

Dynamic LOD Cloud Diagram

Editor(s): Name Surname, University, Country

Solicited review(s): Name Surname, University, Country

Open review(s): Name Surname, University, Country

Ciro Baron Neto , Sebastian Hellmann , Diego Esteves , Luca Matteis , Natanael Arndt ,
Andre Valdestilhas

*AKSW/BIS Universitat Leipzig,
Germany*

E-mail: {cbaron, hellmann, esteves, arndt, valdestilhas}@informatik.uni-leipzig.de

Abstract. The concept of Linked Open Data has been growing significantly, this growth comes with some problems such as data update and a way to filter them. This paper conducts a survey about approaches regarding the concept of LOD Cloud Diagram, a list of requirements and proposes an alternative solution that aims to provide a LOD Cloud Diagram updated in real time and allowing to apply make-filter on specific topics generating specific diagrams.

Keywords: Linked Open Data, Cloud Diagram, Semantic Web,

1. Introduction

Actually the current LOD Cloud image is generated and maintained by Richard Cyganiak from Centre for Data Analytics at NUI Galway and Anja Jentzsch from HPI company [10]. This Diagram is generated by one person, through sending emails and putting this information in a database.

The LOD Cloud Diagram [20] is not updated and doesn't allow use of filters.

There is already previous research on the concept of LOD Cloud Diagram and Visualization of Linked Open Data, such as in [12,15], but this work is also intended to serve as a complement to this research.

1.1. Directed link definition

Let $D = \{d_1, d_2, \dots, d_n\}$ be a set of datasets in a RDF file.

We call a dataset linked if for two disjoint datasets d_i and d_j the $\langle \text{some word} \dots \text{ask ciro or sebastian} \rangle$:

$d = \{ \langle s, p, o \rangle 1, \langle s, p, o \rangle 2, \dots, \langle s, p, o \rangle n \}$ A dataset is a set of triples, where d is a Dataset and s is a subject, p is a predicate and o is an object.

$d_{oi} = \{o_1, o_2, o_3, \dots, o_n\}$

Where d_{oi} is the list of objects.

$d_{si} = \{s_1, s_2, s_3, \dots, s_n\}$

Where d_{si} is the list of subjects.

$\text{count}(d_{oi} \cap d_{sj}) \geq 50$

Then there is a directed link from d_i to d_j .

1.2. Requirements

- gathering information from various sources and try to use as much information as available in void, sd and co
- generate dataid for each dataset
- building a catalog of dataset meta-data
- filtering by different properties of the datasets -
> we should also ask claus and integrate facets (demo:)
- Dynamic generation of cloud diagram picture
- the picture could visualize more properties than just size, links and topic area (government, linguistic, social, geo, ...) e.g. it could colorize the datasets by used language, set the size in proportion to how strongly it is interlinked, create a "heatmap" which visualizes, when the dataset was last updated, or any other property of the dataset

2. Related Work and Existing approaches

There are several existing approaches about Visualizing Data 2D, 3D, Flash, Flex, Mobile, among others [2], [12], [15]. This works are relevant to our work, but our main goal are more focused to LOD CCloud Diagram, in order to do a new Visualization of this Diagram that are updated and allow make-filters.

2.1. *lod-cloud.net: The Linking Open Data cloud diagram*

This work is our main motivation, because here we try to provide some improvements about the actual LOD Cloud Diagram [20]. The problems that we will try to provide some solution are: Up-to-dateness, because according the authors [20], **"This image shows datasets that have been published in Linked Data format, by contributors to the Linking Open Data community project and other individuals and organisations. It is based on metadata collected and curated by contributors to the Data Hub as well as on metadata extracted from a crawl of the Linked Data web conducted in April 2014."**

Therefore, we can notice that this Graph is not updated automatically, because the last updated was: 2014-08-30, and of course, not allowing filters, because is an static image.

2.2. *VoID-graph*

VoID-graph [19] is a standalone web-tool that, given a VoID description, can visualize a diagram similar to the LOD cloud [20], is built using Open Web standards such as JavaScript, HTML5, CSS3 and SVG. Users can simply download the source-code and run the index.html file to utilize the tool from the webSite. Compared to other software, which requires the configuration of specific runtimes and the installation of a variety of dependencies, VoID-graph can more easily be hosted and installed on a variety of systems.

2.3. *LOD Cloud VoID Generator*

This work [11] is responsible by generate a VoID description of all datasets in the lodcloud group on the Data Hub.

This project is about generating a VoID description of all the datasets in the LOD Cloud diagram. It currently

generates an RDF dump containing these descriptions, available online here:¹

The LOD Cloud diagram is a pictorial of datasets published in linked data format on the Web. Metadata for these datasets is recorded in the Data Hub, an open directory of datasets. The lodcloud group contains metadata for all the datasets in the LOD Cloud diagram. VoID is an RDF vocabulary for expressing metadata about such datasets in RDF format. At some point, this project was intended to produce not just an RDF dump, but also RDF and HTML descriptions of each entity described in the dump. That effort stalled, and is removed from the current codebase, but can still be found in the feature-html branch.

Source-code for lod-cloud.net [20], can be found here:², the website of the Linked Open Data Cloud diagram [20].

2.4. *DataHub2Void*

There are an work [9] that generates a VoID description of all datasets in the lodcloud group on the Data Hub.

2.5. *LOD for all*

This work [18] provides an LOD Cloud Diagram, that have some particularities, such as, when you keep you mouse pointer under an bubble, a kind of sub graph is showed. Allow make filters as well. Therefore this feature brings an easy way comparing with others that makes this work be relevant. This work is licensed under a Creative Commons Attribution 3.0 Unported License. creative commons license. Copyright © 2014 FUJITSU LABORATORIES LTD.

Includes additional datasets beyond those that are officially part of the "lodcloud" group. Also include datasets that have been tagged with the "lod" tag (according to metadata collected from Datahub) and that follow the license criteria explained above.

Is interactive. You can filter datasets based on several criteria, zoom and pan around the diagram and select a dataset with a click to inspect its links to other datasets, or with a double-click to see the dataset details page.

¹<http://lod-cloud.net/data/void.ttl>

²<https://github.com/lod-cloud/lod-cloud>

2.6. Towards a Dynamic Linked Data Observatory (DynLOD)

The Dynamic Linked Data Observatory [17] is a framework to monitor Linked Data over an extended period of time. The core goal of our work is to collect frequent, continuous snapshots of a subset of the Web of Data that is interesting for further study and experimentation, with an aim to capture raw data about the dynamics of Linked Data. The resulting corpora will be made openly and continuously available to the Linked Data research community.

Until now, they provide weekly crawls and updated.

The newest crawl dumps should be available for download ³ roughly one day after they got crawled.

This work [17] proposes an way to monitoring Linked Open Data around the Web. One possible problem is that data format, because we cannot found an specific format such as turtle, VoId Description or another one.

2.7. Protovis

Protovis [1] was already been analysed before in [14], but as complement we can tell that is a graphical approach to visualizing the LOD Cloud diagram using free and open-source under the BSD License. It uses JavaScript and SVG for web-native visualizations and no plug-in required. The Protovis team is now developing a new visualization library called D3.js [6], with improved support for animation and interaction. There is no way filter and the graph is manually updated.

2.8. Czech LOD Cloud

Czech LOD (CzLOD) cloud diagram [16] show a set of interlinked LOD datasets that they have published in your research group trying to show the benefits of the Linked Data Visualization Model (LDVM) for LOD visualization. The data are not automatic updated and we cannot found a way to make filters.

2.9. lodlive

LODLive [8] is an environment that generates an Graph based in some filter in some predefined dataSets like dbPedia.org and freebase.org, you also can input an URI. This work allow make filters, but the result graph/Graph is not properly similar to LOD Cloud Di-

agram [20]. An suggestion maybe could be put a kind of automatically search for all datasets, could be direct in DataHub.io.

2.10. rkbexplorer

The rkbexplorer [13] presents a tool that provides an graphical environment. An good point is that according the literature ⁴ that shows is possible to generate a LOD Cloud Diagram using this tool and also allowing filters.

2.11. Callimachus

This work initiallly [4] is an tool for visualizing Linked Data, but still here because this tool can reproduce the LOD Cloud Diagram as well, but in his web-Site ⁵ is explained that is not a intention of this tool. Therefore, this is not a good way to adote this tool, unless we have only this way.

2.12. Gephi: Semantic Web Crawling Visualization

This approach [3] provide some examples and visualizations of semantic web crawls using Gephi ⁶ as graph visualiser and LDSpider as semantic crawler, that allow filters, but not provide automatic update. This work can be used as LOD Cloud Digram as well, because as lod-cloud.net use LDSpider to crawling the data.

3. Criteria

In order to provide a explanation about the criteria adopted in this work which allow to evaluate tools and approaches.

Data Updated Automatically: The work needs to provide data that are automatically updated, in order to provide always an updated and reliable LOD Cloud Diagram.

Allow filter: In order to provide an way to make sub LOD CLOUD diagrams about specific subjects, for example, an sub LOD Cloud about Librarians.

³<http://swse.deri.org/dylodo/data/>

⁴<http://lists.w3.org/Archives/Public/public-lod/2009Jan/0061.html>

⁵<http://semanticweb.com/tag/callimachus-project>

⁶<http://gephi.github.io/>

Table 1
Selected approaches and criteria

Approach / Tool	Criteria		
	Updated	Allow Filter	Web
lod-cloud.net			x
LOD for all		x	x
DynLOD	x		x
Protovis			x
VoID			x
LODLive		x	
rkbexplorer		x	x
Callimachus		x	x
Gephi			
Crawling Visualization		x	

4. Linked Data Browsers

Linked Data Browsers can be an way to visualise LOD-cloud diagrams, thus, the work by Dadzie 2011 [12] shows key requirements and approaches, such as two tables represented by figures 1 and 2 that can represent some criteria comparing 15 tools to visualising Linked Data.

5. Steps to have you own LOD Cloud Diagram

This steps are based in a tutorial provided by Luca Matteis [19] ⁷.

1. Download all metadata about datasets already stored in <http://datahub.io/>. Use this script:
8

to get a void.ttl file

2. Load void.ttl into your RDF storage server.
3. Make this query:

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX tag:<http://www.holygoat.co.uk/owl/redwood/0.1/tags/>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
```

⁷<http://markmail.org/search/?q=+list%3Aorg.w3.public-lod+cloud+visualisation#query:lod%20cloud%20visualisation+page:1+mid:xiwz5envzoeqbq3x+state:results>

⁸<https://github.com/lod-cloud/datahub2void>

```
PREFIX void:<http://rdfs.org/ns/void#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms:<http://purl.org/dc/terms/>
select * ?s a void:Linkset. ?s void:linkPredicate
skos:exactMatch. ?s void:objectsTarget ?otarget.
?s void:subjectsTarget ?target. ?s void:triples
?triples.
```

4. Export results of this query as N-triples.
5. Download <https://github.com/lmatteis/void-graph>.
6. Open file index.html in your browser. Inside text area field, paste all n-triples you have already exported.
7. That is all. You must have your own LOD - Cloud in a couple seconds.

6. Problem Analysis

As stated in Section 2 all of the tools we investigated have update intervals of over 4 months and need a lot of manual effort to keep meta data quality fresh.

We will use an short story in order to explain our contribution in a clear way.

6.1. Story

I am a Greek Librarian and I want to find, filter and classify all datasets that is related to Greece (its people and locations), the greek language and any greek things (e.g. movies, books). Then I want to publish this classification and create my own LOD cloud images to help my colleagues (other Greek librarians) to browse all relevant datasets. I can spend the next three weeks full time to work on a software tool to plow through all the data.

6.2. How to solve

The problem about find, filter and classify datasets we provide an web interface on DataId-Graph where you can find, filter and classify your datasets in a easy an intuitive way create you own LOD cloud diagram.

Table 2
Comparison of functionality in the text-based LD browsers

Usability Criterion	LD Browser						
	Dipper	Disco	Marbles	Piggy Bank	Sig.ma	URI Burner	Zitgist
Visual presentation	–	–	–	–	–	–	–
Data overview							
Detail on demand	x*	x*	x*	x	x*	x*	x
Highlight links in data			x*	x*			
Support for scalability					x		x*
Query (formal syntax)			x*			x	
Query (forms / keyword)				x			
Filtering					x		x
History				x	x*		x*
Presentation templates	x*		x		x		x
Keyword / DMI entry point				x	x	x	
Non-domain specific		x	x	x	x	x	x
Faceted Search / Browse				x			x
Publication / Syndication				x			
Edit underlying data							
Reusable output	x			x	x	x	x*
Target – Lay-users				x			
Target – Tech-users	x	x	x		x	x	x

Fig. 1. Criteria about LD Browsers (1-2) [12]

7. DataId

DataID [7] is an proposal of best-practice for LOD dataset descriptions which utilize RDF files hosted together with the datasets, under the same domain. The data model is based on the widely used DCAT and VoID vocabularies, as well as supporting tools to create and publish DataIDs. The DataID data model provides a uniform way to describe general metadata of datasets.

The Figure 3 shows the overall structure of DataID including the main features.

→ Semantics paper (ask Ciro); Copy and paste from DataID paper ?

7.1. How to index dataids for the new LOD Cloud image

7.2. Implementation

We have five steps to complete the task as you can see in Figure 4 that shows an fluxogram with all steps covered by our idea.

1. **Receive DataId:** In this step we will receive the DataId file, and do validations.

2. **Parse DataId:** After the reception of this DataId, it is possible to read and start to record the data, then we make a search for distributions.
3. **Check for modifications:** In this step we will use HTTP headers to find an signal that the data has modified, such as date or size.
4. **Compare Distributions:** Now we know that there are some modification, then, we need to compare distributions in order to know the right one, thus we use an probabilistic data structure called Bloom Filter [5] under DataId files to find modification on linksets and distributions.
5. **Update DataID:** In case of changes detected by BloomFilter, we provide an update under the DataId file, for exeample, Linksets, because the relations are based on linksets, therefore, after update DataId file, but the main file remains with the same name, thus, the URL is the same, then we have an updated DataId-Graph.

7.2.1. How Bloom Filters works

The Bloom filter is an very important part of our work, because we can have an LOD Cloud Diagram updated and efficiently.

The Figure 5 shows an example of a Bloom filter, representing the set { x, y, z }. The colored arrows

Table 3
Comparison of functionality in the LD browsers with visual representations available

Usability Criterion	LD Browser							
	DBpedia Mobile	Fenfire	Isa Viz	LESS	OpenLink	RDF Gravity	RelFinder	Tabulator
Visual presentation	x	x	x	x	x	x	x	x
RDF graph view		x	x		x	x	x	
Visual overview	x*		x			x		
Data overview							x*	
Detail on demand	x	x	x	x*	x	x	x	x
Highlight links in data		x	x		x	x	x	x*
Support for scalability	x*	x	x		x*	x*	x	
Query (formal syntax)	x*			x	x*	x	x	x
Query (forms / keyword)	x		x		x	x		x
Filtering	x	x*	x		x	x	x	
History	x		x		x			x*
Presentation templates	x		x	x	x			
Keyword / DMI entry point	x				x		x	
Non-domain specific		x	x	x	x	x		x
Faceted Search / Browse								
Publication / Syndication	x	x	x					x
Edit underlying data		x*	x					x
Reusable output		x	x	x				
Target – Lay-users	x	x*	x	x*	x*	x*	x*	
Target – Tech-users		x	x	x	x	x	x	x

Fig. 2. Criteria about LD Browsers (2-2) [12]

show the positions in the bit array that each set element is mapped to. The element w is not in the set $\{x, y, z\}$, because it hashes to one bit-array position containing 0. For this figure, $m = 18$ and $k = 3$.

Explaining in more details, instead of set $\{x, y, z\}$ in our case we have set $\{\text{Linkset1}, \text{Linkset2}, \dots, \text{LinksetN}\}$. Then we will use this data structure to compare linksets and to know how many links we have among our DataId Files.

7.3. DataID-Graph: Visualizing complex LOD DataSets

The idea about Visualize Linked Datasets on the Web was originated by Lucas Matteis on his work called Void-graph. In this work we will provide some modifications that allows visualize DataID. DataID complements the Void structure because DataID has VoId fields as well and provides an structure for representing Datasets more complete. <Ask MARTIN, CIRO, SEBASTIAN, DIEGO, etc..>

The Idea behind DataID [7] is that can provides an structure to represent DataSets. DataID is a best-practice for LOD dataset descriptions which utilize

RDF files hosted together with the datasets, under the same domain. DataID has an data model, which is based on the widely used DCAT and VoID vocabularies, as well as supporting tools to create and publish DataIDs and use cases that show the benefits of providing semantically rich metadata for complex datasets. One characteristic that makes possible to visualize DataID's as an Graph is the use of VoID vocabularies, this is the main feature that makes possible to visualise DataIDs, with some modifications in the VoID-graph [19] that will be explained in this paper.

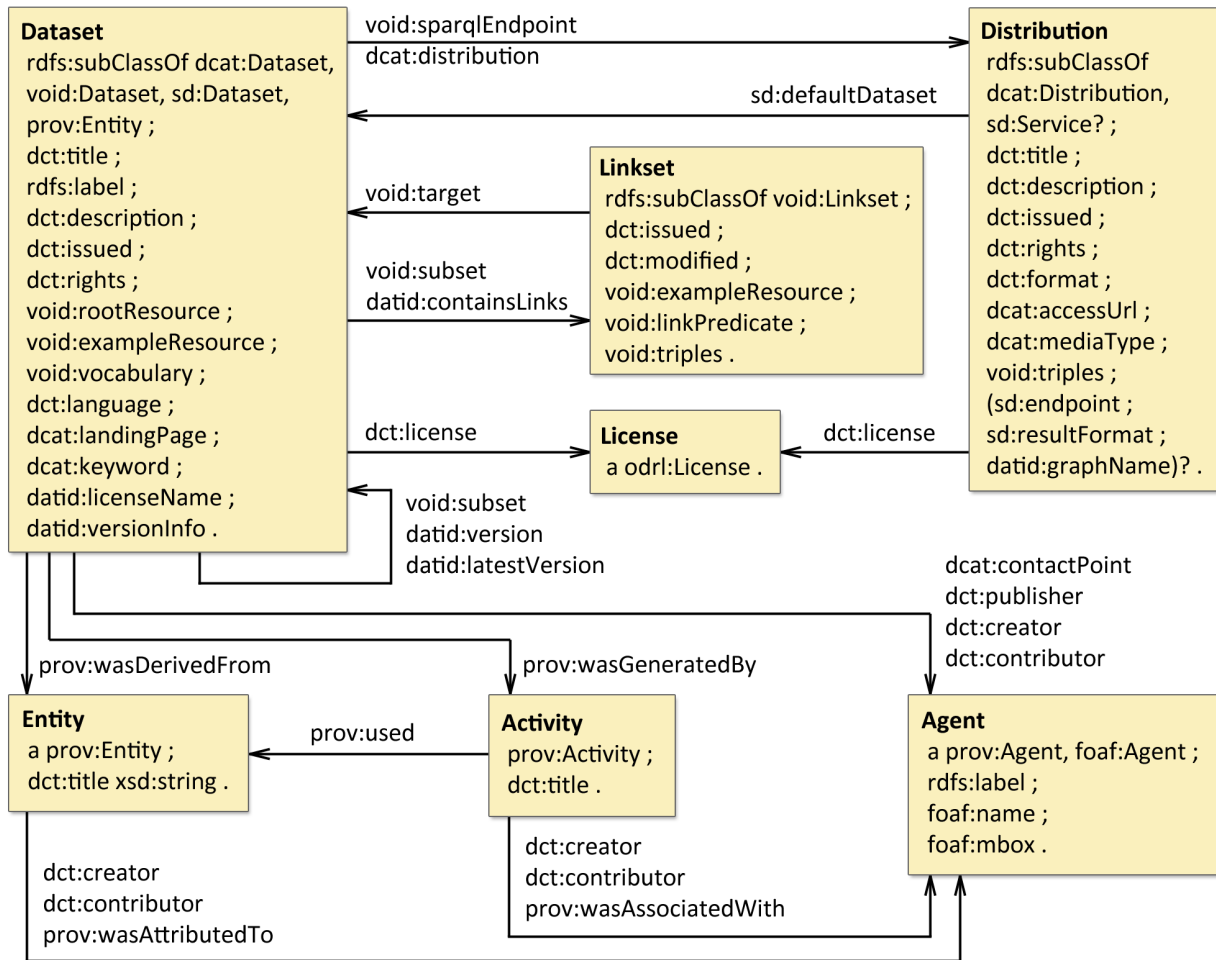
7.3.1. Architecture

The architecture of DataID-graph is very similar to VoID-graph [19], as presented in Figure 6.

—————>>>>

8. LOD Cloud results

—> Discuss with Ciro and Diego.



```

@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix sd: <http://www.w3.org/ns/sparql-service-description#> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix datid: <http://dataid.dbpedia.org/ns#> .

```

Fig. 3. DataID vocabulary diagram

9. Evaluation

→ Discuss with Ciro and Diego. (Discussion : Discuss) (Andre)

Experimental evaluation: (Andre)

We can realize that our model present an improvement...

e.g. looking at the fact, that we could visualize any property of the datasets: e.g. it could colorize the datasets by used language, set the size in proportion to how strongly it is interlinked, create a “heatmap”

which visualizes, when the dataset was last updated, ... this was not available so far and comes with a lot of work if done manually. With our (semi-)automatic approach nearly anything can be visualized

Table 2
Evaluation

NameDataID	QtdTriples	TimeToCreate	%FalsePositive	QtdLinksets
x	x	x	x	x

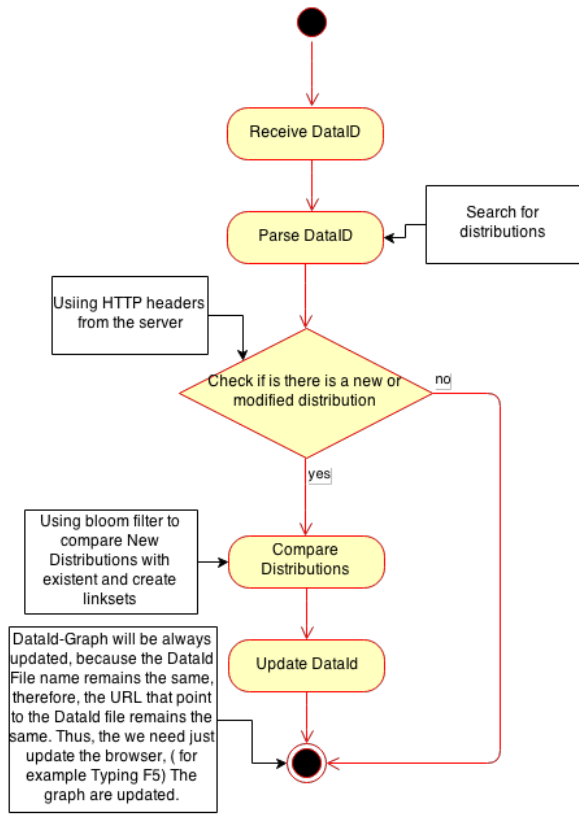


Fig. 4. Fluxogram about whole process.

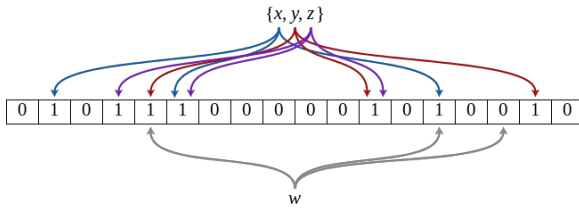


Fig. 5. Bloom filter.

10. Some reasons for this work were identified as:

1. Do something for people that no have enough knowledge about RDF nor Ontologies.
2. Visualising Data can help to identify errors.
3. Make Visual Queries and Filters without know SPARQL, because using SPARQL, before you need to understand the syntax and data content structure.

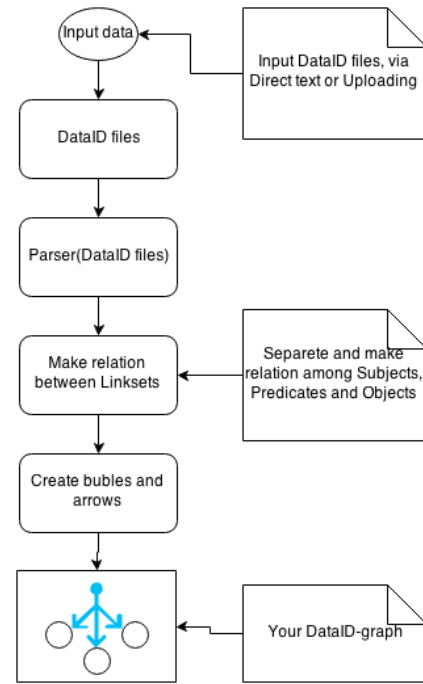


Fig. 6. Activity Diagram for DataId-graph.

11. Contributions

We will describe the contributions of this work can provide:

- **Dynamic LOD Cloud Diagram:** This work brings an LOD Cloud Diagram really updated and allowing filters, as we notice that does not exists an work that do the same thing.
- **A survey:** In the Related works section you can see an updated survey about approaches regarding the concept of LOD Cloud Diagram, that compares and makes an short evaluation about this works. and we also provide a list of requirements.
- **DataId-Graph** A way to visualize Datasets using VoId descriptions to represent an extension to the void-graph, because DataId-Graph shows an structure to describe datasets more complete than only VoId descriptions. DataId uses Void vocabularies as part of your structure, this feature makes possible to visualize DataId's. The source code⁹ about DataId-graph is available to download.
- **Efficient comparison of linksets:** This part of this work compare distributions in order to know the right one, thus we use an probabilistic data

⁹<https://github.com/firmao/void-graph.git>

structure called Bloom Filter [5] under DataID files to find modification on linksets and distributions. Therefore we can notice, according Evaluation section that using Bloom filters we can have a better performance that justifies his use. This is an main point under this work, because without this performance that bloom filters brings this work probably will be impracticable.

12. Conclusions and Future Work

Some conclusion... → Future work: Use Radix tree to reduce de hard disk utilization.

References

- [1] Protovis: Linked open data graph. 2012. available in: <http://inkdroid.org/lod-graph/>.
- [2] K. Andrews. Information visualisation. 2011. available in: <http://courses.iicm.tugraz.at/ivis/surveys/>.
- [3] K. Banman. Semantic web crawling visualization.2014. available in: <https://github.com/kdbanman/visualcrawl>.
- [4] S. Battle, D. Wood, J. Leigh, and L. Ruth. The callimachus project: Rdfa as a web template language, 2012.
- [5] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, July 1970.
- [6] M. Bostock. D3.js - data-driven documents, 2012.
- [7] M. Brümmer, C. Baron, I. Ermilov, M. Freudenberg, D. Kontokostas, and S. Hellmann. DataID: Towards semantically rich metadata for complex datasets. In *Proceedings of the 10th International Conference on Semantic Systems, SEM '14*, pages 84–91. ACM, 2014.
- [8] D. V. Camarda, S. Mazzini, and A. Antonuccio. Lodlive, exploring the web of data. In *Proceedings of the 8th International Conference on Semantic Systems, I-SEMANTICS '12*, pages 197–200, New York, NY, USA, 2012. ACM.
- [9] R. Cyganiak. Datahub2void. 2014. available in: <https://github.com/lod-cloud/datahub2void>.
- [10] R. Cyganiak. The linking open data cloud diagram. 2014. available in: <http://lod-cloud.net/>.
- [11] R. Cyganiak. Lod cloud void generator. 2014. available in: <https://github.com/lod-cloud/datahub2void>.
- [12] A.-S. Dadzie and M. Rowe. Approaches to visualising linked data: A survey. *Semant. web*, 2(2):89–124, Apr. 2011.
- [13] H. Glaser, I. C. Millard, and A. Jaffri. Rkbexplorer.com: A knowledge driven infrastructure for linked data providers. In *Proceedings of the 5th European Semantic Web Conference on The Semantic Web: Research and Applications, ESWC'08*, pages 797–801, Berlin, Heidelberg, 2008. Springer-Verlag.
- [14] R. Kaiser. Visualizing open data. 2011. available in: <http://courses.iicm.tugraz.at/ivis/surveys/ss2011/g4-survey-open-data-vis.pdf>.
- [15] S. Khushro, F. Jabeen, S. R. Mashwani, and I. Alam. Linked open data: Towards the realization of semantic web - a review. *Indian Journal of Science and Technology*, 7(6), 2014.
- [16] J. Klímek, J. Helmich, and M. Necaský. Application of the linked data visualization model on real world data from the czech lod cloud. In *LDOW*, 2014.
- [17] T. Käfer, J. Umbrich, A. Hogan, and A. Polleres. Towards a dynamic linked data observatory. In *In LDOW at WWW*, 2012.
- [18] V. Lee. Lod for all. 2014. available in: <http://lod4all.net>.
- [19] L. Matteis. VoID-graph: Visualize linked datasets on the web, 2014.
- [20] C. B. Max Schmachtenberg and H. Paulheim. Adoption of the linked data best practices in different topical domains. volume 8796. Springer Internat. Publ., 2014.