Editorial

Visualisation of Linked Data – Reprise

Aba-Sah Dadzie [a,*] and Emmanuel Pietriga [b]

[a] *KMi, The Open University, UK*
E-mail: aba-sah.dadzie@open.ac.uk
[b] *INRIA, France*
E-mail: emmanuel.pietriga@inria.fr

Abstract. Linked Data promises to serve as a disruptor of traditional approaches to data management and use, promoting the push from the traditional Web of documents to a Web of data. The ability for data consumers to adopt a follow your nose approach, traversing links defined within a dataset or across independently-curated datasets, is an essential feature of this new Web of Data, enabling richer knowledge retrieval thanks to synthesis across multiple sources of, and views on, inter-related datasets. But for the Web of Data to be successful, we must design novel ways of interacting with the corresponding very large amounts of complex, interlinked, multi-dimensional data throughout its management cycle. The design of user interfaces for Linked Data, and more specifically interfaces that represent the data visually, play a central role in this respect. Contributions to this special issue on Linked Data visualisation investigate different approaches to harnessing visualisation as a tool for exploratory discovery and basic-to-advanced analysis. The papers in this volume illustrate the design and construction of intuitive means for end-users to obtain new insight and gather more knowledge, as they follow links defined across datasets over the Web of Data.

Keywords: Linked Data, linked open data, visualisation, visual analysis, exploratory discovery, visual querying

1. Introduction

Linked Data is a key component of the Semantic Web vision. The ability for data consumers to adopt a *follow your nose* approach, traversing links defined within a dataset or across independently-curated datasets, is an essential feature of what we now call the *Web of Data*, enabling richer knowledge retrieval thanks to synthesis across multiple sources of, and views on, inter-related datasets. Since its early days, Linked Data (LD) promised to serve as a disruptor of traditional approaches to data management and use, promoting the push from the traditional *Web of documents* to this *Web of data* [3,4,40]. But the advantages of following LD principles have now become even more apparent [21,39], as the boundary between the data available on the (online) Web and users' own (personal) data gets ever more fuzzy. Users in our in-

creasingly data- and knowledge-driven world are becoming more and more dependent on applications that build upon the capability to transparently fetch heterogeneous yet implicitly connected data from multiple, independent sources. It has become clear that we must rethink how those users interact with the very large amounts of this complex, interlinked, multi-dimensional data, throughout its management cycle, from generation to capture, enrichment in use and reuse, and sharing beyond its original context.

The Linked Data community introduces its foundational concept as follows [30]:

> "*Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.*"

This sentiment is echoed across academia, state organisations and industry [see 2,3,13,22,39,40,53,54, among others]. But despite the initiative's recognised

---

*Corresponding author. E-mail: aba-sah.dadzie@open.ac.uk.

value, and even though it is now well past its infancy, Linked (Open) Data[1] extension and adoption beyond the Semantic Web (SW) community remains limited. This is due largely to three key challenges:

1. The very large amount of available LD, which keeps increasing [see, e.g., 27,39,54]. For instance, as at mid-2016 Ermilov et al. [14] report 9,960 RDF datasets containing over 130 billion triples, with connectedness up to 40% from 3% in 2011. These measures originate from *LOD-Stats* [34], which provides the most up-to-date set of statistics about LOD datasets referenced on *datahub.io* [50] and selected public and open government stores.

2. The inherently heterogeneous nature of LD, comprising independently-curated datasets backed by both standard and custom ontologies, with variation in granularity, quality, completeness and adherence to the five-star model of Linked Data design [3]. This is further complicated by differences in source and target domains, and the impact this has on data structure and content.

3. The focus in the SW community on automated, machine-driven data processing, which has so far resulted in data management and use solutions that require an advanced level of technical know-how to properly generate, clean and link new data into the Linked Data cloud. Importantly, this has also translated into the need for a correspondingly advanced level of technical know-how at the other end of the workflow, in end-user-driven exploration, analysis and presentation to diverse audiences of LD-backed content. Even though custom tools can be, and often are, developed for specific use cases and datasets [2,9,37,47,54], this situation constitutes a significant obstacle to the adoption of LD as a framework by a wider range of data producers and consumers.

Linked Data already spans a wide range of application areas, a strong indication that its potential value is already largely acknowledged. The open challenges mentioned here, among others, are being tackled by the research community which, beyond the publication of academic papers, is developing and making publicly available numerous pieces of software to test different approaches to the management and use of LD. The adoption and promotion of LD can also be observed at institutional and governmental levels as part of various *open data* initiatives, as well as in industry, where efforts include investigating LD as a way to more efficiently handle *big data* [39,40].

At the simpler level of individual data consumers, *lay* (non-expert) users may eventually employ Linked Data as a portal to timely, context-based knowledge for everyday activities, doing so either implicitly or explicitly. Such activities could range from simple ones, such as comparing the price of groceries or planning a journey using public transport, to more complex ones involving important decisions, such as purchasing life insurance or a new home. LD has the potential to make such activities more efficient by giving users access to a rich Web of interlinked data relevant to the task at hand; however, this should not be accompanied by technical barriers that are difficult to overcome by a majority of users.

The design of user interfaces for LD [22,24,54], and more specifically interfaces that represent the data visually, play a central role in that respect. Well-designed visualisations harness the powerful capabilities of the human perceptual system, providing users with rich representations of the data. Combined with appropriate interaction techniques, they enable users to navigate through, and make sense of, large and complex datasets, help spot outliers and anomalies, recognise patterns, identify trends [7,49]. Visualisation can provide effective support for confirming hypotheses, and favours deriving new insight. Visual analytics systems complement human cognitive and perceptual abilities with automated data processing and mining techniques [9,39,47,55]. Such systems not only support the presentation of data, but processes for the full data management cycle as well, from data capture to analysis, enrichment and reuse.

However, visualisation in and of itself is not an absolute solution to the challenges inherent to complex data sense-making, analysis and reuse. To contribute to the realisation of LD's potential, visualisations have to be tailored to specific tasks, and effectively support users in the performance of these tasks by providing visual representations of the data at a relevant level of detail and abstraction. Following Shneiderman's mantra [49], users should be able to navigate from an overview of the data to detailed representations of specific items of interest. But the system should also provide **relevant** representations of the data at these various levels of detail [7]. This is a major challenge, considering the *follow-your-nose* approach to data exploration mentioned earlier, and given the potentially

---

[1]We use LD and LOD interchangeably throughout this paper, the latter mainly to stress where data is also open to public (re)use.

high level of heterogeneity of the interlinked datasets involved in most LD-backed applications. Only then will the results of visual exploration and analysis lead to new insights about LD, with an increased level of confidence in drawing conclusions from the data and the supporting visual analysis, and in decision-making based on the knowledge thus acquired.

We review the state of the art in Section 2, and look at increasing efforts to evaluate front-end tools with target end-users in Section 3. In Section 4 we discuss on-going research in Linked Data visualisation and the range of challenges tackled by contributors to the special issue, as part of the process followed in designing and building effective visual presentation and analysis solutions for Linked Data. We conclude in Section 5 with a vision for the role of visualisation in advancing research in, analysis of, and applications based on Linked Data.

2. The state of the art: Linked Data visualisation

Tackling the challenges associated with the representation of Linked Data and its visual exploratory analysis requires considering related, supporting technologies. Halpin and Bria [17], Klímek et al. [27], Mai et al. [37], Mitchell and Wilson [39], for instance, highlight reliance on data structuring, typically using ontologies as a backbone; the identification and encoding of links via ontology mapping and alignment; and the need to integrate such technologies into the visual discovery and analysis *pipeline*. Formal querying is a powerful tool for extracting specific information either from one or across multiple datasets. However, formulating any but very simple queries in SPARQL (the *de facto* RDF query language [52]), or any other formal query language, is tedious even for SW experts. It is also error-prone, especially with each additional graph and optional or required relationship [9,39]. Further, the textual results returned by query engines, which typically take the form of machine-readable RDF, are not easily interpreted by users, even if they have a good level of domain understanding. Even when considering expert users, meaningful overviews are required to obtain a good understanding of data structure, content and relationships for any but very small result sets. Predefined query templates serve as an aid here, and as a foundation on which more intuitive support for guided querying and question-answering may be built.

Valsecchi et al. [54] demonstrate the advantages gained in identifying suitable entry points to LD exploration, as a key initial step to improving access, especially for non-SW and non-domain experts. Benedetti et al. [2], Haag et al. [16] discuss the need to support visual querying to overcome technical challenges in the formulation of formal queries by lay users and, as complexity grows, even technical experts. Mai et al. [37] further demonstrate the added value in providing multiple visual perspectives that support the *follow your nose* paradigm when exploring query results across multiple, heterogeneous, albeit linked, datasets. Halpin and Bria [17] describe the process followed to map how Digital Social Innovation is being used throughout the European Union (EU) across multiple aspects of people's lives, and measure the resulting social impact. They illustrate the benefits gained in employing SW technology to overcome several challenges – challenges posed by differences in structure, granularity and completeness of the data sources involved. Indeed, multiple sources have to be queried to capture all data required to create a DSI map (of the EU), a process that may be used to generate a linked open dataset synthesising this data.

Beyond tools that help initiate exploration by identifying entry points into LD, users should also be provided with support for identifying and selecting more focused resources. Depending on the use case, this might involve increasingly customised (or customisable) tools for in-depth exploration and analysis, along with support for reviewing content and sharing both this content and the results of discovery and analysis beyond its initial point of generation or use. Toward this aim, it may be instructive to follow best practice and verified frameworks and models for the design of visualisation and visual analytics such as in [47,49,55], adapting these where necessary for the specific case of LD and SW data in general. As in [9], we use Shneiderman's seminal *visual information-seeking mantra* as a guiding principle in evaluating the usability and utility of visualisation-driven approaches and tools for different users and their tasks – looking at support for: (i) generating overviews on the underlying and related data; (ii) filtering out noise and less important data in order to focus on ROIs (Regions of Interest); and (iii) visualising ROIs in detail.

We review selected work in the SW community that illustrate advancement of the state of the art in Linked Data visualisation in the five years since the 2011 survey about LD visualisation [9], looking at models, approaches and tools that focus on one or more of the following visualisation-driven tasks:

– identification of suitable entry points into this large, complex data space,
– navigation inside individual, and across multiple, linked datasets to support exploratory discovery,
– analysis of data structure and alignment,
– basic to advanced data querying and question-answering,
– sense-making and guided, in-depth analysis,
– content enrichment through data annotation and identification or derivation of new links,
– presentation and sharing of data and results derived from their analysis to different audiences.

This review of the state of the art differs from the 2011 survey [9] in two key areas. First, here we look in more detail at the larger visual exploration and analysis process, examining supporting technologies that, e.g., minimise how much pre-processing is required before the data can be visualised. Secondly, we consider text-based browsers only where they make significant use of graphics to generate effective visual representations. This is not to detract from the value and utility of text-based browsers in general, but to allow more in-depth focus on the merits and challenges of visualisation-driven approaches.

Sections 2.1, 2.2 and 2.3 look at the roles of each of (i) ontologies, (ii) formal querying and exploration and (iii) content analysis and presentation respectively, and their role in the broader task of LD visualisation. However, the discussion that follows shows a good amount of overlap between all three aspects, reinforcing the benefits in integrating these component parts of the process and/or building seamless pipelines between them.

2.1. The role of ontologies in Linked Data visualisation

Dowd [13], Hitzler and Janowicz [21], Ivanova et al. [24], Mitchell and Wilson [39], Welter et al. [58], among others, highlight the importance of ontologies in the effective use of SW technology and other data-driven research and applications. Ontologies provide structure and a practical means to capture, encode, categorise, index and retrieve data. Focusing specifically on LD, ontologies are a critical element in linking independently-created datasets. By also supporting the encoding of colloquial, domain- and community-specific use of language and terminology using commonly-agreed, formal terms and concepts, ontologies aid correct interpretation of content

and identification of both implicit and explicit links across data using non-standard or unambiguous language and terminology [18].

Research on and about the use of ontologies is well established and significantly predates LD. However, employing ontologies to structure LD and support the creation of links within and between linked datasets, as with other ontology engineering tasks, is mostly restricted to technical and domain experts. This is often because of the dearth of effective user interfaces (UIs), which are needed to reduce complexity in all but the simplest of ontological structures. This complexity increases when considering the alignment and mapping of ontologies and LD [24,27,36,37], and impacts correctness and interpretation of data and links between those data. While ideally all LD would follow the five-star LOD design guidelines [3], in practice, the quality and granularity of data on the Web varies significantly. Ontologies are especially useful in such cases, where independently curated datasets with incompatible structure and representation – but with related content – must be linked. Research on ontology visualisation is therefore both relevant to, and often closely linked to, research on LD visualisation and interactive analysis.

Weise et al. [57], for instance, extend VOWL, the Visual Notation for OWL Ontologies [36], to visualise linked dataset schemas. They aim at supporting users in getting an understanding of LD content, especially where datasets do not conform strictly to defined ontological structures or other defined schemas. Their web-based tool, LD-VOWL, follows principles of ontology engineering to extract schema information using a set of SPARQL queries, and visualises the results using a graph layout. Kazemzadeh et al. [26] demonstrate the benefits of defining a unifying ontology to allow the synthesis of biological data, to support more complete and far-reaching analysis of the large amounts of detailed-but-disconnected experimental data generated in this field. Halpin and Bria [17] collated data to map digital social innovation across the EU. They concluded that an ontology provided the most flexible means for capturing, in a single structure, the varied information required to map this information space. They also used ontologies as a support for queries required during the subsequent analysis steps, involving the Linked Data generated by merging the multiple data resources describing this activity. Further, they found that structuring data using an ontology allowed them to continue to add new data "on the fly", including their own survey data. Finally, it also enabled them

to create social networks from the information gleaned about individuals and organisations, due to the inherent network structure of the LD encoded in RDF.

We see a cyclical relationship between ontologies and LD, with ontologies guiding initial structuring during data capture, and the structured LD feeding back into ontology evolution. This can be seen as a virtuous circle, where LD and the backing ontologies contribute to refining one another. This process further aids alignment and linking of disparate, yet related datasets that reuse the ontologies in question.

2.2. *Visual querying & exploratory discovery*

Haag et al. [16], recognising the challenges faced by users in querying LD and RDF in general without good knowledge of the SPARQL query language [52], extend the concept of "Filter/Flow Graphs" [59] to enable relatively sophisticated visual SPARQL querying. The authors carry out a usability evaluation with a small group of users familiar with the Semantic Web, but who were not SW experts. Their findings demonstrate high learnability of the UI, with users self-correcting as the visual representation illustrates the effect of flow construction on the result-set obtained. Benedetti et al. [2] in the *LODeX* prototype, which is targeted at both non-technical users and technical experts, explore requirements for easier construction of SPARQL queries. *LODeX*, which serves as a query construction guide for users, combines a detailed table view on focused elements with a graph layout summarising the schema of linked datasets.

Mai et al. [37] tackle the challenges inherent in linking and exploring "data silos". Among these challenges, they study the need for seamless transition between different perspectives on heterogeneous data, and questions related to fusing differences in data in terms of accuracy and granularity, by using a federated approach to exploration. By following the links within a dataset and across datasets to related data, they provide user support for successively revealing more detail about an entity of interest while switching between coupled representations of query results as tables and graphs. When items in the result-set feature geolocation information, the data may additionally be plotted on a map.

Valsecchi et al. [54] use a cartographic metaphor to generate overviews, aiming to help users obtain an understanding of data structure and content. They rely on interaction techniques that follow Shneiderman's visual information-seeking mantra [49]. The visual lay-

out maps types and sub-types defined in backing ontologies to pseudo-geographical *regions* and their subdivisions. Instances of a type are mapped to *cities* within their (parent) region. Disconnected (isolated) types and untyped data lie on *islands* outside the *mainland* cluster of inter-linked data. From this initial overview, keyword search and interaction with the map (including click and zoom) transparently trigger SPARQL queries to retrieve more detailed information about ROIs and support jumping via links to other related entities in the same region or in other regions. The user may swap between layers over the base map to reveal additional visual cues about data density and structure (e.g., data and object properties, class depth within the ontology).

Similarly, Halpin and Bria [17], mentioned earlier, use a map of the European Union to represent information about digital social innovation (DSI) captured as Linked Data. Complementing SPARQL queries backed by a dedicated ontology with a combination of map-based representation and social network layout showing the links between collaborating organisations, they were able to carry out a relatively detailed analysis that revealed both new information about DSI and areas on the map where information was missing. The ultimate aim of the project is to feed the results of the analysis into EU and state policy on opening up access to funding and other resources to improve DSI.

2.3. *Visual content analysis & presentation*

Being a large, continuously growing, heterogeneous, multi-dimensional collection of independently generated, albeit inter-linked, datasets, LOD contributes to today's *big data* [39], bringing with it both value and challenges for content analysis. One aspect of managing LD involves the collection of statistics and other metrics that describe the data itself, helping to obtain an overview of content and structure and a means for identifying points of entry into the data. Open review and access for the Semantic Web Journal [23] allow capture of the full review and editorial process in addition to metadata (e.g., topics) about paper content as Linked Data. Hu et al. [22] illustrate the benefits in this approach that simplifies linking to further information such as author networks, to enable, among others, tracking current and predicting future trends in research. The authors use a *semantically-enabled journal portal* [31] to present scientometrics derived from analysis of data retrieved from the SPARQL endpoint for this dataset and the results of

statistical, trend and topical analysis. The interactive portal uses a number of visualisation types in a dashboard to provide multiple perspectives for browsing content and the metrics derived from the data. By linking to other LD describing research activity, the dashboard also illustrates spatial distribution of author influence based on network and citation analysis.

The field of Bioinformatics emerged as a result of the introduction of computational techniques to aid the analysis and use of the very large, open, multi-media databases generated in the Biological and Life Sciences, due to the introduction of advanced technology for experimentation. This has enabled significant research and discoveries such as the mapping of the human genome [51], which in turn have led to the generation of further data, enabling further scientific research. Welter et al. [58] discuss the benefits of adopting SW technology to aid data structuring and increase automation in the periodic update of the growing catalogue for the Human Genome project [see 51]. They demonstrate additional benefits such as a reduction in the reliance on manual generation of representative visual overviews of the data, and, further, the ability to create additional, interactive visualisations by appropriating other technology, using the data encoded to simplify and widen (re)use. Welter et al. [58] illustrate extensibility of the approach to capture relevant information from other related, open biological data stores and publications, as well as seamless linking to and more expressive querying of data both within and outwith the field. For example, by extending their ontologies, domain experts are able to retrieve key information not captured in experimental data and study results, such as the ethnicity of participants, by linking to geographical and ethnology databases.

Kazemzadeh et al. [26] also discuss the challenges in obtaining overviews of biological data generated in different ways and for different purposes. Although structured using standard ontologies, these data seldom define explicit links between related attributes. The authors therefore make the case for unifying frameworks that support synthesis of and analysis across these datasets, to enable more comprehensive analysis that considers relevant information captured by different experiments and in disconnected datasets. They use the construction of a custom domain ontology describing links across data of interest and LD principles to demonstrate how this may be achieved, to improve the discovery of biological interaction at the molecular level, a key component of research into human diseases. To support domain experts who may

have difficulty formulating formal (SPARQL) queries, Kazemzadeh et al. [26] provide an interactive visualisation dashboard to enable intuitive search and exploration, employing pre-specified query templates that return results using a colour-coded, force-based network layout that depicts component type, interaction type, and strength of the relationship between components.

Another area outside the SW community where LD plays a key role is in *data journalism*, due to the clearly recognised benefits in following (structured) links between related-but-independent sources of data and knowledge. LD can contribute to improving exploration and enriching information retrieval, complementing new approaches to curating and delivering news stories that threaten to disrupt traditional journalistic practice. Dowd [13] highlights the need to (re)train and upskill a new generation of investigative journalists to enable them to transform *big* and *open* data into basic to advanced graphics and interactive visualisation, to meet expectations for redefining and augmenting traditional news reporting and storytelling. The study also recognises how LOD, along with the use of data collected via social media platforms, enables the ordinary "interested" public to aggregate existing and/or create and report their own news stories, with both professional and such "lay journalists" typically appropriating the medium to provide additional context for stories using statistical charts and geo-visualisation. Professionals with access to high-end display hardware such as smart screens are further able to improve their art by overlaying this related data and corresponding analytics on the visual story.

Data journalism, whether by professionals, lay journalists, as well as other content creators and aggregators such as bloggers, employs tools for data exploration, retrieval, recommendation and annotation. *Refer* [53] is an example of a text-based, visual exploration prototype that uses Named Entity Linking to retrieve and recommend entities for annotating text using data retrieved from DBpedia [12,29]. By exploiting the links present in LD, *refer* provides end-users with support for navigating from their start point to explore other related entities, by traversing data across four categories, colour-coded to aid recognition as `Person`, `Place`, `Event` and the general, abstract concept of `Thing`.

Javed et al. [25], similarly, illustrate the importance of intuitive navigation of LD in *VIZ-VIVO*, which uses radial network maps and sunburst diagrams to gen-

erate information overviews describing research communities, by building on the structure of a custom ontology. The aim is to support exploratory navigation through the data graph, and, as a result, richer, context-driven information retrieval. Javed et al. [25] demonstrate how this approach enables users to answer questions about interaction within research communities and measure the impact of research at different levels.

Rakhmawati et al. [45] illustrate the application of LD principles to financial data reporting and the collection of information on crowdsourcing, to promote more equal distribution of funds to mosques in Indonesia, where such institutions also serve a social function. The authors use a web-based portal and a Twitter portal to collect data about individual and joint contributions to mosques, and merge the financial data into a (centralised) LD store. This approach enables further data enrichment by linking to geolocation and other relevant information, allowing the aggregated financial information to be displayed on a map of mosques coupled with statistical charts displaying more detail, highlighting where non-uniform distribution of funding occurs. Rakhmawati et al. [45] indicate the potential to carry out further analysis as their approach simplifies the porting of the resulting LD to other applications.

3. Evaluating usability & utility

Recognition of the need to support human end-users (in addition to the focus on machines) is receiving greater traction within the SW community, especially for applications targeted at lay users and even domain experts who are not necessarily technological experts. This is critical to extending reach of the technology beyond the SW community, and especially important as expectations of end users grow with advances in technology. Further, as data size and complexity increase exponentially with ubiquitous use of said technology [39,58, among others], the need for intuitive, user- and task-focused tools for navigating through the noise to identify and make effective use of relevant knowledge, within the user's context of use and field of reference, increases. This is true for domain and technical experts and lay users, the latter of whom range from technophobic or low technology awareness, to those who routinely use a range of advanced technological devices and software in their personal, social and working lives.

One benefit in the pipeline or integrated approach to tackling requirements for LD visualisation is that the process itself serves as a knowledge retrieval task and a form of evaluation of both the approach and the tools and techniques that feed into it [see, e.g., 21,37,47]. The development team can only progress to the next stage in what is often an iterative pipeline or workflow if they have been successful in: (a) correctly identifying relevant resources; (b) extracting data and metadata that meet the user's context and the requirements of their task or end goal; (c) preprocessing this data, where necessary, to feed into the visualisation tools available and the analysis required; (d) carrying out initial exploratory analysis, using analytical and/or mining models, statistical and/or visualisation-driven methods. These initial steps will confirm whether the input data and preprocessing will feed into solving the problem at hand, and therefore guide the identification of suitable approaches for more detailed analysis, or may need to be reviewed to improve the process or correct issues found.

We therefore now see an increase in usability evaluation of tools built on SW technology. Further, the open data initiative has spurred the sharing of what is typically considered personal or in-house research data (such as interim analysis and evaluation data) along with final analysis results, to encourage verification of research by third parties and contribute to benchmark datasets. Tietz et al. [53], for instance, make available online anonymised evaluation data for their tool *refer*. They evaluate usability of the text visualisation tool with a range of regular Web users, a quarter of whom described themselves as LD experts, while the others had passing to no knowledge of Linked Data. The study found that while the UI was, overall, seen as useful, and navigation sufficiently intuitive to allow both user types to discover the information sought, some UI features for revealing additional information were not recognised especially by the lay users. An interesting finding: criteria for categorising the data and proposing recommendations, a key component in obtaining more complete understanding of relationships revealed as users navigate through data, was assumed by some users to be the whole document, rather than the entity with the focus. This highlights a significant knowledge gap between SW experts and lay users, with the push in the SW from a Web of *documents* to a Web of *data*, and the finer granularity available with the latter.

Benedetti et al. [2] in what they describe as qualitative evaluation, involving both lay and technical users, report high satisfaction and accuracy across both

user types for tasks requiring browsing of their visual schema summaries (91% with up to 5 min to complete), and successful task completion based on visual query generation (90% with up to 15 min to complete). It should be noted that the authors restrict the tasks to those that can rely on browsing the schema summaries, and do not compare their results with benchmarks, nor mention whether or not relevant benchmarks exist.

Valsecchi et al. [54] target technological experts and evaluate their tool only with users who have a Computer Science background. Participants worked with two maps generated from highly inter-connected DBpedia [12] data and very disconnected Linked-MDB [19,32] data, respectively. The evaluation solicited qualitative and quantitative information on intuitiveness and learnability of the UI, and how well it supported users in obtaining an understanding of data structure, including connectedness and content. Results were mixed: success and correctness in completing tasks ranged from 65% to 100%, with higher scores overall for the DBpedia layout. However, participants were able to discover only some of the key functions, including a subset of those required for navigation. A key finding was the importance of providing additional textual detail to complement and confirm the conclusions drawn from the visual overviews. This is not an unusual finding: Ware [56], for instance, illustrates how text may be used to complement visual thinking and understanding.

We conclude this section with a summary in Table 1 of prototypes built and new approaches described in this section and in Section 2, by comparing these according to a set of criteria for usable visualisation, derived from guidelines for interactive visualisation and visual information seeking found in [49], [see also 9]. It should be noted that the assessment is made using authors' descriptions where an implemented prototype is not available publicly; such cases are highlighted with an asterisk after the tool name.

4. The special issue

The call for papers for a special issue of the Semantic Web Journal on "Visual Exploration and Analysis of Linked Data" was posted in June 2014. We received 13 formal expressions of interest and 11 complete submissions. Of these, 6 were accepted for publication, addressing between them the three key components identified in the call, that feed into effective Linked Data visualisation (see also Section 2).

Linked Data, open or closed, evokes discussions about big data, within the SW community and further afield, in government and in industry, due to its size, complexity, heterogeneity, dynamicity, distributed nature and differences in granularity, quality and veracity. But LD goes beyond simply being big, serving as an *enabler* with potential to address some of the key issues in managing vast amounts of data. While not required to follow a schema, a situation that contributes to challenges in its use, the inherent structure of LD supports identification of the links within a dataset and across to other independently generated data, enabling more seamless exploration and richer information discovery. The authors in this special issue highlight the importance of following guidelines and best practice for curating Linked Data as well as those for visual information presentation and analysis, to work toward realising the full potential of LD.

Fu et al. discuss the need to evaluate visualisation-based techniques used in ontology understanding, a task that contributes to the process of ontology mapping, which is in turn a contributor to linking independent datasets. The authors argue that such evaluation is key to the design of usable, intuitive, user-facing tools.

Nuzzolese et al. mine the link structure in LD to rank importance of data associated with a knowledge entity of interest, based on popularity with respect to the contributions of the crowd to encyclopaedic knowledge. The results are used to generate visual summaries about the topic of interest and support step-wise navigation to other related data. Bikakis et al. aggregate LD hierarchically based on its backing ontological structure, to improve navigation through the data using a variety of visualisation techniques built on top of the resulting aggregates. Scheider et al. harness the familiar metaphor provided by maps to improve navigation through LD, and hence, exploration and querying of LD with spatio-temporal attributes.

A number of one-off visualisation solutions exist. However, low to no reusability limits their usefulness. An important challenge being tackled is, therefore, the design of workflows, pipelines and integrated, multi-perspective solutions that tackle different stages in the complete Linked Data management cycle, typically focusing on data selection and pre-processing, initial exploration and a selection of tools for more detailed analysis of ROIs and for presenting the results of analyses. Brasoveanu et al. describe the task-centred process they followed to build a visualisation dashboard for decision support, comprising a set of coupled tools for aggregation and analysis of linked open statistical

Table 1

Features addressed for the LD visualisation approaches and tools reviewed. "x*": partial implementation or limited scope in tool/approach; " ": feature is not provided but could be implemented for the tool/approach; "–": feature not relevant to or within scope for tool/approach

Usability Criterion	LD-VOWL	Sparql-FilterFlow	LODeX	NHGRI GWAS Diagram	LinkedPPI	EarthCube GeoLink	OpenData-ZIS*	Linked Data Maps	VIZ-VIVO*	SWJ Sciento-metrics	refer
Visual overview	x			x*		x	x	x	x		
Multiple perspectives			x			x	x	x		x	x
Coordinated views			x			x	x				
Data overview										x*	
Graph visualisation	x		x			x			x*		
Ontology/RDF graph view	x	x	x		x	x					
Detail on demand	x	x*	x	x	x	x	x	x	x*	x	x
RDF URIs & labels	x*	x	x			x					
Highlight links in data	x	x	x	x	x	x		x	x	x	x
Support for scalability	x*					x*	x*	x	x*	x	x*
Query (formal syntax)			x								
Query (forms/keyword)		x*	x		x*	x	x*	x			x
Visual querying	x	x	x					x*			x*
Filtering	x	x	x	x		x	x	x	x	x*	x
History									x*		x*
Presentation templates	x	x			x			x			x
Keyword/DMI entry point		x	x		x	x		x			x
Non-domain specific	x	x	x			x*	x*	x	x	x*	x
Reusable output				x*			x			x	
Target – Lay-users		x*					x	x*	x*	x	x*
Target – Domain experts		x		x	x	x			x		x
Target – Tech-users	x	x	x			x*		x		x	x*

data. Del Rio et al. extend the five-star methodology for generating high quality LD. They define a seamlessness metric for visual analytics, with the goal of improving reuse of the interim and final results obtained during visual analysis, by embedding semantic data into the process and the analyses' output.

Cartography is an oft-used metaphor when generating visualisations, as it takes advantage of users' familiarity with and consequent ability to read maps to aid exploratory discovery. For example, a user may wish to explore Napoleon's march to Russia in the War of 1812 [38]. The classic map requires mainly human perception to understand its content; expert knowledge in history and cartography may provide some advantages but are not strong requirements. However, making the visualisation and its content searchable in a digital data store requires additional annotations, both textual and using visual overlays, to be stored with the map. Scheider, Degbelo, Lemmens, van Elzakker, Zimmerhof, Kostic, Jones & Banhatti in *Exploratory Querying of SPARQL Endpoints in Space and Time* examine challenges of this type, faced by users in for-
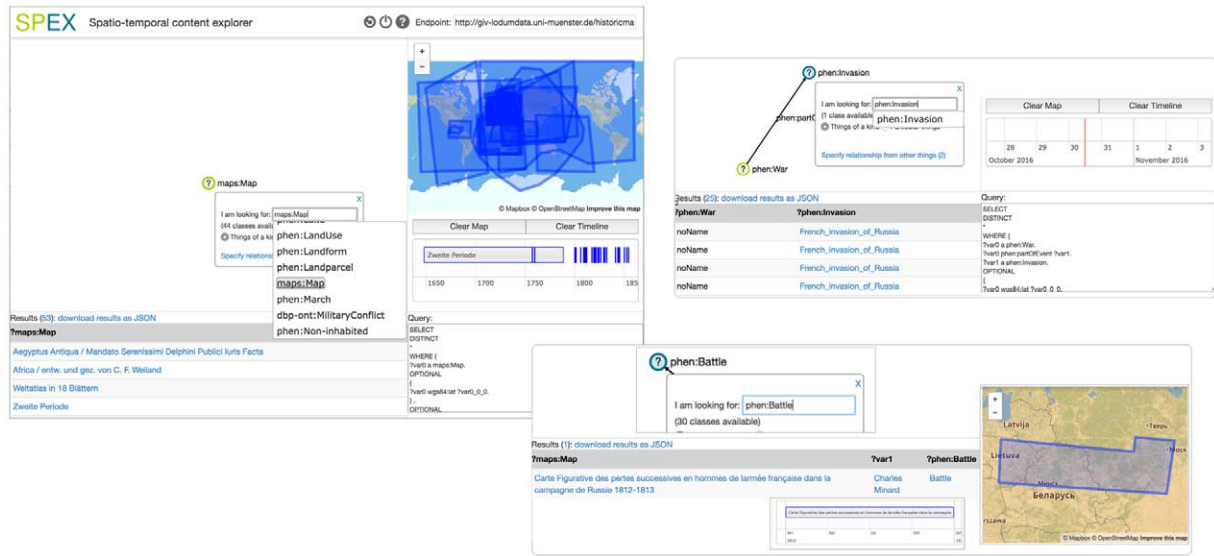
Fig. 1. Building queries using the click-and-select UI in SPEX, the visual query editor built to support non-technical end users exploring Linked Data.

mal querying of data stores as part of the exploration and question-answering process. Even assuming such annotations exist, end users, whether technological experts or lay persons, require additional expertise to successfully retrieve information required to answer their questions; namely, the need for expertise in formal query languages such as SPARQL, at least basic understanding of data structure and content, and depending on the data types involved, understanding of complex data constructs.

The authors therefore look at two data types – *spatial* (physical or geographical **space**) and *temporal* (**time**) – that occur frequently across many domains. Scheider et al. employ *space* and *time* as features that may be used to explore individual data sets and as a bridge between data silos, by harnessing the everyday metaphor of maps coupled with a timeline. They also follow Shneiderman's mantra [49] to define a set of design requirements to support intuitive, visual exploration. They aim to enable a broad range of users, that do not necessarily have much domain or technological expertise, to progressively build and edit queries as they explore linked datasets along attributes defining "space and time".

Scheider et al. describe a qualitative usability evaluation carried out with six participants: two librarians, two PhD students in Geography, and two ITC lecturers. The goal of this study was to assess the extent to which their prototype *SPEX*, the **SP**atio-temporal **C**ontent **E**xplorer (shown in Fig. 1), satisfies the design guidelines they specify. The authors report between 45 and 90 minutes for familiarisation with the tool using a printout of its help file, followed by 30 to 45 minutes to complete a set of exploratory tasks. Participants made good use of all query features, including functionality for reuse of previous results. However, none were successful at completing all tasks, with an overall success rate of 52%. The authors note, though, that even where participants were not successful in completing the tasks, in some cases they tried alternative strategies to answer the questions, with some of these alternatives leading to partial success or a correct answer(s).

Key challenges faced by the participants were related to the unfamiliar use of terminology and labeling of data – which, derived directly from the RDF, would be relatively easily interpreted by a SW expert, but did not translate easily otherwise, especially for the more ambiguous terms. A simple example is distinguishing between a map *entity* and the (visual) *geographical* map. Other challenges were due to user interface design issues such as what were found to be redundant steps to trigger events in the dialogs, and how to generate sub-queries using the timeline.

The authors conclude the paper with future directions for research to more effectively translate their design guidelines into more intuitive support for a non-technical, lay user audience.

Ontologies typically serve as a backbone for structuring Linked Data, and play a key role in defining

links within a dataset and across datasets to other related data. Tools for constructing and exploring ontologies and especially for mapping and alignment of concepts and properties therefore contribute to effective management and (re)use of LD. Fu, Noy & Storey in *Eye Tracking the User Experience – An Evaluation of Ontology Visualization Techniques* recognise a gap between work to develop ontology visualisation tools and the usability evaluation required to ensure best practices are followed in the design and development of such tools. They highlight the resulting negative impact on the appropriation of useful and usable techniques that map to user expertise and knowledge requirements in ontology engineering and related knowledge management tasks, Linked Data being one obvious example.

Fu et al. use eye tracking to collect empirical evidence about usability in a comparative study of two of the most widely used visualisation techniques in ontology understanding: indented lists and trees. The authors argue that eye tracking, being able to directly capture user actions, provides more in-depth measures, at a micro level, to help determine *why* an approach or tool may be effective, as opposed to reliance only on traditional measures of performance during usability evaluation such as time to completion and success rate, that provide indirect means for capturing usability.

The study examined four hypotheses: (i) search, (ii) information processing, (iii) ability to reduce cognitive load, (iv) why a particular technique supports more efficient and/or effective use by particular types of users.

The study involved 36 participants, graduates and undergraduates from a wide range of disciplines within the Sciences. They were instructed to perform mapping tasks between two sets of ontologies; each session lasted approximately 2 hours. Participants were required to verify whether existing mappings were correct and complete (identifying missing mappings otherwise). Visualisations of each ontology were provided to aid task completion, which relied on exploration of each pair of ontologies, both created from standard benchmark datasets. One pair was simple and described conferences. The second pair from the biomedical domain, describing organisms, was more complex. The authors note deliberate effort to reduce participant bias due to familiarisation with the data domains and the visualisation techniques, in order to simulate as closely as possible challenges faced by novices (with no knowledge of SW technologies and especially ontologies). This restriction allowed measurement of the

visualisation support for completing the task while reducing confounding factors, on the basis that domain- or SW experts would require less intuitive support to complete the tasks successfully. Further, the visualisations were generated to be as representative of current applications as possible, by implementing the most typical cues for navigation and browsing, focusing on ROIs, filtering and hiding data. Also, each task started with only the root node visible, so that participants had full control over their navigation paths.

The results, based on 31 participants (five were discarded due to incomplete eye tracking data), found the indented lists to be more effective for searching, while the graphs better supported information processing. Participants completed tasks more quickly using the lists, at least in part due to the more compact visualisation requiring less time to scan. However, they did not improve on accuracy over the use of the graphs. Evidence on which of the two resulted in higher cognitive load was inconclusive, with no statistical difference between the two visualisation techniques. Overall, participants spent more time on information search than in processing; Fu et al. suggest that more research be conducted about building support for search in ontology visualisation.

The paper concludes with pointers to a more detailed investigation of affordances for information search and processing in ontology visualisation tools, including more efficient use of screen space and investigation of the specific requirements of different user types across a wider set of visualisation techniques.

In *Aemoo: Linked Data Exploration based on Knowledge Patterns* Nuzzolese, Presutti, Gangemi, Peroni & Ciancarini extend previous work on the use of knowledge patterns to counter challenges faced by humans and machines attempting to explore the knowledge content of Linked Data. Nuzzolese et al. argue that encyclopaedic knowledge patterns (EKPs), being derived from a collective intelligence store spanning multiple domains, provide a "cognitively sound" base from which to carry out exploratory discovery of complex, often dynamic, heterogeneous data such as LD.

EKPs are essentially mini-ontologies that define a class, and by mining the links to instances in Wikipedia, retrieve relationships to other classes ranked based on popularity or frequency of use. EKPs map to how humans [contributors to Wikipedia] view (knowledge about) entities and inter-relationships within this knowledge. By retrieving the most important/highly ranked relationships (above a variable threshold),
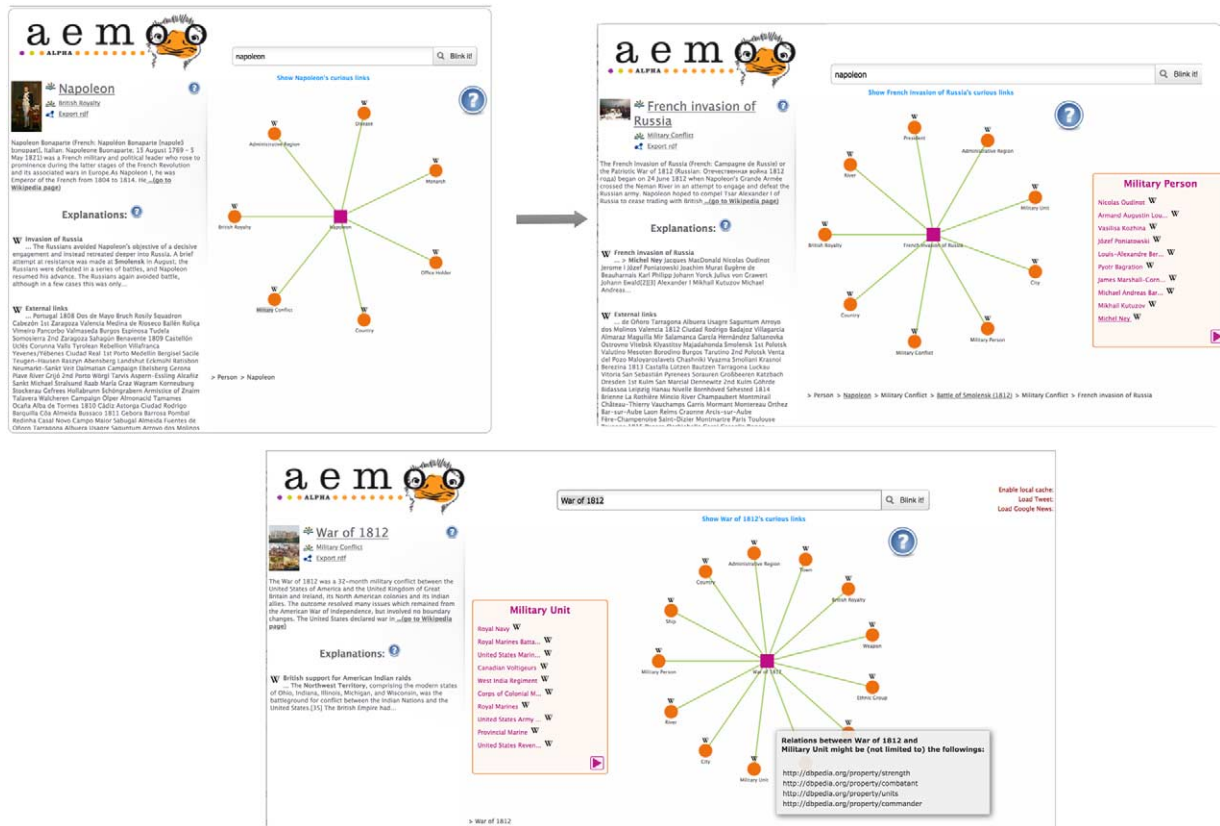
Fig. 2. Illustrating from left to right, navigation from the EKP for the entity retrieved for the search term "Napoleon" to the EKP corresponding to the "French Invasion of Russia" in "The War of 1812" (see lower snapshot), through an entity of type *Military Conflict*.

EKPs can be used to summarise knowledge about an entity.

Nuzzolese et al. describe *Aemoo* (see Fig. 2), a tool targeted at users with neither particular knowledge of the structure or content of a data store, nor expertise in SW technologies. Aemoo demonstrates how EKPs may be used to explore knowledge bases. The aim is to help users focus on the most "important" information about an entity and then navigate to other related information. The authors employ a concept map coupled with a text-based detail window, arguing that mapping the visualisation to the data structure of EKPs aids end-users in forming mental models that support exploration. However, acknowledging potential challenges in the use of the large, complex, unwieldy graphs entailed by the inter-linking of large-scale, multi-dimensional datasets, they restrict their concept maps to a depth one level below the root. Their approach filters out less relevant or important information, thereby also reducing noise. Because Aemoo queries Wikipedia on-the-fly, the information retrieved

reflects all updates to the knowledge base. This is in addition to dynamic data retrieved from Twitter and Google news that correspond to the entity in question. Aemoo includes a switch that allows users to retrieve what is described as *peculiar knowledge* or *curiosities* – entities linked to the focus that fall below the popularity threshold. This enables users to also visualise information about an entity that would usually not be considered important or relevant.

Nuzzolese et al. present the results of a comparative evaluation, taking Google as a baseline for exploratory search, and RelFinder [35] to test Aemoo's automatic summarisation against manual filtering and existing approaches to visual querying and exploration. Thirty-two non-technical undergraduates from two universities completed three tasks using Aemoo and, alternately, one of the other two tools. The authors report that each of the three tools performed better than the other two for one task – RelFinder at summarisation, Aemoo at discovering related entities, and Google at discovering relations between entities. Overall, how-
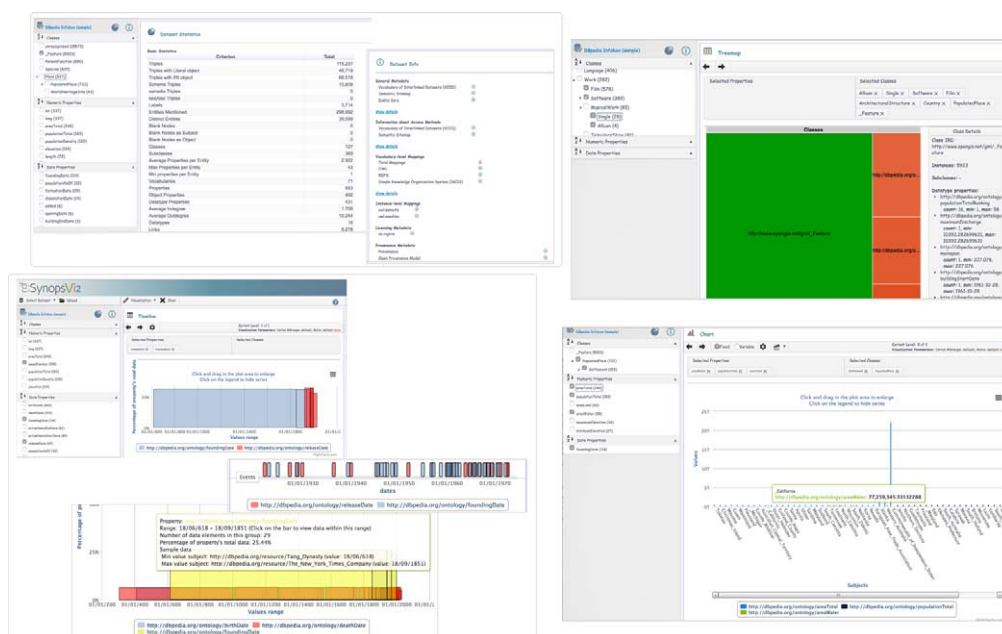
Fig. 3. Visualising the results of hierarchical aggregation of Linked Data in the *SynopsViz* prototype. Clockwise from top, left: dataset statistics; selected classes in a treemap; selected spatial attributes. Bottom, left: middle snapshot zooms in to show detail, before zooming back out and including another temporal attribute.

ever, Aemoo outperformed both RelFinder and Google search, with RelFinder performing worst. While Aemoo also scored higher for learnability and usability, based on the SUS questionnaire [see 6], there was no significant difference with the other tools.

Based on these evaluation results, Nuzzolese et al. are considering, among others, reusing functionality in RelFinder to present relationships in Aemoo. They conclude with pointers to additional work to improve presentation of, and navigation in, knowledge bases.

Bikakis, Papastefanatos, Skourla & Sellis in *A Hierarchical Aggregation Framework for Efficient Multilevel Visual Exploration and Analysis* tackle the challenges faced in exploratory analysis of today's big, heterogeneous, dynamic data, to support immediate and on-going analysis online. Bikakis et al. posit that a hierarchical approach to data processing and aggregation enables customisable construction of data overviews on the fly, from both static and especially dynamic, numerical and temporal data. Followed by support for identifying and drilling into the details of ROIs, the authors aim to provide support for end users in navigating through and making sense of *big data*.

Their generic model takes as input data regardless of structure and generates output structured to allow the use of varied visualisation techniques for presentation and analysis, hierarchical or not, complemented by statistical summaries of the input data. The authors note also that while they focus on data encoded as RDF, the approach is reusable for data in other forms, provided it is numerical and/or temporal.

Figure 3 shows *SynopsViz*, a prototype built to demonstrate the results of applying this hierarchical aggregation model to Linked Data. Three visualisation options, a timeline, a treemap and chart views, are provided for visualising the output.

Bikakis et al. assess the effectiveness and efficiency of their approach by comparing it with a baseline described as *FLAT* – displaying all detail for a dataset. The test measured (initial) response time for data preprocessing and subsequent response time for displaying data of interest, using the DBpedia 2014 dataset [10]. Construction and response time for the two hierarchical approaches presented were roughly equivalent, with minor improvement for the range-based over the content-based tree. For data up to 10,000 triples, FLAT performance approaches that of hierarchical aggregation. Significant cost in construction of the hierarchical approaches was found to be due to communication, notable for small datasets, but decreasing proportionally for larger datasets due to additional time required to render the latter. To improve overall re-

sponse time, the authors also enable incremental construction for the hierarchical approach. With respect to data properties, the test results show that FLAT was restricted to a maximum of 305,000 triples, while the hierarchical approach successfully processed all data subsets. Further, the latter was found to outperform FLAT for all tests, with increasing difference in response time with dataset size.

Usability evaluation of the SynopsViz tool was also carried out with ten participants with a background in Computer Science. The evaluation involved three sets of tasks across all three approaches to measure users' ability to (i) discover specific information, (ii) discover data within defined ranges and (iii) compare and rank data subsets. The comparative evaluation was carried out with two datasets, one containing 970 triples and the second almost 38,000. For all three approaches, the authors found significant differences in task completion time for the larger dataset size, due to the much longer time spent navigating through data structure and content. The hierarchical approaches significantly outperformed FLAT for the larger datasets, because data clustering reduced the number of navigation steps required to find matching responses. For the smaller dataset, however, time to completion for FLAT approached that of the hierarchical approach, due mainly to the need for a larger number of navigation steps through the hierarchy. The range task was also particularly difficult to perform using the FLAT approach – cognitive load may have contributed to this as participants had to browse a significant portion of the data (displayed in detail) to complete the task. The range-based and content-based approaches differed slightly in performance time depending on which approach better matched data clustering to task type. The authors note also a key concern of participants that hampered navigation: getting lost during exploration.

The authors conclude with pointers to future work, including a hybrid version that would employ a weighted combination of each of the content-based and range-based approaches to aggregation.

Del Rio, Lebo, Fisher & Salisbury in *A Five-Star Rating Scheme to Assess Application Seamlessness* assess the cost of coupling multiple visualisation tools together on different platforms, with the aim to promote seamless reuse of visualisation solutions and therefore more effective analysis.

The authors use the *munging* task in Visual Analytics (VA) – processing of input data to fit a specific tool's requirements – to illustrate variation in cost due to data structure and semantics, from high-cost

1–3 star *mundane* data to low-cost, inter-connected, semantically-enriched, 4–5 star data. They extend the PROV ontology [28] to capture data and the analytical process, and to illustrate how inter-connected, semantic, Linked Data may be used to build a bridge between independent datasets and the results of analyses, thus lowering the cost of performing those analyses.

Del Rio et al. describe an application scenario where a student with technological expertise analyses data from multiple stores to obtain information about decommissioned satellites and other equipment abandoned in space. The scenario illustrates the need to switch constantly between semantically-rich data describing the student's focus, and more mundane representations of this data in visual form, that are easier to interpret but whose content cannot be reused for further analysis. The analysis is therefore constrained, with significant cost incurred by moving between mental models of the data formed as the analysis progresses and alternative representations that break these models. Some of these require further data munging, resulting in attention loss with respect to the data and their semantics. The student then passes on her initial results to a fellow student, who reviews and extends the analysis to confirm her results and answer open questions. The scenario illustrates further difficulty reusing the initial results, due to limited provenance information about the data and the lack of traces about the analytical process followed. Cost again increases when the first student attempts to reuse the analysis results of her colleague; while the new visualisations provide further information, they are not directly connected to the underlying linked and semantically-enriched data.

Del Rio et al. utilise the outcomes of this scenario to derive an *application seamlessness* metric that takes into account the initial cost of analysis and the cost of reusing the results of prior analyses. The metric is defined using ontology restrictions that assign costs ranging from one to five stars. The costs reflect the impact of employing a set of applications or following specific approaches in the data analysis process. The proposal harnesses the strengths of both the VA and SW communities for managing data to avoid *pain points* and reduce cost in data analysis, instead increasing effectiveness. A key component of the approach is maintaining the links between the analyst, the analytical tools in use, and input data (including previous analysis results).

The authors conclude with proposals for future work to take into account also analysts' expertise and background in measuring the cost of using selected tools or

approaches. They also discuss refining their guidelines to capture more effectively the cost difference between individual star ratings, as opposed to ranges, in their seamlessness metric.

Today's inter-connected, data-driven economy requires decision-makers to consider the impact of multiple external factors using data from a variety of sources and increasingly outside their domain, in order to feed wider contextual data into informed decision-making. However, the cost of access to third party data, especially where proprietary, and support for effective analysis of the resulting heterogeneous datasets, are often beyond the means of smaller organisations and research institutions. Brasoveanu, Sabou, Scharl, Hubmann-Haidvogel & Fischl in *Visualizing Statistical Linked Knowledge Sources for Decision Support* look at the potential of the growing collection of statistical Linked Open Datasets to fill this need. These datasets, including the Eurostat database [11] and the World Bank Catalog [41] are built on the RDF Data Cube Vocabulary [8].

Brasoveanu et al. identify a set of challenges inherent to the exploration and analysis of large-scale, heterogeneous, inter-connected data, made more complex by the unique challenges found in statistical LD analysis – including variation in data structure and content, inconsistent and poor use of data schemas, and frequent failures of dynamic data portals such as SPARQL endpoints. They therefore highlight the need for integrated visual, multi-perspective analysis solutions that enable complex question-answering over such data, that scale, and that are reusable beyond specific use cases.

The authors describe the iterative, scenario-driven process they followed through requirements collection, design and prototyping, following best practices and guidelines for visual analysis, with a focus on the specific requirements of Linked Data and statistical data. This process resulted in a dashboard containing multiple, coordinated components for data preprocessing and integration, visualisation-driven analysis, and for effective presentation of the often also multi-dimensional analysis results. Based on their findings – challenges faced and successful application of research – they propose a set of guidelines and workflows for tackling the challenges in building decision support systems over Linked Statistical Data.

Brasoveanu et al. illustrate the advantages in providing multiple perspectives on the data (slices) and the results of analyses using a set of scenarios about decision support in the tourism and telecommunica-

tions industries. They then summarise the results of exploratory evaluation of their *ETIHQ* dashboard, to measure usability against traditional approaches to cross-domain analysis for decision support, and identify additional functionality required by target users. Sixteen participants took part in the evaluation, all in their usual work environment: ten tourism experts based in different countries and six researchers. They completed two activities with the dashboard, comprising preset and user-defined tasks, and a third activity using their usual working tools. The *ETIHQ* dashboard was observed to perform significantly better than manual (traditional) approaches to question-answering, with participants providing more detailed and precise responses, accompanied by a gain in task completion time of almost 30%. Recommendations for improvements to the dashboard included, among others, the use of a simpler workflow, the implementation of search features, and the use of domain terminology (as opposed to terminology reflecting the use of SW technology for encoding data). Participants also requested the inclusion of news and social media data.

The authors conclude with a description of updates to the dashboard based on feedback from the usability evaluation.

We summarise in Table 2 the contributions in this special issue by looking at the set of visual presentation and analysis features that was also used in Table 1 to compare other related work. For completeness, two additional features relevant to LD visualisation and analysis are included: *faceted search/browse* and support for *editing underlying data*. Neither of these features was implemented in any of the tools in Table 1, and we only see limited implementation here.

4.1. Reviewers

This special issue would not have been possible without the support of our reviewers, who provided detailed and informed feedback to authors throughout the multiple iterations on each paper. We thank all reviewers for their contribution to the special issue, and list them below in alphabetical order.

James Burton University of Brighton, England, UK
Mark Gahegan University of Auckland, New Zealand
Roberto García González Universitat de Lleida, Spain
Luc Girardin Macrofocus GmbH, Switzerland
Florian Haag University of Stuttgart, Germany
John Howse University of Brighton, England, UK
Valentina Ivanova Linköping University, Sweden

Table 2

Design features for visual presentation and analysis of Linked Data addressed in contributions to the special issue. "x*": partial implementation/limited scope; " ": could be implemented; "–": not relevant/out of scope

Usability Criterion	Scheider et al.	Fu et al. *	Nuzzolese et al.	Bikakis et al.	Del Rio et al. *	Brasoveanu et al. *
Visual overview	x*	x*	x*	x	x	x
Multiple perspectives	x	x		x	x	x
Coordinated views	x*				x	x
Data overview		–			x	x*
Graph visualisation	x	x	x			
Ontology/RDF graph view	x	x	x*			
Detail on demand	x	x	x	x	x	x
RDF URIs & labels	x		x	x		
Highlight links in data		x	x		x	x
Support for scalability				x	x*	x*
Query (formal syntax)	x	–				
Query (forms/keyword)	x		x			x
Visual querying	x		x			x
Faceted Search/Browse				x*		x
Filtering	x*	x*		x	x	x
History			x	x	x	
Presentation templates						x
Keyword/DMI entry point	x		x	x	x*	x
Non-domain specific	x	x	x	x	x	x
Edit underlying data					x*	
Reusable output					x*	x*
Target – Lay-users	x*		x			x*
Target – Domain experts		x*	x		x	x
Target – Tech-users	x	x*	x*	x	x	

Tomi Kauppinen Aalto University School of Science, Finland
Steffen Lohmann Fraunhofer IAIS, Germany
Heiko Paulheim University of Mannheim, Germany
Jan Polowinski TU Dresden, Germany
Mariano Rico Universidad Politecnica de Madrid, Spain
Bernhard Schandl mySugr GmbH, Austria
Thomas Wischgoll Wright State University, USA

5. Future challenges for Linked Data visualisation

In their introduction to the Semantic Web Journal's first set of LD description papers in 2013, Hitzler and Janowicz [21] identify Linked Data as *"enablers in research"*, stating its importance. We reiterate this statement confidently: even if many challenges remain, LD has significant value as a rich, structured source of knowledge for both research and practical applications.

The challenges discussed in this special issue revolve mainly around making LD accessible to end-users, primarily through interactive, visual representations of this novel form of data whose characteristics are quite unique, as discussed earlier. While these characteristics are what make LD such *enablers*, they are also the reason why it is proving so difficult to design effective user interfaces for LD and for the Web of Data at large.

As demonstrated in the papers in this special issue, we continue to see advances in research on the use of visualisation for making LD more accessible. Efforts have been moving from producing simple, basic demonstrators of LD browsing to more elaborate tools that integrate data capture, exploration, analysis and/or presentation in a visualisation-driven front-end. However, there is still significant room for improvement.

Hendler [20] in 2008 raised the "chicken-and-egg problem" regarding the adoption of Semantic Web technology outside the community. The technology was not sufficiently proven at the time, even though RDF and SPARQL had been introduced several years before, and examples of success using ontologies to describe data already existed. Government, state-run

institutions and industry were reluctant to dedicate resources to encoding and sharing their varied data to grow this new "Web of Data", but data was required to feed into the development of sophisticated tools for managing and interacting with this new Web. Seven years later, Neish [40] describes several examples of successful use of Linked Data in different domains outside the Semantic Web community, in government, education and industry. He stresses, however, that links across independently created datasets rarely span across domains. Further, technological barriers continue to hamper the adoption of LD principles in data generation and sharing, with a corresponding increase in economic and other risks following these principles for data management.

User interface design for LD is still in its infancy. This is reflected in publications describing design and prototypes developed within the research community for LD visualisation, which are targeted predominantly at workshops rather than journals or research tracks in conferences. However, workshops dedicated to discussion of research on interactive visualisation tools for LD are one avenue that will help foster research as the generation and consumption of LD increases, and as LD use spreads over more and more application areas. Interestingly, Neish [40] concludes that incremental adoption of LD may help avoid the acknowledged "chicken-and-egg problem", by ensuring that the technology is well developed and proven before the Linked Data initiative starts to play a key role in helping to resolve newer challenges such as those posed by today's complex, distributed, heterogeneous, big data [see also 39,58].

5.1. Breaking away from graph visualisation

The review of the state of the art (Section 2) shows a predominance of LD representations based on graph visualisation techniques. This is probably due to the following key factors that encourage the use of node-link diagrams to depict tree and graph structures: (i) ontologies are often hierarchically structured, rooted at `Thing` or another general, abstract topic or domain concept; (ii) RDF's data model is a directed labeled graph, *"ipso facto we use graphs to represent it"* [48]; (iii) network analysis is one of the more common visualisation-driven tasks carried out within the field, to explore, e.g., collaborations and other interrelationships between researchers and within research data, and social networks at large. It is thus not particularly surprising that many attempts have been made over the years to visualise these tree and graph structures. Such approaches, however, have had limited success; even those that have explored alternatives to the typical representation based on node-link diagrams, such as *OntoTrix*, which investigated the potential of a hybrid representation involving adjacency matrices [1]. While node-link diagrams may be effective representations when considering very small graphs, they fail to provide meaningful visualisations beyond a few dozen nodes and edges, even when styled to better convey the resources' and properties' semantics [43]. Benedetti et al. [2], for instance, report that evaluation participants wondered how usable their graphical schema-based summaries would be for large datasets. In this special issue, Nuzzolese et al. acknowledge this challenge by restricting the concept maps they use to visualise encyclopaedic knowledge to only a single level below the root (and focus), in addition to setting a threshold below which data directly linked to the focus is hidden.

Indeed, node-link diagrams quickly become very cluttered and illegible, and generate a lot of *viscosity* [15] both in terms of interactive exploration and content manipulation. Furthermore, and perhaps more importantly, exposing the data's graph structure might not always be relevant. Doing so may therefore have the unintended consequence of over-emphasising the data model to the detriment of the content. In many cases, the model is of little importance to users, and it is interesting to observe that early attempts at creating general-purpose LD browsers, no matter the targeted level of audience expertise, have tended to rely on presentation solutions such as Fresnel [44], that do not expose the graph structure to users, but rather provide *lenses* on this structure that employ, for example, metaphors that relate more closely to end users' tasks and context.

More recent work, including some of the articles in this special issue, are continuing to investigate this avenue, moving away from node-link diagrams and representations of the low-level graph structure in general. Instead, LD front-ends now tend to provide richer, more versatile sets of user interface components, each designed for different types of data attributes: interactive map components for resources featuring geolocation information, timelines for resources featuring temporal information, etc. Network visualisations can be part of such suites of UI components, but they should no longer serve as the main representation of the data. They should rather provide the means to reveal some particular aspects of the data, and be used

only when appropriate, i.e., when the way the data is structured at a low level is of importance to the user, or when higher-level hierarchical or network structures can be derived from it. We see this in, e.g., Halpin and Bria [17], who in their exploration of DSI overlay links between institutions on a map of the EU to reveal the collaboration network resulting from, among others, joint projects. This view affords policy-makers, the key target audience of the study's outcomes, to gain a quick overview on regions of active participation in DSI, and how institutions may be working together in this regard. A second visualisation showing a traditional social network allows a focus on network analysis. However, while based on the same data, it loses the additional contextual cues (found in the map), which are of much greater value to users, both policy makers and the general public interested in these questions.

5.2. Bridging knowledge and semantic gaps between communities of practice

Visualisation, whether purely for presenting data or as part of the visual analytics process, is rarely void of text [56, among others]. At its most basic, text provides descriptive labels of elements in the view. Appropriately used, supplementary text provides additional detail that increases understanding of complex data and helps to confirm conclusions drawn from visualisation. In the visual analytics process, users will often also interact with the underlying data using snapshots (for Linked Data, encoded in text as RDF), in summary form, or encoded into filters, to vary input or examine detail in ROIs.

A challenge faced by both lay users and domain experts outside the Semantic Web in such cases, however, is the correct interpretation of such text [18]. Usability evaluation (see Section 3) reveals this to be due largely to differences in mental models of data between tool developers and end users, when developers default to presentation using SW terminology and conventions. For instance, `prefix:Entity` or `prefix:Property` are correctly interpreted by a SW expert, with the (ontology) prefix providing additional, discriminating context. For a non-expert, this only increases visual clutter and cognitive load scanning the view; ontology prefixes use abbreviations and terse forms that require domain knowledge to interpret correctly. Entity, data types and properties vary between self-descriptive labels and terse, technical terminology, both of which are sometimes ambiguous to lay users.

Carpendale [7] illustrates the importance of grounding visualisation in users' context, to enable

and empower them through the resulting visualisations and underlying data. Hart and Dolbear [18], Neish [40] discuss the potential for LD to be used to bridge differences in terminology, standards and conventions across domains and specialised fields, to ensure that solutions that employ LD map to users' context rather than to the technology. It is important that LD visualisation is designed *with*, not just *for*, the user. Petrelli et al. [42] illustrate how a process of user-centred design, closely involving target users in the collection of requirements and the design process, revealed marked differences in expectations and terminology use between two highly specialised communities of practice. Cognisant of this, they employed SW principles to bridge this (not unusual) gap. This ensured that their visualisation-driven solutions fully captured the target users' environment, and translated data structure and content such that it enabled interaction with the content in ways that mapped to users' mental models, goals and expectations. Further, by effectively employing SW technologies, the tools helped to trigger new insight, by enabling the end users to carry out more in-depth investigation of their own data, enriched with related knowledge extracted from additional sources.

5.3. Achieving better synergy with the InfoVis and Visual Analytics communities

Important concepts to take into account in user interface design are simplicity and representativeness. Graph visualisation quickly becomes complex, and as discussed earlier, the low-level structure of RDF graphs, or even the higher-level network formed by LD datasets, will often not be the most relevant representation of the data in terms of content. The graph structure provides much expressive power and facilitates interlinking, but it has been designed with software agents in mind, not human users. The fact that graph visualisation offers a direct mapping to SW experts' mental model [33] of the data has probably contributed significantly to its early use, and as more and more SW users were exposed to such visualisations, the corresponding increase in visual literacy [5,46,56] may have in turn contributed to restricting the range of visualisation techniques considered for LD and the Web of Data at large.

Research in visualisation may arguably be reaching a plateau when it comes to the design of new visualisation techniques [see, e.g., 55]. But there is quite a wide range of existing visualisation techniques that have not yet been considered by the SW community,

that could potentially provide more effective support at representing LD and enable users to interact more intuitively with it. Conversely, better synergy with the visualisation and visual analytics communities could be achieved by getting that community to consider LD as an opportunity to enable more open-ended visual exploration, discovery and analytic processes, thanks to the unique properties of this form of data.

Bridging the gap between these communities is important. However this will only be beneficial if further appropriation of existing and new research in visualisation and visual analytics results in demonstrable, positive impact on research and applications in the LD community [39,55]. Following best practices and guidelines in verified visualisation and visual analysis frameworks and models should help to work successfully toward this goal. Further, because visualisation design guidelines target development for the human end-user, they inherently include an element of user-centred design principles, reinforcing the benefits in weaving these into the process of LD visualisation.

van Wijk [55] examines the particular challenges for effective, insightful visualisation of the increasingly large amounts of complex data by technological and domain experts as well as lay users, by considering "The Value of Visualization" from the perspectives of (i) efficiency and effectiveness or utility in technology; (ii) cost and benefit in economics; and (iii) as a field within the academic disciplines of Art and Science. The paper was written just over ten years ago, as marked advances in technology had started to trigger the "data explosion" now described as *big data*. We see today another step in technological advancement; its findings, however, still remain very relevant, with one that is key to LD – extending research findings beyond the specific academic community and translating these to practical application in the wider world. The paper also presents a seminal model of data-and context-driven, user-focused visualisation that has been extended more recently to capture also the challenges in triggering insight and supporting sense-making and knowledge acquisition through exploration and perception in the iterative visual analytics process [see 47].

We have used the term *value* with reference to Linked Data several times in this paper. However, whether we will in reality reap the value of LD is directly impacted by the cost of realising this value. Cost in terms of time, financial cost, human effort and skill – technological and domain, and other resources required to extract this value and make timely, effective use of it in different contexts, through the use of visualisation [55]. Mitchell and Wilson [39] refer to LD as a "broker" that may be used both to reduce cost and increase the value of today's big data. Hitzler and Janowicz [21] identify LD as an "enabler". Neish [40] illustrates how LD is enabling richer, more effective use of independent and related datasets.

Contributors to this special issue on Linked Data visualisation confirm the potential of LD as a unifier and a bridge across large scale, complex, distributed, heterogeneous data and independent tools. By harnessing visualisation as a tool for exploratory discovery and basic-to-advanced analysis, the papers in this volume illustrate the design and construction of intuitive means for target end users, from a variety of domains and with limited-to-advanced knowledge of technology, to obtain new insights and enrich existing knowledge, as they follow links defined across Linked Data. We look forward to a growing library of shared knowledge and visualisation-driven tools that break down technological barriers, promoting instead richer exploration and intuitive, insightful analysis of users' personal context, myriad, shared situations and complex problems captured in Linked Data, and enable end users to draw confident conclusions about data and situations and add value to their everyday, knowledge-driven tasks.

Acknowledgement

References

[1] B. Bach, E. Pietriga and I. Liccardi, Visualizing populated ontologies with OntoTrix, *International Journal on Semantic Web and Information Systems (IJSWIS)* **9**(4) (2013), 17–40. doi:10.4018/ijswis.2013100102.

[2] F. Benedetti, S. Bergamaschi and L. Po, Visual querying LOD sources with LODeX, in: *K-CAP 2015: Proc., 8th International Conference on Knowledge Capture*, K. Barker and J.M. Gómez-Pérez, eds, 2015, pp. 12:1–12:8. doi:10.1145/2815833.2815849.

[3] T. Berners-Lee, Linked Data – Design issues, 2006, https://www.w3.org/DesignIssues/LinkedData, (Accessed 14.10.2016).

[4] C. Bizer, The emerging web of Linked Data, *IEEE Intelligent Systems* **24**(5) (2009), 87–92. doi:10.1109/MIS.2009.102.

[5] E. Blevis, E. Churchill, W. Odom, J. Pierce, D. Roedl and R. Wakkary, Visual thinking & digital imagery, in: *CHI EA '12: Extended Abstracts, 2012 International Conference on Hu-*

man Factors in Computing Systems, J.A. Konstan E.H. Chi and K. Höök, eds, 2012, pp. 2715–2718. doi:10.1145/2212776. 2212703.

[6] J. Brooke, SUS: A retrospective, *Journal of Usability Studies* **8**(2) (2013), 29–40.

[7] S. Carpendale, Innovations in visualization, in: *GI'13: Proc., Graphics Interface 2013*, F.F. Samavati and K. Hawkey, eds, Canadian Information Processing Society, 2013, pp. 1–8.

[8] R. Cyganiak, D. Reynolds and J. Tennison, The RDF data cube vocabulary, Technical report, The World Wide Web Consortium (W3C), 2014, http://www.w3.org/TR/vocab-data-cube.

[9] A.-S. Dadzie and M. Rowe, Approaches to visualising Linked Data: A survey, *Semantic Web* **2**(2) (2011), 89–124. doi:10.3233/SW-2011-0037.

[10] Data set 2014 – DBpedia, http://wiki.dbpedia.org/data-set-2014, (Accessed 26.10.2016).

[11] Database – Eurostat, http://ec.europa.eu/eurostat/data/database, (Accessed 26.10.2016).

[12] DBpedia, http://dbpedia.org, (Accessed 17.10.2016).

[13] C. Dowd, The new order of news and social media enterprises: Visualisations, linked data, and new methods and practices in journalism, *Communication Research and Practice* **2**(1) (2016), 97–110. doi:10.1080/22041451.2016. 1155339.

[14] I. Ermilov, J. Lehmann, M. Martin and S. Auer, LODStats: The data web census dataset, in: *ISWC 2016: The Semantic Web – Proc., 15th International Semantic Web Conference, Part II*, P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck and Y. Gil, eds, 2016, pp. 38–46. doi:10.1007/978-3-319-46547-0_5.

[15] T.R.G. Green and M. Petre, Usability analysis of visual programming environments: A 'cognitive dimensions' framework, *Journal of Visual Languages & Computing* **7**(2) (1996), 131–174. doi:10.1006/jvlc.1996.0009.

[16] F. Haag, S. Lohmann, S. Bold and T. Ertl, Visual SPARQL querying based on extended filter/flow graphs, in: *AVI'14: Proc., International Working Conference on Advanced Visual Interfaces*, P. Paolini and F. Garzotto, eds, 2014, pp. 305–312. doi:10.1145/2598153.2598185.

[17] H. Halpin and F. Bria, Crowdmapping digital social innovation with linked data, in: *ESWC 2015: The Semantic Web. Latest Advances and New Domains – 12th European Semantic Web Conference*, F. Gandon, M. Sabou, H. Sack, C. d'Amato, P. Cudré-Mauroux and A. Zimmermann, eds, 2015, pp. 606–620. doi:10.1007/978-3-319-18818-8_37.

[18] G. Hart and C. Dolbear, Geographic information in an open world, *Linked Data: A Geographic Perspective*, CRC Press, 2013, Chapter 4. doi:10.1201/b13877-5.

[19] O. Hassanzadeh and M.P. Consens, Linked movie data base, in: *LDOW 2009: Proc., Linked Data on the Web Workshop at WWW'2009*, C. Bizer, T. Heath, T. Berners-Lee and K. Idehen, eds, 2009, http://ceur-ws.org/Vol-538/ldow2009_paper12.pdf.

[20] J. Hendler, Web 3.0: Chicken farms on the Semantic Web, *Computer* **41**(1) (2008), 106–108. doi:10.1109/MC.2008.34.

[21] P. Hitzler and K. Janowicz, Linked Data, Big Data, and the 4th paradigm, *Semantic Web* **4**(3) (July 2013), 233–235. doi:10.3233/SW-130117.

[22] Y. Hu, K. Janowicz, G. McKenzie, K. Sengupta and P. Hitzler, A linked-data-driven and semantically-enabled journal

portal for scientometrics, in: *ISWC 2013: The Semantic Web – Proc., 12th International Semantic Web Conference, Part II*, H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J.X. Parreira, L. Aroyo, N. Noy, C. Welty and K. Janowicz, eds, 2013, pp. 114–129. doi:10.1007/978-3-642-41338-4_8.

[23] IOS Press Journal: Semantic Web – Interoperability, Usability, Applicability, http://semantic-web-journal.net, (Accessed 19.10.2016).

[24] V. Ivanova, P. Lambrix, S. Lohmann and C. Pesquita (eds), *VOILA! 2016: Proc., ISWC2016 International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data*, 2016, http://ceur-ws.org/Vol-1704.

[25] M. Javed, S. Payette, J. Blake and T. Worrall, VIZ-VIVO: Towards visualizations-driven linked data navigation, in: *VOILA! 2016: Proc., ISWC2016 International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data*, V. Ivanova et al., eds, 2016, pp. 80–92. http://ceur-ws.org/Vol-1704/paper7.pdf.

[26] L. Kazemzadeh, M.R. Kamdar, O.D. Beyan, S. Decker and F. Barry, LinkedPPI: Enabling intuitive, integrative protein–protein interaction discovery, in: *LISC2014: Proc., 4th International Workshop on Linked Science – Making Sense Out of Data*, J. Zhao, M. van Erp, C. Keßler, T. Kauppinen, J. van Ossenbruggen and W.R. van Hage, eds, 2014, pp. 48–59, http://ceur-ws.org/Vol-1282/lisc2014_submission_4.pdf.

[27] J. Klímek, M. Nečaský, B. Kostov, M. Blaško and P. Křemen, Efficient exploration of linked data cloud, in: *DATA-2015: Proc., 4th International Conference on Data Management Technologies and Applications*, M. Helfert, A. Holzinger, O. Belo and C. Francalanci, eds, 2015, pp. 255–261. doi:10.5220/0005558002550261.

[28] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik and J. Zhao, PROV-O: The PROV Ontology, Technical report, The World Wide Web Consortium (W3C), 2013, https://www.w3.org/TR/sparql11-overview.

[29] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer and C. Bizer, DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web* **6**(2) (2015), 167–195. doi:10.3233/SW-140134.

[30] Linked Data – Connect distributed data across the Web, http://linkeddata.org, (Accessed 21.10.2016).

[31] Linked Data portal for the Semantic Web Journal, http://semantic-web-journal.com/SWJPortal, (Accessed 19.10.2016).

[32] LinkedMDB, http://linkedmdb.org, (Accessed 19.10.2016).

[33] Z. Liu and J.T. Stasko, Mental models, visual reasoning and interaction in information visualization: A top-down perspective, *IEEE Transactions on Visualization and Computer Graphics* **16**(6) (2010), 999–1008. doi:10.1109/TVCG.2010. 177.

[34] LODStats project, http://lodstats.aksw.org, (Accessed 14.10.2016).

[35] S. Lohmann, P. Heim, T. Stegemann and J. Ziegler, The RelFinder user interface: Interactive exploration of relationships between objects of interest, in: *IUI'11: Proc., 15th International Conference on Intelligent User Interfaces*, C. Rich, Q. Yang, M. Cavazza and M.X. Zhou, eds, 2010, pp. 421–422. ISBN 978-1-60558-515-4. doi:10.1145/1719970. 1720052.

[36] S. Lohmann, S. Negru, F. Haag and T. Ertl, Visualizing ontologies with VOWL, *Semantic Web* **7**(4) (2015), 399–419. doi:10.3233/SW-150200.

[37] G. Mai, K. Janowicz, Y. Hu and G. McKenzie, A Linked Data driven visual interface for the multi-perspective exploration of data across repositories, in: *VOILA! 2016: Proc., ISWC2016 International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data*, V. Ivanova et al., eds, 2016, pp. 93–101. http://ceur-ws.org/Vol-1704/paper?pdf.

[38] C.-J. Minard, Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 (Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812–1813), in: *Tableaux graphiques et cartes figuratives*, Ecole nationale des ponts et chaussées, Fol. 10975, Autogr.Regnier et Dourdet (Paris) edition, 1862. Available at Bibliothèque numérique patrimoniale des ponts et chaussés, (Accessed 28.10.2016).

[39] I. Mitchell and M. Wilson, Linked Data – Connecting and exploiting big data, Technical report, Fujitsu Services Limited, 2012. https://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-(v1.0).pdf.

[40] P. Neish, Linked Data: What is it and why should you care? *The Australian Library Journal* **64**(1) (2015), 3–10. doi:10.1080/00049670.2014.974004.

[41] Open Data Catalog – The World Bank, http://datacatalog.worldbank.org/, (Accessed 26.10.2016).

[42] D. Petrelli, A.-S. Dadzie and V. Lanfranchi, Mediating between AI and highly specialized users, *AI Magazine* **30**(4) (2009), 95–102. Special Issue on "Usable AI".

[43] E. Pietriga, Semantic web data visualization with graph style sheets, in: *SoftVis'06: Proc., 2006 ACM Symposium on Software Visualization*, E. Kraemer, M.M. Burnett and S. Diehl, eds, 2006, pp. 177–178. doi:10.1145/1148493.1148532.

[44] E. Pietriga, C. Bizer, D. Karger and R. Lee, Fresnel – A browser-independent presentation vocabulary for RDF, in: *ISWC2006: The Semantic Web, Proc., 5th International Semantic Web Conference*, I.F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold and L. Aroyo, eds, 2006, pp. 158–171. doi:10.1007/11926078_12.

[45] N.A. Rakhmawati, R.P. Wibowo and M.I.H. Amir, Visualisation application development for mosque financial report using linked data and crowd-sourcing, in: *Procedia Computer Science: The Third Information Systems International Conference 2015*, Vol. 72, 2015, pp. 374–381. doi:10.1016/j.procs.2015.12.152.

[46] P. Ruchikachorn and K. Mueller, Learning visualizations by analogy: Promoting visual literacy through visualization morphing, *IEEE Transactions on Visualization and Computer Graphics* **21**(9) (2015), 1028–1044. doi:10.1109/TVCG.2015.2413786.

[47] D. Sacha, A. Stoffel, F. Stoffel, B.C. Kwon, G. Ellis and D.A. Keim, Knowledge generation model for visual analytics, *IEEE Transactions on Visualization and Computer Graph-*

ics **20**(12) (2014), 1604–1613. doi:10.1109/TVCG.2014.2346481.

[48] m.c. schraefel and D. Karger, The pathetic fallacy of RDF, in: *SWUI 2006: Proc., 3rd International Semantic Web User Interaction Workshop*, 2006, http://swui.semanticweb.org/swui06/papers/Karger/Pathetic_Fallacy.html.

[49] B. Shneiderman, The eyes have it: A task by data type taxonomy for information visualizations, in: *Proc., 1996 IEEE Symposium on Visual Languages*, IEEE, IEEE Computer Society, 1996, pp. 336–343. doi:10.1109/VL.1996.545307.

[50] The Datahub, http://datahub.io, (Accessed 14.10.2016).

[51] The International Human Genome Mapping Consortium, A physical map of the human genome, *Nature* (2001), 934–941. doi:10.1038/35057157.

[52] The W3C SPARQL Working Group, SPARQL 1.1 overview, W3C recommendation 21 March 2013, Technical report, The World Wide Web Consortium (W3C), 2013, https://www.w3.org/TR/sparql11-overview.

[53] T. Tietz, J. Jäger, J. Waitelonis and H. Sack, Semantic annotation and information visualization for blogposts with refer, in: *VOILA! 2016: Proc., ISWC2016 International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data*, V. Ivanova et al., eds, 2016, pp. 28–40. http://ceur-ws.org/Vol-1704/paper3.pdf.

[54] F. Valsecchi, M. Abrate, C. Bacciu, M. Tesconi and A. Marchetti, Linked Data maps: Providing a visual entry point for the exploration of datasets, in: *IESD 2015: Proc., 4th International Workshop on Intelligent Exploration of Semantic Data*, D. Thakker, D. Schwabe, K. Kozaki, R. Garcia, M. Brambilla and V. Dimitrova, eds, 2015, http://ceur-ws.org/Vol-1472/IESD_2015_paper_2.pdf.

[55] J.J. van Wijk, The value of visualization, in: *IEEE VIS 05: Proc., Visualization 2005*, Oct. 2005, pp. 79–86. doi:10.1109/VISUAL.2005.1532781.

[56] C. Ware, *Visual Thinking: For Design*, Morgan Kaufmann Publishers Inc., Burlington, USA, 2010.

[57] M. Weise, S. Lohmann and F. Haag, LD-VOWL: Extracting and visualizing schema information for linked data endpoints, in: *VOILA! 2016: Proc., ISWC2016 International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data*, V. Ivanova et al., eds, 2016, pp. 120–127. http://ceur-ws.org/Vol-1704/paper11.pdf.

[58] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson, The NHGRI GWAS catalog, a curated resource of SNP-trait associations, *Nucleic Acids Research* **42**(D1) (2014), D1001–D1006. doi:10.1093/nar/gkt1229.

[59] D. Young and B. Shneiderman, A graphical filter/flow representation of boolean queries: A prototype implementation and evaluation, *Journal of the American Society for Information Science* **44**(6) (July 1993), 327–339. doi:10.1002/(SICI)1097-4571(199307)44:6<327::AID-ASI3>3.0.CO;2-J.