Understanding the Use of *The Vistorian*:
Complementing Logs with Context Mini-Questionnaires

Vanessa Serrano Molinero*
IQS School of Management, Spain

Benjamin Bach†
Univ. of Edinburgh, UK

Catherine Plaisant‡
Univ. of Maryland, USA

Nicole Dufournaud§
EHESS, France
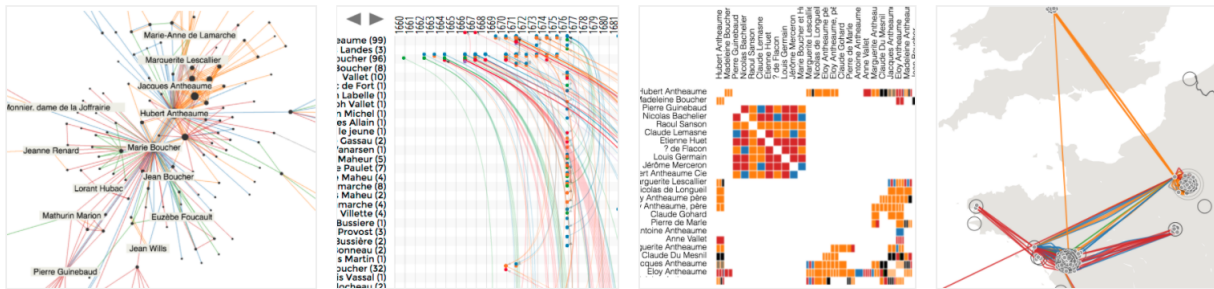
Jean-Daniel Fekete¶
Inria, France

Figure 1: The four visualizations provided by the Vistorian: node-link, TimeArcs, matrix, and map with network. All these visualizations provide pan and zoom, filtering, navigation over time, as well as specific interactions.

ABSTRACT

The Vistorian is a web-based visual analytics tool including four different interactive visualizations. It allows digital humanists to analyze complex geolocated and temporal networks of individuals. A research prototype is now available to researchers. The challenge we try to address is: could we improve our understanding of how digital humanities research prototypes are being used *"in the wild"*? Standard usage logs are insufficient since they do not capture users' intent or the reasons why they might struggle with a prototype. Here, we designed a novel lightweight combination of usage logs and mini-questionnaires attempting to consistently capture user intent and usage context. The paper first describes the Vistorian, then introduces our combined log and questionnaire methodology—with design principles and screen mockups. The technique will be pilot tested this summer, and deployed in the fall for evaluation with historians and their students.

Keywords: Network visualization, logs, understanding users, history, digital humanities.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces; D.2.5 [Software Engineering]: Testing and Debugging—Tracing.

1 GOAL OF THE PROJECT

The number of interactive visualization tools for the digital humanities community continues to grow, as well as the community of researchers studying their use. While commercial and open source tools such as Tableau [19] or NodeXL [13] provide access to traditional visualizations such as scatterplots, barcharts, treemaps, or

*e-mail: vanessa.serrano@iqs.edu
†e-mail: benj.bach@gmail.com
‡e-mail: plaisant@cs.umd.edu
§e-mail: Nicole.Dufournaud@gmail.com
¶e-mail: Jean-Daniel.Fekete@inria.fr

node-link diagrams, researchers typically propose new visualizations that include novel visual representations, innovative interaction, and insightful analytics that are not familiar to humanists — and most users in general.

Our team has been successful at making research tools available to the public, such as Treemaps [17], EventFlow [15], the InfoVis Toolkit [11], Matrix Cubes [3], Small MultiPiles [1] and Time-Curves [4] but we find it difficult to understand 1) how our tools are being used "in the wild" and eventually 2) how to help users transition to the newer techniques and features of our tools.

Several natural approaches come to mind [14] such as surveys, personal interviews or extensive automatic logging. Surveys may be useful but tend to collect subjective information born from multiple sessions and weeks of experience with a tool. Direct observations or interviews of users is not practical because it is too time consuming for typical dissemination effort with limited financial support. Finally, automated logging of all user activities may provide heaps of data but has well documented limitations [12]. This challenge was discussed extensively during the recent IEEE VIS workshop on "Logging Interactive Visualizations & Visualizing Interaction Logs"[1]: automatic logging is fundamentally useless without some knowledge of user intent and context at the time of use.

Intent might correspond to the analytics task users are attempting to complete (e.g. distant reading, trying to answer a particular question, contrasting, looking for specific details) or to the context of use (e.g. demoing to a colleague, preparing figures for a paper, or actual analysis). Capturing intent and context requires a combination of analysis of low level log data — which is error prone—, and data entry by the user — which may be annoying to them. Can a compromise be found? The next questions become: How much intent and context information is necessary (type, level of details, timing) for the automatic logging to become useful? Could capturing intent enhance the analysis process or at least not hurt it? And thinking further in the future: Can improved understanding of usage data facilitate the introduction new features or tools to users?

In this paper we first describe the Vistorian, then present a new approach for complementing usage logs with automated context questionnaires. We describe our guiding design principles and mock-
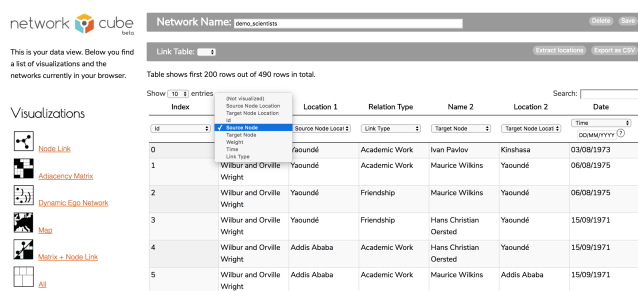
---

[1] ttp://livvil.github.io/workshop/

Figure 2: Data table (linktable) in the Vistorian, with table headers assigning network semantics to columns.

Figure 3: Multiple views in the Vistorian (from left to right): legend, node-link, matrix.

ups of the interface. The technique will be pilot tested and iteratively refined this summer, and deployed in the fall. Our hope is to show that this lightweight technique is well adapted to monitoring and understanding the use of new tools in the historian community, and the digital humanities community in general.

2 THE VISTORIAN

The Vistorian (http://vistorian.net) allows professional historians and students to analyze complex geolocated and multivariate networks of individuals varying over time [2]. Fig. 1 shows the four current visualizations with a social network: node-link diagram, time arcs [8], adjacency matrix, and a map with the links overlaid. The latter allows visualizing networks where entities are geolocated e.g. networks made of correspondence letters between actors moving in space, such as the Republic of Letters [7]. Link colors could indicate a type of letter, or more generally relation types (e.g. family, work, finance).

Network analysis and visualization have recently become popular among historians. Other visualizations than node-link diagrams are not yet part of their standard method, and visualizations are typically used to illustrate findings rather than to explore data. The Vistorian aims to provide a set of network visualization techniques standard in information visualization but not available in present tools such as Gephi [5].

To match its particular audience—domain scientists with little or no experience in computing—the Vistorian is based on four design principles: **simplicity**, **privacy**, **openness** and **extensibility**: interface and data handling has been kept as simple as possible with no requirement to write code or perform complex specification operations. Though the Vistorian is an online platform, it functions without server in the back-end, except for the file server hosting the pages. Instead, user's data is stored in the persistent local storage of the browser, remaining private on the user's machine. Only when clearing the browser cache, data gets deleted. Finally, the Vistorian is open source (https://github.com/networkcube/networkcube) and integrated into the networkcube project [2]. The underlying architecture allows fast prototyping and integration of novel visualizations in D3 and WebGL which are then available to everyone using the the Vistorian online.

2.1 Using the Vistorian

To use the Vistorian, historians provide one or two tables: a *node-table* containing information about the network nodes (e.g., name, location, type) and/or a link-table (source, target, type, timestamp, etc.). Tables are uploaded as csv files and manually assigned network semantics (e.g. source node, target node, link weight, Fig. 2). If the data contains geographic locations—names of cities, places, and countries—the Vistorian searches a geonames service (currently https://nominatim.openstreetmap.org) and retrieves the places' coordinates according to a best-match strategy.
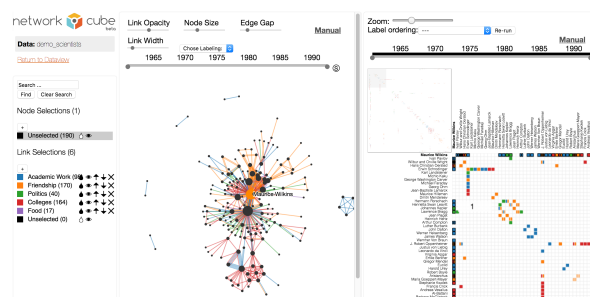
Users can also correct geographical coordinates by specifying longitude/latitude directly in the data table, or by helping the geoname service by regularizing place names (e.g. Brest, *France*).

After assigning semantics to columns, a menu list (Fig. 2, left) shows the individual visualizations and combined view configurations available. Clicking on one of the icons opens a new tab in the browser with that visualization plus a legend. For example, Fig. 3 shows a split view with a node-link and a matrix visualization for the same network. Hovering over the elements in one visualization highlights elements in the respective other view; moving the time-slider (the gray bar above the visualizations) filter the time. It can be synchronized across the views if desired. Elements can be filtered, and different colors can be applied from the legend. Interactions related to these filtering and coloring are automatically propagated to all Vistorian views currently open in the browser, i.e. across all tabs and windows.

As each view in the Vistorian runs in its own tab or window users can open the visualizations they require for a specific task, eventually making use of multiple screens and large displays, while maintaining full brushing-and-linking functionality.

2.2 Usage Scenario

A typical usage scenario with the Vistorian involves the following steps: loading data (linktable, nodetable or both), assigning a semantic to appropriate columns, and opening visualizations. Then, exploration can be conducted from one or several visualizations. For example, Fig. 4 shows the non-standard "Time Arcs" visualization [8] used to explore the social network of a merchant in France in the 17th century [10], extracted from legal acts preserved by French archives. According to standard prejudice, women were not active in commerce at that time, or maybe trading locally and at a short scale. The visualization reveals exactly the contrary: the network changed substantially year after year, according to the kind of trade Marie Boucher conducts. The registers where these trades are recorded mention boats full of wheat traded with the Netherlands, and arming and sending ships to the Caribbeans. The TimeArcs visualization is more appropriate than a node-link diagram filtered by time (as shown in Fig. 5) since it reveals the whole activity evolution in one chart, up to the decreasing activity towards the end of her life.

This example illustrates the value of of the Vistorian to assess the evolution of the activity of a person. However, it has required many trials and errors for our historian collaborator to learn the Vistorian and its visualizations, and understand which visualization is appropriate for which question. We want to understand the main challenges historians face when learning the Vistorian and in what terms they express their questions: We are hopeful that logs and context mini-questionnaire will provide us a better level of understanding.
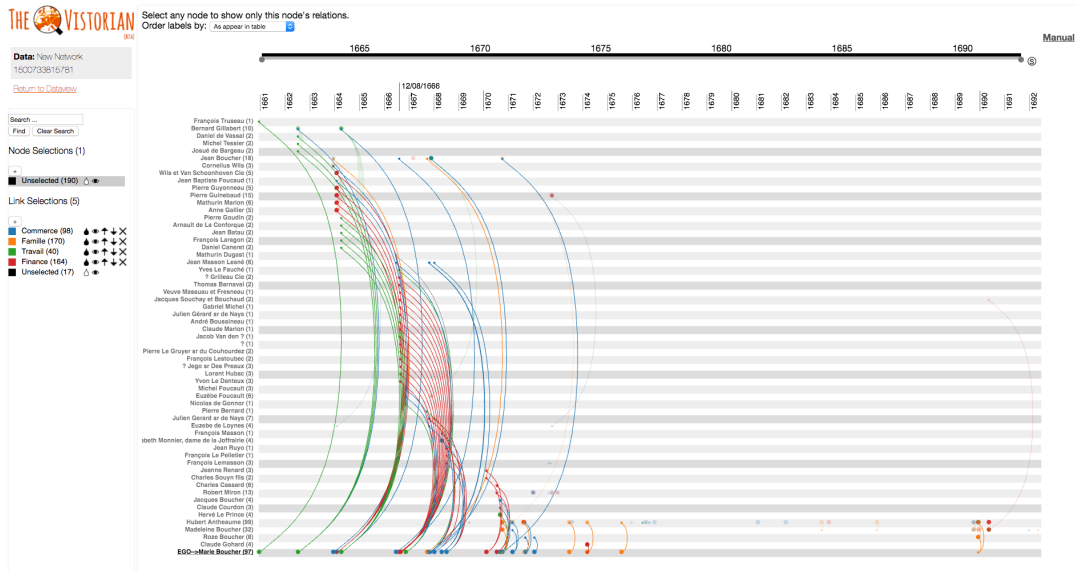
Figure 4: The evolution of the social network of Marie Boucher (bottom line), a merchant in the 17th Century in France, visualized using the Time Arcs technique [8]. The links are colored according to the type of relations as shown in the legend and filtering boxes on the left pane.

Figure 5: Exploring temporal periods of the network by defining ranges on the time-slider.

3 COMPLEMENTING LOGS WITH CONTEXT MINI-QUESTIONNAIRES

We will now describe how we intent to understand how the Vistorian is used "in the wild" using a combination of logs and mini-questionnaires triggered by the use of the tool.

The design of the usage logs is fairly standard in information visualization, as we collect events generated by user activities. We used a two pronged approach. Some members of our team listed what could be collected easily (e.g. buttons pressed and options chosen), while others listed high level concepts we wanted to be able to capture (e.g. detecting different work sessions from the same user) and questions we had (e.g. are users using the matrix for small datasets too? Do they make use of the time-zoom to animate over time at all? Rarely? A lot?). Additional events were captured or generated from low-level ones to match our needs. We designed a hierarchy of events to organize low level events so we could later export the data into an event analytic tool such as EventFlow [15]. Some of the events captured correspond to when mini-questionnaires are answered, or when they remain unanswered after a certain delay.

3.1 Logging

Our logging infrastructure relies on two components: a trace library and a web server recording traces and serving accesses to the traces. We have been using this infrastructure for several years and it is available open source at https://github.com/INRIA/intertrace. The trace library most main function is: traceEvent(category, action, label, value).

The parameters are transmitted to the server to be recorded with two additional information: a session identifier and a time-stamp

(absolute time with millisecond accuracy). The *category* classifies the event purpose (e.g. sent by the tracer itself, or from the browser's behavior, a widget, a lower-level interaction, etc.), the *action* is the name of the function traced (e.g. color-node), the *label* is optional and can provide details about the action (e.g. the identifier of the node), and the value can provide additional information (e.g. the new color assigned to the node).

In the case of the Vistorian, the trace is only sent when users have given their permission to be logged, and the session identifier is made of a unique identifier (UUID) concatenated with the identifier of the visualized network. Therefore, we can monitor the activity of users working on specific datasets. The session identifier is created once and saved as a cookie to identify sessions over multiple connections over days or months. We explicitly ask users who agree to participate for their session identifiers and keep them safely separately from the trace database.

3.2 Mini-Questionnaires

To capture user intent and usage context we designed mini-questionnaires used the following principles:

Provide *incentive* to answer. This may be the most important principle: there should be some type of direct benefit to the user for providing intent and context to the research team. In the case of the Vistorian we linked the mini questionnaires with a new feature that emails users a screenshot of the application. In the future, as an even higher incentive, we plan to include state information so users could go back to the corresponding screen.

Use *safe* and flexible recording method. From prior experience with the release of digital humanities research prototypes in the wild [6, 9, 16] we have learned that 1) users only use prototypes or save comments if they are entirely confident that their work will not be lost, and 2) they are also worried about saving research materials on unknown servers or in the hands of untrusted individuals. For that reason we decided to use email as a communication mechanism. It is familiar, its mechanism and limitations are understood, and all parties receiving the information are clearly identified (and customizable).

Figure 6: Mockup of the Mini-Questionnaire to capture context information, for use in the Vistorian. The mini-questionnaire is tagged to the "mail me a screenshot" function, providing incentive. A single question is added with a small number of options, and in most cases it can be answered in a single action (one double click). If the answer is [Analyze data] then part 2 (Fig. 7) is displayed.

Use *short* questionnaires. Digital Humanities users (like most users) are busy doing their work, do not want to be bothered by lengthy questionnaires, and do not want to be interrupted often. Therefore the proposed questionnaires are very shorts (about 10 buttons maximum for questions and options seemed a good 1st number.)

Provide one-click *shortcuts*. Even if the questionnaires are short, shortcuts reduce user burden.

Allow users to skip the questionnaire. Users should be allowed to skip when too busy.

Complement the event log but remain independent While the information from the questionnaires relates to the detailed activity recorded in the logs it should remain usable on its own. Conversely, the log should not rely on users answering a questionnaire.

Two new buttons were added to each screen of the Vistorian interface: [Mail me a screenshot] and [Report end of activity]. The basic mini-questionnaire is displayed when users ask for a screenshot (Fig. 6). The single question of the basic questionnaire is designed to allow us to separate the valuable log data (i.e. the period of analysis) from all the many other types of use that occur with research software (i.e. testing, bug reporting, demos, training). When users actually answer "analyze data" a second mini-questionnaire is provided (Fig. 7). In this early phase of the research we will rely on open ended comments to gather information on analysis activities. This will later be replaced by more specific questions based on coded open ended comments.

A simpler variant of the questionnaire (without the screenshot options) is displayed when users click on the [Report end of activity] or after a period of inactivity has been detected.

Following a Multidimensional In-depth Long-term Case Studies methodology [18] we ask users of the Vistorian who agree to participate in our case study to be first interviewed for us to understand

Figure 7: Mockup of the second part of the Mini-Questionnaire.

their goals. They are asked to sign an approved user consent form explaining that logs, screenshots and their comments will only be sent to us (but that they can change their mind later on and stop participating); that all the log data remains anonymous and we will treat their email comments or screenshots as confidential unless they give express consent for specific material (to be used later on, e.g. in a publication). We will also inform them that they will be given the choice to be named/thanked explicitly in our publications—or not if they want to remain anonymous.

4 DISCUSSION AND CONCLUSION

We hope that this novel hybrid method with logs, mini mini-questionnaires and screenshots will allow us to better understand the questions historians are trying to address, the difficulties they face; and to determine the best ways to explain new visualizations to them appropriately for different contexts and data. Logging alone cannot provide us this information and interviews cannot realistically be conducted with enough users.

In turn we hope that it will help facilitate the adoption of advanced visualization tools by historians who have complex questions and data but are not trained to manage this data and use visualizations as exploration tools.

In this paper we described the Vistorian, a tool developed for historians to analyze complex network data, and a lightweight approach to complement automatic usage logging with min-questionnaires. While this work with historians is in its early stages we believe it may also useful for the evaluation of other digital humanities projects, and potentially any visual analytics projects. We hope this method may in turn increase the likelihood of successful dissemination of novel visualization techniques.

REFERENCES

[1] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. In *Computer Graphics Forum*, vol. 34, pp. 31–40. Wiley Online Library, 2015. doi: 10.1111/cgf.12615

[2] B. Bach, N. Henry Riche, R. Fernandez, E. Giannisakis, B. Lee, and J.-D. Fekete. NetworkCube: Bringing Dynamic Network Visualizations to Domain Scientists. Posters of the Conference on Information Visualization (InfoVis), Oct. 2015. Poster.

[3] B. Bach, E. Pietriga, and J.-D. Fekete. Visualizing Dynamic Networks with Matrix Cubes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 877–886. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288.2557010

[4] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE transactions on visualization and computer graphics*, 22(1):559–568, 2016. doi: 10.1109/TVCG.2015.2467851

[5] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[6] N. Boukhelifa, M. Bryant, N. Bulatović, I. Čukić, J.-D. Fekete, M. Knežević, J. Lehmann, D. Stuart, and C. Thiel. The CENDARI Infrastructure. *J. Comput. Cult. Herit.*, 2017. to appear.

[7] G. Ceserani. Interactive Visualizations for British Architects on the Grand Tour in eighteenth-century Italy. Stanford Digital Repository, 2015.

[8] T. N. Dang, N. Pendar, and A. G. Forbes. TimeArcs: Visualizing Fluctuations in Dynamic Networks. *Computer Graphics Forum*, 2016. doi: 10.1111/cgf.12882

[9] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant. Discovering interesting usage patterns in text collections: Integrating text mining with visualization. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, pp. 213–222. ACM, New York, NY, USA, 2007. doi: 10.1145/1321440.1321473

[10] N. Dufournaud, B. Michon, B. Bach, and P. Cristofoli. L'analyse des réseaux, une aide à penser: réflexions sur les stratégies économique et sociale de Marie Boucher, marchande à Nantes au XVIIe siècle. In D. Picco, ed., *Femmes et réseaux dans les sociétés modernes et contemporaines. Réalités et représentations*. PU Bordeaux, Bordeaux, to appear.

[11] J.-D. Fekete. The InfoVis Toolkit. In I. Press, ed., *Proceedings of the IEEE Symposium on Information Visualization*, pp. 167–174. IEEE Computer Society, Austin, TX, United States, Oct. 2004. doi: 10.1109/INFOVIS.2004.64

[12] C. Grimes, D. Tang, and D. Russell. Query logs alone are not enough. In *WWW 2007 Workshop on Query Log Analysis: Social and Technological Changes*, 2007.

[13] D. Hansen, B. Shneiderman, and M. A. Smith. *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann, 2010.

[14] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012. doi: 10.1109/TVCG.2011.279

[15] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, Dec. 2013. doi: 10.1109/TVCG.2013.200

[16] C. Plaisant, J. Rose, B. Yu, L. Auvil, M. G. Kirschenbaum, M. N. Smith, T. Clement, and G. Lord. Exploring erotics in emily dickinson's correspondence with text mining and visual interfaces. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '06, pp. 141–150. ACM, New York, NY, USA, 2006. doi: 10.1145/1141753.1141781

[17] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11(1):92–99, Jan. 1992. doi: 10.1145/102377.115768

[18] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pp. 1–7. ACM, New York, NY, USA, 2006. doi: 10.1145/1168149.1168158

[19] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional databases. *Commun. ACM*, 51(11):75–84, Nov. 2008. doi: 10.1145/1400214.1400234