

<b>Institution: DE MONTFORT UNIVERSITY</b>
<b>Unit of Assessment: 29 English Language and Literature</b>
<b>Title of case study: Editing Literary-Historical Manuscripts</b>
<p><b>1. Summary of the impact</b> (indicative maximum 100 words)</p> <p>Contributing to the preservation of literary materials through innovative use of technology, DMU's Centre for Technology and the Arts (CTA) – subsequently renamed the Centre for Textual Studies (CTS) – pioneered new digital techniques for analysing, editing and presenting literary-historical manuscripts of international significance. These techniques revolutionized the scholarly task of capturing data about manuscripts, permitting new kinds of analyses, editing and dissemination, now widely practised to facilitate public access and cultural enrichment. In particular, the CTA/CTS invented a manuscript description standard taken up by major libraries across the world, the International Standards Organization (ISO) via the Text Encoding Initiative (TEI) and by commercial publishers.</p>
<p><b>2. Underpinning research</b> (indicative maximum 500 words)</p> <p>DMU's CTA was established in 1998 to develop research into electronic editorial practice under the leadership of Peter <b>Robinson</b> (1993–2004), working with Norman <b>Blake</b> (1998–2004). The CTA produced editions of Chaucer's <i>Canterbury Tales</i> based on fresh analysis of all the early manuscripts (HRB funded £100,000 1996–99, AHRB funded £394,000 1999–2004, Leverhulme Trust funded £60,000 1995–98). The prevailing digital document encoding standard at the time – the Text Encoding Initiative (TEI) P3 standard – could not cope with the problems of encoding 400–600 year old English manuscripts. By examining hundreds of relevant manuscripts, <b>Robinson</b> derived a new taxonomy of contextualizing information that could be extended by repetition within elements to produce manuscript descriptions of unprecedented length and sophistication. The new standard for manuscript description was formally defined via the newly created Extensible Markup Language (XML), the advanced features of which it was amongst the first to exploit. In 1999 <b>Robinson's</b> edition of <i>The Wife of Bath's Prologue</i> on CD-ROM won the Beatrix White Prize for the best contribution to Medieval and Renaissance scholarship.</p> <p><b>Robinson</b> recognized that were his standard to be adopted by libraries worldwide it would enable the creation of an online union catalogue of all their manuscript holdings. In 1999, he initiated the "Manuscript Access through Standards for Electronic Records" (MASTER) project, funded by the EU (376,000 euro 2001–03) and the European Science Foundation (100,000 French francs 1999). MASTER, with DMU as lead partner of six collaborating European universities, promulgated <b>Robinson's</b> standard and developed and distributed the necessary software for its adoption. Details of MASTER's achievements are archived at <a href="http://master.dmu.ac.uk">http://master.dmu.ac.uk</a> (accessed 13/08/13).</p> <p>The CTA developed other research projects deploying new approaches to scholarly editing enabled by <b>Robinson's</b> work, anticipating the revolution in digital publishing and making the core documents of our literary heritage accessible to all. In 2005, a collaboration with the British Library, the University of Wales Bangor and Keio University, Japan, was established to work on manuscripts of Malory's <i>Le Morte Darthur</i> <a href="http://malory.dmu.ac.uk">http://malory.dmu.ac.uk</a> (accessed 13/08/13). In 2004, the CTA was renamed the Centre for Textual Studies by its new Director Peter <b>Shillingsburg</b> (2002–10) and launched further projects, including <b>Briggs</b> (1994–08)/<b>Shillingsburg's</b> 'Woolf "Time Passes"' project (Leverhulme, £52,780 2006) and partnership projects with UCL on Dante's <i>Commedia</i> (AHRB 2002–04), with Princeton on the <i>Monarchia</i> (AHRB 2003–04), and with the University of Liverpool (AHRB 2003–04).</p> <p>The CTA/CTS developed a wholly new kind of study adapted from the methods of evolutionary biology, called Phylogenetic Analysis using Parsimony (PAUP). Treating random word alterations that occur during scribal transcription as analogous to random genetic alterations in biological</p>

## Impact case study (REF3b)

processes, PAUP produces 'family trees' showing the manuscripts' linguistic relatedness, called cladograms. This research led to a 2008 landmark article in the world's most-cited and influential interdisciplinary science journal, *Nature* (circulation 53,000, readership 424,000).

The CTA/CTS developed a new publication technology – the XML-based Anastasia system (2000) – enabling literary-historical scholars to combine complex manuscripts and transcriptions in a single digital package. This technology was first delivered by CD-ROMs and later over the Internet. <http://www.sd-editions.com/anastasia> (accessed 13/08/13).

### 3. References to the research (indicative maximum of six references)

- 1) Peter Robinson: "A Stemmatic Analysis of the Fifteenth-Century Witnesses to *The Wife of Bath's Prologue*" in N. F. Blake and Peter Robinson (eds.) *Canterbury Tales Project Occasional Papers II* (London: The Office for Humanities Communication, 1997): pp. 69–132.
- 2) Elizabeth Solopova: "Chaucer's Metre and Scribal Editing in the Early Manuscripts of *The Canterbury Tales*" in N. F. Blake and P. M. Robinson (eds.) *Canterbury Tales Project Occasional Papers II* (London: The Office for Humanities Communication, 1997): pp. 143–164.
- 3) Norman Blake: "Editing the *Canterbury Tales*: Preliminary Observations" *Anglia: Zeitschrift für Englische Philologie* 116 (1998): 198–214.
- 4) Norman Blake "The Links in the *Canterbury Tales*" in Susan Powell, Jeremy J. Smith and Derek Pearsal (eds) *New Perspectives on Middle English Texts: A Festschrift for R. A. Waldron* (Woodbridge: Brewer, 2000): 107–18.
- 5) Takako Kato: "Irregular Textual Divisions in Caxton's *Morte Darthur*: Paraphs and Chapter Divisions" *Poetica: An International Journal of Linguistic-Literary Studies*, (53), 2000, 15–38. (2000)
- 6) Peter Shillingsburg: "Private Reading, Public Writing: W. M. Thackeray, Mrs. Grundy, and the Market" *Variants: The Journal of the European Society for Textual Scholarship*, (2-3), 2004, 271-78. (2004)

Evidence of Quality: all peer reviewed. (Outputs available on request)

### 4. Details of the impact (indicative maximum 750 words)

Over the census period, the **reach** and **significance** of the CTA/CTS research programme have been **global**. The importance of the *Canterbury Tales* project to **cultural heritage preservation** is acknowledged by the National Library of Wales, the repository of the Hengwrt Chaucer, the earliest extant Chaucer manuscript. A digitised version of the manuscript was prepared with DMU, using the techniques developed in the CTA/CTS, and enabled the Library to better understand and preserve the fragile parchment of one of its greatest treasures, providing the widest textual archaeology of the Hengwrt manuscript and its scribe explicitly for 'the benefit of the public'. Visitor numbers now total approx 100,000 annually with over two million remote visitors to its site. Building on the CTA/CTS research, the Library is now established as one of Europe's leading centres for digitisation, and its three-year strategy for agile electronic access 2011-14 is recognized as a model for facilitating access to recorded cultures.

In the census period, the CTA/CTS work on a standardized system called MASTER for manuscript description achieved catalogue integration across the holdings of the Royal Library in the Hague, the Arnamagnaean Institute in Copenhagen, L'Institut de recherche et d'histoire des textes in Paris, the National Library of the Czech Republic, Prague, the Bodleian Library, Oxford, the British Library, the Vatican Library and the Biblioteca Ambrosiana, Milan. This amounts to 61,000 manuscripts – the core of Europe's rare manuscript holdings – being combined into a single searchable catalogue using the CTA/CTS standard.

**Impact case study (REF3b)**

The latest manifestation of this is the EU-funded 5.57M euro ENRICH project (2007-09), which explicitly builds on the basis of the CTA/CTS achievement in this field. This created a base for a digital library of European cultural heritage, which made available more than five million digitized pages for a wide range of non-academic users, including 'libraries, museums and archives, policy makers and general interest users'. Examples include the National Library of the Czech Republic, Prague, Centro per comunicazione e l'integrazione dei media, Florence, SYSTRAN S.A., Paris, and Biblioteca Nacional de España, Madrid, Spain.

The XML-based Anastasia system enabled scholars working on manuscript-based historical and literary projects to publish their editions without additional technical support, and it has been adopted by those working on ancient Biblical manuscripts of major **theological significance**. These include Digital Nestle-Aland, which is an electronic version of the standard scholarly edition of the Greek New Testament published in Stuttgart 2012 (ISBN 978-3-438-05140-0), and the online edition of the *Codex Sinaiticus*, one of the most important books in the world, produced in 2009 in an international partnership between the British Library, the National Library of Russia, St Catherine's Monastery and Leipzig University Library.

From 2007, this CTA/CTS standard for manuscript description officially became the whole world's standard. The Text Encoding Initiative (TEI) is an affiliated organization of the International Standards Organization (ISO) – itself a body with consultative status on the United Nations Economic and Social Council – and it has adopted the CTA/CTS standard. In 2002, the P4 version of the TEI guidelines (an interim release) adopted the CTA/CTS approach to produce its first systematic instructions on encoding manuscript descriptions, and the latest version, P5 published in 2007, fully incorporates the CTA/CTS work. The CTA/CTS-initiated MASTER project thus became the model for other projects attempting to build inter-library union catalogues.

Because virtually all commercial publication of scholarly works uses the TEI guidelines, the CTA/CTS standard has become ubiquitous in commercial publications based on manuscript sources. Virtually every publisher uses TEI encoding when handling complex manuscript materials. Typical examples include Cambridge University Press, which used the CTA/CTS standard to publish two volumes of the letters of Samuel Beckett in 2009–11 (ISBNs 978-0521867931, 978-0521867948). In 2009, Leo Jansen, Hans Luijten and Nienke Bakker edited the artist's letters using the CTA/CTS standard for the Van Gogh Museum in Amsterdam (ISBN 978-0500238653). 2,000 pages of Isaac Newton's manuscripts have been digitized using the CTA/CTS standard as part of Indiana University Press's 'The Chymistry of Isaac Newton' project. Between 2009 and 2011, AMS Press of New York published three volumes of the work of James Fenimore Cooper using the CTA/CTS standard (ISBNs 978-0404644673, 978-0404644666 and 978-0404644796). A full list of projects using the TEI standard is given in section (5) below; all those involving manuscript work use the CTA/CTS standard developed at De Montfort.

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

For more information about the Hengwrt Chaucer, its importance and DMU's role in its digitisation, please see <http://www.llgc.org.uk/?id=257> (accessed 13/08/13). For more information about the libraries digitisation programme, please see <http://www.llgc.org.uk/index.php?id=122> and the links therein (accessed 13/08/13). For the reference to 'benefit of the public', please see [http://www.llgc.org.uk/fileadmin/documents/pdf/nlw\\_strategy\\_s.pdf](http://www.llgc.org.uk/fileadmin/documents/pdf/nlw_strategy_s.pdf) (accessed 22/08/13)

Evidence for the catalogue integration across Europe can be seen from the MASTER project – please see <http://master.dmu.ac.uk/>

For further information about the ENRICH project, please see <http://enrich.manuscriptorium.com/> This describes the project "ENRICH: Towards a European Digital Library of Manuscripts", which built on MASTER to bring in more libraries and further develop the CTA/CTS manuscript encoding standard (accessed 13/08/13).

To see examples of the internationally important books which have been put into the public domain

**Impact case study (REF3b)**

as a consequence of this research, please see the following links (both accessed 13/08/13):

- [www.codexsinaiticus.org](http://www.codexsinaiticus.org) This describes the scope and progress of the work on the Codex Sinaiticus, building on its adoption of the CTA/CTS encoding standard and its publication software.
- <http://nestlealand.uni-muenster.de> This describes the University of Munster Institute for New Testament Textual Research's adoption of the CTA/CTS encoding standard and its publication software.

Evidence for becoming the world's standard: developing the Text Encoding Initiative:

- [http://www.tei-c.org/About/Archive\\_new/Master/Hermes/index.htm](http://www.tei-c.org/About/Archive_new/Master/Hermes/index.htm) This describes how the TEI first began its adoption of the manuscript description standards created by the CTA/CTS MASTER project (accessed 13/08/13). See the sections "1 Introduction" and "2 Background" for the leading role of MASTER within the international collaboration that led to TEI P5's incorporation of the CTA/CTS work.
- [http://www.tei-c.org/About/Archive\\_new/Master/Reference/oldindex.html](http://www.tei-c.org/About/Archive_new/Master/Reference/oldindex.html) This describes the TEI's final adoption of the manuscript description standards created by the CTA/CTS MASTER project. See in particular the "Intro's" assertion "In its present form, the system documented here was produced as a major deliverable of the MASTER (Manuscript Access through Standardization of Electronic Records) project, funded by the European Union Framework IV program from January 1999 to June 2001."
- For the current TEI guidelines, please see <http://www.tei-c.org/index.xml> (accessed 13/08/13). This shows that the CTA/CTS research on manuscript description became the basis of the TEI's standard and hence the standard for all the world, including virtually all commercial publishing of manuscript materials, which is almost invariably achieved using TEI P5.
- <http://www.tei-c.org/Activities/Projects> This is a list of the projects using the TEI standard that is based on the CTA/CTS manuscript encoding standard. (accessed 13/08/23)

In addition, Barbrook, Adrian C, Christopher J. Howe, Norman Blake and Peter Robinson "The Phylogeny of *The Canterbury Tales*" *Nature* 394 (1998): 839 is an article in the world's most widely read *science* journal, showing the impact of the CTA/CTS work beyond its disciplinary boundaries.