Combining Cultural Heritage Related Web Resources in 3D Information Landscapes

Doron Goldfarb   Max Arends   Josef Froschauer   Dieter Merkl   Martin Weingartner

Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11, 1040 Vienna / Austria
vsem@ec.tuwien.ac.at

In recent years, the number of data collections that are publicly available via the Internet has dramatically increased. Web based semantic databases like Wikipedia derivate DbPedia extend the functionality of Wikipedia by allowing semantic annotations to the information that is usually in form of free text. Since these data collections also cover broad topics related to the Cultural Heritage (CH) domain, they might be well suited to serve as data sources for Cultural Information Systems. Besides the mentioned community based knowledge bases, a number of dedicated Web galleries, for example the Web Gallery of Art (WGA), offer huge digital collections of artworks and relevant metadata. Moreover, renowned institutions like the Getty Foundation provide rich vocabularies of CH related terms that are rather aimed at professional users. We believe that a combination of such different sources might provide interesting insights that would not be available when using each source alone. We therefore propose a system that combines these sources and displays them by using methods from Information Visualisation. We apply the metaphor of 3D Information Landscapes, based on a graph visualisation showing artists and their artworks as nodes and their mutual relationships such as teacher/student, parent/child etc. as edges. The resulting network is drawn in chronological order, thus allowing users to explore art history in a new way.

Information Visualisation, Cultural Heritage, 3D Virtual Environments, Semantic Web

1. INTRODUCTION

By the time when André Malraux wrote his widely acclaimed essay "Le musée imaginaire" (Malraux 1947) he most likely did not foresee the implications that the emergence of the World Wide Web (WWW), whose hypertext based structure had at about the same time been envisioned by his contemporary Vannevar Bush (Bush 1945), would eventually have to the museums of the future.

Indeed, from today's point of view, the appearance of the WWW has had great impact on the way today's museums present themselves as institutions and how they deal with their audiences (Bearman & Trant 2002). Additionally, this development also had great impact on the wide availability of (digital) Cultural Heritage (CH) resources in general.

The recent advances of the read/write Web, better known as Web 2.0, have further increased the impact of IT on CH institutions. Seeing institutions shifting their interpretation policies towards bidirectional communication (Coldicutt & Streten 2005) with their visitors does reflect the trend towards democratisation of content production and communication as can be seen in community based knowledge-bases such as Wikipedia.

With more and more CH institutions - galleries, libraries, archives and museums (GLAMs) - sharing their collections and related contextual information online, and - most likely also driven by the increased availability – a growing number of independent users contributing to collaborative CH resources, new questions arise on what to do with such massive collections of data. The discussion "professionally curated" vs. "crowd-sourced" knowledge is ongoing and will one day probably reach an equilibrium in accepting that a combination of both approaches might be the best way to go. This is because, on the one hand, professional research and curation are the indisputable vital core of establishing new knowledge about CH artefacts, their (historic) context and function, while, on the other hand,

community based knowledge benefits from the participation of a huge number of people, thus opening up new views and personal experiences that might not be uncovered otherwise (Oomen & Aroyo 2011).

Given the huge effort that is being invested in establishing interoperability standards for data interchange, the large-scale „global database" of combined museum collections and related content has yet to come. But nevertheless, in a smaller scale, it is already possible to explore the possibilities that arise from the combination of different data sources into a larger body of information.

While "traditional" means of presenting CH artefacts on the Web rather focus on displaying the artefacts "one by one", linking them together according to well established criteria like "same artist", "same epoch", "same style" etc., the availability of additional context information poses a challenge to find interesting new concepts of presentation.

In this paper, we envision a novel way how to convey Cultural content that is available on the Web. We show how an available collection of artworks can be augmented with contextual information from both institutional and community based sources. Then we discuss the benefit of our approach towards presenting the resulting knowledge base within a 3D visualisation environment instead of relying on traditional Web site design.

The first section contains an overview of our motivation for our approach and related work. Then we give a brief description of the data sources that we are currently using and how we combine them. Afterwards, we introduce the 3D environment in which the knowledge is presented, focusing on the Information Landscape approach and its implications. This is followed by a discussion of the current state of the project and unsolved problems. Finally, we briefly mention interesting issues that we would like to cover in the future.

2. BACKGROUND

There have been numerous approaches to present digital Cultural Heritage, therefore we give a very brief overview capturing the most prominent solutions, focusing on the presentation of CH artefacts online on the WWW - for a thorough discussion see for example (Arends et al. 2011)

Early efforts in the 1990's were mainly focused on offline CD-Roms that were usually on sale in museum shops, offering access to either virtual

walkthroughs through the museum's premises or authored multimedia presentations focusing on the biography and œuvre of specific artists. With the advent of the Web 1.0 a number of online "virtual museums" appeared on the Web, both maintained by independent enthusiasts and early adopting institutions showing more than just their address and opening hours, relying mostly on traditional html-based presentations but occasionally also on Quicktime VR or even VRML based 3D environments.

The advent of dynamic website generation technologies like php enabled GLAMs to put (parts) of their collections online by offering Web based query front-ends to their collection databases. This was a first step to customise the displayed content according to the user's interest, offering possibilities to search and explore the museum's content.

With the shift towards Flash-based presentations and web application architectures, more sophisticated approaches like the Heilbrunn Timeline of Art by the Metropolitan Museum of Art [i] became possible, providing users with increased interactivity such as scrollable timelines with contextual and chronological information and interactive maps.

The advent of Web 2.0 technologies eventually introduced social features like tagging, rating and comments, encouraging some "early adopter" institutions to shift the focus from strictly curatorial presentation to alternative contextualisations provided through user generated vocabularies such as folksonomies.

The current state of the art is characterised by Web presences of GLAMs with integrated Web 2.0 features (e.g. Brooklyn museum [ii]), collaborative art history portals like Smarthistory [iii] and platforms that offer access to shared institutional data, as for example the Google Art Project [iv] combining collections from 17 institutions across the world, offering high-resolution displays of artworks and Google Street View like walkthroughs through the institution's premises.

Given these recent developments, we want to focus on three key aspects of which two - inter-institutional data sharing and user collaboration - are already actively being adopted within the field of digital CH, while the third - 3D user interfaces - will most likely follow in a short while, given the recent wave of 3D displays pushed on the market. In addition, will briefly introduce a use case - visualising art history - that we chose as starting point for our prototype, integrating the aspects "shared data" - "collaborative data" and 3D interfaces.

2.1 Data shared by institutions

Throughout the recent years, the idea of the "Semantic Web[v]" has gained more and more attention. It reflects the dream of a global knowledge base containing all the information that is already present within the "current" Web, but in machine readable form instead of its current state of being mainly represented as free-text, therefore largely only accessible for human readers. Moreover, it also calls for opening up the so-called "data-silos" hidden in the "deep-web", that is, all the information that is stored in separated databases that are, at best, only accessible through query front-ends.

With respect to GLAMs, this idea implies that once they agreed on common data standards, they would be able to share their information about their artefacts with each other, but also with the rest of the world, accessible by individual machines using dedicated approaches like RDF[vi] and related Query Languages. This could lead to new insights in research, but also to more democratic access for the public.

The International Council of Museums (ICOM) maintains a working group that has been dedicated to establishing such a standard, the CIDOC-CRM[vii]. GLAMs around the world are encouraged to implement the standard for better interoperability.

Other groups like the Getty Research Institute put their effort into controlled vocabularies such as the Getty Union List of Artist Names (ULAN)[viii], encouraging indexers to use a common vocabulary for descriptive data. Together with initiatives like the Visual Resources Association (VRA)[ix] they moreover provide data standards in form of core attributes and best-practice models for describing CH artefacts in a consistent manner, with the aim to make inter-institutional data exchange less error-prone and time consuming.

Of course, the high diversity of CH in general, its truly global scale arising from the multitude of the world's cultures, makes it almost impossible to agree on single standards that are able to capture every detail. Nevertheless, as recent projects like Google Art Project have shown, it is an exciting idea to get access to a shared collection of CH artefacts - although in the present case, rather limited with respect to what could be possible.

2.2 Collaboratively maintained data collections

Under the umbrella term "Web 2.0" a number of technologies - or better: usage paradigms - have emerged, introducing the user that was before rather perceived as a mere consumer of information as an active participant in the creation of Web content. As previously mentioned, a number of GLAMs have already implemented core social features on their Web presences - such as tagging, rating, comments etc.. Some institutions even share parts of their collections on social platforms like Flickr Commons, actively seeking for participation by external users. Nevertheless, an ongoing discussion is centred on the mutual implications arising from the intersections between "classical" institutions and collaborative platforms like Wikipedia. The Wikipedia:GLAM/BM[x] project for example is a pilot bringing active Wikipedia users and professionals from the British Museum together in order to find and evaluate strategies for collaboration.

Indeed, the quality present in collaborative platforms is far better than one might assume, as such these platforms have developed sophisticated mechanisms of self-control and quality assurance. It is therefore of interest to take data from such platforms into consideration for CH information systems.

2.3 3D Web Interfaces

While the previously mentioned examples mostly utilise common 2D Web interfaces, a number of approaches have also proposed 3D interfaces as means of Web based communication of CH artefacts. The umbrella term "Web3D" covers technologies that enable access to the WWW through 3D interfaces, either using standalone applications or browser plugins. The recent introduction of a common standard WebGL[xi] tries to foster the integration of 3D technology directly into Web browsers, thus eliminating the need for standalone applications or special plugins.

Significant research has been dealing with the platform Second Life[xii], enabling users to wander 3D virtual worlds, created by themselves or by others, in form of an avatar, being able to interact with other visitors, track their actions and interests. A number of institutions have built virtual presences within the platform, as for example the Dresden Gallery offering a complete virtual representation of its premises. See (Hazan 2010) for an in-depth discussion of CH in Second Life.

While platforms like Second Life - although extending the possibilities of the real world through, for example, flying - rather aim at reproducing the physical appearance of the real world, other approaches utilise more abstract, Information Visualisation methods. (Ruffaldi et al. 2008) for example propose a system that is based on Information Landscapes, a virtual environment where the surroundings - the "landscape" - are composed from the underlying data only. Data itself

defines such an environment, its boundaries, its appearance, thus enabling visitors to directly immerse in the content.

With regard to abstract Information Visualisation, the advantages and disadvantages of 3D interfaces have been thoroughly discussed - see (Card et al. 1999) for basic considerations and (Tavanti & Lind 2001) or (Risden et al. 2000) for 2D vs 3D comparison studies - obtaining mixed results on their usability, depending on the concrete task and the available technology. Given the recently revived trend towards 3D content and consumer-level, affordable stereoscopic 3D display technologies, it can be assumed that 3D user interfaces will again get into the focus of attention in the near future.

2.4 A Use Case: Visualising Art History

The issues mentioned throughout this section made us think about a use case for building a prototype that addresses the combination of available institutional data-sources with data from collaborative platforms, using 3D interfaces for its presentation to an audience that consists of individual persons with interest in the field, so to say, the typical Wikipedia users.

We were inspired by works such as (Kwastek 2003), proposing the development of new, interactive visual interfaces as a chance for the field of Art History to benefit from the extended possibilities offered by digital media. The overview of historic approaches to visual representations of historical time in (Davis et al. 2010) further strengthened our idea to start with a prototype that aims at visualising Art History using a chronological approach.

Last but not least, the available data sources, especially data from the Getty Union List of Artist Names (ULAN) and data from the collaborative Genealogy of Influence[xiii] project suggested to focus on visualising Art History as a chronological flow of relations between historic actors.

We therefore implemented a prototype that visualises the relations between artists, patrons and other involved people, mutually associated with each other by either being members of the same family or being situated in teacher/student, principal/agent or more general relationships. The relationships are displayed in a 3D visualisation, allowing users to immerse into the chronological flow of Art History, following paths through works of art along the personal relationships of their respective creators.

3. DATA SOURCES

Throughout this section, we describe the data sources that we currently use in our prototype and give reasons for our specific choices. The last subsection gives a brief overview of our technical infrastructure regarding the data integration. In general, we employ data standards that follow the principles of the idea of the Semantic Web, namely RDF based data representation and storage, because their open nature is well suited to the integration of different data-sources.

3.1 The Union List of Artist Names (ULAN)

In order to visualise relations between artists, we needed reliable biographical information about their life. The Getty ULAN provides such data that is maintained by CH professionals and is regularly updated with new or updated facts. The list of preferred/non-preferred names for each artist is a necessary feature in order to identify one and the same person across different data-collections. The birth/death date and place information enables to locate the lifespan of the person within chronological and geographical boundaries. Including the multitude of different associative relationships (teacher/student, parent/child etc.) that are stored for a significant fraction of the featured persons, the ULAN thus forms the basis of our chronological visualisation.

3.2 Extending the ULAN with collaborative data

Due to its professional orientation, the ULAN only contains connections that are reasonable from a professional point of view. Thus, the relationships between persons in the ULAN are limited to traceable Art Historical facts.

We believe, however, that the casual user might be less rigorous regarding the strict verifiability of relations, but would also consider general influenced/influenced by artist to artist relationships that have been identified by other users with comparable background knowledge and interest.

Therefore, at this point, we integrate relationship data from collaborative platforms into the network graph from the ULAN, thus extending it with potentially less scientifically sound, but probably nevertheless interesting facts about the relationships between artists. One benefit of the extension are relationships that span beyond generations of artists, as Wikipedia for example contains the information that Pierre-Auguste Renoir was influenced by Raphael Sanzio, a relationship not covered by the ULAN.

We use the semantic Wikipedia derivative DBpedia[xiv] and Freebase[xv] for that purpose. Both platforms are based on collaborative editing, the former being extracted from Wikipedia into a semantic machine-readable format, the latter providing a machine-readable semantic database that can be extended by everyone in a Wikipedia-like manner. Figure 1 illustrates an example for such a combination of relationships.
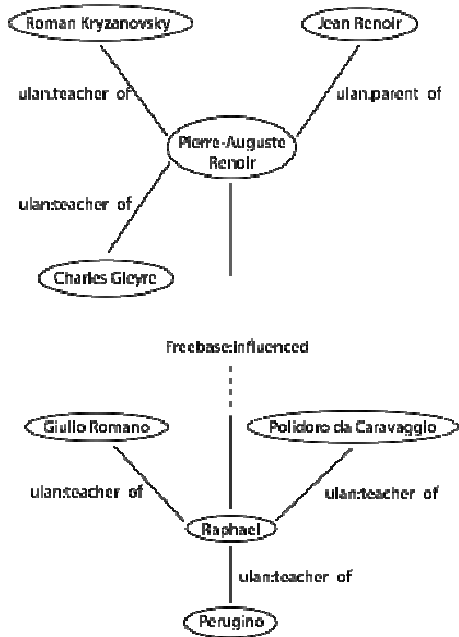

Figure 1: Extending ULAN relationships with Freebase

3.3 The Web Gallery of Art (WGA)

Displaying relevant artworks for each featured artist is an essential element of the visualisation. As image collection, we chose the Web Gallery of Art[xvi] (WGA) as a starting point - eventually, our system should incorporate visual resources from different data-sources. The WGA collection offers digital reproductions and metadata on currently more than 26,000 artworks and more than 2,900 artists covering a timespan between approximately 1000-1850. The metadata provided is, however, limited to information on artist name and biographical data, artwork title, its current location and rough content classification such as type of artwork (painting, sculpture, etc.), general theme of the artwork (religious, still-life, genre, etc.) and the school (Italian, Flemish, German etc.).

Using the list of alternative names for artists in the ULAN, we were able to match a huge proportion of artists from the WGA with the ULAN. We integrate the two remaining Getty vocabularies (TGN & AAT) in order to match the current location of a WGA artwork and its related type, theme and school information respectively.

3.4 Technical infrastructure

We chose the data storage method and platform following the approach proposed in (Wielemaker et al. 2008), as it already provides a framework for an RDF based integration of the Getty Vocabularies with visual resources that are described using the VRA Core standard. Moreover, we use their open-source platform ClioPatria[xvii] as data storage platform, as it provides a standard SPARQL endpoint and various APIs for data retrieval and presentation.

4. THE INFORMATION LANDSCAPE

The goal of the visualisation is to provide a 3D immersive environment where users can travel through time, following the chronological relationships between artists while looking at their related artworks, thus choosing their own paths through Art History.

4.1 Construction of the Landscape

The directed graph based nature of the extended ULAN relationships - artists are nodes and their relationships are the directed edges - immediately suggests the choice of a graph drawing algorithm for visualisation. The visualised network appears like a landscape that is directly constructed from the relationships from the underlying data.

The Information Landscape metaphor is a rather old principle of displaying hierarchical structures within 3D immersive environments (Andrews et al. 1996) but has proved to be effective with respect to alternative 3D approaches for comparable data-collections (Wiss & Carr 1999). We therefore had to choose an appropriate graph layout algorithm in order to transform the data relations into a chronological timeline representation. We chose to implement an approach based on the well established DAG algorithm (Gansner et al. 1993) for drawing layered directed graphs, assigning the artists birth dates to layers, leading to a layered graph representation with each layer representing an arbitrary timespan in history. Figure 2 shows a layered graph representing the close relationships around Leonardo Da Vinci.

The layout created by the algorithm is flat by default, so the height dimension can be chosen to encode additional characteristics of the data. As a starting point, we chose to encode the number of connections to and from a node as its height, therefore visibly promoting nodes that have more connections than other nodes, forming hills in the landscape. This reflects the concept that an artist with more incoming and outgoing relationships played a more important role in Art History. This is,

of course, a strong simplification of reality, as artists with many documented relationships were not necessarily more influential than other artists with no or only few documented relations. This, however, is just an initial encoding and can easily be changed to other measures, as for example the number of produced artworks.
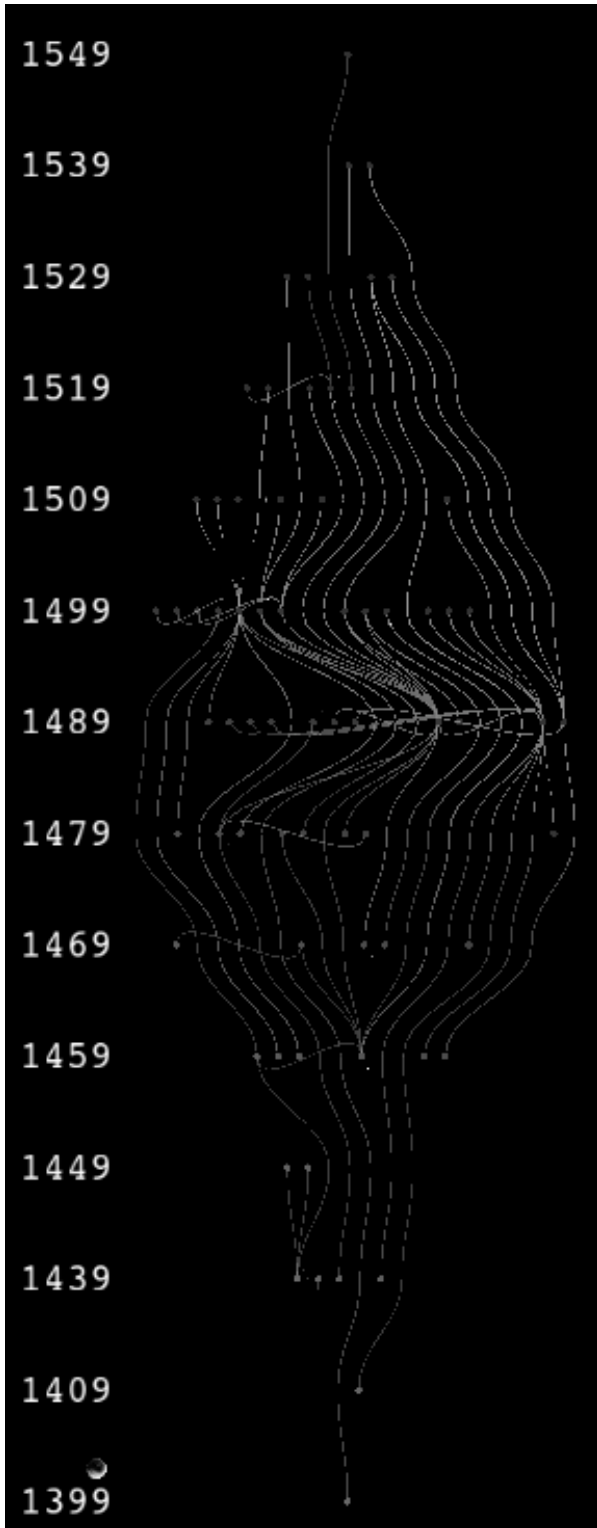

Figure 2: Relationships of Leonardo Da Vinci

The resulting Information Landscape is shown in Figure 3. The view is directed along the time axis, looking "into the future". Colouring of the graph's edges is used to denote the different types of relationships like teacher/student, patron of, etc.. Around the nodes representing artists, a selection of their artworks is presented. When a visitor gets close to the artwork, labels with respective metadata are displayed.

4.2. Technical infrastructure

We use the Unity Game Engine[xviii] for the realisation of the Visualisation, as it provides a powerful, yet low-cost programming environment that is suitable for interactive 3D environments. The availability of a browser plugin enables users to access the environment with a standard Web-browser.

5. NAVIGATION & INTERACTION

A visit to the environment is currently started with entering a name of an artist into a text-box. An autocomplete feature displays matching names while the user is typing. After clicking on a name, the data is downloaded from the Web and he Landscape is built.

We provide three distinct navigation aids. A label in the lower left corner of the screen displays the historical date that the current position of the user corresponds with. An overview map on the right side of the screen provides a birds view of the whole graph, making it easier for the user to locate him/herself in the timeline. A compass needle is shown in the upper left corner, providing instant feedback on the user's orientation with respect to the flow of time. See Figure 3.

Currently, we have implemented two modes of navigation. The first mode allows the user to move freely, pointing the mouse to the desired direction and moving forwards/backwards upon left/right mouse button input. The second mode automates user's movements by allowing him/her to click on a relation between two nodes, automatically translating the user's position following the direction of the relationship in form of an animation.

6. DISCUSSION

We bellieve that the immersive 3D representation of relationships between Art Historical actors provides a new form of experiencing Art History. Since the user experience is directly focused on persons, artworks and their relationships, we expect this setting to motivate people for playful

exploration. The current state of the prototype reflects our initial ideas about how such an environment can be realised. We are planning to perform a qualitative evaluation of the environment in order to get feedback regarding wanted features and the effectivity of the navigation.

We are aware of a number of limitations that we want to address in future. The position of the artist nodes in the timeline is currently determined by their birth date, which is only a very rough measure, as the duration of life of different humans may vary greatly. One solution could be to take the average date between birth and death as position for the node in the timeline, another solution could be to step away from the node based representation and choose lines reflecting the lifespan of a person. This would, however, have a strong impact on the visual representation of the network.

We will continue to look for interesting sources to be included in our knowledge-base, the next step will be to identify portraits of the featured people within DBpedia. We expect the combination of the ULAN data with data from collaborative platforms to open up new paths through Art History, from classic art to contemporary art, from well-known artists to almost unknown artists. Evaluating the reasonableness of such an approach is therefore also mandatory.

7. CONCLUSION

We have presented a novel way of experiencing Art History online, by combining different data sources and presenting the resulting body of knowledge in form of a 3D Information Landscape. We are aware of the current limitations of the approach and are looking forward to refine the system from user evaluation.

8. ACKNOWLEDGEMENTS

Figure 3: The Information Landscape

9. REFERENCES

Andrews, K., Pichler, M. and Wolf, P. (1996) Towards rich information landscapes for visualising Web spaces, In *Proceedings of the IEEE Symposium on Information Visualisation '96*, San Francisco, CA, Oct. 1996, pp. 62-63, 121. IEEE Computer Society, Washington, DC, USA.

Arends, M., Goldfarb, D., Merkl, D. and Weingartner, M. (2011) Museums on the Web: Interaction with Visitors, In G. Styliaras, D. Koukopoulos and F. Lazarinis (eds), *Handbook of Research on Technologies and Cultural Heritage: Applications and Environments*, IGI-Global, Hershey, PA, USA.

Bearman, D. and Trant, J. (2002) Cyberspace is Our Space. In J. Trant and D. Bearman (eds.), *Museums and the Web 2002: Proceedings,* Archives & Museum Informatics, Toronto.

Bush, V. (1945) As we may Think. In *The Atlantic Monthly,* 176(1), pp. 101-108, July 1945.

Card, S., Mackinlay, J. and Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think.* Morgan Stanley, San Francisco, CA, USA.

Coldicutt, R. and Streten, K. (2005) Democratize And Distribute: Achieving A Many-To-Many Content Model, In J. Trant and D. Bearman (eds.), *Museums and the Web 2005: Proceedings*, Archives & Museum Informatics, Toronto.

Davis, S. B., Bevan, E. and Kudikov, A. (2010) Just in time: defining historical chronographics, In Seal, A. with J. Bowen and K. Ng (eds), *EVA London 2010: Electronic Visualisation & the Arts. Proceedings*, London 6-8 July, BCS, 2010.

Gansner, E. R., Koutsofios, E., North, S. C. and Vo, K.-P. (1993) A Technique for Drawing Directed Graphs, In *IEEE Transactions on Software Engineering*, 19(3), pp. 214-230, IEEE Press Piscataway, NJ, USA.

Hazan, S. (2010) Musing the Metaverse, In Ioannides, M., Alonzo, A., Georgopoulos, A., Kalisperis, L., Brown, A. and Pitzalis, D. (eds), *Heritage in the Digital Era*, Multi-Science Publishing, Brentwood, Esse, UK.

Kwastek, K. (2003) Visualising Art History, In *Convergent Practises, New Approaches to Art and visual culture, CHArt Proceedings*, 2003, CHArt, London, UK.

Malraux, A. (1947) *Le Musèe imaginaire*, Albert Skira AG., Geneva, CH.

Oomen, J. and Aroyo, L. (2011) Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges, To be published in *Proceedings of the 5th International Conference on Communities & Technologies*, Brisbane.

Risden, K., and Czerwinski, M. P. (2000) An initial examination of easue of use for 2D and 3D information visualizations of web content, In *International Journal of Human-Computer Studies*, 53, pp. 695-714.

Ruffaldi, E., Evangelista, C., Neri, V., Carrozzino, M. and Bergamasco, M. (2008) Design of Information Landscapes for Cultural Heritage Content, In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, 10-12 September 2008, Athens, pp. 113-119, ACM, New York, NY, USA

Tavanti, M. and Lind, M. (2001) 2D vs 3D, Implications on Spatial Memory, In *Proceedings of the IEEE Symposium on Information Visualization '01*, San Diego, Ca, October 22-23, 2001, pp.139-145, IEEE Computer Society, Washington, DC, USA.

Wielemaker, J., Hildebrand, M., van Ossenbruggen, J. and Schreiber, G. (2008) Thesaurus-Based Search in Large Heterogeneous Collections. In *Proceedings of the International Semantic Web Conference 2008*. LNCS 5318, pp. 695-708, Springer-Verlag, Berlin - Heidelberg.

Wiss, U. and Carr, D. A. (1999) An Empirical Study of Task Support in 3D Information Visualizations, In *Proceedings of the IEEE Symposium on Information Visualization '99*, July 14-16, 1999, London, UK, pp. 392-399, IEEE Computer Society, Washington, DC, USA.

[i] http://www.metmuseum.org/toah/
[ii] http://www.brooklynmuseum.org/
[iii] http://smarthistory.org/
[iv] http://www.googleartproject.com/
[v] http://www.w3.org/RDF/FAQ
[vi] http://www.w3.org/RDF/
[vii] http://www.cidoc-crm.org/
[viii] http://www.getty.edu/research/tools/vocabularies
[ix] http://www.vraweb.org/projects/vracore4/
[x] http://en.wikipedia.org/wiki/Wikipedia:GLAM/BM
[xi] http://www.khronos.org/registry/webgl/specs/latest/
[xii] http://secondlife.com/
[xiii] http://www.freebase.com/view/influence
[xiv] http://dbpedia.org/About
[xv] http://www.freebase.com
[xvi] http://www.wga.hu
[xvii] http://e-culture.multimedian.nl/software/ClioPatria.shtml
[xviii] http://unity3d.com/