



CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2013

Week 5

Today's Class

- Digitisation
- Text Encoding

Discussion Analysis

Theban Mapping Project

DIGH 400 - Introduction to Digital Humanities Research

Online research exercise

Jules Verne and H.G.Wells

- find as many different editions of
 - Around the World in Eighty Days
 - A Journey to the Centre of the Earth
 - The Time Machine
 - The War of the Worlds
- record availability, copyright notice, PDF download option, original source for digitised copy...
- create a collaborative spreadsheet and share with group



Assessment

group project work (15%)

- group developed and maintained website
- promotes collaborative work (different tasks for different skills)
- maintain and develop web resource on digital humanities
- includes latest news, publications, projects and reviews, information on DH Centers and programmes...
- weekly blog to inform class of research, what's been done, needs to be done...

<http://www.digital-humanities.com>

Why do we bother with digitisation?

Benefits and Costs

- loss of analogue information to precise digitisation

"We should be cautious about letting the radiance of the bright future blind us to the limitations of this new technology"

(Smith, Abby. 1999. "Why Digitize?")

- dependent upon how much information you actually gather whilst digitising
 - density of data
 - frequency of sampling
- breadth or depth of information gathered is also a consideration

DIGH 400 - Introduction to Digital Humanities Research



Why do we bother with digitisation?

The ongoing saga of 'Analogue vs Digital'

- consideration of cost, time, transmission and storage
- loss of information from analogue to digital
- benefits of digital
- perceived benefits of analogue

"...reducing it to 0s and 1s is not unlike a root canal: by extracting the nerves, the tooth is killed in order to save it."

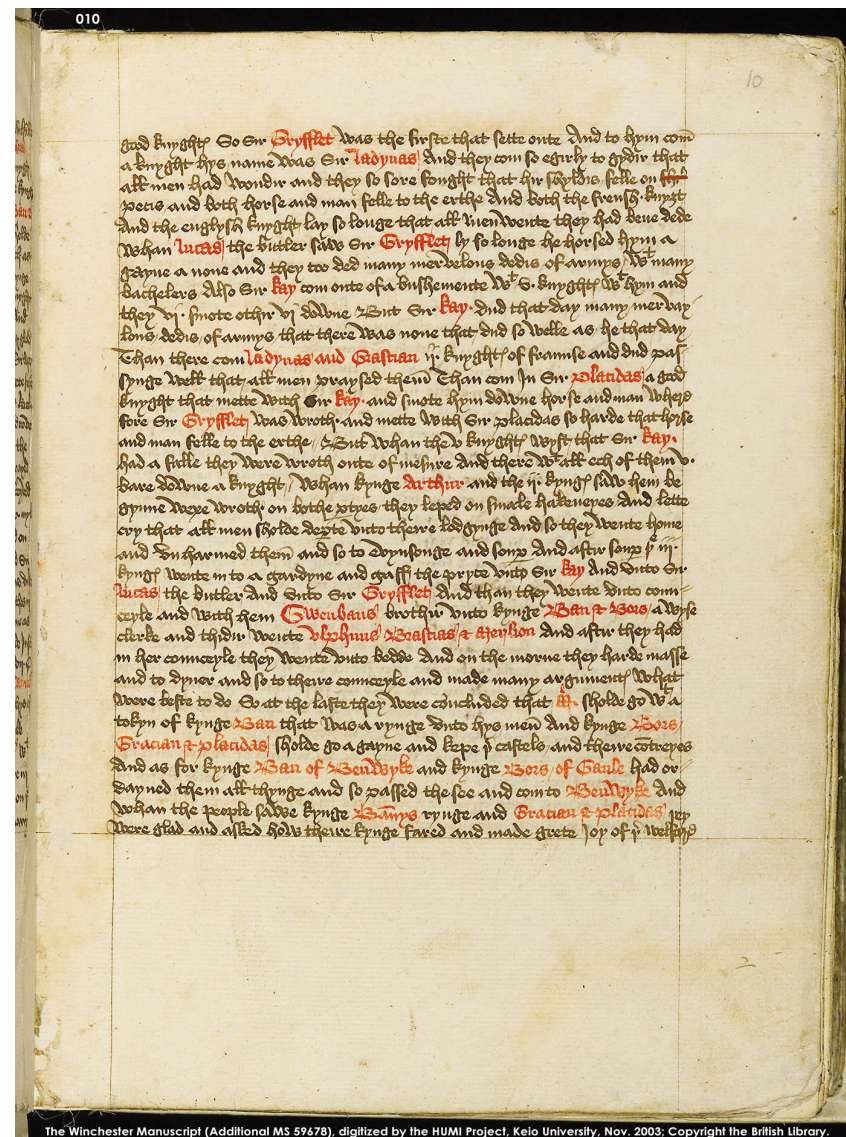
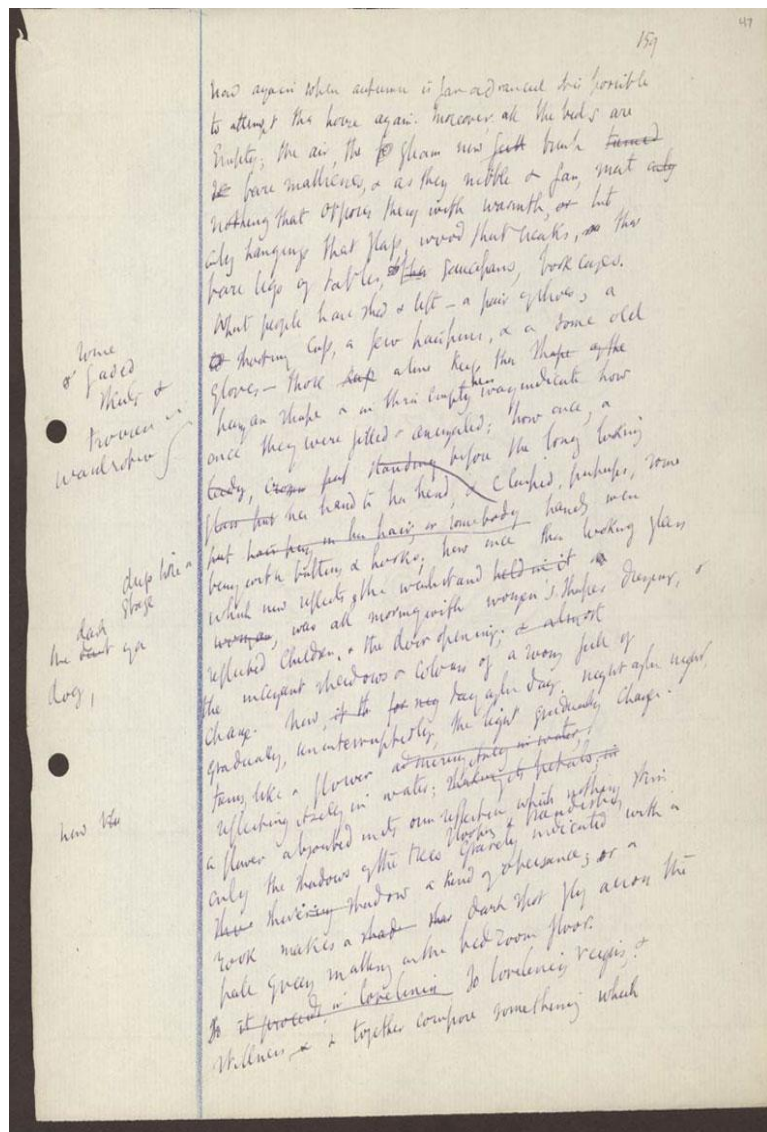
(Unknown. 2004. "The Great Analog Versus Digital Debate". VoicePrint Online)

Why do we bother with digitisation?

The ongoing saga of 'Analogue vs Digital' - Textual Considerations

- comparable concerns with music in textual digitisation
- 'density of data' is still a concern with text
- should we encode texts with style, punctuation, notes...
- how much do we encode, in particular with manuscripts
- considerations of discipline, project, and personal preferences
- variation in requirements dependent upon manuscript
eg: modern vs manuscript

DIGH 400 - Introduction to Digital Humanities Research



Why do we bother with digitisation?

- many advantages to the process
 - ease of access
 - preservation
 - increased opportunities for collaboration...
 - search benefits
- consider the many benefits of remote access to this material
- breadth of users
- universal, democratic access to material
- is it worthwhile digitising an underused archive or resource due to these potential benefits?

The act of digitisation

Digitising Text

- image of a printed page
 - scan or photograph a page of text
- advantages might include:
 - relatively quick and ease to produce
 - close representation of original
 - add tools for ease of manipulation
- disadvantages might include:
 - limited search options within image
 - file size is considerably larger compared to text
 - hardware considerations

The act of digitisation

Digitising Text - shall we use markup?

- another option for digitisation of textual material
- advantages such as complete machine readability
- markup may take many different forms
- all deal with the classification of components of a document
- more sophisticated capture structural and descriptive aspects

"explicit (to a machine) what is implicit (to a person)"

(Burnard, Lou. "Digital Texts with XML and the TEI". Text Encoding Initiative)

The act of digitisation

Digitising Text - conforming to a standard

- exciting opportunities are possible when we all conform to a standard
- document markup predates the internet and computers
- separation of content from format
- GML and SGML
- open and flexible but also complicated, time-consuming and expensive
- conforming to the concept of TEI