A Visual Exploration Workflow As Enabler for the Exploitation of Linked Open Data

Laurens De Vocht, Anastasia Dimou, Jonas Breuer, Mathias Van Compernolle, Ruben Verborgh, Erik Mannens, Peter Mechant, and Rik Van De Walle

In: Proceedings of the 3rd International Conference on Intelligent Exploration of Semantic Data - Volume 1279, 30–41, 2014.

http://dl.acm.org/citation.cfm?id=2877799.2877803

A Visual Exploration Workflow as Enabler for the Exploitation of Linked Open Data

Laurens De Vocht[1], Anastasia Dimou[1], Jonas Breuer[2], Mathias Van Compernolle[3], Ruben Verborgh[1], Erik Mannens[1], Peter Mechant[3], and Rik Van de Walle[1]

[1] iMinds - Ghent University, Multimedia Lab, Ghent, Belgium
{firstname.lastname}@ugent.be
[2] iMinds - Vrije Universiteit Brussel, SMIT, Brussels, Belgium
{firstname.lastname}@iminds.be
[3] iMinds - Ghent University, MICT, Ghent, Belgium
{firstname.lastname}@ugent.be

Abstract. Semantically annotating and interlinking Open Data results in Linked Open Data which concisely and unambiguously describes a knowledge domain. However, the uptake of the Linked Data depends on its usefulness to non-Semantic Web experts. Failing to support data consumers to understand the added-value of Linked Data and possible exploitation opportunities could inhibit its diffusion. In this paper, we propose an interactive visual workflow for discovering and exploring Linked Open Data. We implemented the workflow considering academic library metadata and carried out a qualitative evaluation. We assessed the workflow's potential impact on data consumers which bridges the offer: published Linked Open Data; and the demand as requests for: (i) higher quality data; and (ii) more applications that re-use data. More than 70% of the 34 test users agreed that the workflow fulfills its goal: it facilitates non-Semantic Web experts to understand the potential of Linked Open Data.

1 Introduction

A lack of in-depth understanding of the underlying semantic technology limits most Web users to interpret and query Linked Open Data (LOD). Data consumers do not straightforwardly perceive that the enriched and machine-understandable representation of the information, the offered cognitive prosthetics, could help in solving complex problems [21]. As Dadzie and Rowe [3] mentioned, the uptake of the Linked Data by a mainstream audience depends on its utility to non-Semantic Web experts; failing to support them in understanding the LOD potential could inhibit its swift adoption and positive impact on society. Therefore, non-Semantic Web experts need the means to discover and to explore LOD more intelligently.

We expect that data consumers will form their demands for more LOD of better quality and for targeted applications benefiting from LOD, if they are aware of the potential value. Until now, observation of data was sufficient as data is often fragmented by default in different sources, even if sets of these data combined describe a certain knowledge domain. In the case of LOD though, their linked nature should be revealed. Exploring LOD implies that data consumers become aware of the available content and the links between them to really appreciate the "potential" they have at their disposal.

Furthermore, Janssen et al. [12] argued that information may appear to be irrelevant when viewed in isolation, but when linked and analyzed collectively can result in new

insights. Goedertier et al.[4] confirmed that there is a positive trend to open up public data. The Web accords to consumers the infrastructure and means for interaction with the data. However, the study also confirms that the re-use of provided LOGD is still limited. Finally, Genie Stowers[5] outlines the following reasons to make use of data visualization: (i) for storytelling and making information more visible; (ii) for simplifying, clarifying and analyzing data and making plans, within a more data-driven policy-cycle.

As LOD are typically represented as graphs, exploring its visualization is a way to allow Web users to implicitly compose queries, identify links between resources and intuitively discover new relevant pieces of information [3]. This addresses the need to reveal links, made available in the LOD, which were not explicitly defined in the original data.

From the perspective of the initiating government, there are clear benefits of such an initiative; in the particular case of research information data, public institutions make investments by funding research and they need to assess the return on such investment. Further employing the example of academic library metadata illustrates that researchers and organizations undertaking research (universities etc.) benefit foremost themselves from having qualitative data; this data they can then provide to a common LOD environment provided by a public sector body, where the public and (especially) the government can access and explore the data, thereby generating value.

In this paper, we present an interactive visual graph-based exploration workflow over published LOD and we assess how this workflow supports data consumers to discover and explore the data: how useful it is for data consumers to discover the data and explore the links between them. Our uppermost goal is to assess the users' satisfaction and thus the potential of such a visual workflow to act as enabler for LOD exploitation. The remainder of the paper is structured as follows: in section 2, we describe the current state of the art regarding LOD visualizations, information exploration and what LOD means in the societal context. In Section 3, we describe the exploration workflow, in Section 4 we describe the system architecture that implements the workflow and in Section 5 we present the evaluation. In Section 6 we summarize our conclusions and we suggest future work.

2 State of the Art

As our proposed solution is multi-component and multi-dimensional, we outline the state of the art of the different relevant fields below. Because our workflow relies on visualizations, we review the state of the art regarding LOD visualizations and academic metadata visualizations, as this is the domain of our use case. We outline the information exploration techniques applied to our workflow and summarize some insights already formed relevant to the LOD socio-economic potential.

2.1 Information Exploration

The resolution of vague or complex information problems requires exploratory behaviours, for instance: multiple publishers providing resources. We examine how exploratory search and exploratory data analysis applied to LOD can fruitfully solve such problems. Exploratory Data Analysis (EDA) [18] allows the data itself to reveal its underlying model and its relationships without requiring any formal statistical modeling and inference (non hypothesis-driven). Graphical EDA employs a variety of techniques

[4] https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd

[5] http://www.businessofgovernment.org/report/use-data-visualization-government

to present the underlying data, maximizes the insight into a dataset and uncovers the underlying data patterns, allowing the users to discover the resources in the dataset.

Exploratory search, on the other hand, describes either the problem context that motivates the search or the process by which the search is conducted [14]. The users start from a vague but still goal-oriented defined information need and are able to refine their need upon the availability of new information to address it, with a mix of keyword look-up, expanding or rearranging the search context, filtering and analysis. During exploratory searches and analysis, it is likely that the problem context becomes better understood, allowing the searchers to make more informed decisions about interaction or information use [21].

2.2 Linked Open Data Visualizations.

The "Linked Data Visualization Model" (LDVM) allows to connect different datasets, data extractions and visualizations in a dynamic way [1] rather than focusing on a single platform. Furthermore, the use of coordinated views to facilitate integration of visualizations [16], is a complex process to decide on visualization methods to successfully aid seeking and discovery of information [15]. Coordinated views are provided to align multiple perspectives on a dataset. In "VisLink", "coordinated views" link existing visualizations [2], while we use a coordinated view to align the narrowing and broadening views in the workflow. Tvarozek et al. [19] empower users with access to semantic information spaces via an exploratory browser. At the end of an exploration session users need to start a new search, a history view allows users to step back.

Dadzie and Rowe [3] concluded in their study on Linked Data interfaces and visualizations that only a limited number was available at the time of writing and each of them focuses on a separate aspect to support users. They highlighted the issue, an important motivation for our workflow as well, that without good quality LOD there is little motivation to build such interfaces for end-users while these interfaces are needed to locate and retrieve LOD in the first place. To address this issue, we aim to have visualizations for diverse domain specific solutions. However, these visualizations remain generic because they are built for any case specific queries against published LOD.

2.3 Academic Data Visualizations

In the past there were attempts to visualize research networks but most of them did not rely on LOD. The below mentioned works based on research LOD consider visualizations as a supportive mean to the presented information. None of them focus on the visualizations and, therefore they do not take into consideration a methodological approach for the data exploration as such.

The Semantic Web Journal published its own Drupal-based journal management system [11] focusing on providing a novel user interface. Among others, they provide graph-based research networks that visualize the emerging research networks as researchers author papers together or they review the different submissions. "Arnet-Miner" [17], on the other hand, distinguishes between the networks (star graph of co-authors) and the communities of researchers (simple graphs). Finally, "TalkExplorer" [20] takes into consideration bookmarks and tags for the visualizations of the research groups and puts the focus on providing recommendations rather than exploring the underlying dataset. In our workflow we make abstraction of the query creation process and use of pre-defined query templates to facilitate the creation of the visualizations.

3 Exploration Workflow

Data consumers need *Information Analysis* and *Synthesis* (IAS) to acquire knowledge and form inquiries for a domain unknown to them, consisting of complex and disparate data. Corresponding *Information Exploration* techniques can facilitate the exploration of the available datasets as mentioned in Section 2.1. To this end, we considered for our workflow *exploratory data analysis* to assist data consumers to analyze the available dataset and *exploratory search* to facilitate them synthesizing complex queries. We applied each technique to visualisations and implemented them.

3.1 Defining the Workflow

The visual workflow is streamlined through a coordinated view of the two different parts centralizing the link focused on a specific resource that binds them. This way, our workflow enables users to discover, search and analyze LOD. Figure 1 shows how users start with an overview of the dataset (Figure 1*a*) through which the users "dive" in more narrow perspectives (Figure 1*b*) by selecting a group to find out details and see the internal relations of the subdivisions (Figure 1*b*). A coordinated view (Figure 1*c*) of selected resources leads them through a broadened view (Figure 1*d*) by exploring relations of these resources.

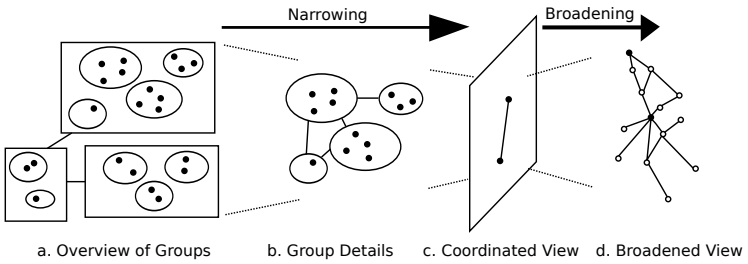

Fig. 1: Narrowing views (*a, b*) allow data consumers to analyse the dataset. The coordinated view (*c*) allows perspective switching in the workflow. Broadening views (*d*) allow data consumers to explore the interlinked information beyond the dataset's boundaries.

a,b. Narrowing Views. The narrowing views aim to familiarize the data consumer with a certain dataset, as they are not aware of its context. Since there is no explicit assumption regarding its content, as EDA ordains, the dataset itself reveals its underlying model and the relationships between its resources. Given the "unlimited" extent of a dataset, the initial view is focused on this certain dataset and its broader concepts are demonstrated. Deeper exploration, following the links until reaching the resources that can not be further decomposed, helps to better comprehend the available data.

c. Coordinated View. As the users, supported by the visualizations, narrow down to more detailed resources (a certain resource or the links between two resources), they reach the resources that cannot be further decomposed and thus act as the coordinated view. Starting from this view, data consumers, being aware of the underlying dataset, start exploring the dataset. The coordinated view form a "bridge" between the narrow view and the broader view, which exploits existing links amongst resources across different datasets.

d. Broadening Views. In the case of *broadening* views, data consumers find novel relations between existing and known resources interacting with the visualizations of

the LOD. The possible views are not limited to the data of the narrowing view but the links to other datasets are also revealed and visualized if considered relevant. It is a new way to search and explore the information. This way, users get an overview by using an approach that visualizes interactively search process in an aligned linked data knowledge base of related resources.

3.2 Applying Information Exploration Techniques to the Workflow

The narrowing view is achieved based on exploratory analysis techniques [18] applied to the dataset. Without any formal modeling or assumption about the underlying dataset, the main concepts and their relationships are gradually revealed. Subsequent views narrow the broader concepts and reveal more details about the relations among the concepts. The broadening view is achieved using exploratory search techniques [14] over individual instances of the LOD. Users iterate over individual concepts, their direct neighbors and their relationships. Iteratively expanding and focusing the visualization leads to more insight into selected concepts in the datasets.

4 Implementation

We developed a demo where we implemented each step of the workflow: the narrowing, coordinated and broadening views; as described in Section 4.1. The visualizations used complementary datasets (see Section 4.2) whose intersection serves for the coordinated view. In Figure 2 we apply the workflow to a use case of academic library metadata which is generalizable to other types of LOD and domains.

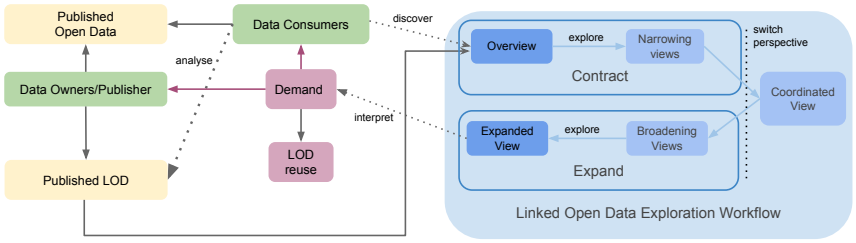

Fig. 2: The visualizations are part of the LOD exploration workflow which can be applied for the exploration of different datasets, in this case academic library metadata.

4.1 Visualization Tools

We implemented the workflow using two tools we developed for visualizing LOD. This results in the graph based exploration interface supporting narrowing views by *LOD/VizSuite* and broadening views by *ResXplorer* over a coordinated view.

LOD/VizSuite The goal of the LOD Visualization Suite (LOD/VizSuite)[6] is to create an easily customizable visualization framework on top of LOD. LOD/ VizSuite aims to be data and schema agnostic, therefore it can be easily transferable to visualize different datasets. Its functionality is based on SPARQL queries which are published as SPARQL templates. Parameters can be passed to the SPARQL template at request time, which replace placeholders to construct a valid SPARQL query. The SPARQL templates are published at a DataTank[7] instance, a RESTful (Linked) Open Data management system

[6] http://ewi.mmlab.be/academic
[7] http://thedatatank.com

which publishes data on the Web. LOD/VizSuite exposes research and collaboration networks, communities of practice in a certain discipline and timelines to monitor a discipline's evolution over time [7].

ResXplorer The goal of ResXplorer[8] is that researchers can find novel relations between existing known items such as authors, publications or conferences. Users interact with a visualization of resources [5] using an interface combining an optimized pathfinding algorithm (the Everything is Connected Engine) [4] with Web 2.0 technologies (such as JQuery and Django). The result is a semantic search tool providing both a technical demonstration and a visualization that is applicable to many other applications beyond academic library metadata.

4.2 Datasets

LOD/VizSuite provides visualizations based on the LOD provided by the *"Research Information Linked Open Data"* (RILOD)[9] data-set. RILOD is the result of the integration of heterogeneous sources related to research in Flanders, ending up in a rich and diverse dataset. The datasets contains resources of researchers from the region of Flanders, their publications and projects, which are associated with the corresponding research groups and institutes, and classified under the IWETO Discipline classification[10]. ResXplorer makes use of the *"Digital Bibliography and Library Project"*[11] (DBLP), an on-line reference for bibliographic information on major computer science publications [13]. The binding between RILOD and DBLP is their content's intersection: the same researchers and publications appear in both datasets.

4.3 Embedding Visualizations in the Workflow

In this section, we explain how we embedded the visualizations to implement the three types of views of our exploration workflow. Figure 3 shows the visualizations embedded in the workflow.

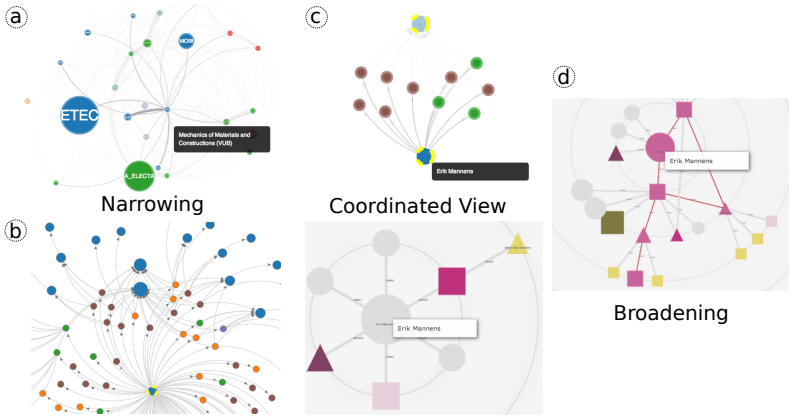

Fig. 3: Corresponding with steps *a, b, c, d* in Figure 1, users narrow down from disciplines (*a*) to research groups and further to the individual researchers in this group (*b*). To find out relations between researchers they select two researchers and, using the coordinated view (*c*), shift to the broadening view and expand to resources beyond their research community (*d*).

[8] http://www.ResXplorer.org
[9] http://ewilod.be/ewilod/html/sparql-test.html
[10] http://ewilod.be/ewilod/lod/0.1/ontology/taxonomies/iwetoDisciplineCodes
[11] http://dblp.l3s.de/

The broadest concepts, which cover all the dataset, are chosen for the *overview view*. The *overview view* serves the users to discover the *main concepts* of the dataset, the *strength of the relations* between them (Figure 3a) and the *diversification of the total number* of the instances that constitute the broader concept. From the overview, the users discover the narrower entities (Figure 3b). *Broader views* are achieved by aggregating narrower entities using SPARQL queries that select and group them. In our use case, visualizations that provide an *overview view* of a research discipline are achieved by aggregating the researchers under their research groups and providing their collaboration links considering their co-publications. The research groups are demonstrated as graph nodes that diversify in size depending on the total number of projects they have, while the strength of the links depends on their researchers' co-publications.

As a research group is the aggregation of individual researchers, an end user can further narrow down and view the researchers and their publications (*decomposed views*). This is the narrowest view which acts as the *coordinated view* (Figure 3c). In our use case such a resource can be an individual researcher or the links between two researchers whose extensive collaboration network is demonstrated. As the end users view the network formed around a researcher or the exhaustive list of paths between two researchers, they can be transposed to the corresponding view of the broadening part of the workflow.

While exploring the *broadening view*, data consumers are not limited to the data of the dataset but their exploration is enhanced with links to other datasets of the LOD cloud that might be relevant to their exploration (e.g., DBLP in our use case). We show that visualizing resources, such as conferences, publications and proceedings, expose affinities between users and those resources (Figure 3d). We characterize each affinity, between users and resources, by the amount of shared interests and other commonalities.

5 Evaluation

We evaluated the workflow at "iMinds The Conference"[12], a yearly gathering for researchers active in several aspects of digital innovation. The conference's audience was identical to the use cases' target groups, namely researchers and R&D policy makers. Data was gathered using a multi-method approach. We used the think-aloud protocol during the experiments to collect feedback from the participants and we recorded the screen actions of participants using QTrace[13].

Test users received no information about the tool in advance. They were asked to execute some assignments and to fill in a questionnaire afterwards. During the user test, users were asked to think aloud. The observer recorded comments and took notes. In the additional framework, we asked people coming over at the booth, but who did not participate in the user test, to fill in the same questionnaire after receiving an explanation about the workflow and the tools. The survey included (i) perceived aims of each step in the workflow; (ii) a series of statements on a five-point Likert scale measuring the usefulness, learnability, complexity, explorability, transparency as perceived by the respondents; and (iii) open auestions on insight in quality of the used data and suggestions. Each test took about 30 to 45 minutes. The questionnaire took an additional 5 to 10 minutes to be completed. Apart from the 17 users who participated in the evaluation, 17 additional users participated in the evaluation by filling out the same questionnaire after receiving information about the visualizations, giving us richer data for the survey analysis. In total, 34 users participated in the survey.

[12] http://conference2013.iminds.be/

[13] http://www.qasymphony.com/qtrace.html

We kept the audience of the assessment broad by conducting also semi-structured interviews with various stakeholders. All of them are likely to be affected by the impact and value of accessible and explorable LOD. The use case was situated in the context of research information. Thus interviewees are active for the Flemish government department of Research and Innovation, the Department of Research policy from Ghent University, and, from a commercial point of view, in the domain of Business Development & Academic Relations.

In our previous work [6], we evaluated how appealing our visual workflow and the visualization tools are to the end users by assessing the productivity and precision of the narrowing part and the complexity and searchability of the broadening view. Our evaluation showed that the implemented visualisations were capable of assisting the end-users to interpret the visualisations, thus adequate for the scope they were designed. In this work, we evaluate the workflow in regard to the scope it was implemented for.

5.1 Visualizations in the Workflow

We observed how the test users executed the assignments and we asked them to think aloud. The test users were asked (i) to start from their preferred research discipline (overview view), (ii) to go on towards their preferred research group and researchers and, explore their collaborations (explore the links of the narrower views) and (iii) to explore the links of one of the researchers that they concluded at while they navigated to broader views (broadening view).

Observations. The think-aloud analysis gave us information regarding the perception of the visualisations by the participants, during the executions of the assignments. The observers asked test users to search for their own name; or when it did not appear in the dataset; to searh for a research subject they are familiar with (for example related to their own research). Users had to indicate differences in the size of nodes and they received an explanation on the meaning of the node sizes. Afterterwards they had to indicate if the visualization appeared to them as a good representation of the real-world situation. Via this direct feedback, we concluded that test users are able to reason via the tools: for example or by appointing missing research groups in the visualizations, by putting the size of the nodes into discussion in LOD/VizSuite.

The observations give us further insights regarding how the users expect the exploration to happen: clicking on the broad views, e.g., clicking research groups within disciplines, they expect they get an intermediate overview and each step forward in the workflow can give additional input to explore. Once they realize the fact of the narrowed view and the effect of the coordinated view, data consumers are able to fully comprehend the workflow and start with simple reasoning that supports the intention of the exploratory search tool. We observed that, once test users comprehend the exploration workflow, they better accept the visualized data and become more able to form their exploration path. This affects their exploration behavior: they use the different features to get further insights (search queries, top affinity suggestions, or expanding via node clicking). Although complexity rises within the visualizations during the explorations (earlier explored data stay visualized and taken into account), test users understand the potential of visualizing academic data and can name how they are related to another researcher via conferences or publications from intermediate researchers. Test users declared that further input could bring in additional points of interests.

Survey. To evaluate the exploration of the LOD, we asked the test users and twenty extra respondents their impression of the views using a questionnaire. To determine the impact and quality of the workflow considering their use for data consumers, we analyzed how the users explored and perceived the visualizations in the corresponding views. We especially measured the perceived usefulness and learnability and how the participants estimate the potential of the visualizations.

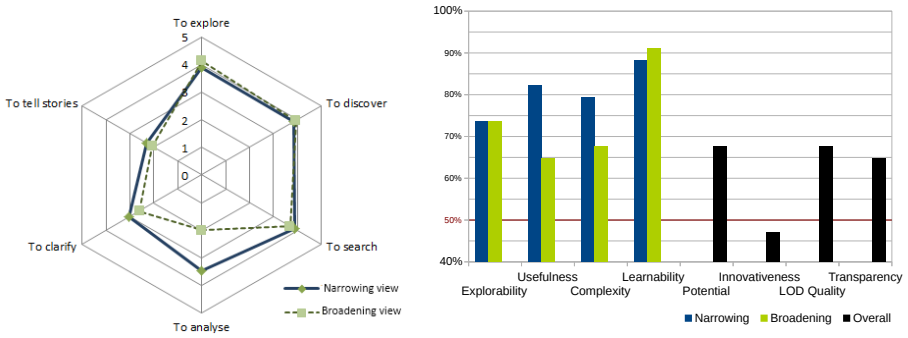

Fig. 4: **User perceived goals of the views (left)** indicate that the narrowing view is perceived to be more suitable for analysis while the broadening view scores slightly better for exploration. **User satisfaction for the workflow (right)** shows that overall the views don't seem to expose innovation. In terms of uselness and complexity the users are very satisfied with the narrowing, they need some time to learn how the broadening works.

Visual workflow's goals. To understand how the users perceive this visual workflow and its goals, we asked them to score possible purposes of use. As displayed in Figure 4, the respondents indeed perceived both the narrowing and broadening views as adequate tools to explore, discover and search. The broadening view is considered by the respondents as being a tool for exploration in the first place and discovery in the second place. The narrowing view is considered as a means to explore and to search.

Usefulness and Explorability. Test users agree that the visualizations is useful in terms of what it exposes 22 out of 34 (65%) agree for the broadening views and 28 out of 34 (82%) for the narrowing views. 28 out of the 34 respondents (74%) agree or strongly agree that they understand the displayed relations of the broadening view are presented as an optimized selection of all results. Although, respondents stay rather undecided when it comes to the limitations: 16 out of 34 respondents (47%) agree or strongly agree that it is useful that the number of visualized resources and relations are fixed to the 7 most relevant resources, whereas 11 out of 34 (32%) disagree or strongly disagree on this. Finally, the respondents strongly agree that both the broadening view and the narrowing views support them gaining insights into the published data, but they were less confident in the case of the narrowing view.

Complexity and Learnability. The majority of the respondents agreed that they can interpret the complexity of the visualizations both for the narrowing views, 27 out of 34

(79%), and the broadening views, 23 out of 34 (68%). Most of the respondents agree with the statement that after a learning period (we did not measure this) they get familiar with the visualizations in the narrowing view and they can get benefit out of it 30 out of 34 (88%) and even more of them agree for the broadening exploration 31 out of 34 (91%).

Workflow Potential and Innovativeness. The respondents were asked how they perceive the potential of the workflow for LOD exploration. 23 of 34 test users (68%) state that the visual exploration workflow clearly helps them to understand the potential of Open Data. 16 out of 34 (47%) respondents agreed or strongly agreed that the visualizations help to get insight into innovation. However, 11 out of 34 (32%) respondents remained undecided.

Linked Open Data Quality and Transparency. 23 out 34 (68%) respondents agreed that the workflow can serve as an encouraging factor to provide more open data and of better quality. The workflow imposes the data consumers to behave in a different and unfamiliar way, but at the end, they get more familiar with the underlying LOD which can bring in new unexplored information. This can explain the mixed response on the visualizations as an encouraging factor to gain data of better quality. 22 out of 34 respondents (65%) agreed that the government becomes more transparent if the public is able to explore the pubished LOD with such visualizations. Once they get familiar with the narrowing and coordinated view, our respondents state it helps them in the discovery and exploration of the data. The broadening part helps them to gain new insights in the dataset and explore links with data published to the broader LOD cloud.

5.2 Workflow applied to Linked Open Data and Academic Metadata

The observations lead us to the conclusion that wider *ecosystems* need to be taken into account to analyze and facilitate value creation around Linked Open Data. The term *ecosystem* is applicable here, because it emphasizes interconnections and interdependencies between the different actors and roles.

As illustrated in Figure 2, consumers can discover published LOD and its meaning through the exploration workflow. As this adds value to their own activities, they are likely to increase the demand, and address data publishers to improve respective infrastructure. To assess the potential of LOD, the process of linking open data in a broader multi-stakeholder context needs also to be taken into account. For this assessment, Linked Open Government Data (LOGD) [8] and the significance of visualization and exploration for governments and policy-making plays a decisive role. Geiger and Von Lucke [9] evaluated opportunities for government as a positive paradigm shift, for instance, new ways of political legitimation or innovation and modernization of administration in an increasingly open world. Higher scalability and quality in the production of LOGD entails that governments need to be seen both as data supplier and as data consumer [10].

Here, the much discussed impact of LOD as a driver for economic growth and innovation is most applicable[14] . Thus, each actor must have different incentives for participating in LOD initiatives: Governments might aim for efficiency; companies for generating revenue; universities for better research etc. However, the interaction and interdependence of all involved parties can finally impact on more data, with better quality, and benefits for all affected parties. This potential is not always revealed because of various challenges which form roadblocks to the realization of that potential. The problem

[14] http://www.slideshare.net/DeloitteAnalytics/open-data-13509015

is that, even when data is published adequately, there will neither be value nor impact arising from the development, as long as citizens, businesses, public servants etc. do not use it. This means that the challenges of achieving the potential, which LOD clearly holds, go beyond its mere publication.

An interview, undertaken with the responsible functionary of the initiating government department, clarified their motivations in relation to Open Government Data. To achieve the desired objectives, a Linked Open Data environment for research information was created, based on the spirit that data, which is publicly financed, should be properly accessible to the public. However, its capacities were still limited and information was not as complete as it could be. Data was published in an infrastructure to facilitate exploration and ultimately to add value, but still accessibility remains curtailed, in particular for non-expert users. Therefore, the development of visual exploration workflows on top of the LOD was encouraged, providing accessibility to complex data, and understanding of the meaningful correlations emerging from the links.

6 Conclusions and Future Work

Interactive user interfaces based on enhanced visualizations allow users to have a unique, multifaceted experience when combined with techniques for information exploration. Such interface is demonstrated in our graph-based workflow for the exploration of a dataset. According to our respondents, they, as data consumers, become acquainted with the underlying dataset and the workflow can bring in new unexplored information and knowledge as soon they familiarized themselves with the workflow. The respondents, all active in the occupation of academics (namely domain experts) remained satisfied, and thus we verify that the workflow is applicable to academic library metadata and generalizable to the exploration of any other LOD. Overall, offering such visualizations on top of LOD: turns the potential of expressing a demand for LOD exploitation more likely; and it increases the demand for data of better quality, both in terms of context and semantics.

Considering the results of the demo's evaluation, we are planning an integrated solution that aligns with the users' workflow as they explore the information. Additionally, we would like to apply the proposed exploration workflow to other datasets and evaluate its applicability to other domains such as cases of exploring other type of actors and their actions' output but also even broader. Considering the test users' feedback, we will improve several aspects of the implementation. A task based evaluation, could show that the proposed workflow indeed facilitates tasks that can not be carried out otherwise, therefore reinforcing the "potential of linked data". We plan to support multigraphs introducing multiple relations between displayed resources and subgraphs and direct expansion of the aggregated resources in order to achieve a more consistent exploration over the available data. We aim to keep the exploration workflow as data and schema agnostic so it can be easily transferable to other datasets. Finally, we plan a thorough evaluation of the integrated workflow applied to diverse datasets.

7 Acknowledgement

The research activities described in this paper were funded by Ghent University, the Flemish Department of Economy, Science and Innovation (EWI), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

References

1. J. M. Brunetti, S. Auer, and R. García. The linked data visualization model. In *International Semantic Web Conference (Posters & Demos)*, 2012.
2. C. Collins and S. Carpendale. VisLink: Revealing relationships amongst visualizations. *IEEE Trans. on Visualization and Computer Graphics (Proc. of the IEEE Conf. on Information Visualization (InfoVis))*, 13(6), 2007.
3. A.-S. Dadzie and M. Rowe. Approaches to visualising Linked Data: A survey. *Semantic Web*, 2(2), 2011.
4. L. De Vocht, S. Coppens, R. Verborgh, M. Vander Sande, E. Mannens, and R. Van de Walle. Discovering meaningful connections between resources in the Web of Data. In *Proceedings of the 6th Workshop on Linked Data on the Web*. CEUR-WS, 2013.
5. L. De Vocht, E. Mannens, R. Van de Walle, S. Softic, and M. Ebner. A search interface for researchers to explore affinities in a Linked Data knowledge base. In *Proceedings of the 12th International Semantic Web Conference Posters and Demonstrations Track*. CEUR-WS, 2013.
6. A. Dimou, L. De Vocht, M. Van Compernolle, E. Mannens, P. Mechant, and R. Van de Walle. A Visual Workflow to Explore the Web of Data for Scholars. In *Proceedings of the 1st Workshop on Big Scholar: Towards the Web of Scholars*, 2014.
7. A. Dimou, L. De Vocht, G. Van Grootel, L. Van Campe, J. Latour, E. Mannens, P. Mechant, and R. Van de Walle. Visualizing the information of a Linked Open Data enabled Research Information System. In *Proceedings of the Current Research Information Systems Conference*, 2014.
8. L. Ding, V. Peristeras, and M. Hausenblas. Linked open government data. *Intelligent Systems, IEEE*, 27(3):11–15, 2012.
9. C. P. Geiger and J. von Lucke. Open Government and Linked Open Government Data. *Journal of eDemocracy and Open Government*, 4(2), 2012.
10. A. Graves. Creation of visualizations based on Linked Data. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*. ACM, 2013.
11. Y. Hu, K. Janowicz, G. McKenzie, K. Sengupta, and P. Hitzler. A linked-data-driven and semantically-enabled journal portal for scientometrics. In *The Semantic Web - ISWC 2013*, volume 8219 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013.
12. M. Janssen, Y. Charalabidis, and A. Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 2012.
13. M. Ley. The DBLP computer science bibliography: Evolution, research issues, perspectives. In *String Processing and Information Retrieval*. Springer, 2002.
14. G. Marchionini. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 2006.
15. J. C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*. IEEE, 2007.
16. M. Scherr. Multiple and coordinated views in information visualization. *Trends in Information Visualization*, 2008.
17. J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: Extraction and mining of academic social networks. KDD '08. ACM, 2008.
18. J. Tukey. *Exploratory Data Analysis*. Addison-Wesley series in behavioral sciences. Addison-Wesley Publishing Company, 1977.
19. M. Tvarožek et al. Exploratory search in the adaptive social semantic web. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, 3(1), 2011.
20. K. Verbert, D. Parra, P. Brusilovsky, and E. Duval. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, 2013.
21. R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 2009.