Digital Historical Research: An Ideal Case Study for Networked Visualisation Research

Michael Hawkins (University of Sussex / University of Cambridge), Simon Julier (UCL), Robert Iliffe (University of Sussex), Tim Weyrich (UCL), Melissa Terras (UCL), Claire Warwick (UCL)

Our Thesis

- Historians are a kind of intelligence analyst

- Therefore:
  - If we can design tools to aid the historical analysis process,
  - We can discover underlying principles and best practices for many other visual analysis tasks as well

What Do Historians Do?

- A historian seeks *plausible explanations* for a set of facts

- These explanations links networks of evidence together

- The links are based on *informed opinion*

- Multiple explanations are always possible

A Brief History of Newton

- 4 January 1643 - 31 March 1726

- Professor at Cambridge until 1696

- Warden of Royal Mint until 1726

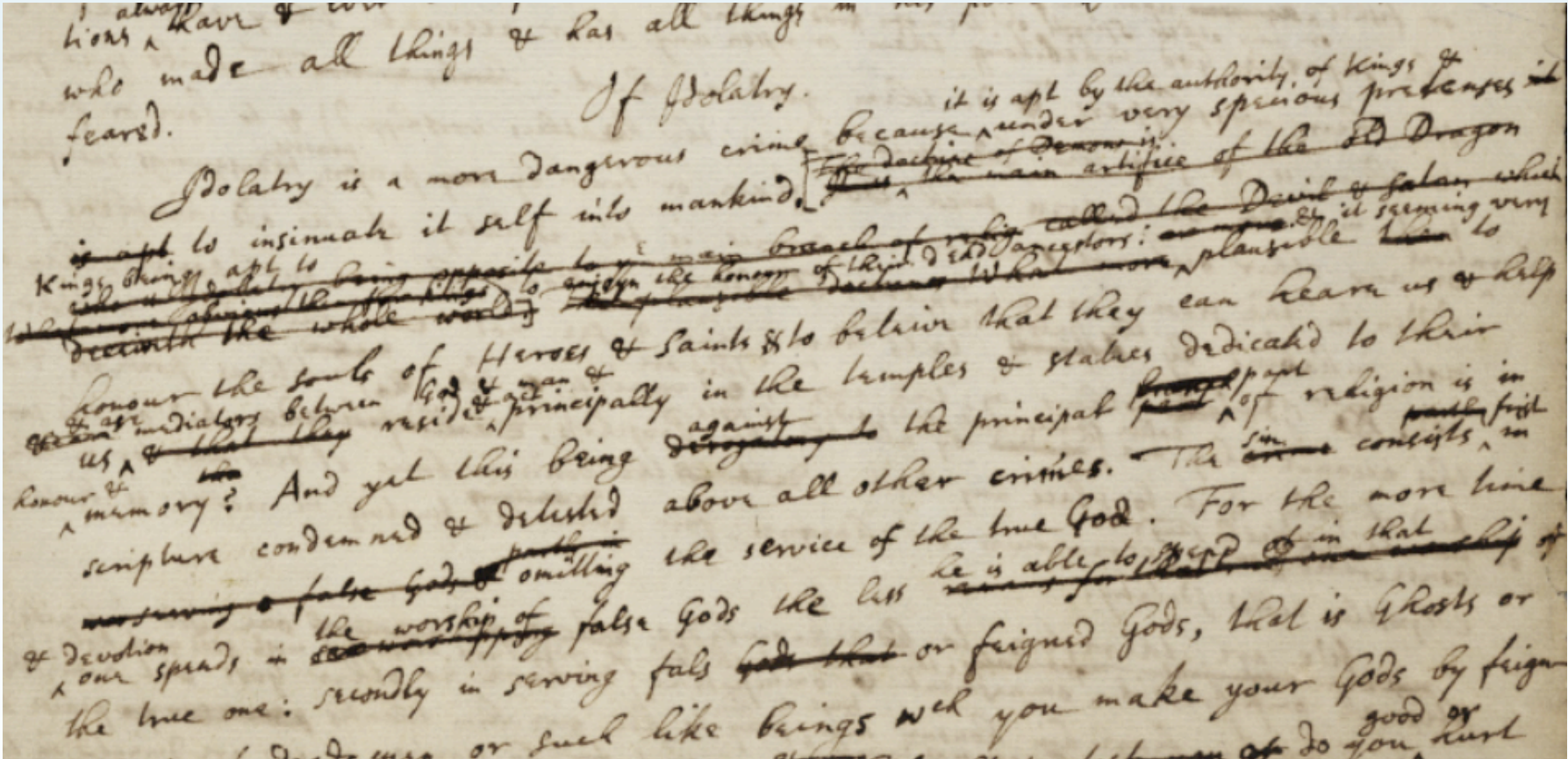- Wrote intensely about theological issues for much of his life
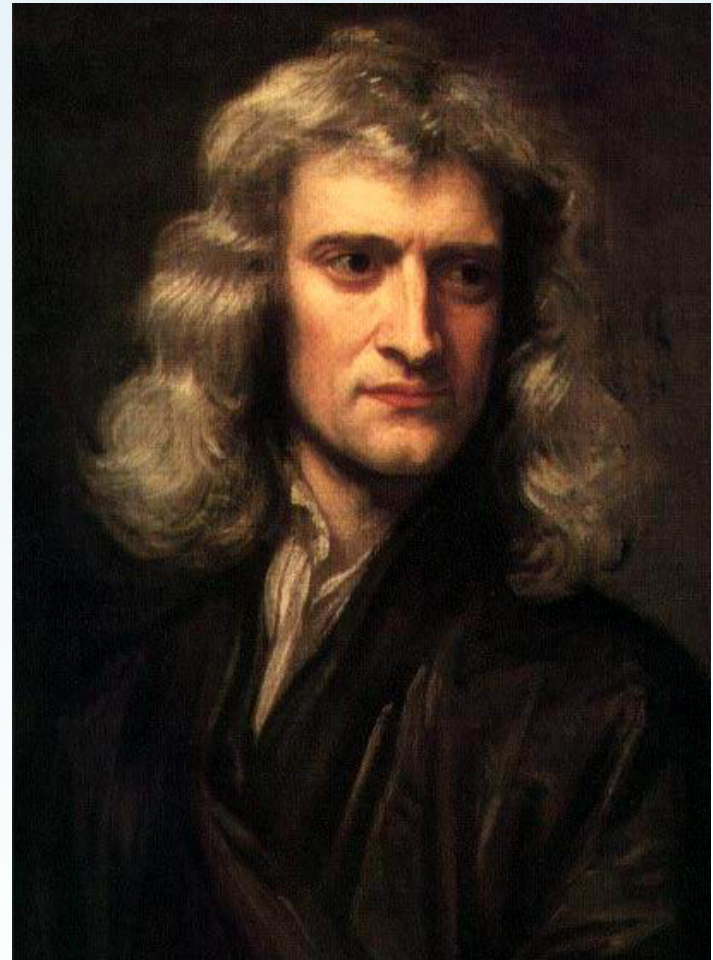
Godfrey Kneller portrait of Newton, 1689

The Newton Corpus

- Transcribed by the Newton Project, University of Sussex

- Consists of more than five million words on a range of topics including:
  - Alchemy
  - Mathematics
  - Physics
  - Theology
  - Religious history
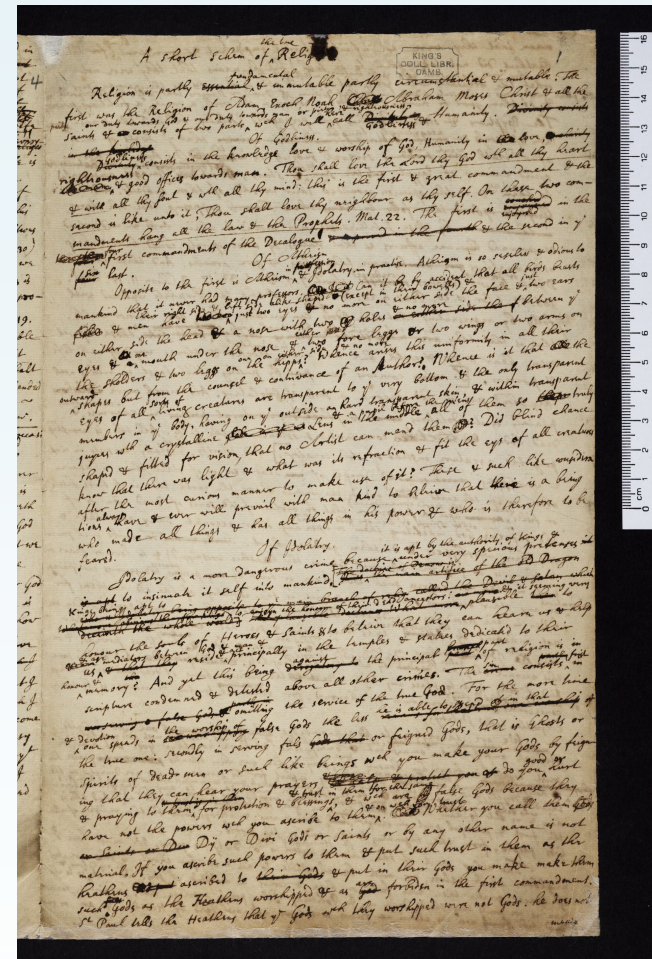
Difficulties with Newton

Some of the many; one document has more than 22,000 insertions and deletions

Structures of Documents

- Documents often "forked":
  - Started with bits of one document
  - Created later revisions using ambiguous cross-linking symbols

- Documents linked with other activities:
  - Later theological documents include mathematical notes on the revision to the *Principia*
  - Partial drafts of manuscripts also appear on the back, or even between the lines, of draft letters to Treasury Officers

Location of the Documents

- Individuals in the 18th century sifted through material to develop publishable manuscripts

- Large private collections were broken up in the 19th century

- Wide-spread sale of even small fragments across the entire globe in the 20th century

Dealing with Newton

- Given these difficulties, the corpus cannot be read like a book

- Visual analytics techniques required

Structure of a Visual Analytics System
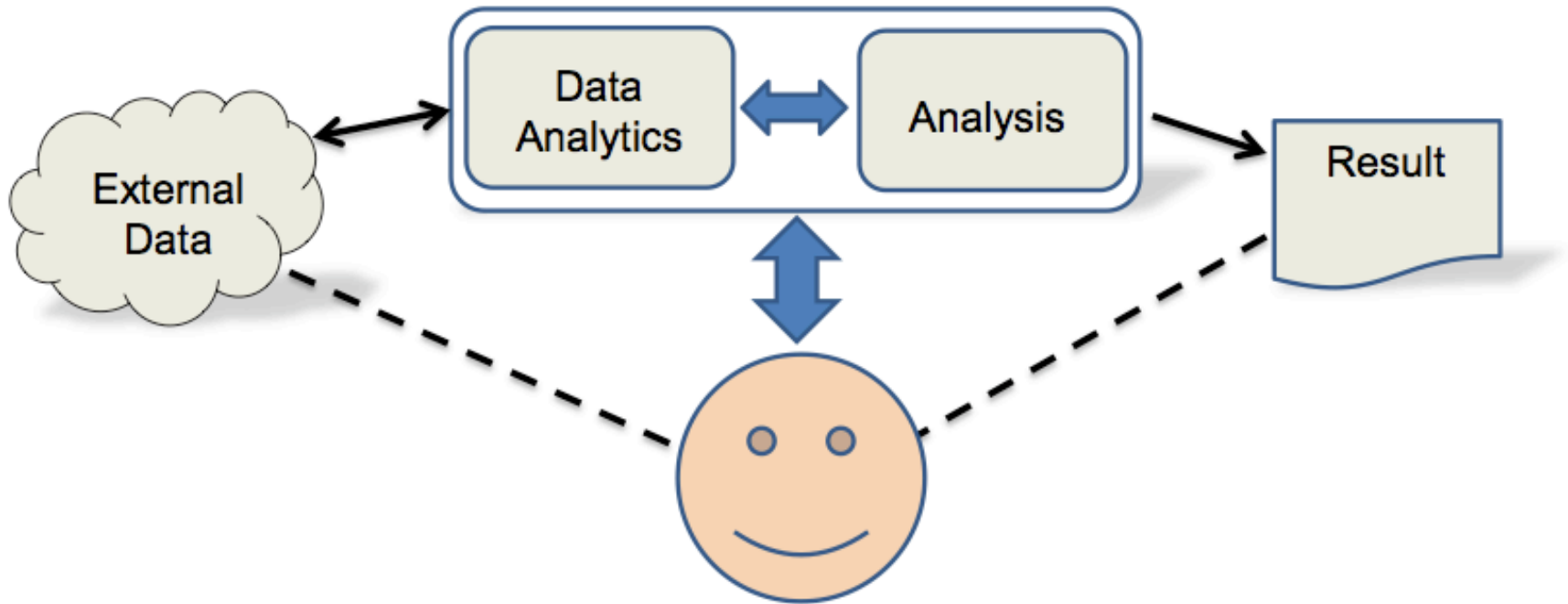
From "National Visualization and Analytics Center: Evolving to Meet World Needs", Richard May, VASW 2010

External Data Transcribed via XML

```
<body>
<div>
<pb xml:id="p001r" n="1r"/>
<head rend="center">Paradoxical Questions <lb type="intentional"/>concerning <choice><abbr>y<hi rend="superscript">e</hi></abbr><expan>the</expan></choice> morals &amp; actions of <lb/>Athanasius &amp; his followers</head>
<p><add place="interlinear" hand="#jc" indicator="yes">See this copy by Sir I. Newton ending with these <lb/>words ought to be corrected</add></p>
<div>
<head rend="center">- - Quæst 1. <lb type="intentional"/>Whether <choice><abbr>y<hi rend="superscript">t</hi></abbr><expan>that</expan></choice> ignominious death of Arius <supplied reason="damage">in a bog-house</supplied> <lb/>was not a story put about by Athanasius abo<supplied reason="damage">ve twenty years</supplied> <lb/>after his death?</head>
<p>How Arius died I reccon a question of <supplied reason="damage">no <del type="strikethrough">great</del> moment <add place="supralinear" indicator="yes">but</add></supplied> <lb/>because it <del type="strikethrough">leads in</del> <add place="supralinear" indicator="yes"><del type="strikethrough">gives light</del></add> <add place="supralinear" indicator="yes">leads</add> to other things of moment <supplied reason="damage">I chuse to</supplied> <lb/><del type="strikethrough">propound it</del> begin with it. <del type="strikethrough">This man</del> <add place="supralinear" indicator="yes">We are told in history <supplied reason="damage"><choice><abbr>y<hi rend="superscript">t</hi></abbr><expan>that</expan></choice> he</supplied></add> was <supplied reason="damage">excommunicated by</supplied> <lb/>the Council of Nice &amp; <del type="strikethrough">&amp; be <add place="supralinear" indicator="yes">being banished by</add> <lb/>upon <supplied reason="damage">recanting</supplied> was restored again into commu<supplied reason="damage">nion</supplied> <gap reason="damage" extent="unclear"/> <lb/>of Tyre &amp; then</del> <add place="supralinear" indicator="yes"><choice><sic>banishished</sic><corr>banished</corr></choice></add> by <supplied reason="damage">the Emperor Constantine <choice><abbr>y<hi rend="superscript">e</hi></abbr><expan>the</expan></choice>
```

Visualise Different Document Views

Diplomatic text

Normalized text

Visualise Keyword Linkages

"Newton Spectrum", Henry Florence, Mayooran Nambiarooran, Ahmed El-Abiary, Yosef Dalbah, Dani Abdallah, Shannon D'Souza

The Missing Link…
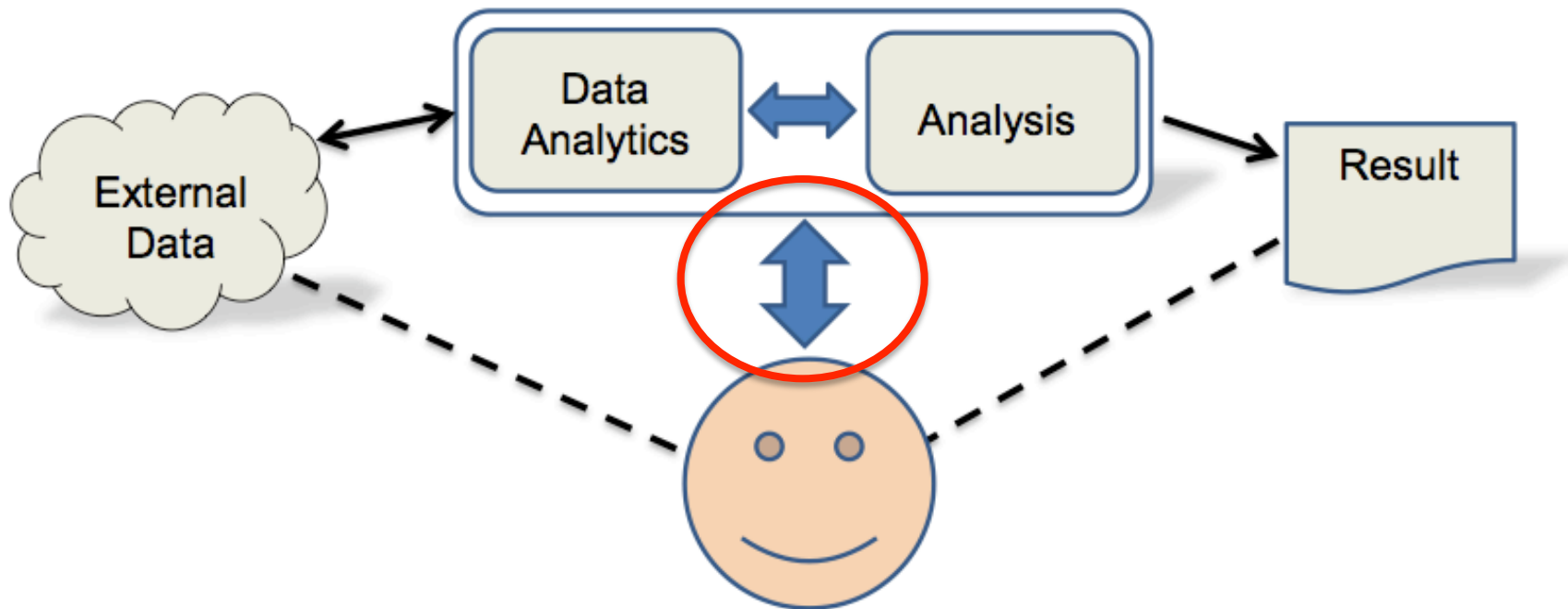
From "National Visualization and Analytics Center: Evolving to Meet World Needs", Richard May, VASW 2010

The Missing Link…

- Historical research means that the final, high-level arbiter of the quality of the inference is the human
- The system needs to support their mental model of activity
- These include:
  - Store parallel, competing hypotheses and means to browse between them
  - Store the assumptions inherent in creating those hypotheses
  - Provide methods to compare and contrast the differing network structures upon which they are based

Where Next?

- We are currently actively looking into two areas:

- Sensemaking:
  - Study of how people orient themselves in new environments

- Retroduction:
  - Study of how people develop novel hypotheses from surprising data

Summary

- Historical research involves looking for (often tenuous) relationships in mountains of data

- Most relationships have to be established by informed opinion

- Visual analysis tools can be used to help manage and display the information

- However, there are still huge challenges in how such systems interact with the mental models of end users