PROVIDEDH PROGRESSIVE VISUAL DECISION-MAKING IN DIGITAL HUMANITIES

PROGRESSIVE VISUAL DECISION-MAKING IN DIGITAL HUMANITIES

TOWARDS UNCERTAINTY ANNOTATION IN TEI

Authors: Alejandro Benito (University of Salamanca), Michelle Doran (Trinity College Dublin), Jennifer Edmond (Trinity College Dublin), Michał Kozak (Poznań Supercomputing and Networking Center), Cezary Mazurek (Poznań Supercomputing and Networking Center), Alejandro Rodríguez (University of Salamanca), Roberto Therón (University of Salamanca)

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

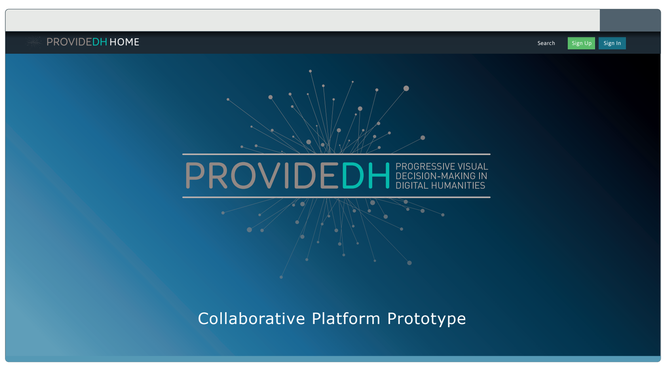
The PROVIDEDH project revolves around **two key scientific questions**:

1 — Is it possible to **assess the degree of uncertainty** a particular body of data leaves or introduces vis-a-vis any research question posed to it, tracking the points in the process at which uncertainty is introduced and the impact of the application of different computational models?

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?

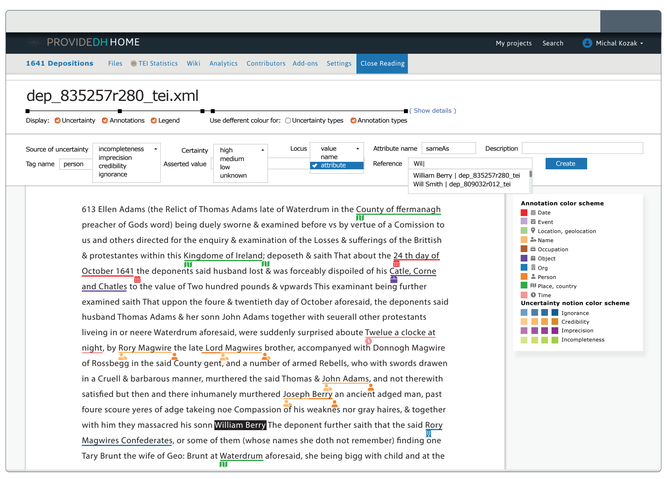
COLLABORATIVE PLATFORM

Our project seeks to promote a debate around uncertainty Digital Humanities datasets. We argue this debate can be enabled by means of a collaborative platform that is able to collect, organize and share the users' different perspectives on their own data or that of others.



COLLABORATIVE TEXT ANNOTATION

An user-friendly interface of annotating electronic text is included in the close reading module of our collaborative platform. Additionally, potential directions of visually tracking these annotations (provenace of the dataset) are included. The user can thus decide which snapshot of the dataset is better suited to the current research task.

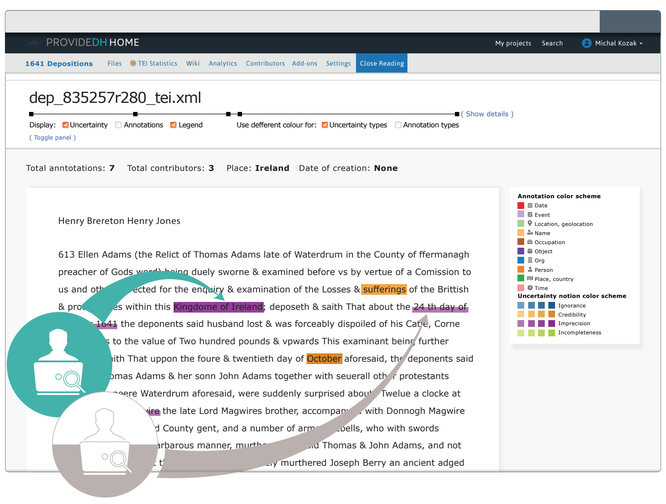


PROGRESSIVE VISUALIZATION

Computations in massive datasets may take too long to obtain a satisfactory user experience. On the other hand, our methodology invites researchers to explore their data from multiple different perspectives. We employ a novel visualization paradigm to produce user-steerable visualizations in order to tackle latency issues.

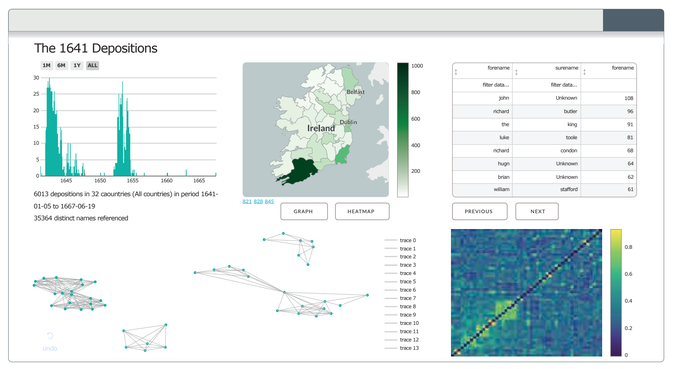
UNCERTAINTY ANNOTATION

Solutions for many possible cases are considered, like annotating the same or intersecting texts by many people with different profiles, or annotating already tagged entities by many people with different perspectives of those entities. A taxonomy of uncertainty sources have been proposed for the users to categorize their uncertainty annotations.



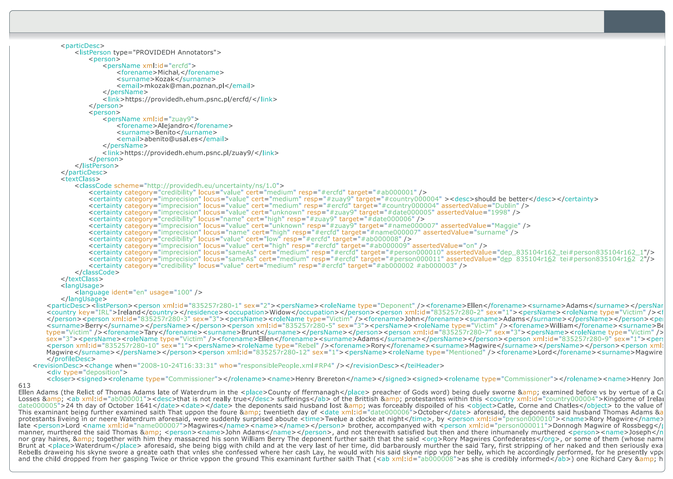
UNCERTAINTY VISUALIZATION

Our user-centered design process ensures the proposed visualizations are able to capture and expose the inherent uncertainty in a different range of humanistic research questions. Our first examples employ the 1641 Irish Depositions dataset and have been created in close collaborations with experts in the field.



UNCERTAINTY IN TEI

A standard way to annotate DH research datasets is through the use of TEI tags. Although the TEI provides various methods to indicate that some aspects of the encoded text are problematic or uncertain, it is not a common practice and uncertainty remains a challenge. We aim at bringing uncertainty to the surface by associating the existing TEI elements with the developed taxonomy of sources of uncertainty.



Contact / Roberto Therón (theron@usal.es)
For more details go to: **http://providedh.eu**