PROVIDEDH

PROgressive VIsual DEcision-making for Digital Humanities

Roberto Theron
Universidad de Salamanca

MOTIVATION

During the whole lifecycle of any Digital Humanities 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 aims to provide visual interactive tools that convey the degree of uncertainty of the datasets and computational models used behind, designed to progressively adapt the visualizations to incorporate the new, more complete or more accurate data.

SCIENTIFIC BACKGROUND

PROVIDEDH PROJECT

TWO KEY SCIENTIFIC QUESTIONS

1. 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?

2. 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?

THE PROBLEM OF BIG DATA IN DH

- Increasing
  - Availability of computational methods
  - Number of interconnected devices through the Internet
  - Size and complexity of analyzed data sets
    - There's increasingly less interest in visualizing small datasets separately
- Motivates
  - Application of data-management algorithms
  - Fast replacement of old, paper-based research methodologies
- There is a need for
  - Novel computer techniques that deal with this problematic

USER INTERACTION

A delay of one second is the upper bound for computer responses after which users lose focus on their current train of thought [1,2,3,4,5]

[1] E. Zgraggen, A. Galakatos, A. Crotty, J. D. Fekete, and T. Kraska, "How Progressive Visualizations Affect Exploratory Analysis," IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 8, pp. 1977–1987, Aug. 2017.

[2] S. C. Seow, Designing and engineering time: The psychology of time perception in software. Addison-Wesley Professional, 2008.

[3] J. Nielsen, "Powers of 10: Time scales in user experience," Retrieved January, vol. 5, p. 2015, 2009.

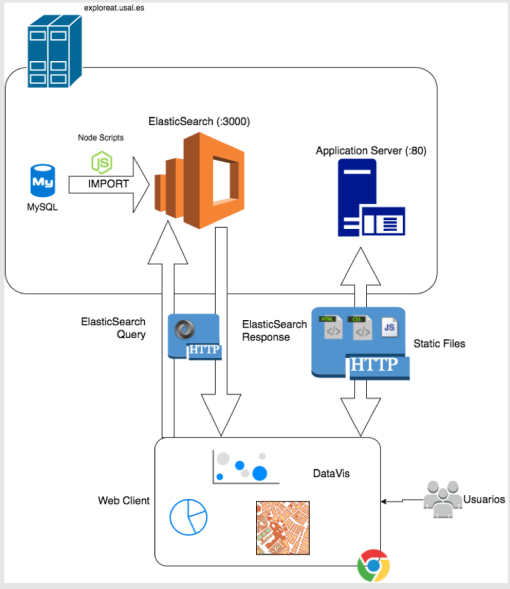
[4] S. K. Card, G. G. Robertson, and J. D. Mackinlay, "The information visualizer, an information workspace," in Proceedings of the SIGCHI Conference on Human factors in computing systems. ACM, 1991, pp. 181–186.

[5] B. Shneiderman, "Response time and display rate in human performance with computers," ACM Computing Surveys (CSUR), vol. 16, no. 3, pp. 265–285, 1984.

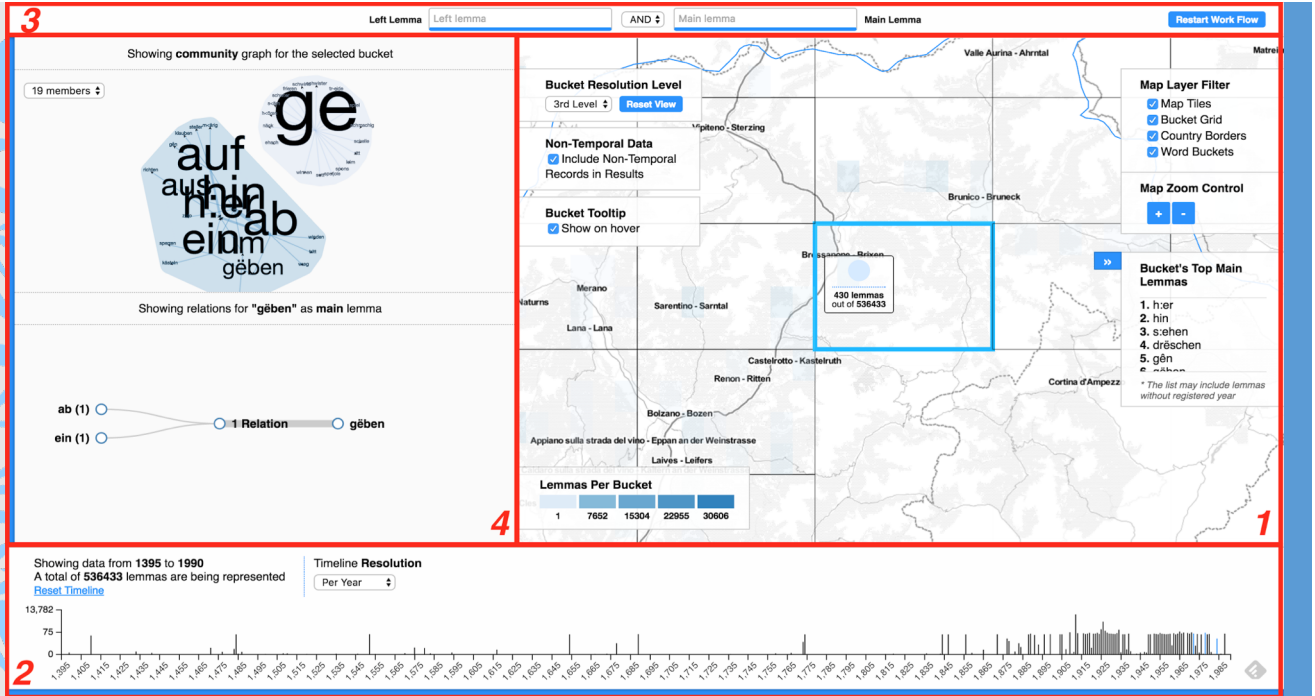
AN ARCHITECTURE PROPOSAL FOR DH*

1. There is not a standard architecture in DH

2. Our proposal is oriented towards:
   a. Deal with **big amounts** of records (>1M records)
   b. Enhance **user experience**
   c. **Reactive** components
   d. Supports **interactive time frames**



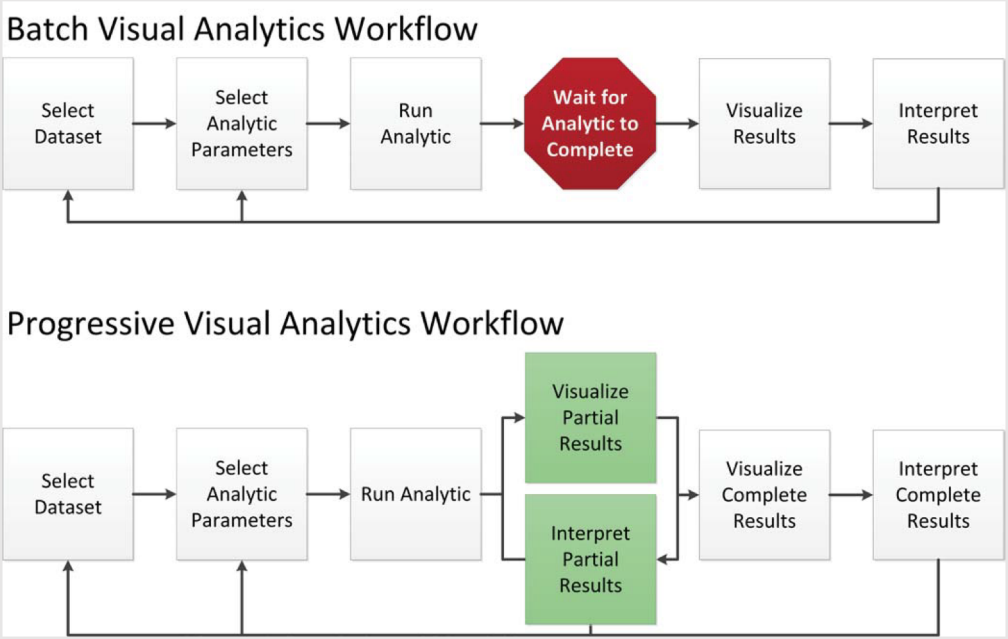
*From previous collaboration among partners of the consurtium.
—Project **exploreAT!** (http://exploreat.usal.es)



*From previous collaboration among partners of the consortium —Project **exploreAT!** (http://exploreat.usal.es)

PROGRESSIVE VISUAL ANALYSIS

- Processing data in incremental chunks rather than in one full piece allows for a quicker exploration of results that are refined over time increasing the precision of the visualized information progressively
- The communication between parts improves when employing progressive visualizations, as results are not displayed only after the whole data set has been processed in a blocking way, but rather on small iterative and detail-improving steps
- Progressive visual analytics depends on adapting analytical algorithms to produce meaningful partial results and enable analyst intervention without sacrificing computational speed
- Prioritizes subspaces of interest



Stolper et Al.: Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics (2014)

WHY ARE PROGRESSIVE VISUAL ANALYSIS AND UNCERTAINTY RELATED?

- The concepts behind progressive analytics and visualizations can also be applied to uncertainties in the data, as it relies on the same principles at its core
- In both cases data is shown to the analysts in incompleteness, due to it not having been fully processed yet (progressive analytics) or because of the lack of information (uncertainty)
- In both cases the data can become clearer to understand with time, either by keeping processing it periodically until it is fully covered, or by adding information to it to fill the uncertain gaps present in it → Make these differences visible

KEY CHALLENGES

PROVIDEDH PROJECT

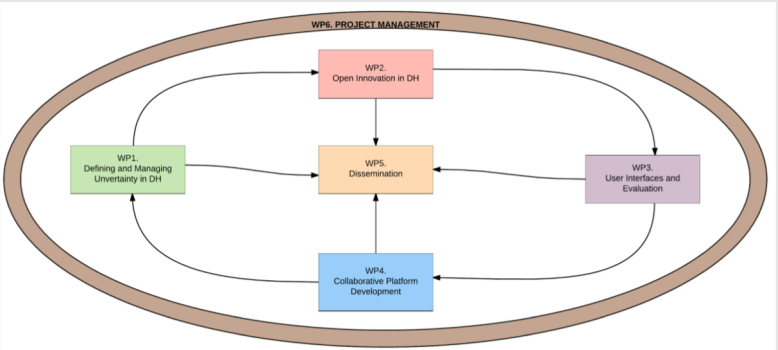
AIMS OF PROVIDEDH

- **Categorise all sources of uncertainty** that can affect the DH practice
- **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
- **Propose a (software) framework** that makes use of the uncertainty metrics, so any given representation of the data can be analysed
- **Clarify what DH tasks need decision-making** and build a pipeline oriented towards the handling of uncertainty

OVERVIEW

PROVIDEDH PROJECT

WORK PACKAGES



PROVIDEH GANTT		MONTH																																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
WP1	Defining and managing Uncertainty in DH																																				
WP2	Open innovation in the Humanities																																				
WP3	User interfaces and Evaluation																																				
WP4	Collaborative platform development																																				
WP5	Dissemination and Exploitation of Results																																				
WP6	Project Management																																				
	Milestones		X		X		X			X		X	X				X		X			X		X	X				X		X				X	X

POTENTIAL IMPACT

PROVIDEDH PROJECT

IMPACT

Both DH and the visualisation communities can benefit from the knowledge acquired by the PROVIDEDH project, which will deal with the adaptation of visual analytics to communicate sources of uncertainty to a very different context; this, in turn would transfer back out to other disciplines and communities

OUTCOMES (I)

- A **taxonomy of sources of uncertainty** that may appear during the lifecycle of a DH project. From its inception, collection, preparation, analysis, exposure and presentation, the different sources of uncertainty will be identified and understood.
- A **framework of metrics** that can be computed to assess the **degree of uncertainty** with which a particular cultural object or dataset is stored and that of the derived knowledge.
- A characterization of the particular needs and **roles of stakeholders beyond scholars within an open innovation science context**, which contribute data and derived knowledge with different levels of expertise or familiarity with the datasets that, in turn, bring different uncertainty sources to the collections.
- A **framework for the progressive visual analysis** of cultural datasets able to convey the change in the datasets and its corresponding uncertainty.

OUTCOMES (II)

- A **framework to embed uncertainty metrics in the visualizations** and or models most commonly used in DH, that takes into account the additional uncertainty that such representations and models introduce including the way they are perceived by users.
- A set of **multimodal alternatives** that enhance the user experience when dealing with uncertainty.
- An **open collaborative platform** in which progressive visual analysis for decision making under uncertainty can be performed for cultural data collections of general interest for the society.
- An **'uncertainty toolkit'** that takes the results of this work on humanities data and transfers their findings outward, exploring their implications for other sectors and user profiles.

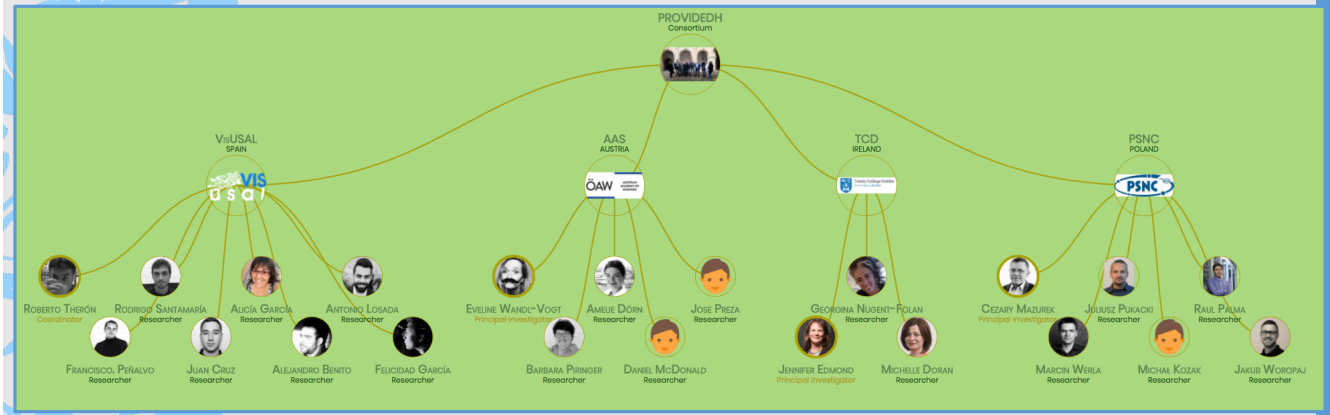
PARTNERS

PROVIDEDH PROJECT

PARTNERS

- University of Salamanca (SPAIN)
  — Visual Analytics Leader
  PI and Consortium Coordinator: Roberto Therón
- Austrian Academy of Sciences (AUSTRIA)
  — Open Innovation in Digital Humanities Leader
  PI: Eveline Wandl-Vogt
- Trinity College Dublin, Faculty of Arts Humanities and Social Sciences (IRELAND)
  — Engagement and Exploitation Leader
  PI: Jennifer Edmond
- Poznan Supercomputing and Networking Center (POLAND)
  — Infrastructures and Platform Leader
  PI: Cezary Mazurek

PARTNERS


PROVIDEDH
Consortium
VISUAL
SPAIN
AAS
AUSTRIA
TCD
IRELAND
PSNC
POLAND
Roberto Therón
Coordinator
Rodrigo Santamaría
Researcher
Alicia García
Researcher
Antonio Losada
Researcher
Francisco Peñalvo
Researcher
Juan Cruz
Researcher
Alejandro Benito
Researcher
Felicidad García
Researcher
Eveline Wandl-Vogt
Principal Investigator
Amelie Dörn
Researcher
Jose Preza
Researcher
Barbara Piringer
Researcher
Daniel McDonald
Researcher
Georgina Nugent-Folan
Researcher
Jennifer Edmond
Principal Investigator
Michelle Doran
Researcher
Cezary Mazurek
Principal Investigator
Juliusz Pukacki
Researcher
Raul Palma
Researcher
Marcin Werla
Researcher
Michał Kozak
Researcher
Jakub Woropaj
Researcher

PARTNERS

Kick-off Meeting at Salamanca (Spain)