

## Old wine, new wineskins

Smith, Catherine

*License:*

Other (please specify with Rights Statement)

*Document Version*

Version created as part of publication process; publisher's layout; not normally made publicly available

*Citation for published version (Harvard):*

Smith, C 2019, 'Old wine, new wineskins: digital tools for editing the New Testament', *Wissenschaftliche Untersuchungen zum Neuen Testament*, vol. 417, pp. 407-434.

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

Checked for eligibility: 25/02/2019

Smith, C. (2019) Old wine, new wineskins: Digital tools for editing the New Testament. In G. V. Allen (Ed. ), The future of New Testament textual scholarship (pp. 247-264). Tübingen, Germany: Mohr Siebeck.

For non commercial use only.

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# The Future of New Testament Textual Scholarship

From H. C. Hoskier to the Editio Critica Maior  
and Beyond

Edited by  
Garrick V. Allen

Mohr Siebeck

Author's e-offprint with publisher's permission

GARRICK V. ALLEN, born 1988; 2015 PhD; Lecturer in New Testament Studies at Dublin City University, Republic of Ireland.  
orcid.org/0000-0003-4428-1775

ISBN 978-3-16-156662-2 / 978-3-16-156663-9  
DOI 10.1628/978-3-16-156663-9

ISSN 0512-1604 / eISSN 2568-7476  
(Wissenschaftliche Untersuchungen zum Neuen Testament)

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data are available at <http://dnb.dnb.de>.

© 2019 Mohr Siebeck Tübingen, Germany. [www.mohrsiebeck.com](http://www.mohrsiebeck.com)

This book may not be reproduced, in whole or in part, in any form (beyond that permitted by copyright law) without the publisher's written permission. This applies particularly to reproductions, translations and storage and processing in electronic systems.

The book was printed on non-aging paper by Gulde Druck in Tübingen, and bound by Großbuchbinderei Spinner in Ottersweier.

Printed in Germany.

Author's e-offprint with publisher's permission

## Table of Contents

Preface.....	IX
--------------	----

### Intellectual History of Textual Scholarship

*Garrick V. Allen*

The Patient Collator and the Philology of the Beyond: H. C. Hoskier and the New Testament .....	3
--	---

*Juan Hernández Jr.*

Hoskier's Contribution to the Apocalypse's Textual History: Collations, Polyglots, Groupings.....	39
--	----

*Martin Karrer*

Herman Charles Hoskier and the Textual Criticism of Revelation.....	51
---	----

*Jan Krans*

Hoskier in the Spiritual World .....	69
--------------------------------------	----

*Jennifer Wright Knust*

On Textual Nostalgia: Herman C. Hoskier's Collation of Evangelium 604 (London, British Library Egerton 2610; GA 700) Revisited.....	79
--	----

*Peter J. Gurry*

'A Book Worth Publishing': The Making of Westcott and Hort's Greek New Testament (1881) .....	103
--	-----

### The Status Quaestionis and Future of Textual Scholarship

*Stanley E. Porter*

The Domains of Textual Criticism and the Future of Textual Scholarship....	131
--	-----

Table of Contents

*Gregory Peter Fewster*

Finding Your Place: Developing Cross-reference Systems  
in Late Antique Biblical Codices ..... 155

*Christina M. Kreinecker*

Papyrology, Papyrological Commentary, and the Future  
of New Testament Textual Scholarship ..... 181

*Jacob W. Peterson*

Patterns of Correction as Paratext: A New Approach with Papyrus 46  
as a Test Case..... 201

*Dirk Jongkind*

Redactional Elements in the Text of Codex B ..... 231

*H. A. G. Houghton*

The Garland of Howth (Vetus Latina 28): A Neglected Old Latin  
Witness in Matthew..... 247

*Curt Niccum*

Hoskier and His (Per)Version of the Ethiopic..... 265

*Thomas J. Kraus*

Ostraca and Talismans: The Story of Two Former Text-Critical  
Categories and What to do with Them Today..... 283

*An-Ting Yi*

The Critical Apparatus of Stephanus' Greek New Testament of 1550:  
Early Printed Editions and Textual Scholarship..... 305

*Tommy Wasserman*

Methods of Evaluating Textual Relationships: From Bengel  
to the CBGM and Beyond..... 333

*J. K. Elliott*

Thoroughgoing Eclectic Textual Criticism: Manuscripts and Variants  
of Revelation..... 363

*Jill Unkel*

Speaking in Tongues: Collecting the Chester Beatty Biblical Manuscripts... 379

*Table of Contents*

Editing the New Testament in a Digital Age

<i>D. C. Parker</i> The Future of the Critical Edition .....	395
<i>Catherine Smith</i> Old Wine, New Wineskins: Digital Tools for Editing the New Testament....	407
<i>Klaus Wachtel</i> The Development of the Coherence-Based Genealogical Method (CBGM), its Place in Textual Scholarship, and Digital Editing.....	435
<i>Annette Hüffmeier</i> Apparatus Construction: Philological Methodology and Technical Realization.....	447
Contributors .....	461
Bibliography .....	463
Ancient Sources Index.....	503
Modern Author Index .....	513
Subject Index .....	521

# Old Wine, New Wineskins: Digital Tools for Editing the New Testament

*Catherine Smith*

The use of computers in New Testament textual criticism goes back almost as far as the first use of a computer for any kind of text processing. That first project is generally agreed to be the *Index Thomisticus* begun in 1949 by Roberto Busa.<sup>1</sup> Busa's own account of the work has become something of a founding myth of digital humanities.<sup>2</sup> Part of the story recounts his interaction with another pioneer of the discipline, John W. Ellison. Ellison was the first person to use a computer to address problems of New Testament textual criticism and began his work only a year after Busa.<sup>3</sup> While technology has changed considerably in the intervening decades, the themes and motivations that were key in Ellison's work are still present in the discussion of our use of digital tools today. This chapter discusses the development of computing in the discipline from this early adoption to our increased reliance on it today. It examines the growing expectations of software and some of the challenges that must be addressed if our data and tools are to continue to adapt to the ever-changing digital environment.

---

<sup>1</sup> Susan Hockey, "The History of Humanities Computing," in *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth (Malden: Blackwell, 2004), 4; Steven E. Jones, *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards* (New York: Routledge, 2016).

<sup>2</sup> Roberto Busa, "The Annals of Humanities Computing: The Index Thomisticus," *Computers and the Humanities* 14/2 (1980): 83–90; Jones, *Roberto Busa*, 1–3.

<sup>3</sup> Robert A. Kraft, "The Use of Computers in New Testament Textual Criticism," in *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis*, ed. B. D. Ehrman and M. W. Holmes (Grand Rapids: Eerdmans, 1995), 268.

## A Selective History of Electronic Editing

The magic of myth and legend has come true in our time. One types the correct incantation on a keyboard, and a display screen comes to life, showing things that never were nor could be (1975).<sup>4</sup>

Ellison's doctoral work focused on New Testament textual criticism but he is probably best known for his work on the first computer-generated concordance of the Bible, and it is in this role that he and Busa crossed paths.<sup>5</sup> They began their work independently and, as far as I can tell, each without knowledge of the other; as Busa himself says, these ideas tend to arise from the *milieu*.<sup>6</sup> They were also working with different technologies. Busa was using IBM punch card machines with all data entry and manipulation performed using punched cards.<sup>7</sup> He initially estimated that his project would require a total of 13 million punched cards (weighing about 500 tonnes and taking up 108 cubic metres of space). Ellison used the Remington Rand Univac machine and, while he did use punched cards for some of the initial input verification, the main data entry, processing instructions, and storage all used magnetic tape.<sup>8</sup> The story goes that Busa read

---

<sup>4</sup> Frederick P. Brooks Jr., *The Mythical Man-Month: Essays on Software Engineering*, anniversary ed. (Boston: Addison-Wesley, 1975); (repr., 1995 with four new chapters), 7–8.

<sup>5</sup> Meredith Hindley, "The Rise of the Machines. NEH and the Digital Humanities: The Early Years," *Humanities: The Magazine of the National Endowment for the Humanities* 34/4 (2013), <https://www.neh.gov/humanities/2013/julyaugust/feature/the-rise-the-machines> [accessed 9 March 2018]. Depending on whether or not punched card machines are classified as computers, Ellison could have been the first person to produce the first *computer* generated concordance of anything. Oakman cites Ellison's concordance as the starting point for humanities computing (Robert L. Oakman, *Computer Methods for Literary Research*, 2<sup>nd</sup> ed. [Athens: University of Georgia Press, 1984], ix), which is also noted by Edward Vanhoutte, "The Gates of Hell: History and Definition of Digital | Humanities | Computing," in *Defining Digital Humanities: A Reader*, ed. M. Terras, J. Nyhan, and E. Vanhoutte (Farnham: Ashgate, 2013), 131. Hindley specifically credits Ellison with making "the crucial jump from accounting machines to computers" (Hindley, "The Rise of the Machines"). Along these lines Clair M. Cook, "Automation Comes to the Bible," *Computers and Automation* 7/3 (1958): 18 (repr. of the same article in *Christian Century* 54/30 [1957]) reports that Ellison experimented with punched card machines for several years before deciding they were not suitable for his text-critical work and turning to the electronic computer.

<sup>6</sup> Busa, "The Annals of Humanities Computing," 84.

<sup>7</sup> Busa worked in collaboration with H. Paul Tasman, a Senior Engineer at IBM. Jones, *Roberto Busa*, 18.

<sup>8</sup> "Bible Labor of Years is Done in 400 Hours," *Life* (February 1957): 92.

Gardner Soule, "The Machine that Indexed the Bible," *Popular Science* (November 1956): 174. Ellison's work on the RSV concordance was completed alongside Remington Rand engineers John Graham (algorithm design) and Al Bosgang (programming).



about Ellison's work in a newspaper which led to the two meeting.<sup>9</sup> Busa's own account is as follows:

One day I learned from the newspapers that an Episcopalian minister, Rev. John W. Ellison who was preparing a concordance of the Revised Version of the Bible...had used Remington magnetic tapes...I went to shake hands with him and said: "You are a great ally of mine!" Immediately after I went to IBM: "See what Remington is doing?" Since that time the processing of the IT [*Index Thomisticus*] has been done mainly by computers and punched card equipment was used only peripherally.<sup>10</sup>

Recent research by Steven Jones suggests that the relationship between the two was more significant than this summary account implies. The Busa archive has an entire folder relating to Ellison which contains documents and correspondence that illustrate what Jones describes as "a true intellectual collaboration" including sharing details of their methodologies.<sup>11</sup> Ellison's concordance of the RSV was completed in 1955 but it took two years to typeset and was published in 1957.<sup>12</sup> Busa's completed concordance of Thomas Aquinas' writings was published in 56 volumes between 1974 and 1980 and only required six million punched cards thanks to the use of magnetic tape.<sup>13</sup>

---

<sup>9</sup> Thomas Nelson Winter, "Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance," *The Classical Bulletin* 7/1 (1999): 13 suggests that this meeting might have happened in 1954, and he is followed in that by Vanhoutte ("The Gates of Hell," 127). Winter seems to assume this date from a subscript in a paper by Busa, which lists a number of articles in the press that relate to Ellison's work, the earliest being from 1954 (Roberto Busa, "The Use of Punched Cards in Linguistic Analysis," in *Punched Cards: Their Applications to Science and Industry*, ed. R. S. Casey et al., 2<sup>nd</sup> ed. [New York: Reinhold Publishing Corporation, 1958], 373). However, this article in *Time* ("According to Mark IV," *Time* [9 August 1954]) concerns Ellison's doctoral work on Textual Criticism and does not mention Remington Rand (since he was using the Harvard Mark IV), magnetic tapes, or his work on the concordance. Recent research by Jones using the Busa archive suggests that the men had no contact until 1956 with the initial contact made in May of that year, their correspondence continuing through the Autumn (Jones, *Roberto Busa*, 101). There were a few articles in the press in early 1956 that could have been the trigger for contact. A very brief announcement of the concordance project appeared in the *New York Times* in February, which mentions the Univac ("Bible Index Computed," *The New York Times*, 17 February 1956), and a technically detailed article in the IBM Systems magazine in March–April (William R. McCully, "Univac Compiles a Complete Bible Concordance," *Systems Magazine* [March–April 1956]: 22–3).

<sup>10</sup> Busa, "The Annals of Humanities Computing," 85.

<sup>11</sup> Jones, *Roberto Busa*, 101.

<sup>12</sup> The typesetting was not done by the Univac. The output was printed with a Uniprinter (which took about 1,000 hours) and the pages were then sent to the typesetters who "follow[ed] the usual procedure for the manufacture of a book" according to McCully, "Univac Compiles a Complete Bible Concordance," 23. See John W. Ellison, *Nelson's Complete Concordance of the Revised Standard Version Bible* (New York: Thomas Nelson, 1957).

<sup>13</sup> Roberto Busa, "Concluding a Life's Safari from Punched Cards to World Wide Web," in *The Digital Demotic: Selected Papers from Digital Resources in the Humanities 1997*,

Ellison's work on textual criticism started before his work on the concordance. In 1950, Ellison approached Howard Aiken at the Harvard Computational Laboratory to ask for processing time on the university computer for his project, apparently the first humanities scholar to do so.<sup>14</sup> In 1951 Ellison was granted time on the (still not quite operational) Harvard Mark IV computer and began learning how to program it.<sup>15</sup> He had a vision to transform the way data was collected, processed, and shared by textual critics. He observed that while "[t]he differences between any given pair of manuscripts are finite and immutable"<sup>16</sup> each scholar who studied a manuscript started from scratch by comparing it to a standard printed text and a selection of other manuscripts before publishing the results.<sup>17</sup> He suggested that, in order to deal with the large volume of New Testament manuscripts, data could instead be collected collaboratively according to agreed methods and categories and stored in look-up tables akin to the logarithm tables used for mathematics.<sup>18</sup> He also wanted to see whether using all the points of variation in a small selection of verses could find the same family groupings that had been established by traditional methods using the "significant" readings from the full text.<sup>19</sup>

The ideas behind this computer-assisted method were that the results would be reproducible across a range of New Testament texts and that the work could be built on by adding the data from other manuscripts to those already studied without having to start the entire comparison process again. In keeping with his principles of data reuse Ellison did not make any manuscript collations himself, but instead obtained permission to use those being prepared by the American committee of the IGNTP for their edition of Luke.<sup>20</sup> In addition to the 81 IGNTP collations, Ellison also used 200 collations of Luke 10:1–15 prepared by Merrill Parvis and it is this additional data that determined the selection of these 15 verses for his research.<sup>21</sup> Each of these collations would have recorded, for a

---

Office for Humanities Communication Publication 10, ed. L. Burnard, M. Deegan, and H. Short (London: Office for Humanities Communication, 1998), 3.

<sup>14</sup> John W. Ellison, "Computers and the Testaments," in *Computers in Humanistic Research, Readings and Perspectives*, Prentice-Hall Series in Automatic Computation, ed. E. A. Bowles (Englewood Cliffs: Prentice-Hall, 1967), 163.

<sup>15</sup> While Busa and Ellison both worked with programmers on their concordances, Ellison wrote his own algorithms for his text-critical work. Ellison, "Computers and the Testaments," 163–64.

<sup>16</sup> John W. Ellison, "The Use of Electronic Computers in the Study of the Greek New Testament Text" (PhD diss., Harvard Divinity School, 1957), summary 6. Ellison's thesis is quoted in this article by the kind permission of his next of kin.

<sup>17</sup> Ellison, "Computers and the Testaments," 162.

<sup>18</sup> Ellison, "The Use of Electronic Computers," 124, summary 6.

<sup>19</sup> Ellison, "Computers and the Testaments," 162.

<sup>20</sup> This was later published as IGNTP, *The Gospel according to Luke*, The New Testament in Greek 3 (Oxford: Clarendon, 1983–1987).

<sup>21</sup> Ellison, "The Use of Electronic Computers," 63–4, 67.

single manuscript, all of the places where its text differed from the *Textus Receptus* (TR).<sup>22</sup> Ellison encoded all of the differences in what was effectively a spreadsheet:<sup>23</sup> the first four columns pinpointed the unit number and the position of the variant within the unit.<sup>24</sup> The fifth column identified the variant reading and the final two columns were used to classify the variants into one of eight categories: substitutions, omissions, additions, inversions, spelling of proper nouns, changes of inflectional form, itacisms, and other spelling variations.<sup>25</sup> The computer was then used to compare each pair of manuscripts to see whether they agreed or disagreed in their deviation from or agreement with the TR, and to count the number of times they did in each of the variant categories. The output from the computer took the form of a table for each manuscript showing the number of differences between it and all of the other manuscripts subdivided by the nature of the variation.<sup>26</sup>

By analysing the patterns of agreements and disagreements in these tables Ellison found that 307 of the 309 manuscripts used in his study fell into the same groupings established by the older methods and that the “‘insignificant’ readings seemed to reveal as much as the ‘significant’ ones when they were all studied.”<sup>27</sup> Ellison’s work is often cited in literature discussing the use of computers for textual criticism,<sup>28</sup> but it does not seem to have been taken into consideration in the development of the profiling methods that were later developed for New Testament manuscripts,<sup>29</sup> perhaps to Ellison’s frustration. Several years after the

<sup>22</sup> The IGNTP used reproductions of an edition published by Oxford University Press in 1873.

<sup>23</sup> The technology available meant that it was actually encoded and stored as a seven digit number. Ellison, “The Use of Electronic Computers,” 43.

<sup>24</sup> The section of text had been split into 65 units by Parvis, and this was adopted by Ellison for his input scheme. Ellison “The Use of Electronic Computers,” 67.

<sup>25</sup> Ellison also had a system for recording variations within variations which is left out here for simplicity. Ellison, “The Use of Electronic Computers,” 42–4.

<sup>26</sup> Ellison, “Computers and the Testaments,” 164. The data calculated by Ellison, and on which his conclusions are based, is essentially the same data that would be used in the Coherence Based Genealogical Method (CBGM) to establish pargenealogical coherence; the categories used by Ellison would even allow spelling variants to be excluded from the calculation. See Tommy Wasserman and Peter J. Gurry, *A New Approach to Textual Criticism: An Introduction to the Coherence-Based Genealogical Method* (Atlanta: SBL, 2017), 27–42.

<sup>27</sup> Ellison, “Computers and the Testaments,” 165–66.

<sup>28</sup> Kraft, “The Use of Computers,” 268, 274; Bruce M. Metzger and Bart D. Ehrman, *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 4<sup>th</sup> ed. (Oxford: Oxford University Press, 2005), 240.

<sup>29</sup> It is not possible to say for certain why this is the case. Part of the agreement made with the IGNTP regarding use of their data was that Ellison was not allowed to publish his results until the edition of Luke had been published (Ellison, “The Use of Electronic Computers,” 63–4). In 1950, around the time Ellison was starting his work the expected publication date for the Luke volume was 1954 (American Editorial Committee of the IGNTP,

completion of his doctoral work he writes that “the scholars who are most involved in textual studies at present feel that the subjective intuition of the scholar is going to make the greatest discoveries. They are still at work with the older methods, essentially counting on their fingers!”<sup>30</sup> Nevertheless, the themes present in Ellison’s work (collaboration, agreed standards, data sharing and reuse, and computational techniques) are still key to the IGNTP today in its work towards the *Editio Critica Maior* (ECM).

Computers have played an important part in the preparation of the ECM since its inception, at least as early as 1970.<sup>31</sup> There are definitely similarities with Ellison’s work but I have not been able to establish a direct line of influence. Under the direction of Kurt Aland at the Institut für Neutestamentliche Textforschung (INTF), and with support from Wilhelm Ott from the Tübingen computing centre and Bonifatius Fischer from the Vetus Latina Institute, work began on selecting the manuscripts for inclusion in the ECM based on collations of selected test passages.<sup>32</sup> Although the manual selection of passages was what Ellison was trying avoid, the computational task was very similar. The selected passages were manually collated and the results were then encoded for processing by the computer. Each variant reading for each of the 1,000 selected passages had a punched card which contained the identification number for the

---

“Proposed Publication of the Manuscript Evidence for the Text of the Greek New Testament,” *The Bible Translator* 1 [1950]: 171), but the volume was not actually completed until 1987 which perhaps means that Ellison’s work was not as widely known as it might otherwise have been. An article originally published in the same year Ellison was awarded his doctorate seems to suggest that the method Ellison used was copyrighted; Cook, “Automation Comes to the Bible,” 16 notes that “Mr Ellison’s copyrighted *Methods of Using Digital Computers*.” This fact also may have hampered its further use, although this seems to go against Ellison’s overall philosophy. Of course it could also be the case that the discipline was simply not ready to trust this approach, in particular the use of a full section of the text rather than passages carefully selected by experienced scholars. However, doctoral research by Andrew Edmondson at the University of Birmingham suggests that Ellison may have been correct in his proposition. Using phylogenetic software, Edmondson has shown that family groups can be established using collation data from chapters that do not contain any of the readings considered to be significant in defining the family group. See Andrew Edmondson, “Discovering New Family Trees of Biblical Manuscripts,” Poster presented at the University of Birmingham Research Poster Conference (16 June 2015).

<sup>30</sup> Ellison, “Computers and the Testaments,” 166.

<sup>31</sup> Bonifatius Fischer, “The Use of Computers in New Testament Studies, with Special Reference to Textual Criticism,” *JTS* 21/2 (1970): 307–308; Kurt Aland, “Novi Testamenti Graeci Editio Maior Critica,” *NTS* 16/2 (1970): 163–77.

<sup>32</sup> There is no published record of exactly how these passages were selected. Richards describes the details he managed to find in W. Larry Richards, “An Analysis of Aland’s Teststellen in 1 John,” *NTS* 44 (1998): 27–30.

passage, the text of the reading and all of the witnesses that attest the reading.<sup>33</sup> As with Ellison's work, the computer was used to calculate the similarities between the text of the different manuscripts and these calculations form the basis of the selection of witnesses for inclusion in the ECM.<sup>34</sup> A similar workflow was originally planned for the preparation of the final apparatus, although by the time the processing was actually done, desktop computers and databases were available and were used instead.<sup>35</sup>

Several computer programs existed by the early 1970s which could handle machine collation of texts, but none could have coped with the complexity of the New Testament tradition.<sup>36</sup> As early as the 1950s, Ellison proposed a system in which several component collation programs could be linked together and run in a single operation.<sup>37</sup> Ott took a similar approach in the 1970s,<sup>38</sup> but it was not really until personal computers began to take over from main frames in the 1980s that the idea of fully-integrated software support for critical editing could be considered. Ott suggests that it is important "that the software used [1] cover all the steps of critical editing [2] is flexible enough to be adapted to the user's respective requirements [and 3] be a tool that can be handled easily and safely by people who are...non-programmers."<sup>39</sup>

---

<sup>33</sup> Wilhelm Ott, "Computers and Applications in Textual Criticism," in *The Computer and Literary Studies*, ed. A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith (Edinburgh: Edinburgh University Press, 1973), 201.

<sup>34</sup> For examples of the output formats see Ott, "Computers and Applications," 207–23. The full results of the analysis are also published in the volumes of *Text und Textwert*. The data processing of the test passages was begun by Ott at the Tübingen computing centre, but was later transferred to the Münster computer centre with the programming completed by Annette Benduhn-Mertz under the academic direction of Gerd Mink. See Kurt Aland, "Vorwort," in *Text und Textwert der griechischen Handschriften des Neuen Testaments, I. Die Katholischen Briefe, Band 1: Das Material*, ed. K. Aland (Berlin: De Gruyter, 1987), VII.

<sup>35</sup> Ott, "Computers and Applications," 202; Fischer "The Use of Computers," 307; Klaus Wachtel, "Editing the Greek New Testament on the Threshold of the Twenty-First Century," *Literary and Linguistic Computing* 15/1 (2000): 43–4 and personal communication.

<sup>36</sup> For a summary of the programs available, see Penny Gilbert, "Automatic Collation: A Technique for Medieval Texts," *Computers and the Humanities* 7/3 (1973): 139–47; for an idea of some of the problems with the programs see Robert L. Oakman, "The Present State of Computerized Collation: A Review Article," *Proof: The Yearbook of American Bibliography and Textual Studies* 2 (1972): 333–48.

<sup>37</sup> Ellison, "The Use of Electronic Computers," 123–24

<sup>38</sup> Julianne Nyhan and Andrew Flinn, "The University Was Still Taking Account of Universitas Scientiarum: Wilhelm Ott and Julianne Nyhan," in *Computation and the Humanities: Towards an Oral History of Digital Humanities*, ed. J. Nyhan and A. Flinn (Berlin: Springer, 2016), 66–7.

<sup>39</sup> Wilhelm Ott, "Software Requirements for Computer-aided Editing," in *Editing, Publishing, and Computer Technology: Papers given at the Twentieth Annual Conference on Editorial Problems, University of Toronto, 2–3 November 1984*, ed. S. Butler and W. P. Stoneman (New York: AMS Press, 1988), 103.

Peter Shillingsburg suggests similar system requirements, writing that a good computer assisted process should be “an integrated system in which the output from any one stage of computer work is usable as input for the next stage”; that it should cover all stage of editing from examining source texts to typesetting and printing; that all the mechanical and repetitive work is done by the computer; and that “every stage in the process is interruptible, reviewable, revisable, and, if necessary, repeatable.”<sup>40</sup> This last point is equivalent to Ott’s second point concerning flexibility, as Shillingsburg emphasises:

Editing is an art, not a science, and computers are only machines. As editors, we do not want our decisions to be influenced by the limitations of the tools we use. Any artist or craftsman worth his salt who lacks a tool to do a certain thing gets or makes one that will.<sup>41</sup>

To summarise: what is required is a fully integrated tool which is completely flexible and configurable by each editor to suit their particular editorial preferences without the need to do any programming, and that can deal with any textual tradition and support all stages of the editorial process from analysing the source texts to typesetting the final apparatus. That is a wonderful concept, but (although not everyone has given up on it) it is an ideal, which I and others do not consider to be achievable.<sup>42</sup> Pierazzo suggests that “it is impossible to develop a single framework that will satisfy all the possible use cases that characterize the variegated panorama of textual scholarship.”<sup>43</sup> And,

[w]hile it is (relatively) simple to model the way one scholar works, or even a small group, it is much more complicated to model the working methods of a large community of textual scholars, due to the different theoretical approaches to their workflow, the different types of editorial product they aim to produce, and their national and disciplinary habits and idiosyncrasies.<sup>44</sup>

Joris van Zundert suggests that such frameworks are “waiting for a horde of uniformly behaving humanities scholars that will never come”<sup>45</sup> and that “the very wish to cater to everyone pushes the designers toward generalization, and thus necessarily away from delivering data models specific enough to be useful to

---

<sup>40</sup> Peter L. Shillingsburg, *Scholarly Editing in the Computer Age: Theory and Practice*, 3<sup>rd</sup> ed. (Ann Arbor: University of Michigan Press, 1996), 138–39.

<sup>41</sup> Shillingsburg, *Scholarly Editing in the Computer Age*, 139.

<sup>42</sup> Peter Robinson’s Textual Communities platform seems to be aiming for something along these lines. See Peter Robinson, “Some Principles for Making Collaborative Scholarly Editions in Digital Form,” *Digital Humanities Quarterly* 11/2 (2017): §38.

<sup>43</sup> Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods* (Farnham: Ashgate, 2015), 116.

<sup>44</sup> Pierazzo, *Digital Scholarly Editing*, 111.

<sup>45</sup> Joris van Zundert, “If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities,” *Historical Social Research/Historische Sozialforschung* 37/3 (2012): 165–66.



anyone.”<sup>46</sup> This tension is commonly found when computers are introduced into the humanities. It is difficult to strike a balance between the usability and reusability of technical components and the unique character and requirements of each project, between generalizability and specificity,<sup>47</sup> between standards and innovation.

In a book review from 1980, Lou Burnard observes that one of the first discoveries “the novice in the art of literary computing always makes is that it has all been done before. The reinvention of the wheel, and its proud presentation at international symposia, is a familiar event.”<sup>48</sup> And that is, of course, exactly what I am doing now. The paradigm shift for the IGNTP and the ECM happened long before any of the interfaces discussed below existed. Since 1997 the editorial teams have been committed to a digital workflow based on full text transcriptions. For transcription and collation they used Collate2, one of the few integrated systems that attempted to strike the balance between the general and the specific.<sup>49</sup> In 1997 Collate had been under development for about a decade and Collate2 had been released the previous year, yet it still had to be extended to accommodate the requirements of the New Testament.<sup>50</sup> With support from Peter Robinson, the INTF team successfully used Collate2 to produce the ECM of 1 John (2003), the other Johannine Epistles and Jude (2005), and Acts (2017).<sup>51</sup> When it became clear that Collate2 would need to be replaced, a number of projects were initiated for the purpose. CollateX was developed as part of the *Interedition* project to replace the core collation functionality, and the Workspace for Collaborative Editing project addressed the interface elements required for the ECM.<sup>52</sup>

---

<sup>46</sup> van Zundert, “If You Build it,” 172.

<sup>47</sup> Max Kemman and Martijn Kleppe, “User Required? On the Value of User Research in the Digital Humanities,” in *Selected Papers from the CLARIN 2014 Conference, October 24–25, 2014, Soesterberg, The Netherlands* (Linköping: Linköping University Electronic Press, 2015), 71.

<sup>48</sup> Lou D. Burnard, “At Home with the Hardware,” *The Times Literary Supplement* 4024 (1980): 533.

<sup>49</sup> Peter Robinson, *Collate: Interactive Collation of Large Textual Traditions, Version 2* (Computer Program distributed by the Oxford University Centre for Humanities Computing: Oxford, 1994). Wilhelm Ott’s TUSTEP would be another example. See [https://tustep.uni-tuebingen.de/tustep\\_eng.html](https://tustep.uni-tuebingen.de/tustep_eng.html) [accessed 25 July 2018].

<sup>50</sup> D. C. Parker, “Electronic Religious Texts: The Gospel of John,” in *Electronic Textual Editing*, ed. L. Burnard, K. O’Brien O’Keefe, and J. Unsworth (New York: Modern Language Association of America, 2006), 202–203.

<sup>51</sup> H. A. G. Houghton et al., “The *Editio Critica Maior* of the Greek New Testament: Twenty Years of Digital Collaboration,” *Digital Philology* (forthcoming).

<sup>52</sup> Interedition was funded by a European Science Foundation COST action from 2008–2012 and involved many partners including ITSEE and the INTF. The Workspace for Collaborative Editing was jointly funded by the AHRC and the DFG from 2010–2013.

In what follows I will focus on the transcription and collation stages of the digital workflow, since these are the stages that were most transformed by its adoption.<sup>53</sup> When the IGNTP produced its edition of Luke's Gospel transcription and collation were a single task. The only words of a manuscript that were transcribed were those that differed from the text with which they were being collated.<sup>54</sup> These 250 or so collations then had to be combined to make the edition, not an easy task on paper even with some actual cut-and-paste.<sup>55</sup> Since 1997, there have been two distinct stages: first, each manuscript has been transcribed in full (although this actually entails adapting a base text to reflect the text in the witness) and then all of the transcriptions are automatically collated and the result edited for publication. The tools used to support these two stages of the workflow will be discussed in this paper. Each has taken a different approach to resolving the tension between the editorial task and the computer. Their introduction has not changed the workflow; they are just new wineskins for the same old data.

## Transcription

The use of computers in classics, although it has made great strides, is still faced with the fact that before almost any kind of work can be done, a corpus must be prepared so that a machine can understand it; optical scanners hold hope for the future, but at the present, we are still at a stage where much manual labor is necessary. It is to be hoped that duplication of such comparatively tedious work can be avoided as far as possible, leaving time free for developing more interesting and imaginative applications of computers to classical studies (1968).<sup>56</sup>

Transcription is still the first task for many digital editing projects. While the optical scanners predicted above are now widely used for printed texts, nearly 50 years on from this quotation the situation for manuscripts is not radically different.<sup>57</sup> Data entry is at least easier now than it once was. If errors were made in

---

<sup>53</sup> For a description of the full workflow see H. A. G Houghton and C. J. Smith, "Digital Editing and the Greek New Testament," in *Ancient Worlds in Digital Culture*, Digital Biblical Studies 1, ed. C. Clivaz, P. Dilley, and D. Hamidović (Leiden: Brill, 2016), 110–27.

<sup>54</sup> For a detailed description of how manual collations are made see D. C. Parker, *An Introduction to the New Testament Manuscripts and Their Texts* (Cambridge: Cambridge University Press, 2008), 95–100.

<sup>55</sup> Ulrich Schmid, "Transmitting the New Testament Online," in *Text Comparison and Digital Creativity: The Production of Presence and Meaning in Digital Text Scholarship*, ed. W. Th. van Peursen, E. Thoutenhoofd, and A. van der Weel (Leiden: Brill, 2010), 193–95.

<sup>56</sup> Stephen V. F. Waite, "Computers and the Classics," *Computers and the Humanities* 3/1 (1968): 29.

<sup>57</sup> The READ project (<https://read.transkribus.eu>) is working with many partners to improve Handwritten Text Recognition (HTR) using the tool Transkribus. While this has potential to change the situation in the future, the large amount of training data required means



transcriptions on paper tape, the options available for correcting them were very similar to those used by scribes in manuscripts. The typists could either move back to the problem character and overwrite it with a special delete character that the computer would ignore (rather like adding a dot above a letter) or type a special character at the end of the line to show that the whole line should be ignored, followed by a number to ignore multiple lines (rather like the inverted commas that can be found in some manuscripts marking complete deletion).<sup>58</sup> With electronic data entry and storage the practical challenges of transcribing have eased, but at the same time our expectations of transcriptions, and by extension transcribers, have grown.

TEI has become the *de facto* means of encoding text transcriptions in the humanities and it encapsulates the tension between standardisation and flexibility. People have pointed out that – despite its subtitle – the flexibility allowed within the TEI guidelines means that it does not work as a format for interchange or interoperability and also makes it complex to use as an encoding tool.<sup>59</sup> The TEI guidelines encompass such a wide range of possibilities that, as John Lavagnino says, if you chose the TEI for your project, that is not the end of choice; “there are further questions about exactly how you use it.”<sup>60</sup> Not only must each project decide what features of their manuscripts to encode but also define the subset of TEI they will use to encode them and how that subset is to be applied. Andrews describes an Interedition-sponsored thinktank at the 2011 TEI Members’ Meeting about interoperability within the TEI. She reports:

It became clear over the course of the meeting that flexibility and customizability is currently much more important to textual scholars than the sort of standardization that would allow for true progress toward digital critical editions.<sup>61</sup>

TEI is often referred to as a standard. Indeed, this is mentioned in the first of the Poughkeepsie principles which formed the basis of the TEI, which states: “The guidelines are intended to provide a standard format for data interchange in

---

it is unlikely to be a practical solution for the New Testament where we already have a reliable text on which to base our transcriptions.

<sup>58</sup> Wilhelm Ott, “Transcription and Correction of Texts on Paper Tape. Experiences in Preparing the Latin Bible Text for the Computer,” *Revue Informatique et Statistique dans Les Sciences Humaines* 2 (1970): 57–8.

<sup>59</sup> See for example Desmond Schmidt, “Towards an Interoperable Digital Scholarly Edition,” *Journal of the Text Encoding Initiative* 7 (2014), <https://journals.openedition.org/jtei/979> [accessed 9 March 2018] and Martin Holmes, “Whatever Happened to Interchange?” *Digital Scholarship in the Humanities* 32 suppl\_1 (2017): i63–i68.

<sup>60</sup> John Lavagnino, “When Not to Use TEI,” in *Electronic Textual Editing*, ed. L. Burnard, K. O’Brien O’Keefe, and J. Unsworth (New York: Modern Language Association of America, 2006), 335.

<sup>61</sup> Tara L. Andrews, “The Third Way,” *Variants* 10 (2013): 63.

humanities research.”<sup>62</sup> However the ability for scholars to customise the guidelines is also one of the founding design principles.<sup>63</sup> Ide and Sperberg-McQueen (an editor of the initial TEI guidelines), describe the situation as follows:

The TEI made an early commitment to formulating its encoding scheme as a set of *guidelines*, rather than as a *standard*, for encoding literary and linguistic materials. This reflects first of all a commitment to preserving the intellectual autonomy of researchers who encode texts electronically: by constraining the allowable forms and intellectual content of electronic texts, a strictly normative standard would risk constraining their ability to reflect the particular intellectual commitments of individual researchers, and thus in some ways the types of intellectual work which can conveniently be undertaken. A wholly permissive encoding scheme, on the other hand, risks failing to provide a usable basis for the sharing and reuse of expensive textual resources. It also encourages pointless variation in encoding methods arising not from different intellectual approaches to textual research, but from merely random differences in the use of the scheme.<sup>64</sup>

They go on to say that “[t]he TEI has tried to steer a middle course.”<sup>65</sup> It could be argued that this middle course has still failed when it comes to interchange, but the breadth and flexibility offered by the guidelines is surely the reason that TEI has become so widely used. Before the ECM teams began using TEI, the importance of flexibility within agreed parameters as a means to facilitate collaboration had already been recognised, as David Parker notes:

The use of agreed computer methodologies makes it possible for separate projects to work together without losing their identity. Thus, IGNTP and INTF have been able to share certain common objectives without losing their individuality. This balance is important, for if everyone did things exactly the same way, textual editing would be greatly impoverished.<sup>66</sup>

While the customisation options of the TEI are an important factor in its popularity within the humanities, they do cause problems for the creation of generic tools and software that can support its use.<sup>67</sup> The use of XML as the current expression of TEI means that all of the standard XML editing tools can be used: these tools can make XML easier to work with by providing syntax highlighting and validation support. However, using them still requires a very high level of XML and TEI knowledge and a lot of discipline on the part of the transcriber.

---

<sup>62</sup> James Cummings, “The Text Encoding Initiative and the Study of Literature,” in *A Companion to Digital Literary Studies*, ed. R. Siemens and S. Schreibman (Oxford: Blackwell, 2008), 453.

<sup>63</sup> Cummings, “The Text Encoding Initiative,” 472.

<sup>64</sup> Nancy M. Ide and C. Michael Sperberg-McQueen, “The TEI: History, Goals, and Future,” *Computers and the Humanities* 29/1 (1995): 8.

<sup>65</sup> Ide and Sperberg-McQueen, “The TEI,” 8.

<sup>66</sup> Parker, “Electronic Religious Texts,” 205. The importance of the electronic workflow, in particular the transcription format, for aiding collaboration is echoed in Wachtel, “Editing the Greek New Testament,” 48.

<sup>67</sup> Pierazzo, *Digital Scholarly Editing*, 112, 119–20.

When a critical edition is created electronically, the transcriptions provide the foundation on which everything else is built: it is impossible to create a good quality edition from poor quality transcriptions. However, as this is one of the most time consuming phases it is also the most likely to be crowd-sourced and that requires careful consideration of the tools used. *Transcribe Bentham*, the flagship crowd-sourcing project in this area, has been actively engaging volunteers to make transcriptions of Bentham's writings, including many unpublished manuscripts, since 2010.<sup>68</sup> While not compulsory, the use of basic TEI markup in the transcriptions is encouraged and a transcription toolbar is provided to help users add the appropriate tags.<sup>69</sup> The toolbar makes it easier for transcribers to add tags and the team found that many people did choose to add the TEI annotation, in particular the "super-transcribers" (the small percentage of volunteers that produce the vast majority of the work).<sup>70</sup> However, they also found that the visual presence of the tags in the transcription interface was putting off potential volunteers. Consequently, as part of the READ project, the *Transcribe Bentham* team are working on a "what-you-see-is-what-you-get" (WYSIWYG) editor to replace the existing toolbar so that in future the XML tags will be hidden.<sup>71</sup> They predict that having such an editor will increase the rate at which transcriptions are made as well as improving recruitment and retention of volunteers and therefore the investment in creating the editor will be worth it.<sup>72</sup>

This idea of a payoff of investment is an important point and goes hand in hand with the tension around specific versus general requirements. The use of the toolbar and the introduction of the WYSIWYG editor necessarily restricts the tags that can be used but, for *Transcribe Bentham*, the restriction of tags and the time invested to make the editor are both compromises worth making in order to achieve their goals.

Another system currently in development that takes a very different approach to transcription is the Textual Communities platform.<sup>73</sup> Textual Communities is

<sup>68</sup> Details of the *Transcribe Bentham* project can be found at <http://blogs.ucl.ac.uk/transcribe-bentham> [accessed 4 March 2018].

<sup>69</sup> Tim Causer and Melissa Terras, "Crowdsourcing Bentham: Beyond the Traditional Boundaries of Academic History," *International Journal of Humanities and Arts Computing* 8/1 (2014): 52.

<sup>70</sup> Causer and Terras, "Crowdsourcing Bentham," 49, 52.

<sup>71</sup> Tim Causer and Valerie Wallace, "Building a Volunteer Community: Results and Findings from *Transcribe Bentham*," *Digital Humanities Quarterly* 6/2 (2012): §62–4; Louise Seaward, "The Crowd, the Volunteers and the Supertranscribers. Building and Supporting an Online User Community for the Bentham Edition," presentation read at the co:op READ Convention Marburg 2016. Available online at <https://coop.hypotheses.org/224> [accessed 4 March 2018].

<sup>72</sup> Tim Causer, Justin Tonra, and Valerie Wallace, "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *The Collected Works of Jeremy Bentham*," *Literary and Linguistic Computing* 27/2 (2012): 131.

<sup>73</sup> Available at <http://www.textualcommunities.org> [accessed 25 July 2018].

designed to be a general platform to support all stages of the editorial process allowing anyone to make a digital edition without specialist technical support.<sup>74</sup> When discussing transcription, Robinson explains that “this system does not attempt to hide the XML: we think it helpful for the editor and transcribers to see exactly what encoding is being applied to the document.”<sup>75</sup> Despite this positive spin, the prominence of the XML in the editor is part of the compromise inherent in trying to develop a transcription tool that works for any edition using the full range of TEI.<sup>76</sup> In order to use this generic tool successfully projects must invest time and resources into mastering TEI-XML and passing these skills on to their transcribers. An excellent example of this is the work of the *Estoria de Espanna* (*Estoria*) project, one of the six projects that began using the Textual Communities system as a trial before its formal release.<sup>77</sup> In addition to the transcriptions made by members of the *Estoria* team, they had some success with a small crowd sourcing trial using the Textual Communities transcription interface.<sup>78</sup> While the team did find that the XML initially put off some potential volunteers, carefully-staged training materials that gradually exposed volunteers to the XML annotations required were successful in encouraging participation.<sup>79</sup> The materials produced to support this endeavour are extensive and represent a considerable investment by the project.<sup>80</sup> In addition to the development of the supporting material, time also had to be invested in providing feedback to volunteer transcribers, which the project team admits may cause problems if the crowd-sourcing scheme were to be expanded in the future.<sup>81</sup>

The IGNTP was using volunteers to make transcriptions long before the term crowd-sourcing even existed. Volunteers helped produce the IGNTP edition of Luke’s Gospel (published in two volumes in 1984 and 1987) by making collations of manuscripts. Photographic images of the manuscripts would be posted

---

<sup>74</sup> Peter Robinson, “Some Principles for Making Collaborative Scholarly Editions,” §38. The project was also partly inspired by the growing interest in web-based collaborative editing and crowd sourcing. See Peter Robinson “The background to the Textual Communities Project,” accessible at <http://wiki.usask.ca/display/TC/The+Background+to+the+Textual+Communities+Project> [accessed 25 July 2018].

<sup>75</sup> Robinson, “Some Principles for Making Collaborative Scholarly Editions,” §40.

<sup>76</sup> In order to support the full scope of TEI Robinson supports any XML in the Textual Communities transcription interface. Robinson, “Some Principles for Making Collaborative Scholarly Editions,” §40.

<sup>77</sup> Robinson, “Some Principles for Making Collaborative Scholarly Editions,” n.17. The *Estoria de Espanna* was led by Aengus Ward at the University of Birmingham (<http://estoria.bham.ac.uk>, accessed 4 March 2018).

<sup>78</sup> Polly Duxfield, “Transcribing the *Estoria de Espanna* Using Crowdsourcing: Strategies and Aspirations,” *Magnificat Cultura i Literatura Medievals* 2 (2015): 131.

<sup>79</sup> Duxfield, “Transcribing the *Estoria de Espanna*,” 142–44.

<sup>80</sup> The canvas course created by the team on the Canvas platform can be seen here: <https://canvas.bham.ac.uk/courses/6673> [accessed 4 March 2018].

<sup>81</sup> Duxfield, “Transcribing the *Estoria de Espanna*,” 145.

out to the volunteers who would make a collation of the witness against the agreed base text and then post their collations back to the editorial team. When full text electronic transcriptions replaced individual collations, the IGNTP volunteers along with a team of paid students and staff in Münster and Birmingham were able to make transcriptions in plain text files using the relatively simple Collate2 markup scheme.<sup>82</sup> However, with the move to XML both the IGNTP and INTF teams felt that it was too much to ask transcribers to be able to work directly in XML and that the quality of the transcriptions would be adversely affected by the added visual distraction of XML tags. As the New Testament has such a long transmission history and so many transcriptions remain to be made for the ECM, it was worth investing the time and resources to create a bespoke transcription tool that removed the need for transcribers to learn XML. As part of the Workspace for Collaborative Editing project developers Martin Sievers and Gan Yu from the Center for Digital Humanities at the University of Trier produced the Online Transcription Editor (OTE). A screenshot of the OTE within the Workspace can be seen in Figure 1.

The OTE was not written from scratch but is an extension to the open-source rich text editor TinyMCE.<sup>83</sup> It is possible to transcribe the full text into the editor but for the New Testament we always use a base text which the transcriber modifies to match the text of the manuscript.<sup>84</sup> For simple annotations such as adding line breaks shortcuts are included: whenever the return key is pressed a line break will be added. For more complex annotation, transcribers use a series of menus and submenus to add the annotation to the text without ever encountering any XML. Mouseovers inform the user of any additional details: for example, the text that can be seen in square brackets in Figure 1 is text that has been supplied by the transcriber. The mouseover (shown in Figure 2) informs the user of the reason the text has been supplied and the source from which the supplied characters were taken. To add the annotation, the user highlights the section of text involved and then clicks on the appropriate menu.

---

<sup>82</sup> The transcription guidelines for the Collate format used by the Principio Project are available online. D. C. Parker et al., *The Principio Project: Procedures for Transcribing Witnesses* (2001), <http://epapers.bham.ac.uk/2948>, 9–17 [accessed 4 March 2018].

<sup>83</sup> <https://www.tinymce.com> [accessed 4 March 2018].

<sup>84</sup> H. A. G. Houghton, “Electronic Transcriptions of New Testament Manuscripts and their Accuracy, Documentation and Publication,” in *Ancient Manuscripts in Digital Culture: Visualisation, Data, Communication*, Digital Biblical Studies 3, ed. D. Hamidović, C. Clivaz, and S. Savant (Leiden: Brill, forthcoming).

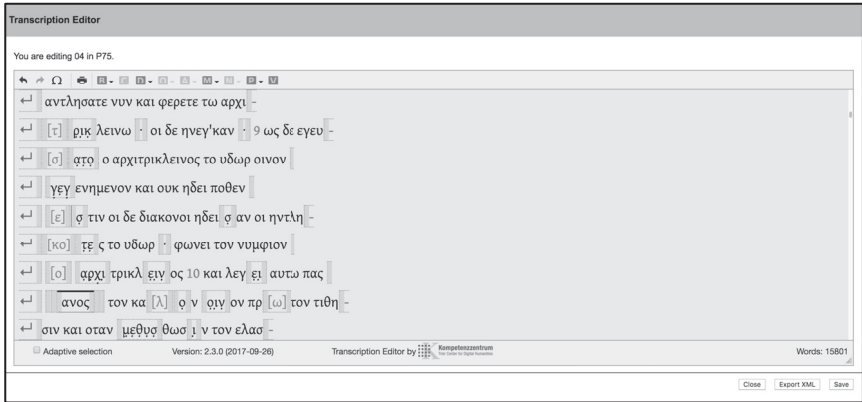


Fig. 1. Screenshot of the OTE within the Workspace for Collaborative Editing

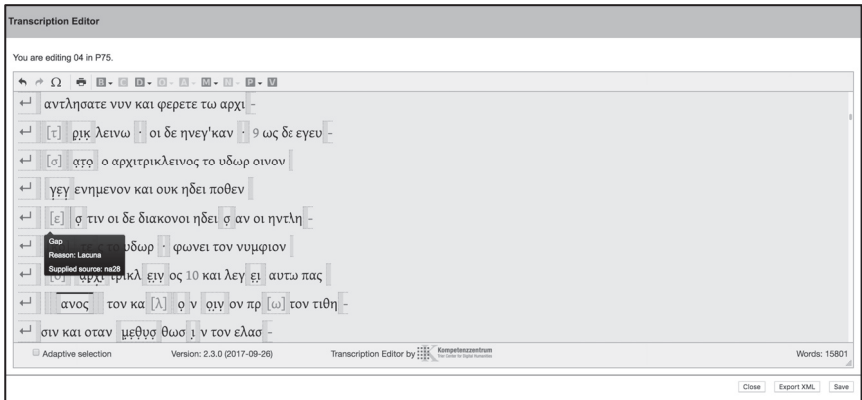


Fig. 2. Example of mouseover in the OTE

This will sometimes directly add the annotation but at other times will open a form to allow further details to be selected or entered. For more complex annotations that require further text to be added, such as indicating a correction, a version of the OTE is provided within the form so that the same editor can be used to add annotation to the text of the correction. The correction form is shown in Figure 3.<sup>85</sup>

<sup>85</sup> The user experience of OTE is described in more detail in Darius Müller, “Zur Elektronischen Transkription von Apokalypsehandschriften: Bericht zum Arbeitsstand,” in *Studien zum Text der Apokalypse II*, ANTF 50, ed. M. Sigismund and D. Müller (Berlin: De Gruyter, 2017), 19–30.

Fig. 3. Form used to add corrections

The OTE only supports the particular customisation of the TEI used by the INTF and the IGNTP for the production of transcriptions for the ECM. While some of the customisation, such as adding corrections, might be applicable to other traditions, some features are specific to biblical texts. For example to establish the structure of the text books, chapters, verses and lections are used. Other supported features such as the annotation of *nomina sacra*, Eusebian canons, and Ammonian sections are also specifically biblical in their nature. The editor for the Greek New Testament will undoubtedly be lacking support for features common in other traditions. For example, researchers from the Multimedia Yasna project are working with the same developers to produce a similar version of the OTE to support their transcriptions of Avestan Yasna. Both the Avestan language and the text of the Yasna require further customisation of the editor which would be irrelevant for the Greek text of the New Testament in the same way that many of the features in the New Testament editor are not relevant to the Yasna project.

Developing the tool was an investment of time and resources, but it would have taken a greater amount of time to train all of our transcribers to become fluent in TEI, and the need to learn this encoding might even have put off some transcribers (as the *Transcribe Bentham* project found). Using the editor does not mean that our transcribers require no training at all, but it does mean that their training can focus on palaeographical skills and the correct identification and annotation of the features of the manuscript. For smaller projects with fewer transcriptions to make, or fewer transcribers to teach, a more generic tool like that in Textual Communities might be more suitable: in that case, investment would instead need to be put into the training and/or learning of TEI and XML. There is a balance that must be found and each editing project will find it in a different place.



## Collation

The collation of manuscripts requires the infuriating accuracy of a pedant and the obsessive stamina of an idiot. It is therefore an ideal task for a computer (1989).<sup>86</sup>

One of the first elements of Collate to be replaced was the core collation algorithm. CollateX was written by Ronald Haentjens Dekker as part of the Interedition project and is specifically intended to be a general tool. The design of the program follows the Gothenburg model (also devised by the Interedition project team) in its separation of the collation process into distinct stages: tokenisation, regularisation, alignment, analysis, and visualisation.<sup>87</sup> It is intended that any of the stages can be taken over by another service, should that be necessary or desirable. As a result CollateX can receive input in many formats, has a number of settings that can be used to change its behaviour, and offers several output formats. The flexibility of CollateX is a great asset and is an important part of its appeal to a broad user base. But as a command line tool it is not particularly easy for non-programmers to use. The Collation Editor is a wrapper around CollateX that simplifies some of the interaction. In terms of the Gothenburg model, the Collation Editor uses the alignment functions provided by CollateX, along with the regularisation mechanism provided by the JSON input format. It provides its own visualisation and analysis tools allowing the user to regularise tokens, correct the automatic alignment, set variant lengths and order the readings in each variant unit and control certain CollateX settings, such as selecting the alignment algorithm used and changing the settings for the fuzzy matching of strings. Tokenisation, is left to the user (or the platform in which the collation editor is running), since the JSON input format used requires the data to be pre-tokenised and the preparation of the tokenised data is, to an extent, dependent on the collation editor settings required.

Matthew Kirschenbaum suggests that, although interfaces in the Digital Humanities are the primary point of interaction for users, they are often the last thing to be written (normally at a point when the project is running out of both time and money).<sup>88</sup> This was certainly not the case with the collation editor. We already had the computationally difficult aspect in CollateX: what was needed was an easy way for editors to interact with the software and the data, making this an interface design project from the start. That is not to say that this was just a project about making things look nice. In order for the collation editor to be fit for

---

<sup>86</sup> Peter Robinson, "The Collation and Textual Criticism of Icelandic Manuscripts (1): Collation," *Literary and Linguistics Computing* 4/2 (1989): 99.

<sup>87</sup> See the CollateX Documentation available online at <https://collatex.net/doc/> §1 [accessed 4 March 2018].

<sup>88</sup> Matthew G. Kirschenbaum, "'So the Colors Cover the Wires': Interface, Aesthetics, and Usability," in *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth (Malden: Blackwell, 2004), 525.



purpose, it had to encapsulate the complex process of preparing a critical edition and that involved understanding and modelling the process alongside building the interface. The role of the computer in scholarly editing was summarised by Thomas Howard-Hill in 1973 and his conclusions are still relevant to the design of editorial tools today.

The computer is to be the servant of the editor, and the system is to enable the editor to concentrate where his attention is most required, on problems of textual and critical analysis. No computer-based system which distracts the editor's attention from the goal of his activity to the means of achieving it is likely to find acceptance.<sup>89</sup>

This view aligns with both our policy of hiding the XML from our transcribers and with our approach to designing the collation.

Deciding on the layout for the collation output was central to the design of the Collation Editor. CollateX itself offers a number of output options including alignment tables with one column (or row) per witness, a JSON based table structure and TEI compliant XML. In addition it provides an SVG serialisation of its internal data structure, a directed acyclic graph.<sup>90</sup> In this view, the words of the texts are shown in the nodes in the graph and the lines that connect them (in graph terminology, the edges) are labelled with the witnesses. The text of a witness can be read by following the path using the connecting lines and reading every node the path travels through. This visualisation has also been extended to provide a means of editing the collation output, such as regularising word forms or joining words together into a single variation unit.<sup>91</sup> While Stefan Jänicke and his collaborators consider the readability of the CollateX variant graph to be an improvement on previous proposals they suggest that there is still room for improvement.<sup>92</sup> They propose five “design rules” to improve the readability of the CollateX variant graph which they have implemented in their javascript library TRAViz.<sup>93</sup> The rules proposed are: use font size to give an indication of the number of witnesses attesting a particular reading (the larger the font the more witnesses read the word); avoid “backwards edges” so that all texts read fully from

---

<sup>89</sup> Thomas H. Howard-Hill, “A Practical Scheme for Editing Critical Texts with the Aid of a Computer,” *Proof: The Yearbook of American Bibliography and Textual Studies* 3 (1973): 336–37.

<sup>90</sup> Ronald Haentjens Dekker and Gregor Middell, “Computer-Supported Collation with CollateX,” conference paper at *Supporting Digital Humanities 2011: Answering the Unaskable*. Copenhagen, Denmark (2011).

<sup>91</sup> Tara L. Andrews and Joris van Zundert, “An Interactive Interface for Text Variant Graph Models,” *Digital Humanities* (2013) <http://dh2013.unl.edu/abstracts/ab-379.html> [accessed 4 March 2018].

<sup>92</sup> Stefan Jänicke, Annette Geßner, Greta Franzini, Melissa Terras, Simon Mahony, and Gerik Scheuermann, “TRAViz: A Visualization for Variant Graphs,” *Digital Scholarship in the Humanities* 30 suppl\_1 (2015): i83–i99.

<sup>93</sup> The code, tutorials and examples of the tool in action can be found on the project website <http://www.traviz.vizcovery.org/> [accessed 4 March 2018].

left to right; use colours rather than labels for edges (connecting lines); bundle the most common edges together into a single thicker line; insert line breaks to avoid horizontal scrolling. As with the CollateX variant graph, TRAViz provides editing features to detach words and realign them as well as joining nodes together. One area of improvement made by TRAViz does not seem to be covered in the rules. This is the vertical alignment of the words and the layout of the edges which combine to make the pattern of shared readings and variants immediately obvious in the outer shape of the graph: it gets wider where there are lots of variants and narrower where more witnesses agree. The examples provided in the paper and on the TRAViz website are certainly visually appealing, but as with the CollateX variant graph, once the number of witnesses starts to rise or witnesses become more divergent, it becomes harder to make sense of the data.<sup>94</sup> The use of colour also necessarily limits the number of witnesses TRAViz can meaningfully display.

The need to collate upwards of 200 witnesses at a time meant a different approach was required for the New Testament. The layout used can be seen in Figure 4. Internally, CollateX performs baseless collation and both of the visualisations already discussed above have also followed this model. In our layout a base text is used to organise the variants, although we still benefit from the results of the baseless collation. This also has the advantage of ensuring that there are no backward edges. Any of the witnesses in the collation can be used as the base text so long as it contains the verse being collated.<sup>95</sup> For the purposes of the ECM the base text used is the text of NA<sup>28</sup> (one of the standard hand editions of the New Testament) with any verses not present in the main text added (for example John 5:4). The top line of the display shows the base text with the index numbers used in the ECM (even numbers for words and odd numbers for spaces). Below this the variant readings are presented: each variant unit is in a separate box and each reading is labelled with a letter.<sup>96</sup> The *a* reading is always the reading of the base text. Lacunose and omitted readings are shown at the bottom of each unit. The witnesses that attest each reading can be seen by hovering over the reading text; the reading of a particular witness can be highlighted by selecting the siglum in the menu at the bottom of the page. The layout gives a similar overall outline to the TRAViz display, making it easy to assess the scale of the variation. We have not followed the recommendation of Jänicke et al. to use line breaks rather than horizontal scrolling. This is because we have to be able to handle overlapping variants. These are longer variant units, often ones which have a difference in word order, which are moved into secondary apparatus lines. Having line

---

<sup>94</sup> For an example of such a graph see Jänicke et al., “TRAViz,” i96 *Fig. 17*.

<sup>95</sup> The collation editor can actually display the variants if the base text is lacunose for the verse but the data cannot be edited and manipulated properly, since all of the variants are treated as additions.

<sup>96</sup> The use of letters to identify readings follows the convention used in the ECM.

breaks as well as these extra lines of apparatus would make the display much more difficult to understand.

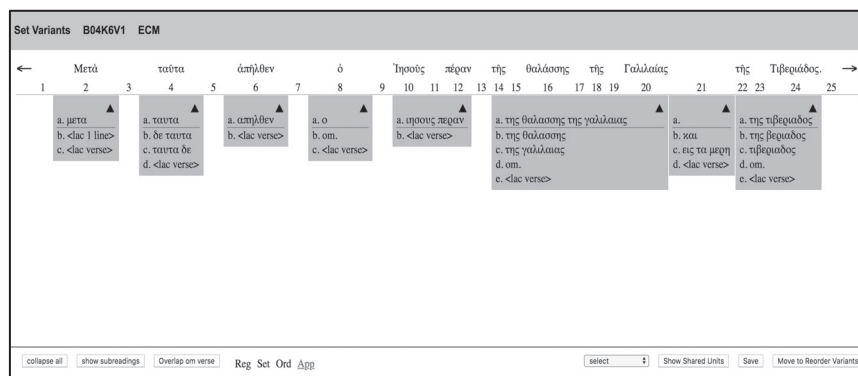


Fig. 4. Screen shot of the collation editor layout

The most important functional aspect of the Collation Editor is the ability to edit and correct the results from CollateX. Collate was the first program to treat computerised collation as an interactive process between the computer and the editor. When Robinson wrote Collate, he wrote it for the Macintosh computer, as that was the only system at the time that could reliably support what he refers to as “point and click” menus.<sup>97</sup> Although cutting edge technology at the time, these menus forced the scholar to interact at a level at least once removed from the data itself. Today there are many more options than were available in the late 1980s. The aim with the Collation Editor has been to keep the data itself central and to keep the interface as invisible as possible.<sup>98</sup> To this end the collation editor uses “drag and drop” and the right click context menu for the majority of the interaction, allowing editing to happen through direct manipulation of the text. For example, in the regularisation interface, to create a rule the user drags the word to be regularised onto the target word and this brings up a menu which allows the scope and type of the rule to be specified. When setting variants, one unit can be dragged and dropped onto an adjacent one to join them together; they can be split again by right clicking on the unit and selecting “split words” from the context menu. If words have been aligned incorrectly, this can be corrected in the set variants interface. To move an individual reading to a different location

<sup>97</sup> Peter Robinson, “A History of Collate,” *Scholarly Digital Editions* 29 September 2014: <http://scholarlydigitaleditions.blogspot.co.uk/2014/09/the-history-of-collate.html> [accessed 4 March 2018].

<sup>98</sup> This follows the advice from Negroponte who states “the secret to interface design [is to] make it go away.” Nicholas Negroponte, *Being Digital* (London: Hodder & Stoughton, 1995), 93. Originally cited in Kirschenbaum, “So the Colors Cover the Wires,” 524.

the readings are split using the right click context menu and then each reading can be moved independently. To move a reading to a new unit, it is dropped onto the appropriate unit. To move it to a space without a unit (to be treated as an addition) it is dropped on the number for the space and a new unit is created.

While there is no such thing as an intuitive interface,<sup>99</sup> this direct manipulation of the data has proved to be a successful model for interaction. This became apparent in one of our regular project meetings during a demonstration of a newly implemented solution to the overlapping variants problem. The example being used for the demonstration had been provided to me as an example of an overlapping variant at a previous meeting. Following the demonstration I was expecting either confirmation that the interface was fine or, more likely, a list of changes or new features that were required. What ensued was a long discussion of whether or not the beginning of John 6:23 should be actually treated as an overlapping variant after all. At the time this was rather frustrating, the issue at hand was not whether John 6:23 was an example of an overlapping variant, but rather whether the interface allowed the editors to record an overlapping variant in the correct way. On reflection, the response was perfect. The editors had just been introduced to a brand new feature of a still relatively unfamiliar interface and they had not noticed the interface, but only data. And that is exactly what Howard-Hill was talking about: the interface should not distract the user from the data being edited.

Initially, the Collation Editor was only intended to work within the suite of tools developed as the Workspace for Collaborative Editing. However, it became clear that other editing frameworks also needed a way for their users to interact with CollateX, so the code has been adapted to work in different architectures. There is also a standalone version that works with the local file system and requires no further infrastructure.<sup>100</sup> Broadening the user base for the editor in this way also introduced the requirement that it work with a much wider range of traditions, beyond the Greek and Latin New Testament as was the original intention. The fact that the editor was written to work with such a complex tradition and with two languages from the start, meant that some of the features had also been configurable from an early stage: examples of this include the base text and the list of witnesses to be used in each project. Other features, such as the types of regularisation rules, were made configurable early on for practical reasons of project management. For example, the editorial teams did not agree on the types of rules they wanted to use for regularisation so these were set in the configuration files on a project-by-project basis. This pattern has now been extended and

---

<sup>99</sup> J. Raskin, "Intuitive Equals Familiar," *Communications of the ACM*, 37/9 (1994): 17–8.

<sup>100</sup> This work was carried out in the summer of 2015 along with Troy Griffiths, the programmer of the NTVMR, one of the other systems which sought access to the collation editor.

all of the Greek, Latin, and New Testament specific settings have been moved into configuration files. It is hoped, having started with such a large and complex tradition, that the collation editor will only require changes to the configuration files in order to handle other traditions. Only time will tell whether that is a realistic hope.<sup>101</sup>

The result of all of these changes is that, although it will not be suitable for every edition ever made, the Collation Editor is now a more generally applicable tool. It is working in the NTVMR and in the Textual Communities platform and the standalone version was used to make the edition of the *Estoria de Espanna*.<sup>102</sup> There is, of course, a downside to the flexibility introduced into the editor by generalising it in this way. Writing the service files necessary to have it work with different architectures is the first challenge, although most projects will have technical support if they are building their own architecture which may mitigate this issue. More of a problem is the complexity of the configuration process for different traditions and editorial practices. Certain features may require additional code to be written and almost all have an impact on the way the data must be prepared for use. An additional complication is the flexibility of TEI which means that the collation editor cannot currently provide any help with the transformation of data from TEI to the JSON format which it uses. It is simply not possible to predict what features of a text a user will encode, how they will encode them, and what they expect the Collation Editor to be able to do with the result. It would of course be possible to set out rules for what TEI can and cannot be used with a set of automatic import scripts, but that would start to reintroduce use-case restrictions and also go against the principle that has made TEI so widely used. The result is that a lot of time and effort must go into working out what an editor wants to do with their edition and preparing the data and configuration files in such a way that it is all possible. In many cases the technical skills needed to configure the editor are great, perhaps more than might be required to use CollateX on the command line. That is the price to be paid for a tool, particularly one with a graphical user interface, that can be adapted to the specific requirement of multiple textual traditions and editorial approaches.

---

<sup>101</sup> Users from other traditions have already encountered frustrations around tokenisation differences between manuscripts which makes it difficult for them in the early stages of collation. While I maintain that tokenisation is not a direct responsibility of the collation editor (see the opening of this section), I accept that this might need to be addressed in future developments.

<sup>102</sup> See description above and Aengus Ward, “The Estoria de Espanna Digital: collating medieval prose – challenges...and more challenges,” *Digital Philology* (forthcoming). The edition is available here: <http://estoria.bham.ac.uk/edition/> [accessed 4 March 2018].

## Preservation and Sustainability

If I have hitherto seemed tedious, and superfluously attentive to detail, my excuse is that I have sought to present a faithful view of every characteristic of my manuscript – to faithfully edit it in short – besides the mere notation of an addition to or an omission from the *textus receptus*. The advantage of this method, combined with a really accurate collation, is obvious. It presents the whole codex to those at a distance, and does away with that never-ending *re-examination* of documents, which has been going on – at such a cost of precious time – ever since the days of Henry Stephens (1890).<sup>103</sup>

Concerns about preservation and sustainability are not new. As the quotation above shows, in 1890 H. C. Hoskier was already thinking about data reuse and was encouraging others to build on his work rather than starting afresh. With the advent of computers this became more of a practical reality: data sharing and reuse was a key motivation for Ellison in his work and he himself reused data created by others; when Fischer was making his concordance of the Vulgate, Busa (via Ott) provided a copy of the lemmatised Latin lexicon that had been created for his concordance project in exchange for information about the words used in the Vulgate, but not by Thomas Aquinas;<sup>104</sup> it was possible to purchase the Greek New Testament on punched cards or paper tape from 1969;<sup>105</sup> and in 1976 the Oxford Text Archive was established in order to preserve machine readable texts and make them available to the academic community.<sup>106</sup> However, the limited storage technology available meant that having access to these texts did not guarantee they could be easily reused for other purposes. There were also already concerns about the quality of the data. For example, an ESTS committee in 1970 called for all existing copies of tapes prepared by Andrew Q. Morton (possibly for his authorship studies) to be proofread and for a system to be established to certify which of the copies had been checked. It also pointed out that the data did not include chapter and verse references, which might be needed, and suggested that diacritical signs and “at least major textual variants” be added.<sup>107</sup> Adding additional data was not straightforward in the days of punched cards and paper tape since, as Fischer points out, “the marks can hardly be fitted in afterwards without copying the whole text again.”<sup>108</sup>

While digital storage, rather than just digital processing, has certainly made data more extensible and therefore easier to reuse, preservation remains a

---

<sup>103</sup> H. C. Hoskier, *A full Account and Collation of the Greek Cursive Codex Evangelium 604* (London: David Nutt, 1890), xxviii.

<sup>104</sup> Nyhan and Flinn, “The University Was Still Taking Account of Universitas Scientiarum,” 59.

<sup>105</sup> The purchase price for the tapes/cards was £12 in 1969 which is the equivalent of about £190 today. Fischer, “The Use of Computers,” 298.

<sup>106</sup> Hockey, “The History of Humanities Computing,” 8.

<sup>107</sup> Kenneth Grayston, “Computers and the New Testament,” *NTS* 17/4 (1971): 479–80.

<sup>108</sup> Fischer, “The Use of Computers,” 298.

challenge.<sup>109</sup> With the amount of digital data being produced it is also a growing challenge. As Pierazzo points out, it is not possible to keep everything. This makes it important to have sensible policies about what to keep and how to keep it.<sup>110</sup> One of the key elements to preservation is to preserve the data separately from its display interface and associated tools.<sup>111</sup> The data will almost certainly be useful long after the tools that made it have gone. We have already seen this with our own editions. The transcriptions made in plain text using Collate2 were saved independently of the system and were easy to convert to XML so as to be ready for the new tools.<sup>112</sup> The editions published with Anastasia, the publishing tool for Collate, were only stored within the Anastasia system and although for the most part we did manage to recover the apparatus data, it took a lot of time and some data was lost. Had the apparatus been stored separately as data and not just as part of the display system this would have been much easier to do.

The format chosen for the preservation of data is also important. As Cummings points out, despite its problems for interchange, TEI-XML is still a good format for long-term data preservation.<sup>113</sup> Pierazzo suggests that the use of any standard is a positive thing in and of itself, for a number of reasons. If more people use a particular standard there will be more interest in ensuring that that standard is maintained, and if for any reason the standard becomes obsolete and “a critical mass of data exists in one format, tools will be produced to migrate it.”<sup>114</sup> Holmes advocates producing different preservation formats to improve the reusability of data, suggesting that we produce a simplified version of our TEI texts using a restricted tag set such as TEI Lite or TEI Simple.<sup>115</sup> We also need to think about appropriate licensing.<sup>116</sup> In addition to their original purpose of providing early dissemination of academic research, institutional repositories have proved to be a good place to preserve and disseminate XML documents and

---

<sup>109</sup> See Pierazzo, *Digital Scholarly Editing*, 169–92; William Kilbride, “Saving the Bits: Digital Humanities Forever?” in *A New Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth (Oxford: John Wiley, 2016), 408–19.

<sup>110</sup> Pierazzo, *Digital Scholarly Editing*, 170–71.

<sup>111</sup> Marilyn Deegan, “Collection and Preservation of an Electronic Edition,” in *Electronic Textual Editing*, ed. L. Burnard, K. O’Brien O’Keeffe, and J. Unsworth (New York: Modern Language Association of America, 2006), 366.

<sup>112</sup> Around 1,000 transcriptions made by the INTF and IGNTF have been converted from plain text to XML. That would have been a lot of data to lose.

<sup>113</sup> Cummings, “The Text Encoding Initiative,” 473.

<sup>114</sup> Pierazzo, *Digital Scholarly Editing*, 171.

<sup>115</sup> Holmes, “Whatever Happened to Interchange?” i65.

<sup>116</sup> See Robinson, “Some Principles for Making Collaborative Scholarly Editions,” §36 on the problems with the Creative Commons non-commercial clause and Houghton, “Electronic Transcriptions,” on the removal of the share-alike condition. In November 2017 the IGNTF took the decision to release its transcriptions under a CC-BY license.



other digital assets.<sup>117</sup> A new initiative from the UK research councils aims to make openly available the data collected during the course of publicly-funded research; this is the data equivalent to open access in publishing.<sup>118</sup> As part of this, institutions are required to provide the infrastructure to preserve this data and make it discoverable. At Birmingham this infrastructure is known as the Research Data Archive and forms part of the Birmingham Environment for Academic Research (BEAR).<sup>119</sup> It allows the long-term archiving of data and allows it to be related to published outputs. This is ideal for the ECM which produces a printed volume as alongside the electronic data.

While preservation options have improved, sustainability remains a challenge. Pierazzo suggests that bespoke tools are inherently unsustainable, implying that more generic tools are more sustainable.<sup>120</sup> It may well be the case that tools of interest to a larger user base, if they are genuinely useful to those users, are more likely to be sustained for longer but it does not mean they are necessarily more sustainable. Sustainability is about more than a large user base: the latter does not necessarily translate into a large development community, particularly with tools in the Humanities. Collate2 had a relatively large user base, but still only one developer; it was also reasonably well documented, yet it became unsustainable because of changes in hardware beyond any developer's control.<sup>121</sup> In ITSEE we stopped using Anastasia partly because of changes to the web server software but also because we discovered other options which we found easier to use. That does not mean that either of those tools was not needed at the time they were written. They were needed, and they have changed the way we edit the New Testament, but they have served their purpose and now we have better tools. The ones that come next will probably be better still.

One of the ways that has been suggested for improving the sustainability of tools and, at the same time, to resolve the tension between generalisation and specific requirements is to use a modular approach. This was the model used for the development of CollateX: it focused on microservices that model individual steps in the editing process but can be joined together into larger workflows.<sup>122</sup> These smaller elements have many advantages over large scale infrastructures. The focus on a single task means they can be flexible without getting too complex, they are more easily sustained and replaced if necessary, and they can be

---

<sup>117</sup> Alex Ball, *Preservation and Curation in Institutional Repositories* (Edinburgh: Digital Curation Centre, 2010), 5–6.

<sup>118</sup> “Concordat on Open Research Data” (2016) <http://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/> [accessed 25 July 2018].

<sup>119</sup> See <https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/index.aspx> [accessed 4 March 2018].

<sup>120</sup> Pierazzo, *Digital Scholarly Editing*, 112.

<sup>121</sup> The INTF actually kept a copy of Collate running until 2015 by preserving an old pre-Intel Macintosh computer as its host. Houghton et al., “The *Editio Critica Maior*.”

<sup>122</sup> van Zundert “If You Build it,” 181.



chained together in customised ways to support a much wider user base. Pierazzo calls for a similar approach, using the term bricks for the component parts, and suggests that these could be joined together with only a small amount of configuration in order to create bespoke workflows for individual editorial projects.<sup>123</sup> This approach achieves maximum reuse without compromising individual expression. Of course you still need programmers to pipe the data from one brick to another and an architecture such as the NTVMR, Textual Communities, or the Workspace for Collaborative Editing to control the workflow. These can all share the bricks that work for them and more easily replace the ones that do not by finding or writing an alternative. The smaller the bricks, the better of course, but it depends on the task: some will need to be larger than others because of the tasks they have to perform. The Collation Editor, for example, is a wrapper around CollateX making it bigger than might be ideal but still discrete enough to be used in different workflows. I like to think of the ideal in terms of the mantra for setting variants for the ECM: each tool should be “as small as possible and as large as necessary.”

When it comes to sustainability and preservation our efforts should go into the data and the final edition, not into the tools used to create them. This data can be reused in the next generation of tools that will invariably come along. No one can predict how technology will change or how that technology will change editorial practice: in 1950 it was predicted that America might ultimately need a total of 12 computers!<sup>124</sup> Just two years before the adoption of the full text transcription workflow by the ECM teams, Robert Kraft hailed Robinson’s Collate as “an excellent program,” but implied that the work involved in making the individual transcriptions is too great a task for the New Testament for it to be worth exploring.<sup>125</sup>

Because the technology available for building and running tools will always be changing, we should worry less about their long-term survival. This is not to say that we should give up on sustainability altogether. We should follow sustainable software practices and use appropriate data standards; we should ensure that tools are documented, for future developers as well as users;<sup>126</sup> that the code is open source and made available in a repository such as sourceforge or github, preferably with a Digital Object Identifier (DOI); we should publish papers that

---

<sup>123</sup> Pierazzo, *Digital Scholarly Editing*, 116–17.

<sup>124</sup> Edmund A. Bowles, “Preface,” in *Computers in Humanistic Research, Readings and Perspectives*, Prentice-Hall Series in Automatic Computation, ed. E. A. Bowles (Englewood Cliffs: Prentice-Hall, 1967), v.

<sup>125</sup> Kraft, “The Use of Computers,” 271–72.

<sup>126</sup> As Stephen Ramsay, “Hard Constraints: Designing Software,” in *A New Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth (Oxford: John Wiley, 2016), 456 says, “‘Read the source’ is an absurd substitution for documentation.”

explain why the tools were needed, what the tools do and how they do it.<sup>127</sup> By following these sustainable development practices we preserve the knowledge that went into the tools and we can build on that knowledge in the future. We will lose tools along the way, which is both inevitable and acceptable: we do not have all of Busa's punched cards, magnetic tapes, and calculating machines, but his concordance is still available online as part of the *Corpus Thomisticum* project. We should not stop building bespoke tools if they are needed just because they might not be as quite as sustainable as more general ones. Van Zundert even suggests that it is at the level of specific domain problems that methodological innovation is found.<sup>128</sup> In short, we should let concerns about sustainability inform *how* we develop, but we should not let them control *what* we develop.

New Testament textual criticism has a long history of embracing technology and that trajectory is continuing. The concept of full text transcriptions and machine collation is now fully embedded in our workflow. The second-generation transcription and collation tools have already become central to the production of the ECM. Work is underway in a collaboration between the INTF and the Cologne Center for eHumanities (CCeH) to develop a second generation of tools to support the Coherence-Based Genealogical Method. Again, these tools will not change the method itself, but rather provide new, interactive interfaces: more new wine skins for existing data and techniques. Alongside this, new computation techniques are also being developed. Andrew Edmondson at the University of Birmingham is investigating the use of phylogenetic techniques for establishing manuscript relations using the university's high performance computing cluster, which like the Harvard Mark IV used by Ellison, is large enough for people to walk around inside. Technology will always be changing, and new technology brings with it new possibilities. If history is anything to go by, New Testament textual critics will continue to make good use of it.

---

<sup>127</sup> A journal has been launched to support the publishing of papers in this area motivated by the problems encountered with getting such papers published elsewhere, see <https://openresearchsoftware.metajnl.com> [accessed 4 March 2018]. Having a platform for publication however does not solve the problem of the low status attached to such papers within University departments.

<sup>128</sup> van Zundert "If You Build it," 167–68.