

# Manuscripts Online

Friday 18th November 2011  
Humanities Research Institute

Present: Mike Pidd, Kathy Rogers, Sharon Howard, Estelle Stubbs, Orietta Da Rold, Wendy Scase, John Thompson, Linne Mooney, Alastair Dunning  
Apologies: Jeremy Smith

## 1. Collaboration Agreements (partner agreements)

First draft of this is in progress at Sheffield.

Need from members of Editorial board (EB)

- guidance on costs (travel costs, workshops, conference) from partners' institutional funding
- contact information for each institution (have Birmingham and Leicester so far)

## 2. Project Plan, blog and key JISC documentation

Draft project plan is on Google Docs (needs to be updated by MP)

WS raised issue with use of term 'primary sources' as most of the sources aren't (i.e. Wendy drew the distinction between primary sources as real, physical objects that are evidence for the past as opposed to secondary sources that are studies of the primary sources). The project will use the term 'primary resources' to refer to its electronic data.

OR / EB to review Dissemination section in particular

Use blog/twitter for early publicity and recruitment of testers, as well as drawing on Quadrivium  
Peter Stokes (DigiPal) has been in contact with OR expressing interest

Blog has been set up - SH/MP to organise domain name

Blog should also be used for making public internal project documentation: eg. the project plan and minutes of meetings (edit out anything private)

OR and MP to discuss timetable for blog posts

## 3. Overview of the workflow

4. Q1: - What should be the scope of Manuscripts Online in terms of the data to be included?

5. Q2: - Do we include data which is in Middle English only, or do we include texts in Latin, French etc?

6. Q3: - What will the NLP process identify?

a. problem of non-Latin/non-standard characters and extent to which we try to standardise (distinction here between NLP analysis and representation in public website)

Lack of consistency between different sources -> will need to review samples of data by resource to make decisions

- use GDocs spreadsheet for recording information about non-standard characters, with a view to developing a translation table.

- importance of documenting for users the decisions that are made

- process to be blogged

b. NLP and other languages

- extent to which we can identify names/places/dates in the data

- what other languages will we try to identify?: Latin, Anglo Norman etc. Need to establish a definitive set of languages.

- Substantial proportion of the data is Latin

- We aim to at least start by identifying what is non-English text

- OR to meet with MP and KR to go over the detailed groundwork for this

i. Will need dictionaries/wordlists - OR/EB to identify online/digital resources

ii. files with mixed language as samples [OR has transcriptions marked up for different languages]

iii. information about word endings/language structure

Start with Latin as an exercise for other languages

Don't need to assume what language any dataset will be, but do need to know language structures to be able to identify them.

Further resources needed - thesauruses, technical terms (eg for illuminations)

Online codicological dictionary

Possible use of activity data from user searches - HRI to discuss whether this would be feasible

Geographical search (in addition to simple places search) whereby a search for, say, Worcester will also return hits relating to villages in the vicinity of Worcester.

- should be technically possible, though possible performance issues - HRI to look at feasibility

- how it works for users - will need to be included in user testing

- Sources for developing geographical base data, particularly the relationship of, say, villages to larger organisational entities: VCH, OS and Taxatio

## 7. Search engine specification

Questions/issues arising from flowchart

### Keyword w/ variants search

Which languages apart from English? Will need dictionaries  
Ideal would be to switch on multiple dictionaries for comparison  
There may be performance issues - we can only try it and see  
Will need to provide warnings to users  
Consider carefully what the defaults will be

OR/EB to identify suitable variant dictionaries  
Latin - need both classical and medieval  
Anglo-Norman (David Trotter)

### Search filters

MP - document on BaseCamp for discussion

## 8. Visual design and user focus groups

To be tendered to a design agency  
- MP will shortlist 3 names to be approved by EB and draw up a specification of the work

User focus groups

- The Quadrivium workshop is in February, which may be a bit late for this purpose
- 1st user group in Birmingham, ideally January - OR/WS to firm up on dates
- 2nd user group in Sheffield during November 2012
- Need a range of users, reflecting desired users of final site: ug/pg/postdoc; different fields (lang/lit/etc); librarians; local history groups

## 9. Datasets and content providers

Importance of getting contracts moving early in order to get data

High priority to contact Paul Schaffner (Michigan) re MED, Compendium and EEBO-TCP (SH)  
Other non-partner sources: Europa Inventa, METS

SH to contact members of EB re their own data and MTAs  
MP to send SH contacts/correspondence for other providers

## **10. Material Transfer Agreements (content licences)**

Template agreement has been drafted.  
Proposed that Sheffield will take legal responsibility  
WS/OR agreed with this, especially as the EB might change composition in the future

## **11. Resource Descriptions**

MP will implement these shortly.  
EB - fill in as soon as possible, especially data bundle 1.

## **12. Stakeholder panel**

MP to invite:

Tim Hitchcock  
Jane Winters .  
Caroline Palmer.  
Ian Johnson  
Aleksandr Drozdov  
Kathleen Doyle  
Andrew Prescott  
Toby Burrows  
Rob Newman (ProQuest)  
Adam Farquhar (BL Head of Digital Library)  
Elmar Mittler (Europa Regia)  
Peter Ainsworth  
Alastair Dunning (JISC)

## **13. Quadrivium Workshops**

Some additional funding at Glasgow (check with JS) - extra workshops may be possible

17 February '12 - Glasgow  
January 2012 (specifically user testing) - Birmingham  
Early November 2012 (specifically user testing) - Sheffield

## **14. Dissemination Plan**

MP to start whiteboard on Basecamp for sharing ideas  
- conference presentations, digital humanities as well as medieval

One day conference - Leicester, late January '13

## **15. International data linking opportunities**

Not covered within the duration of the meeting, although there was a general consensus to explore data linking opportunities with initiatives such as Europa Regia

## **16. Subsequent Meetings**

SH will do a Doodle Poll to organise future meetings.  
Everyone agreed to hold all future meetings in Sheffield.