New Methodologies
for effective exploitation of
Digital Manuscript Corpora

Wendy Scase

W.L.Scase@bham.ac.uk

Existing Digital Resources

- Images

- Calendars

- Catalogues

- Descriptions

- Transcriptions

Methodological Challenges (1)

- **Creating linked** and **interoperable corpora** and new tools to enable new uses for digital MS resources
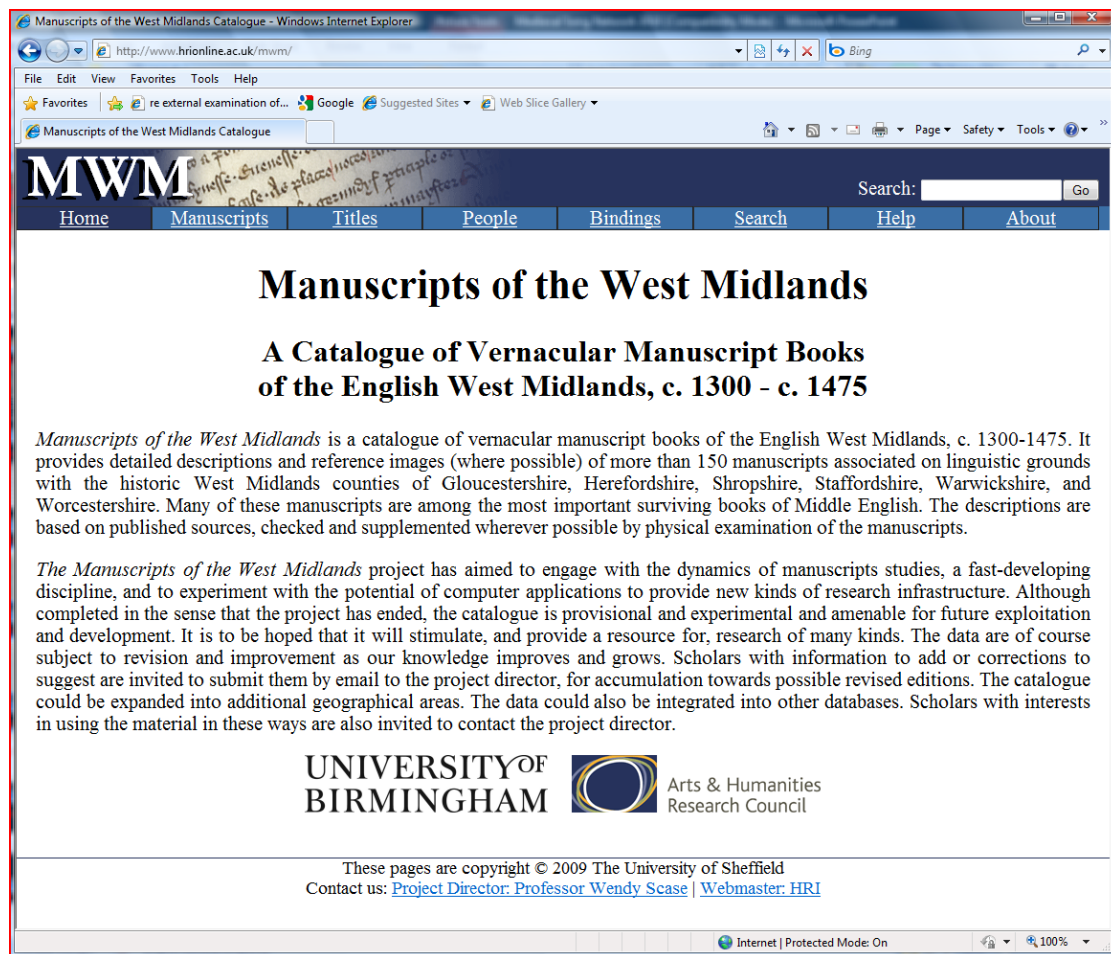
Problem discussed at ESF Exploratory Workshop in Birmingham 2009 'Applying Semantic Web Technologies to Medieval Manuscript Research'

Challenges (2)

- **Methods for analysis** of digital corpora

Case Study of two Corpora:
Manuscripts of the West Midlands of England

MWM - The Manuscripts of the West Midlands Project

Manuscripts of the West Midlands

A Catalogue of Vernacular Manuscript Books of the English West Midlands, c. 1300 - c. 1475

Manuscripts of the West Midlands is a catalogue of vernacular manuscript books of the English West Midlands, c. 1300-1475. It provides detailed descriptions and reference images (where possible) of more than 150 manuscripts associated on linguistic grounds with the historic West Midlands counties of Gloucestershire, Herefordshire, Shropshire, Staffordshire, Warwickshire, and Worcestershire. Many of these manuscripts are among the most important surviving books of Middle English. The descriptions are based on published sources, checked and supplemented wherever possible by physical examination of the manuscripts.

The Manuscripts of the West Midlands project has aimed to engage with the dynamics of manuscripts studies, a fast-developing discipline, and to experiment with the potential of computer applications to provide new kinds of research infrastructure. Although completed in the sense that the project has ended, the catalogue is provisional and experimental and amenable for future exploitation and development. It is to be hoped that it will stimulate, and provide a resource for, research of many kinds. The data are of course subject to revision and improvement as our knowledge improves and grows. Scholars with information to add or corrections to suggest are invited to submit them by email to the project director, for accumulation towards possible revised editions. The catalogue could be expanded into additional geographical areas. The data could also be integrated into other databases. Scholars with interests in using the material in these ways are also invited to contact the project director.

UNIVERSITY OF BIRMINGHAM   Arts & Humanities Research Council

These pages are copyright © 2009 The University of Sheffield
Contact us: Project Director: Professor Wendy Scase | Webmaster: HRI

A corpus of metadata and images of MSS in West Midlands English.

Project aims: to resource study of late medieval vernacular MS production.

www.hrionline.ac.uk/mwm/

MWM: description of St John's College Cambridge MS B.12

Cambridge, St. John's College, 34

| **Shelfmark** | **Date and Language** | **Contents** |
| **Physical Description** | **History** | **Record History** | **Bibliography** |

Shelfmark
Country: England
Settlement: Cambridge
Repository: St. John's College
Idno: 34
AltName: B.12

Back to Top

Date and Language
Date: s. xvin
Language: English
Dialect: Scribal Dialect: Herefordshire. Linguistic Atlas Grid Reference: 361 244, LP 7450 (McIntosh, Samuels, and Benskin 1986, p. 199).

Back to Top

Contents
An early fifteenth-century copy of the *Confessio Amantis*.

Item: 4
Locus: The manuscript has not been foliated.

MWM: description of St John's College Cambridge MS B.12 (2)

<u>Physical Description</u>

Form: Codex

Support: Parchment

Extent: 310 x 230 mm

Collation: 1-27 8 with catchwords. Quire 12 wants 7 (stub remains, initial letters visible). Quire 14 wants 1 (stub remains, initial letters visible).

Layout: Pricking of small rounded holes in the header (10 mm from the edge), footer (21 mm from the edge) and external margin (15 mm from the side and 22 mm from the top edge) to indicate the framing space. Writing space: double columns: c. 82 x 23 mm with 39 lines.

Writing: One scribe writing in a neat Anglicana Formata hand. In the Latin summaries the hand becomes smaller and more angular, whilst retaining the same degree of cursiveness. Each line terminates with a square-shaped full stop. Double compartment *a*; diamond-shaped lobed *d* with square ascender curving 90 degrees from the left to the right, ending in an otiose hairline; 8-shaped *g* with rounded head, the descender closing into a rounded lobe, finishing in a squarish point attached to the head. This does not extend below the level of writing. There is a small tail on the right of the head; straight descender with a tail on the top left side of the *p*, rounded head with shaft starting from the right towards the left, finishing just past the meeting point with the descender; long and short *r* are used both in medial and final position; long *r* in conjunction with 'er', short *r* with 'or'; long and sigma *s* used regularly; no rounded *s*; dotted þ; the stem of þ is slightly bent towards the left by 45 degrees, ending with a small tail, the shaft does not extend past the stem/descender, it is used regularly in initial and final position; open *w* with the first stem terminating in a small tail and the second back stroke extending above the level of the letter turning towards right with a sharp angle and terminating with a tail. Two closed lobes finish the letter on the right side. Capital letters are decorated with double lines across the ascenders, otiose dots in the lobes of letters. Body height: 3 mm in the English text; 2 mm in the Latin.

Decoration: The first page has a full border decorated in blue, red and gold with leaf motifs prominent on the external margin. First letter of the first word of the prologue is a two-line capital in gold, with red and blue background. No other decoration except for two-line Lombardic capitals in blue with red penwork decoration to indicate the beginning of each section. Initial letters of each line are tinted in yellow. Other Litterae Notabiliores written in blue or red, no paraph marks.

Metadata + Images
(metadata are enhanced with images but licensing images a problem)
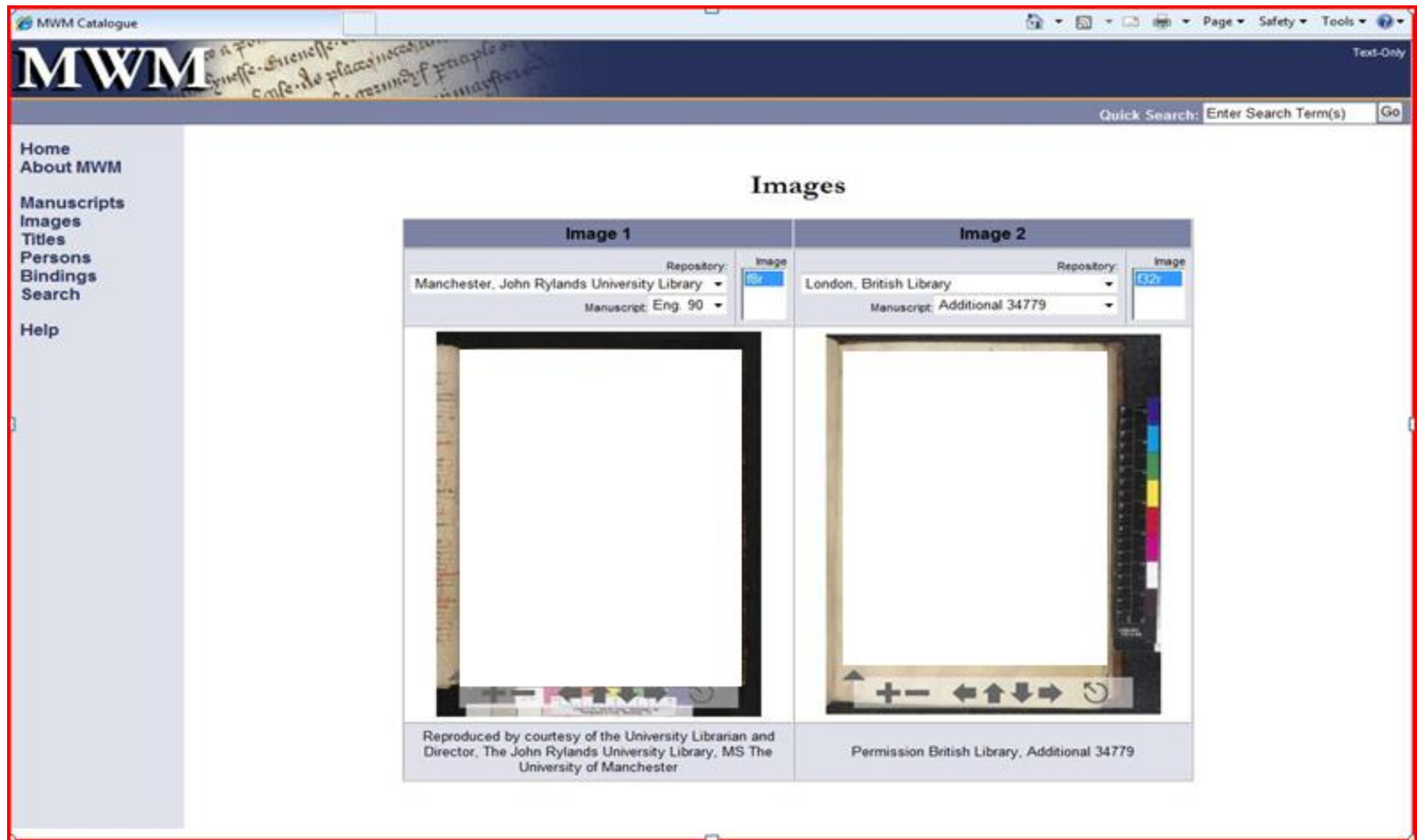
2 Midlands *Prick of Conscience* MSS

The Vernon Manuscript Project

Oxford, Bodleian Library, MS Eng. Poet. a .1, c. 1400, in English West Midlands dialect

A FACSIMILE EDITION OF
THE VERNON MANUSCRIPT

OXFORD, BODLEIAN LIBRARY
MS. ENG. POET. A.1

Editor Wendy Scase
Software Nick Kennedy

START
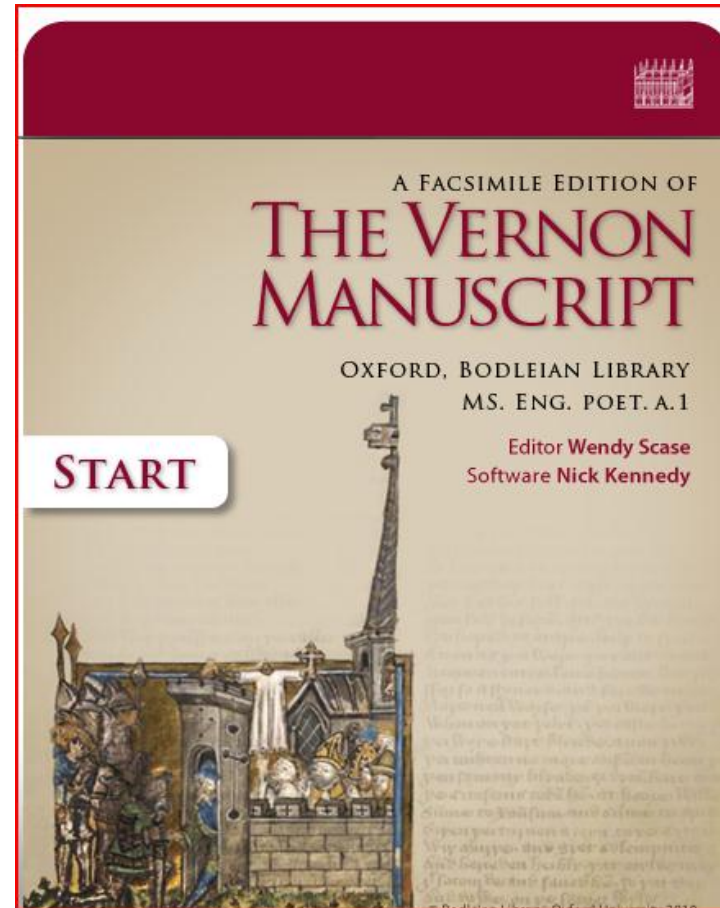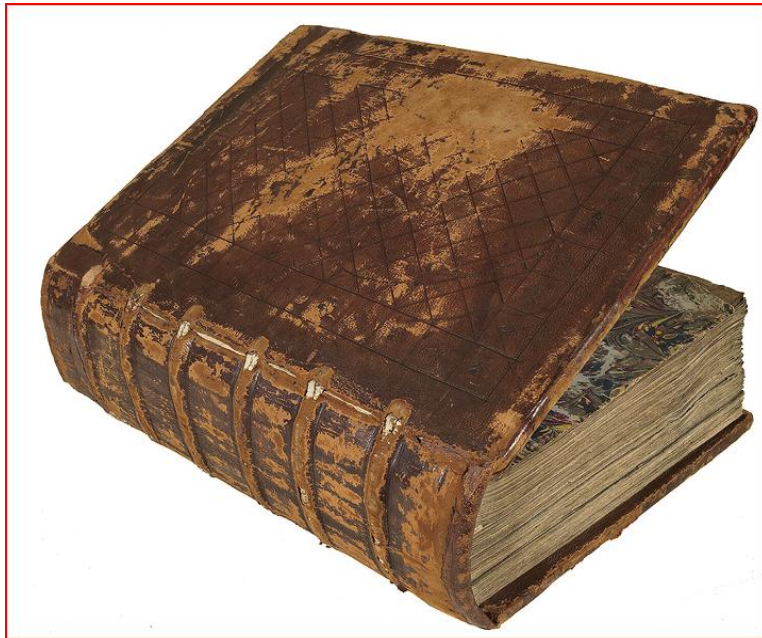
www.birmingham.ac.uk/vernonmanuscript

The Vernon Manuscript Project has received support from Arts & Humanities Research Council

A Medieval 'Data Deluge'

350 folios (69 folios lost)

c. 370 different items – all devotional and religious, prose and
verse, almost all in Middle English

Word count: 2.5 x Tolstoy's *War and Peace*

16 Miniatures (c. 52 lost)

Digitising the Data Deluge: Aims

Metadata + images (images alone are of limited use)

An accurate text to parallel the images in the edition

A corpus of linked data exploitable for research

Policies on Data Capture

Working with digital files, we could record:

Changes of ink

Sharpening of pen

Letter forms

Marks of abbreviation

Policies on Data Capture (2)

Considerations:

Equivalent to transcribing *War and Peace* 2.5 times

Three years to do it

Over-riding aims are accuracy and consistency

A team effort

Policies on Data Capture (3)

Decisions:

Resist interpretation; don't embed assumptions

Have simple rules and stick to them (even if the result looks odd, for example capitalisation and word division)

Record data in format that can be re-used, added to, developed (in xml)

Allocation of data to Transcription and Description files –not strict TEI

Transcription Files: Scribal (?) Corrections

'Samul' corrected with
   superscript 'e' and caret
   mark

(fol. 397v col. a)
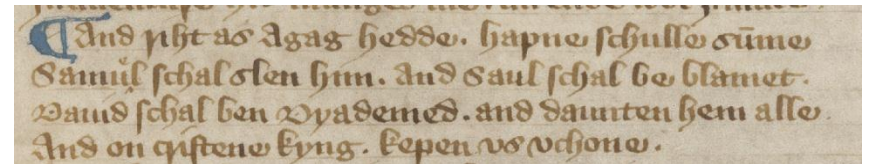
Page Description File: Guide Words and Letters

faint text for rubricator to insert 'Primus passus de dowel' (fol. 400v, col. B)
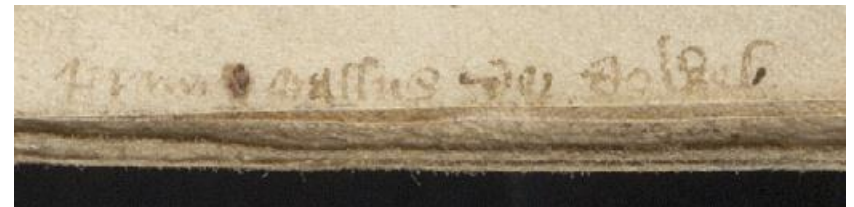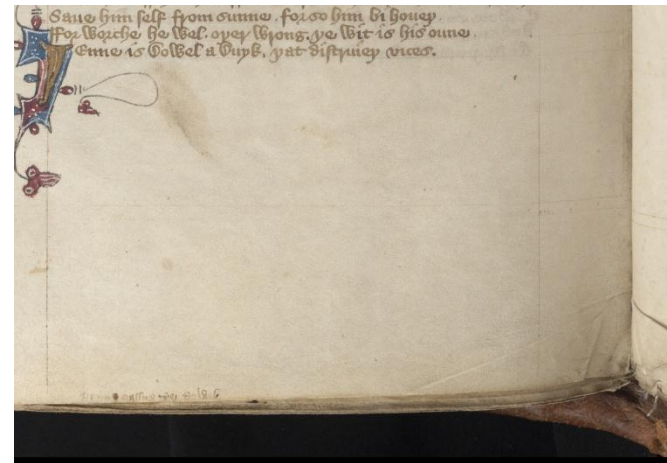
Page Description File: Assembly

catchword to ensure correct ordering of quires

below, 'hic sequitur vnus quaternus et vltimus' – note by supervisor (?) to binder (?)

(bottom edge guide text for rubric)
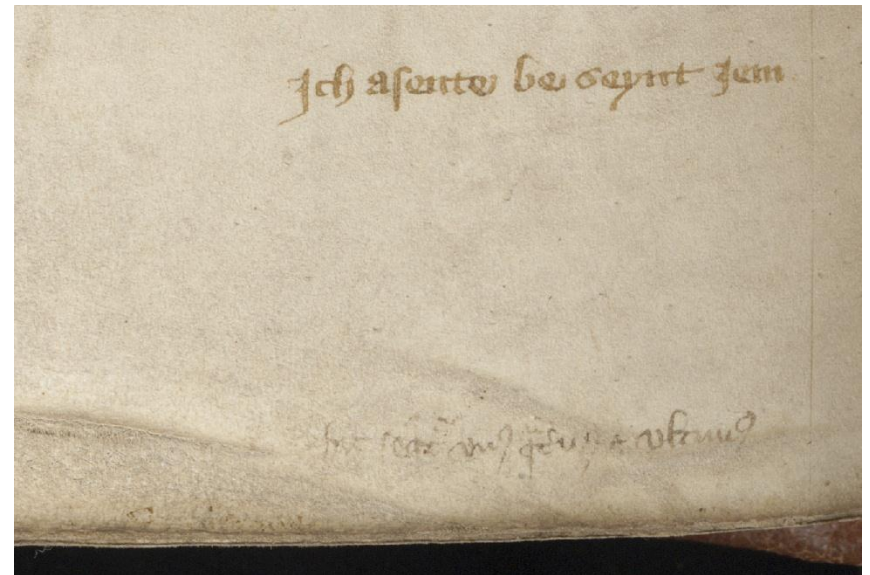
Vre lord and seint Ypolyt . þis hors saue and loke.
Now bidde we seint Ypolyt . þat he vre erende beode.
Þat God haue merci of vs alle . and of alle þat habbeþ neode. AMEN.

SEinte Marie godes Moder . fro þe Apostles was heo nouht.
Þo þe holigost on witsonenday . among hem was i brouht.
Er he was in hire aliȝt . ac noþeles heo was þere.
Heo ne dorste for þe giwes . from þe Apostles for fere.
Ac siþen þo þe Apostles wende . euerich in his syde.
To preche aboute cristendam . in þe lond wel wyde.
Vr ladi was feble and old . heo mihte hem suwe nouht.
Heo Mournede ek for hire sone . on him was al hire þouht.
Bi syde þe hul of Syon. in to an hous heo wende.
In hire beodes heo wonede þere . til vr lord aftur hire sende
Þeo Aposteles come ofte hire to . hire cumfort to make.
And nomeliche seint Ion þe Ewangelist . for heo was him bi take.
Heo þouhte so muche on hire sone . þat whon heo mihte i seo.
Eny stude and þer neih come . þat he hedde onne i beo.
Hire þouȝte heo was in loye i nouh . þerfore heo wende ofte.
In studes as he hedde ibeo . priueliche and softe.
As Men doþ ȝit in pilgrimage . in gladnesse of heore mode
In þe stude þat he i Baptiȝed was . and i do on þe Roode
In þe stude þat he i buried was . and to heuene wende.
Whon heo to eny of þeose come . hire herte bi gon to amende
Neih seke men ofte heo was . to cunforte hem of sore.
Hire folewede vche goodnesse . for heo was ful of milce and ore.
Such lyf aftur hire sone deþ . twelue ȝeer heo ladde neih.
And euere heo was in gret mournynge . þat heo hire sone ne seih.
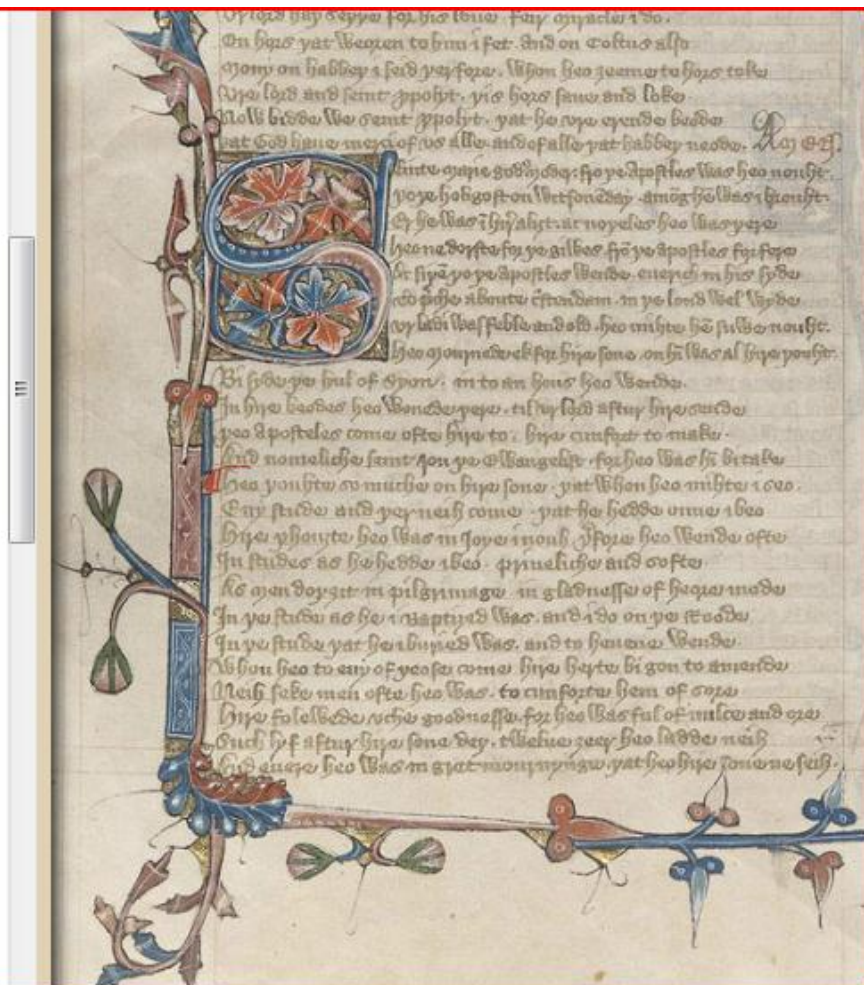
Using the Vernon files for palaeographical research

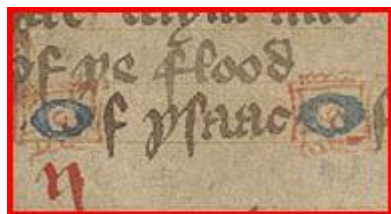Palaeographical analysis distinguishes two scribes, A & B

Linguistic analysis locates B in N Worcs and A in SW Worcs
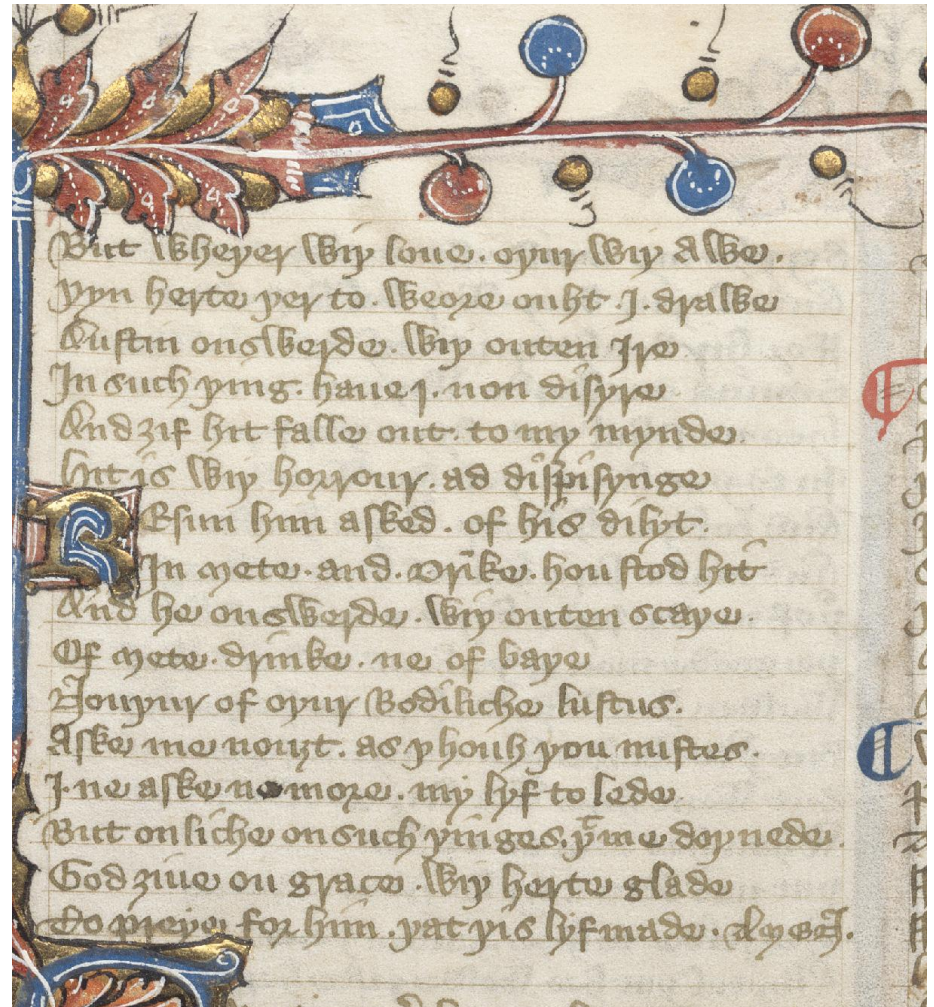
Qu. 1: **Did Scribe A Decorate his Text?**
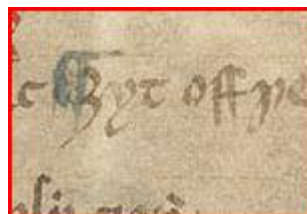
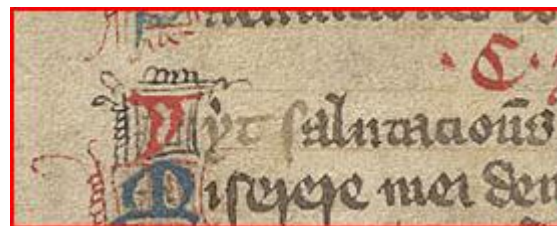Scribe A and Flourisher form <Y> and <Þ> (thorn) indistinctly

Scribe A	Flourisher	Scribe A	Flourisher

But Scribe A uses <Y> only for /i/ (e.g. *Ysaac*)

For /j/ (471 instances) Scribe A uses <ȝ> (yogh) (e.g. *ȝyt* = yet)

Whereas for /j/ Flourisher also uses <Y> (e.g. *Yyt* = yet)

Answer: No!

Qu. 2: Provenance of Flourisher?

Scribe A is located in West Midlands on linguistic grounds

Flourisher's graphemic profile (<Y> = <Þ>) is associated with areas *outside* the West Midlands

Further Research Possibilities

- Could also use transcription + description + images to investigate, for example, what constrains scribe's variations in his practice – e.g. why does he abbreviate here and not there? Why this form of a letter here and that there?

- Creation of and linking to new corpora e.g. Simeon MS (BL Addit. 22283) would support further research along these lines

Next: Enlarging the Corpus
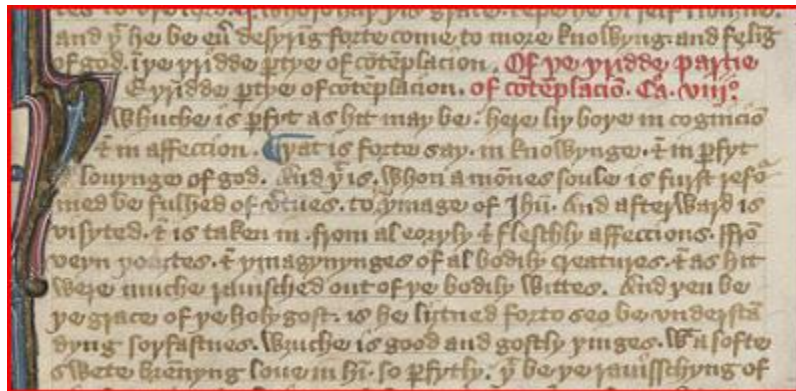
Vernon scribe B's *Scale of Perfection* in Vernon and Simeon (fol. 152r continued by Simeon Scribe 3)
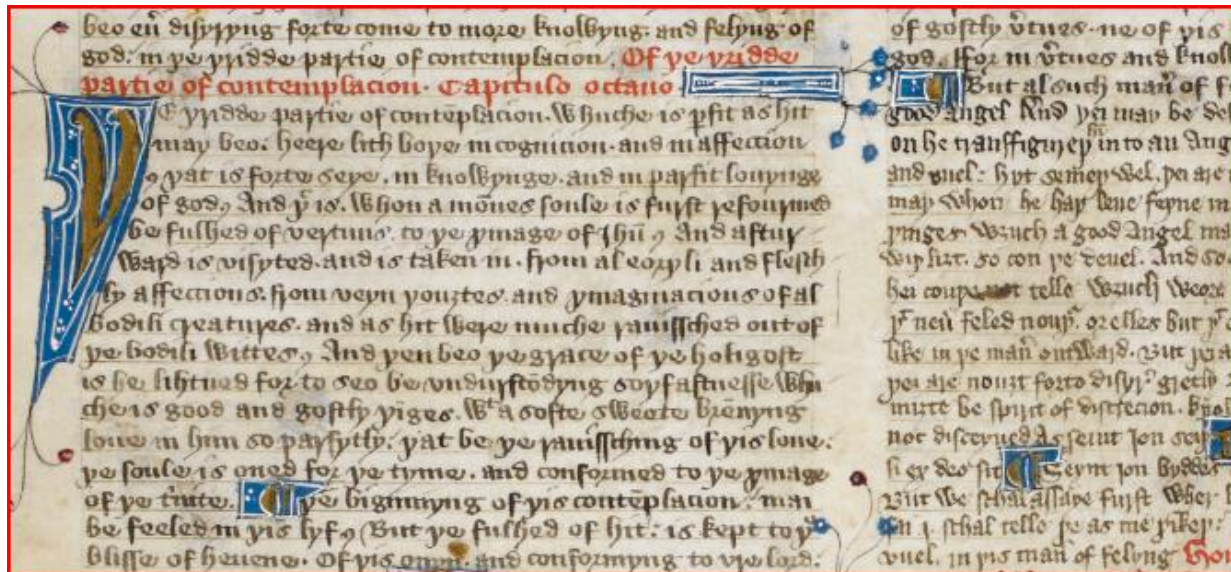
Exploiting Corpora to Enhance Palaeographical Research

Vernon and MWM projects offer examples of how

- **linking corpora of different genres within and across resources**

- **and analysis through cross-referencing of palaeographical methods with other methodologies (e.g. linguistic profiling)**

can lead to insights