



Trinity College Dublin  
Coláiste na Tríonóide, Baile Átha Cliath  
The University of Dublin



Maynooth  
University  
National University  
of Ireland Maynooth



# Uncertainty in Humanities Research Datasets

Workshop Report

# Uncertainty in Humanities Research Datasets Workshop Report

---

**Edited by**  
Michelle Doran  
Jennifer Edmond

## Contents

<b>Executive Summary</b>	<b>2</b>
Why focus on Uncertainty in Digital Humanities Research?	2
<b>Workshop Background</b>	<b>3</b>
<b>Workshop Organisation</b>	<b>3</b>
<b>Recommendations</b>	<b>4</b>
General Recommendations	4
For the Platform Development	4
For the Taxonomy	4
Other Issues	5
<b>Next Steps</b>	<b>5</b>
<b>Acknowledgements</b>	<b>6</b>
<b>Appendix 1: Workshop Programme</b>	<b>7</b>
<b>Appendix 2: Taxonomy of Uncertainty V2.0</b>	<b>12</b>
<b>Appendix 3: List of Attendees</b>	<b>13</b>



## Executive Summary

This report arises from a one day workshop co-hosted by the [Humanities Research Institute at Maynooth University](#), the [Trinity College Dublin Centre for Digital Humanities](#) and the consortium members of the interdisciplinary [PROgressive Visual Decision-Making in Digital Humanities](#) (PROVIDEDH) project on Wednesday, 6 March 2019. The aim of the workshop was to explore the topic of uncertainty in Humanities research datasets and to draw recommendations from key stakeholders within the academic community as to how the PROVIDEDH project can best achieve its research objectives.

The PROVIDEDH project is a three-year project funded within the [CHIST-ERA](#) call 2016 for the topic “[Visual Analytics for Decision Making under Uncertainty – VADMU](#).” The project aims to give Digital Humanities (DH) scholars a space to explore and assess the completeness and evolution of digital research objects, the degree of uncertainty that the models applied to the data incorporate, and to share their perspectives and insights with the project’s broad range of stakeholders.

The workshop was attended by humanities researchers and computer scientists from two Irish universities (Trinity College Dublin and National University of Ireland Maynooth) as well as textual scholars from the University of Birmingham and representatives of the Irish, Polish and Spanish [PROVIDEDH partners](#) (further details of attendees available in **Appendix Three**).

## Why focus on Uncertainty in Digital Humanities Research?

The 2016 CHIST-ERA call states that “the central challenge of big data analytics is to understand their reliability, scope, and accuracy, and to communicate them to users in an intuitive manner”. Digital Humanities scholars generally acknowledge the fact that “knowing what archival data exists, in which form and where” is a huge problem, rendering it impossible to judge the completeness or accuracy of a given digital resource, and in turn, of the results. Digital Humanities’ methods and research questions require different approaches than others, which enable both the complexity and the noise surrounding the signal to be maintained. Therefore both Digital Humanities and the data visualisation community 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.

## Workshop Background

Between July and September 2018, a series of user scenarios and user stories were created from the results of a series of informal interviews conducted at Trinity College Dublin (TCD) with researchers working with the [1641 Depositions](#) dataset and/or related datasets all relating to the study of the history of Early Modern Ireland, the aim of which was to inform the development of the first iteration of the PROVIDEDH research platform. These user scenarios explore the place of uncertainty in the historical research processes that engage these and related records commonly used in coordination with each other.

At the encouragement of the TCD team, the PROVIDEDH project was supplied with copies of the original data from the digital edition of the 1641 Depositions, which was launched in 2009/10. These datasets were used in the creation of the nascent PROVIDEDH platform.

In November of the same year, members of the PROVIDEDH TCD team attended a locally hosted workshop titled 'Breaking Silos'. The purpose of the workshop was to develop a methodology for establishing ontologies to identify people and places in historical research across multiple datasets and platforms. Many of the datasets under consideration had also informed the creation of the PROVIDEDH user scenarios and user stories and the theme of uncertainty was pervasive throughout the workshop discussions. Following this workshop, the TCD team members approached the organisers with a view to organising a follow-on workshop to further explore this theme and it was agreed to co-host an interdisciplinary workshop on the subject.

## Workshop Organisation

The workshop revolved around two sessions. Both sessions included presentations and group discussions. The aim of the morning session was to articulate different perspectives on the role of uncertainty in Digital Humanities research. In the afternoon, participants were given the opportunity to offer feedback on both the project's nascent taxonomy of sources of uncertainty in DH and the approach adopted by the PROVIDEDH platform developers as well as to establish the needs and requirements of the historian, and indeed other humanities scholars, for

the next iteration of the digital research environment. The abstracts of the presentations are available in **Appendix One**.

## Recommendations

This section presents a summary of the recommendations that arose from the workshop. General recommendations have been drawn from the discussions in the morning sessions, and more specific project-related recommendations come from the afternoon feedback session.

### General Recommendations

#### For the Platform Development

- **It is important that the developers of research environments document the design process.** Understanding the methodologies employed in the development of Digital Humanities platforms necessitates the production of far more documentation concerning the provenance of the research environment than platform developers are accustomed to producing.
- The **relationship between the annotation environment (i.e. data analysis) and the data visualisation(s)** is unclear and requires further **clarification**.
- The question of the **authority of the annotator** was repeatedly raised. It was recommended that annotators should have a **user profile**. In the case of academic annotators, profiles could be linked to their **ORCID**.
- **Use cases should be provided to researchers to demonstrate how the platform enhances research.** This is particularly true of the historical research process where it was felt that the platform may not be able to make a meaningful contribution.

#### For the Taxonomy

- **The non-credibility label of the taxonomy requires revision as it currently conflates error and uncertainty** - error being binary (i.e. not uncertain, rather incorrect).
- **Instances of conflicting correct answers were highlighted in the discussion of the taxonomy.** In this regard working with probabilities rather than error may be beneficial as it is important to indicate which answer(s) are more likely to be correct without

excluding the possibility of alternatives. The extent to which this would be adopted by the research community was called into question, with the further recommendation that should it be incorporated into the platform it would require extensive user testing.

- **It was agreed that the taxonomy needs to provide for the identification of both the type of uncertainty and the source of uncertainty in the research process.** To that end, a matrix-based taxonomy may be of the most use to researchers.

## Other Issues

- **It is important to acknowledge that there exist a number of social barriers which will greatly impact the success of the project objectives.** The following emerged throughout the workshop discussions:
  - The issue of data citation was raised. Humanistic data curation and preparation is both a time-consuming and epistemically demanding process. An overlay system using such data therefore raises issues of both credit and bias, requiring robust mechanisms for traceability to be built in.
  - The historical research process necessitates an interrogation of primary source material and concern was expressed that distant reading tools such as the data visualisation tool of the PROVIDEDH platform distances the researcher from the objects of their research.
  - There is a general tendency among Arts and Humanities scholars to favour and reward lone scholarship. How might they be encouraged to share their data for others to work with as is required by a community annotation environment? How might we reward and recognise the labour involved in the creation of the original datasets?

## Next Steps

In the next phase of the project, the project consortium will aim to build on the workshop recommendations through the following steps:

1. To revise and publish the taxonomy of uncertainty;
2. To review the project platform and to incorporate the feedback;

3. Following this, the workshop attendees will be asked to test the revised platform and to offer feedback to the platform developers;
4. To develop case studies through the incorporation of the taxonomy into the the user scenarios which informed the first iteration of the research environment.

## Acknowledgements

**The members of the PROVIDEDH would like to thank the following:**

The staff in the Maynooth University Arts and Humanities Institute for their support in hosting the workshop on which this report is based. In particular, we would like to acknowledge the assistance of Stavros Angelis and Ann Donoghue for their administrative support as well as the generosity of the Director of the Institute, Prof. Thomas O'Connor, in providing funding for workshop catering.

Furthermore, we would like to extend our thanks to the speakers at the workshop Jack Kavanagh and Chris Beausang.

Lastly, we would like to thank each of the workshop attendees for generously giving their time and for their contributions throughout the workshop.

Appendix 1

Workshop Programme (including abstracts)



## Morning Session

The aim of the morning session (plenary) is to articulate different perspectives on the role of uncertainty in Digital Humanities research.

---

**10:30-11:00 Registration and Coffee**

**11:00-11:10 Introduction**

**11:10-11:30 Exploring the Place of Uncertainty in the Historical Research Process**

The focus of the paper is a set of user scenarios and user stories which were created from the results of a series of informal interviews conducted at Trinity College Dublin with researchers working with the 1641 Depositions dataset (<http://1641.tcd.ie/>) and/or related datasets all relating to the study of the history of Early Modern Ireland. The user scenarios explore the place of uncertainty in the historical research processes that engage these and related records commonly used in coordination with each other. The aim of the paper is to identify patterns which emerge within the narratives that can provide direction for the caretakers of cultural heritage datasets to support the resolution of the tasks that were captured in the interviews, chief among which are the challenges of accessibility and interoperability.

**Speaker** Dr. Michelle Doran, PROVIDEDH Research Fellow, Trinity College Dublin Centre for Digital Humanities.

**11:30-11:50 Uncertain data and findings: exploring the digitised records of the Military Archives of Ireland**

The Military Archives of Ireland since the late 2010s have released a large volume of archival information in a digitised format onto the web via the [militaryarchives.ie](http://militaryarchives.ie) website. The process of storing, cleaning and re-arranging this digitised data into a coherent structure has presented challenges for research and the presentation of coherent findings. Two examples of this will be explored

in the discussion of uncertain data: the Irish Army Census, 1922 and a dataset of ex-British Army personnel in the National Army created from a collection that scanned as part of archival preservation and is only available via CD-ROM.

**Speaker** Jack Kavanagh, Doctoral Candidate, Maynooth University Arts and Humanities Institute

**11:50-12:10** **Uncertainty in Quantitative Literary Criticism**

Many schools within contemporary literary criticism focus on broader cultural trends, at the expense of the individual component parts which together compose a text's structure. Statistical analysis meanwhile, requires formalised and unambiguous data structures which can often seem reductive to literary scholars trained in critical theory and historicisation. Finally, there is little consensus within quantitative literary criticism as to what a successful 'result' actually is. This presentation will consider these challenges and propose a middle ground via network analysis and community detection algorithms in a bid to move the critical discourse forward.

**Speaker** Chris Beausang, Doctoral Candidate, Maynooth University Arts and Humanities Institute

**12:10-13:00** **Questions and Answers**

**13:00-14:00** **Lunch**

## Afternoon Session

The aims of the afternoon session are to get feedback from participants on both the project's nascent taxonomy of sources of uncertainty in Digital Humanities and the approach adopted by the PROVIDEDH platform developers as well as to establish the needs and requirements of the historian, and indeed other humanities scholars, for the next iteration of the digital research environment. The session will feature a series of recent works on the 1641 Depositions Datasets orientated towards the interactive exploration and collaborative annotation of the texts of the collection.

---

### 14:00-14:45 Interactive Visualisation of the 1641 Depositions Dataset

The session will consist of the following:

- Presentation of results of the preliminary analysis of the dataset with an emphasis on describing sources of uncertainty, structural holes and other traits of interest for the PROVIDEDH project.
- A showcase of spatial and dynamic network analysis prototypes. The use cases will include:
  - The disambiguation of persons' names and toponyms
  - The visual representation of the narrative structure of a deposition
  - Live uncertainty tuning.
- A presentation of an interface prototype for the collaborative annotation of uncertainty.

**Speakers** Prof Roberto Theron, Group Manager of the Visual Analytics and Information Visualization Research Group (VisUsal), Computer Science and Automation, PROVIDEDH Principal Investigator, University of Salamanca and Alejandro Benito, PROVIDEDH Research Fellow, Doctoral Candidate, Computer Science and Automation, University of Salamanca.

### 14:45-15:00 PROVIDEDH Taxonomy of Uncertainty V2.

This session will present the project consortium's nascent taxonomy of uncertainty in Digital Humanities which seeks to translate established notions of epistemic and aleatory uncertainty in spatial data into relevant categorisations of uncertainty for present-day Digital Humanities research and research environments.

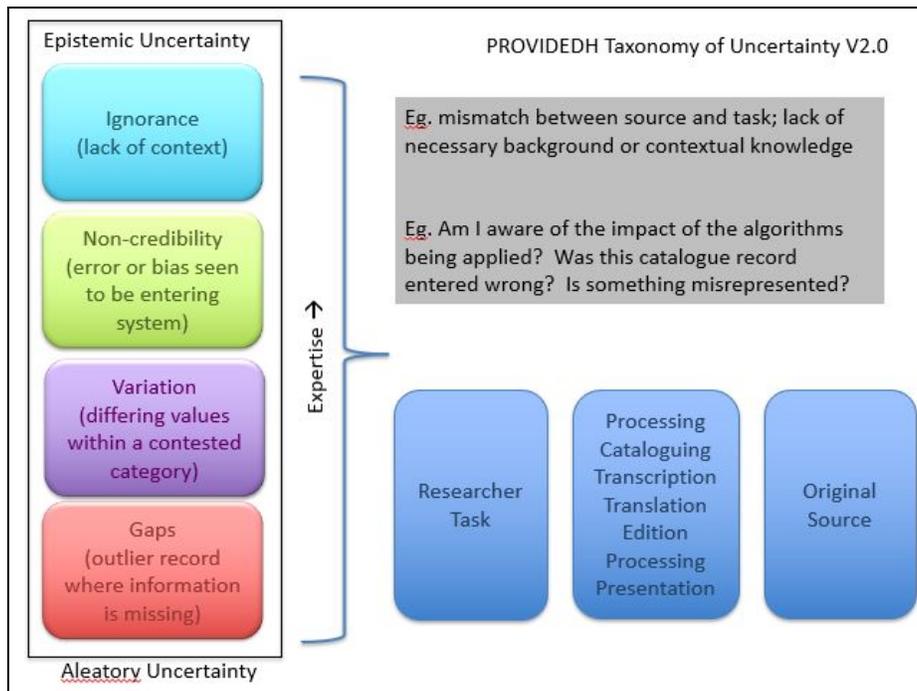
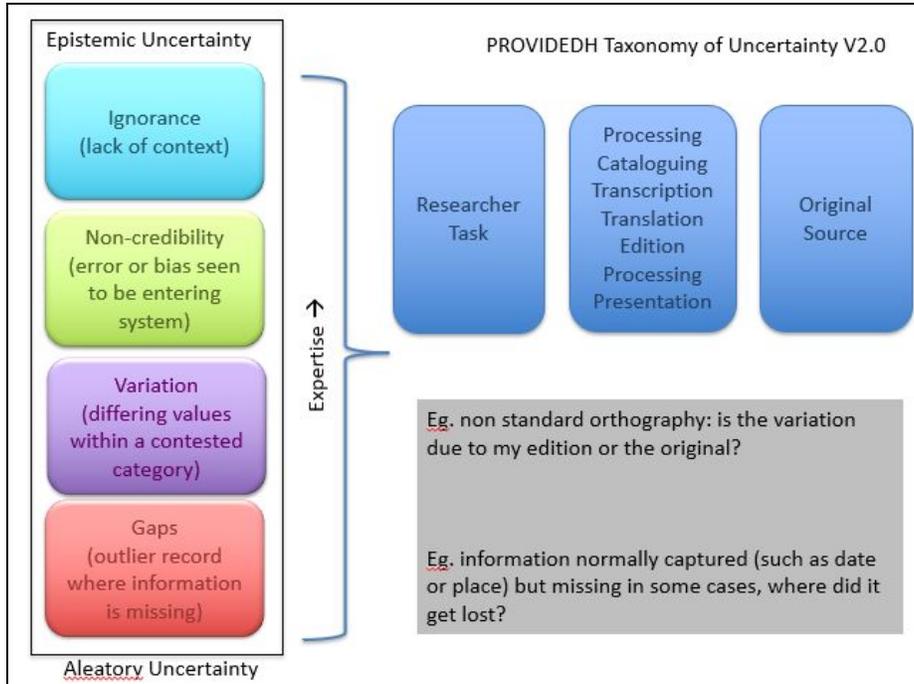
**Speaker** Prof Jennifer Edmond, Director of the Centre for Digital Humanities, PROVIDEDH WP Lead, Trinity College Dublin.

**15:00-15:50** **Feedback to PROVIDEDH Consortium Members from Workshop Attendees**

**15:50-16:00** **Closing Remarks**

Appendix 2

PROVIDEDH Taxonomy of Uncertainty V2.0



Appendix 3

List of Attendees

Name	Institute	Position
Prof. Helen Abbott	University of Birmingham	Professor of Modern Languages
Stavros Angelis	National University of Ireland, Maynooth	Arts and Humanities Institute Technology Officer
Alejandro Benito	University of Salamanca	Research Fellow, Doctoral Candidate, Computer Science and Automation
Chris Beausang	National University of Ireland, Maynooth	Doctoral Candidate, Digital Arts & Humanities
Dr. David Brown	University of Dublin, Trinity College	Research Fellow in History
Prof. Costis Dallas	University of Toronto	Associate Professor in the Faculty of Information
Dr. Michelle Doran	University of Dublin, Trinity College	Research Fellow in Digital Humanities
Prof. Jennifer Edmond	University of Dublin, Trinity College	Associate Professor in Digital Humanities
Jack Kavanagh	National University of Ireland, Maynooth	Doctoral Candidate, Digital Arts & Humanities
John Keating	National University of Ireland, Maynooth	Senior Lecturer, Computer Science
Michał Kozak	Poznan Supercomputing and Networking Center	Knowledge Platforms Team Leader at Poznan Supercomputing and Networking Center
Cezary Mazurek	Poznan Supercomputing and Networking Center	Director of Poznan Supercomputing and Networking Center
Shane McGarry	National University of Ireland, Maynooth	Doctoral Candidate, Computer Science
Dr Rebecca Mitchell	University of Birmingham	Reader in Victorian Literature and Culture, Head of Research, EDACS

Prof. Thomas O'Connor	National University of Ireland, Maynooth	Director of Arts & Humanities Institute
Prof. Micheál Ó Siochrú	University of Dublin, Trinity College	Head of Department of History
Prof Roberto Theron	University of Salamanca	Group Manager of the Visual Analytics and Information Visualization Research Group (VisUsal), Computer Science and Automation
Dr Aengus Ward	University of Birmingham	Professor in Medieval Iberian Studies