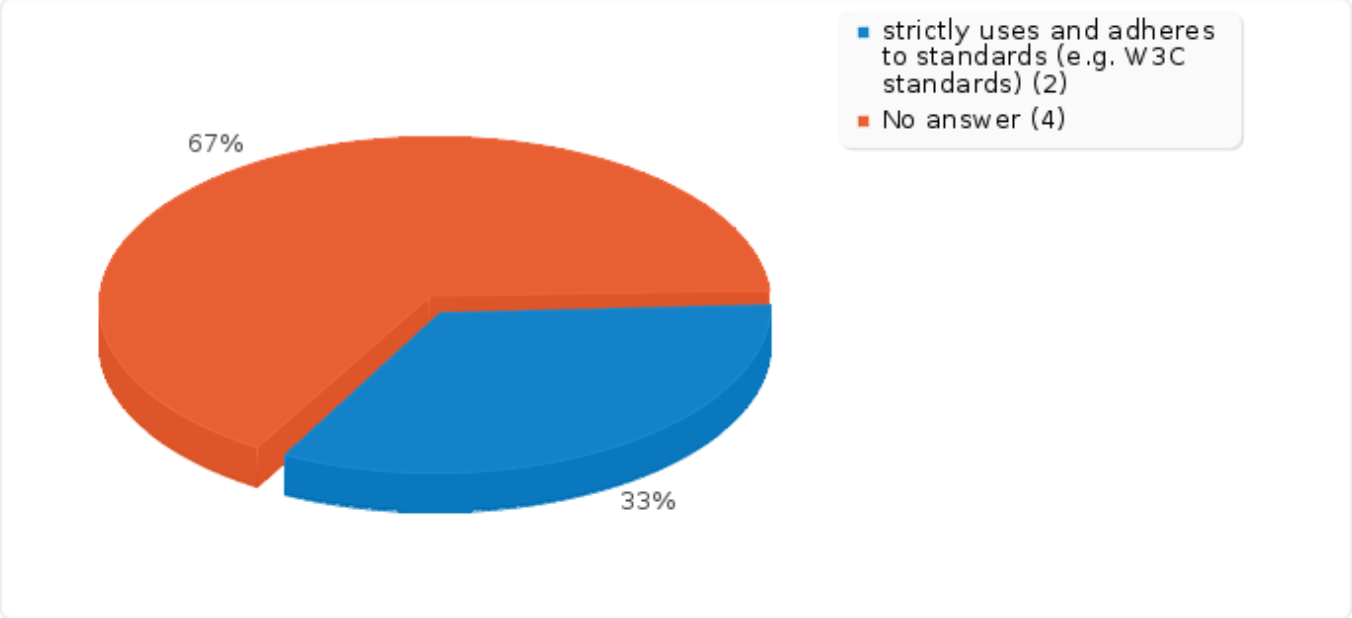
Field summary for 4.1.7

How do you rate the standards compliance and compatibility of the tool?

Answer	Count	Percentage
strictly uses and adheres to standards (e.g. W3C standards) (A1)	2	33.33%
deviates from standards, but you consider those deviations necessary (please explain) (A2)	0	0.00%
does not adhere to standards (A3)	0	0.00%
No answer	4	66.67%

Field summary for 4.1.7

How do you rate the standards compliance and compatibility of the tool?



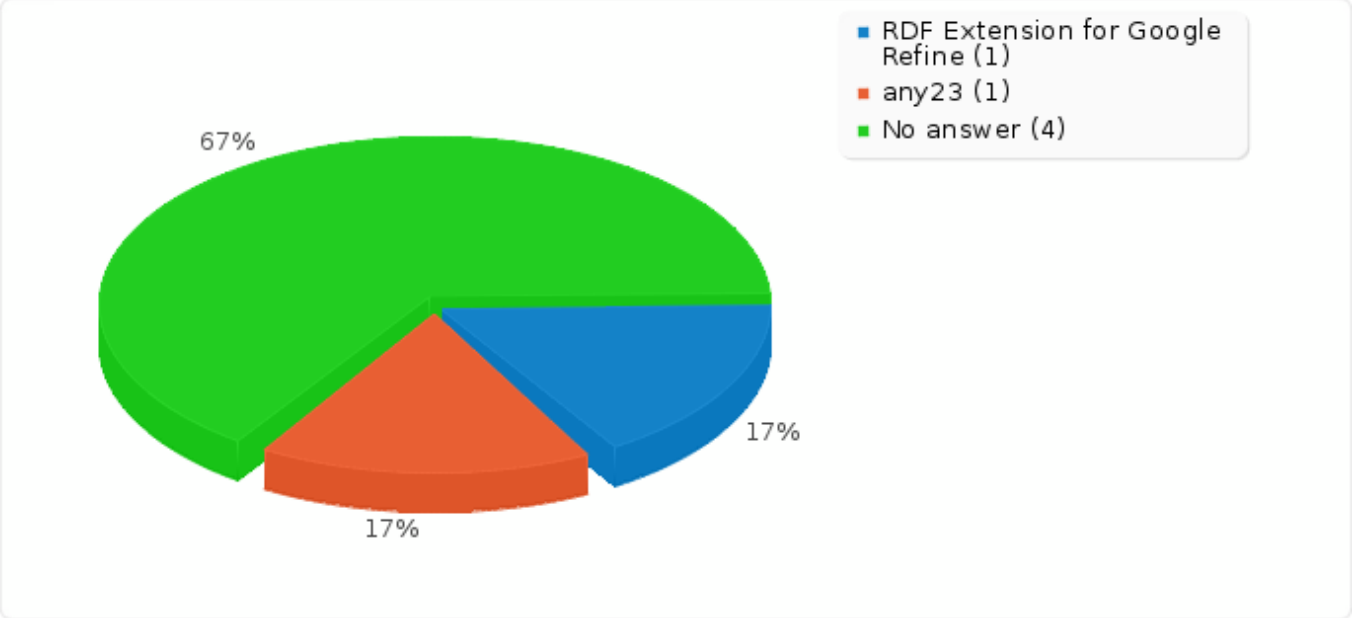
Field summary for 5.1.0

Name of the third selected Tool

Answer	Count	Percentage
Neologism (A1)	0	0.00%
D2R Server (A2)	0	0.00%
RDF Extension for Google Refine (A3)	1	16.67%
Pubby (A4)	0	0.00%
LDSpider (A5)	0	0.00%
Sindice (A6)	0	0.00%
any23 (A7)	1	16.67%
R2R Framework (A8)	0	0.00%
Silk Link Discovery Framework (A9)	0	0.00%
Sig.ma (A10)	0	0.00%
RelFinder (A11)	0	0.00%
OntoWiki (A12)	0	0.00%
SPARQL Views (A13)	0	0.00%
No answer	4	66.67%

Field summary for 5.1.0

Name of the third selected Tool



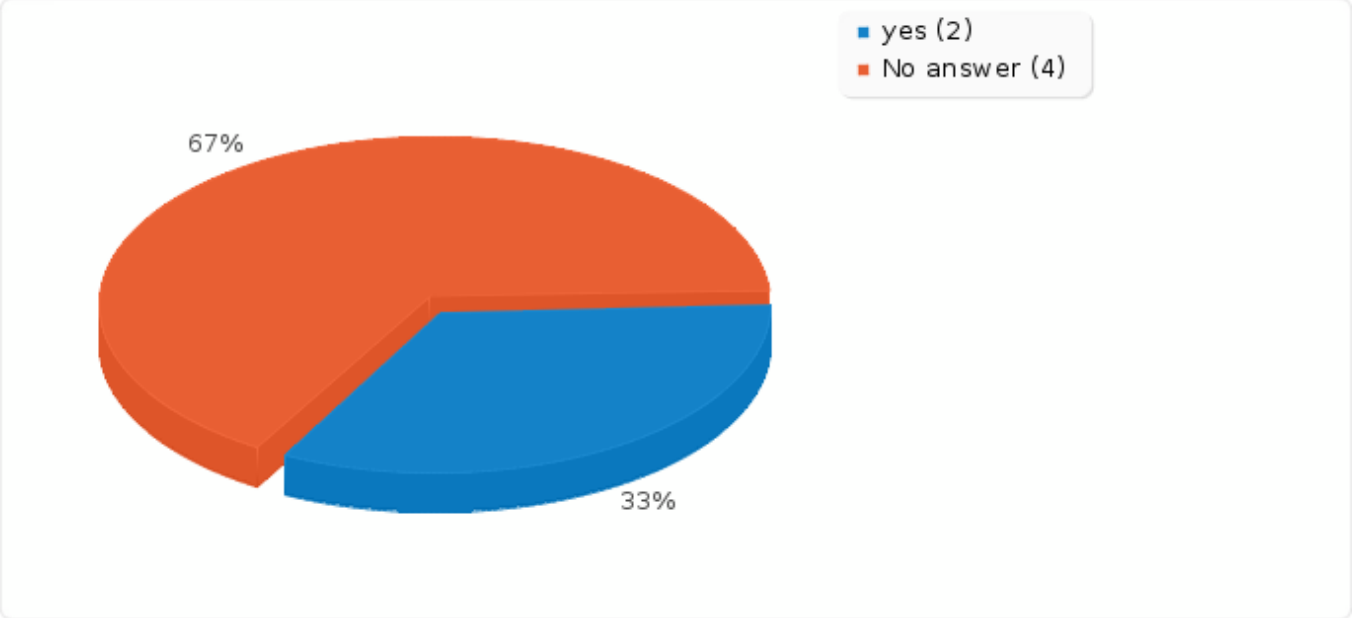
Field summary for 5.1.4

Did you install and run the tool?

Answer	Count	Percentage
yes (A1)	2	33.33%
tried, but did not get it working (within the time I had) (A2)	0	0.00%
no (A3)	0	0.00%
No answer	4	66.67%

Field summary for 5.1.4

Did you install and run the tool?



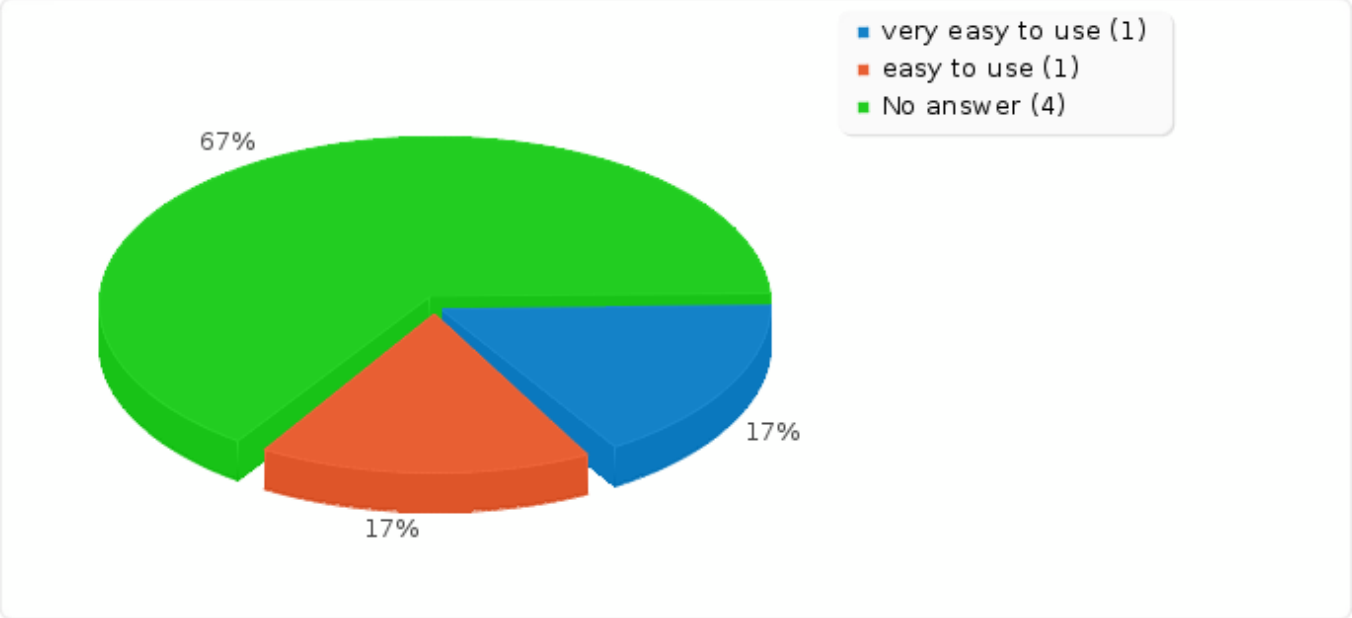
Field summary for 5.1.5

Do you consider the tool easy to use?

Answer	Count	Percentage
very easy to use (A1)	1	16.67%
easy to use (A2)	1	16.67%
moderately easy to use (A3)	0	0.00%
difficult to use (A4)	0	0.00%
has severe usability issues (A5)	0	0.00%
No answer	4	66.67%

Field summary for 5.1.5

Do you consider the tool easy to use?



Field summary for 5.1.7

How do you rate the standards compliance and compatibility of the tool?

Answer	Count	Percentage
strictly uses and adheres to standards (e.g. W3C standards) (A1)	0	0.00%
deviates from standards, but you consider those deviations necessary (please explain) (A2)	1	16.67%
does not adhere to standards (A3)	0	0.00%
No answer	5	83.33%

Field summary for 5.1.7

How do you rate the standards compliance and compatibility of the tool?