

## Analyzing and Visualizing Uncertain Knowledge: Introducing the PROVIDEDH Open Science Platform

Alejandro Benito, Michelle Doran, Jennifer Edmond, Michał Kozak,  
Cezary Mazurek, Alejandro Rodríguez, Roberto Therón, Eveline Wandl-Vogt

*What is text, really? TEI and beyond (TEI 2019) September 16 – 20, University of Graz, Austria*



## PROVIDEDH

PROgressive VIsual DEcision-Making in Digital Humanities (PROVIDEDH) project:

- a three-year project funded within the CHIST-ERA call 2016 for the topic “Visual Analytics for Decision Making under Uncertainty – VADMU.”
- 4 partners (2 technical and 2 humanistic):
  - The Research Group on Visual Analytics, University of Salamanca, Spain
  - Poznań Supercomputing and Networking Center, Polish Academy of Sciences, Poland
  - Trinity Long Room Hub Arts & Humanities Research Institute, Trinity College Dublin, Ireland
  - Austrian Centre for Digital Humanities of the Austrian Academy of Sciences, Austria
- Underlying uncertainty in DH research data affects decision-making and persists during the project’s lifecycle. This uncertainty will always be present. **Thus, efforts in providing technical support for humanistic research should focus on managing and making it more transparent, rather than removing it.**

## Aleatoric and Epistemic Uncertainty

Anyone using uncertain information needs to think carefully about the possible sources of uncertainty, and how they may be addressed (by probability theory, belief function theory, fuzzy set theory, etc.)

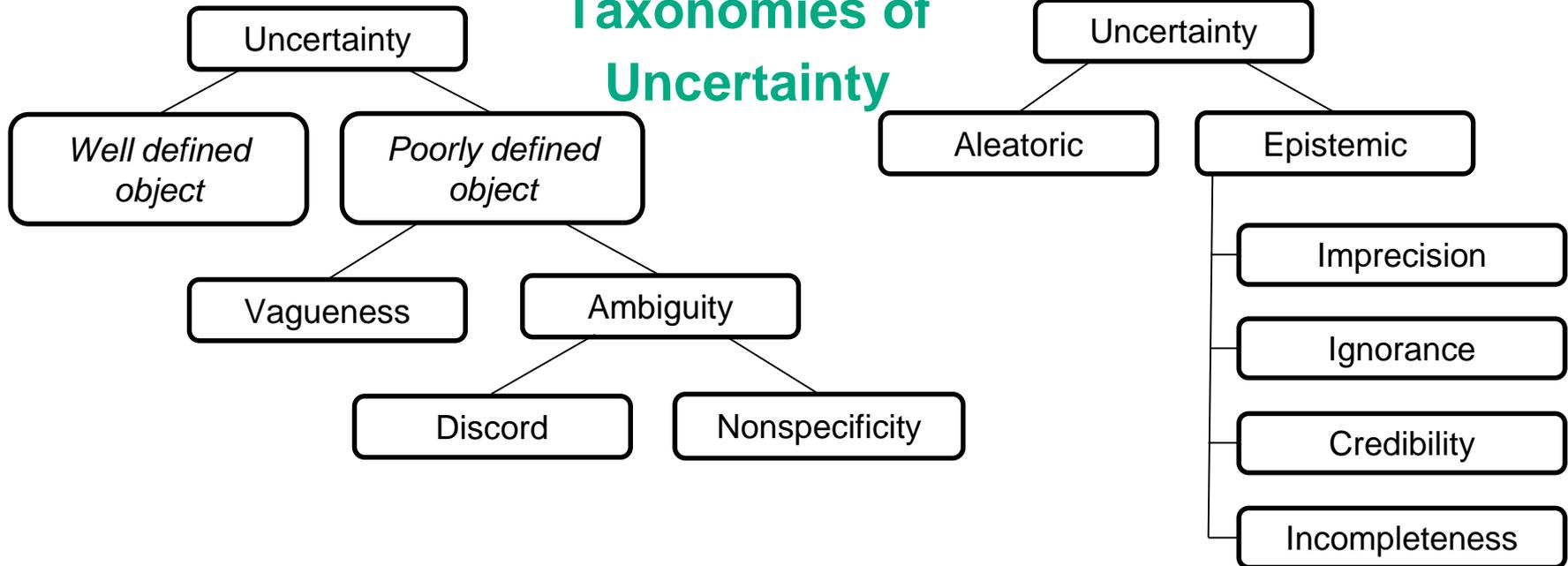
In the literature (sources of) uncertainty is often classified into two categories: aleatoric and epistemic.

- **Aleatoric** uncertainty, which is regarded as inherent in the phenomenon and can not be reduced. It exists due to the random nature of physical events.
- **Epistemic** uncertainty, which results from our incomplete knowledge and could in principle be reduced, although this may be impractical, not possible in the framework of available time and resources, or many similar reasons. It is associated with the user performing the analysis, his ignorance or the human nature of making mistakes.

Peter F. Fisher. Models of uncertainty in spatial data. Geographical information systems, 1, 191–205, 1999.

Christophe Simon, Philippe Weber, Mohamed Sallak. Data Uncertainty and Important Measures; John Wiley & Sons, 2018.

## Taxonomies of Uncertainty



## Taxonomy of Epistemic Uncertainty

- **Imprecision** corresponds to the inability to express the true value because the absence of experimental values does not allow the definition of a probability distribution or because it is difficult to obtain the exact value of a measure.
- **Ignorance** is related to the fact that information could have been incorrectly assessed by the person gathering or organizing the data. It is also possible that people, not fully sure about how to deal with data, ignore some information and generate uncertainty during the evaluation and decision processes.
- **Incompleteness** corresponds to the fact that not all situations are covered. Often it is impossible to know every possible option available.
- **Credibility** (discord) concerns the weight that an agent can attach to its judgment. This concept can be linked to that of biased opinions, which are related to personal visions of the landscape, which can make for significant variations between different groups and individuals, given their backgrounds.

...

## Uncertainty annotation in TEI

Locating and tracing uncertainty through the evolution of a textual corpus can be done with the use of TEI tags [Guidelines for Electronic Text Encoding and Interchange, Section 21.1.2 Structured Indications of Uncertainty]

- **<certainty>**
- **<precision>**
- **@cert**

However, the use these method is not a common practice.

We aim to promote a wider use of these tags

- by providing a user-friendly interface for collaborative annotating texts with uncertainty,
- by associating the existing TEI elements with the developed taxonomy of uncertainty,
- by implementing an environment for progressive visualisations of uncertainty of annotated data (at first we are implementing an interactive application for one type of imprecision – name entities with many name variations)

## <certainty> - category of uncertainty

In our platform we would like to use the <certainty> element. Unfortunately it has no attribute where we can note the category of uncertainty. So we would like to add such attribute to the specification:

```
<elementSpec ident="certainty" module="certainty" mode="change">
  <attList>
    <attDef ident="category" mode="add">
      <desc xml:lang="en">indicates the category of uncertainty</desc>
      <datatype><dataRef key="teidata.enumerated"/></datatype>
      <valList type="semi" mode="change">
        <valltem mode="add" ident="imprecision"/>
        <valltem mode="add" ident="ignorance"/>
        <valltem mode="add" ident="incompleteness"/>
        <valltem mode="add" ident="credibility"/>
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

## <certainty> - @locus

Currently the locus element indicates more exactly the aspect concerning which certainty is being expressed: specifically, whether the markup is correctly located, whether the correct element or attribute name has been used, or whether the content of the element or attribute is correct, etc. It's closed list of the following attributes:

- **name**: uncertainty concerns whether the name of the element or attribute used is correctly applied
- **start**: uncertainty concerns whether the start of the element is correctly identified,
- **end**: uncertainty concerns whether the end of the element is correctly identified,
- **location**: uncertainty concerns both the start and the end of the element,
- **value**: uncertainty concerns the content (for an element) or the value (for an attribute).

Two of these option are ambiguous. In case of name and value, we don't know exactly whether the uncertainty concerns the name or attribute and which attribute. For automatic processing of XML files it's a problem. Of course, TEI has a solution for it by using a pair of attributes "target" and "match". But we would like a bit more 😊

## <certainty> - @locus

Fragment of the Deposition 821026r012:

```
<p>... And this deponent was credibly told by one dorothy Hanse, and divers others that came from
  <location>
    <placeName>
      <settlement xml:id="settlement00001" type="city">Cashell</settlement>
    </placeName>
  </location>, That the Rebels had killed ...
</p>
```

Annotating that Cashell is misspelled (and it might be a different city) is quite simple:

```
<certainty category="imprecision" locus="value" target="#settlement00001" assertedValue="Cashel"/>
```

How to annotate that the "type" attribute is wrongly applied?

## <certainty> - doubts about an attribute

```
<p>... And this deponent was credibly told by one dorothy Hanse, and divers others that came from  
<location>  
  <placeName>  
    <settlement xml:id="settlement00001" type="city">Cashell</settlement>  
  </placeName>  
</location>, That the Rebels had killed ...  
</p>
```

```
<certainty category="imprecision" locus="value" target="#settlement00001" assertedValue="town"/>
```

## <certainty> - doubts about an attribute

```
<p>... And this deponent was credibly told by one dorothy Hanse, and divers others that came from  
<location>  
  <placeName>  
    <settlement xml:id="settlement00001" type="city">Cashell</settlement>  
  </placeName>  
</location>, That the Rebels had killed ...  
</p>
```

ambiguous  
?

```
<certainty category="imprecision" locus="value" target="#settlement00001" assertedValue="town"/>
```

## <certainty> - doubts about an attribute

```
<p>... And this deponent was credibly told by one dorothy Hanse, and divers others that came from
<location>
  <placeName>
    <settlement xml:id="settlement00001" type="city">Cashell</settlement>
  </placeName>
</location>, That the Rebels had killed ...
</p>
```

```
<certainty category="imprecision" locus="value" target="#settlement00001" assertedValue="town"/>
```

```
<certainty category="imprecision" locus="value" target="#settlement00001" match="@type" assertedValue="town"/>
```

## att.scoping: @target and @match

The pair of att.scoping attributes provides **attributes for selecting particular elements within a document**:

- **target**: points at one or several elements or sets of elements by means of one or more data pointers, using the URI syntax.
- **match**: supplies an arbitrary XPath expression which identifies a set of nodes, selected within the context identified by the target attribute if this is supplied, or within the context of the parent element if it is not.

An XPath expression identifies a set of nodes in an XML document. But what if this set does not exist in the document, but we would like to refer to it?

## Use-case of unifying people (entities)

Fragment of the Deposition 823180r162:

```
<p>... a rebellious maner the wife of  
<person xml:id="person00002">Anselmus Adams</person> (&amp; her two children)...  
</p>
```

Fragment of the Deposition 824148r131:

```
<p>... then <add place="inline">&amp; there</add> murdred the wife of  
<person xml:id="person00005">Ancelmus Adams</person> (and two children of his)...  
</p>
```

We would like to annotate with a high degree of certainty that these two people are the same person.

```
<certainty category="imprecision" locus="value" target="#person00002" match="@sameAs"  
assertedValue="824148r131#person00005"/>
```

## Use-case of unifying people (entities)

Fragment of the Deposition 823180r162:

```
<p>... a rebellious maner the wife of  
<person xml:id="person00002">Anselmus Adams</person> (& her two children)...  
</p>
```

Fragment of the Deposition 824148r131:

```
<p>... then <add place="inline">& there</add> muredred the wife of  
<person xml:id="person00005">Ancelmus Adams</person> (and two children of his)...  
</p>
```

We would like to annotate with a high degree of certainty that these two people are the same person.

```
<certainty category="imprecision" locus="value" target="#person00002" match="@sameAs"  
assertedValue="824148r131#person00005"/>
```

serious  
doubts

empty set of  
nodes

## Use-case of unifying people (entities)

Fragment of the Deposition 823180r162:

```
<p>... a rebellious maner the wife of  
<person xml:id="person00002">Anselmus Adams</person> (& her two children)...  
</p>
```

Fragment of the Deposition 824148r131:

```
<p>... then <add place="inline">& there</add> muredred the wife of  
<person xml:id="person00005">Ancelmus Adams</person> (and two children of his)...  
</p>
```

We would like to annotate with a high degree of certainty that these two people are the same person.

```
<certainty category="imprecision" locus="sameAs" target="#person00002"  
assertedValue="824148r131#person00005"/>
```



solution

## <certainty> - changes for locus

```
<elementSpec ident="certainty" module="certainty" mode="change">
```

```
<attList>
```

```
<attDef ident="category" mode="change">
```

<desc xml:lang="en">indicates more exactly the aspect concerning which certainty is being expressed: specifically, whether the markup is correctly located, whether the correct element or attribute name has been used, or whether the content of the element or attribute is correct, etc. Doubts about attributes and their values can be reflected by assigning values beyond the semi-open list, i.e. by assigning the name of an attribute.</desc>

```
<datatype><dataRef key="teidata.enumerated"/></datatype>
```

```
<valList type="semi" mode="change">
```

```
<valltem mode="change" ident="name"/>
```

```
<valltem mode="change" ident="start"/>
```

```
<valltem mode="change" ident="end"/>
```

```
<valltem mode="change" ident="location"/>
```

```
<valltem mode="change" ident="value"/>
```

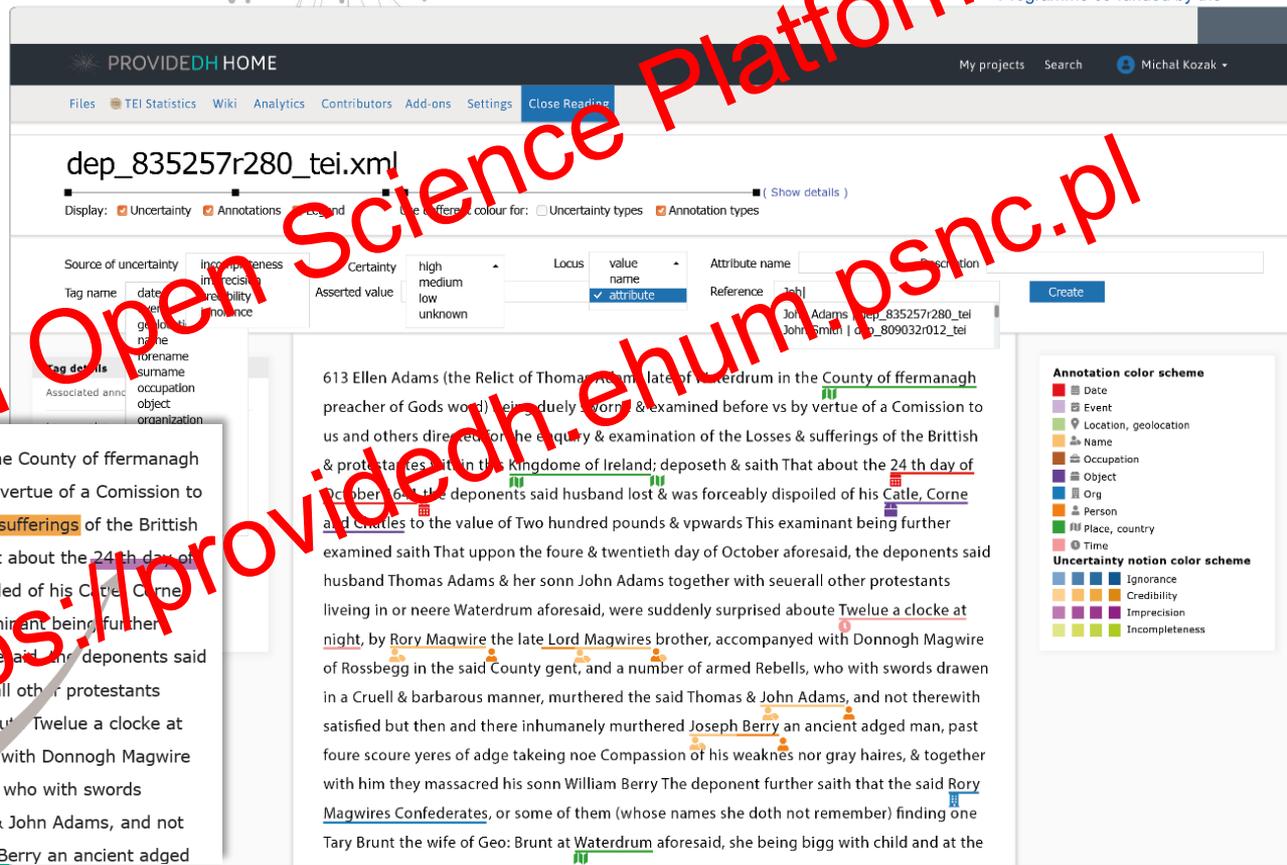
```
</valList>
```

```
</attDef>
```

```
</attList>
```

# Collaborative Text Annotation

PROVIDEDH Open Science Platform  
<https://providedh.ehum.psnc.pl>



The screenshot shows the PROVIDEDH HOME interface. At the top, there are navigation links: Files, TEI Statistics, Wiki, Analytics, Contributors, Add-ons, Settings, and Close Reading. The main content area displays a document titled 'dep\_835257r280\_tei.xml'. Below the title, there are display options for 'Uncertainty' and 'Annotations'. A table below the document shows metadata for the text, including 'Source of uncertainty', 'Incompleteness in precision', 'Certainty', 'Locus', 'value name', and 'Attribute name'. The table has columns for 'Tag name', 'date', 'incompleteness in precision', 'Certainty', 'Locus', 'value name', 'Attribute name', and 'Description'. The 'value name' column is currently set to 'attribute'. To the right of the table, there is a 'Create' button. Below the table, there is a text document with annotations. The text is from a historical document and includes phrases like '613 Ellen Adams (the Relict of Thomas Adams late of Waterdrum in the County of ffermanagh preacher of Gods word) being duly sworn & examined before vs by vertue of a Comission to us and others directed for the enquiry & examination of the Losses & sufferings of the British & protestants in the Kingdome of Ireland; deposeeth & saith That about the 24 th day of October 64 the deponents said husband lost & was forcibly dispoiled of his Cattle, Corne and Cuddeles to the value of Two hundred pounds & vpwards This examinant being further examined saith That upon the foure & twentieth day of October aforesaid, the deponents said husband Thomas Adams & her sonn John Adams together with seuerall other protestants living in or neere Waterdrum aforesaid, were suddenly surprised aboute Twelue a clocke at night, by Rory Magwire the late Lord Magwires brother, accompanied with Donnogh Magwire of Rossbegg in the said County gent, and a number of armed Rebels, who with swords drawn in a Cruell & barbarous manner, murdered the said Thomas & John Adams, and not therewith satisfied but then and there inhumanely murdered Joseph Berry an ancient adged man, past foure scoure yeres of adge taking noe Compassion of his weaknes nor gray haire, & together with him they massacred his sonn William Berry The deponent further saith that the said Rory Magwires Confederates, or some of them (whose names she doth not remember) finding one Tary Brunt the wife of Geo: Brunt at Waterdrum aforesaid, she being bigg with child and at the

**Annotation color scheme**

- Date
- Event
- Location, geolocation
- Name
- Occupation
- Object
- Org
- Person
- Place, country
- Time

**Uncertainty notion color scheme**

- Ignorance
- Credibility
- Imprecision
- Incompleteness



# TEI outcome

```

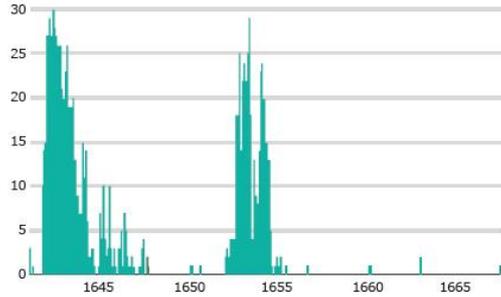
<particDesc>
  <listPerson type="PROVIDEDH Annotators">
    <person>
      <persName xml:id="ercfd">
        <forename>Michał,</forename>
        <surname>Kozak</surname>
        <email>mkozak@man.poznan.pl</email>
      </persName>
      <link>https://providedh.ehum.psn.pl/ercfd/</link>
    </person>
    <person>
      <persName xml:id="zuay9">
        <forename>Alejandro</forename>
        <surname>Benito</surname>
        <email>abenito@usal.es</email>
      </persName>
      <link>https://providedh.ehum.psn.pl/zuay9/</link>
    </person>
  </listPerson>
</particDesc>
<textClass>
  <classCode scheme="http://providedh.eu/uncertainty/ns/1.0">
    <certainty category="credibility" locus="value" cert="medium" resp="#ercfd" target="#ab000001" />
    <certainty category="imprecision" locus="value" cert="medium" resp="#zuay9" target="#country000004" ><desc>should be better</desc></certainty>
    <certainty category="imprecision" locus="value" cert="medium" resp="#ercfd" target="#country000004" assertedValue="Dublin" />
    <certainty category="imprecision" locus="value" cert="unknown" resp="#zuay9" target="#date000005" assertedValue="1998" />
    <certainty category="credibility" locus="name" cert="high" resp="#zuay9" target="#date000006" />
    <certainty category="imprecision" locus="value" cert="unknown" resp="#zuay9" target="#name000007" assertedValue="Maggie" />
    <certainty category="imprecision" locus="name" cert="high" resp="#ercfd" target="#name000007" assertedValue="surname" />
    <certainty category="credibility" locus="value" cert="low" resp="#ercfd" target="#ab000008" />
    <certainty category="imprecision" locus="value" cert="high" resp="#ercfd" target="#ab000009" assertedValue="on" />
    <certainty category="imprecision" locus="sameAs" cert="medium" resp="#ercfd" target="#person000010" assertedValue="dep_835104r162_tei#person835104r162_1" />
    <certainty category="imprecision" locus="sameAs" cert="medium" resp="#ercfd" target="#person000011" assertedValue="dep_835104r162_tei#person835104r162_2" />
    <certainty category="credibility" locus="value" cert="medium" resp="#ercfd" target="#ab000002 #ab000003" />
  </classCode>
</textClass>
<langUsage>
  <language ident="en" usage="100" />
</langUsage>
</profileDesc>
<revisionDesc><change when="2008-10-24T16:33:31" who="responsiblePeople.xml#RP4" /></revisionDesc></teiHeader>
613
Ellen Adams (the Relict of Thomas Adams late of Waterdrum in the <place>County of ffermanagh</place> preacher of Gods word) being duely sworn &amp; examined before vs by vertue of f
Losses &amp; <ab xml:id="ab000001"><desc>that is not really true</desc> sufferings</ab> of the Brittish &amp; protestantes within this <country xml:id="country000004">Kingdome of f
date000005">24 th day of October 1641</date><date></date> the deponents said husband lost &amp; was forcibly spoiled of his <object>Cattle, Come and Chatles</object> to the valu
This examinant being further examined saith That upon the foure &amp; twentieth day of <date xml:id="date000006">October</date> aforesaid, the deponents said husband Thomas Adam
protestants livinge in or neere Waterdrum aforesaid, were suddenly surprised aboute <time>Twelve a clocke at night</time>, by <person xml:id="person000010"><name>Rory Magwire</na
late <person>Lord <name xml:id="name000007">Magwires</name></person> brother, accompanied with <person xml:id="person000011">Donnogh Magwire of Rossgeb
manner, murdered the said Thomas &amp; <person><name>John Adams</name></person>, and not therewith satisfied but then and there inhumanely murdered <person><name>Joseph
nor gray haire, &amp; together with him they massacred his sonn William Berry The deponent further saith that the said <org>Rory Magwires Confederates</org>, or some of them (whose r
Brunt at <place>Waterdrum</place> aforesaid, she beinge bigg with child and at the very last of her time, did barbarously murder the said Tary, first stripping of her naked and then seriously
Rebells drawing his skyne swore a greate oath that vnles she confessed where her cash Lay, he would with his said skyne ripp vpp her belly, which he accordingly performed, for he presently
and the child dropped from her gasping Twice or thrice vppon the ground This examinant further saith That (<ab xml:id="ab000008">as she is credibly informed</ab>) one Richard Cary &

```



### The 1641 Depositions

1M 6M 1Y ALL



6013 depositions in 32 caountries (All countries) in period 1641-01-05 to 1667-06-19

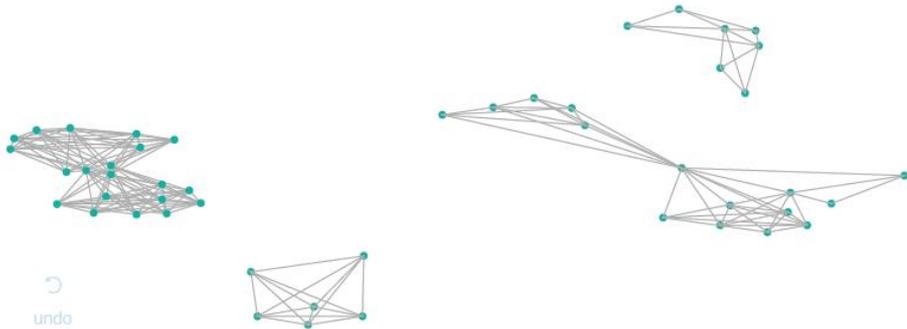
35364 distinct names referenced



821 828 845

GRAPH

HEATMAP

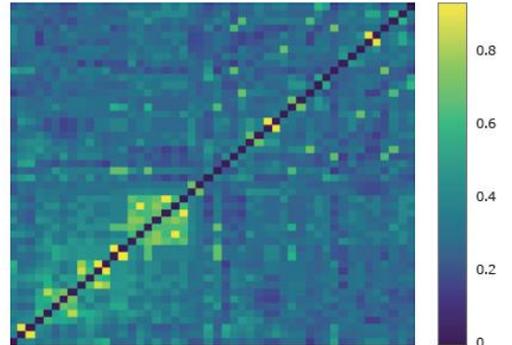


- trace 0
- trace 1
- trace 2
- trace 3
- trace 4
- trace 5
- trace 6
- trace 7
- trace 8
- trace 9
- trace 10
- trace 11
- trace 12
- trace 13

↓	forename	↓	surname	↓	forename
	filter data...		filter data...		
	john		Unknown		108
	richard		butler		96
	the		king		91
	luke		toole		81
	richard		condon		68
	hugn		Unknown		64
	brian		Unknown		62
	william		stafford		61

PREVIOUS

NEXT





Programme co-funded by the  
EUROPEAN UNION

**Thank you for your attention!**