Annotated Facsimile Editions: Defining macro-level structure for image-based electronic editions

Neal Audenaert and Richard Furuta
Department of Computer Science, Texas A&M University, College Station, TX 77843, USA

Abstract

Annotated Facsimile Edition (AFED) is a high-level model for representing macro-level structure in digital facsimiles. AFED models a facsimile as a set of images with multiple orderings or collations. The structure of these collations are encoded by 'annotations' that define a range of images in the collation and describe the properties of the content object identified by the annotations (for example, chapter, paragraph, page, poem). Separate annotation streams encode multiple analytical perspectives, for example, the physical structure of the edition (volumes, pages, and lines) and the poetic structure (poems, titles, epigraphs, and stanzas). Annotations within a single analytical perspective—but not those from different perspectives—follow a hierarchical structure. We discuss our initial results in implementing AFED and using it to deploy a reading interface for AJAX enabled rich-client Web applications. The primary contribution of our work is a general-purpose model for representing digital facsimiles that focuses on the major conceptual structures present among the contents of documents drawn from a wide range of sources. AFED provides a highly flexible model that can serve as a substrate for developing tools designed to support visual document editing during the exploratory stages of scholarly research.

Correspondence:
Neal Audenaert, Department of Computer Science, Texas A&M University, 301 Bright Building, College Station, TX 77843, USA.
E-mail:
neal@csdl.tamu.edu

1 Introduction

Facsimile images form a major component in many digital editing projects. Well-known projects, such as the *Blake Archive* (Eaves *et al.*, 2007) and the *Rossetti Archive* (McGann, 2007) use facsimile images as the primary entry point to accessing the visually rich texts in their collections. Even for projects focused on transcribed electronic editions, it is now standard practice to include high-resolution facsimiles.

Encoding standards and text-processing toolkits have been the focus of significant research. Tools, standards, and formal models for encoding information in image-based editions have only recently begun to receive attention. Most work in this area has centered on the digitization and presentation of visual materials (Viscomi, 2002) or detailed markup and encoding of information within a single image (Lecolinet *et al.*, 2002; Kiernan *et al.*, 2005; Dekhtyar *et al.*, 2006). Comparatively little work has been done on modeling the large-scale structure of facsimile editions. Typically, the reading interface that presents a facsimile determines its structure.

Separating the software used to model data from that used to build user interfaces has well-known

Literary and Linguistic Computing, Vol. 24, No. 2, 2009.
doi:10.1093/llc/fqp008 Advance Access Published on 27 April 2009

advantages for both engineering and digital humanities practices. To achieve this separation, it is necessary to develop a model of a facsimile edition that is independent of the interface used to present that edition.

In this article, we present a unified approach for representing linguistic, structural, and graphical content of a text as an Annotated Facsimile Edition (AFED). This model grows out of our experience with several digital facsimile edition projects over more than a decade, including the *Cervantes Project* (Furuta *et al.*, 2001), the *Digital Donne* (Monroy *et al.*, 2007a), and the *Nautical Archaeology Digital Library* (Monroy *et al.*, 2007b). Our work on these projects has emphasized the need for an intuitive conceptual model of a digital facsimile. This model can then serve as the basis for a core software module that can be used across projects without requiring extensive modification by software developers. Drawing on our prior work, we have distilled four primary goals for such a model:

Openness: scholars' focused research needs are highly specific, vary widely between disciplines, and change over time. The model must accommodate new information needs as they arise.
Nonhierarchical: facsimile editions contain some information that should be presented hierarchically, but they cannot be adequately represented as a single, properly nested hierarchy.
Restructuring: a facsimile is a representation of the physical form of a document, but the model should enable applications to restructure the original form to meet specific needs.
Alignment: comparison between varying representations of the same work is a fundamental task of humanities research. The model must support alignment between facsimiles of different copies of a work.

2 Annotated Facsimile Editions

AFED models the macro-level structure of digital facsimiles. AFED is based on the intuition that a digital facsimile can be represented as multiple image streams with annotations over each stream. These annotations describe the semantic structure and content of the document. Figure 1 shows a simplified diagram illustrating a two-volume edition of collected poems. Annotations, depicted as arcs between images, encode the structure of the document and the properties of the structural elements they represent. Separate annotation streams encode multiple analytical perspectives. For example, in Fig. 1, the annotations shown below the image stream describe the physical structure of the edition (volumes, pages, and lines) while the annotations shown above the image stream describe the poetic structure (poems, titles, epigraphs, and stanzas).

Several tools and standards have focused on providing support for detailed annotation of individual images and for linking those images with encoded texts. These tools include the Electronic Edition Production Toolkit (EEPT), the Image Markup Tool, and AXE (Kiernan *et al.*, 2005; Carlin *et al.*, 2005; Reside, 2007). The TEI P5 guidelines now provide modules for working with facsimile

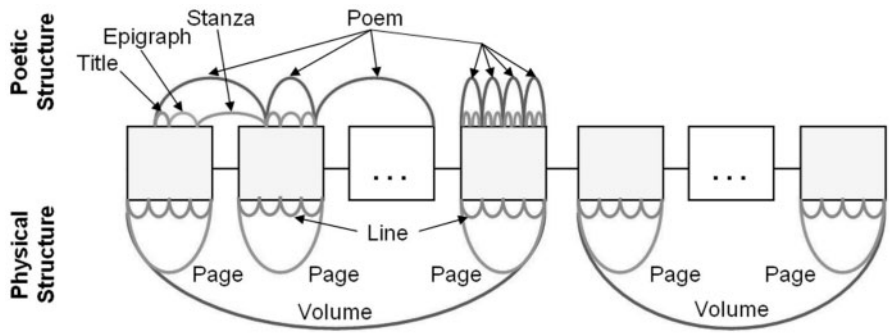

Fig. 1 A simplified diagram showing an edition of collected poems (in two volumes) represented as an AFED

images and supporting basic markup of those images. When editors need nonrectangular markup, the guidelines allow the inclusion of SVG-based annotations in a mixed namespace document.

In contrast, instead of focusing on individual images or on linking images to specific portions of encoded texts, we have developed AFED to provide an image-centric approach to modeling the structure of digital facsimiles. We emphasize the macro-level document structures needed to facilitate navigating and reading documents rather than detailed scholarly markup. For example, readers may want to navigate a novel by chapters, turn to a specific volume of an encyclopedia, or compare different versions of a poem that was published in several anthologies. Structures at this level typically span multiple images in a facsimile. The fact that this is not always the case—multiple short poems that appear on a single page are one counter-example—requires that AFED also permit the marking of multiple objects with a single image.

While AFED permits annotation of content within individual images, the model does not specify the semantics of these annotations beyond establishing a relative order of annotations within a given annotation stream. In order to achieve a general solution, we have designed AFED to be agnostic with respect to the contents of facsimile images. An example of a relatively simple macro-level markup task—identifying pages in different facsimiles—illustrates some of the difficulties in developing a general-purpose solution.

Most facsimiles rely on images either of individual pages or of book openings consisting of two pages in a single image. In the former case, marking a page simply requires identifying individual images; in the latter case, it requires the identification of a region within an image. In other cases, however, individual pages may span multiple images. An editor may want to mark individual leaves as pages, each of which has verso and recto sides from two different images. Pages from badly damaged manuscripts may consist of multiple, independently photographed fragments. Facsimiles of some documents, scrolls for example, may not have pages at all.

In order to account for the nearly infinite variety of documents and editorial needs, AFED does not model the details of a document's content. Despite this, the ability to support sub-image level annotations, along with the flexible mechanisms provided for describing annotation content (discussed in detail below), allow AFED to seamlessly represent documents structures that range from extremely coarse-grained (such as volumes) to extremely fine-grained (such as individual words and characters).

2.1 Image stream

The image stream intuitively corresponds to the sequential ordering of page images in a traditional printed book. These images, however, need not represent actual 'pages'. An image might show a variety of artifacts including an opening of a book, a fragment of a scroll, or an unbound leaf of manuscript notes. While it is natural to treat facsimile images sequentially, any particular linear sequence represents an editorial decision—a decision that may not be explicitly represented by the structure of the physical document. For example, an editor may choose to arrange an edition of letters according to the date written, recipient, or thematic content.

In addition to ambiguity of the original document, editors may also want to represent known or hypothesized historical states of a document. For example, one shipbuilding treatise that we are working with as part of our nautical archeology digital library has been rebound several times. Originally, the current document was written as three separate books. As the needs of ship-builders changed, these books were reassembled with different page orders. In its current form, many of the pages have three numbers, one from each of the different bindings. Other pages have lost one or more of their page numbers. Consequently, we need the ability to represent at least three, partially known orderings for the pages in this book.

Rather than assuming that the images of a digital facsimile have a single natural ordering, we have modeled facsimiles as an unordered collection of images. Editors impose sequential orderings (collations) on these images. These collations, or

Table 1 Information represented by an annotation

Annotation management	
Type	The name of this type of annotation, e.g. page, volume, chapter, poem, and stanza.
Collation	The specific image collation that the start and stop indices refer to.
Start index	The index into the collation where this annotation starts.
Stop index	The index into the collation where this annotation ends.
Sequence	A number for resolving the sequence of multiple annotations on the same page.
Content	
Canonical name	A canonical name that uniquely identifies this content relative to a domain-specific classification scheme.
Display name	The name to be displayed when referring to an instance of this annotation.
Properties	A set of key/value pairs providing domain-specific information about the annotation.
Transcriptions	A set of transcriptions of the content that this annotation specifies.
Relationships	
Parent	A reference to the parent of this annotation.
Children	A list of references to the children of this annotation.
CrossRefs	Lists of related annotations. Each list is identified by a key that indicates how the annotations in the list are related to this annotation.

image streams, provide the basic structural substrate to which annotations are attached. Each facsimile must have at least one collation, and each collation may contain an arbitrary sub-set of the images in a facsimile.

2.2 Annotations

Annotations, depicted intuitively in Fig. 1 as arcs over the image stream, are the primary means for representing structural and linguistic content in AFED. An annotation identifies a range of images in a collation and specifies properties about those images. Table 1 lists the information specified by each annotation. Properties in italics are optional. As shown in this table, annotations support three main categories of information: annotation management, content, and relationships.

2.2.1 *Annotation management*

Information in the annotation management category specifies the type of an annotation and identifies the images referenced by the annotation. Each annotation is an instance of a particular annotation type. This is analogous to elements in XML documents, such as the 'p' element or the 'quote' element. The type defines the analytical perspective of the annotation (for example, physical or narrative structure) and provides a description of the content that is by an annotation similar to the way an XML schema or DTD defines the attributes and valid child nodes of an XML element.

In addition to identifying the type of annotation, information in this category also specifies the starting and ending images relative to a particular collation. Often, multiple annotations such as lines or sentences may appear on a single page. To ensure that annotations are properly ordered, each annotation may optionally include a sequence number. AFED is agnostic as to the precise semantics of sequence number. It may be defined relative to a single page (for example, line numbers for the lines of text on a page), to a larger section (for example, line numbers for lines of in an epic poem), or an entire document (for example, page numbers in a facsimile that shows two pages per image). The only requirement is that the sequence number must define the relative order of multiple annotations of the same type on the page.

2.2.2 Content

The content category describes the object referenced by an annotation. This includes the name of the referenced object, various user-defined properties object, and transcriptions of the object's textual content. Annotations support two naming conventions. One is the name that should be shown in user interfaces when an annotation is displayed. Depending on the interface design and task, this

name might appear as a tool tip when a user hovers over an annotation, as a page title when a user is viewing the details of an annotation, or in any other way needed by an interface designer. An annotation's name may be generic and generated automatically (for example, an editing interface may choose to assign the name 'paragraph' to all paragraph annotations) or it may be specific to the object being annotated (for example, the title of a book chapter).

In order to facilitate comparisons between documents, annotations also allow an editor to specify a canonical name according to a domain-specific naming convention. The scholarly communities that study many documents have developed canonical numbering schemes to support comparison between different copies of the same work—Bible verses are one example. The canonical name of an annotation can be used to automatically find and retrieve related annotations from different documents. Since canonical names usually do not match the name given to the referenced item by the artifact itself and are rarely appropriate for display to a general audience, AFED stores these names separately from those intended for display in general-purpose user interfaces.

The main tool for describing the content of an annotation is a set of key/value properties used to encode descriptive metadata. Each annotation type specifies a set of fields or keys to which values may be assigned. For each field, users (including both human editors and automated tools) may specify one or more value. In order to establish the authority of metadata supplied using these properties and to facilitate collaborative projects, values may be associated with the user responsible for assigning the specified value and the date the information was entered. As mentioned previously, annotation types define known fields rather than enumerating all possible fields. Consequently, user interfaces built around AFED may allow users to add new fields on the fly. This approach frees users from the requirement that they know in advance all of the types of data that will be associated with any particular annotation. New fields types can be added quickly to express formative or uncertain hypotheses about the data. Over time, authorized users may update the fields that are defined by the annotation type to include unofficial fields that have proven to be particularly useful.

While the primary purpose for annotation properties is to enable human readable metadata, a secondary use is the storage of information for use by applications. For example, a program could use a property field to store the coordinates of the image region that is represented by an annotation such as the bounding box around a line of text. AFED specifies only that field values be character strings, the specific format of those strings is unspecified. Applications may store values as XML-encoded data, comma separated values, or other character-based representations.

AFED also provides basic support for transcribing content. In general, we expect that the best practice for integrating detailed transcriptions will be to link to or from TEI-encoded documents. For convenience, however, AFED provides basic support for transcribing annotated content. Since different users or editorial objectives will require different types of transcriptions, multiple transcriptions can be associated with each annotation. For example, one user may provide paleographic and normalized transcriptions of the diary of a seventeenth-century conquistador, while another adds translations into English and French.

2.2.3 *Relationships*

Finally, information in the relationships category describes the relationships between different annotations. Two types of relational structures are provided: cross references and hierarchies. Annotations may reference other annotations within a document. For example, one edition of *Don Quixote*, the Bowle edition, is published in three volumes, with the third volume containing commentary about the text in the first two volumes. In building a reader's interface for this book, we wanted to be able to display either the original structure of the book or the page from the main text alongside the corresponding page or pages from the commentary. Cross-references provide a mechanism for accomplishing this.

The second type of relationship supported by AFED is hierarchical relationships. Hierarchies suffer from well-known limitations (Renear *et al.*, 1996) and visual rather than logical representations information pose additional difficulties for hierarchical structures. Despite the difficulty of modeling a document within a single tree structure, many features of a text are best represented hierarchically—chapters that contain paragraphs that contain sentences that contain clauses, for example. To support these relationships, AFED allows annotations hierarchical relationships within a single perspective by allowing each annotation to point to a parent annotation and multiple child annotations. These hierarchies must be represented by the physical structure of the annotations—that is, a child annotation's start and end indices must fall within its parent's start and end indices and must not overlap with its siblings. Sequence numbers are used to resolve the potentially ambiguous ordering between annotations that occur on the same page. One key advantage of AFED is that it provides explicit support for hierarchical structures where they are helpful, without requiring their use where they are not.

3 Implementation

We have designed AFED to be an enabling technology for representing facsimiles in an abstract way so that they can be used by multiple applications for different purposes. These applications include fully automatic tools for importing and specifying the basic structure of digital facsimiles, user interfaces for scholarly editing, and reading interfaces both for the general public as well as for scholars. To achieve this, we have developed a prototype implementation of the model and have used this implementation as the backbone for some basic user interfaces. We are currently working to develop some more advanced interfaces and to integrate our AFED-based models into CritSpace, a Web-based environment for conducting exploratory research in the cultural heritage digital libraries.

Initially, we developed a proof-of-concept implementation of AFED. We used this application in the *Cervantes Project* as the basis for a series of digital facsimiles of early editions of the *Don Quixote* owned by Biblioteca Nacional de España. As a part of this project, we developed automated tools to import data from the file system, extracting information from the directory structure and file names about page numbers, illustrations, chapters, volumes, editions, and different copies of the editions. We presented the resulting facsimiles in Web-based reader's interface that allowed users to navigate by pages, chapters, and volumes (for multivolume editions). Some of the digital images available to us showed a book opening while others contained only a single image. In order to provide a consistent interface, for those facsimiles with one page per image we dynamically reconstructed book openings using the properties associated with the page annotation. Since the initial development of this tool, other members of the project have used it to build more facsimile editions.

This implementation omitted many of the details of the full model, but demonstrated the usefulness of the core concept behind AFED: the image stream plus annotations approach to modeling digital facsimiles and implementing user interfaces. We have recently developed a full implementation of AFED as described in this article and have begun to develop user interfaces for it. Currently, we have provided a simple browsing interface, shown in Fig. 2. The pages of the facsimile are displayed in a thumbnail view. When the user clicks on an image, a dialog appears containing a high-resolution image the user can zoom in on in order to see more detail. In this dialog, navigational controls are provided that allow the user to display the next and previous pages. These controls may also be used to select other annotations types to navigate by, for example, chapters or poems depending on the annotations present in the facsimile.

In the near future, we plan to develop an editor's interface that will allow scholars to define new annotation types, create new facsimile collations and manipulate existing ones, add annotations to collations, and edit annotation contents. We are also working to design and implement new reading interfaces.

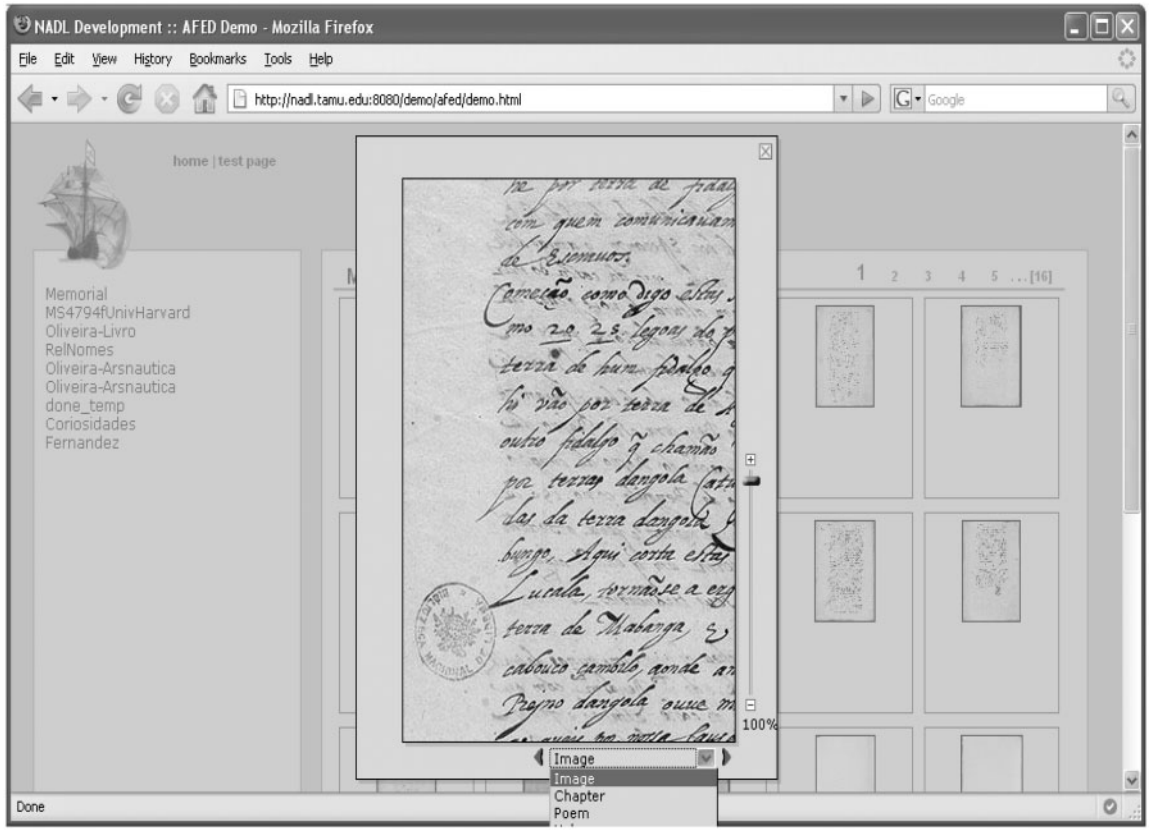

Fig. 2 A simple reader's interface using an AFED-based digital facsimile. The user may navigate to different annotated sections by selecting the type of annotation and clicking the left or right arrows

4 Discussion and Conclusions

AFED provides a framework for describing facsimile editions in a general, intuitive manner. This work forms one component of our long-term research efforts to understand and support scholarly interpretation and creative analysis of visually complex documents. Documents express information as a combination of written words, graphical elements, and the arrangement of these content objects in a particular media. The spatial arrangement and visual attributes (for example, color, font, size, and orientation) combine to form the visual grammar of a document or class of documents. This visual grammar provides components of the document's meaning in conjunction with the actual words or images of the document.

The relationship between the visual grammar of a document and its words and images varies as a continuum in (at least) two dimensions. On one axis, the visual grammar may be expressed more or less formally. For example, a journal article usually requires that documents follow specific layout guidelines, whereas providing a complete description, *let alone* specification, of the works of William Blake would be difficult at best.

On the second axis, the visual grammar may be more or less integral to the meaning of the document. Changing the format of a journal article from that typically used by *Literary and Linguistic Computing* to the format preferred by Springer's *Lecture Notes in Computer Science* is unlikely to cause us to think that these two documents are somehow fundamentally different. On the other

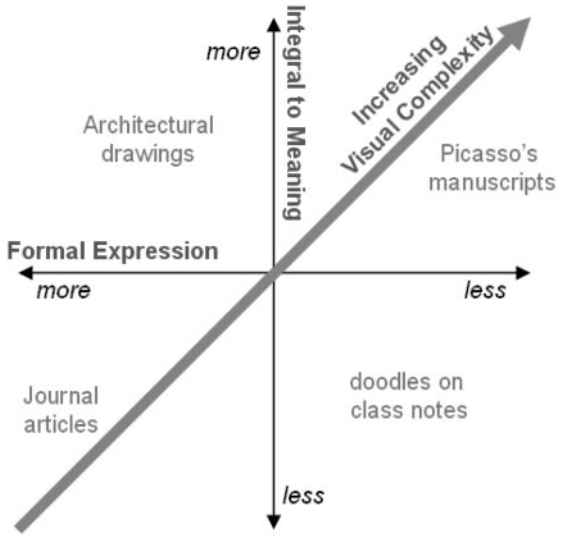

Fig. 3 Graph showing how the relationship between a document and its visual grammar varies along two dimensions

hand, the visual grammar of an architectural design document is central to its meaning.

As illustrated in Fig. 3, visual complexity increases as the visual grammar becomes less formal and more tightly integrated with the meaning of the text. It is important to note, though, that visual complexity does not depend only on the visual properties of the document. All documents convey information both in terms of the logical, textual content as well as the way in which that textual content is arranged on a page alongside (and structured by) other graphical features of the document. The importance of the actual words and images relative to the other visual content of a document depends on how that document is used. For example, the visual grammar of a printed book may not be particularly relevant when read as a medieval philosophical treatise. A scholar interested in the history of book making, however, may care more about the font size, illustrated letters, and layout of footnotes and scholia—the visual elements will take priority over the narrative content of the text. Consequently, the visual complexity of a document depends on a combination of the goals and objectives of the various people using the document and the properties of the document itself.

Supporting the early stage, exploratory research involving visually complex documents requires tools that provide lightweight access to the original documents and facilitates the rapid expression, revision, and refinement of tentatively posed hypotheses. Traditional transcription (and all transcription-based textual encoding), on the other hand, requires that an editor first identify the set of tokens and carefully identify and record a single sequence of tokens that will represent the original document (Huitfeldt and Sperberg-McQueen, 2008). This process of selection and flattening not only adds editorial value, but also limits access to the full scope potential readings of the original work.

We have designed AFED to help support a visual editing paradigm that provides computational support for editing cultural heritage documents while requiring minimal formalization early in the research process. Kiernan (2007) is careful to distinguish between image-based scholarly editions and 'plain old facsimiles'. In designing and implementing AFED, we have focused our attention on understanding and modeling the structure of plain old facsimiles. Our general-purpose model represents digital facsimiles at a high-level, focusing on the major conceptual structures that are present among the contents in these documents. This approach allows us to construct a highly flexible model for use as a substrate in tools that support exploratory editorial processes.

References

Carlin, C., Haswell, E., and Holms, M. (2005). *Image Markup Tool*. University of Victoria Humanities Computer and Media Center. http://mustard.tapor.uvic.ca/~mholmes/image_markup/ (accessed 23 September 2008).

Dekhtyar, A., Iacob, I. E., Jaromczyk, J. W., Kiernan, K., Moore, N., and Porter, D. C. (2006). Support for XML markup of image-based electronic editions. *International Journal on Digital Libraries*, **6**(1): 55–69.

Eaves, M., Essick, R. N., and Viscomi, J. (eds) (2007). *The William Blake Archive*. http://www.blakearchive.org/(accessed 24 November 2007).

Ferraiolo, J., Jun, F., and Jackson, D. (2003). Scalable Vector Graphics (SVG) 1.1 Specification. World Wide Web Consortium. http://www.w3.org/TR/SVG/ (accessed 23 September 2008).

Furuta, R., Kalasapur, S. S., Kochumman, R., Urbina, E., and Vivancos-Pérez, R. (2001). *The Cervantes Project: Steps to a Customizable and Interlinked On-Line Electronic Variorum Edition Supporting Scholarship, Proceedings of ECDL 2001, LNCS, 2163*. Heidelberg: Springer-Verlag, pp. 71–82.

Huitfeldt, C. and Sperberg-McQueen, C. M. (2008). What is transcription. *Linguistic and Literary Computing*, **23**(3): 395–10.

Kiernan, K., Jaromczyk, J. W., Dekhtyar, A., Porter, D. C., Hawley, K., Bodapati, S., and Iacob, I. E. (2005). The ARCHway Project: architecture for research in computing for humanities through research, teaching, and learning. *Literary and Linguistic Computing*, **20**(Suppl 1): 69–88.

Kiernan, K. (2007). Digital facsimiles in editing. In Burnard, L., O'Keeffe, K., and Unsworth, J. (eds), *Electronic Textual Editing*. New York: Modern Language Association. http://www.tei-c.org/About/Archive_new/ETE/Preview/kiernan.xml (accessed 23 September 2008).

Lecolinet, E., Robert, L., and Role, F. (2002). Text-image coupling for editing literary sources. *Computers and the Humanities*, **36**(1): 49–73.

McGann, J. (2007). *The Complete Writings and Pictures of Dante Gabriel Rossetti*. Institute for Advanced Technology in the Humanities, University of Virginia. http://www.rossettiarchive.org/ (accessed 24 November 2007).

Monroy, C., Furuta, R., and Stringer, G. (2007a). *Digital Donne: Workflow, Editing Tools and the Reader's Interface of a Collection of 17th-century English Poetry, Proceedings of Joint Conference on Digital Libraries JCDL* (Vancouver, BC, 2007), New York, NY, ACM Press, pp. 411–2.

Monroy, C., Furuta, R., and Castro, F. (2007b). *Texts, Illustrations, and Physical Objects: The Case of Ancient Shipbuilding Treatises, Proceedings of ECDL 2007, LNCS, 4675*. Heidelberg: Springer-Verlag, pp. 198–209.

Renear, A., Mylonas, E., and Durand, D. (1996). Refining our notion of what text really is: the problem of overlapping hierarchies. In Ide, N. and Hockey, S. (eds), *Research in Humanities Computing*. Oxford: Oxford University Press.

Reside, D. (2007). *The AXE Tool Suite: Tagging Across Time and Space, Proceedings of Digital Humanities*. Urbana-Champaign, Graduate School of Library and Information Science, University of Illinois, pp. 178–9.

Viscomi, J. (2002). Digital facsimiles: reading the William Blake Archive. In Kirschenbaum, M. (ed.), *Image-based Humanities Computing*. Special issue of *Computers and the Humanities*, **36**(1): 27–48.