

Humanities Collections and Reference Resources
National Endowment for the Humanities.

Year 1 Interim Report

Period covered: 1 May 2016-30 April 2017

Grant number: PW-234705-16

Grant title: The Middle English Compendium Reborn: Expanding Content, Improving Access, and Stabilizing Technology

Project Director: Paul Schaffner

Project institution: University of Michigan, Ann Arbor

Date: 24 May 2017

Narrative

Overview

The purpose of this grant, and this project, is threefold. The largest task we set ourselves was to improve the data, revising, correcting, and enlarging all three components of the Middle English Compendium (Middle English Dictionary, Bibliography of sources, and Corpus of primary texts), and in so doing not only to update them once, but to open them to further regular updates in future. Second, we needed to refine the data so as to improve both its potential for search and its potential for productive interaction with sibling dictionaries and similar resources. Third, in parallel with data improvement and refinement, we needed to specify and produce a platform onto which the data could be migrated, a platform that will allow the regular updates and expanded interaction that we have prepared it for. Aside from the third point, which was deliberately left to year two, we have made considerable progress in both revising and refining the data, progress which affects all three MEC components. In some areas we have exceeded expectations (especially in the addition of texts to the Corpus, the examination of newly published material, and in the exposure of hidden spellings); in some we are just where we should be (especially the very time-consuming task of resolving the issues raised by thousands of 'Supplement slips'), and in some areas we have fallen a bit behind (e.g., in revising the 'map' of MED to OED entries).

This project comprises a great number and variety of specific tasks, which are duly listed below in an itemized progress report. For this overview, suffice it to say that the broad outlines of the project remain exactly as originally proposed and that our progress in implementing it fulfills our promises and expectations. Additions to the Corpus and Bibliography are nearly complete; refinements of the Dictionary data are well on their way; and additions and corrections to the Dictionary itself -- the most significant and labor-intensive improvement -- are exactly on track. So far, of the 54,000 MED entries, 4,800 have been modified in some way, of which approximately 3,300 have been 'supplemented' by the addition of some 4,500 additional quotations, 300 by the addition of an explanatory note or change of definition. We have created 693 wholly new entries. In each case, what we have accomplished represents roughly half of what the two-year total would be, so our estimates of the project's scope have so far been borne out.

Circumstances and experience have led to some slight changes of work plan and perhaps also of the project's proportions and priorities, as described below.

Changes to the work plan reflect my evolving judgment (as manager) of the skills and aptitudes of the available staff, as well as their expressed preferences. Thus the "OED map" task, which was originally intended to be distributed amongst the staff, has been taken on by one student

worker who is good at it, a fact which also has the effect of prolonging it well into year two, as has the surprising fact that the pre-existing map that we were counting on tweaking simply does not exist for the alphabet's largest letter (S-), so we are waiting on OED to supply it, a minor setback. Similarly, the task of extracting significant quotations from recent editions has been taken on by another student who excels at that. The two regular staff members (Latta and Logarbo) have gladly surrendered these tasks and have in turn specialized in the editing of Corpus texts and the resolution of Supplement slips respectively. In general, staff members have preferred more keenly than I expected to specialize in a narrower set of tasks and to focus on them, rather than spread their attention more thinly and broadly. I have been content to follow this model.

Changes to priorities and proportions have been subtle, mostly the result of two circumstances: first, that the Supplement slips have proven more various, and more problematic, than we hoped, and therefore a bit more labor-intensive to process (see Appendix). Many slips, for example, say simply, 'modify definition' -- but fail to say how, or why. Others say, 'this is a new spelling' -- but it isn't. Or 'See notes on p. 72' -- without indicating the burden of said notes, or indeed in which book we should hope to find page 72. Each slip raises its own issues, and each needs to be resolved, lest we pass on inaccurate or incomplete information. Nevertheless, we are 65% of the way through the main body of slips, which is good--but would like to have been further along. We have also left very few slips behind. Starting this week, we plan to leave a higher percentage of problematic slips behind, in order to insure that we get through them all, and leave no clear and simple corrections unimplemented.

Secondly, Middle English scholarship has been more productive during the last twenty years than we thought, more editions have appeared than we expected, and the pace appears to be accelerating, especially as regards hitherto neglected areas such as scientific and medical texts and local documents. Rather than examine thirty new editions for potential expansions to the Middle English lexicon, as expected, we have examined a hundred and are not done yet. In keeping with our forward-looking focus on continual updating and ongoing relevance, I have thought fit to allocate appropriate resources (mostly student-worker hours) to this surge in scholarship. Three works in particular, two arriving this past year and one due to appear next week, threaten to consume some of these resources, but also promise to make the MED more relevant than ever: the first are two new editions: Mark Clarke's edition of hundreds of recipes for dyes, paints, etc., mostly unedited till now; and alongside it, Juhani Norri's two-volume *Dictionary of Medical Vocabulary 1375-1550* (2016), an outstanding work, keyed to the MED, containing many technical terms and technical senses missed by MED. Aided by word lists supplied to MED by Professor Norri, I have a plan to refer MED readers to the Norri *Dictionary* whenever that would be helpful, and acknowledge the vocabulary he has unearthed. The third 'work' is the online corpus of Middle English Local Documents (MELD, Stavanger, Norway), which will debut at the end of May with more than two thousand hitherto unavailable and unexamined ME documents. This project offers the potential for genuine collaboration of MELD with MED: we can provide targets for their lemmatisation process; they can provide grist for our lexicographic mill. All three of these works fall under the category of 'recent editions of ME

works' and if nothing else show that Middle English is far from a moribund field, and is therefore all the more in need of a dynamic MED; the last two, along with such things as the revived online Anglo-Norman Dictionary, with which I have also been in contact, have the potential to position MED not just as a standalone resource but as a node in a network of related resources, which is our real objective. They therefore may take on a somewhat (I stress: only somewhat) higher priority than a few of the low-priority tasks designated as such in the original proposal, such as resolving the last few unlinked bibliographic references in the MED, or updating cross-references to standard reference books.

Itemized Progress Report

(The goals described in each case are final goals, to be completed by 30 April 2018.)

A. Expansion of the Corpus of Middle English

1. Create new texts

Goal: Create approximately 100 new searchable texts.

Results to date: Complete. 143 new texts have been keyed, proof-read, and edited to completion. These additional texts enlarge the online Corpus by about a third in terms of bytes, but more than double its size, as regards texts and titles.

Still to do: convert the XML to a format compatible with our online system. This is a routine operation, to be scheduled in due course this coming year.

Status: on track, more texts than promised completed, a little ahead of schedule

2. Document new texts

Goal: Though not explicitly mentioned in our work plan, we do need to create a bibliographic header (effectively: a catalog record) for each text.

Results to date: a bare title list was prepared prior to keying.

Still to do: the bare list needs to be scaled up to fuller catalog records, the work of about a week.

Status: in progress.

B. Expansion of the Bibliography of primary sources.

1. Add new quotable editions.

Goal: Add "approximately 30" select recent editions to the Bibliography component of the Compendium. This will allow said editions to be quoted in the MED, and will allow these editions' editorial notes to be combed for corrections to MED.

Results to date: 100 text editions have been added to the Bibliography, and about 20 existing entries have been corrected. Most of the editions appearing in the main publishers' series have been added.

Still to do: add new editions as they appear (one arrived just this week), search for editions appearing outside of the main series.

Status: original goal far exceeded, but still more to do. See also the comments in the overview about Norri's *Dictionary* and the Middle English Local Documents corpus.

2. Generate new "stencils"

Goal: A stencil is a datable combination of work, manuscript, and edition in the form of a bibliographic short form. We had no estimate of how many of these might be entailed by the addition of new editions.

Results: 786 new 'stencils' have been created.

Still to do: the final number of quotable new stencils may approach 1,000.

Status: original goal was never quantified, but results exceed expectations already.

C. Revision of the Middle English Dictionary

1. Prepare the data for revision.

Goal: "Create a set of 'hooks' in the eMED that will allow the incorporation of supplementary material" and "create a standard for ranking these inclusions on degree of incorporation and degree of reliability."

Results to date: the existing schema for the MED was modified to allow various kinds of supplementation, including an attribute designed to specify not the degree of completion, but the additional treatment necessary for full integration into the dictionary. I.e., we are attaching to each supplemented entry coded information describing what still needs to be done to it. Implicitly this expresses not just the degree of integration, but the degree of reliability, e.g. "needs to be reviewed" indicates less reliability than "no further review necessary."

Status: done.

2. Take a rough inventory of the Supplement slips

Goal: sort the slips as they arrive from the archives into a semblance of order, assign them a rough priority.

Results: done.

Status: done.

3. Create new Supplement slips.

Goal: scour the notes, etc., of newly available editions for examples that supplement or correct what was in MED beforehand. In effect, create new virtual 'supplement' slips from the latest editions. It was not possible to estimate how many of these there might be.

Results: We have created approximately 300 virtual slips, many of them for new words. A few particularly vexing but promising editions remain unharvested, editions chock full of technical vocabulary but which will have to be checked word-by-word, because their editors have not thought fit to say anything useful in the notes.

Status: in progress, approaching completion (except for Clarke, Norri, and MELD).

4. Append and/or implement corrections and additions to existing Dictionary entries.

This is the most important of our tasks, and may be broken down thus by information source:

(1) Pencilled corrections in a print copy of the MED belonging to the old MED production department.

Goal: incorporate all of them into the MED.

Results to date: done.

Status: done.

(2) "Supplement slips" created by MED editors over decades and set aside at the time for a putative future MED Supplement.

Goal: "Incorporate at least at a minimal level the bulk of the available Supplement slips (leaving aside the most problematic) into the text of the eMED."

Results to date: the Supplement Slip boxes when retrieved from the University archives turned out to consist of several different sorts of material, including one box of bottom-of-the-barrel detritus (which we never planned to tackle); three boxes of unorganized and unprocessed slips (low priority); three boxes of entire words, mostly 'i-prefix' words omitted from MED due to a policy misjudgment (taken on by Professor Lewis, with some success, slated to be entered by students in the coming year); and 17 boxes of sorted Supplement slips, arranged by entry headword. These are our chief concern. 11 of the 17 boxes have been gone through this year and their information either attached to the appropriate entry or (occasionally) deliberately marked as problematic. *In sum:* if all of these boxes are put together, including the low-priority ones, they number 24, of which we have successfully addressed exactly half. And we are two thirds of the way to our chief goal, the capture and preservation of the organized Supplement material.

Still to do: the remaining six boxes. Optionally: the unorganized three.

(3) the new 'virtual' slips extracted from recent editions.

Goal: "incorporate ... to at least the minimal level, all such 'new' supplement material."

Results: Only about twenty quotations from this source have so far been attached to the appropriate entries, but since the slips were created by us, with this use in mind, the test batch were readily converted into actual revisions to MED. The remainder should take only a few days.

Status: on track, awaiting the completion of the edition-combing process.

D. Enhancement of search capabilities through data refinement

1. Supply MED - OED links and Modern English lookup forms.

Goal: Complete the mapping of MED entries to OED entries and sub-lemmas, supplying also to each MED entry (where different) a modern English lookup form. Spot check the existing 33,000 links supplied by OED; manually add links to the remaining 21,000 entries.

Results: roughly 40,000 MED entries now map to an OED headword (where one exists) as the result of 7,000 manual mappings.

Still to do: 14,000 remain to be linked, of which 4,000 lie in the "S-" volume, where the original OED automated link generator broke. We hope to persuade OED to re-run their script on these entries, and to supply the remainder manually.

Status: in progress, a little behind.

2. Expose more implicit and abbreviated spellings to full search

Goal: Expand at least the compressed spelling variants (found in approximately 13,000 entries) to make them searchable and incorporate those expanded forms into the MED data.

Results: The form sections of all the entries were extracted and modified by a combination of automatic and manual means. All of the most manageable compressed spelling lists have been expanded; 188,431 spellings were formerly searchable; we've increased that to 231,258.

Still to do: the extracted spelling lists need to be reattached to their original entries (a matter of writing the right script to accomplish that, a day's work). Beyond that, exposing further spellings to search will require a very different approach, and, though valuable, is not within the remit of this grant.

Status: essentially complete.

3. Link many of the approximately 3,000 remaining unlinked quotations in the MED to an authoritative form and entry in the Bibliography.

Goal: Of the nearly one million quotations in the Dictionary, all but a few thousand are authority-controlled and linked to the corresponding Bibliography entry. Our goal was to cut this number in half, researching, correcting, and normalizing cryptic and often erroneous references.

Results: not yet begun.

Status: not done; low priority.

4. Update external cross-references within the Bibliography

Goal: at least experiment with adding and updating links to such external reference works as the latest volume of the Manual of Writings in Middle English; the two new Indexes of Middle English Verse (one in print, one online); the online version of the Linguistic Atlas of Late Medieval English, and the Linguistic Atlas of Early Medieval English.

Results: not yet begun.

Status: not done; low priority, as 'experiment' indicates..

E. "MEC 2.0"

Goal: Design and implement a functional platform for MEC congruent with other developments within the Library's digital architecture.

Results: this is a specifically 'year two' goal. A specification is in the works, but is not yet complete.

Status: awaiting year two.

APPENDIX

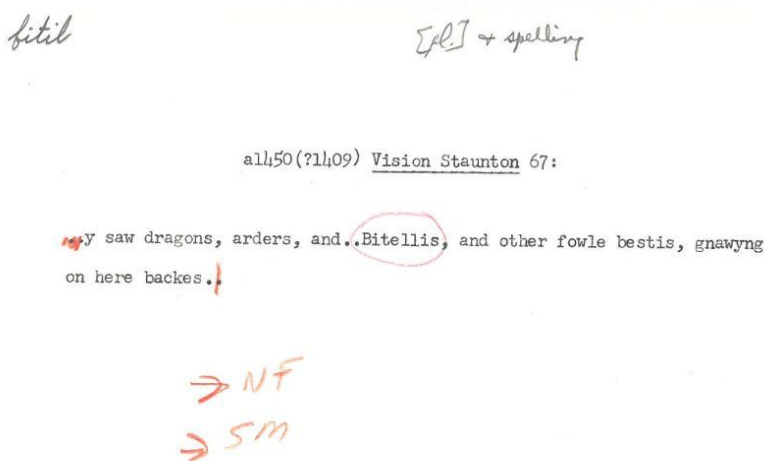
1. Sample of a new 'virtual Supplement' slip from a recent edition.
2. Samples of simple and problematic original Supplement slips.
3. Sample project announcement.

Appendix 1. Sample virtual Supplement slip.

```
F:\MARKUP\mec\work\pfs\SUPPL_VEN_ANAT.sgm
439 </eg>
440 </entryfree>
441
442 <entryfree>
443 <form><orth>grain</orth> <pos>n.</pos></form>
444 <eg>
445 <cit md="1475"><bibl><stncl rid="HYP.XXX135.20170226T154930"><date>a1475</date>
<title>Saliceto(2)</title><ms>Wel 290</ms></stncl> 40/313,315</bibl><q>Yt accordyth
&thorn;at the kutyng &thorn;at be done or owen to be done yn the place of the vpper chanel
and of the nose owen to be done after the granes of &thorn;e place, why in thys membre the
greynes goone after the goyng of the synews and of the muscles of the face and of the nose
</q>
446 </cit>
447 <cit md="1475"><bibl><stncl rid="HYP.XXX135.20170226T154930"><date>a1475</date>
<title>Saliceto(2)</title><ms>Wel 290</ms></stncl> 37/190,192,194,196</bibl><q>The
kutynges whych schuld be mayde, hyt ys not nedeful to be in euery membre after the graynes,
as yn the forhede..although the greynes gone ouerthwert the kutynges; and namely depe
owen to be done after the contrary of the greynes, for the synews of the forhede gone after the
lengthe and after the greynes.</q>
448 <note></note>
449 </cit>
450 </eg>
451 <note>The sense here appears to correspond not merely to OD sense 14 ('texture,
composition') but specifically to sense 15 ('the longitudinal arrangement of fibres or particles,
in lines or veins more or less parallel along which the material is more easily cloven or cut
than in any other direction'); if so it is a sense not otherwise attested in MED, and attested in
OD only from the late 16th century. </note><note></note>
452 </entryfree>
453
```

Appendix 2. Sample original (paper) Supplement slips

a. A relatively simple Supplement slip. (NF = “new form”; SM = “supplementary material”)



This adds a spelling, and the first non-glossary quotation, to the entry “bitil” n., appended as below -- but the observant editor will have noticed that the spelling “arders” (pl. of “nadder” (modern “adder”)) is also unattested, so the same quotation should be appended to that entry.

```

1 <?xml version="1.0" encoding="US-ASCII" ?>
2 <?xml-stylesheet type="text/css" href="http://www-personal.umich.edu/~pfs/mec/code/med.css"?>
3 <!DOCTYPE MED SYSTEM "http://www-personal.umich.edu/~pfs/mec/code/med.xml.dtd">
4 <MED>
5 <ENTRYFREE ID="MED4945" SEQ="4945">
6 <FORM><ORTH>bitil</ORTH> <POS>n.</POS> Also <ORTH>bitil</ORTH>, <ORTH>b&macr;&dotb;til
  </ORTH>.</FORM>
7 <ETYM><LANG>OE</LANG> <HI REND="B">bitela</HI> (?orig. `a biting vermin or insect').</ETYM>
8 <SENSE>
9 <DEF>Apparently applied to various kinds of vermin and insects: bug, beetle.</DEF>
10 <EG>
11 <CIT CD="1440"><BIBL><STNCL RID="hyp.934.19981211T105002"><DATE>(1440)</DATE> <TITLE>
  PParv.</TITLE><MS>Hrl 221</MS></STNCL> 37</BIBL>
12 <Q>Bytylle [vrr. bytyyl, bityl], worme: Buboscus.</Q></CIT>
13 <CIT MD="1500"><BIBL><STNCL RID="hyp.256.19981211T105002"><DATE>?a1500</DATE> <TITLE>
  Lndsb.Nominale</TITLE> <MS>Lndsb</MS></STNCL> 767</BIBL>
14 <Q>Nomina vermium..Hic carembes: a betylle.</Q></CIT>
15 <CIT MD="1500"><BIBL><STNCL RID="hyp.1641.19981211T105002"><DATE>?a1500</DATE> <TITLE>
  Trin-C.LEDict.Suppl.</TITLE><MS>Trin-C O.5.4</MS></STNCL> 609</BIBL>
16 <Q>A bytylle [ <TITLE>Trin-C.LEDict.</TITLE>]: Scarabo: a scharabot].</Q></CIT>
17 </EG>
18 </SENSE>
19 <supplement todo="BNI">
20 <eg>
21 <cit><bibl><stncl><date>a1450(?1409)</date> <title>Vision Staunton</title></stncl> 67</bibl>
22 <q>y saw dragons, arders, and..Bitellis, and other fowle bestis, gnawying on here backes.</q></cit>
23 <note>New form: bitellis (pl.)</note>
24 </eg>
25 </supplement>
26 </ENTRYFREE>
27 </MED>
  
```

b. A relatively messy Supplement slip, with four alternate identifications of the same word.

Jape
 c1475 (c1450) Idley Instr. 2.A. 2217
 God..graunt eche wedded man.. To abide at hom with
 a groyne or a chappe.
 → NE chappe — add as
 → Ed. gloss: with grumble and rap.
 → "Jaw of a pig" — form of 'jape' (?)
 per SMK
 or Noun & Chopper V. (i) ?

..very dubiously converted to a new queried entry 'chap' n.:

```

F:\MARKUP\mec\work\MED-working_copy\c\MED07219.5-chap_n.xml
1 <?xml version="1.0" encoding="US-ASCII" ?>
2 <?xmlstylesheet type="text/css" href=
  "http://www-personal.umich.edu/~pfs/mec/code/med.css"?>
3 <!DOCTYPE MED SYSTEM "http://www-personal.umich.edu/~pfs/mec/code/med.xml.dtd">
4 <MED>
5 <ENTRYFREE ID="MED07219.5" SEQ="7219.5" RESP="pfs" DUB="Y">
6 <FORM><ORTH>chap</ORTH> <POS>n.</POS> Also <ORTH>chappe</ORTH>.</FORM>
7 <SENSE>
8 <DEF>?The jaw ?of a pig.</DEF>
9 <EG>
10 <CIT CD="1450" MD="1475"><BIBL><STNCL><DATE>c1475(c1450)</DATE> Idley
  <TITLE>Instr.</TITLE></STNCL>2.A.2217</BIBL>
11 <Q>God..graunt eche wedded man such a grace To abide at hom with a groyne or a
  chappe, Till the worlde be turned into anothir shappe.</Q></CIT>
12 </EG>
13 </SENSE>
14 <NOTE>Stub entry.</NOTE>
15 </ENTRYFREE>
16 <supplement todo="ERI">
17 <note>The phrase 'a groyne or a chappe' is a crux. The treatment of 'chappe' here suits
  the food metaphors in this passage; agrees with the MED's previous treatment of 'groyne'
  s.v. groyne n. ('pig snout'); and assumes the continuity of the byform 'chap' recorded in
  OED as chap n.(2), dating in OD only to the mid-sixteenth century. The editor however, by
  glossing the phrase 'with grumble and rap' appears to identify 'groyne' as a form of grone
  ('groan') and 'chappe' as a form of 'chop' ('blow'), against which stands the rhyme. A third
  alternative is to take chappe as a form of jape n., the phrase meaning something like 'with
  a groan or a joke.</note>
18 </supplement>
19 </MED>
20
  
```

Appendix 3: sample project announcement (this one directed to the electronic newsletter of the Dictionary Society of North America).

We at the University of Michigan Library are pleased to report that revision of the [Middle English Dictionary](#) and its associated resources has begun, after a 20-year hiatus, thanks to a two-year grant (2016-18) awarded under the Humanities Collections and Reference Resources program of the National Endowment for the Humanities. Our hope is that this effort will set in motion a process of ongoing revision, but our immediate goals are to perform only the most urgent updates, publishing as much supplementary information as we can, even when we can not afford to integrate it fully into the existing online Dictionary. We have been attaching to the appropriate entries additional quotes and notes taken chiefly from two sources: (1) about 20,000 'supplement slips' set aside over decades by the MED editors in hope that an MED Supplement would some day be compiled. These have hitherto lain largely inaccessible in boxes in the university archives. And (2) a much smaller number of new virtual supplement slips drawn from editorial notes in recent editions of ME texts, notes of the sort that typically declare, "this word (or sense, or spelling) omitted by MED." We have also been adding those more recent editions (or some of them) to the Dictionary's Bibliography, and will be roughly doubling the number of primary texts hosted on the Middle English Compendium site. [Corrections and additions either to MED or to its Bibliography are welcome.](#)

We have also continued the long process of removing artifacts of its print origins from the Dictionary, in favor of data that is more amenable to being searched and manipulated by computer. We are expanding some of the implicit and contracted spellings given in the print MED, thereby exposing them to search; and we are completing a nearly comprehensive map of MED to OED lemmas and sub-lemmas, thereby facilitating not only a seamless link to the corresponding OED entry but potentially also a lookup search of the MED itself by modern English equivalent (mostly to be identified with the OED headword).

Our larger goal is to transform the MED from an essentially static resource to an essentially dynamic one, in two respects, editorially, adding the ability to accept, store, and publish additions and corrections; and technically, preparing for the migration of the Dictionary and Compendium data to new platforms (to avoid technical obsolescence) and new functionality, which will allow the data to be opened up for reuse by other products and projects, thus fulfilling the vision that we always had for the MEC as a node in a network of historical dictionaries, electronic editions, text portals, and other linguistic resources.

[Paul Schaffner](#) (P.I.); [John Latta](#) (editor); [Mona Logarbo](#) (editor); [Robert E. Lewis](#) (MED chief editor emeritus; volunteer editor).