Facing the Deep:
The Orlando Project Delivery System 1.0

Susan Brown (University of Guelph), Patricia Clements, Isobel Grundy, Jeffrey Antoniuk, Sharon Farnel, Jane Haslett (University of Alberta), and Kathryn Carter (Wilfrid Laurier University, Brantford)¹
sbrown@uoguelph.ca

Abstract
This paper reflects on the interplay between DTD design and that of the delivery system of the Orlando Project, an intensively encoded body of born digital materials in women’s literary history. The project developed and refined an extensive content-oriented SGML tagset before any material had been written, and without specific delivery plans. First describing the project’s XML delivery system, the paper reflects on such issues as the relationship of DTD structure to delivery, the results of our user testing to date, and the challenges posed by navigation and hyperlinking in the production of a new model of digital scholarship.

KEYWORDS: Text encoding, Knowledge representation, XML, Databases, Delivery systems, User testing, Hyperlinking, Literary history.

This paper reflects on the interplay between DTD design and the shaping of the Orlando Project’s delivery system. The project is completing an intensively encoded body of digitally original materials on women’s writing in the British Isles from the beginnings to beyond the 1960s. Text markup generally proceeds either from the desire to represent a pre-existing text according to a rigorous representation standard such as the Text Encoding Initiative, or (and often also) from a relatively clear sense of what formatting the markup will be used to produce or what functionality (such as hyperlinking or searchability) it will support. Orlando initially developed and refined an extensive content-oriented tagset in Standard Generalized Markup Language (or SGML) to reflect its researchers’ priorities in literary history. Document Type Definition (or DTD) design began before any material had been written, and without specific plans for how the encoded text would be delivered to end users. The project’s DTDs therefore represented a set of abstract priorities for producing a literary history. Our XML delivery system was developed years later, intensively from 2002.
Prompted by deep questions about the shape and function of electronic scholarship, the project thus undertook a challenging experiment in the transformation of literary history to digital form, working from principle to practice in the implementation and delivery of SGML markup.

1. Background and DTD Design

First a necessary overview: Orlando was conceived of as an experiment in interpretive markup more than a decade ago, when graphical interface browsers were in their infancy and it was an open question whether SGML and the TEI would be established as the standards they have become for academic scholarship in the humanities. The core literary team of Susan Brown, Patricia Clements, and Isobel Grundy were new to humanities computing and, taking advice from co-investigator Susan Hockey about the future direction of such work, jumped headlong into DTD development with quite different aims from those of existing text encoding projects. These aims were to try to meet some of the challenges that had been leveled at conventional literary history, by exploiting electronic text’s capaciousness and flexibility, its potential to be multi-linear and multi-voiced, and its pressure towards self-consciousness. From SGML we wanted not only platform independence and a flexible yet familiar visual organization of the text, but also to produce a transparently self-reflexive structure for conducting literary history in new ways.

So in the design of our DTDs the project members faced a number of challenges. The TEI and existing projects helped us decide on basic structural hierarchies for our documents, but took us only a certain distance towards content markup. The literary team had no experience in humanities computing, and we had no documents for document analysis: this was a deliberate decision since we wished to make technological tools a shaping factor in the research. Beyond the conviction of the potential offered by going electronic, we had no precise delivery plans. Instead, we defined factors important to our understanding of the history women’s writing in the British Isles, tried to decide if they made sense to tag, and if so define how to tag them. Then we related these concepts to each other hierarchically and reconciled the TEI-based structural hierarchy to the hierarchies of the conceptual tags, although it might be more accurate to say that the DTDs interleave, rather than reconciling, these two quite different kinds of hierarchies. And as Figures 3 and 4 (below) demonstrate, the hierarchies function quite differently in the two main DTDs.
The process of DTD development, testing, and revision took intensive collaborative work from 1996 to 1998, although selective tweaking has been ongoing. Throughout this process, although we would occasionally blue sky it, we tended to postpone questions of delivery, since the technical ground was shifting under our feet as we worked. It was unclear, for instance, whether or not XML, which enabled the delivery of SGML documents the World Wide Web, would take off, and if so whether it would offer the best mode of delivery.

The project devised 5 DTDs. Simple ones for chronology items and bibliographical information were conceived with the clearest sense of delivery: the granularity of these materials made them obvious candidates for manipulation by users to produce on-the-fly bibliographies and chronologies. There is also a quite simple DTD for general topic entries. The heart of the project lies in the paired DTDS for encoding discussions of writers’ lives and those of their texts and literary careers. These we conceived as accounts in continuous prose, along the lines of encyclopedia entries. We envisioned that these documents would contribute to the delivery system in a somewhat dynamic fashion, particularly insofar as they would embed events that would become part of an overall chronology, and we assumed some degree of hyperlinking among the materials. Altogether, there are currently 201 different tags and 102 different attributes in the project’s DTDS. Those unique to Orlando are concentrated in the Life and Writing tagsets. To give some sense of the intensiveness of the tagging, in a corpus of about 145 MB of raw XML, excluding internal-use tags such as those for research notes, there are above 1,400,000 tags.

This topic map (Fig. 1) from the delivery system shows the top-level content tags of the DTD for encoding discussions of writers’ lives. It uses Div1s to organize documents according to broadly conceived topics such as Birth, Cultural Formation, Education, Occupation, and so on. These tags often have unique subtags and/or attributes, such as geographical heritage, school, religious denomination, or job. Documentation of each of the tags appears in the sidebar on this screen, with brief definitions, information on related (nested) tags and attributes, and examples of how the tag has been used in the system.

Although the number of unique tags in each DTD is roughly equal (69 for Life and 64 for writing), the Writing DTD (Fig. 2) appears more complex, organized as it is into the three major content tags of textual features, reception, and production, each notionally having numerous subtags. Despite its apparent complexity, however, the writing DTD allows
considerably greater freedom in writing than the biography DTD.

The DTD makes heavy use of inclusions to allow all the subtags from any
of these three organizing categories to be employed within another one; for instance, a discussion of the responses to a text may wish to mention a reviewer’s response to one of its characters. Although the entire discussion may occur within a <RECEPTION> tag, a document author can still use the <CHARACTERNAME> tag, despite the fact that it is conceptually, if not strictly hierarchically, nested within <TEXTUALFEATURES>.

This creation of a quite flat DTD runs somewhat counter to the influential view among proponents of descriptive markup that the components of markup, which “have an intrinsic direct connection with the intellectual content of the text,” are best understood as an ordered hierarchy of content objects (Renear 224). It may be related to the type of “objects” represented by much of the Orlando Project’s content markup, which tend to be different and more interpretive textual phenomena than, say, chapters, examples, or even names. At any rate, the decision was driven by our sense that the imbrication of issues in discussions of writings and literary careers was, for the purposes of a literary history at least, greater than in biographical ones, and that to permit readable prose we had to allow for what were in essence overlapping hierarchies.

Interestingly, the contrasting structures of these two DTDs have resulted in quite different organizing principles within the documents written with them. All documents tend to treat their material roughly chronologically. Life documents are structured by the sixteen <DIV1>-level content tags shown in Fig.1, although the order and frequency of occurrence vary from one document to another. Writing documents, on the other hand, are primarily organized neither by the three major tags of production, reception, and textual features, which recur numerous times in most documents, nor by subtags, but according to the categories of writings, or issues, or particular texts discussed in the document. These last are linked through the use of a <TEXTSCOPE> tag to the Div2s in which they occur and to their respective bibliographical entries.

The difference between the two document structures is easily seen in an outline (Fig. 3 and Fig. 4) of tag use in a fairly small document pair, that for Augusta Ada Byron. These outlines show the tagging structures down to the tags immediately below the <DIV1> level; these tags indicate the conceptual organization of the document. The first <DIV1>s in both cases are expanded as far as the <DIV2> level, below which substantive discussion generally resides.
Fig. 3: Outline of high-level tags in the Byron Life document (29K)

<BIOGRAPHY>
<ORLANDOHEADER>
<DIV0>
<DIV1>
<HEADING>
<BIRTH>
<DIV2>
<CULTURAL FORMATION>
<FAMILY>
<EDUCATION>
<HEALTH>
<LOCATION>
<INTIMATERELATIONSHIPS>
<LEISUREANDSOCIETY>
<FAMILY>
<FRIENDSASSOCIATES>
<OCUPATION>
<EDUCATION>
<LOCATION>
<FAMILY>
<FAMILY>
<HEALTH>
<HEALTH>
<INTIMATERELATIONSHIPS>
<FAMILY>
<WEALTH>
<HEALTH>
<DEATH>

Note that the tags immediately above nest each within a <DIV1>, with <DIV2>s below them.
Fig. 4: Outline of high-level tags in the Byron Writing document (19K)

\[\text{<WRITING>}
\text{<ORLANDOHEADER>}
\text{<DIV0>}
\text{<HEADING>}
\text{<AUTHORSUMMARY>}
\text{<DIV1>}
\text{<HEADING><TITLE TITLETYPE="MONOGRAPHIC"> A Sketch of the Analytical Engine</TITLE></HEADING>}
\text{<DIV2>}
\text{<TEXTSCOPE>}
\text{<PRODUCTION>}
\text{<TEXTUAL-FEATURES>}
\text{<RECEPTION>}
\text{<HEADING>Other Writings</HEADING>}
\text{<HEADING>Responses to Ada</HEADING>}
\text{<DIV2>}
\text{<TEXTSCOPE>}
\text{<PRODUCTION>}
\text{<TEXTUAL-FEATURES>}
\text{<RECEPTION>}
\text{<HEADING>Other Writings</HEADING>}
\text{<HEADING>Responses to Ada</HEADING>}
\text{</DIV1>}
\text{</DIV0>}
\text{</WRITING>}

Note that the headings here each nest within a \texttt{<DIV1>}.

Each \texttt{<DIV1>} contains one or more \texttt{<DIV2>}s, which may contain any or all of the three tags \texttt{<PRODUCTION>}, \texttt{<TEXTUAL-FEATURES>}, and \texttt{<RECEPTION>}. Thus, the priorities of this literary history project which seeks to map the production, dissemination, consumption, and interrelation of texts by women in relation to the material and discursive conditions of their lives, are represented in two DTDs with quite diverse document structures. In addition to the Life and Writing documents, there is also the more granular material of chronology, bibliography, and topic entries.

2. Orlando Delivery System

In 2002 we turned seriously to devising a delivery system for the body of encoded texts we were creating with these DTDs. Some early ideas, such as wedding place tags with a geographical information system to produce maps, remain unrealized, and we would dearly like to expand on the functionality we have been able to achieve. But we are currently focusing on polishing this first delivery system. And now, in our second round
of testing, we have occasion to reflect on the interplay between our initial abstract aims, the DTDs we designed, and delivery.

One major difference is that the system is more dynamic than we foresaw. The Orlando Project had its genesis in the work done by Virginia Blain, Patricia Clements, and Isobel Grundy on the Feminist Companion to Literature in English, an alphabetical reference guide composed of brief biocritical entries on women writers. The advantages of a more expansive and electronically searchable form of publication were immediately evident, but the Life and Writing essays that were conceived of as forming the core of the Orlando Project’s electronic literary history were conceived initially as relatively static, which is to say we assumed that most readers of the Project would read them sequentially, as they would a print entry. This expectation is reflected in the delivery system to the extent that document pairs tagged with the Life and Writing DTDs are grouped with related material to make up what we refer to as “entries”, available as tabbed screens.

However, the first view (Fig. 5) is not directly of the Life and Writing documents themselves. The Overview screen takes the form of a table composed of headings and key portions of documents in order to offer to users
an overview of a writer’s significance and the material available on her. Thus, the paragraph in prose at the top of this screen renders the contents of the <AUTHORSUMMARY> tag that heads most writing documents (but which, as the system has developed, has come not to be rendered at all in the delivery system view of the writing documents themselves), a case in point of the instability of the concept of a “document” in an electronic environment. The brief chronology of milestones on the left is a subset of the total events available on the writer, generated by the system according to tags reflecting structural significance in the life and career.

This page also functions as a navigational tool. The Life High-lights and Writing Highlights provide the contents of the <HEADING>s from the documents themselves, giving users a sense of their contents and the ability to hyperlink to that portion of the document. The tabs provide views of the two core documents by themselves or in parallel frames, a chronology comprised of all the events contained within the core documents plus any others within the system (freestanding or from other writers’ documents) that refer to this writer, a set of links to all other references to her (or him?) within the textbase, and a list of her written works entered in the bibliography.

In sum, what comprises an ‘entry’ within the Orlando Project is composed of more than just the Life/Writing document pairs, and the view of an ‘entry’ is decidedly different from anything project members envisioned at the outset. It has emerged from the set of constraints and enablements generated from a complex set of factors—including the conjunction of our intellectual and disciplinary backgrounds and experiences, our interest in producing a new kind of literary history, and the funding which enabled us to undertake a large-scale computing project—and largely embedded in the DTDs. Within the delivery system generally, materials are sectioned and dispersed to make the system responsive to users. Chunks become part of other chronologies or bibliographies. Thus Ada Byron’s birth event is visible on her overview screen (as well as being embedded in the view of its source—her Life document—and in her own timeline), but if a user follows the hyperlink on the name of her father and then, on his Links screen, looks at his timeline (Fig. 6), her birth also appears, differently contextualized.

The more dynamic treatment of our documents has also led us to devise as a major feature of the home page (Fig. 7) a browsing feature that offers a regularly changing random set of hyperlinks offering serendipitous paths into project materials.
Fig. 6: George Gordon Lord Byron Links screen, portion of timeline view

Fig. 7: Orlando Delivery System 1.0 home page
Fig. 8: Bessie Rayner Parkes Links screen, with politics, and specific response links visible

These currently lead either to a screen of chronological items or to the entry for an individual writer. The fact that this seems to be the single most controversial feature of the delivery system brings home the challenge of handling hyperlinks, which register very differently with different people, even within a fairly focused user community. Everywhere else within the system, automated hyperlinks are created and handled consistently for all multiple instances of names, places, organizations, and titles. The content tagging allows us to organize multiple links according to the context in which they occur. Thus any hyperlink in the system leads to a Links screen, in which hyperlinks are organized according to their occurrence within particular Life or Writing content tags (Fig. 8).

Clicking on a particular link then takes the user to the relevant portion of that writer’s document. A user can thus move through the project materials in an informed and selective way. The timeline, bibliography, and search results tabs on the Links screen allow users to pursue hyperlinks in a range of other ways.

Users can of course search on tag occurrences or tag contents. We had envisioned this in the DTD design stage as a primary benefit of the
tagging. For instance, one can search for intertextual references to Jane Austen.

Results (Fig. 9) are entire Div2s from entries with the hits highlighted. These can be read as a set and in dialogue with their immediate tagging context, which is provided on the right. This right-column, showing where in the immediate tagging structure the hit occurs, provides contextual information that may suggest to users how to further refine a search by specifying an attribute or attribute value, or by using a tag-within-tag query. It is always possible to consult the tagged text, by clicking on the “show markup” link. From these search results, if a user wants more context, clicking on “profile” provides the contents of the <AUTHORSUMMARY> tag; clicking on the author’s name links to the source entry at the point of the excerpt.

Fig: 9: Partial Search Results for <NAME> containing "Austen" within <INTERTEXTUALITY> tag

The People entry point also relies on the markup to enable searches for people who wrote a particular genre, held a certain kind of job, or are associated with a place. Results come as a list of names, for access directly to entries, or as a set of excerpts (Fig. 10) that can be read independently.
This entry point thus makes some of the functionality of the tag searches available with a less daunting interface.

Fig. 10: Partial set of excerpts from documents accessed by selecting “scientific writing” from genre picklist in People entry point

3. User Feedback

The challenge to the average student or scholar of English posed by the project’s unusual form of knowledge representation came home with the results of our first round of user testing on a functional system. We have, despite the alienating effect that such language has on those in our home discipline of English studies, increasingly come to refer to our prospective readers as “users”. This is for us a necessary reminder of the paradigm shift in which Orlando is participating: as our testing of the delivery system confirmed, many of the behaviours associated with the use of print reference works do not carry over to this medium. Although the concepts represented by our tags would be recognizable to other literary historical scholars and indeed to most undergraduate students in English, their implementation in this new medium and the unfamiliar way in which the system brings them into relationship with each other have profound impacts.
User testing made it clear that we needed to declutter our search screens as much as possible, and to clarify search logic. Our testers suggested that we needed to resist trying to elucidate all the nuances of the system on the search panels. Nevertheless, they also expressed a strong desire for more help and documentation, hopefully indicating that they will seek details and nuance after they have achieved basic familiarity.

As a consequence, while streamlining the screens, we have instituted help at a range of levels. We are currently developing a more integrated help system architecture to allow users to move between different help features, search the documentation, consider some instructive case studies, and consult FAQs about how to use the system. Together, we hope these features will help users navigate the Orlando materials effectively.

3.1 Navigating

Navigation was from the outset a major concern in our delivery design. This is a practical implication of Jerome McGann’s insight into the radiance of textuality that characterizes both print and electronic texts and makes designing tools for their electronic dissemination and study so challenging. As he says, “Every document, every moment in every document, conceals (or reveals) an indeterminate set of interfaces that open into alternate spaces and temporal relations” (2001, 181). Interacting with a complex electronic textual resource on the World Wide Web makes a huge array of navigational options both within and beyond that resource immediately available. In revising the Orlando Project’s delivery system we’ve tried to strengthen our strategies to orient users. These include: hierarchical organization of the site into the home page and three major entry points; the tabbed screens format stressing relationships between sets of materials; consistent formatting and reuse of screen elements; keeping screen proliferation to a minimum; and, with the help of our designer Stan Ruecker, providing a sense of prospect on the textbase as a whole.

With the exception of the views of entries, the major functionality of the Orlando Delivery System, and particularly the links and tag search results screens, produce a series of excerpts that at first glance might seem to resemble the scattershot outcome of a Google™ search. They are, however, considerably more focused and organized than that, emerging as they do from a body of materials produced according to research, writing, and tagging protocols and returned not simply on the basis of semantic content but on the basis of the tagging structure. Experimenting with the alterna-
tive option of free text searching is likely to show how much this method aids the extraction of meaningful information. Interestingly, the structural differences between the two major DTDs outlined above have not seemed to us to significantly impact the tag search or the organized linking.

As noted above, exploiting the tagging for the Links screens provides us with a solution to the problem of a densely interlinked electronic environment that leaves the user in control. As Ward Tietz argues, “When we initiate a link, we have made a decision to proceed, to cross over, but if there are many possible link-choices and paths, the consequences of the choice intensify” (509). Although the Links screens refuse immediate gratification, the celerity associated with the hyperlink, they provide considerably greater user control than many other hyperlinking systems, and higher quality returns on the user’s risk investment in following the link.

Our testing indicates, not surprisingly, that being able to move through materials effectively is crucial: none of our testers reported having read individual entries from start to finish. This suggests that granularity will be a major criterion for the production of scholarship in electronic form and that its structure—which will determine how well one can manage the relationships among its components—will therefore become of paramount importance.

Thankfully, although the Orlando Project was conceived in large part around the entries, the structure of the documents means that chunks of them—notably the chronology structures and Divs2—can be databased, indexed in relation to the tagging hierarchy, and retrieved to produce the dynamism that users seem to desire. We decided relatively early that Divs2s and chronology structures would be irreducible units for search and delivery. However, the length of some of our Divs2s (which result from the way the <TEXTSCOPE> tag functions within the Writing DTD and the delivery system) has led us to experiment with the return of abbreviated or “short form” search results which provide shorter but less contextualized result sets.

Thus, although there is no linking narrative, and indeed a set of excerpts may in some cases seem quite heterogeneous or even tending towards opposing conclusions, this new form of literary history (embedded in and produced by the encoding and its interaction with semantic content) produces results that are meaningful, often in ways unanticipated by the user. For, as Jerome McGann argues of the Collex toolset being developed by the Applied Research in ‘Patacriticism’ initiative, the very act of placing materials in relationship to one another, remediating them
and juxtaposing them, makes an argument (2004). In this case, the argument is formed in the interaction between the document content and tagging structure devised and applied by the Orlando project team members, and the user’s own preoccupations, inquiries, and pathways through the materials. As Alan Liu has remarked of the Romantic Chronology project Laura Mandell and he co-edited, “an information-rich matrix . . . makes the relational database not just a means but a paradigm of knowledge” (310). Altogether, the Delivery System represents the first fruits of Orlando’s efforts to produce a user-oriented electronic literary history. Although we did not foresee what shape delivery would take, the content tagging we devised, together with our reliance on the TEI for basic structure and a set of core tags to facilitate hyperlinking, provides the basis for dynamically produced sets of informative and interpretive text about women’s writing in the British Isles. The delivery system aims to make the structuring and retrieval principles evident to the user, without letting them overwhelm, so that the conclusions users take away from their engagement with the materials can be informed by an awareness of the tagging. Results produced by the delivery system emerge from the interaction between user interests and the tagging: they are the conclusions of a new kind of inquiry within the context of shifting paradigms of knowledge representation and, with it, of scholarly production and dissemination in the humanities. Orlando’s tagging is highly interpretive. We hope it will, as a result, shift the focus of user attention away from received notions of information retrieval towards consideration of the processes of organizing, representing, and interpreting materials in electronic form. We hope it will lead to reflection on the constructedness of literary history, and on ways that history can be reconstituted.

We need urgently to engage scholars who are not computing humanists in the debate over how scholarly work is represented electronically. If, as Sean Latham argues, “the digital archive partially resurrects the complex embeddedness of cultural practices” (412), the Orlando Project insists on the complex embeddedness of literary historical information, relations, observations, and judgments. Latham hails the potential of “a critical practice that involves reading frenetically without losing sight of the depths such motion might otherwise obscure” (424). Designed to foster such a critical practice in our use of secondary as well as primary electronic texts, the public “face” of Orlando aims to help our larger intellectual community face the immense challenges of the move to digital scholarly text.
Notes

1 The Orlando Project acknowledges the Social Sciences and Humanities Research Council of Canada, the Canadian Foundation for Innovation, and the Universities of Alberta and Guelph for their generous support of the Project.

2 For instance, the first freeware SGML web interface, SoftQuad’s Panorama, was only released in 1995.


5 Our current work on the delivery system involves revising tag and attribute names to make them more easily grasped by users, as well as eliminating tags that have not been frequently used, so this count is down from a previous high of 245 tags and 116 attributes. For a range of reasons, some tags we initially envisioned have not in practice turned out to be frequently used in the history that we actually produced. An example is the <HISTORICALTERMCONTEXTDATE> attribute we had applied to a group of elements related to such issues as nationality, ethnicity, and so on. This attribute was considered to offer a helpful way of tracking, and signaling, the fact that identity categories change over time. However, since it is in practice quite difficult to assert when a particular historical term comes into usage and when it ends, and one would generally wish to contextualize and support such assertions discursively, this attribute was used very rarely.

6 On the development of the delivery system, see also Brown, Clements, Elio, and Grundy, “Between Markup and Delivery; or, Tomorrow’s Electronic Text Today.” Work on delivery has been undertaken by the Project’s core team and achieved by various team members working over the years. Work on the project’s delivery system continues, so views of it provided here will differ from the release version. Note also that incorporation of Topics into the delivery system will take place at some point in the future.

7 Because a literary history of women in the British Isles is in a sense a history of the entire field of writing, viewed with a particular set of concerns and emphases, the textbase necessarily includes treatment of male writers and writers who were not active in Britain, some with entries and some with more distributed treatment.

8 The home page provides links to introductory screens with tips for getting started, as well as providing introductory searches. How Orlando Works, also available from the home page, provides a fuller introduction. Each entry point has a general
help page explaining how it works, and a context-sensitive help feature (currently available through a right-click of the mouse, although this mechanism is being reconsidered) elucidates specific screen elements.

\(^9\) The principle of prospect is an adaptation, for a new medium and with expanded potential, of a traditional aspect of design for print. See Ruecker, Homich, and Sinclair, “Multi-level Document Visualization.”
Works Cited


