

# Models for Publishing Academic Hypertexts

Peter J. Nürnberg and David L. Hicks

Department of Computer Science, Aalborg University Esbjerg

Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark

<http://cs.aue.auc.dk/{pnuern,hicks}@cs.aue.auc.dk>

**Abstract.** Hypertext has grown rapidly in popularity, from being a relatively obscure concept in the late 1980's to near ubiquity with the advent and spread of the World Wide Web. Many forms of information are naturally and easily expressed as hypertexts, especially those that draw connections among various works. Academic work is relatively highly structured work, both internally, and in how it is related to existing work. Nonetheless, aside from a few systems that provide simple footnote and reference linking, no serious attempts have been made to publish academic articles in hypertextual form, with robust, first-class structures comprising a large part of the product. Previous investigations into the requirements for supporting the publishing of academic articles as fully functional hypertexts have noted numerous reasons for slow progress toward this goal. In this paper, we look closely at models for publishing hypertexts. We do so by pointing out several axes along which publishing models may differ, providing a systematic way in which to discuss these models. We then examine three particular models, situating them within our framework, and describing some of their implications. Examining these instances allows us to hypothesize more generally about the implications of choices posed by our framework.

## 1 Introduction

Hypertext is a medium for making connections explicit. Most people are familiar with hypertext through the World Wide Web. However, hypertext systems have been in use for many decades (e.g., NLS [3], KMS [1]), and were considered theoretically by Bush over a half-century ago [2]. Whereas in traditional, linear media, connections between parts of texts are made implicitly (e.g., through the use of footnotes or citations), hypertext allows these connections to be made explicit by

instantiating traversable links or structures between related parts of texts. In traditional media, connections are followed manually by the reader (e.g., scanning to the bottom of a page to find a footnote or going to the library to find a reference); in hypermedia, links may be followed by a simple interface action (e.g., clicking the mouse on blue, underlined text).<sup>1</sup>

Hypertext is an ideal medium for complex, highly structured information, such as academic articles. In fact, in his seminal article [2], Bush uses a running example of research articles rendered in hypertext as an argument in favor of this new medium. Because a hallmark of quality research is its positioning within the field relative to related work, academic articles are often highly interconnected. However, to date, few serious attempts have been made to allow authors to publish articles hypertextually. Generally, published academic articles are paper-based or contain only the most rudimentary hypertextual capabilities. Previously [6], we have discussed reasons why this might be so. In this paper, we look at more detail at some of the possible models of publishing with an eye toward how these models might support hypertext publishing.

First, we consider the space of publishing models by describing a number of axes along which these models may differ. Secondly, we populate this space with three representative model instances. Finally, we draw some preliminary conclusions about potential models to support hypertextual publishing.

## 2 Publishing model space

Publishing models may vary in a number of ways. We attempt to organize the space of publishing models by defining a number of axes along which models may differ. We begin by presenting some definitions used throughout the remainder of the paper. Then, we take a closer look at our model space.

### 2.1 Definitions

We use the following definitions throughout the paper. Our usage is intended to be inline with common usage of these terms, but our notion of *publisher* may be somewhat unfamiliar to some readers. We believe our usage reflects current trends, however.

- *Content*: primarily academic research articles, but secondarily, other content such as appendices, reviews, editorials, etc.
- *Author*: a generator of content.

---

<sup>1</sup> Hypermedia is generally taken to include richer, more complicated structures than simply user-traversable links (see [5] for more examples). However, for purposes of this article, one may assume that user-traversable links are a defining feature of hypertext.

- *Reader*: a consumer of content.
- *Publisher*: a reviewer/endorser of content. Traditionally, a publisher is the institution responsible for printing and distributing content. However, increasingly, publishers (especially those of online content) are primarily responsible for some sort of quality assurance on content with which they are associated.

## 2.2 Model space axes

There are many axes along which models for publishing academic hypertexts may vary. In this section, we describe several of these. Although this list is not complete, we feel it provides a reasonable starting point for further discussions. For each axis, we briefly describe the universe of possible values, and give an example of the meaning of each axis value. The axes are not presented in any particular order. For quick reference, Tab. 1 summarizes the axes and their values.

**Table 1.** Summary of model space axes and possible values

axis name	possible values
primary content delivery format	<i>paper</i> <i>renderable electronic</i> <i>non-renderable electronic</i>
content state	<i>living</i> <i>dead</i>
storage	<i>reader stored</i> <i>publisher stored</i> <i>author stored</i>
primary commodity	<i>copy ownership</i> <i>subscription</i> <i>pay-per-view</i>
publication cost bearer	<i>reader pays</i> <i>author pays</i> <i>advertisers pay</i> <i>publisher pays</i>
secondary content status	<i>undifferentiated</i> <i>differentiated</i>
review timing	<i>pre-publication</i> <i>post-publication</i>

**Primary content delivery format.** Content may be delivered in several different formats. However, in most models, one format will be considered primary, while others are derived.

- *Paper.* Hardcopy of the content is considered primary, while electronic versions are secondary or derived. For example, traditional print journals deliver paper hardcopy primarily, even if they may also provide electronic versions of much of the content delivered in paper form.
- *Renderable electronic.* While the primary delivery format is electronic, linear renditions<sup>2</sup> are easily generated. For example, a journal may primarily distribute PDF files, which may be easily printed.
- *Non-renderable electronic.* Not only is the primary format electronic, but no easy method for rendering the content into a linear rendition exists. For example, a journal may distribute complex, interwoven hypertexts on-line.

**Content state.** Content may either be characterized in one of two states.

- *Living.* Content may be modified after publication. For example, a journal may allow modifications and corrections to be made to a published article, assumedly marking these modifications and corrections as such. *Living* content implies *electronic* (renderable or not) primary delivery format.
- *Dead.* Content is “frozen” after publication, meaning “logical modifications” or corrections are made in the form of follow-up articles, errata, or other such separate pieces that refer back to the original work.

**Storage.** Content may be stored in a variety of places.

- *Reader stored.* Content is stored by the reader. For example, traditional paper-based journals are generated by the publisher and then sent to readers. This option is meant to include storage by such “reader service institutions” such as libraries.
- *Publisher stored.* Content is stored by the publisher. For example, a journal publisher may store PDF files, which then may be retrieved by readers.
- *Author stored.* Content is stored by the author of the content. In this case, the publisher acts as a “seal of approval” on content, perhaps by affixing a digital signature to author-provided content, indicating, for example, that the said content was reviewed by the publishing institution.

<sup>2</sup> By “linear rendition”, we mean to include any of several traditional linear formats, such as printed hardcopy, speech audio stream, braille tape, etc.

**Primary commodity.** Publishers may sell content in different ways. This axis is heavily dependent upon primary content delivery format.

- *Copy ownership.* A reader may purchase the ownership of a copy of content. For example, a reader of a traditional journal may pay for a volume of the journal. Copy ownership usually implies primary paper content, but may apply to electronic media as well (e.g., buying a CD-ROM).
- *Subscription.* A reader may purchase the right to view content for an unlimited number of times within a limited time frame. For example, a reader of an electronic journal may pay for one month of access to the journal's content. Subscription implies some type of electronic content (renderable or not), since paper content is sold rather than leased.
- *Pay-per-view.* A reader may purchase the right to view content for a limited number of times (over a presumably unlimited time frame). For example, a reader of an electronic journal may pay for 20 accesses to the journal's content. Pay-per-view often implies non-renderable content, since otherwise, readers may "cheat" the system by collecting renditions of content.

**Publication cost bearer.** Different entities may bear the costs associated with publication. For this axis, any combination of the value below are possible.

- *Reader pays.* A reader may (indirectly) bear the cost of publication. For example, part of the cost that a reader of a traditional paper journal pays to the publisher may be used to offset publication expenses incurred by the publisher.
- *Author pays.* The author may bear the cost of publication. For example, many conferences essentially charge publishing authors through conference fees, page charges, etc.
- *Advertisers pay.* Advertisers may bear the cost of publication. For example, a traditional paper-based journal may offset publication costs through advertising income.
- *Publisher pays.* The publisher may bear all costs associated with publication. The publisher may expect that other derived benefit outweighs the publication costs. For example, the publisher of an electronic journal may not charge anyone for the cost of maintaining the journal.

**Secondary content status.** Secondary content, such as reviews, editorials, errata, etc., may be treated in several different ways.

- *Undifferentiated.* Secondary content may be treated identically to primary content. For example, an electronic journal may publish reviews in the same manner as it publishes reviewed articles. By this, we mean that secondary content is managed as primary content with regard to primary delivery format, content state, storage, publication cost bearer, and primary commodity.

- *Differentiated*. Secondary content may be treated differently than primary content. For example, an electronic journal may offer articles via pay per view, but offer editorials free of charge. Differentiated status may vary in degree. That is, secondary content in some models may be highly differentiated, while in others, it may only be slightly differentiated.

**Review timing.** Reviews for academic articles may be handled in two ways.

- *Pre-publication*. Primary academic content may be reviewed before publication. For example, most traditional paper journals first review submissions before publishing them.
- *Post-publication*. Primary academic content may be published before being reviewed. Assumedly, such content is marked as not having yet been reviewed. For example, some electronic journals “publish” all submissions and then solicit reviews. Such articles are marked as “under review”. After the review process, articles are marked as either “accepted” or “rejected”. Post-publication review implies *living* content state.

### 3 Populating the model space

In this section, we consider three models of publishing academic hypertexts in light of the axes discussed above. We have chosen these models in order to give good coverage of the model space, not to endorse any particular model. We also consider more generally the success of these models so far in supporting hypertext publishing, and/or the likelihood of being able to support this aim in the future.

#### 3.1 Traditional model

By “traditional model”, we mean the publishing model that has, until recently, been used by the vast majority of journals. In our model space, the traditional model is defined as follows.

- *Paper* primary content delivery format: this model stresses paper delivery, even in the presence of possible accessible online versions.
- *Dead* content: this model delivers dead, unmodifiable content.
- *Reader stored* content: this model delivers content to readers (or reader institutions, such as libraries), who are then responsible for storage.
- *Copy ownership*: readers pay for the right to own a fixed number of copies of the content.



- A combination of *reader pays*, *publisher pays*, and sometimes *advertisers pay* pay options: the cost of publication is borne by some combination of the publisher and reader, and possibly also by advertisers. In this case, publication cost includes costs for typesetting and physical delivery of content.
- *Undifferentiated* secondary content status: delivered content makes no functional distinction between primary academic and secondary content.
- *Pre-publication* reviews: only articles vetted by the review process are published.

This model cannot support hypertext publishing, since paper is an inadequate delivery medium for hypertextual content.

### 3.2 Emerging e-publishing model

By "emerging e-publishing model," we mean the publishing model that is being adopted by a growing number of online journals. In our model space, the emerging e-publishing model is defined as follows:

- *Renderable electronic* primary content delivery format: this model stresses electronic delivery of content which is essentially inherently linear. This is often equivalent simply to providing softcopy of paper content.
- *Dead* content: this model usually does not allow modifications to published material, often because the "underlying" content is itself paper-based, and therefore not easily modified.
- *Publisher stored* content: this model usually implies that the publisher sets up a web site or similar electronic repository for storage of content.
- *Subscription*: readers pay for the right to retrieve content an unlimited number of times within a limited time frame.
- A combination of *reader pays*, *publisher pays*, and sometimes *advertisers pay* pay options: the cost of publication is borne by some combination of the publisher and reader, and possibly also by advertisers. In this case, publication cost includes storage and accessibility assurance costs.
- *Undifferentiated* secondary content status: delivered content makes no functional distinction between primary academic and secondary content.
- *Pre-publication* reviews: only articles vetted by the review process are published.

This model can support hypertext publishing, but the hypertextual capabilities of content in such a model are likely to be relatively impoverished. This is because content is usually "backed" by hardcopy, which lacks hypertextual capabilities. The hypertext functionality presented in this model tends to be that which can be automatically generated by computer, such as limited forms of footnote following and navigation from citations to associated bibliographic entries.

### 3.3 Hypothetical model

By “hypothetical model”, we mean a publishing model that explores some parts of the model space unexplored by the previous two models. We do not necessarily endorse this model, but believe it is instructive to consider. In our model space, the hypothetical model is defined as follows:

- *Non-renderable electronic* primary content delivery format: this model places more stress hypertextual facets of the content than on paper renditions.
- *Living* content: this model allows updates to content, meaning that readers buy access (either through subscription or pay-per-view) to changing content bases.
- *Author stored* content: this model views the primary role of publishers as vetting organizations that digitally sign and ensure availability of author stored content.
- *Pay-per-view*: readers pay for the right to view content a fixed number of times over an unlimited time span.
- *Author pays* option: the cost of publication is borne by the author. In this case, publication cost includes storage and accessibility assurance costs.
- *Differentiated* secondary content status: secondary content, including reviews and editorials, is delivered free of charge in this model.
- *Post-publication* reviews: articles may be posted by authors at any time. Publishers may act as portals to unvetted content by clearly marking it as such. After a review process, content is reclassified as either accepted or rejected.

This model can potentially be used to support hypertext publishing. Focusing on electronic content that need not be renderable allows for rich, interactive hypertextual capabilities. Additionally, because content is stored (hosted) and paid for by authors, the particular functionality offered by a given article is not constrained to a standard set of functionality provided by a publisher or available on a reader's machine. Potentially, an open set of server-side (author-hosted) computations may support an open set of hypertextual interactions with content. Secondary, standardized content might be associated with a set of standard functionality and perhaps even be hosted by the publisher. As has been previously shown [4], hypertext can be used to support versioning, offering one possible implementation technology for living documents.

## 4 Conclusions

In this paper, we have examined various models of publishing hypertextual academic articles. We have delineated a model space and briefly examined three possible models, with a special view toward how these models support rich hypertextual capabilities. Our discussion has shown that not only do a number of “obvious” factors (e.g., primary content delivery format) influence a model's ability to support



hypertext publishing, but that a number of less obvious factors (e.g., who stores content, who pays for publication) also exert influence. Factors such as primary commodity were not shown to impact the ability to publish hypertexts.

## References

1. Akscyn, R. M., McCracken, D., Yoder, E.: KMS: A distributed hypermedia system for managing knowledge in organization. *Comm. ACM* 31(7) (Jul 1988), 820-835.
2. Bush, V.: As we may think. *Atlantic Monthly* 176(1) (Jul 1945), 101-108.
3. Engelbart, D. C.: Augmenting human intellect: A conceptual framework. SRI Tech. Rep. AFOSR-3223 Contract AF 49(638)-1024 (Oct 1962), Palo Alto, CA.
4. Hicks, D., Leggett, J., Schnase, J., Nürnberg, P.: A hypermedia version control framework. *ACM Trans. Inf. Sys.* 16(2) (Apr 1998).
5. Nürnberg, P. J., Wiil, U. K., Leggett, J. J.: Structuring facilities in digital libraries. *Proc. Euro. Conf. on Dig. Lib. 1998* (Crete, Greece, Sep).
6. Nürnberg, P. J., Hicks, D. L.: Peer-reviewed, publishable hypertexts: a first look. *Proc. of OHS 7* (S. Reich, ed.), LNCS 2266, Springer Verlag, Berlin, 2001.