



CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2014

Week 5

Today's Class

- SciFi Authors
- Digital Humanities website
- Digitisation (Part 1)

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



[Scifi Authors website](#) | [Scifi Authors sheets](#)

DIGH 400 - Introduction to Digital Humanities Research

[Digital Humanities Website](#)

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."

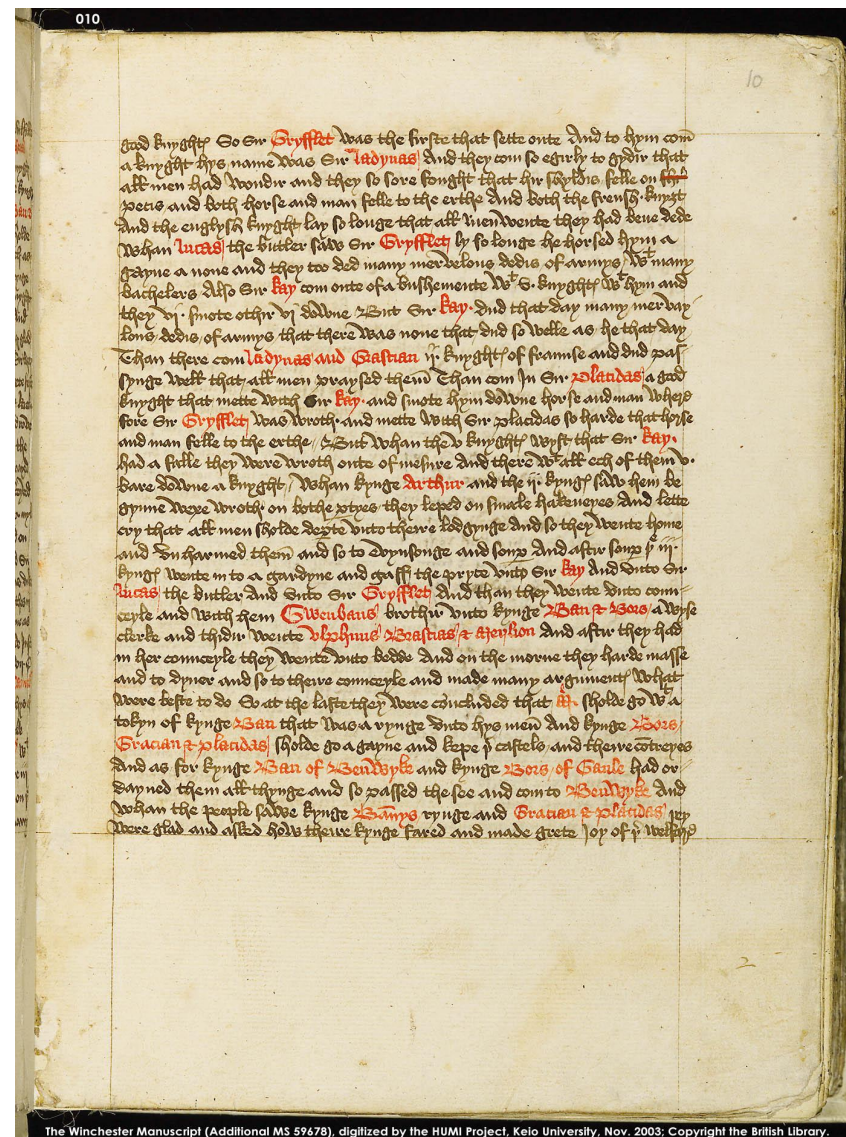
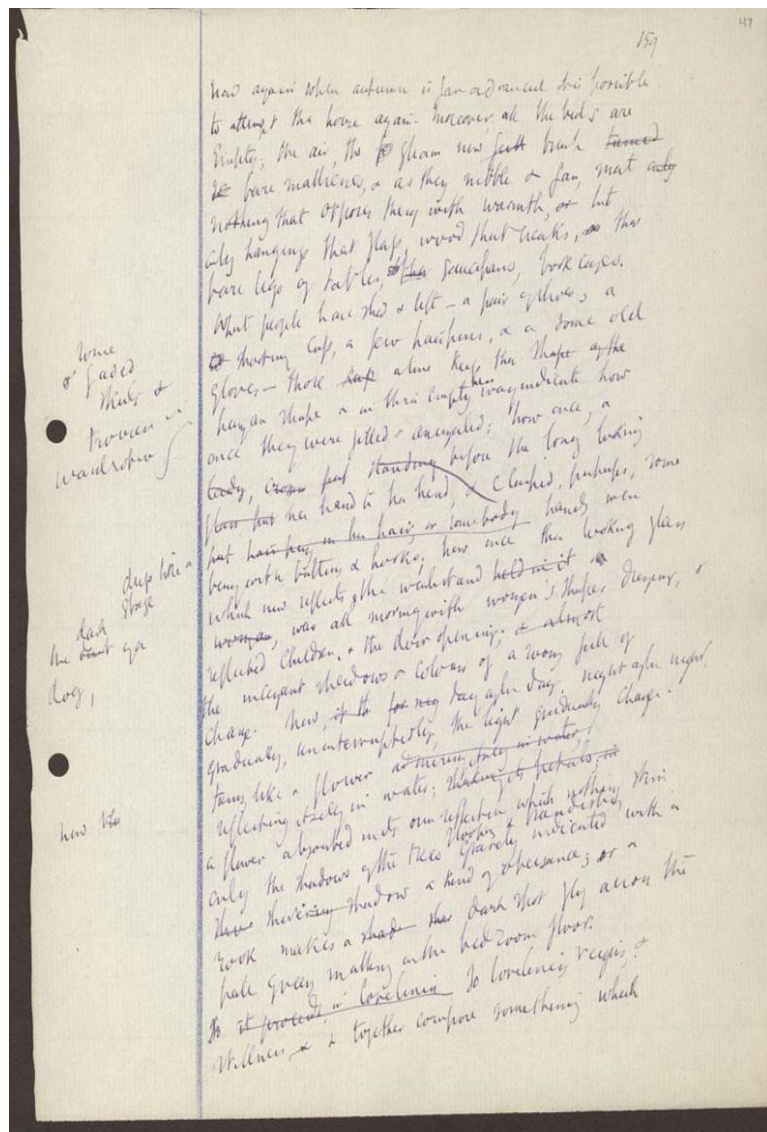
(Unknown. 2004. "The Great Analog Versus Digital Debate". VoicePrint Online)

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

The act of digitisation

Digitising Text - Options for making your own digital text

- import, automate, or type your own
- scan or photograph analogue to digital
- minimum 300dpi scan or better 600dpi
- book scanners and cradles are also available
- OCR texts
- OCR limitations include poor performance for
 - non-Latin characters
 - small print
 - certain fonts
 - complex page layouts or tables
 - mathematical or chemical symbols
 - most texts pre-19th century

The act of digitisation

Digital Images

- again we can use scanners and cameras
- more specialised material with cradle and book scanners
- consider fidelity to the original and long term preservation
- quality of digital result dependent upon
 - quality of original
 - digitising method employed
 - sample rate of digital to analogue
- minimum 300 DPI, preferably 600 DPI
- uncompressed format, such as TIFF
- optimise to JPEG for online publication

The act of digitisation

Digital Images - metadata

- metadata added to images
- in particular important when the files are not text searchable
- standards such as [Dublin Core](#) or [METS](#)
- technical information about the file such as
 - resolution
 - compression (if applicable)
 - scanning process
 - copyright and access rights
 - authorship of the image
 - and much more...

god knyght So **Er Gryffet** was the firste that sette oute And to hym com
 a knyght his name was **Er Ladynas** And they com so egerly to godir that
 all men had wondir and they so sore fought that the knyghts felle on the
 peire and both horse and man felle to the erthe And both the knyghts dyed
 And the knyghts lay so longe that all men wente they had bene dede
 Whan **Lucas** the butler sawe **Er Gryffet** by so longe he horsed hym a
 gayne a none and they too ded many mervelous dedes of armys And many
 bachelers Also **Er Ray** com oute of a bushemente And **Er Gryffet** was hym and
 they dy. smote othir by dolne But **Er Ray** did that day many mer-
 velous dedes of armys that there was none that did so well as he that day
 Whan there com **Ladynas** and **Gastrian** the knyght of fransse and did pas-
 sange well that all men prayd thend Whan com in **Er plandras** a god
 knyght that mette both **Er Ray** and smote hym dolne horse and man where
 fore **Er Gryffet** was wroth and mette both **Er plandras** so harde that horse
 and man felle to the erthe But whan the knyghts dyed that **Er Ray**
 had a felle they were wroth oute of meschance and there was all ech of them
 have dolne a knyght Whan kynge **Arthur** and the knyghts sawe hem be-
 gyne they were wroth on both parties they leped on smale hatenyes and sette
 cry that all men sholde dreynt unto thers to dreynt and so they wente home
 and engharmed them and so to dounsonge and song And after song the
 knyght wente in to a gardyne and cast the pryte unto **Er Ray** and unto **Er**
Lucas the butler and unto **Er Gryffet** And than they wente into com-
 cepte and with hem **Gwenhau** brother unto kynge **Ban** and **Bois** a wyse
 clerk and thidre wente **Blodynne** **Er Gastrian** and **Er Arthur** And after they had
 in her conceyte they wente into bedde And on the morne they had masse
 and to dner and so to thers conceyte and made many argument What
 were best to do At the laste they were concluded that they sholde go to a
 tobyn of kynge **Ban** that was a vnyge unto the men And kynge **Bois**
Er Gastrian and **Er plandras** sholde go a gayne and kepe the castels and thers cotreys
 And as for kynge **Ban** of **Beutysse** and kynge **Bois** of **Gaulle** had or-
 dained them all thynge and so passed the see and com to **Beutysse** And
 whan the people sawe kynge **Ban** vnyge and **Er Gastrian** and **Er plandras** they
 were glad and asked how thers kynge fared and made grete joy of the knyghts

DIGH 400 - Introduction to Digital Humanities Research

