GlamMapping Trove

Betti, A.; Castermans, T.; Speckmann, B.; van den Berg, H.; Verbeek, K.

Publication date
2016

Document Version
Final published version

Published in
VALA2016 Proceedings

Citation for published version (APA):
Betti, A., Castermans, T., Speckmann, B., van den Berg, H., & Verbeek, K. (2016). GlamMapping Trove. In *VALA2016 Proceedings: The VALA2016 18th Biennial Conference and Exhibition was held at the Melbourne Convention and Exhibition Centre, Melbourne, Australia from 9–11 February 2016* VALA. http://www.vala.org.au/vala2016-proceedings/vala2016-session-8-betti

GlamMapping Trove

Arianna Betti
Professor and Chair (Philosophy of Language)
University of Amsterdam, Department of Philosophy & Institute of Logic, Language and Computation
A.Betti@uva.nl

Thom Castermans
PhD student (Computer Science)
Eindhoven University of Technology
t.h.a.castermans@tue.nl

Bettina Speckmann
Professor (Computer Science)
Eindhoven University of Technology
b.speckmann@tue.nl

Hein van den Berg
Assistant Professor (Philosophy)
VU Amsterdam & University of Groningen
h2.vanden.berg@vu.nl

Kevin Verbeek
Assistant Professor (Computer Science)
Eindhoven University of Technology
k.a.b.verbeek@tue.nl

Abstract:
This paper presents the current state of development of GlamMap, a visualisation tool that displays library metadata on an interactive, computer-generated geographic map. The focus in the paper is on the most crucial improvement achieved in the development of the tool: GlamMapping Trove. The visualisation of Trove's sixty-million book records is possible thanks to an improved database structure, more efficient data retrieval, and more scalable visualisation algorithms. The paper analyses problems encountered in visualising massive datasets, describes remaining challenges for the tool, and presents a use case demonstrating GlamMap's ability to serve researchers in the history of ideas.

1. Introduction

The fields of information visualisation and visual analytics within computer science develop computer-supported, interactive, visual representations that allow users to extract meaning from large and heterogeneous data sets. While such visual techniques have become common practice in the sciences, they are little employed by libraries, despite similar increases in available data. Something similar is witnessed in areas of the humanities that depend crucially on library resources for their data, such as book history, philosophy, or literature studies. Take for instance a humanities projects in which *geo*visualisation is particularly important, one in which researchers wish to map places of publications of certain books in certain years to unveil certain important centres of book production. Such projects would typically employ remarkably basic computer-supported visualisations, often simply pins on a free-to-use Google map layer. Such visualisations are a far cry from state-of-the-art research solutions in visual analytics. What could be the cause of lack of awareness of such research solutions? The cause seems directly related to problematic aspects of visualisation software tools currently on offer, as existing tools present shortcomings in one or more of the following aspects: scalability, speed, openness, insightfulness, flexibility & user-friendliness.

From the point of view of libraries and large-scale library aggregators such as WorldCat or Trove, it is particularly important to employ technological solutions for visual analytics that are capable of addressing all these problems at once. Only in this manner can the changing needs of a large, varied group of users of possibly interlinked large-scale collections of (meta)data be adequately satisfied.

This paper presents the current state of development of the visualisation software GlamMap as a tool with an unrivalled combination of scalability, speed, openness (as it is non-proprietary), insightfulness, flexibility, and user-friendliness. GlamMap employs iterative development, whereby an early prototype is iteratively refined in design and functionality, based on user feedback. It is now possible to present a version with several important improvements compared to the first prototype of GlamMap presented at VALA 2014 (Betti et al. 2014).

The focus here is on the most crucial improvement recently achieved in the technical development of the software: GlamMapping Trove. The improvement is in fact twofold: the visualisation of all of Trove's sixty-million book records, and the new software design, which is better suited to the new algorithm's adaptive scaling for large-scale datasets. A new use case based on the Trove dataset as an example of a large-scale bibliographic dataset is presented, to demonstrate adaptive scaling and flexibility to serve heterogeneous users and needs. An analysis of problems encountered in acquiring large datasets to test the scalability of the software and consequent repercussions on certain groups of users is also presented. An analysis of open challenges and future perspectives for the software is provided, and the paper concludes with a brief description of the history and user evaluation of GlamMap.

2. Libraries, researchers and cutting-edge technology

Visual analytics techniques that allow the extraction of meaning from large and heterogeneous data sets have become common practice in the sciences (see e.g. Kellogg et al. 2008; Preim & Oeltze 2008; Van Mameren, Peterman & Wuite 2008). However, such techniques are little employed by libraries, despite similar increases in available data. In view of this, it is little wonder that similarly low levels of engagement with state-of-the art technologies can be witnessed in the practice of library users whose work substantially depend on accessibility and (re)usability of bibliographic (meta)data - researchers such as historians of philosophy, intellectual historians, scholars in comparative literature, and in book studies (henceforth this group of users is what is meant when talking about ´text-based humanities researchers').

The work of this group of library users is most directly affected by the extent to which libraries are able to offer quick and easy access to high-quality bibliographic data for exploration. One specific type of use case involving library metadata for this user group are geovisualisation projects, i.e. projects in which datasets including locations (e.g. locations of printers, places of intellectual production) are interpreted as offering significant information for a certain research question, when visualised on a geographical map. One such use case is highlighted in section 5 below.

Geovisualisation projects of this kind tend to employ, typically, remarkably basic computer-supported visualisations, often simply pins on a free-to-use Google map layer.[1] Such visualisations are a far cry from state-of-the-art research solutions in visual analytics. Why are basic solutions so widely adopted? What could be the cause of lack of awareness of state-of-the-art research solutions? The answer to this question is rather complex, as disparate factors play a role. Among the latter could confidently be listed a number of sociological, organisational and public policy-related factors. These factors include the sociology of collaboration among research groups in academia from different fields and their different practices, the complexity of interaction between research groups and institutions outside academia in the business and cultural heritage, otherwise known as GLAM (galleries, libraries, archives and museums) sector, trends in governmental public sector funding policy, both local and across the globe, and the nature and the managerial organisation of the GLAM sector.[2] Next to these factors, however, it is important to highlight a significant cluster of causes that seem to be more specifically related to certain problematic aspects of visualisation software tools currently available on the market, among which are some causes of a purely technological nature.

A combination of research on visualisation tools, expert user testing, peer-group evaluations and market research that has been conducted within several projects between 2010 and 2015 on the basis of interdisciplinary research meetings, internal evaluation, workshops and interviews[3] has shown the following. Libraries are interested in new opportunities to visualise information and data they have on maps. The main rationale for this is to improve the accessibility of their information for society, and to improve awareness for their information towards citizens. Likewise, humanities researchers are interested in accessing library data in a way which is fast, immediate, flexible, insightful, attractively designed and, most importantly, require low resources overall (little money, little time, and low technological skills).[4]

These desiderata prove unmatched by the most common solutions available on the market. Currently, the market shows a gap between traditional GIS professional commercial products, such as ESRI's GeoWeb/ArcGIS[5] or Pitney Bowes' MapInfo[6] and the free-to-use, lightweight, internet-based, open-to-all Google Maps, and the (slightly) more sophisticated CartoDB[7] or Palladio[8]. Existing visualisation tools present shortcomings in one or more of the following aspects: scalability, speed, openness, insightfulness, flexibility & user-friendliness. The fact that existing tools show these problematic aspects seems to relate in an interesting way to the sociological, organisational and public policy-related factors mentioned above. For instance, one participant at one of the GlamMap workshops from a national library has pointed out the following two difficulties as far as the implementation of innovation in the library world is concerned. First, GLAM users tend to be conservative (and thus innovation-resistant). For these users, the threshold to employing new technological solutions, especially solutions that are not out-of-the-box and need instead to be developed or at least adapted for a research purpose in an interdisciplinary setting, is typically too high. Second, due to funding limitation, both as regards staff and non-staff resources, when it comes to engaging in new projects geared towards research and innovation, libraries need to make choices, and at the moment priority is given to digitisation (i.e. focus on bringing full-text data in the library system).[9]

When directly asked to evaluate GlamMap, researchers judged the tool to be attractive for researchers, because it is as free as putting pins on Google Maps, but it was judged to be more flexible, insightful and beautiful. With respect to its alternatives, GlamMap is interesting for libraries because it has an unrivalled combination of scalability (especially as regards clustering of massive amounts of data), speed, flexibility, openness, user-friendliness and insightfulness. The current prototype of GlamMap described in section 4 is currently able to visualise all of the 60 million records of Trove, the biggest aggregated repository of bibliographic metadata of the Australian continent, in a fast, attractively designed, and effective way. There are grounds to conclude that the use of visualisation tools such as GlamMap in library (project) portals - especially aggregators, and consortia such as Europeana - might contribute to spreading awareness among users as to the existence of more sophisticated visualisation solutions, and might result in increased demand for more insightful, state-of-the-art visualisation technologies.

At the moment, GlamMap is being further developed within the project *Visual Analytics for the World's Library Data* (2015-2019). In this phase, the authors have chosen to develop GlamMap within strategic partnerships, and to give priority to develop fully customised solutions for software and services for specific groups of users such as researchers, libraries, and other GLAM actors. There will be regular releases of the well-researched and well-tested components of GlamMap as free software, while in the meantime developing and testing of new innovative features in a closed environment will continue. The most recent focus of development, as will become clear in section 4, has been on scalability.

3. Related Work

There are different ways to visualise library metadata. Typical search engines provide search results as lists, which are often long and hard to interpret, because they provide little context and clues to the magnitude of the result. Considerable effort has been put into systems to visualise search engine result pages. Most of these approaches rely on a hierarchy in the metadata, such as the Dewey Decimal classification or other schemes. Compact visualisation of hierarchies can, for example, be attained by treemaps, which are confined to a predefined rectangular region (Clarkson et al. 2009). Another area that has received much attention is topic visualisation over large document collections (see, e.g., Collins et al. 2009; Cui et al. 2011; Oelke et al. 2014). Visual topic analysis systems help users to explore and understand topic evolutions. Sets of independent topics can be visualised by tag clouds (Collins et al. 2009), but this technique is less suitable for tracking evolution of multiple, dependent topics. Cui et al. (2011) have proposed a river-flow-based visual metaphor to handle the latter case. The Bohemian Bookshelf (Thudt et al. 2012) is a system that is aimed at discovery in digital book collections. The system does not work with text-based queries, but provides the users with a playful environment that can be considered a digital parallel to browsing bookshelves. The idea is that this encourages and supports serendipitous discoveries.

Another, rather intuitive way to visualise library data, is to represent data geographically, by depicting, for example, the place of publication of documents on a geographic map. This approach has guided the development of GlamMap. This approach is also adopted by the Alexandria Digital Library Project (Smith 1996; Ancona et al. 2002), GeoVIBE (Cai 2002), The Digital Public Library of America (DPLA 2015), OCLC's mapFAST prototype (OCLC 2015), and Europeana's 4D interface (e4D) prototype for Europeana (Europeana 2015). As shown at VALA 2014, GlamMap differs from these geographic and information retrieval tools by providing advanced zooming techniques that help users to preserve their mental map while interactively exploring the visualisation (Betti et al. 2014). Bibliographic items published at a certain geographic location are visually represented by square clusters of differing sizes. For example, when inspecting all the items published in the area of London (England), the user is provided with a preview of the sub-clusters that will result from zooming in, such as clusters of items published in Bath, Wolverhampton and Shaftesbury. The user can subsequently explore these various sub-clusters by further zooming in. Through these advanced zooming techniques, which are not provided by other tools, GlamMap provides the user with both a general overview of items published in large geographical areas, and with the ability to interactively explore individual items published at specific geographic locations. Finally, as mentioned, there exist (commercial) tools such as CartoDB (https://cartodb.com/) and Palladio (http://hdlab.stanford.edu/projects/palladio/) that allow users to geographically visualise data on a map. GlamMap provides more sophisticated and informative visualisations than these tools, insofar as GlamMap, as opposed to CartoDB and Palladio, provides users with a higher quality clustering of data and visualises multiple attributes of data (e.g., year of publication, place of publication, number of publications).

4. GlamMapping Trove

As already demonstrated at VALA 2014 (Betti et al. 2014), the first prototype of GlamMap was perfectly able to interactively visualise the bibliographic metadata of a dataset of several thousands of bibliographic records, like Wilhelm Risse's *Bibliographica Logica* (Risse 1965), a standard reference work on the history of logic. However, the size of this dataset pales in comparison to the datasets of large-scale aggregators such as WorldCat or Trove, which consist of tens or hundreds of millions of records. Unfortunately, the first prototype of GlamMap was not designed to be scalable to such datasets. Significant changes to GlamMap's system and design were needed, even just to show the data of Trove, and further challenges remain for complete interactive exploration of such very large datasets. The current version of GlamMap supports interactive zooming and panning of the complete Trove data set (see Figures 1 and 2 or explore Trove in GlamMap using the interactive prototype at http://glammap.net/glamdev/maps/5[10]).

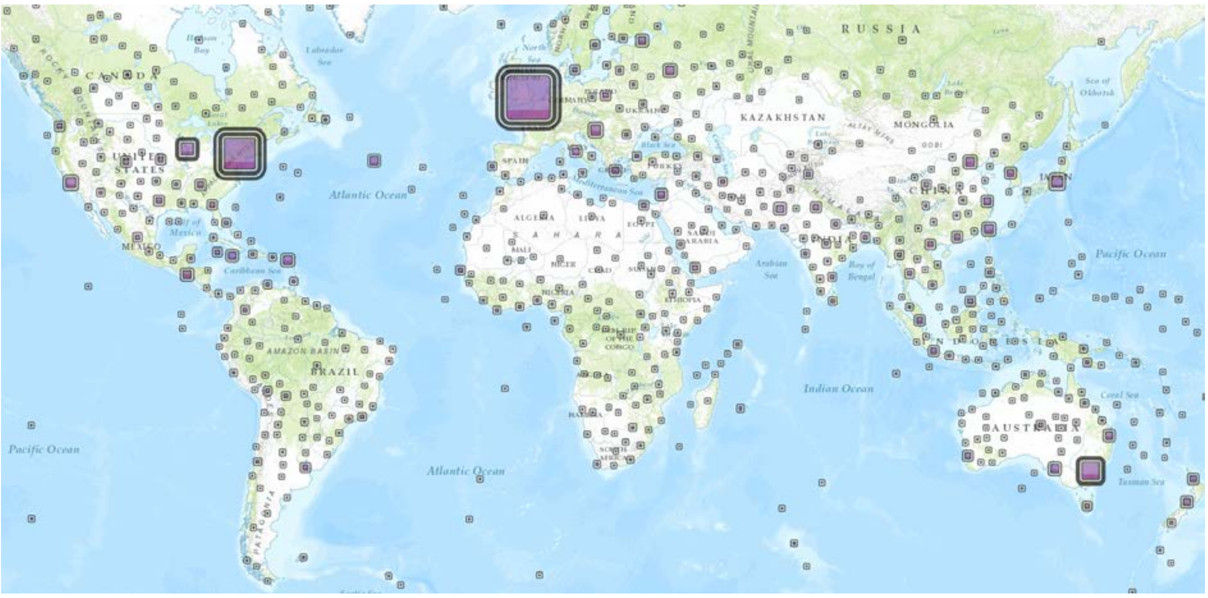

Figure 1: GlamMapping Trove.

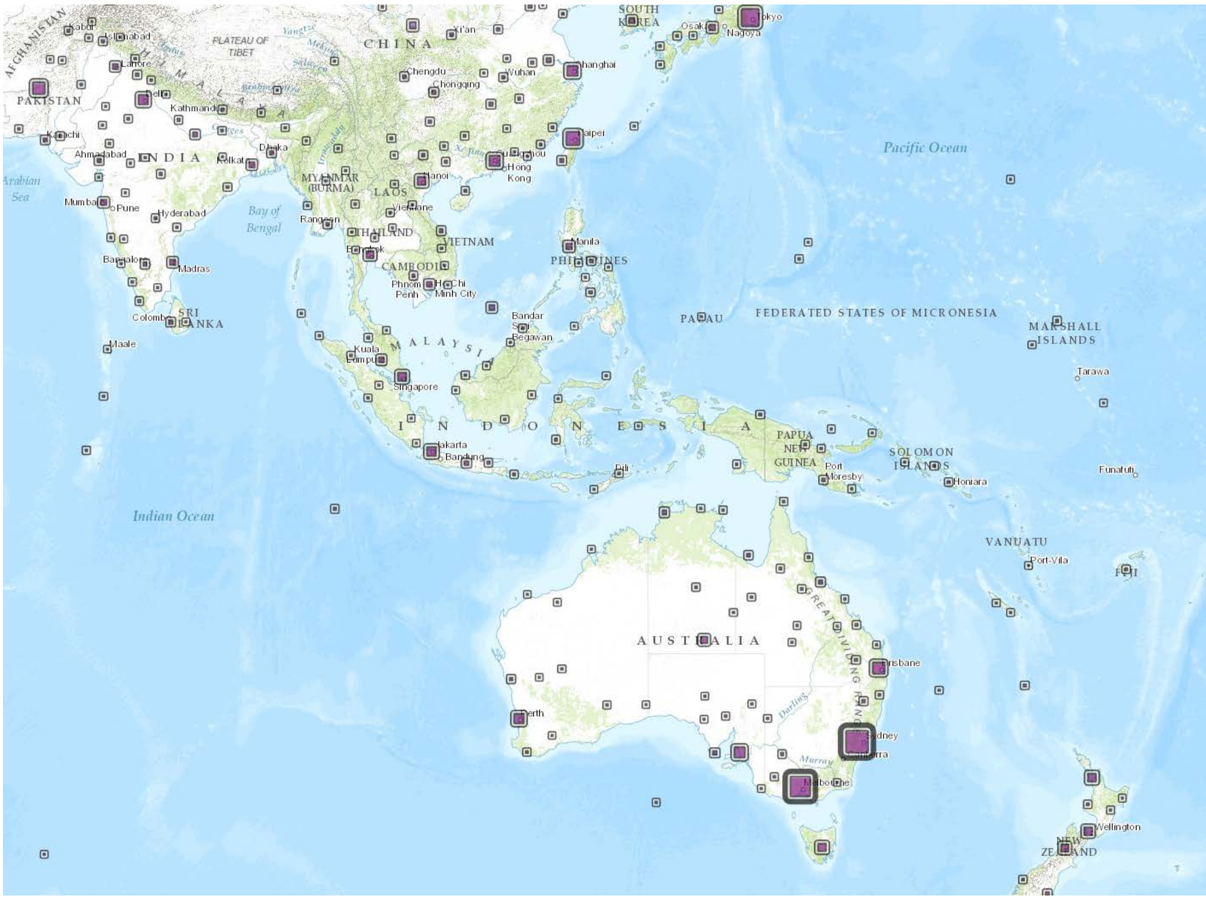

Figure 2: GlamMapping Trove (detail).

The scalability of GlamMap is influenced by two main aspects: performance and visual design. The performance aspect is clear: obtaining and visualising the data for datasets with millions of records is much slower than for datasets with thousands of records. Nonetheless, it is important to offer the user an interactive visualisation of the data; at the least, he or she needs to be able to smoothly pan and zoom to relevant locations. Furthermore, the user should clearly not have to wait for hours upon starting the application to view Trove. Offering such a responsive and interactive visualisation of Trove required a significant redesign of GlamMap's system, in particular its database.

The visual design aspect of scalability concerns the symbols shown on the map. In the first prototype, the size of a symbol corresponds directly to the number of records at that location, where the area of a symbol grows linearly with respect to the number of records. However, large datasets like Trove tend to have a much larger variance with respect to the number of records at different locations. The linear scaling thus results in a visualisation essentially consisting of one large symbol, prohibiting useful insight into the data. It was therefore necessary to employ other scaling techniques to provide meaningful visualisations of large datasets.

4.1 Improving performance

The first challenge of GlamMapping Trove was to actually import the Trove dataset into the project's database. Although this aspect is not directly relevant to the visualisation of the data, it is an important aspect for allowing users to upload their own data to GlamMap efficiently, one of the project's design goals. It is aimed to allow users to upload their own BibTeX (http://www.bibtex.org/) files, from which the system extracts records with associated locations. The location names are translated into coordinates automatically, so that the records can be displayed on the map similarly to how Trove is visualised. The import functionality is still experimental at the moment, but functional. It can be tested at http://glammap.net/glamdev/bib_jobs: users can create an account and log in to be able to upload their BibTeX files for visualisation. These files cannot be too large, though; datasets the size of Trove require special treatment to be imported efficiently.

Importing Trove presented two main problems: (1) efficiently updating the GlamMap database and (2) geocoding the locations. In the first prototype of GlamMap, new records of an imported dataset were inserted one-by-one into the database through a database interface library, and each record was properly validated so that the database structure would remain consistent. For Trove, however, importing the dataset in this way was prohibitively slow. By combining several (about 10,000) update operations into large batch operations, avoiding automatic database validation, and generally bypassing slow operations of the database interface library, it was possible to reduce the computation time by a factor of 2,000. Given that it still takes 6 hours to import Trove, this was a significant and vital improvement.

Additionally, since the Trove dataset contains only location names and not coordinates, it was necessary to translate location names into coordinates, also known as geocoding. Here it is important not to perform geocoding for each record separately, as was the case in the first prototype. Many records refer to the same location, and thus need only a single geocoding operation. Therefore, the first step is to group records with the same location name, and simply perform geocoding afterwards, only once per group. Still, geocoding is quite slow as the process is limited by the rate limit of the geocoding service being used (only 30,000 queries per day). Therefore, currently under investigation are more direct options of geocoding that are not limited in this way.

The next challenge is to visualise the dataset in the GlamMap application. In order to achieve that, it is necessary to extract the relevant data from the database. In particular, to draw the symbols on the map, it is necessary to group the records by location and further categorise them by publication year. In the first prototype, this information was computed on the fly. However, the data in GlamMap's database is organised according to the FRBR (Functional Requirements for Bibliographic Records) hierarchy by the IFLA (International Federation of Library Associations and Institutions), which distinguishes a *work*, its *expressions*, *manifestations* and *items*. Although this is a structured way to store bibliographic records, it does not allow efficient access to the data. To extract the relevant data from GlamMap’s database, it was necessary to join all levels of the hierarchy into one table, which is a costly operation. Furthermore, the grouping and categorisation of records still needed to be performed after the join, causing very slow access to the relevant data. In the current version of GlamMap, the join, grouping, and categorisation is already precomputed and stored as such in the database. This allows for much more efficient access to the relevant data. The drawback is that it is harder to add more data to the database and that the database requires more storage space. Nonetheless, these drawbacks are clearly outweighed by the benefits for large datasets such as Trove.

The final challenge with respect to performance is the clustering of symbols, which is essential to provide a clear overview of the data. The clustering algorithm in the first prototype proved to be much too slow to deal with the Trove dataset. The main reason for this poor performance was the very accurate computation of symbol sizes. In the version of GlamMap that was presented at VALA 2014, every book was shown as a separate rectangle in the symbols. The position of the rectangles in the symbol was carefully calculated, so that rectangles would not be obstructed by the rounded corners of the symbol. This operation is quite costly and for datasets the size of Trove unfortunately impossible to perform on the fly. It was, therefore, decided to trade accuracy for performance. Records are now simply represented by squares (matching the shape of the symbol) and do not necessarily need to fit within the rounded corners of the symbol. Furthermore, the system no longer draws a separate rectangle for each book, but instead draws all books from the same timeframe as a single polygon. This again improves performance, while keeping the impression the data makes intact. Although the resulting symbol size and representation is not as accurate as in the first prototype, it can be computed orders of magnitude faster, and the differences are not noticeable for a dataset the size of Trove. A side-by-side comparison between the symbols is shown below.

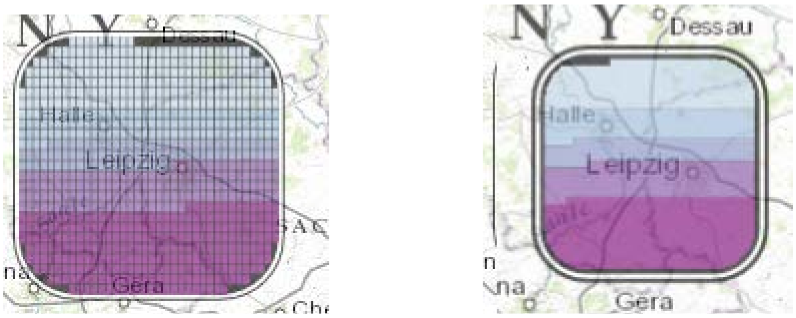

Figure 3: Old and new symbols.

4.2 Improving the visual design

As already mentioned above, symbol sizes cannot be scaled linearly when GlamMapping Trove. To avoid symbols becoming too large, several compression levels are introduced. The idea is simple: symbols using higher compression levels are displayed smaller than they would be with linear scaling. Higher compression levels correspond to smaller scale factors. The compression level of a symbol is determined by the number of records represented by the symbol, using certain specified threshold values. Symbols at the lowest compression level (level 0) use the standard linear scaling. For Trove, there are two more compression levels, one with threshold value 10^6 and scale factor 25/39 (level 1), and one with threshold value 10^7 and scale factor 4/9 (level 2). Using these compression levels, GlamMap offers a much better overview of the data.

Naturally, to avoid confusion, it must be clear at which compression level a particular symbol is. Therefore, a simple visual encoding is used to communicate the compression level of a symbol. Symbols at compression level 1 have a thicker outer border, and symbols at compression level 2 have a double thick border. This visual encoding allows a user to easily distinguish between the different compression levels.

Figure 4: From left to right, compression level 0, 1, and 2.

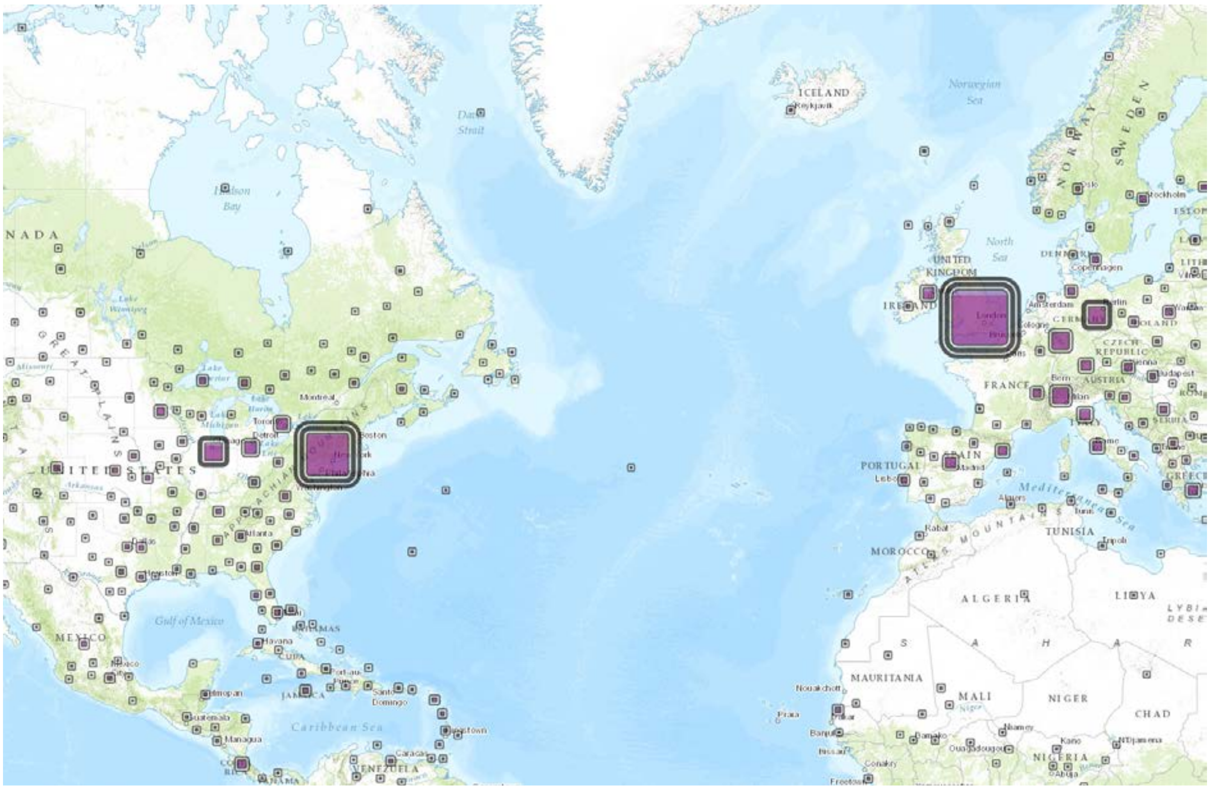

Figure 5: The symbols for Washington and London are drawn using compressions level 2. Without compression, they would fill most of the map.

4.3 Future challenges

Although the current version of GlamMap is able to visualise Trove interactively, there are still several features missing and certain aspects that can be improved. Firstly, the geocoding is still not working as well as the authors would like. The locations of Trove records are often not clearly specified in the data, making it hard for many Trove records to be geocoded (currently about 5% of the Trove records cannot be geocoded at all). Furthermore, some locations are geocoded incorrectly, often due to ambiguity or misspellings (or any other problem in the location description). This results, for example, in a significant number of records being mapped to the South Pole, which seems unlikely to be correct (see Figure 6).[11] It is hoped to reduce these problems in the future by developing improved geocoding. However, in the end, GlamMap is also limited by the quality of the data.

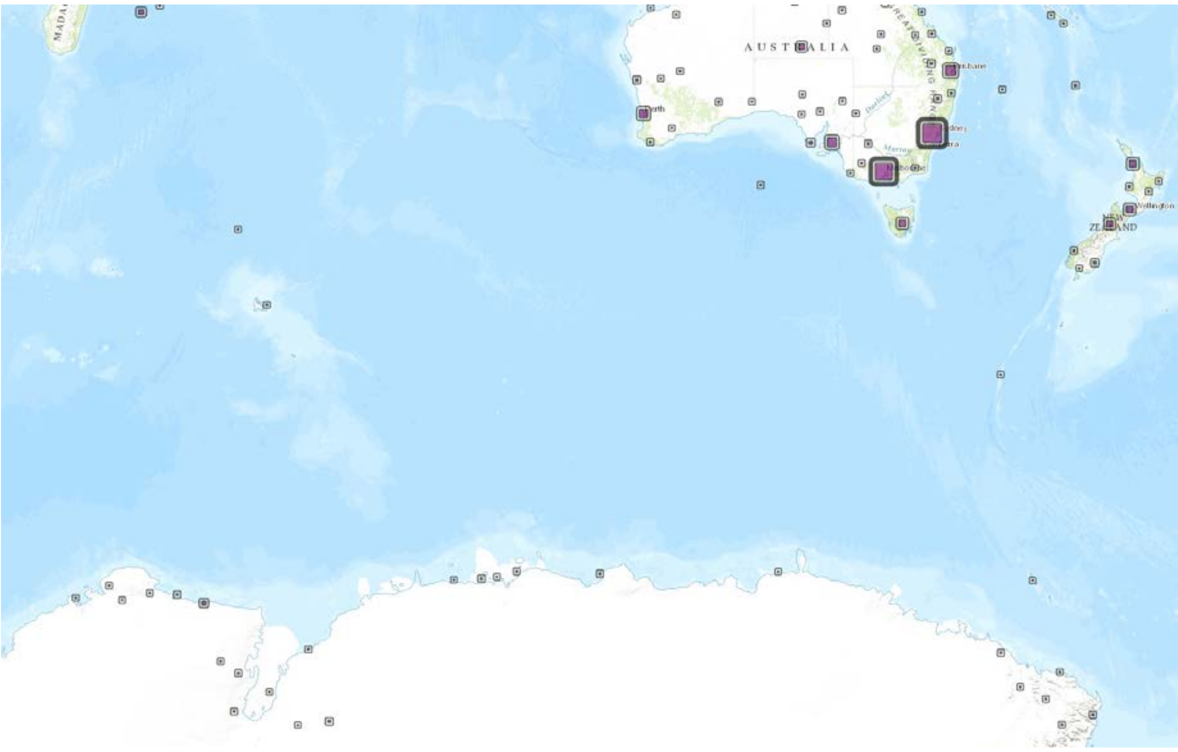

Figure 6: Records are geocoded to have been published at the South Pole.

The system cannot yet interactively filter the Trove dataset, that is, search for authors, or titles or within ranges of publication dates. Even though GlamMap has the tools to perform such data filtering, it was necessary to disable these filtering options for Trove for now as they are prohibitively slow. Even with the new optimised database structure, it takes about a minute to filter on publication year, and much longer to filter on author or title. Given the optimisations already applied to the database, it is likely that the project will require a completely different (distributed) system architecture to achieve near-interactive performance for data filtering on Trove.

Finally, the clustering algorithm that needs to be executed after extracting the data from the database also needs further improvement. This algorithm is particularly slow when the data contains thousands of different locations. The authors are currently investigating new algorithmic techniques to improve the performance of the clustering algorithm.

5. Application: History of Ideas

The authors' previous paper documented the benefits of GlamMap for researchers in the humanities, in particular researchers working within history and philosophy of science (Betti et al. 2014). This paper will demonstrate the utility of the novel visualisation of the Trove dataset for researchers working within the history of ideas.

The history of ideas is a discipline largely founded by the historian Arthur Lovejoy (1873-1962), which traces how concepts change through time. Historians of ideas often identify a concept that has been widely discussed in the history of Western thought and trace how the meaning of this concept has changed, taking into account different authors operating within different historical periods and contexts (van den Berg et al. 2014). For example, historians of ideas interested in the development of biology may describe how the concept of *evolution* has changed from antiquity up to the twentieth century, whereas historians of ideas interested in political science may describe how the concept of *democracy* has been construed in different historical periods.

Like many types of research undertaken in the humanities, research in the history of ideas is largely interpretative and strongly text-based. Historians of ideas need access to massive amounts of textual resources, such as books, notebooks, diaries, and letters, in order to describe how certain concepts are construed by authors. Thus, returning to the previous example, the historian of ideas working on the history of biology may want to identify all of the textual sources in which the biologists Alfred Russel Wallace (1823-1913) and Charles Darwin (1809-1882) discuss the concept of evolution, in order to subsequently describe similarities and differences between the ideas of these two authors. Finding textual sources can be laborious and time-consuming. This is because the quantity of textual sources studied by historians of ideas is extremely large and because many important historical sources are not known by the historian before conducting research. Finding and identifying textual sources that contain relevant information is an important part of the research conducted by the historian of ideas. It is not easy, for example, to discover all the eighteenth and nineteenth-century texts that discuss the concept of evolution, and identifying all of these texts is a major part of the work done by historians of ideas.

Visualisations of large bibliographic databases, such as Trove, are very helpful to the historian of ideas in order to identify both known and unknown historical texts. The larger the database the better: ideally, historians of ideas want to have access to *all* the (bibliographic data of) writings published in a certain historical period. Through the GlamMap visualisation of Trove, a large library database aggregator, historians of ideas gain a quick visual overview of most or all of the publications of a certain author, or, indeed, of a large number of the total publications in a certain historical period. This visual representation of large quantities of bibliographic data is easier to inspect than list-based representations of these data. Thus, for example, once GlamMap allows for searching for the works of Alfred Russel Wallace, it will provide researchers with a quick visual overview of a large number of works published by Wallace. By inspecting the visualisation, historians of ideas can quickly identify both known and unknown works, and can quickly identify translations of these works. In short: the GlamMap visualisation of Trove provides historians of ideas with a tool that allows them to easily inspect and identify both known and novel research data.

It is further possible to illustrate the manner in which the GlamMap visualisation of Trove provides historians of ideas with access to novel research data by giving a concrete example of research conducted in the history of *philosophy*. Suppose a historian wishes to investigate the impact of the writings of the famous eighteenth-century philosopher Immanuel Kant (1724-1804). The historian of ideas will typically want to describe the intellectual context within which Kant operated, and will then want to identify successors of Kant that somehow utilised his ideas. GlamMap provides an interesting way to accomplish these tasks. The historian can use geographical information to reconstruct the intellectual *milieu* of Kant's writings and identify successors of Kant who utilised his philosophical ideas. Throughout his life, Kant worked in the city of Königsberg, Prussia (now the city of Kaliningrad, Russia). By clicking on Königsberg, the GlamMap visualisation of Trove provides a quick overview of all the writings published in Königsberg (partially depicted below).

	und Jurisprudenz (Trove API)		
Hoffmann, Johann Gottfried, 1765-1847.	Das Interesse des Menschen und Bürgers bei den bestehenden Zunftverfassunge (Trove API)	koenigsberg	1803
Krug, Wilhelm Traugott, 1770-1842.	System der theoretischen Philosophie (Trove API)	koenigsberg	1825
Johann Heinrich Westphal 1794-1831	Logarithmische Tafeln von Westphal (Trove API)	koenigsberg	1821
Krug, Wilhelm Traugott, 1770-1842.	System der theoretischen Philosophie (Trove API)	koenigsberg	1825
Herbart, Johann Friedrich, 1776-1841.	Allgemeine Metaphysik, nebst den Anfängen der philosophischen Naturlehre von Johann Friedrich Herbart (Trove API)	koenigsberg	1828
Herbart, Johann Friedrich, 1776-	Allgemeine Metaphysik, nebst	koenigsberg	1828

Figure 7: subset of works published in Königsberg.

Among the publications listed are books written by Wilhelm Traugott Krug (1770-1842) and Johann Friedrich Herbert (1776-1841). Krug and Herbert are little known philosophers who both succeeded Kant in the chair of logic and metaphysics at the University of Königsberg, and who developed philosophical ideas in reaction to Kant. By allowing for an intuitive and easy search by geographic location, the GlamMap visualisation of Trove thus allows historians of ideas to identify research data that enable the study of the way in which Kant's philosophy was shaped and transformed by his successors at the university of Königsberg, providing context and insight into the immediate reception of Kant's philosophy

6. Conclusion

This paper has described the current state of development of GlamMap, a geovisualisation tool that displays library metadata on an interactive map. At VALA 2014, the authors presented a prototype of GlamMap that allowed for the interactive visualisation of over seven thousand bibliographic records. GlamMap has been improved to enable the visualisation of the approximately sixty million records of Trove, the biggest repository of bibliographic metadata of Australia. The visualisation of Trove's sixty million records required the development of an improved database structure, more efficient data retrieval methods, and more scalable visualisation algorithms. The paper has provided a use case that demonstrates that the visualisation of Trove allows researchers in the humanities, more specifically historians of ideas, to quickly search, navigate and explore the dataset, and to identify novel research data. In the future, it is planned to improve the geocoding of Trove's records, and to enable users to interactively filter the Trove dataset by enabling them to search for authors, titles, and for publications within ranges of publication dates.

Acknowledgements

We thank Jasper Brekelmans for help with writing this paper. A. Betti, T. Castermans, B. Speckmann & K. Verbeek are supported by the project *Visual Analytics for the World's Library Data*, funded by the Netherlands Organisation for Scientific Research (NWO). H. van den Berg is supported by Europeana eCloud, #325091.

References

Ancona, D, Freeston, M, Smith, T & Fabrikant, S 2002, 'Visual explorations for the Alexandria Digital Earth Prototype', in K Börner & C Chen (eds.), *Visual Interfaces to Digital Libraries,* Springer-Verlag, Berlin & Heidelberg, pp. 199– 213.

Betti, A, Gerrits, D.H.P, Speckmann, B & van den Berg, H 2014, 'GlamMap: Visualizing Library Metadata', *Proceedings of VALA 2014 - Libraries, Technologies, and the Future*. http://bit.ly/1oWsPcj

Betti, A & van den Berg, H 2014, '@PhilosTEI - Final user evaluation report' (November 2014) http://axiom.humanities.uva.nl/philosteireport.pdf

Bulger, M, Meyer, E.T, de la Flor, G, Terras, M, Wyatt, S, Jirotka, M, Eccles, K & Madsen, C 2011, *Reinventing research? Information practices in the humanities - A Research Information Network Report*, viewed 27 September 2015, http://www.rin.ac.uk/system/files/attachments/Humanities_Case_Studies_for_screen_2_0.pdf

Cai, G 2002, 'GeoVIBE: A visual interface for geographic digital libraries', in K Börner & C Chen (eds.), *Visual Interfaces to digital libraries*, Springer-Verlag, Berlin & Heidelberg, pp. 171–187.

Clarkson, E.C, Desai, K & Foley, J.D 2009, `ResultMaps: Visualization for search interfaces', *IEEE Trans. Visualization and Graphics*, vol. 15, no. 6, pp. 1057-1064.

Collins, C, Viegas, F & Wattenberg, M 2009, `Parallel tag clouds to explore and analyze faceted text corpora', *Proceedings of the IEEE Symposium Visual Analytics Science and Technology* (VAST), pp. 91–98.

Cui, W, Lui, S, Tan, L, Shi, C, Song, Y, Gao, Z.J, Tong, X & Qu, H 2011, 'TextFlow: towards better understanding of evolving topics in text', *IEEE Trans. Visualization and Graphics*, vol. 17, no. 12, pp. 2412-2421.

Europeana Foundation 2015, 'Europeana 4D', viewed 28 September 2015, http://labs.europeana.eu/apps/europeana4d.

DPLA 2015, 'DPLA by County and State', viewed 28 September 2015, http://dp.la/apps/14.

Kellogg, L.H, Bawden, G.W, Bernardin, T, Billen, M, Cowgill, E, Hamann, B, Jadamec, M, Kreylos, O, Staadt, O & Sumner, D 2008, 'Interactive visualization to advance earthquake simulation', in K.F Tiampo, D.K Weatherley & S.A Weinstein (eds.), *Earthquakes: Simulations, Sources and Tsunamis*, Birkhäuser Verlag, Basel, pp. 621–633.

OCLC 2015, 'mapFast', viewed 28 September 2015, http://www.oclc.org/research/themes/data-science/mapfast.html.

Oelke, D, Strobelt, H, Rohrdantz, D, Gurevych, I & Deussen, O 2014, 'Comparative Exploration of Document Collections: a Visual Analytics Approach', *Computer Graphics Forum*, vol. 33, no. 3, pp. 201-210.

Preim, B & Oeltze, S 2008, '3D visualization of vasculature: an overview', in Linsen, H Hagen & B Hamann (eds.), *Visualization in medicine and life sciences*, Springer-Verlag, Berlin & Heidelberg, pp. 39–59.

Risse, W 1965, *Bibliographia logica*, Georg Olms, Hildesheim.

Smith, T.R 1996, 'A digital library for geographically referenced materials', *Computer*, vol. 29, no. 5, pp. 54–60.

Thudt, A, Hinrichs, U & Carpendale, S 2012, 'The Bohemian Bookshelf: supporting serendipitous book discoveries through information visualization', *Proceedings SIGCHI Conference on Human Factors in Computing Systems* (CHI' 12), pp. 1461-1470

Van den Berg, H, Parra, G, Jentzsch, A, Drakos, A & Duval, E 2014. 'Studying the History of Philosophical Ideas: support tools for research discovery, navigation, and awareness'. *Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business* (*I-know 2014*), Article no. 12. DOI: 10.1145/2637748.2638412.

Van Mameren, J, Peterman, E.J.G & Wuite, G.J.L 2008, 'See me, feel me: methods to concurrently visualize and manipulate single DNA molecules and associated proteins', *Nucleic Acids Research*, vol. 36, no. 13, pp. 4381–4389.

Appendix I - History and Evaluation of GlamMap

GlamMap has been developed within six projects starting from 2010: four grants from the The Young Academy (De Jonge Akademie) of the Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen) (2010, *Mapping Scientific Philosophy in Europe*; 2011, *Mapping Philosophy 2;* 2012, *GlamMap glammed up;* 2013, *SalVing GlamMap* 2013), one ERC (European Research Council) Proof of Concept of the ERC Starting Grant Tarski's Revolution: A New History and the current project, Visual Analytics for the World's Library Data, funded by the Netherlands Organisation for Scientific Research (NWO) in partnership with OCLC.

The six projects have involved eight main researchers, including a core group of expert users in the humanities from the group of Arianna Betti (University of Amsterdam, Concepts in Motion - Axiom Group) and researchers in computer science from the Applied Geometric Algorithms group of Bettina Speckmann (Technical University of Eindhoven) working in close cooperation.

In the first phase of development (2010-2013), evaluation of GlamMap has been done by the core group of expert users from the field of philosophy. In the second phase (2013-15) the evaluation has been conducted in two additional ways: (1) via presentations in interdisciplinary (inter)national meetings involving a global spectrum of researchers in the sciences, arts and humanities, (2) via market research (based on desk research, email exchanges, phone interviews and workshops) with a diverse cohort of linguists, archaeologists, scientometrists,[12] book studies scholars, experts in innovation studies, experts from the GLAM sector (from National libraries, digital cultural heritage organisations, museums, and multi-library projects), officers from governmental organisations such as patent offices, mostly from the Netherlands. The workshops took place on June 27th, and August 29th, 2014.

Endnotes

[1] Alternative tools employed are CartoDB, or as this blogger puts it referring to ´Google Maps on steroids´. http://petercarrjones.com/projects/mapping-protest-voices-of-the-santa-barbara-oil-spill/ or Palladio (see e.g. this collection of student projects in history & new media, http://rrchnm.org/robertson/hist696f14/assignments/student-projects/)

[2] An anonymous referee has suggested that other factors might be (1) the existence of products on the market without justifiable utility for libraries and others; (2) lack of personnel in institutions to implement and utilise new and existing resources; (3) local, homegrown, or other resources already in use to meet specific objectives. The authors agree that these factors are likely play a role. In particular, as to (1), the users within the GlamMap projects have signalled discrepancies between the rationale of tools offered through the web portal of libraries and interface design, and their own needs while searching for information. For instance, FRBR-isation of bibliographic data turns out to be of much higher priority for researchers (including students), than tools such as Catalogue Plus or EBSCO Discovery Services.

[3] For a report on evaluations, see Appendix I.

[4] For more on the practices of humanities researchers in relation to new technologies, see Burger et al. 2011.

[5] http://www.esri.com/news/arcuser/0206/geoweb.html

[6] http://www.pitneybowes.com/us/location-intelligence/geographic-information-systems.html

[7] https://cartodb.com/

[8] http://hdlab.stanford.edu/projects/palladio/

[9] From a humanities perspective, the focus on digitisation is justified both in view of the first point (conservatism), and given the priority humanities researchers themselves indicate right now, i.e. the building of digital corpora. Building a corpus of electronic sources is felt to be an absolute presupposition for text-based humanities researchers to engage in new computational methodologies, cf. also again the RIN 2011 report quoted in note 3 above, and Betti & van den Berg 2014. The traits of the typical way of working of a specific group of researchers in philosophical history of ideas are described in van den Berg et al 2014.

[10] Please note that currently GlamMap is fully tested and operational only in the Chrome browser.

[11] Since we currently employ only one reliable geocoding source, there is at present no method of detecting whether records are wrongly geocoded to 'unlikely' locations e.g. Antarctica, but we can exclude mistakes in the geocoding of books to locations such as London or New York. Most books for which we don't know the location would end up at coordinates (0, 0), which is located somewhere in the ocean (https://www.google.nl/maps/@0,0,7z?hl=nl).

[12] Scientometrics can be broadly defined as the study of science, technology, and innovation from a quantitative perspective.