The Ghost in the Machine
Revisiting an Old Model for the Dynamic Generation of Digital Editions

Daniel Paul O'Donnell

"The Electronic *Cædmon's Hymn* Editorial Method" (1998)

In 1998, a few months into the preparation of my electronic edition of the Old English poem *Cædmon's Hymn* (O'Donnell forthcoming), I published a brief prospectus on the "editorial method" I intended to follow in my future work (O'Donnell 1998). Less a true *editorial* method than a proposed workflow and list of specifications, the prospectus called for the development of an interactive edition-processor by which "users will [...] be able to generate mediated ('critical') texts on the fly by choosing the editorial approach which best suits their individual research or study needs" (O'Donnell 1998, ¶ 1).

The heart of the prospectus was a diagram of the "Editorial Process Schema" I intended to follow (figure 1). The edition was to be based on TEI (P2) SGML-encoded diplomatic transcriptions of all twenty-one known witnesses to the poem. Its output was to consist of dynamically generated "HTML/XML" display texts that would allow users access to different views of the underlying textual data depending on their specific interests: e.g. editions containing reconstructions of archetypal texts, student texts based on witnesses showing the simplest vocabulary and grammar, "best text" editions of individual witnesses or recensions, etc. The production of these display texts was to be handled by a series of SGML "filters" or "virtual editions" that would be populated by the unspecified processor used to format and display the final output.

The Electronic *Cædmon's Hymn*
Editorial Process Schema

Witness A	Witness B	Witness C	Witness D	Witness E	Witness F

Witness Archive Level
Facsimiles and Diplomatic transcriptions in TEI Markup

Virtual Text A — Reconstruction of 'Author's Original'	Virtual Text B — Normalised text of Single Witness	Virtual Text C — Simplified Student Edition	Virtual Text D — Edition of Dialect A Recension	Virtual Text E — Edition of Dialect B Recension

Filter Level
Empty SGML (TEI) 'virtual texts'

Processor Level
Formatting and conversions of TEI text

Gets actual readings from witnesses as noted in virtual texts
Supplies emendations/alternative readings
Normalises spelling (as necessary); add modern punctuation as specified in virtual eds.
Formats display and converts SGML coding to HTML/XML coding

Display Text A — Reconstruction of 'Author's Original'	Display Text B — Normalised text of Single Witness	Display Text C — Simplified Student Edition	Display Text D — Edition of Dialect A Recension	Display Text E — Edition of Dialect B Recension

Display Level
HTML/XML formatted texts.

User

Note: all witness affiliations and editorial decisions are hypothetical
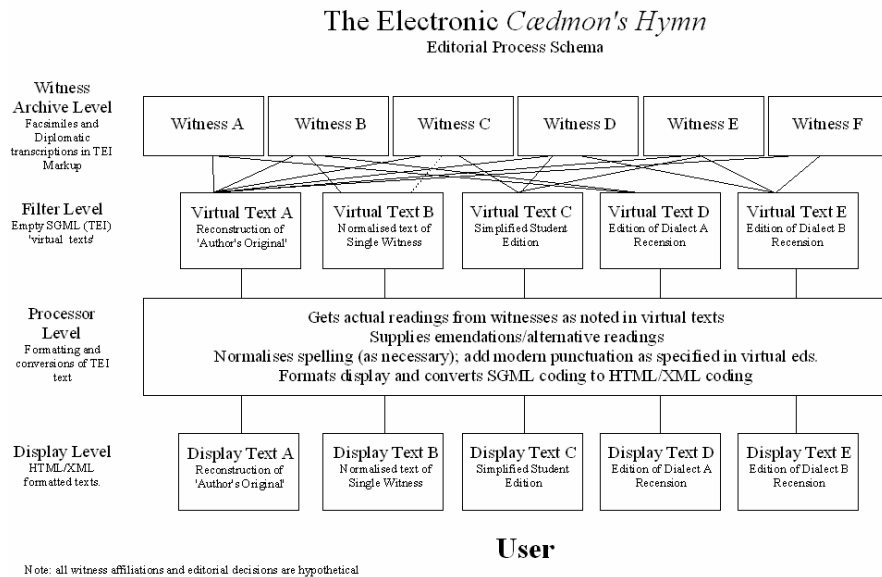
Figure 1. Editorial Process Schema (O'Donnell 1998)

Goals

The initial impetus for this approach was practical. Although it is quite short, *Cædmon's Hymn* has a relatively complex textual history for an Anglo-Saxon poem. Even in print, it has always been edited as a multi-text. The standard print edition (Dobbie 1942) reproduces two editorial versions of the poem without commenting on their relative priority. Few other studies have managed to be even this decisive. Dobbie's text was the last (before my forthcoming edition) to attempt to produce critical texts based on the entire manuscript tradition. Most editions before and since have concentrated on individual recensions or groups of witnesses.[1] Anticipating great difficulty in proof-reading an electronic edition that might have several editorial texts and multiple textual apparatus,[2] I was at this early stage keenly interested in reducing the opportunity for typographical error. A workflow that would allow me to generate a number of

different critical texts from a single set of diplomatic transcriptions without retyping was for this reason an early desideratum.

This convenience, however, was not to come at the expense of editorial content: a second important goal of my prospectus was to find an explicit home for the editor in what Murray McGillivray recently had described as a "post-critical" world (McGillivray 1994; see also Ross 1996; McGann 1997). In medieval English textual studies in 1998, indeed, this post-critical world seemed to be fast approaching: the first volume of the *Canterbury Tales Project*, with its revolutionary approach to electronic collation and stemmatics and a lightly-edited guide text, had been published two years earlier (Robinson 1996). Forthcoming publications from the *Piers Plowman Electronic Archive* (Adams *et al.* 2000) and *Electronic Beowulf* (Kiernan 1999) projects, similarly, promised a much heavier emphasis on the manuscript archive (and less interest in the critical text) than their more traditional predecessors. My initial work with the *Cædmon's Hymn* manuscripts (e.g. O'Donnell 1996a; O'Donnell 1996b), however, had convinced me that there was a significant need in the case of this text for both user access to the witness archive and editorial guidance in the interpretation of this primary evidence – or, as Mats Dahlström later would point out, that the two approaches had complementary strengths and weaknesses:

> *The single editor's authoritative control in the printed SE [Scholarly Edition], manifested in e.g. the versional prerogative, isn't necessarily of a tyrannical nature. Conversely, the much spoken-of hypermedia database exhibiting all versions of a work, enabling the user to choose freely between them and to construct his or her "own" version or edition, pre-supposes a most highly competent user, and puts a rather heavy burden on him or her. Rather, this kind of ultra-eclectic archive can result in the user feeling disoriented and even lost in hyperspace. Where printed SE:s tend to bury rival versions deep down in the variant apparatuses, the document architecture of extreme hypertext SE:s, consequential to the very nature of digitally realised hypertext, threatens to bury the user deep among the mass of potential virtuality.* (Dahlström 2000, 17)

Keen as I was to spare myself some unnecessary typing, I did not want this saving to come at the expense of providing access to the "insights and competent judgement" (Dahlström 2000, 17) I hoped to acquire in several years' close contact with the manuscript evidence. What I needed, in other words, was a system in which the computer would generate, but a human edit, the final display texts presented to the reader.

Theory

In order to accomplish these goals, the prospectus proposed splitting the editorial process into distinct phases: a transcription phase, in which human scholars recorded information about the text as it appeared in the primary sources (the "Witness Archive"); an editorial ("Filtering") phase, in which a human editor designed a template by which a display text was to be produced from the available textual evidence ("Virtual Editions"); a processing phase, in which a computer applied these filters to the Witness Archive; and a presentation phase, in which the resultant output was presented to the reader. The first and second stages were to be the domains of the human editor; the third and fourth that of the computer. An important element of this approach was the assumption that the human editor, even in traditional print sources, functioned largely as a rules-based interpreter of textual data – or as I (in retrospect unfortunately) phrased it, could be "reduced to a set of programming instructions"[3] – in much the same way as a database report extracts and format specific information from the underlying data table of a database:

> *In my view, the editor of a critical edition is understood as being functionally equivalent to a filter separating the final reader from the uninterpreted data contained in the raw witnesses. Depending on the nature of the instructions this processor is given, different types of manipulation will occur in producing the final critical edition. An editor interested in producing a student edition of the poem, for example, can be understood to be manipulating the data according to the instructions "choose the easiest (most sensible) readings and ignore those which raise advanced textual problems"; an editor interested in producing the "original" text can be seen as a processor performing the instruction "choose readings from the earliest manuscript(s) when these are available*

and sensible; emend or normalise readings as required"; and an editor interested in producing an edition of a specific dialectal version of a text is working to the instruction "choose readings from manuscripts belong to dialect x; when these are not available, reconstruct or emend readings from other manuscripts, ensuring that they conform to the spelling rules of the dialect". (O'Donnell 1998, ¶¶ 4 f.)

Advantages

From a theoretical perspective, the main advantage of this approach was that it provided an explicit location for the encoding of editorial knowledge – as distinct from textual information about primary sources, or formatting information about the final display. By separating the markup used to describe a text's editorial form from that used to describe its physical manifestation in the witnesses, or its final appearance to the end user, this method made it easier in principle both to describe phenomena at a given level in intrinsically appropriate terms and to modify, reuse, or revise information at each level without necessarily having to alter other aspects of the edition design – in much the same way as the development of structural markup languages themselves had freed text encoders from worrying unduly about final display. Scholars working on a diplomatic transcription of a manuscript in this model would be able to describe its contents without having to ensure that their markup followed the same semantic conventions (or even DTD) as that used at the editorial or display levels.

Just as importantly, the approach was, in theory at least, infinitely extensible. Because it separated transcription from editorial activity, and because it attempted to encode editorial activity as a series of filters, users were, in principle, free to ignore, adapt, add to, or replace the work of the original editor. Scholars interested in statistical or corpus work might choose to work with raw SGML data collected in the witness archive; those interested in alternative editorial interpretations might wish to provide their own filters; those wishing to output the textual data to different media or using different display formats were free to adapt or substitute a different processor. Espen S. Ore recently has discussed how well-made and suitably-detailed transcriptions of source material might be used or adapted profitably by other scholars and projects as the basis

for derivative work (Ore 2004); from a theoretical perspective the "Editorial Method" proposed for use in *Cædmon's Hymn* offered an early model for how such a process might be built into an edition's design. Indeed, the method in principle allowed editors of new works to operate in the other direction as well: by building appropriate filters, editors of original electronic editions could attempt to model the editorial decisions of their print-based predecessors, or apply techniques developed for other texts to their own material.[4]

Implementation (1998)

Despite its theoretical attractiveness, the implementation of this model proved, in 1998, to be technically quite difficult. The main problem was access to technology capable of the type of filtering envisioned at the Virtual Edition level. In the original model, these "editions" were supposed to be able both to extract readings from multiple source documents (the individual witness transcriptions) and to translate their markup from the diplomatic encoding used in the original transcriptions to that required by the new context – as a reading used in the main text of a critical edition, say, or a form cited in an apparatus entry, textual note, or introductory paragraph. This type of transformation was not in and of itself impossible to carry out at the time: some SGML production environments and several computer languages (e.g. DSSSL or, more generally, Perl and other scripting languages) could be used to support most of what I wanted to do; in the days before XSL, however, such solutions were either very much cutting edge, or very expensive in time and/or resources. As a single scholar without a dedicated technical staff or funding to purchase commercial operating systems, I was unable to take full advantage of the relatively few transformation options then available.

The solution I hit upon instead involved dividing the transformation task into two distinct steps (extraction and translation) and adding an extra processing level between the witness and virtual edition levels in my original schema:
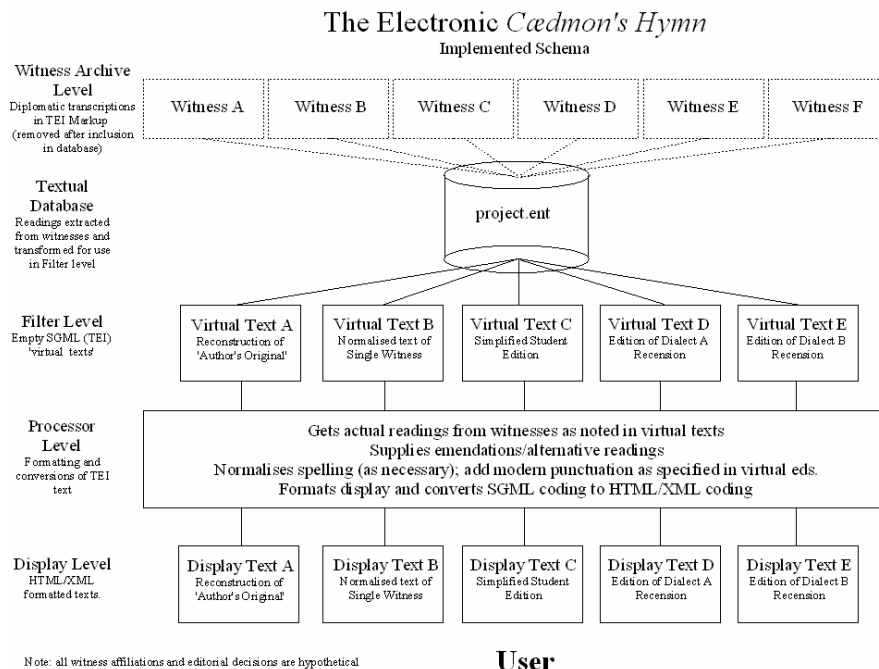
The Electronic *Cædmon's Hymn*
Implemented Schema

Witness Archive Level
Diplomatic transcriptions in TEI Markup (removed after inclusion in database)

| Witness A | Witness B | Witness C | Witness D | Witness E | Witness F |

Textual Database
Readings extracted from witnesses and transformed for use in Filter level

project.ent

Filter Level
Empty SGML (TEI) 'virtual texts'

| Virtual Text A Reconstruction of 'Author's Original' | Virtual Text B Normalised text of Single Witness | Virtual Text C Simplified Student Edition | Virtual Text D Edition of Dialect A Recension | Virtual Text E Edition of Dialect B Recension |

Processor Level
Formatting and conversions of TEI text

Gets actual readings from witnesses as noted in virtual texts
Supplies emendations/alternative readings
Normalises spelling (as necessary); add modern punctuation as specified in virtual eds.
Formats display and converts SGML coding to HTML/XML coding

Display Level
HTML/XML formatted texts.

| Display Text A Reconstruction of 'Author's Original' | Display Text B Normalised text of Single Witness | Display Text C Simplified Student Edition | Display Text D Edition of Dialect A Recension | Display Text E Edition of Dialect B Recension |

Note: all witness affiliations and editorial decisions are hypothetical

User

Figure 2. Implemented Schema

Instead of acting as the locus of the transformation, the editorial filters in this revised model provided a context for text that had been previously extracted from the witness archive and transformed for use in such circumstances. The text these filters called upon was stored in a textual database as part of the project's entity extension file (`project.ent`, see Sperberg-McQueen and Burnard 2004, § 3.3), and hence resident in the project DTD. The database itself was built by extracting marked-up readings from the original witness transcription files (using grep) and converting them (using macros and similar scripts) to entities that could be called by name anywhere in the project. Transformations involving a change in markup syntax or semantics (e.g. from a diplomatic-linguistic definition of a word in witness transcriptions to a syntactic and morphological definition in the edition files) also generally were performed in this DTD extension file.

First two lines of a TEI SGML transcription of Cædmon's Hymn witness T1:

```
<l id="t1.1" n="1">
 <seg type="MSWord" id="t1.1a.1">Nu<space extent="0"></seg>
 <seg type="MSWord" id="t1.1a.2"><damage type="stain"
        degree="moderate">sculon</damage><space></seg>
 <note id="t1.1a.3.n" type="transcription" target="t1.1a.2
        t1.1a.4 t1.1b.1 t1.2b.3 t1.3a.1 t1.4a.1 t1.4a.2
        t1.4b.1 t1.6a.1 t1.6a.2 t1.7b.1 t1.7b.2
        t1.9b.2">&copyOft1.1a.2;...&copyOft1.9b.2;] Large
        stain obscures some text down inside (right) margin
        of p. 195 in facsimile. Most text is readable,
        however.</note>
 <seg type="MSWord" id="t1.1a.3"><damage type="stain"
        degree="moderate">herigean</damage><space></seg>
 <caesura>
 <seg type="MSWord" id="t1.1b.1"><damage type="stain"
        degree="light">he</damage>ofon<lb>rices<space></seg>
 <seg type="MSWord" id="t1.1b.2">&wynn;eard<space></seg>
</l>
<l id="t1.2" n="2">
 <seg type="MSWord" id="t1.2a.1">meotodes<space></seg>
 <seg type="MSWord" id="t1.2a.2">me<corr sic="u"
        cert="50%"><del rend="overwriting">u</del><add
        rend="overwriting"
        place="intralinear">a</add></corr>hte<space></seg>
 <note type="transcription" id="t1.2a.2.n" target="t1.2a.2"
        resp=dpod>&copyOft1.2a.2;] Corrected from
        <foreign>meuhte</foreign>?</note>
 <caesura>
 <seg type="MSWord" id="t1.2b.1">&tyronianNota;<space
        extent="0"></seg>
 <seg type="MSWord" id="t1.2b.2">his<space></seg>
 <seg type="MSWord" id="t1.2b.3"><damage type="stain"
        degree="severe"><unclear reason="stain in facsimile"
        cert="90%">mod</unclear></damage><damage type="stain"
        degree="moderate">ge&thorn;anc</damage><space></seg>
 <note type="transcription" id="t1.2b.3.n"
        target="t1.2b.3">&copyOft1.2b.3;] <c>mod</c> obscured by
        stain in facsimile.</note>
</l>
```

Same text after conversion to entity format (Information from the original l*,* w*,* caesura*, and* note *elements are stored separately).*

```
<!ENTITY t1.1a.1 'Nu<space type="wordBoundary" extent="0">'>
<!ENTITY t1.1a.2 'sc<damage type="stain"
        rend="beginning">ulon</damage><space
        type="wordBoundary" extent="1">'>
```

```
<!ENTITY t1.1a.3 '<damage type="stain"
         rend="middle">her&inodot;gean</damage><space
         type="wordBoundary" extent="1">'>
<!ENTITY t1.1b.1 '<damage type="stain"
         rend="end">heo</damage>fon<lb>r&inodot;ces<space
         type="wordBoundary" extent="1">'>
<!ENTITY t1.1b.2 '&mswynn;eard<space type="wordBoundary" extent="1">'>
<!ENTITY t1.2a.1 'meotodes<space type="wordBoundary" extent="1">'>
<!ENTITY t1.2a.2 'me<damage type="stain"
         rend="complete">a</damage>hte<space
         type="wordBoundary" extent="1">'>
<!ENTITY t1.2b.1 '<abbr type="scribal"
         expan="ond/and/end">&tyronianNota;</abbr><expan
         type="scribal">ond</expan><space
         type="wordBoundary" extent="0">'>
<!ENTITY t1.2b.2 'h&inodot;s<space type="wordBoundary" extent="1">'>
<!ENTITY t1.2b.3 '<damage type="stain"
         rend="beginning"><unclear
         rend="complete">mod</unclear>ge&thorn;anc</damage><space
         type="wordBoundary" extent="1">'>
```

Same text after conversion to editorial format for use in editions.

```
<!ENTITY ex.1a.1 'Nu'>
<!ENTITY ex.1a.2 'sculon'>
<!ENTITY ex.1a.3 'herigean'>
<!ENTITY ex.1b.1 'heofonrices'>
<!ENTITY ex.1b.2 '&edwynn;eard'>
<!ENTITY ex.2a.1 'meotodes'>
<!ENTITY ex.2a.2 'meahte'>
<!ENTITY ex.2b.1 'ond'>
<!ENTITY ex.2b.2 'his'>
<!ENTITY ex.2b.3 'modge&thorn;anc'>
```

Citation from the text of T1 (bold) in an introductory chapter (simplified for demonstration purposes).

```
<p id="CH6.420" n="6.42">Old English <mentioned
        lang="ANG">swe</mentioned>, <mentioned
        lang="ANG">sw&aelig;</mentioned>, <mentioned
        lang="ANG">swa</mentioned> appears as <mentioned
        rend="postcorrection"
        lang="ANG">&t1.3b.1;</mentioned> (&carmsx;
        <mentioned rend="postcorrection"
        lang="ANG">&ar.3b.1;</mentioned>) in all West-Saxon
        witnesses of the poem on its sole occurrence in 3b.
        The expected West-Saxon development is <mentioned
        lang="ANG">sw&aelig;</mentioned>, found in early
        West-Saxon. As in most dialects, however,
        <mentioned lang="ANG">swa</mentioned> develops
        irregularly in the later period. <mentioned
```

```
lang="ANG">Swa</mentioned> is the usual late
West-Saxon reflex (see &hogg1992;,
&sect; 3.25, n. 3).</p>
```

Citation from the text of T1 (bold) in a textual apparatus (simplified for demonstration purposes)

```
<app id="EX.1A.1.APP" n="1" from="EX.1A.1">
  <lem id="EX.1A.1.LEM" n="1a">&ex.1a.1;</lem>
  <rdggrp>
   <rdggrp>
    <rdggrp>
     <rdg id="T1.1A.1.RDG"
         wit="T1">&t1.1a.1;</rdg><wit><xptr doc="t1"
         from="T1.1A.1" n="T1" rend="eorthan"></wit>
     <rdg id="O.1A.1.RDG" wit="O (Pre-Correction)"><seg
         rend="precorrection">&o.1a.1;</seg></rdg><wit><xptr
         doc="o" from="O.1A.1" n="O (Pre-Correction)"
         rend="eorthan"></wit>
    </rdggrp>
   </rdggrp>
   <rdggrp>
    <rdggrp>
     <rdg id="N.1A.1.RDG" wit="N">&n.1a.1;</rdg><wit><xptr
         doc="n" from="N.1A.1" n="N" rend="eorthan"></wit>
    </rdggrp>
   </rdggrp>
   <rdggrp>
    <rdggrp>
     <rdg id="B1.1A.1.RDG"
         wit="B1">&b1.1a.1;&b1.1a.2;</rdg><wit><xptr doc="b1"
         from="B1.1A.1" n="B1" rend="eorthan"></wit>
     <rdg id="TO.1A.1.RDG"
         wit="To">&to.1a.1;&to.1a.2;</rdg><wit><xptr
         doc="to" from="TO.1A.1" n="To"
         rend="eorthan"></wit>
     <rdg sameas="O.1A.1.RDG" wit="O
         (Post-Correction)"><seg
         rend="postcorrection">&o.1a.1;&o.1a.2;</seg></rdg><wit><xptr
         doc="o" from="O.1A.1" n="O (Post-Correction)"
         rend="eorthan"></wit>
     <rdg id="CA.1A.1.RDG"
         wit="Ca">&ca.1a.1;&ca.1a.2;</rdg><wit><xptr
         doc="ca" from="CA.1A.1" n="Ca"
         rend="eorthan"></wit>
    </rdggrp>
   </rdggrp>
  </rdggrp>
 </app>
```

Implementation (2005)

The solutions I developed in 1998 to the problem of SGML transformation are no longer of intrinsic interest to Humanities Computing specialists except, perhaps, from a historical perspective. With the publication of the first XSL draft in November 1999, and, especially, the subsequent rapid integration of XSL and XML into commercial and academic digital practice, editors soon had far more powerful languages and tools available to accomplish the same ends.

Where my solutions are valuable, however, is as proof-of-concept. By dividing the editorial process into distinct phases, I was able to achieve, albeit imperfectly, both my original goals: no Old English text from the primary witnesses was input more than once in my edition and I did to a certain extent find in the "Virtual Editions" an appropriate and explicit locus for the encoding of editorial information.

With the use of XSLT, however, it is possible to improve upon this approach in both practice and theory. In practical terms, XSLT functions and instructions such as `document()` and `xsl:result-document` eliminate the need for a pre-compiled textual database: scholars using XSLT today can work, as I originally had hoped to, directly with the original witness transcriptions, extracting readings, processing them, and outputing them to different display texts using a single language and processor – and indeed perhaps even a single set of stylesheets.

In theoretical terms, moreover, the adoption of XSLT helps clarify an ambiguity in my original proposal. Because, in 1998, I saw the process of generating an edition largely as a question of *translation* from diplomatic to editorial encoding, my original model distinguished between the first two levels on largely semantic grounds. The Witness Archive was the level that was used to store primary readings from the poem's manuscripts; the filter or Virtual Edition level was used to store everything else, from transformations necessary to translate witness readings into editorial forms to secondary textual content such as introductory chapters, glossary entries, and bibliography.

In XSLT terms, however, there is no significant reason for maintaining such a distinction: to the stylesheet, both types of content are simply raw material for the transformation. What this raw material is, where it came from, or who its author is, are irrelevant to the stylesheet's task of

organisation, adaptation, interpretation, and re-presentation. While poor quality or poorly constructed data will affect the ultimate quality of its output, data composition and encoding remain, in the XSLT world, distinct operations from transformation.

This is significant because it helps us refine our theoretical model of the editorial process and further isolate the place where editorial intelligence is encoded in a digital edition. For organisation, adaptation, interpretation, and re-presentation are the defining tasks of the scholarly editor as much as they are that of the XSLT stylesheet. Change the way a standard set of textual data is interpreted, organised, adapted, or presented, and you change the nature of the final "edition". Editions of literary works are usually based on very similar sets of primary data – there is only one *Beowulf* manuscript, after all, and even better attested works usually have a relatively small set of textually significant witnesses, editions, or recensions. What differences arise between modern editions of literary texts tend for the most part to hinge on the reinterpretation of existing evidence, rather than any real change in the available data.[5] In traditional editions, the evidence for this observation can be obscured by the fact that the "editor" also usually is responsible for much of the secondary textual content. That the observation is true, however, is demonstrated by emerging forms of digital editions in which the editorial function is largely distinct from that of content creation: multi-generational and derivative editions such as those discussed by Ore (2004), as well as interactive models such as that proposed by the Virtual Humanities Lab (e.g. Armstrong & Zafrin 2005), or examples in which users reinterpret data in already existing corpora or databases (e.g. Green 2005).

Taken together, this suggests that my 1998 model was correct in its division of the editorial process into distinct tasks, but imprecise in its understanding of the editorial function.
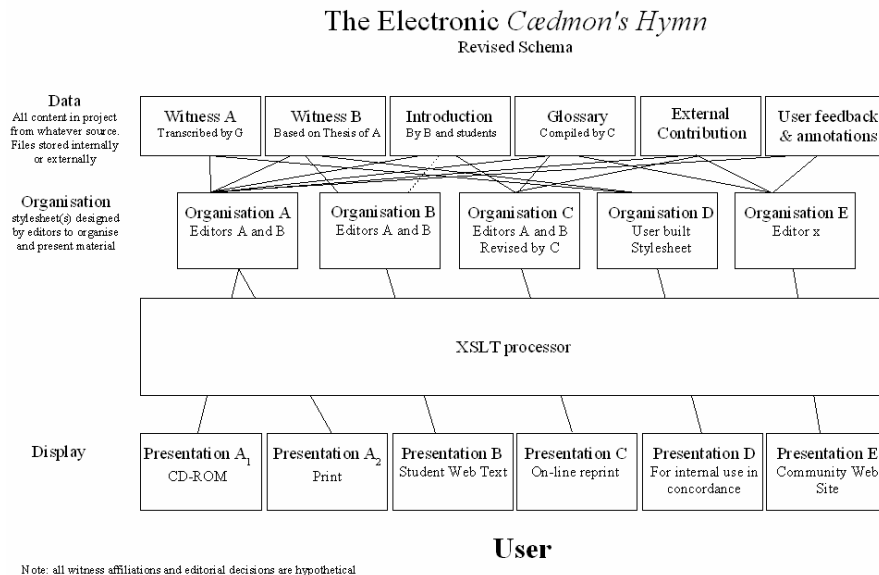
Figure 3. Revised Schema

In the revised version, the original "Witness Archive" is now reconceived of more generally as a collection of textual data used in the edition, regardless of source or type. This data is then organised, interpreted, adapted, and prepared for presentation using stylesheets (and perhaps other organisational tools) provided by an "editor" – regardless of whether this "editor" is the person responsible for assembling and/or authoring the original content, an invited collaborator, or even an end user. As in the original model, this reorganisation is then presented using an appropriate display media.

Conclusion

Technical advances of the last eight years have greatly improved our ability to extract and manipulate textual data – and our ability to build editions in ways simply impossible in print. The model for the editorial

process proposed in O'Donnell (1998) represented an early attempt to understand how the new technology might affect the way editors work, and, more importantly, how this technology might be harnessed more efficiently. With suitable modifications to reflect our field's growing sophistication, the model appears to have stood the test of time, and proven itself easily adapted to include approaches developed since its original publication. From my perspective, however, a real sign of strength is that it continues to satisfy my original two goals: it suggests a method for avoiding reinputting primary source documents, and it provides a description of the locus of editorial activity; in an increasingly collaborative and interactive scholarly world, it appears that the ghost in the machine may reside in the stylesheet.

Daniel Paul O'Donnell is an Associate Professor of English at the University of Lethbridge, Alberta, Canada. He is also director of the Digital Medievalist Project <http://www.digitalmedievalist.org/> and editor of Cædmon's Hymn: A Multimedia Study, Edition, and Archive *(D.S. Brewer, forthcoming 2005). His research interests include Old English poetry, Textual and Editorial Studies, Humanities Computing, and the History of the Book.*
E-mail: daniel.odonnell@uleth.ca
Web page: http://people.uleth.ca/~daniel.odonnell/

Notes

1. A bibliography of studies and editions of *Cædmon's Hymn* can be found in O'Donnell (forthcoming).
2. In the event, the final text of O'Donnell (forthcoming) has eight critical editions, all of which have several apparatus, and "semi-normalised" editions of all twenty-one witnesses.
3. This choice was unfortunate, as it seems that it led to my model being understood far more radically than I intended (e.g. in Dahlström 2000, 17, cited above). A perhaps better formulation would be that editors (print and digital) function in a manner analogous to (and perhaps reproducable in) progamming instructions.
4. In practice, of course, this type of modelling would work best in the case of simple, linguistically oriented exemplars. It becomes increasingly difficult – though still theoretically possible – with more complex or highly eclectic editorial approaches. A rule-based replica of Kane and Donaldson (1988), for example, is possible probably *only* in theory.
5. While this obviously does not apply in those few cases in which editions are made after the discovery of significant new textual evidence, such discoveries are few and far between. Most editorial differences are the result of a reinterpretation of essentially similar sets of textual data.

References

ADAMS, ROBERT, HOYT N. DUGGAN, ERIC ELIASON, RALPH HANNA III, JOHN PRICE-WILKIN & THORLAC TURVILLE-PETRE, EDS. (2000). *The Piers Plowman Electronic Archive, vol. 1: Corpus Christi College, Oxford MS 201 (F)*. (SEENET, series A.1.) Ann Arbor: SEENET & University of Michigan Press.

ARMSTRONG, GUYDA & VIKA ZAFRIN (2005). "Towards the Electronic *Esposizioni*: The Challenges of the Online Commentary." *Digital Medievalist* 1.1. <http://www.digitalmedievalist.org/article.cfm?RecID=1> [2005-06-21]

DAHLSTRÖM, MATS (2000). "Drowning by Versions." *Human IT* 4.4: 7-38. <http://www.hb.se/bhs/ith/4-00/md.htm> [2005-06-21]

DOBBIE, ELLIOTT VAN KIRK, ED. (1942). *The Anglo-Saxon Minor Poems*. (The Anglo-Saxon Poetic Records 6.) New York: Columbia University Press.

GREEN, JONATHAN (2005). "Opening the Illustrated Incunable Short Title Catalog on CD-ROM: An End-user's Approach to an Essential Database." *Digital Medievalist* 1.1. <http://www.digitalmedievalist.org/article.cfm?RecID=4> [2005-06-21]

KANE, GEORGE & E. TALBOT DONALDSON, EDS. (1988). *Piers Plowman: The B Version, Will's Visions of Piers Plowman, Do Well, Do Better and Do Best.* Rev. ed. London: Athlone Press.

KIERNAN, KEVIN S., ED. (1999). *The Electronic Beowulf.* London: British Library.

MCGANN, JEROME J. (1997). "The Rationale of Hypertext." *Electronic Text: Investigations in Method and Theory*. Ed. K. Sutherland. Oxford: Clarendon. 19-46. Preprint available at: <http://jefferson.village.virginia.edu/public/jjm2f/rationale.html> [2005-06-21]

MCGILLIVRAY, MURRAY (1994). "Towards a Post-Critical Edition: Theory, Hypertext, and the Presentation of Middle English Works." *Text* 7: 175-199.

O'DONNELL, DANIEL PAUL (1996a). "Manuscript Variation in Multiple Recension Old English Poetry: The Technical Problem and Poetical Art." Unpubl. PhD Diss. New Haven, CT: Yale University.

O'DONNELL, DANIEL PAUL (1996b). "A Northumbrian Version of *Cædmon's Hymn* (*Eordu* Recension) in Brussels Bibliothèque Royale MS 8245-57, ff. 62r2-v1: Identification, Edition, and Filiation." *Beda Venerabilis: Historian, Monk, and Northumbrian*. Eds. L. A. R. J. Houwen & A. A. MacDonald. Groningen: Egbert Forsten. 139-165.

O'DONNELL, DANIEL PAUL (1998). "The Electronic *Cædmon's Hymn*: Editorial Method." <http://people.uleth.ca/~daniel.odonnell/research/caedmon-job.html> [2005-06-21]. The original URL for this page (cited, for e.g. in Dahlström 2000) was <http://home.uleth.ca/~daniel.odonnell/caedmon-job.html> It is reprinted below.

O'DONNELL, DANIEL PAUL (Forthcoming). *Cædmon's Hymn: A Multimedia Study, Edition, and Witness Archive*. Cambridge: D. S. Brewer.

ORE, ESPEN S. (2004). "Monkey Business – or What is an Edition?" *Literary and Linguistic Computing* 19.1: 35-44.

ROBINSON, PETER, ED. (1996). *The Wife of Bath's Prologue on CD-ROM*. Cambridge: Cambridge University Press.

ROSS, CHARLES L. (1996). "The Electronic Text and the Death of the Critical Edition." *The Literary Text in the Digital Age*. Ed. R. Finneran. Ann Arbor: University of Michigan Press. 225-231.

SPERBERG-MCQUEEN, C. M. & LOU BURNARD (2004). *TEI P4: Guidelines for Electronic Text Encoding and Interchange*. TEI/ACH/ACL/ALLC. <http://www.tei-c.org/P4X/> [2005-06-21]

Appendix: O'Donnell (1998)

The following is a reprint of O'Donnell (1998). It has been reformatted for publication, but is otherwise unchanged from the original text with the exception of closing brackets that were missing from some of the code examples in the original and that have been added here. The Editorial Schema diagram has been redrawn without any deliberate substantive alteration. The original low resolution version can be found at <http://people.uleth.ca/~daniel.odonnell/research/caedmon-job.html>.

The Electronic *Cædmon's Hymn*
Editorial Method

Daniel Paul O'Donnell

The Electronic *Cædmon's Hymn* will be an archive based, virtual critical edition. This means users will:

- have access to colour facsimiles and diplomatic transcriptions of all surviving witnesses
- be able to generate mediated ('critical') texts on-the-fly by choosing the editorial approach which best suits their individual research or study needs.

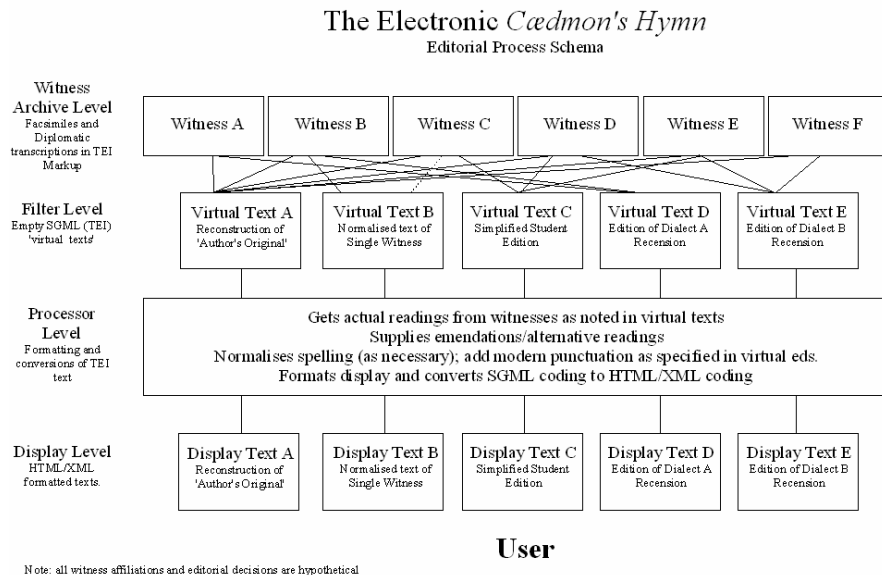The following is a rough schema describing how the edition will work:

Figure 1.

This schema reflects my developing view of the editing process. The terms (Witness Level, Processor Level, etc.) are defined further below.

In my view, the editor of a critical edition is understood as being functionally equivalent to a filter separating the final reader from the uninterpreted data contained in the raw witnesses. Depending on the nature of the instructions this processor is given, different types of manipulation will occur in producing the final critical edition. An editor interested in producing a student edition of the poem, for example, can be understood to be manipulating the data according to the instructions *choose the easiest (most sensible) readings and ignore those which raise advanced textual problems*; an editor interested in producing the 'original' text can be seen as a processor performing the instruction *choose readings from the earliest manuscript(s) when these are available and sensible; emend or normalise readings as required*; and an editor interested in producing an edition of a specific dialectal version of a text is working to the instruct-

tion *choose readings from manuscripts belong to dialect* x; *when these are not available, reconstruct or emend readings from other manuscripts, ensuring that they conform to the spelling rules of the dialect.*

If editors can be reduced to a set of programming instructions, then it ought to be possible, in an electronic edition, to automate the manipulations necessary to produce various kinds of critical texts. In the above schema, I have attempted to do so. Instead of producing a final interpretation of 'the text', I instead divide the editorial process into a series of discrete steps:

- the assembly of witnesses (the **witness archive level**);
- the construction of the filter through which the raw data from the witness archive is to viewed (the **virtual editions** or the **filter level**)
- the physical assembly and formatting of the chosen variant readings from the witnesses (the **processing level**);
- the final 'printed' text as it appears in the user's browser (the **display level**).

Because the critical edition is not seen as an actual text but rather as a simple view of the raw data, different textual approaches are understood as being complementary rather than competing. It is possible to have multiple 'views' coexisting within a single edition. Readers will be expected to choose the view most appropriate to the type of work they wish to do. For research requiring a reconstruction of the hypothetical 'author's original', a 'reconstruction filter' might be applied; a student can apply the 'student edition filter' and get a readable simplified text. And the oral-formulaicist can apply the 'single manuscript x filter' and get a formatted edition of the readings of a single manuscript.

Because different things are expected of the different levels, each layer has its own format and protocol. Because all layers are essential to the development of the text, all would be included on the CDRom containing the edition. Users could program their own filters at the filter level, or change the processing instructions to use other layouts or formats; they could also conduct statistical experiments and the like on the raw SGML texts in the witness archive or filter level as needed.

Witness Archive

The witness archive consists of facsimiles and diplomatic transcriptions of all relevant witnesses marked up in SGML (TEI) format. TEI is better for this initial stage of the mark-up because it is so verbose. Information completely unnecessary to formatting – linguistic, historical, metrical, etc. – can be included for use search programs and manipulation by other scholars.

The following is a sample from a marked-up transcription at the witness archive level:

```
<l id="ld.1" n="1">
        <w>Nu</w>
        <w>&wynn;e</w><space extent=0>
        <w>sceolan</w>
        <w>herian</w>
        <w><del type="underlined">herian</del></w>
        <caesura>
        <w>heo<lb><add hand="editorial" cert="90">f</add>on
        <space extent=1>rices</w>
        <w>&wynn;eard</w>.<space extent=0>
</l>
```

Virtual Editions

Virtual Editions are the filters that contain the editorial processing instructions. They are not so much texts in themselves as records of the intellectual processes by which a critical text interprets the underlying data contained in the witness archive. They are SGML (TEI) encoding documents which provide a map of which witness readings are to be used in which critical texts. For most readings in most types of editions, these instructions will consist of empty elements using the 'sameAs' and 'copyOf' attributes to indicate which witness is to provide a specific reading: e.g. `<w copyOf=CaW2></w>` where CaW2 is the identifier for the reading of a specific word from manuscript Ca. One of the advantages of this method is that eliminates one potential source of error (cutting and pasting from the diplomatic transcriptions into the critical editions); it also allows for the near instantaneous integration of new manuscript readings into the finished editions – changes in the witness transcriptions are automatically incorporated in the final texts via the filter.

In some cases, the elements will contain emendations or normalisation instructions: e.g. `<w sameAs=CaW2>þa<w>`.

The sample is from a virtual edition. It specifies that line 1 of this critical text is to be taken verbatim from manuscript ld (i.e. the text reproduced above):

```
<l id="Early.1" n="1" copyOf="ld.1"></l>
```

Processing Level and Display Texts

The 'Virtual Editions' are a record of the decisions made by an editor in producing his or her text rather than a record of the text itself. Because they consists for the most part of references to specific readings in other files, the virtual editions will be next-to-unreadable to the human eye.

Turning these instructions into readable, formatted text is the function of the next layer – in which the processing instructions implied by the virtual layer are applied and in which final formatting is applied. This processing is carried out using a transformation type processor – like Jade – in which the virtual text is filled in with actual readings from the witness archive, and these readings then formatted with punctuation and capitalisation etc. as required.

The final display text is HTML or XML. While this will involve a necessary loss of information – most TEI tags have nothing to do with formatting, few HTML tags have much to do with content – it is more than compensated for by the ability to include the bells and whistles which make a text useful to human readers: HTML browsers are as a rule better and more user friendly than SGML browsers. Users who need to do computer analysis of the texts can always use the TEI encoded witness transcriptions or virtual editions.

Here is my guess as to how HTML would display the same line in the final edition (a critical apparatus would normally also be attached at this layer containing variant readings from other manuscripts [built up from the manuscript archive using the 'copyOf' attribute rather than by cutting and pasting]; notes would discuss the various corrections etc. ignored in the reading text of this view):

```
<P>Nu we sceolan herian heofonrices weard</P>
```