Enhancing Scholarly Publishing in the Humanities and Social Sciences: Innovation through Hybrid Forms of Publication

Paper prepared for PKS Scholarly Publishing Conference 2011

26-28 September 2011

by Nicholas W. Jankowski, Clifford Tatum, Zuotian Tatum, Andrea Scharnhorst

e-Humanities Group
Royal Netherlands Academy of Arts and Sciences, KNAW
Cruquiusweg 31
1019 AT Amsterdam, NL
Enhanced Publications Project Web site

ABSTRACT

Enhancing scholarly publication involves presentation in a Web environment with interlinking of the 'objects' of a document: datasets, supplementary materials, secondary analyses, and post-publication interventions. Development of the Semantic Web is aimed at facilitating long-term content structure through standardized meta-data formats intended to improve interoperability between concepts and terms within and across knowledge domains. At the same time, ad-hoc scholarly discourse, facilitated by the participatory dynamics of Web 2.0 applications, contributes to an emergent content structure through compliance with open Web standards. While the top-down Semantic Web and bottom-up intertextuality structure are not inherently incompatible, their differences have implications for the design, use, and diffusion of enhanced scholarly publications. In this paper we illustrate a hybrid approach that employs Semantic Web techniques while focusing on practices entailed in contemporary intertextual discourses. This approach is applied to four books prepared for traditional academic publishers; the Web sites and functions for three of these books are described in this paper.

INTRODUCTION

Scholars in the humanities and social sciences are increasingly considering possibilities to make research available on the Web. Instruments for data collection and analysis, datasets and metadata describing this material, conference papers and project reports are appearing in Web-based repositories. One area lagging behind in this trend, however, is development of Web venues that integrate the traditionally published book with the diverse materials related to an overall research project. 'Enhanced publications' is a term reflecting such integration and a range of initiatives have been supported by the SURFfoundation in the Netherlands to develop such forms of publication.

One of the six projects included in this initiative is 'Enhancing Scholarly Publishing in the Humanities and Social Sciences'. Initiated in January 2011, Enhancing Scholarly Publishing was designed to prepare Web sites for four traditionally-published scholarly books and, in the process, to utilize a model of and tools for enhanced publications developed by SURF. The project concluded at the end of June 2011. This paper notes the objectives and accomplishments during that period, reflects on some of the challenges encountered, and sketches paths meriting further exploration.

The project had three objectives: (1) to develop hybrid forms of publications, (2) to develop a database allowing for aggregation of content attributes and associations across the individual book Web sites, and (3) to disseminate the results of the experiences in preparing enhanced publications. Regarding the first objective, a template constructed within the WordPress platform was to be used to construct Web sites to complement the four selected books. These Web sites were to contain a broad range of features:

- Supplementary resources (e.g., links, blogs, chapter appendices, author profiles);
- Chapter visualizations (e.g., animations, figures, tables) in color;
- Hyperlinks, both internal and external to the book texts;
- Author updating of site materials;
- Search features.

Regarding the second objective, a database was to be developed that would allow for aggregation of content attributes and associations across the individual book Web sites, such that topical relationships, intellectual underpinnings, and contextual factors could be made explicit. Fundamental to this approach is a focus on Webbased texts as dynamic and evolving discourses rather than completed works ready to be archived.

The third objective of the project was to disseminate the results of the experiences in preparing enhanced publications. This objective involved sharing experiences through a range of activities: conference papers and presentations, such as this paper for the PKS conference; journal publications, such as a theme issue of *New Media & Society*; and discussions with publishers and other players involved in scholarly publishing.

Four books were selected for inclusion in the project: three edited anthologies and a single-author university-level textbook:

- Jankowski, N. W. (2009). e-Research: Transformation in Scholarly Practice. New York: Routledge.
- Wouters, P., Beaulieu, A., Scharnhorst, A., & Wyatt, S. (forthcoming, under review by MIT Press). Virtual Knowledge.
- Park, D., Jankowski, N. W., & Jones, S. (2011). The Long History of New Media: Technology, Historiography, and Newness Contextualized. New York: Peter Lang.
- Jankowski, N. W. (forthcoming 2012). *Digital Media: Concepts & Issues, Research & Resources.* Cambridge UK: Polity Press.

The conceptual foundations of the Web sites and the underlying principles of the database architecture are elaborated in the next section. Web sites complementing three of the four scholarly book publications noted above are presented in the subsequent section (the fourth book is insufficiently developed to merit presentation at this time). Finally, in the concluding section we reflect on the overall project and note areas where further work is to be undertaken.

PLATFORM DESIGN AND FEATURES

As an important mode of scholarly communication, particularly in the humanities, the academic book format has experienced relatively little enhancement from the affordances of digital media, networked content, and database technologies. In this project we leverage the potential for creating enhancements for the book in its present form. The basic architecture for this project includes a database-driven Web site for each of four traditionally-published books, along with a central aggregation database that facilitates synchronization features and will in a future stage of development include queries within and across the four books.

As reflected in Figure 1, a conceptual diagram of the project, content on each book Web site is managed with a local database, connected to a central database. In this way, the linkage is established for aggregation and each book Web site retains an individual Web presence with local content management and storage. In addition to the five Web sites (four book Web sites and the central project Web site), a central accomplishment for the project is the launch of Semantic WordPress for Digital Scholarship, termed Semantic WORDS, and two specially tailored open source plugins designed to introduce traditionally published books to Web-based scholarly communication.

Each of the four book Web sites contains a broad range of features intended to enhance the printed versions of these books, including supplementary resources, visualizations, intertextual linking of content, and formal

structuring of content using Semantic Web ontologies. In a subsequent phase, further development is planned of the central database in order to facilitate aggregation of content across the individual book Web sites, such that object relationships, discursive threads, and contextual factors can be made explicit.

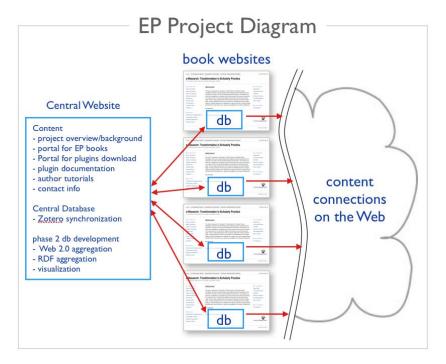


Figure 1: Schematic Diagram of EP Project

The following Web sites have been developed as part of this project:

- <u>Central Project Web site</u>;
- <u>e-Research</u> book Web site;
- Virtual Knowledge book Web site;
- Long History of New Media book Web site.

In this project we employed the <u>WordPress</u> Content Management System (CMS) as the foundation for the Web sites, both for its relative ubiquity and ease of use. According to the recent World Wide Web Technology Survey (<u>W³Techs</u>), the WordPress platform is used by 14% of the largest one million Web sites on the Internet, which amounts to more than half of the Web sites using a content management system.

The development strategy for this project focused in the first place on academic practice: how academics in the social sciences and humanities conduct scholarship. This priority informed development-specific functionality and interfaces. To this end, the relative ease of using WordPress played an important role. Figure 2 displays the functional modules of the platform in three clusters: the WordPress software (upper left), community-developed WordPress plugins (lower left) and the custom-developed plugins (right).

Semantic WordPress for Digital Scholarship

The Semantic WordPress for Digital Scholarship framework (<u>Semantic WORDS</u>) is the basis for this hybrid platform, which leverages Web 2.0 participatory modes of scholarly communication combined with formalized content structures imposed by Semantic Web formats; see Figures 2 and 4. The Semantic WORDS software is developed as open source for use with the WordPress Content Management Software, which is also open

¹ For background and rationale of this hybrid approach to enhanced publications, see Tatum, (10 June 2011) '<u>Web 2.0</u> and/or Semantic Web?' available on the Digital Scholarship site.

source. Semantic WORDS is comprised of two custom plugins that are integrated with WordPress and <u>Zotero</u>, the open source, Web annotation and citation management system.

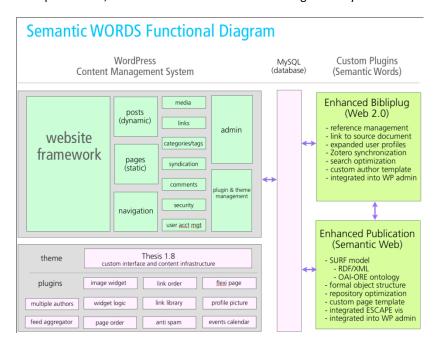


Figure 2: Semantic WORDS Functional diagram

Enhanced Bibliplug

The first of the Semantic WORDS plugins, Enhanced Bibliplug, provides a suite of features for authors, which are focused on organization and publication of academic content on the Web. Features include custom page templates for academic texts, integration with Zotero for citation management, and expanded author profile pages for CV content management, such as publications, presentations, projects and other related career accomplishments. In addition to providing authors with advanced tools for publishing on the Web, Bibliplug facilitates visibility (e.g., in search engines) of relationships between and among researchers, institutions, and both formal and informal scholarly communication.

Bibliplug was first developed for the Virtual Knowledge Studio (VKS) in 2009, and is still in use on some dozen project-related Web sites. At the time of development, the goal was to create a central repository for all researchers affiliated with the Studio to organize their academic work. The initial design included: (a) database schema for storing bibliographical references, (b) administration pages to manage the references, and (c) short code for easy retrieval of references based on author, year, and publication type.

In this project we further developed the plugin and re-released it as Enhanced Bibliplug. Added functionality includes: (a) the ability to connect and synchronize with Zotero accounts (see Figure 3), (b) a custom Author page template to display user's academic title and affiliation, bio, and CV content such as publications and presentations, (c) the ability to export bibliography data in RDF format based on the SPAR ontologies, and (d) the ability to group references based on categories and tags.

Enhanced Publication for WordPress

The second plugin, Enhanced Publication for WordPress, works in parallel with Bibliplug. Added content is simultaneously structured in Semantic Web formats based on academic publishing ontologies. Unlike many Semantic Web applications, this plugin includes integration of a visualization feature, such that object relationships can be browsed with the InContext application developed by the SURFfoundation. The central function of this plugin is to describe a WordPress site as an OAI-ORE aggregated book (an enhanced publication). In this structure, we convert WordPress pages into book chapters and use various other plugins to facilitate and describe reference lists, authors and editors, and attachments. For visualizing the content object relationships, we employ SURF's InContext Visualiser; see Figure 5.

This combination of Web site and book objects was not readily compatible with any single ontology. We therefore selected a list of related ontologies to describe the full content of the aggregation. Following is a list of ontologies used in creating a new ontology that satisfied this need:

- RDF: Resource Description Framework Ontology;
- OAI-ORE Vocabulary for Resource Aggregation;
- DCTerms: Dublin Core Metadata Ontology;
- FOAF: Friend of a Friend Ontology;
- FRBR: Functional Requirements for Bibliographic Records Ontology;
- SWAN: Provenance, Authoring and Versioning in scientific discourse Ontology;
- RES: Academic Researchers Ontology;
- BiRO: Bibliographic Reference Ontology;
- FaBiO: FRBR-aligned Bibliographic Ontology;
- PRISM: Publishing Requirements for Industry Standard Metadata Ontology;
- ESCAPE-Display: Vocabulary for describing inverse relationship of FOAF.

In addition to our custom plugins, we use several plugins from among the wide range of open source plugins developed by the WordPress community. Eleven of these are depicted in the diagram shown in Figure 2. We use another three plugins to augment functionality in our custom plugins: <u>Co-Authors Plus</u>, <u>Ninja Page Categories and Tags</u>, and <u>User Avatar</u>.²

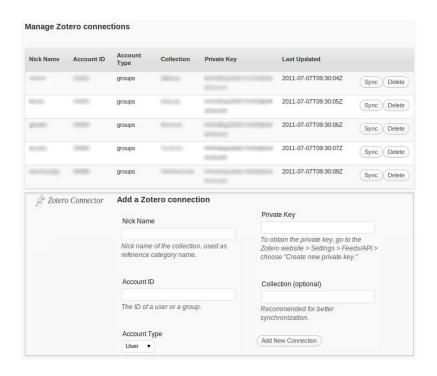


Figure 3: Zotero Admin page in WordPress

⁻

² The WordPress community has produced more than 15,000 plugins; see http://wordpress.org/extend/plugins/.

Aggregation Structure

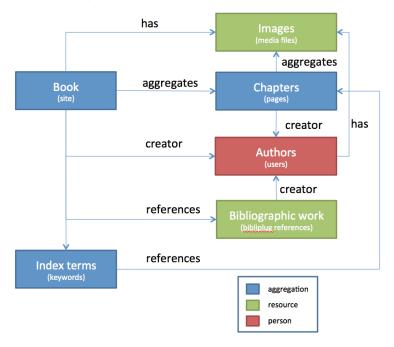


Figure 4: Semantic WORDS Aggregation Structure

InContext Visualization

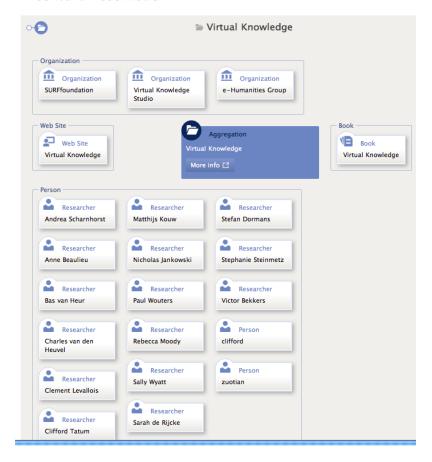


Figure 5: InContext Visualization – Screenshot from the Virtual Knowledge Web site

ENHANCING SCHOLARLY PUBLICATIONS

This section presents work undertaken in developing the Web sites for three traditionally published, or to be published, books noted above. Inasmuch as there are specific details and features related to each, they are presented in separate subdivisions of this section. All of the book Web sites, however, are based on a uniform template constructed within the WordPress content management system described in the previous section.

Book 1: e-Research: Transformation in Scholarly Practice

The book *e-Research: Transformation in Scholarly Practice* was released mid-2009 by Routledge and reflects the characteristic features of traditionally published and specialized scholarly monographs: hardcover, black text printed on white paper, and figures reproduced in tones of gray. There is no use of color in the book, other than on the cover. A Web-based enhanced version of this publication could include myriad features associated with Web sites, such as:

- Illustrations, figures and tables, in color;
- Internal hyperlinks between sections of the book;
- External hyperlinks to related Internet-based materials;
- Supplementary resources for book chapters (e.g. recent publications, multimedia and other materials).

Many additional features are also possible:

- Interlinking index terms with book text;
- Chapter references with hyperlinks;
- Author search via Google Scholar for other publications;
- Key word search for similar publications;
- Periodic updating of material by chapter authors;
- Comment and blog functions facilitating interactions between readers and authors.

Routledge granted permission to place the text of the book on the Web site, and this has allowed us to illustrate how the chapters will be presented in both pdf and html file formats. At this time, two chapters have been prepared in this fashion. Figure 7 shows the Web site page with links to presentations given by authors; Figure 8 illustrates information on related books and links to sites associated with these publications. Figure 9 depicts author information from the database created for the book.

Although preparation of the Web site complementing this book is well underway, the text for all 17 chapters has not yet been uploaded to the site. Once completed, these chapter presentations will also include the following functionalities:

- Search function through chapter texts;
- Hyperlinks embedded in chapters;
- Pop-up figures and tables in chapters.

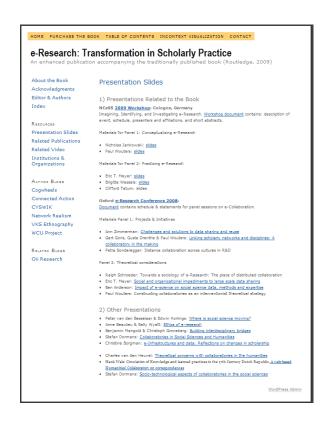


Figure 7: Links to presentations given by contributors to e-Research



Figure 8: Publications related to e-Research

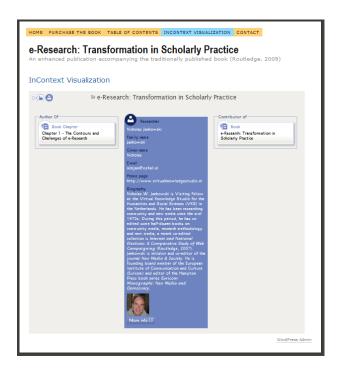


Figure 9: Author information from e-Research database

Book 2: Virtual Knowledge

The book manuscript *Virtual Knowledge* is based on research prepared by scholars associated with the Virtual Knowledge Studio for the Humanities and Social Sciences (VKS), established by the Royal Netherlands Academy of Arts and Sciences (KNAW) in 2006. The primary objective of the VKS³ was to facilitate innovative research practices in the humanities and social sciences, and the book *Virtual Knowledge* is designed to reflect that aim. Contributions came from scholars associated with the three divisions of the VKS: in Amsterdam, Rotterdam (Erasmus Studio), and Maastricht (Maastricht Studio). One function of the book project was to enhance cohesion among the wide array of VKS projects and to foster interactions among staff at the three divisions of VKS.

From its conception, VKS intended to initiate and conduct new research practices, and to engage with on-going innovative practices of other researchers. In this regard, VKS researchers were both 'makers' and 'observers' of new digital scholarship. Two notions central to Science and Technology Studies (STS), which constituted the home discipline of many of the central members of the VKS, are practice and community. These notions are reflected in the preparation of *Virtual Knowledge* and in the complementing Web site. Regarding the book, three workshops were conducted during preparation of chapters; regarding the Web site, one workshop was held related to preparing and uploading contact for the site.

Based on interactions during preparation of the book, it was decided to prepare a Web complement to the print volume. Several considerations contributed to adoption of this idea:

- To continue interactions among authors;
- To support the formation of a community around ideas expressed in the book;
- To embed the book in an emerging environment of similar books;
- To disseminate and promote of the book.

To support preparation of an enhanced publication for the book and to explore how preparation of such a Web complement might facilitate the previously mentioned community function, a workshop was organized in April

³ The VKS concluded operation on 31 December 2010 and the e-Humanities Group has been mandated to continue the work of the VKS in a more focused manner and under a modified organizational structure. Further aspects of this transition are indicated on the Web site of the e-Humanities Group.

2011 for book contributors. Of the 17 contributors, seven attended the workshop. The event provided opportunity for participants to become familiar with the Web site and the general procedures for uploading information, including bibliographic entries that were submitted with a specially prepared plugin for the WordPress site. The workshop concluded with a general discussion, during which some persons expressed regret at not being involved in an earlier stage in order to contribute to the design process and the user interface with the site. This discussion was continued in a post-workshop survey that allowed all contributors to reflect on the Web site under construction.

The level of contribution during and after the workshop was modest. While content has been uploaded to the site, much remains to be completed. That acknowledged, preparation for the workshop did stimulate members of the project team to complete the Web site template and specially developed plugins for bibliographic entries. Some of the criticisms of the workshop and reservations about an enhanced publication included:

- Inadequate involvement of the book editors in the planning;
- Unclear value of a book Web site for authors;
- Time constraints prevented engagement at the desired level;
- Uncertainty about the utility of some site features, including author photos and videos;
- Technical problems experienced with the site, including functioning of the interface;
- Insufficient support from project team members in using the site.

Some of the positive reactions to the workshop and EP Project included:

- Appreciation for being able to link references;
- Acknowledgment of potential value in creating cohesion of edited collections through an enhanced publication;
- Value of Web site for author visibility;
- Relevancy of site to own research practices.

Negotiations are ongoing with the publisher to which the book manuscript *Virtual Knowledge* has been submitted for consideration. Preliminary reactions reflect interest in publishing the volume and in combining the book with an enhanced publication in the form developed during this project. To this end, a <u>preliminary Web site</u> has been prepared for the book and includes the basic functionalities included in the WordPress template; see Figure 10 for illustration of the page describing the book.

Many of the functionalities for the accompanying Web site will remain important and further work will be required to complete preparation of the content related to these features (e.g., providing supplementary resources such as links, uploading bibliographic entries, completing video films of authors reflecting on their chapters).

It is anticipated that a second workshop for authors may be necessary once arrangements have been made with the publisher regarding preparation of the book. This workshop will build on the experiences of the initial workshop held during this project.



Figure 10: Page 'About the Book' from Virtual Knowledge site

Book 3: The Long History of New Media

The third book that is part of this project on enhanced publishing was released by Peter Lang in May 2011 and is entitled *The Long History of New Media: Technology, Historiography, and Contextualizing Newness*. As with the other books in the project, this is an edited volume and has been prepared and published in a manner reflective of conventional procedures for scholarly publishing. The book was released as a paperback and the cover consists of a designed arrangement of book title and names of editors. The text of the book is printed in black ink on white paper; there are few illustrations and no tables in the book.

The Web site constructed for <u>The Long History of New Media</u> contains a similar set of features as prepared for the other two books and uses the same WordPress template for the site; see Figure 10 illustrating the homepage of the site and Figure 11 containing biographical sketches of contributing authors.

Inasmuch as the book has recently been released by the publisher Peter Lang and no prior arrangement was made for reproducing the full book manuscript, only introductory paragraphs from the chapters have been uploaded to the site, along with the text of the introduction chapter. These texts are temporary additions to the site and require approval from the publisher before the site is formally announced, which is planned for the autumn 2011.

The Web site is under construction and the content will mirror that available on the site for the *e-Research* book, and include the following features:

- Book-related materials: description of book, table of contents, chapter abstracts, figures from chapters, compilation of references, and publisher information.
- Profiles of contributors: photos and bios of authors and editors;
- Supplementary resources: lists of institutions, publications, videos, and presentations related to Web history;
- Topic-related blogs: group blog for authors of the book, individual blogs by book authors, and other blogs relevant to the themes in the book;
- Interlinking index terms with book text;
- Figures reproduced on Web site; figures in color;
- Chapter references with hyperlinks and an overall bibliography for book;
- Author search via Google Scholar for other publications by author;
- Key word search for similar publications based on chapter titles.



Figure 10: Homepage of site for Long History of New Media



Figure 11: Author biographical sketches, Long History of New Media

CONCLUSION

In this project a form of enhanced publication has been developed for traditionally prepared and published scholarship. The undertaking has involved an accelerated learning experience in the terminology and models underlying the initial SURF tender. We accepted the opportunity made available by the tender to achieve our initial and central objectives: preparation of Web sites to complement traditionally published scholarly monographs, along with construction of an overarching database containing materials from those monographs. We were intentionally ambitious in formulation of objectives for the project, and it is not surprising that those objectives were only partially accomplished. Still, the work undertaken reflects considerable accomplishment during the six months available for the project. Much remains to be done, however, and some of the remaining tasks are itemized in this section. Beyond a 'to do' list, this section reflects on the overall project and 'next steps' that build on and extend beyond this SURF project with enhanced publishing.

Original objectives

Three objectives were formulated in the revised proposal submitted in January 2011 by the e-Humanities Group to the SURFfoundation: (1) develop Web sites to complement four printed volumes released by commercial and university-based publishers; (2) develop a database allowing for aggregation of content across the individual book Web sites; and (3) share the experiences in preparing enhanced publications through conventional channels for scholarly communication, informal and formal. Each of these objectives involved a range of specific activities, and a summary of the activities undertaken is presented below.

Activities undertaken

The initial WordPress template for the book Web sites was redesigned to facilitate ease of use by book authors and ensure basic uniformity in the presentation of site content. Plugins for the site were designed, tested, and implemented; these plugins facilitate author uploading of references. Web sites were prepared for each of the four books using a common template, and illustrations of content for each of the books was uploaded to the respective sites. The amount of content uploaded varied per book because of the different phase of completion and 'life cycle' of each book. For example, the book *e-Research* was released two years ago and the publisher agreed to allow the full text of the book to be placed on the Web site. The book *Digital Media*, in contrast, is still being written and only brief descriptions of chapters are made available. A database has been constructed for each of the book titles and these individual databases are integrated into an overall database. Finally, a

range of dissemination activities have been initiated: from individual talks with publishers and authors to formulation of proposals for conference panels and papers.

Remaining tasks

Perhaps the most pressing task remaining to be achieved involves further population of the four book Web sites with content. This is a prerequisite to approaching the authors of the respective books and requesting update of material. We envision availability of adequate content on each of the sites by late 2011, at which time we will invite chapter authors to revise and update the content as necessary. In the case of the book *Virtual Knowledge*, site completion will be performed in consultation with the publisher of the volume which has expressed strong interest in such a complementary Web venue to the printed volume.

Reflections

Three points merit comment in this final section of the report: (1) project duration, (2) database construction, and (3) the original aspiration of this project to accentuate the communicative component of enhancing scholarly publications. First, six months is a very short period of time and is inadequate to achieve the initially formulated objectives. With hindsight, we should have strived for less in order to achieve more. That acknowledged, progress was made across four different scholarly titles for development of complementing sites using a platform with a large and active community of users. Second, we initiated construction of a database for objects from each of the books with some reservation, not knowing whether and how the construction would develop. As it turned out, this component of the project developed well and the work accomplished positively reflects the potential of enhanced publications containing such infrastructure. Third and finally, we underestimated the need for exchange with book editors and authors; although one of the books, *Virtual Knowledge*, was well-placed regarding short communication lines and author interest, we nevertheless encountered considerable reservation at the time of a specially planned workshop, partially reflected in limited attendance and subsequent uploading of content. The 'lesson' from this experience is that the need for personal communication among central players should not be underestimated.

In conclusion, it may be appropriate to mention that all members of the e-Humanities Group Enhanced Publication Project are extending the achievements of this project into other arenas of journal and book publishing. Other enhanced publications are in preliminary stages of planning as well as preparation of reflective texts on this trajectory of scholarly communication. We welcome the challenges these plans involve and in contributing further to a hybrid form of Web-based and traditional publishing.