

The Electronic Scriptorium: Markup for New Testament Manuscripts

Houghton, H.A.G.

License:

None: All rights reserved

Document Version

Early version, also known as pre-print

Citation for published version (Harvard):

Houghton, HAG 2014, The Electronic Scriptorium: Markup for New Testament Manuscripts. in C Clivaz, D Hamidovic & A Gregory (eds), *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*. Scholarly Communication, vol. 2, Brill, pp. 31-60. <<http://www.brill.com/products/book/digital-humanities-biblical-early-jewish-and-early-christian-studies>>

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

Publisher Rights Statement:

Copyright Brill, the authors. Eligibility checked: 30/07/2014

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 providing details and we will remove access to the work immediately and investigate.

THE ELECTRONIC SCRIPTORIUM: MARKUP FOR NEW TESTAMENT
MANUSCRIPTS

Hugh A.G. Houghton¹
University of Birmingham (UK)

1. *A History of Transcriptions*

Every act of copying is a transcription, even though the end product may differ from its source. When a New Testament manuscript was copied by hand, the exemplar would often have been marked up in advance by an editor; corrections and notes added to the exemplar during its use may also have been incorporated into the new copy by the scribe.² In the early days of printing, most textual editions were effectively transcriptions of single manuscripts, again often marked up for the guidance of the compositors, as can still be seen in certain codices.³ The new printing technology meant that the resultant text was reproduced identically on each occasion. One of the results of this was to establish a standard for subsequent scholarship. The consistency of the printed text meant that it could be used as the basis for a collation: instead of providing the full text of multiple manuscripts of the same work, textual variation could be far more economically expressed as a list of differences from a printed version. This can be seen in the alternative readings reproduced in the margins of Stephanus’ *Textus Receptus* and other editions of the Bible.⁴ While editors may have adjusted the main text to create a composite form based on the most compelling readings from all witnesses sampled, the format of subsequent editions remained relatively stable. Manuscripts of particular importance occasionally merited the print publication of a

¹ The author is Principal Investigator (UK) of the *Workspace for Collaborative Editing*, a project funded by the Arts and Humanities Research Council and the Deutsche Forschungsgemeinschaft. He would like to thank his colleagues Catherine Smith, David Parker, Troy Griffiths and Zeth Green for their comments on an earlier version of this paper and their contribution towards the development of the XML schema itself.

² For an example of this, see Schmid, ‘Scribes and Variants – Sociology and Typology’.

³ The marks are still evident in Oxford, Corpus Christi College MS 26 and MS 27, used for sixteenth-century editions of Basil: see Wilson, *A Descriptive Catalogue of the Greek Manuscripts of Corpus Christi College*.

⁴ Estienne, *Novum Iesu Christi D(omini) N(ostri) Testamentum ex Bibliotheca Regia* (third edition). A set of IGTP guidelines on how to make a collation of a manuscript is explained in Parker, *An Introduction to the New Testament Manuscripts and their Texts*, 95–100.

full transcription, some of which reproduce the page layout and even the letter forms of the original.⁵ Other technological developments led to the production of facsimile editions, usually at even greater cost.⁶ Each critical edition of the New Testament text had to be created afresh, based on a collation of collations, with new typesetting on each occasion (and the potential introduction of new errors). Additional evidence and corrections to previous editions could be incorporated in each printing but, as the number of known witnesses increased, the majority of the information was usually taken on trust from one edition to the next. The scale of such an endeavour meant that no comprehensive edition was produced in the twentieth century to succeed the proliferation of New Testament texts from the second half of the nineteenth century.⁷

The advent of computers offered not just the possibility of storing and retrieving the huge amount of data required but a new paradigm for editing and publication. Collaboration on the *Editio Critica Maior* may be traced back to the adoption by both the INTF and IGNTF of the *Collate* software developed by Peter Robinson in the early 1990s.⁸ The core function of this program is the automatic generation of an apparatus of readings from separate files containing full-text transcriptions of individual witness. This removes the scope for human error in the mechanical task of collating multiple sources. At the same time, it re-focusses attention on the individual documents themselves. A by-product of the gathering of data for the new edition is that the primary sources can be presented in full. The inclusion of information about page layout, such as the extent of each line or the size of individual letters, makes it possible to generate a facsimile of each witness in digital typography. Addition-

⁵ The classic example is Tischendorf's pseudo-facsimile of *Codex Sinaiticus* (von Tischendorf, *Bibliorum Codex Sinaiticus Petropolitanus*) although mention should also be made of editions of insular biblical manuscripts which reproduce characteristic scribal features, such as Gwynn, *Liber Ardmachanus: The Book of Armagh*, and Hoskier, *A New and Complete Edition of the Irish Latin Gospel Codex Usser. 2 or r₂ otherwise known as 'The Garland of Howth'*.

⁶ The earliest biblical example of which I am aware is Rettig's 1836 edition of the St Gall bilingual gospels using lithographic technology: Rettig, *Antiquissimus Quatuor Evangeliorum Canoniorum Codex Sangallensis Graeco-Latinus Interlinearis Nunquam Adhuc Collatus*. Photographic facsimiles became common at the turn of the twentieth century, e.g. Thompson, *Facsimile of the Codex Alexandrinus*, (4 vols.), Lake, *Codex Sinaiticus Petropolitanus: The New Testament, the Epistle of Barnabas and the Shepherd of Hermas*.

⁷ For more on this historical period, see Parker, *Textual Scholarship and the Making of the New Testament*, 106–122.

⁸ Robinson, *Collate: Interactive Collation of Large Textual Traditions*, Version 2 (Computer Program). On the use of *Collate* see, for the INTF, Wachtel, 'Editing the Greek New Testament on the Threshold of the Twenty-first Century' and, for the IGNTF, Parker, 'Electronic Religious Texts: The Gospel of John'.

ally, the production of the collation from electronic files means that the same transcriptions can be re-used in multiple editions, rather than starting afresh on each occasion. In essence, the task of the first digital editors of the New Testament is initially to produce a diplomatic edition of each document: the editing of the work itself is a later stage during which the disparate data from individual witnesses is brought together into a standardised form.⁹

The encoding of the transcriptions for *Collate* followed and extended the conventions developed by Robinson.¹⁰ In addition to changing an electronic base file of the work to correspond to the text found in a particular document, transcribers were able to include information about the physical and textual characteristics using a system of markup. This was divided into four categories:

1. Block markers, indicated by angled brackets: < >
These are present in the base file and indicate the standard division of the work (into book, chapter and verse). They are used for orientation within the text and as the identifiers for collation.
2. Location markers, indicated by pipes: | |
These describe the physical layout of the text in each manuscript, dividing it into pages, columns and lines. They are not taken into account during collation.
3. Tags, indicated by square brackets: []
These designate a portion of text as distinctive in some way, for example identifying section numbers written in the manuscript, capital letters or abbreviations. They may be used to mark sections which are difficult to read or have been reconstructed. They also indicate the readings of different hands, where the original text has been altered by the copyist or subsequent users. An opening tag is placed at the beginning of the relevant portion of text, and a corresponding closing tag at the end.
4. Comments, indicated by braces/curly brackets: { }
These enable transcribers to make observations which are not treated as part of the text for collation. These may include glosses

⁹ For a definition of the key terms 'document', 'work' and 'text' and their application to the New Testament tradition, see Parker, *Textual Scholarship*, 10–14 and 29. On the shift in the task of the editors, see Parker, 'Through a Screen Darkly: Digital Texts and the New Testament', 404.

¹⁰ A full description of the capability of the software and the conventions is provided in Robinson, *Collate 2*. The markup itself was based on the *Oxford Concordance* program produced several years earlier by Susan Hockey and Ian Marriott. A worked example of *Collate* encoding is given by Parker, 'Through a Screen Darkly', 405–407.

or page numbers, or simply provide a commentary on certain readings or the state of the manuscript.

The *Collate* software ran on a Macintosh computer using an operating system before the introduction of the Unix-based OS X in 1999. It also relied on the ASCII character set, with only 95 printing characters. Transcriptions of Greek manuscripts therefore required a font which substituted Greek letters for the standard Roman characters. *Symbol Greek* was the betacode font adopted for this purpose. However, the character mapping of this font led to the substitution of the brackets indicating markup elements with other, often unusual, characters. The result was a file which, although it benefited from an economical system of markup that allowed the transcriber to focus on the biblical text, often appeared impenetrable to the human eye. Furthermore, the use of a betacode font for both text and markup and the creation of tags beyond those originally specified by the software means that, while files could be shared between specialists working in the two collaborating institutions, the encoding was not transparent for external users.

The publication of the transcriptions in an electronic edition relied on the conversion of the plain text files into Standard Generalized Markup Language (SGML) with each Greek letter converted into a separate unique entity in order to enable the presentation of both Roman and Greek characters on the same page. The publication system adopted by both projects was *Anastasia*,¹¹ also developed by Peter Robinson. This ran as a server which converted the SGML data into HyperText Markup Language (HTML) for viewing in a web browser. The tags were converted into HTML elements and then rendered according to a specified scheme (e.g. red characters in square brackets for supplied text, or blue characters for first-hand readings and green characters for corrections): hyperlinks were used to navigate around the edition, and images of each page could also be incorporated. The transcriptions could be viewed in two modes: 'Page Layout', reproducing the organisation of text on each page, and 'Chapter View' with the text arranged by verse. The principal New Testament editions using this software were the New Testament Transcripts prototype (NT Transcripts), published by INTF in 2003, and

¹¹ Robinson, *Anastasia: Analytical System Tools and SGML/XML Integration Applications*, Version 2.0, (Computer Program). The documentation is available online at <http://sd-editions.com/anastasia/index.html>, last accessed 8 May 2013.

the three online editions of different traditions of the Gospel according to John published by the IGNTP in 2007.¹²

Even as these editions were being produced, two significant developments paved the way for future innovations. The first was the widespread adoption of the Unicode character encoding, used by the World Wide Web Consortium (W3C) with the introduction of HTML 4.0 from 1997 and installed as standard in Macintosh OS X. With over a million possible character encodings, this meant that Greek letters (and, from 2005, Greek numerals and other special characters) could be uniquely identified regardless of font, and thus appear in the same file with symbols from other languages without the need for further differentiation. The second was the Text Encoding Initiative (TEI), established as a consortium in 2000, which employed XML as the standard encoding format with effect from its *P4* set of guidelines published in 2001.¹³ However, while its collation engine remained the key to providing the data for editions of the New Testament, *Collate* did not support Unicode, could not collate files in XML (although it could convert transcriptions into this format for publication in *Anastasia*) and did not work natively on Mac OS X. The challenge of creating a successor, called *CollateX*, was undertaken by the *Interedition* COST action funded by the European Science Foundation from 2008 to 2012.¹⁴ In anticipation of this, the IGNTP and INTF started to make transcriptions in Unicode from 2009, using the standard *Collate* markup.¹⁵

The first example of the re-use of electronic New Testament transcriptions was the *Digital Codex Sinaiticus*, an online edition of this fourth-century Greek Bible combining images from all four holding institutions, a complete transcription of the text and a translation.¹⁶ The files of this manuscript prepared by the INTF for the NT Transcripts edition were made available to the project and subsequently enhanced by the addition of information such as Eusebian canons and marginal glosses in order to

¹² All of these were available online at the date of writing, at <http://nttranscripts.uni-muenster.de/> and <http://www.iohannes.com/> (comprising editions of the Greek majuscule manuscripts of John, the *Vetus Latina* manuscripts of John and the Byzantine text of John). *Anastasia* was also used for the publications of the *Canterbury Tales Project* (<http://www.canterburytalesproject.org/>) and other online and CD-ROM editions published by *Scholarly Digital Editions* (<http://sd-editions.com/>).

¹³ Burnard et al., *TEI P4: Guidelines for Electronic Text Encoding and Interchange*.

¹⁴ See <http://www.interedition.eu/>.

¹⁵ The guidelines for transcribers were deposited in the University of Birmingham Institutional Research Archive at <http://epapers.bham.ac.uk/751/>; the latest version (5) was published in October 2012 at <http://epapers.bham.ac.uk/1676/>.

¹⁶ <http://www.codexsinaiticus.org/>. The history of this project is related in Parker, *Codex Sinaiticus. The Story of the World's Oldest Bible*.

match the conventions adopted for the rest of the manuscript. They were then converted to the markup developed for the electronic edition, a customised version of XML, and published online in 2009 using a bespoke system created for the project.¹⁷

In the autumn of 2010, work began on a new online environment to integrate all the tools required by the partners involved in producing the *Editio Critica Maior*. Scheduled for completion in late 2013, the *Workspace for Collaborative Editing* aims to connect each stage of the editorial process.¹⁸ These comprise the initial transcription of manuscripts, the automated collation of witnesses and production of an initial critical apparatus and database, the addition of patristic and versional evidence, the establishment of the initial text (*Ausgangstext*) using the Coherence-Based Genealogical Method and the eventual publication of the *Editio Critica Maior* in print and electronic form. Early on in the project, the decision was made to use TEI-compatible XML as the format for encoding and storing transcriptions. A subset of the latest version of the TEI Guidelines, *P5*, was therefore developed by the project in order to meet the requirements for work towards the *Editio Critica Maior*.¹⁹ These comprised both the conversion of earlier *Collate* files to the necessary standard for incorporation into the new environment and the creation of new transcriptions.

Despite the many advantages of XML for standardisation and storage, the verbose character of the markup makes it very inefficient for transcribers to work directly in this encoding. Instead, one of the components of the *Workspace for Collaborative Editing* is an online Transcription Editor used within a web-browser.²⁰ The interface mimics the display of the XML of transcriptions already published online by the projects, providing a 'what you see is what you get' environment. As with the earlier electronic transcriptions, users do not start from scratch, but are able to choose from a selection of base texts with the standard divisions of the

¹⁷ Parker, *Codex Sinaiticus*, includes a sample of the plain text and XML transcriptions on pp. 178-9. The Old Testament and other writings were transcribed by the project team, following the same methodology. The process of transcription is also described in Houghton, 'The Electronic Transcription of Codex Sinaiticus'. There is documentation on the website: the full transcription can be downloaded from http://codexsinaiticus.org/en/project/transcription_download.aspx.

¹⁸ See further Parker, *Textual Scholarship*, 113–119 and 138–141.

¹⁹ The full version of the *P5* Guidelines is Burnard and Bauman, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. These are available online at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>.

²⁰ This was developed by Thomas Burch, Martin Sievers and Gan Yu of the Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften at the University of Trier (KoZe).

work already in place. The transcribers change this to correspond to the reading of the manuscript. The 'hidden' parts of the markup are supplied in dialogue boxes which appear when an additional element is selected, such as a line break or a correction. Many of these can be added through shortcut keys. One particularly notable feature is the 'editor within an editor' dialogue box which enables transcribers to edit the markup of the text supplied by a corrector, specifying unclear or supplied characters or other types of formatting. The transcription tool is not intended to stand alone but to be integrated into a suite of tools such as the *Workspace for Collaborative Editing* or the *New Testament Virtual Manuscript Room* (NT.VMR),²¹ drawing on other sources of information such as databases with bibliographical details for New Testament manuscripts and indexes of biblical content for digital images of each page.

In the second half of this paper, the XML encoding adopted for the *Workspace for Collaborative Editing* will be described, along with observations on how it has developed over the course of the project and some problems which have been encountered. There is always a balance to be struck regarding the amount of information included in a transcription, which reflects the potential of these initial electronic files for multiple uses. For the purpose of creating a critical edition of a work, details of formatting and layout are superfluous; a 'digital facsimile', however, tries to match the document as closely as possible. The practice of the IGNTP has been to include some information which goes beyond the purely textual, in order both to enable the transcription to form the basis of an electronic facsimile with explanatory information for non-specialist users, and also to have the possibility of more complex searches based on specific phenomena, such as abbreviations or spelling conventions. Of course, electronic transcriptions can always be altered and improved, and it may be that other researchers with, for example, a particular interest in punctuation or textual divisions, will enhance files in this way. Similarly, although the *Workspace for Collaborative Editing* itself was specifically commissioned for collaborative work on New Testament manuscripts leading to the production of a critical edition, it is hoped that, like many of the major developments in textual scholarship which originated in work on the New Testament, it may also be applied to other textual traditions. At each stage, therefore, this potential expansion of

²¹ This is a community portal developed by the INTF for work on New Testament manuscripts: the current version, NT.VMR 2.0, is hosted at <http://ntvmr.uni-muenster.de/>.

material has been kept in mind so that, with minimal adjustment, it should be possible to use the same framework in a different context.

2. An XML Encoding for Manuscript Transcriptions

The XML encoding developed for the IGNTP transcriptions of New Testament manuscripts is a subset of the TEI *P5 Guidelines*.²² Compatibility with the TEI means that generic documentation for the markup is provided online by the TEI. Furthermore, tools developed for use with the entire range of the *P5 Guidelines*, such as those for parsing, visualisation and analysis may be applied directly to these transcriptions. Given the variety of potential combinations of XML elements, however, the selection of a smaller group was necessary in order to ensure a manageable standard format for developing tools within the *Workspace for Collaborative Editing*, displaying the transcriptions with a standard XSLT template and interchange with other encodings, such as JSON for the *CollateX* engine. The choice of elements was informed by the encodings used for the *Digital Codex Sinaiticus* and New Testament transcriptions published using *Anastasia*, along with the TEI *P5 Guidelines*.²³ The initial scheme was issued on 1st December 2010; a revised version was adopted by the IGNTP in early 2011 and published online in the University of Birmingham Institutional Repository (UBIRA).²⁴ Ongoing work has led to alterations, periodically released as revised versions: the guidelines described here are Version 1.4, published in July 2013, which may be downloaded from <http://epapers.bham.ac.uk/1727/>. A schematic overview is also included in the Appendix below.

Unlike the *Collate* markup described above, all XML markup is enclosed within angled brackets: < >. The first group of letters within the brackets identify the *element*. This may be further qualified by *attributes*, whose value is expressed between quotation marks. So, for example, <wn="4" xml:lang="en"> indicates a word element (w) with two attributes: a numerical identifier (n) of 4 and a language value (xml:lang) of English (en). Elements may either be empty, providing punctual information such as a single space or a line break, or may enclose a portion of text. In

²² Burnard and Bauman, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. For a description of how to make an electronic transcription of a New Testament manuscript, see Parker, *An Introduction*, 100–106.

²³ Unfortunately, it was only after the first schema had already been created that I encountered Timothy Finney, ‘Manuscript Markup’.

²⁴ Version 1.1: <http://epapers.bham.ac.uk/738/>.

the latter case, the element must be closed after the last character to which it applies, using a forward-slash within angled brackets, such as `</w>`; for empty elements, the closing slash appears within the same brackets as the rest of the element, as in the element `<pb/>` for page break. In certain cases, multiple elements may apply simultaneously to the same portion of text: this is known as 'nesting'. The overall principle for the transcription of manuscripts is that what appears on the page is transcribed as text and everything else is indicated by markup. This is facilitated by editing programs which display text and markup in different colours, sometimes also differentiating between elements and attributes.²⁵

2.1. Header

The TEI header, `<teiHeader>`, is an obligatory part of an XML transcription, providing details about both the electronic file and also the manuscript transcribed.²⁶ The amount of information provided may vary from project to project: the more that can be entered from a database or series of drop-down menus the better, in order to ensure standardisation of identifiers. Most of the details of Greek New Testament manuscripts are already stored in the electronic version of the Gregory–Aland *Kurzgefasste Liste* available in the NT.VMR,²⁷ while details of the originating project and encoding procedures are common to multiple transcriptions. In order to publish individual transcriptions as self-contained files, however, some information may be repeated as part of the header.

The different 'type' attributes of the `<title>` element reflect the distinctions expressed by the *Documents, Works, Texts* project in developing an ontology environment for the identification of electronic material relating to manuscripts.²⁸ *Document* refers to the manuscript. Although the main description is free-text (e.g. 'Codex Alexandrinus'), two attributes permit the precise identification of the document according to standard systems: the 'key' attribute has the five-digit number used in the electronic *Liste*, which also underlies the IGNTP and INTF file naming

²⁵ For a further introduction to XML, focussing on the hierarchy of elements, see Finney, 'Manuscript Markup', 276-279.

²⁶ The header violates the principle that the text field only contains words from the manuscript: the whole `<teiHeader>` element is an editorial construct, as are editorial notes (discussed below).

²⁷ <http://ntvmr.uni-muenster.de/liste>.

²⁸ This project, funded by the UK Joint Information Services Committee (JISC) ran from 2009–10; as noted above, the differences are described in Parker, *Textual Scholarship and the Making of the New Testament*, 29.

scheme, while the 'n' attribute gives the *siglum* of this witness in the *Editio Critica Maior*.²⁹ *Work* denotes the customary abstract identification of an authorial creation (e.g. 'The Gospel according to John', abbreviated to 'John' in references). The use of *collection* to refer to the whole New Testament puts each work in its wider context although, of course, individual manuscripts vary in the selection of books they contain.

Details of creators, funders, editions and dates are included in order to provide recognition and keep track of the publication of individual transcriptions. As noted above, the IGNTP has a policy of depositing completed files in an Institutional Repository (currently UBIRA). These are freely available, as has subsequently become a condition of research funded by UK Higher Education Funding Councils. Furthermore, they are issued with a Creative Commons licence which permits their re-use.³⁰ Just as photographic facsimiles are of interest to a far wider range of scholarly disciplines than textual editing alone, electronic transcriptions offer the potential for fresh investigation of linguistic and scribal phenomena as well as other types of data analysis.

Bibliographical information about the manuscript can be included within the <sourceDesc> element, incorporating parts of the TEI *P5 Guidelines* module on Manuscript Description (e.g. <msDesc>, <msIdentifier>, <institution>, <repository>). Although much of this information is already present in the *Liste*, its inclusion in individual transcriptions assists in opening them up to discovery by search engines. The use of <altIdentifier> elements permits the identification of the manuscript in a variety of different catalogues or editions. The inclusion of references to the *Leuven Database of Ancient Books* is part of a reciprocal arrangement for the provision of links to these transcriptions in their catalogue.³¹

For the purpose of *Collate*, details of transcription practice and the history of each file were recorded in an initial status note. The former is incorporated into the <encodingDesc> section: the <editorialDecl> element is a free-text field where information on the treatment of features such as punctuation, capitalisation or rubrication may be specified. The section should also include an empty element, <variantEncoding/>, specifying the procedure adopted for handling different readings.³² The file history may be detailed in the <revisionDesc> section where details of each alteration to the file are listed.

²⁹ This naming scheme is explained in Parker, *An Introduction*, 105–106.

³⁰ On Creative Commons licences, see <http://creativecommons.org/>.

³¹ <http://www.trismegistos.org/ldab/>.

³² See further 2.5. below on corrections.

A header which employs the full scope of the TEI *P5 Guidelines* would provide an exhaustive amount of information, from bibliographical references for secondary literature to the enumeration and identification of members of the project team and their interventions in the file. However, a compromise has been made in order to provide enough information to enable each file to stand by itself while not distracting from the task of transcription. For example, it would be good practice to include in the header a list of all scribal hands which worked on the manuscript, either as copyists or correctors. However, the header information is normally added at the start of a transcription, when the nature and composition of the copying team may be unknown; furthermore the identification of scribal hands can require specialist palaeographical expertise. As noted above, one of the benefits of an electronic transcription is the potential for its enhancement by subsequent users. So long as the header provides enough information in a valid form for the identification and re-use of the transcription by external projects, then it will have served its purpose.

2.2. *Divisions of the Work and Document*

Scholars of the New Testament benefit from a commonly-accepted system of divisions into book, chapter and verse brought to completion by Stephanus in the sixteenth century.³³ It is worth remembering that this scheme of chapters and verses is not present in earlier manuscripts, which preserve evidence for a variety of series. These latter divisions may be recorded as paratextual elements (see 2.3 below), but the modern system is used for ease of reference and to connect transcriptions with each other and the critical apparatus. In XML, the longer units of 'book' and 'chapter' are treated as <div> elements; the use of <ab> ('anonymous block') for biblical verses is exemplified in §16.3 of the TEI *P5 Guidelines*. In order to locate each verse, a concatenation of 'n' attributes is created: the biblical book is identified by a two-digit code preceded by B (e.g. 'B04' for John), the chapter by this and the chapter number prefaced by K (e.g. 'B04K5' for John 5) and the verse in a similar manner (e.g. 'B04K5V21' for John 5:21). Although it would be possible to use

³³ It may be noted in passing that there is not always complete agreement between different versions or editions: for variations in the nomenclature of biblical books, see the chapter by Laurence Mellerin in the present volume; differences in the versification of John between the Nestle–Aland Greek text and the Stuttgart Vulgate are listed in Burton et al., *Vetus Latina 19. Iohannes*, 6. The IGNTP and INTF always follow the versification of the Nestle–Aland edition.

an XPath query to search recursively for verse 21 within chapter 5 of book 04, in a plain text editor a single identifier at each level enables easy navigation. In an earlier version of the schema, this value was encoded as a unique attribute ('xml:id'). However, while each biblical verse is unique, within a manuscript it may appear on more than one occasion (such as lectionaries with overlapping readings, bilingual codices, dittography of a longer passage or verses split over page breaks when a transcription is stored by page). For this reason, the 'n' attribute was used, which can be repeated as necessary. In bilingual manuscripts, the use of an 'xml:lang' attribute on each <ab> element permits the differentiation of each version.³⁴

Within each <ab> element, each word is treated as a <w> element and punctuation tokens as <pc> elements. The <w> elements are numbered with an 'n' attribute: this is inherited from *Anastasia*, where it formed part of the information connecting the critical apparatus to the transcription. This attribute may also be used as a link with the presentation of the *Editio Critica Maior*, where each word in the base text is allotted an even number, although this would require the numbers to be added after the collation stage. This attribute could also be used to map translations back to a Greek text, as well as link to external resources such as concordances. Other information may be nested within the <w> elements, such as abbreviations (<abbr>) and formatting information (<hi>), discussed below.

The <div> element may be used for other material, either within or outside the elements denoting 'book' and 'chapter'. Preliminary material, such as the *Letter to Carpianus* or canon tables, may be allocated its own <div> at the same level as a biblical book. Within each book, prefaces and *kephalaia* (lists of chapter titles) may precede the first chapter. Short titles at the beginning or end of the work are also treated in their own right, as <div type="incipit"> and <div type="explicit">. (In Collate, these were assigned verse number 0 in imaginary chapters at the beginning and end of each book so that they could be collated.) A longer

³⁴ The most common languages for New Testament bilinguals are Greek, Latin and Coptic. However, the XML indication of Greek (xml:lang="el") refers to the modern rather than the ancient language. The BCP47 standard adopted elsewhere in the TEI Guidelines indicates that a private use code, such as xml:lang="el-x-koine", would be the best way of indicating New Testament Greek. However, for simplicity, we have adopted the IANA codes of 'grc' for ancient Greek and 'cop' for Coptic (see further <http://www.iana.org/assignments/language-subtag-registry>). Of course, Coptic should be further subdivided into the different dialects but these are already indicated in the SMR *sigla* for Coptic manuscripts, so the use of 'cop' as a blanket language indicator may suffice.

colophon may also be treated as another <div>, subdivided into <ab> or <p> elements as desired.

As already described, a single XML transcription can also be used to generate a textual facsimile of each physical page of the document if information is included on the disposition and formatting of the text. In the TEI *P5 Guidelines* this is done by a series of empty elements which indicate page, column and line breaks: <pb/> <cb/> and <lb/>.³⁵ All these have the 'n' attribute giving an editorial numeration. (These may or may not correspond to any numerals written on the page, which would be transcribed as <fw> elements.) As there is only one instance of each of these numbered elements, they are also given an 'xml:id' attribute which connects them to the document, using a concatenating hierarchy similar to that mentioned earlier for divisions of the work. Thus <lb n="3" xml:id="P246C2L23-044"> identifies page 246, column 2, line 3 of manuscript 044. In order to distinguish between different ways of numbering, the attribute 'type' is included on <pb/> elements, with the value 'page' when each side has a separate number and 'folio' when a page is given a single number and the sides are identified as *recto* or *verso*. Because the papyrological *recto* and *verso* may not correspond to the assigned page numbers, the direction of the fibres may also be specified with the 'subtype' attribute. In the case of palimpsests, information about the different systems of page numeration can be included within the <pb/> element. The attribute 'facs' can be used to give a URL for a digital image of the page. At the level of <lb/>, indentation, *ekthesis* and justification are encoded using the 'rend' attribute.

The downside of this approach to encoding layout, however, is that it leaves open a problem of multiple hierarchies when producing a digital facsimile. As its name implies, the Text Encoding Initiative is primarily concerned with texts (where layout is usually secondary) rather than documents (where the emphasis is on the physical object). Thus the TEI scheme makes it easy to extract a chapter or a verse (as a division of the work), but much more difficult to extract a single page of text, since a page break is a single empty element at the beginning of a page rather than an opening and closing element surrounding a portion of text. The only way of discovering where a page, line or column ends is to go to the next break element and assume that everything between the two is part of the first element. This leads to problems, such as when a number of

³⁵ In addition, the element <gb/> ('gathering break'), to indicate a manuscript's quire structure, was introduced into version 1.8.0 of the *P5 Guidelines* in late 2010.

pages (or lines) may be missing, or at the end of a document or page where there is no closing tag. Furthermore, because page divisions rarely coincide with textual divisions, when considering a single page in isolation the opening words are usually not identified (because they form part of a block which began on the previous page), while the closing block may remain unfinished – and is therefore invalid XML. The fact that breaks are encoded as empty elements also makes it difficult to use a standard transformation scenario to extract a single page of XML to display alongside a facsimile of the manuscript, although this is less of an issue when the transcription is stored in an XML database.

In part, the problem results from the mismatch between the task of transcription, proceeding page by page, image by image, and the abstract divisions of the work. When working from individual digital images (particularly if, as in the NT.VMR, more than one transcriber may be responsible for different pages of a single manuscript) the hierarchy of layout takes precedence over the system of chapters and verses, especially if the base text is automatically supplied using indexing information already collected in a database. However, when the same data is fed into a collation engine, this relies for its reference on the customary sequence of the work. One work-round is to close and reopen `<ab>` and `<div>` elements at the extremities of each page, using the 'part' attribute, to ensure that each page is self-contained. This has the unwelcome result of fragmenting the text: in order to collate it, the various parts of the `<ab>` have to be identified (using the 'n' attribute) and re-assembled. It also places an additional burden on the transcriber by making them responsible for ensuring that all the relevant elements are closed and then reopened at every page break, although in the online interface this can be partially automated. This is the solution adopted in these guidelines.

Other alternatives have been suggested. The minimal level is represented by Finney's use of the empty `<milestone/>` element to record each division of the work in place of containers.³⁶ This has the advantage of not privileging one hierarchy over another, but would require a second stage of processing to extract a particular section by page or biblical text. A maximal level is found in the XML developed by Peter Robinson for the *Digital Codex Sinaiticus*, which not only departs from the TEI *P5 Guidelines* in turning break elements into containers (meaning that text was contained between opening and closing page, column and line elements), but also by subdividing the breaks into pairs of 'start' and 'end'

³⁶ Finney, 'Manuscript Markup', 272, 279-280.

subtypes. This is indicated in the 'id' attribute, with the details of the corresponding other member included as a 'corres' attribute.³⁷ The advantage of this is that marginal material can be transcribed at exactly the point at which it appears in the text: if it is in the top margin of the page, then it would be enclosed by the opening and closing elements of the <pb> for the start of the page; if it is at the end of a line, it would be added between the indications of the <lb> for the end of that line. As a check, an innovative <margin> element was also added inside the break elements, confirming that the text was located in the margin. The disadvantage is a substantial increase in the size and complexity of the file, much of it unnecessary: as marginal material only occurs within a small proportion of break elements, most of the co-ordinated 'start' and 'end' breaks are simply recorded as empty elements. Even though the output XML was generated automatically, it often had to be manipulated by hand. Furthermore, the departure from the TEI *Guidelines* (including the creation of a <margin> element) meant that it was considered unsuitable for the IGNTP transcriptions.

2.3. Paratextual Elements and Formatting

Electronic transcriptions offer the possibility of comparing far more details found in manuscripts than simply the biblical text, so long as these are also included in the transcription. Indeed, the goal of an 'electronic facsimile' is to represent all the information on each page. Features such as page numbers, quire signatures, running titles and other elements of the *mise-en-page* can all be described as <fw> elements. The following explanation is given in §11.6 of the TEI *P5 Guidelines*: 'Although the name derives from the term *forme work*, used in description of early printed documents (the 'forme' being the block used to hold movable type), the fw element may be used for such features of any document, written or printed'.³⁸ In addition to the three types mentioned above, lection titles and chapter titles are also treated as <fw> elements. When an <fw> element includes a value, especially if this is in Greek or Latin numerals, the equivalent value is supplied as an 'n' attribute. Certain paratextual elements are numerals: chapter numbers and the two parts of the Eusebian apparatus (Ammonian Sections and Eusebian Can-

³⁷ This procedure is sometimes known as 'Trojan Markup': see further DeRose, 'Trojan Markup and Other XML Milestone-Tagging Techniques', available at <http://www.mulberrytech.com/overlap/index.html>.

³⁸ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHSK>.

ons) are encoded as <num> elements, again with the value given as an 'n' attribute. The ability to collate or compare series of section numbers, especially those which are unusual, can offer significant information about the relationship of manuscripts.³⁹

The decoration of the text is recorded by the <hi> element. Its principal deployment is to indicate letters written in rubrics or other coloured ink, horizontal lines above or below the text (except where these are a substitute for a missing letter) and oversize letters. The convention developed in *Collate* was to describe the size of large capital letters as a multiple of the height of a standard line: this provides the numerical value for the 'height' attribute. When texts written in minuscule include capital letters within the vertical span of a single line, these are simply transcribed as capitals and not otherwise indicated (although care is sometimes needed to identify unusual features of certain scripts). The recording of decorative lines often has to be approached with caution, in order to distinguish whether they were added by the first hand (as in the case of *nomina sacra* or titles) or a later user: if the latter are transcribed, it may be appropriate to treat them as corrections (or, better, alterations).

2.4. *Lacunae, spaces, supplied and damaged text*

Many New Testament manuscripts, especially the early papyri, have only been partially preserved. It is possible to give an extremely detailed description of the nature and cause of damage or gaps in XML, but this has been restricted in the IGNTP scheme in order to accommodate earlier practice in the plain-text transcriptions made for *Collate*. Lacunae, when the writing material is no longer present and no reconstruction is offered, are expressed by an empty <gap/> element. The approximate size of the gap is given by the combination of two attributes, 'unit' (with values 'page', 'line' or 'char[acter]') and 'extent'. The benefit of 'extent', as defined in the TEI *P5 Guidelines*, is that it can be used to give a range (e.g. '3-4') rather than just a single number. In an earlier version of the guidelines, a 'ghost page' was added to fragmentary manuscripts beginning in the middle of a verse, with the opening <div> and <ab> elements before the <pb/> of the surviving text in order to avoid giving the impression that the lacuna occurs on a page which is completely

³⁹ See, for example, McGurk, 'The Disposition of Numbers in Latin Eusebian Canon Tables'; Popovic, 'Du nouveau sur les Évangiles de Split', especially 290–291; O'Loughlin, 'The Biblical Text of the Book of Deer (C.U.L. li.6.32): Evidence for the Remains of a Division System from its Manuscript Ancestry'.

extant. The adoption of the 'part' attribute means that this is no longer necessary and only extant text need be transcribed.⁴⁰ In addition, a second type of <gap/> element, with the values 'verse' and 'chapter' for the 'unit' attribute, may be added to indicate a textual lacuna. This means that even though the extant page may be physically complete, the incompleteness of the text (with regard to the standard form) can still be recorded. This is also helpful for aligning witnesses in an automatic collation.

In the case of shorter lacunae, the transcriber may choose to supply a reconstruction of the missing text. In this case, the words are enclosed within a <supplied> element, which replaces the <gap/>. Where the parchment is still extant but the text is no longer readable, the same procedure may be followed but with the attribute 'reason="illegible"' rather than 'reason="lacuna"'. The source of the reconstruction may also be specified: values for this attribute may include 'transcriber', a standard text such as 'NA28' or 'Maj', or the *editio princeps*. Where the text is partially extant but the reading is unclear, the relevant characters may be given inside an <unclear> element, with the reason again specified as an attribute.⁴¹ There is also an XML <damage> element which can be used to enclose sections of damaged text: this is not adopted within the IGNTF schema, however, as the use of <unclear> and <supplied> gives a more precise indication of exactly which characters may be read. Detailed guidance on the use and combination of these various elements is given in §11.5.2 of the TEI *P5 Guidelines*. It should be noted that where a copyist has deliberately left a blank space, this should be transcribed using the element <space/>, the dimensions specified in the same way as for a <gap/>.

2.5. Corrections and alternative readings

The practice of the IGNTF and INTF, following procedures established for *Collate*, is to encode variant readings within the flow of the text at the point to which they refer.⁴² It is also a convention that variation units

⁴⁰ Unfortunately, the XML elements <lacunaStart/> and <lacunaEnd/>, which could be employed in transcriptions to indicate the extremities of missing portions, are intended for use in a critical apparatus and can only appear within <rdg> or <lem> elements.

⁴¹ There is the scope to give an indication of degrees of certainty as an XML attribute: see further Finney, 'Manuscript Markup', 270–272. The approach to unclear and supplied text in transcriptions for the *Editio Critica Maior* reflects the requirements of printed editions following the Leiden conventions.

⁴² As noted in 2.1. above, this is indicated in the TEI header by <variantEncoding method="parallel-segmentation" location="internal"/>.

must consist of complete words, in order to assist with the automatic generation of an apparatus. In XML, each variation unit is enclosed within an <app> element, while the respective readings are identified as <rdg> elements. The attributes 'type' and 'hand' identify the nature of the readings and the person responsible. The first element in the sequence is normally <rdg type="orig" hand="firsthand">, followed by all the words in that unit.⁴³ As many corrections to that section of text may follow as are required, usually in chronological order (so far as this can be ascertained) and with each <rdg type="corr"> allocated a separate 'hand' attribute. The entire variation unit must be included within the <rdg> element even if some words are left unchanged by a particular corrector. The numbering of the 'n' attributes of the <w> elements in the first <rdg> continues the sequence of the verse, and is then repeated in each subsequent <rdg>. If a single word is omitted, the relevant <w> element is missed out; if an entire <rdg> is blank (i.e. deletion of the entire text or a first-hand omission) this is indicated by the entity &om; in order to assist with collation and display.⁴⁴

The combination of <app> and <rdg> elements may also be used to transcribe commentary readings or alternative readings, such as those found in the margins of Family 1 minuscule manuscripts. In this case, the reading in the body of the manuscript is identified as 'type="orig"' and the alternative is 'type="alt"' or 'type="comm"' as appropriate, with details of the hand responsible. Recording the position of these readings follows the process detailed for marginal material below: all the words within the <rdg> should be enclosed within <seg type="margin">. The ability to combine all the different types of encoding is a particular strength of XML: changes in punctuation, the form of abbreviations, the decoration of the text and even the position of readings (if a text added in the margin was subsequently written above the line by a later hand) are as easy to describe as alterations to the text itself.

2.6. Abbreviations and non-standard characters

As already observed, details of abbreviations and character forms go beyond the information required in transcriptions for the purpose of cre-

⁴³ It would also have been possible to designate the first reading as the lemma, the element <lem> rather than <rdg>. However, because the earliest reading of a manuscript may not be entirely legible, or may have been immediately changed by the first hand, the hierarchy implicit in the use of *lemma* could be misleading.

⁴⁴ For example, omissions in *Codex Sinaiticus* and Nestle–Aland are indicated by the [†] symbol.

ating a critical edition. Nonetheless, two types of abbreviation found in most New Testament manuscripts are regularly recorded in transcriptions for the *Editio Critica Maior*. These are *nomina sacra* and numerals, which are also usually indicated by an overline. The <abbr> element is added after the opening <w> and followed by the formatting for overline <hi rend="overline">. The three nested elements are closed in reverse order. The type attribute identifies whether the abbreviation is 'nomSac' or 'num'. Other abbreviations may be silently expanded or have the omitted letters supplied within an <ex> element. The latter allows the display to toggle between the visible characters, matching the manuscript, and the full form for reading purposes. Sometimes more than one character may be visible in an abbreviation, such as τ^λ, the standard form for τελος, which could be rendered τ<ex>ε</ex>λ<ex>ος</ex>. However, as other non-letter like symbols are used for abbreviations, it may be better to restrict the 'visible characters' to those which are normal size, as in τ<ex>ε</ex>λος</ex>.

Non-standard characters sometimes feature in a manuscript, often involving abbreviations. Transcribers should, as far as possible, supply the appropriate character in Unicode even if this is not supported by all fonts: examples include ζ and ϙ for the numerals 6 and 90.⁴⁵ When a character is not represented in Unicode, there is the option of using a different glyph (or combination of glyphs) of similar appearance, such as the use of a macron in place of a 'nu-superline', although this is not transparent and should be mapped in a one-to-one relationship for consistency.⁴⁶ A better alternative is to use entities to represent non-Unicode characters. So, for example, the single character abbreviations for Latin *autem* and *enim* in Insular script are transcribed as &autem; and &enim; rather than trying to reproduce them by combinations such as *hr* or ++. Similarly, while most punctuation characters are included in Unicode, even in Greek, non-standard forms (or those which conflict with XML encoding, such as > or ;) can be treated in this way. Ligatures may also be expressed as entities. However, it is necessary to declare all entities which occur in an XML document, as well as their intended representation, above the TEI header in order for the file to parse successfully.

⁴⁵ The code points for these characters are, respectively, U+03DB and U+03D9.

⁴⁶ An alternative would be to expand the *nu*-superline into <ex>ν</ex>, but this does not of itself give the precise form which the abbreviation has taken. Conversely, the *kai*-compendium does have its own glyph in Unicode, ϰ (U+03D7), but IGHNT practice is to expand this as κ<ex>αι</ex>. In the *Digital Codex Sinaiticus*, almost all the unusual marginal punctuation characters were reproduced by combinations of other glyphs (e.g. Quire 59 fol. 5r or Quire 82 fol. 2v).

The treatment of bilingual manuscripts was mentioned in 2.2 above, with an ‘xml:lang’ attribute on each <ab> element. On other occasions when two languages are found in the same manuscript (such as untransliterated Greek words in a Latin codex), these may be enclosed in a <foreign> element, again specifying the language as an attribute. The use of Unicode, however, renders this less of a problem than formerly.

2.7. *Editorial notes*

The descriptiveness of XML markup means that most features of a transcription can be encoded without the need for further comment. A <note>, however, is one of the most flexible TEI elements and allows the editor and transcriber to supply a free-text description. In the IGTP scheme, an editorial note should be added at the point to which it refers and given a unique ‘xml:id’ attribute. A distinction is made between notes intended for publication in an online edition and internal queries raised locally by transcribers, using the ‘type’ attribute.

The <note> element constitutes an exception to the principle that the only text which does not occur within markup brackets is that which is physically extant on the page. Despite its flexibility, however, it seems advisable to restrict <note> to editorial material rather than use it to convey complex written material, such as paratextual elements. A further use for the <note> element is in the editorial identification of lectionary passages or the biblical references for initial chapter or canon titles. In addition, it may be used in conjunction with the empty <handShift/> element to indicate a change of copyist.

2.8. *Marginal material and text positioning*

The shortcomings of the TEI *P5 Guidelines* regarding the encoding of the physical layout of text in a document were mentioned above in 2.2. The description of text in margins is a case in point for New Testament manuscripts. A small selection of elements (including <fw> and <note>) has the attribute ‘place’, where the location of the element may be supplied. However, this is not available for section numbers (when encoded with <num>) or variant readings (apart from additions). Instead, the solution has been to adopt the same approach for all marginal material, which is to designate it as an arbitrary segment, <seg type=“margin”>, which may contain (and be contained by) the majority of XML ele-

ments⁴⁷. The use of a 'subtype' attribute permits the further specification of the precise location (e.g. 'lineleft', 'pagetop' etc.). In addition, the 'n' attribute links it to the xml:id of the relevant page, column or line in that witness (e.g. n="@P7rC1L45-33"). This should enable it to be displayed at the appropriate place on the page as well as being processed at the relevant point in the flow of the text.⁴⁸ According to the same logic, the most suitable place to encode chapter or section numbers is at the beginning of the verse to which they refer rather than the beginning of the line; the downside of this is that the transcriber does not handle them in sequence, unless they are already present in the base text. The <seg> element can also be used to record the position of other displaced text on a page, for example corrections added above the line or the final characters of a word added in blank space after the line below.

3. Conclusion

This paper has attempted to give an account of the development of encoding formats for electronic transcriptions used by the INTF and IGNTP and explain some of the reasoning behind the decisions made with regard to the current scheme of XML encoding. It is hoped that this – as well as the other fruits of the *Workspace for Collaborative Editing* – will be of interest and assistance both to those who use these transcriptions and those working on other textual traditions who are faced with similar problems. The aim has been to describe the approach from a philological rather than a technical standpoint: full documentation for each user group is available when consulting the TEI *P5 Guidelines* themselves, to which the reader is referred. Keeping within the broad guidelines of the TEI brings a welcome uniformity to electronic transcriptions but it is rarely possible to have a 'one size fits all' approach, hence the subset adopted for the *Editio Critica Maior* of the New Testament. As noted, both the transcriptions and guidelines are made available through UBIRA, where future updates may also be found.

The switch to the digital medium has affected not just the format of a transcription, but also the process of transcribing. Starting from an electronic base text, which will usually agree around 90% with the text of the

⁴⁷ The absence of <div>, however, could cause a problem, although the hierarchy of work outlined above expects all text within a <div> to be further located within <ab> or <fw> elements.

⁴⁸ In the *Digital Codex Sinaiticus*, a <ptr/> element was additionally inserted at the relevant place in the margin, linked to the xml:id of the corresponding element.

manuscript under consideration, means that transcribers focus on the differences between the manuscript and the base rather than simply trying to copy the manuscript's text word for word. One advantage of this is the reduction of the potential for harmonisation to a 'mental text' when transcribing, although there is the concomitant danger of failing to notice a difference and leaving the base text unchanged. Secondly, encoding a manuscript to permit its presentation as a 'textual facsimile' means that all formatting information must be explicitly encoded. In a handwritten copy of a text, column and line breaks (and other aspects of the *mise-en-page*) can be reproduced almost without thinking: the task of including this information means that the transcriber inherits the duties of a compositor as well as a copyist. Nonetheless, this attention to format as well as text, to the physical characteristics of a manuscript as well as its contents, is in keeping with the recent emphasis in New Testament textual criticism on manuscripts as documents as well as tradents of the text. The flexibility of presentation offered by XML encoding means that not only can transcriptions serve as 'textual facsimiles' alongside images of a manuscript, but that it will become ever easier to search for and compare physical information (as well as orthographic data and other scribal habits) in the study of New Testament manuscripts.

Finally, the potential of electronic transcriptions for revision, adjustment and re-use means that, with the entry of the New Testament into the digital sphere, a new chapter will have to be written on the transmission of the text. As with the transcriptions themselves and the manuscripts before them, some information will be explicitly included by the transcriber (in the form of editorial notes or comments in the header), while other trends may only become apparent over time. In this way, perhaps, the electronic scriptorium is no different from its predecessors. Even so, while attempts at standardisation often founder in the face of the mass of material, different practices, and varieties of use for the end product, certain innovations have become accepted. Most of these are in the realm of formatting, such as chapter and verse numbers, book titles, or the Eusebian Apparatus. It remains to be seen what lasting effects may be produced by the transition to electronic transcriptions.

References

- Burnard, L., and S. Bauman, eds., *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Charlottesville VA: Text Encoding Initiative Consortium, 2007).
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>.
- Burnard, L., S. Bauman and S. DeRose, eds., *TEI P4: Guidelines for Electronic Text Encoding and Interchange* (Charlottesville VA: Virginia UP, 2001).
- Burton, P.H., H.A.G. Houghton, R.F. MacLachlan and D.C. Parker, eds., *Vetus Latina 19. Iohannes* (Freiburg-im-Breisgau: Herder, 2011–).
- DeRose, S., 'Trojan Markup and Other XML Milestone-Tagging Techniques', in *International Workshop on Markup of Overlapping Structures 2007*.
<http://www.mulberrytech.com/overlap/index.html>.
- Estienne, R. (Stephanus), *Novum Iesu Christi D(omini) N(ostri) Testamentum ex Bibliotheca Regia* (third edition) (Paris: Stephanus, 1550).
- Finney, T.J., 'Manuscript Markup', in *The Freer Biblical Manuscripts: Fresh Studies of An American Treasure Trove*. (SBLTCS 6), ed. L. W. Hurtado (Atlanta: Society of Biblical Literature, 2006), 263–287.
- Gwynn, E.J., *Liber Ardmachanus: The Book of Armagh* (Dublin: Hodges, Figgis and Co., 1913).
- Hoskier, H.C., *A New and Complete Edition of the Irish Latin Gospel Codex Usser. 2 or r₂ otherwise known as 'The Garland of Howth'* (London: Quaritch, 1919).
- Houghton, H.A.G., 'IGNTP Guidelines for XML transcriptions of New Testament manuscripts' (2013).
<http://epapers.bham.ac.uk/1727/>
- , 'The Electronic Transcription of Codex Sinaiticus', in *Синайский кодекс и памятники древней христианской письменности: традиции и инновации в современных исследованиях [Codex Sinaiticus and Old Manuscripts of Early Christian Writing: Traditions and Innovations in Modern Research]*, ed. by E. Krushelnitskaya and Z.L. Levshina. (St Petersburg: National Library of Russia, 2012).
- Lake, K., *Codex Sinaiticus Petropolitanus: The New Testament, the Epistle of Barnabas and the Shepherd of Hermas* (Oxford: Clarendon, 1911).
- McGurk, P., 'The Disposition of Numbers in Latin Eusebian Canon Tables', in *Philologia Sacra I. Aetates und Neues Testament*, ed. by R. Gryson (Freiburg-im-Breisgau: Herder, 1993), 242–58, repr. in *Gospel Books and Early Latin Manuscripts*, P. McGurk (Aldershot: Ashgate, 1998).
- O'Loughlin, T., 'The Biblical Text of the Book of Deer (C.U.L. II.6.32): Evidence for the Remains of a Division System from its Manuscript Ancestry', in *Studies on The Book of Deer*, ed. by K. Forsyth (Dublin: Four Courts Press, 2008), 3–31.
- Parker, D.C., 'Through a Screen Darkly: Digital Texts and the New Testament', *Journal for the Study of the New Testament* 25 (2003), 395–411.
- , 'Electronic Religious Texts: The Gospel of John', in *Electronic Textual Editing*, ed. by L. Burnard, K. O'Brian O'Keefe and J. Unsworth (New York: MLA, 2006), 202–205.
- , *An Introduction to the New Testament Manuscripts and their Texts* (Oxford: Oxford UP, 2008).
- , *Codex Sinaiticus. The Story of the World's Oldest Bible* (London: British Library, 2010).
- , *Textual Scholarship and the Making of the New Testament* (Oxford: Oxford UP, 2012).
- Popovic, V., 'Du nouveau sur les Évangiles de Split', *Bulletin de la société nationale des antiquaires de France* 1990 (1992), 275–93
- Rettig, H.C.M., *Antiquissimus Quatuor Evangeliorum Canonice Codex Sangallensis Graeco-Latinus Interlinearis Nunquam Adhuc Collatus. Ad Similitudinem Ipsius Libri Manu Scripti Accuratissime Delineandum Et Lapidibus Exprimendum* (Zurich: Schulthess, 1836).
- Robinson, P.M.W., *Collate: Interactive Collation of Large Textual Traditions*, Version 2, Computer Program distributed by the Oxford University Centre for Humanities Computing (Oxford, 1994).

- , *Collate 2. Manual* (Oxford: The Computers and Variant Texts Project, 1994).
- , *Anastasia: Analytical System Tools and SGML/XML Integration Applications*, Version 2.0, Computer Program (Scholarly Digital Editions: Nottingham, 2003).
Documentation available online at <http://sd-editions.com/anastasia/index.html>.
- Schmid, U.B., 'Scribes and Variants – Sociology and Typology', in *Textual Variation: Theological and Social Tendencies* (T&S 3.6), ed. by H.A.G. Houghton and D.C. Parker (Gorgias: Piscataway NJ, 2008), 16–23.
- Thompson, E.M., *Facsimile of the Codex Alexandrinus*, (4 vols.) (London: British Museum, 1879-1883).
- von Tischendorf, C., *Biblorum Codex Sinaiticus Petropolitanus* (Leipzig: Giesecke & Devrient, 1862).
- Wachtel, K., 'Editing the Greek New Testament on the Threshold of the Twenty-first Century', *Literary and Linguistic Computing* 15 (2000), 43–50.
- Wilson, N.G., *A Descriptive Catalogue of the Greek Manuscripts of Corpus Christi College, Oxford* (Cambridge: D.S. Brewer, 2011).

APPENDIX : SCHEMATIC OVERVIEW OF THE XML STRUCTURE FOR NEW
TESTAMENT TRANSCRIPTIONS

(based on Version 1.4, July 2013)

```
<TEI>
<teiHeader>
  <fileDesc>
    <titleStmt />
    <editionStmt />
    <publicationStmt />
    <sourceDesc>
      <msDesc />
    </sourceDesc>
  </fileDesc>
  <encodingDesc />
  <revisionDesc />
</teiHeader>

<text>
  <body>
    <pb type="page/folio" n="" xml:id="" />
    <cb n="" xml:id="" />
    <lb n="" xml:id="" />
    All the above location elements <pb> <cb> <lb> may be followed by:
    <seg type="margin" subtype="" n="@...">
      subtypes: pagetop pagebottom pageleft pageright
      coltop colbottom colleft colright lineleft lineright
    <seg type="line" subtype="" n="@...">
      subtypes: above below here overwritten
    <seg type="other" subtype="" n="@...">
      <seg> may contain <fw>, <note> and <num> elements
      as well as further breaks
    <fw type="" />
      types: pageNum quireSig runTitle chapTitle lectTitle
    <num type="" n="" />
      types: chapNum AmmSec EusCan
    <note type="" n="" />
      types: editorial local canonRef
    <lb> may include the following rend attributes:
      centerJust rightJust indent hang

  <div type="book" n="B$$" >
    Optional <div type="preface"> <div type="capitula/kephalaia">
      <div type="incipit"> <div type="explicit">
  <div type="chapter" n="B$$K$" >
  <ab n="B$$K$V$" >
    Contents of <ab> may include:
```

<w /> <pc /> <num /> <note /> <gap /> <fw />

The elements <w> <pc> <num> may contain or be contained within:

<hi> (rend="colour", rend="cap", height="")

<unclear> (reason="")

<supplied> (source="" reason="")

<foreign xml:lang="">

<seg type="margin" subtype="" n="@...">

The elements <w> <pc> <num> may be contained within:

<app><rdg type="" n="" id="">...</rdg></app>

The element <w> may contain:

<abbr type="nomSac/num"> <ex>

The element <note> must contain type="" and may contain id=""

note-types: editorial local canonRef

<gb n="" /> (optional quire element)

</ab></div></div>

</body>

</text>

</TEI>

ILLUSTRATION: SAMPLE XML OF A PAGE OF TRANSCRIPTION OF A NEW TESTAMENT MANUSCRIPT

(Codex Bezae: Gregory–Aland 05, Cambridge, University Library Nn. 2.41, fol. 114v. The transcription may be seen in context alongside the image online in the *Cambridge University Digital Library* <http://cudl.lib.cam.ac.uk/view/MS-NN-00002-00041/198.>)

```
<pb xml:id="P114v-05" n="114v" type="Quire_15-2v" subtype="flesh"
facs="http://cudl.lib.cam.ac.uk/view/MS-NN-00002-00041/198/" />
<div type="book" n="04" part="M"><div type="chapter" n="4" part="M"><ab
n="B04K4V1" xml:lang="grc" part="F">
<seg type="margin" subtype="pagetop"><seg type="margin" sub-
type="lineleft"><fw type="chapRef" rend="ink">ΙΙΙΙ</fw></seg><fw
type="runTitle" place="center"><w n="2">κατ<ex>α</ex></w><w
n="4">ιωαννη</w></fw></lb>

<seg type="margin" subtype="lineleft"><fw type="titlos" rend="M2">της
σαμαρητιδο<ex>ς</ex></fw></seg></seg>

<cb xml:id="P114vC1-05" n="1"/>

<lb xml:id="P114vC1L1-05" n="1"/><w n="20">οτι</w><w n="22"><abbr
type="nomSac"><hi rend="ol">ης</hi></abbr></w><w
n="24">πλειονας</w><w n="26">μαθητας</w><w n="28">ποιει</w>

<lb xml:id="P114vC1L2-05" n="2"/><w n="30">και</w><w
n="32">βαπτισει</w><w n="34">η</w><w
n="36">ιωαννης</w><pc>·</pc><ab n="B04K4V2" xml:lang="grc"><w
n="2">καιτοιγες</w><w n="4">αυτος</w><w n="6"><abbr
type="nomSac"><hi rend="ol">ης</hi></abbr></w>

<lb xml:id="P114vC1L3-05" n="3"/><w n="8">ουκ</w><w
n="10">εβαπτισεν</w><w n="12">αλλ</w><w n="14">οι</w><w
n="16">μαθηται</w><w n="18">αυτου</w>

<lb xml:id="P114vC1L4-05" n="4"/></ab><ab n="B04K4V3"
xml:lang="grc"><seg type="margin" subtype="lineleft"><num type="AmmSec"
n="32"><hi rend="ol">λβ</hi></num></seg><w n="2">αφηκεν</w><w
n="4">την</w><w n="6">ιουδαιαν</w><w n="8">γην</w></seg>
```

type="margin" subtype="lineright"><fw type="lectTitle"
rend="L">τελοξ</fw></seg>

<lb xml:id="P114vC1L5-05" n="5"/><w n="10">και</w><w
n="12">απηλθεν</w><w n="14">παλιν</w><w n="16">ειξ</w><w
n="18">την</w><w n="20">γαλιλαιαν</w>

<lb xml:id="P114vC1L6-05" n="6"/></ab><ab n="B04K4V4"
xml:lang="grc"><seg type="margin" subtype="lineleft"><fw type="lectTitle"
rend="L">ανναγ<lb break="no"/>νοσμα</fw><num type="AmmSec"
n="33"><hi rend="ol">λγ</hi></num></seg><w n="2">εδει</w><w
n="4">δε</w><w n="6">αυτον</w><w n="8">διερχεσθαι</w><w
n="10">δια</w><w n="12">της</w><w n="14">σαμαριας</w>

<lb xml:id="P114vC1L7-05" n="7"/></ab><ab n="B04K4V5"
xml:lang="grc"><w n="2">ερχεται</w><w n="4">ουν</w><w
n="6">ειξ</w><w n="8">πολιν</w><w n="10">της</w><w
n="12">σαμαριας</w>

<lb xml:id="P114vC1L8-05" n="8"/><w n="14">λεγομενην</w><w
n="16">συχαρ</w><pc>·</pc><w n="18">πλησιον</w><w
n="20">του</w><w n="22">χωριου</w>

<lb xml:id="P114vC1L9-05" n="9"/><w n="24">ου</w><w
n="26">εδωκεν</w><w n="28">ιακωβ</w><pc>·</pc><w
n="30">ιωσηφ</w><w n="32">τω</w><w n="34">υιω</w><w
n="36">αυτου</w>

<lb xml:id="P114vC1L10-05" n="10" rend="hang"/></ab><ab n="B04K4V6"
xml:lang="grc"><w n="2">ην</w><w n="4">δε</w><w n="6">εκει</w><w
n="8">πηγη</w><w n="10">του</w><w n="12">ιακωβ</w><pc>:</pc>

<lb xml:id="P114vC1L11-05" n="11"/><w n="14">ο</w><w
n="16">ουν</w><w n="18"><abbr type="nomSac"><hi
rend="ol">ιηξ</hi></abbr></w><w n="20">κεκοπιακως</w>

<lb xml:id="P114vC1L12-05" n="12"/><w n="22">εκ</w><w
n="24">της</w><w n="26">οδοιποριας</w><w n="28">εκαθεξετο</w><w
n="30">ουτως</w>

<lb xml:id="P114vC1L13-05" n="13"/><w n="32">επι</w><w
n="34">τη</w><w n="36">πηγη</w><space unit="char" extent="1"/><w
n="38">ωρα</w><w n="40">ην</w><w n="42">ως</w><w
n="44">εκτη</w>

<lb xml:id="P114vC1L14-05" n="14"/></ab><ab n="B04K4V7"
xml:lang="grc"><w n="2">ερχεται</w><w n="4">γυνη</w><w>

n="6">εκ</w><w n="8">της</w><w n="10">σαμαριας</w><w n="12">αντλησαι</w><w n="14">υδωρ</w>

<lb xml:id="P114vC1L15-05" n="15" rend="hang"/><w n="16">λεγει</w><w n="18">αυτη</w><w n="20">ο</w><w n="22"><abbr type="nomSac"><hi rend="ol">της</hi></abbr></w><w n="24">δος</w><w n="26">μοι</w><w n="28">πειν</w>

<lb xml:id="P114vC1L16-05" n="16"/></ab><ab n="B04K4V8" xml:lang="grc"><w n="2">οι</w><w n="4">γαρ</w><w n="6">μαθηται</w><w n="8">αυτου</w><w n="10">απεληλυθισαν</w>

<lb xml:id="P114vC1L17-05" n="17"/><w n="12">εις</w><w n="14">την</w><w n="16">πολιν</w><pc>·</pc><w n="18">ινα</w><w n="20">τροφας</w><w n="22">αγορασωσιν</w>

<lb xml:id="P114vC1L18-05" n="18"/></ab><ab n="B04K4V9" xml:lang="grc"><w n="2">λεγει</w><w n="4">ουν</w><w n="6">αυτω</w><w n="8">η</w><w n="10">γυνη</w><w n="12">η</w><w n="14">σαμαριτις</w>

<lb xml:id="P114vC1L19-05" n="19"/><w n="16">συ</w><w n="18">ιουδαιος</w><w n="20">ων</w><w n="22">πως</w><w n="24">παρ</w><w n="26">εμου</w>

<lb xml:id="P114vC1L20-05" n="20"/><w n="28">πειν</w><w n="30">αιτεις</w><w n="32">γυναικος</w><w n="34">σαμαρτιδος</w>

<lb xml:id="P114vC1L21-05" n="21" rend="hang"/></ab><ab n="B04K4V10" xml:lang="grc"><w n="2">απεκριθη</w><w n="4">ο</w><w n="6"><abbr type="nomSac"><hi rend="ol">της</hi></abbr></w><w n="8">και</w><w n="10">ειπεν</w><w n="12">αυτη</w>

<lb xml:id="P114vC1L22-05" n="22"/><w n="14">ει</w><w n="16">ηδεις</w><w n="18">την</w><w n="20">δωρεαν</w><w n="22">του</w><w n="24"><abbr type="nomSac"><hi rend="ol">θυ</hi></abbr></w>

<lb xml:id="P114vC1L23-05" n="23"/><w n="26">και</w><w n="28">τις</w><w n="30">εστιν</w><w n="32">ο</w><w n="34">λεγων</w><w n="36">σοι</w><w n="38">δος</w><w n="40">μοι</w><w n="42">πειν</w>

<lb xml:id="P114vC1L24-05" n="24"/><w n="44">συ</w><app><rdg type="orig" hand="p.m."><w n="46">ν</w></rdg><rdg type="corr" hand="A"><w n="46">αν</w></rdg></app><w n="48">ητησας</w><w n="50">αυτον</w>

<lb xml:id="P114vC1L25-05" n="25"/><w n="52">και</w><w n="54">εδωκεν</w><w n="56">αν</w><w n="58">σοι</w><w n="60">υδωρ</w><w n="62">ζων</w>

<lb xml:id="P114vC1L26-05" n="26" rend="hang"/></ab><ab n="B04K4V11" xml:lang="grc"><w n="2">λεγει</w><w n="4">αυτω</w><w n="6">η</w><w n="8">γυνη</w><w n="10"><abbr type="nomSac"><hi rend="ol">κε</hi></abbr></w><w n="12">ουδε</w><w n="14">αντλημα</w><w n="16">εχεις</w>

<lb xml:id="P114vC1L27-05" n="27"/><w n="18">και</w><w n="20">το</w><w n="22">φρααρ</w><w n="24">εστι</w><w n="26">βαθυ</w><pc>·</pc><w n="28">ποθεν</w><w n="30">εχεις</w><w n="32">υδωρ</w><w n="34">ζων</w>

<lb xml:id="P114vC1L28-05" n="28"/></ab><ab n="B04K4V12" xml:lang="grc"><w n="2">μη</w><w n="4">συ</w><w n="6">μειζων</w><w n="8">ει</w><w n="10">του</w><w n="12"><abbr type="nomSac"><hi rend="ol">πρς</hi></abbr></w><w n="14">ημων</w><w n="16">ιακωβ</w>

<lb xml:id="P114vC1L29-05" n="29"/><w n="18">ος</w><w n="20">εδωκεν</w><w n="22">ημειν</w><w n="24">το</w><w n="26">φρααρ</w>

<lb xml:id="P114vC1L30-05" n="30"/><w n="28">και</w><w n="30">αυτος</w><w n="32">εξ</w><w n="34">αυτου</w><w n="36">επιεν</w>

<lb xml:id="P114vC1L31-05" n="31"/><w n="38">και</w><w n="40">οι</w><w n="42">υιοι</w><w n="44">αυτου</w><w n="46">και</w><w n="48">τα</w><w n="50">θρεμματα</w>

<lb xml:id="P114vC1L32-05" n="32" rend="hang"/></ab><ab n="B04K4V13" xml:lang="grc" part="I"><w n="2">απεκριθη</w><w n="4"><abbr type="nomSac"><hi rend="ol">ης</hi></abbr></w><w n="6">και</w><w n="8">επιεν</w><w n="10">αυτη</w><w n="12">πας</w><w n="14">ο</w><w n="16">πεινων</w>

<lb xml:id="P114vC1L33-05" n="33"/><w n="18">εκ</w><w n="20">του</w><w n="22">υδατος</w><w n="24">τουτου</w><pc>·</pc><w n="26">διψησει</w><w n="28">παλιν</w></ab></div></div>