

Progressive Visual Decision Making for Digital Humanities (PROVIDEDH): Conceptual outline and first results

Roberto Theron, Eveline Wandl-Vogt, Jennifer Edmond, Cezary Mazurek

In the recent years, with the pervasiveness of computers and a great variety of electronic devices connected to the Internet, Digital Humanities (DH) as a research field has experienced a great transformation that has permitted the completion of very ambitious projects with large impact in the society beyond the academia. This has resulted in a major economic impact in the cultural and creative industry. A number of new and powerful ICT have made possible the exploitation of a wealth of data (either digitized or digitally born) that have, on the other hand, changed enormously the practice in DH, and exposed novel challenges that must be faced in order to complete any of the said projects. From the creation to the consumption of digital resources, there are new stakeholders, contexts and tasks to consider. The amount of digital resources produced (or digitized), stored, explored, and analysed in any DH project is immensely vast (especially if we take into account the introduction of linked-data), so the traditional humanities tools have to be either substituted or aided with ancillary tools in the form of interactive visualisations or novel user interfaces [3,4,9]. Furthermore, during the whole lifecycle of any DH project –from the data preparation to the actual analysis or exploration phase–, many decisions have to be made in order to yield the desired results that depend on the uncertainty pertaining to both the datasets and the models behind them.

The PROVIDEDH project (PROgressive VISual DEcision-Making in Digital Humanities) aims to provide visual interactive tools that convey the degree of uncertainty of the datasets and computational models used behind them, designed to progressively adapt the visualizations to incorporate the new, more complete or more accurate data. The project would not take only into account scholars, since it is most relevant in DH that the role of citizens has changed enormously. We live in a society that has democratized science, and the number of projects in which the contribution of citizens, either producing or using digital resources, has exploded. The experience gained in other areas of science in which the intervention of computing has been much deeper and constant will be analysed and adapted to the case of humanities. Especially, regarding infrastructures, frameworks, models and tools that can be standardized for the different disciplines in the humanities.

Thanks to the recent advances in ICT, DH now has the possibility to access a large number of different interconnected sources for analysis that allow the design and development of platforms and systems that help to carry out academic works of great scope and impact, while allowing their immediate exposure to society. At the same time, citizens can contribute to the progress of these works through the incorporation of new or better data to the available sources. However, there are many levels of uncertainty throughout the process, from the collection, preparation, processing and presentation of results.

PROVIDEDH aims to give DH scholars a space to explore and assess the completeness and evolution of the involved research objects [7], the degree of uncertainty [2] that the models applied to the data incorporate, and to share their perspectives and insights to other stakeholders of society.

PROVIDEDH revolves around two key scientific questions:

- Is it possible to assess the degree of uncertainty that a particular DH dataset incorporates, tracking such degree during the evolution of the dataset and how it is affected when different computational models are applied to it?
- Is it possible to convey this evolution of uncertainty by means of interactive multimodal interfaces that progressively adapt to the moment in which they are used and permit the users to make decisions accordingly?

Particularly difficult is the situation in DH, since it must be taken into account that any technological system or platform proposed towards giving an affirmative answer to the previous questions has to be made easily accessible by humanists and stakeholders that might not be digitally trained or very familiar with the inherent problem of uncertainty in their day by day practice.

With these two questions and context in mind, the general objectives of the PROVIDEDH project are the following:

- To understand all the sources of uncertainty that can affect the DH practice.
- To develop a set of metrics that convey the degree of uncertainty that research objects, data sets, and collections introduce as well as the different computational models applied to them.
- To propose a framework that makes use of the uncertainty metrics, so any given representation of the data can be assessed according to its degree of uncertainty.
- To propose a Progressive Visual Analytics [1,5,6,8,10] solution that ensures that users can follow the behaviour and degree of uncertainty of the underlying dataset as it evolves, i.e., able to trace changes in data and its inherent uncertainty as well as in the way it is perceived.
- To develop a web-based multimodal collaborative platform for the progressive visual analysis of different DH collections, both for scholars and citizen humanists.
- To trigger the formation of a “community of practice” that humanists can build on to reinforce each other’s efforts to achieve metrics that are practical and of high quality.

In this interaction we aim to introduce PROVIDEDH and present first results.

References:

1. Aigner, W., Miksch, S., Müller, W., Schumann, H., & Tominski, C. (2008). Visual methods for analyzing time-oriented data. *IEEE transactions on visualization and computer graphics*, 14(1), 47-60.
2. Barthelmé, S. (2010). *Visual uncertainty: a bayesian approach* (Doctoral dissertation, Paris 5).
3. Benito, A., Losada, A. G., Therón, R., Dorn, A., Seltmann, M., & Wandl-Vogt, E. (2016, November). A spatio-temporal visual analysis tool for historical dictionaries. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 985-990). ACM.
4. Champion, E. (2015). Seeing is revealing: A critical discussion on visualisation and the digital humanities. *Digital Humanities 2015*.
5. Fekete, J. D., & Primet, R. (2016). Progressive analytics: A computation paradigm for exploratory data analysis. *arXiv preprint arXiv:1607.05162*.
6. Frey, S., Sadlo, F., Ma, K. L., & Ertl, T. (2014). Interactive progressive visualization with space-time error control. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2397-2406.
7. Palma, R., Hołubowicz, P., Corcho, O., Gómez-Pérez, J. M., & Mazurek, C. (2014, May). Rohub—a digital library of research objects supporting scientists towards reproducible science. In *Semantic Web Evaluation Challenge* (pp. 77-82). Springer, Cham.
8. Stolper, C. D., Perer, A., & Gotz, D. (2014). Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1653-1662.
9. Theron, R., & Fontanillo, L. (2015). Diachronic-information visualization in historical dictionaries. *Information Visualization*, 14(2), 111-136.
10. Turkay, C., Kaya, E., Balcisoy, S., & Hauser, H. (2017). Designing progressive and interactive analytics processes for high-dimensional data analysis. *IEEE transactions on visualization and computer graphics*, 23(1), 131-140.

