Cognitive-based Visualization of Semantically Structured Cultural Heritage Data

Kalliopi Kontiza, Antonis Bikakis, Rob Miller

Department of Information Studies, University College London, UK
{k.kontiza.12,a.bikakis,rsm}@ucl.ac.uk

Abstract. We present preliminary findings regarding the increasing use of InfoVis tools and semantically structured data by cultural heritage institutions. This sector faces a number of challenges in developing best practices for publishing Linked Data, including the presentation of their digital cultural heritage collections and the visualization of their multidimensional hidden histories. We suggest that, as these institutions' interest in Semantic Web technologies grows and associated applications are more widely adopted, the need to provide InfoVis tools for efficient overview and exploration of cultural data increases. We postulate that changes in the paradigms for interaction with cultural datasets are also needed, with more focus on users' needs and cognitive processes. We suggest that by taking into account human information processes, better cognitive support can be introduced via InfoVis tools for Linked Data, thus reducing the cognitive load experienced by users.

1 Introduction

The aim of this position paper is to give a brief summary of the current state-of-the-art as regards the use of Information Visualisation (InfoVis) for semantically structured cultural heritage (CH) data, and to suggest some directions for research and development in this area. The Semantic Web (SW), also known as Web 3.0, is the new environment in which cultural digital resources will be exploited. The Resource Description Framework (RDF), which integrates a variety of applications using XML or other machine-readable formats for syntax and URIs for naming, has become the standard data model for describing semantically structured data using statements in the form of triples, which describe resources and can be considered as metadata. Linked Data (LD) is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be identified, and links between related resources can be made. Although the term LD is often used as if it is a specific, well defined technology, it could be better understood as a set of best practices [8]. These aim to get data published on the web in a way that is readable, interpretable and usable by machines, by ensuring that their meaning is explicitly defined by a string of words and markers [6].

LD practices are starting to be introduced as novel and promising approaches to address the specific challenges that the CH sector encounters when publishing "collection" data on the Web [7]. Museums as memory organizations are key players in

preserving CH Tangible[1] objects by storing them with attached metadata. Despite efforts amongst the CH community to analyze the needs of their online 'audiences', no clear understanding about these users and their expectations has yet been gained [8]. Of course, publishing CH contents on the Web cannot replace the physical experience of visiting a museum or an exhibition in reality. But the key question is whether online content, when combined with appropriate Visualization tools, can act as a complementary alternative for access and exploration of cultural data. The hope is that the structure and semantics of RDF representations and standard ontologies, combined with InfoVis techniques, can lead to new ways of thinking about aspects of cognition as emergent properties of the interaction of people with cultural artifacts [16].

We take as a starting point the definition of InfoVis as "the use of computer supported interactive visual representation of data to amplify cognition" [3]. Our long term aim is to investigate why and how InfoVis can assist people in understanding CH information, in particular by exploiting the structure and the semantics of RDF representations and relevant ontologies. More specifically we aim to develop a cognitively based InfoVis model that is geared towards revealing the various temporal, spatial, contextual, conceptual links between different cultural artifacts, their creators and associated events, building upon standard ontologies and vocabularies. From this we hope to develop a set of principles for the design of interactive visual interfaces for exploring and understanding the semantically encoded data used in museum collections.

The rest of the paper is structured as follows. Section 2 provides an overview of some relevant SW terminology and then describes a number of current initiatives that use InfoVis Tools to present semantically structured CH data. Section 3 provides arguments in favor of taking greater account of theories of human cognition when designing CH InfoVis Tools, in order to provide efficient exploration of semantically encoded cultural datasets. Finally, Section 4 summarizes the preceding discussion, and sets out an agenda for future research.

2 Semantic Web and Linked Cultural Data

2.1 Background

Tim Berners-Lee has re-introduced the SW as "a new layer of metadata being build inside the Web" [7]. One of the main principles behind its architecture is that it offers users the ability to share knowledge by constructing meaningful representations on the Web [1]. RDF was published as the first SW standard by the World Wide Web Consortium (W3C) in 1999[2]. But it was only around 2005 when the ideas of SW and LD started to gain momentum and the SW and LD research communities focused on enhancing existing large datasets using simple RDF and lightweight ontologies.

The use of lightweight ontologies based on RDF is one of the reasons for the success of LD in the CH sector. The Dublin Core vocabulary[3] for example, which became popular for expressing metadata as RDF, provides the basis for the Europeana Data Model

[1] Tangible cultural heritage consists of concrete cultural objects, such as artifacts, works of art, buildings and books

[2] http://www.w3.org/RDF/

[3] http://dublincore.org/documents/dcmi-terms/

(EDM)[4] used by Europeana[5] for gathering and publishing metadata for thousands of digitized cultural collections [10]. But even more sophisticated models are provided as simple RDF Schema ontologies. One example is the CIDOC Conceptual Reference Model (CIDOC-CRM)[6], a core ontology for describing the semantics of schema and data structure elements in CH documentation [7]. This was developed as an ambitious attempt to link museum objects across collections ontologically [4]. The CIDOC-CRM aims to facilitate "the integration, mediation and interchange of heterogeneous cultural information, by capturing their richness" [4].

Heath [7] enumerates some of the content related challenges that CH institutions are faced with in order to publish their data collections: their content can be multi-format (text, images, audio, video, collection items, learning objects), multi-topical (art, history, artifacts, traditions, etc.), multi-lingual, multi-cultural and multi-targeted (targeted to both laymen and experts, various ages, etc.). SW standards and LD as best practice could provide a shared basis with which to facilitate content-related and cross-domain semantic interoperability.

Progressive steps have been taken towards LD online publishing. Following the LD principles[7] Tim Berners-Lee also introduced a five-star rating scheme for LD at increasing levels of openness and linkage[8]. For a linked dataset to be accepted by the Open Linked Data Project[9] it must not be subject to commercial licenses or use restrictions.

2.2 Current Initiatives on Linked Data

Figure 1 (larger version at lod-cloud.net) gives an overview of the datasets that have been published online following the Linked Data format and principles. It shows the growing number of CH institutions, web services and museums that have adopted LD standards as a way to turn the rich descriptions of their digital collections into a web of related data. Examples of such institutions include The Getty Research Institute[10] and the Yale Center for British Art[11]. Examples of web services are the Swedish Open

[4] http://www.europeana.eu/schemas/edm/
[5] http://europeana.eu
[6] http://www.cidoc-crm.org
[7] The LD principles formulated by Tim Berners-Lee are: 1) use URIs as names for things, 2) use HTTP URIs, 3) when someone looks up a URI, provide useful information using the standards (RDF, SPARQL), 4) include links to other URIs, so that they can discover more things.
[8] http://www.w3.org/DesignIssues/LinkedData.html
[9] http://linkeddata.org/
[10] http://www.getty.edu/research/tools/vocabularies/lod/
[11] http://britishart.yale.edu/collections/using-collections/technology/linked-open-data

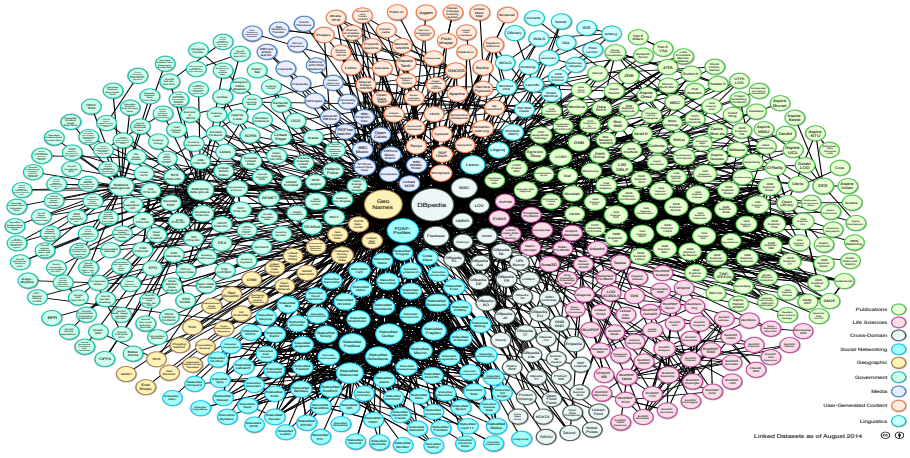

Fig. 1. The Linked Open Data cloud (August 2014) consists of datasets (shown as bubbles) in different domains (colour identified) and mappings between similar sources in the dataset (shown as arcs between bubbles). *Source: Linking Open Data cloud diagram 2014, Schmachtenberg, Bizer, Jentzsch and Cyganiak.* `http://lod-cloud.net/`.

Cultural Heritage (SOCH)[12], Europeana[13], and Museos de España[14]. Specific museums include various Italian Museums[15], The Smithsonian American Art Museum[16], The British Museum[17], The Victoria and Albert Museum[18], and The Rijksmuseum[19].

The integration of semantics-based interaction paradigms showcases the benefits of semantically marked up information, delivered to the end user through a *semantic browsing facility*, and highlights an important feature of the SW for the CH domain; the ability to explore and navigate relationships. The SCULPTEUR project focused initially on the semantic interoperability of LD and further developed the mSpace interaction framework[20], resulting in an interface that provides the users with the ability to organize their information space. The interactive interface aimed to allow users unfamiliar with

[12] A web service used to search and retrieve data from any organization holding information or media relating to Swedish cultural heritage, `http://www.ksamsok.se/in-english/`

[13] Aims to support research of the digitised content of Europe's galleries, museums, libraries and archives by addressing issues such as licencing, interoperability and access, `http://labs.europeana.eu/api/linked-open-data/introduction/`

[14] a directory that allows facet-based searches on a collection with more than 1,500 dynamic pages of public and private Spanish museums, `http://museos.gnoss.com/comunidad/mismuseos/`

[15] `http://www.linkedopendata.it/datasets/musei`

[16] `http://americanart.si.edu/collections/search/lod/about/`

[17] `http://collection.britishmuseum.org/`

[18] `http://www.vam.ac.uk/api/`

[19] `https://www.rijksmuseum.nl/en/api`

[20] `http://www.w3.org/2001/sw/wiki/MSpace`

museum collections to navigate, visualize and explore rich sources of cultural heritage information [19]. In particular, the application of InfoVis tools in mSpace helps people to develop knowledge by exploring relationships in data, and to customise access to the content to suit their individual interests by "slicing, sorting, swaping, information views and multimedia preview cues" [18].

MUSEUMFINLAND (Finnish Museums on the Semantic Web)[21] is a semantic portal for publishing heterogeneous museum collections on the Semantic Web in order to provide the museum visitors with intelligent content-based search and browsing services [9]. The view-based multi-facet search engine exploits the semantic link recommendation system to reveal the underlying semantic context of the collection items and their mutual relation [9]. CultureSampo[22] is a continuation of this work. The CultureSampo interface has been enhanced with InfoVis tools that allow the user to explore cultural heritage through nine semantic perspectives/thematic views: Maps and historical places, relational search, faceted domain-centric browsing, collections, Finnish history, cultural processes and skills, biographies, semantic Kalevala, and Carelia[23]. The main concept behind the interface is "to let users create virtual exhibitions that mimic the way real museums are organized, containing themed exhibition rooms of items and displays that together, through the objects, tell the story of a particular subject" [14].

Through the CHIP (Cultural Heritage Information Personalization) project[24], the Rijksmuseum focused on delivering novel personalization functions for the visitors on the museum's website. The CHIP demonstrator included three innovative components; the Art Recommender, the Tour Wizard and the Mobile Tour Guide, where the SW was deployed to support "the presentation of recommendations by combining different views like a historical timeline, a museum map and a faceted browser"[20]. The evaluations of the CHIP demonstrator provided critical insight in how to further adapt the user interaction facilities and interface to suit user's needs and preferences.

The ResearchSpace project[25] is a contextual search system that allows searching against objects, people, places, events, periods and concepts, providing context by making use of semantically enriched cultural data. The new design of the ResearchSpace search interface is a radical departure from traditional keyword and advanced searching, and can be customized to suit different online audiences.

Finally, the Russian Linked Culture Cloud[26] project is a collaboration between The Russian Museum and ITMO University, and is based on open data related to Russian Culture heritage. SW technologies, enabling enrichment of initial data with other facts of Russian culture, provide an advanced user experience with the help of visualization and navigation through enriched text, interactive timelines and an interactive influence graph [15].

[21] http://www.museosuomi.fi/
[22] http://www.kulttuurisampo.fi/
[23] http://www.kulttuurisampo.fi/about.shtml?lang=en
[24] http://www.chip-project.org
[25] http://www.researchspace.org/home/project-information/design
[26] http://culturecloud.ru/

3 Cognition-based InfoVis Tools for Linked Datasets

As industry's and society's interest in SW technologies grows and the number of widely adopted SW applications increases, there is an opportunity to focus on the properties and characteristics of semantically encoded data that can most readily be used to enhance InfoVis tools. In the case of online museum collections, SW technologies can have a beneficial impact on (a) exploring and navigating *relationships*, as richer semantics highlight the conceptual relationships between artifacts, and (b) *presentation-interaction*, as they offer richer presentation possibilities in terms of browsing and navigation.

Linked data interoperability in the semantic web has recently received much research attention. However, the emphasis has largely been in automating the mapping process to standards, even though the creation of mappings often involves the user. The main Linked Data users are technology experienced, and one reason for this is the lack of appropriate user interfaces and visualizations for non-expert users. Visual approaches are needed to assist various kinds of users, who pursue diverse goals and have individual requirements. InfoVis tools developed using a human-cognition-based model will cetainly improve users' engagement with and understanding of semantically encoded cultural data.

Case studies of museums that have implemented personalization facilities to their Web sites show that understanding is stimulated when the systems use concepts familiar to the user [2]. Visuals help understanding by acting as a frame of reference or as a temporary storage area for human cognitive processes. By providing a larger working set for thinking and analysis they become external cognition aids [12]. Card, Mackinlay and Shneiderman [3] list some key ways in which visuals can amplify cognition:

(i) Increasing memory and processing resources available
(ii) Reducing search for information
(iii) Enhancing the recognition of patterns
(iv) Enabling perceptual inference operations
(v) Using perceptual attention mechanisms for monitoring
(vi) Encoding information in a manipulable medium

One of the main aims of InfoVis is to amplify and augment the cognition of users. But a key challenge of the field is to measure its effectiveness in this respect. The absence of a cognitive-based framework for the evaluation of InfoVis systems makes the significance of achievements in this area difficult to describe, validate and defend. According to [13] even though InfoVis research has matured technically in recent years, an important problem for the field remains the lack of an underlying theory or even a systematic framework for guiding design and investigation. InfoVis classifications have focused on the process of synthesising and displaying data, and how to standardise this from a computational point of view. Considerations of how users interact with the resulting interfaces have only recently been integrated into this area of research. A greater emphasis on human cognition has the potential to uncover new ways of presenting, searching, exploring and visualizing the available semantic data in order to enhance human understanding. For example, the very recent work of Patterson et al [16] links

InfoVis with high-level cognitive processes such as reasoning and thinking, and Liu et al [13] argue that the use of "distributed cognition ... has the potential to serve as a theoretical framework for InfoVis". Other approaches that have revealed potential research opportunities in this area include Greene and Petre's Cognitive Dimensions [5], Johnson-Laird's work on model theory and reasoning using visual notations [11] and Peirce's systems of diagrammatic logic[17].

In the particular context of cultural heritage, a cognitive-based framework for the design of InfoVis interactive systems can improve visitors understanding of collections and their ability to explore the cultural information according to their needs. This will help visitors to discover the interconnectedness of digital cultural collections, enable information to be presented attuned to their interests and background, and therefore increase users interest and engagement with both digital and physical collections.

4 Summary and Conclusions

This paper presented some preliminary findings from an ongoing investigation into the use of InfoVis for semantically structured cultural heritage data. We gave a summary of the current state-of-the-art in this area, followed by an argument for more user-focused research and development, which will draw on models and theories of human cognition. Since the ultimate purpose of InfoVis in this context is to enhance users' ability to explore and understand (cultural) data, its effectiveness must be measured in these terms, and attention must therefore be paid to ordinary users' cognitive processes. Our future research plan is to use data from observational studies together with ideas from cognitive psychology in order to develop a theoretical framework for guiding the design and evaluation of cultural heritage InfoVis tools. Such tools will allow users to explore and interact with cultural data with only minimal additional cognitive load. An additional longer term aim is to apply this framework to produce an actual exemplar for a new generation of online InfoVis tools for exploration of museum data collections.

References

1. Alesso, H.P., Smith, C.F.: Thinking on the Web: Berners-Lee, Goedel, and Turing. Wiley, Hoboken, N.J. (2008)
2. Bowen, J., Fantoni Filippini, S.: Personalization and the Web from a Museum Perspective. In: Museums and the Web 2004: Proceedings. Archives & Museum Informatics, Washington DC / Arlington VA, USA (2004)
3. Card Stuart K. : Readings in information visualization : using vision to think / written and edited by Stuart K. Card, Jock D. Mackinlay, Ben Shneiderman. Morgan Kaufmann series in interactive technologies Y, Morgan Kaufmann Publishers (1999)
4. Eide, O., Felicetti, A., Ore, C.E., DAndrea, A., Holmen, J.: Encoding cultural heritage information for the semantic web. procedures for data integration through cidoc-crm mapping. In: Open Digital Cultural Heritage Systems Conference. p. 47 (2008)
5. Green, T.R.G., Petre, M.: Usability analysis of visual programming environments: a cognitive dimensions framework. Journal of Visual Languages & Computing 7(2), 131–174 (1996)
6. Guerrini, M., Possemato, T.: Linked data: a new alphabet for the semantic web. JLIS. it 4(1), 67 (2013)

7. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space, vol. 1 (2011)
8. Hooland, S.v.: Linked data for libraries, archives and museums: how to clean, link and publish your metadata. Facet Publishing, London (2014)
9. Hyvonen, E., Makela, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinlandFinnish museums on the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 3(2-3), 224–241 (2005)
10. Isaac, A., Baker, T.: Linked data practice at different levels of semantic precision: The perspective of libraries, archives and museums. Bulletin of the American Society for Information Science and Technology 41(4), 34–39 (2015)
11. Johnson-Laird, P.: 8 the history of mental models. Psychology of reasoning: Theoretical and historical perspectives p. 179 (2004)
12. Kerren, A.: Information visualization human-centered Issues and perspectives. Springer (2008)
13. Liu, Z., Stasko, J.T.: Mental models, visual reasoning and interaction in information visualization: A top-down perspective 16(6), 999–1008 (2010)
14. Makela, E., Hyvonen, E., Ruotsalo, T.: How to deal with massively heterogeneous cultural heritage data lessons learned in culturesampo. Semantic Web Interoperability, Usability, Applicability 3(1) (2012)
15. Mouromtsev, D., Haase, P., Cherny, E., Pavlov, D., Andreev, A., Spiridonova, A.: Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing. In: The Semantic Web. Latest Advances and New Domains, pp. 637–651. Springer (2015)
16. Patterson, R.E., Blaha, L.M., Grinstein, G.G., Liggett, K.K., Kaveney, D.E., Sheldon, K.C., Havig, P.R., Moore, J.A.: A human cognition framework for information visualization. Computers & Graphics 42, 42–58 (2014)
17. Pietarinen, A.V.: Peirce and the logic of image. Semiotica 2012(192) (2012)
18. Pitzalis, D., Lahanier, C., Pillay, R., Aitken, G., Russell, A., Smith, D.A., Sinclair, P.A.S., Addis, M.J., Lowe, R., Hafeez, S., others: Semantically exposing existing knowledge repositories: a case study in cultural heritage (2006)
19. Sinclair, P.A.S., Goodall, S., Lewis, P.H., Martinez, K., Addis, M.J.: Concept browsing for multimedia retrieval in the SCULPTEUR project (2005)
20. Wang, Y., Stash, N., Aroyo, L., Gorgels, P., Rutledge, L., Schreiber, G.: Recommendations based on semantically enriched museum collections. Web Semantics: Science, Services and Agents on the World Wide Web 6(4), 283–290 (2008)