Collaborative Authoring, Evolution, and Personalization for a "Transdisciplinary" Textbook

Robert J. Glushko
School of Information
University of California
Berkeley, CA 94720
+1-510-643-2754
glushko@berkeley.edu

ABSTRACT

This article is a case study about a book titled *The Discipline of Organizing*, which proposes a transdisciplinary synthesis of ideas from library and information science, computer science, informatics, cognitive science, business, and other disciplines that "intentionally arrange collections of resources to enable interactions with them."

This case study discusses the interrelationships between the transdisciplinary goal for the book, the process of collaborative authoring required to write it, the novel architecture of the book's content, and the innovative reading experiences in print and ebook formats that are enabled.

The idea that a new discipline is contextualized by more specific concepts and methods inevitably led to a collaboratively-authored book whose design embodies this intellectual architecture. The book's content is organized as a transdisciplinary core with supplemental content identified by discipline. This content model creates a "family of books" with thousands of siblings, any of which can be published in print or as an ebook by filtering on the disciplinary attributes.

This "design-time" customization has been extended to enable "reading-time" personalization for ebook formats. In addition, the rich semantic markup that enables customization and personalization is fodder for further experimentation about "smart textbooks" that can be continuously made smarter by dynamic discovery and inclusion of content.

Categories and Subject Descriptors

H.4.3 [Information Systems Applications]: Communications Applications – *Information browsers*. I.7.2 [Document and Text Processing]: Document Preparation – *Hypertext/Hypermedia*

Keywords

ebook, collaboration, document architecture, customization, personalization, reading experience

1. INTRODUCTION

This article is a case study that describes the challenges and insights that emerged in the design, development, and delivery of a book titled *The Discipline of Organizing* (TDO) [20]. TDO proposes a transdisciplinary synthesis of ideas from library and information science, computer science, informatics, cognitive science, business, and other disciplines that arrange collections of resources to enable interactions with them (Section 3).

A book with the ambitious goal of defining a new discipline must be broad enough to include all the disciplines that contribute to the "transdiscipline" (Section 2) that emerges at their intersection. It must treat each contributing discipline with enough depth so that the new concepts of the emergent discipline can be re-applied meaningfully to discipline-specific concepts and examples. A transdisciplinary book implies collaborative authoring by which experts in different disciplines identify and communicate about the gaps and overlaps between disciplinary concepts (Section 4).

To make TDO both broad and deep without making it bloated and hard to understand required some innovations in book design and implementation (Section 5). The key idea was to tag the book's content by discipline, effectively creating a family of related texts around a common transdisciplinary core. As readers selectively follow hypertext links to include discipline-tagged content, they are in effect tailoring the book for different courses and perspectives (Section 6).

Underlying these capabilities are single-source publishing technologies and methods by which different output programs can select, transform, and assemble a set of formats and editions by exploiting markup in a book's source files. Typically, the number of formats and editions produced this way from a content repository can be counted on one hand: a print edition or two, and maybe a couple of ebook formats. We have steadily enhanced the source markup and the programs that use it to enable the design time configuration of one of many thousands of possible books (Section 7), to allow a book's content and functionality to adapt to the capabilities of its ebook platform (Section 8), and finally to allow readers to

personalize the configuration of content in an ebook as they read it (Section 8).

2. DEFINING DISCIPLINARITY

A book that contains content by authors from more than one discipline can be described as multidisciplinary, interdisciplinary, or transdisciplinary. These terms can be defined to emphasize the relationships among different types of content in the book [32] or to focus on the collaboration among the authors to write the book [13]. We choose the former approach because we consider collaborative authoring an important topic on its own.

Multidisciplinarity in a book is easy to achieve – it simply means that more than one discipline is represented in the book. The disciplinary content could be entirely separated, as in a book with chapters or articles written by authors from different disciplines, and not necessarily about the same topic.

Books more often aim to be interdisciplinary, meaning that different commentators from different disciplines discuss some particular issues or phenomena. Because the authors consider the same topics, their collective contributions are often said to create a "confluence" or "intersection."

Transdisciplinarity, the goal we were striving for in *The Discipline of Organizing*, reflects a higher degree of content integration or synthesis of common underlying concepts, structures, relationships that are identified from the contributing disciplines. In contrast to interdisciplinarity, which describes the confluence of disciplines by enumerating the contribution of each discipline, transdisciplinarity requires that the new disciplinary intersection can be explicitly defined. This explicitness creates a new discipline whose abstractions can be reapplied to each of the disciplines to interpret discipline-specific concepts and examples.

3. THE DISCIPLINE OF ORGANIZING

Organizing is a fundamental issue in many professional fields, but there is only limited agreement in how they approach problems of organizing and in what they seek as their solutions. For example, the field of library and information science has traditionally studied organizing from a public sector and bibliographic perspective, paying careful attention to user requirements for access and preservation, and offering prescriptive methods and solutions. The disciplines of management and industrial organization deal with the organization of human, material, and information resources in contexts shaped by commercial, competitive, and regulatory forces. Computer science and informatics tend to study organizing in the context of information-intensive applications with a focus on process efficiency, system architecture and implementation.

Nevertheless, despite their obvious differences, the books in libraries, the employees of a firm, weather observations in a data repository, and digital songs on a music player are all "resources" – "things with value that can support goal-oriented activity" – that have been intentionally selected and organized. Similarly, despite their obvious differences, libraries, businesses, data repositories, and music collections can all be described as "organizing systems" – each is "an intentionally arranged collection of resources and the interactions they support."

A discipline of organizing complements the conventional disciplinary focus on specific resource and collection types (libraries organize books, museums organize art and artifacts, business systems organize product and customer information)

with a framework that views organizing systems as existing in a multi-dimensional design space in which different types of resources can be considered simultaneously, better exposing the relationships and contrasts among them. There are five groups of design decisions, phrased in generic language to emphasize their broad applicability: *What is being organized? Why? How much? When? How, or by what means?*

A set of concepts and vocabulary about organizing that spans disciplines enables people with narrower disciplinary perspectives to work together and learn from each other. Common intellectual ground is also a prerequisite for educational experiments at a scale beyond that of a single university, such as massive open online courses.

MIT Press published *The Discipline of Organizing* in print and ebook formats in 2013. The published book names seventeen co-authors, led by the author of this article, who also edited the book. Enhanced ebook editions that include several dozen photos, embedded quizzes and discussion questions, and other features that take advantage of digital reading platforms have been published by O'Reilly Media since 2014 ([21],[22]). In 2014 the book was named an "Information Science Book of the Year" by the Association for Information Science and Technology[2].

4. COLLABORATIVE AUTHORING

Almost by definition, a transdisciplinary book requires multiple authors to write it because it can only be written if experts in different disciplines can identify and communicate about the gaps and overlaps between disciplinary concepts. However, most people are specialists in only one area (I-shaped knowledge), a few specialize in two related areas (H-shaped), and fewer still are deep in one area with good knowledge across many other areas (T-shaped) [13].

The feasibility, effectiveness, and longevity of a collaborative authoring effort is shaped by many factors that collectively shape the collaboration process and schedule:

- the origin and extent of shared goals for the collaboration
- the relative professional status of the collaborators
- their relative degree of content expertise
- their relative degree of expertise with authoring and collaboration technology
- the extent of pre-existing personal and professional relationships among hem
- their relative time commitment.

4.1 Patterns for Collaborative Authoring

With this number and variety of factors there are a great many possibilities for bringing authors together to collaborate, and many researchers have attempted to describe and evaluate the configurations that predict successful or unsuccessful collaborations ([30],[31],[33],[38],[39],[42],[53]).

Each pattern for collaborative authoring makes different choices about the intellectual, procedural, technological, and social issues that arise when the collaboration is defined and as it proceeds.

One key contrast is between authoring modes in which the goal of collaborating authors is first to "construct a new reality" by reconceptualization and redefinition, versus authoring modes in which the authors work together to "document a view of reality" that already exists when the collaboration is created

[43]. Reconceptualization and knowledge creation might work best when it involves face or face or high-bandwidth synchronous communication. In contrast, "documenting" work can be defined and structured in ways that enable work to be distributed over place and time. Highly structured collaboration processes yield more consistent results, but of course gives the authors fewer choices about content and organization [15]. Some people describe collaborations that involve reconceptualization as inherently more democratic than those where the collaborating authors develop or flesh-out concepts and structures which they had less of a role in defining.

These findings vary significantly across disciplines. Successful scientific and engineering collaborations can involve large numbers of authors working in both highly structured and loosely structured ways; in the humanities collaborative authoring efforts to reconceptualize knowledge are almost frowned upon.

It is useful, for the *Discipline of Organizing* case study here, to contrast three patterns for collaborative authoring, because the choice of pattern had profound implications for content architecture and for the evolution and maintenance of the book.

Open Collaboration is well known in the literature of collaboration. It is exemplified by the popular and oversimplified notion of how Wikipedia works, or at least how it worked at its beginning: a potentially unbounded number of contributors, who are unlikely to know each other, voluntarily create and edit content about some subject for which they claim expertise ([36],[45]). Authoring guidance in the form of templates or data schemas might exist, but this is more syntactic than semantic. (As Wikipedia grew, contributors took on more specialized roles and more rules for content completeness and consistency were imposed, but this made it more hostile to newcomers ([18],[25]). An academic variant of Open Collaboration is practiced in many conferences or workshops that bring together authors with different perspectives on some topic in the hope that intellectual cross-pollination will occur.

When a group of authors who represent different disciplines or perspectives is formed to develop a shared vision and plan for the book they will write, this pattern might be called Consensus Collaboration. This pattern is more likely to produce successful results than Open Collaboration because its focus is on producing agreement rather than simply presenting different views. For example, "book sprints" are structured collaborations in which a facilitator helps the group reach consensus and write a book quickly [6]. However, if the process for reaching consensus is too bottom-up and democratic, it can sometimes fail because the social pressures to be agreeable in face-to-face encounters cause controversial or complex topics to be avoided. The consensus that is reached can be superficial and insufficient to justify the hard work of actually writing a shared book.

A third approach to collaborative authoring can be called Hierarchical Collaboration. Unlike the "sideways" or "bottom-up" processes of Open and Consensus collaboration, Hierarchical Collaboration is a "top-down" process that begins with a single author, or editor, who does the reconceptualization to create the vision for a collaboration, and who then recruits collaborators with specific complementary expertise to help develop and document the vision. With the goal of the collaboration explicit from the beginning, consensus is implicit when a collaborator accepts the invitation to participate. Hierarchical Collaboration has been the authoring pattern followed to write *The Discipline of Organizing*.

4.2 Collaborative Authoring for *The Discipline of Organizing*

The vision of *The Discipline of Organizing* emerged from the author's experiences teaching a graduate course on Information Organization at the University of California, Berkeley. The idea that organizing is a transdisciplinary concept reflects the author's broad academic and professional experience in information systems design, data and process modeling, cognitive science, and business. The need for this synthesis was reinforced by encounters with practitioners who organize and work with different kinds of resource collections, and with students who converged in the author's Berkeley classroom with more than ten different undergraduate majors.

In 2010, the author proposed to write a book whose conceptual structure reflected the syllabus of his Berkeley course, which by that time contained about thirty lectures, with over a thousand slides of lecture notes. An important goal was to write a book that could be used by other Schools of Information, many of which have roots in library science, while others emphasize digital humanities, informatics, management information systems, business, or user experience design [28]. This goal led to the recruitment of numerous co-authors to write book chapters with the proposal that they would all use the Berkeley lecture notes as starting points to facilitate conceptual continuity. Many of these collaborators were solicited because they had backgrounds and biases that were complementary to those of the author, making them well suited to filling in disciplinary gaps that make the book more balanced and comprehensive. Several of the authors were current or former university professors, and others were graduate students who had worked or would later work in major web firms, web start-ups, consulting organizations, academic and government research labs, and law firms.

This process for collaborative authoring had some critical consequences. The involvement of dozens of co-authors and reviewers with different backgrounds substantially improved the coverage of the book far beyond what a single author could achieve, but it did so by significantly increasing the length of the manuscript. Very often a co-author or reviewer would suggest that some additional content or citation was essential to make the book credible to readers from a particular discipline ("if we don't discuss this, computer scientists won't take this seriously"; "this is part of the information science canon"; "ALA accreditation requires this topic in introductory courses"... and so on).

Initially, chapter authors enthusiastically attempted to incorporate all of the comments, confident that this approach would result in a comprehensive book. But after a while, the relentless desire to create a book that was both broad and deep in disciplinary coverage became tiring to many authors, and the book started to feel more like an encyclopedia than a tightly integrated textbook. Furthermore, some of the collaborators were Berkeley graduate students with appropriate expertise and enthusiasm, but who had finished their graduate work and needed to move on. In retrospect, both of these challenges were predictable given the Hierarchical Collaboration approach; the collaborators had taken on a vision that they had not participated in creating, so there were limits to how hard they would work to achieve it.

Something needed to change, or the book would never be completed.

5. CONTENT ARCHITECTURE

5.1 "Core" and "Supplemental" Content

Many books, especially technical and professional ones, are designed with a "core" stream of content that is augmented by "supplemental" content of various types. The types of supplemental content, the structures that organize it in books, and its presentation and formatting are highly conventional.

Tables, figures, illustrations, and sidebars are often supplemental content, and are usually constrained to appear as close as possible to the core text that mentions them. These types of supplemental content are usually created by the author or by people who are following the author's specifications.

Footnotes, endnotes, annotations, bibliographic citations, glossary entries, and indexes are types of supplemental content that are also closely anchored to particular parts of the core text. Footnotes and annotations are usually constrained to appear on the same page as their text anchor, but the other types of content are more typically arranged at the end of larger text units like chapters or at the end of the book. The author does not typically create some of these types of content, especially indexes.

Appendixes, commentaries, reviews, and case studies are types of supplemental content that are typically associated more coarsely with a book as a whole. People other than the book author also commonly create them, but they do with explicit attention to the core content.

There is no sharp boundary that divides content that can be viewed as supplemental from independently-written content, created without a particular body of core content in mind but which would be useful in understanding or enhancing it. However, content of the latter type is likely to be in sources that are much harder to discover and might have an organization or topical granularity that makes it more difficult to integrate into the core content with any precision. See ([7],[34],[37]) for research and design strategies dealing with these and related challenges in creating "open hypermedia" systems.

5.2 Restructuring TDO into "Core" and "Supplemental" Content

Section 4.2 explained how the goal of multidisciplinary comprehensiveness was undermining the coherence and comprehensibility of the TDO manuscript. At the same time, in late 2011, many of the co-authors had moved on to other jobs and projects, leaving a much smaller authoring group, led by the author of this case study, to finish the book. This gave us an opportunity to rethink and revise the book from end to end, and to devise a new strategy, rather than surrender to the breadth vs. depth challenges.

We decided to restructure the book, to emphasize the transdisciplinary core of the new discipline of organizing, while preserving the disciplinary identity of the concepts, methods, technology, and people that contributed to it. The author of this article edited each chapter to more tightly focus on transdisciplinary content, extracting discipline-specific content into paragraph size chunks, most of which became end-of-chapter notes. Some longer chunks were made into sidebars.

This restructuring identified core content that could stand on its own and provide prerequisite concepts for the discipline-specific content. The conceptual dependences among the topics were carefully teased out, leaving pieces of supplemental content with no mutual dependencies. It is hard to specify precisely how this was accomplished because it was done the old fashioned way, through close reading and analysis of the text, and with careful rewriting to remove dangling conceptual threads. Our familiarity with the text, and knowing the set of disciplines we had sought to incorporate into the book, biased us against trying topic identification or text classification algorithms to identify and classify modules of supplemental content. These techniques have also been used with mixed success by Open Hypermedia and Linked Data researchers [37].

The restructuring effort qualitatively changed the content architecture of the book. Twenty-four percent of the text was moved into about 600 disciplinary-specific chapter endnotes and 100 sidebars. The authors, reviewers, editors, and the publisher all agreed that the book had been substantially improved by the restructuring, and it went to press in early 2013.

5.3 Evolution of the Content Architecture

The plan was for the book to be published in essentially identical print and ebook editions, but the two months required to print, bind, and distribute the former were too long a wait for the ebook editions to stand still. The novel restructuring of the content to create six categories of disciplinary endnotes inspired further refinement of the disciplinary classification. Now there were ten: Library and Information Science, Museums, Archives, Computing, Web, Cognitive Science, Linguistics, Philosophy, Law, and Business.

TDO's transdisciplinarity core with integrated multidisciplinary content enabled it to be adopted quickly as a primary or secondary textbook in a diverse set of university courses in Information Organization, Knowledge Management, Cataloging, Digital Collections, Information Architecture, and Information Systems Design. An enterprising instructor using TDO in an Information Architecture course argued that it would be a better fit if it contained more content that emphasized that perspective, so she wrote an Information Architecture sidebar and about ten endnotes to seed a new category that appeared in the 2nd edition of TDO, published in 2014 ([21],[22]).

The first edition of TDO contained four case studies that used the five design questions for organizing systems (*What is being organized? Why? How much? When? How, or by what means?*) as an outline to enable easy comparison. This design framework quickly became a foundation of many courses that used TDO as a textbook, and the author and a few other instructors have had students write case studies as course assignments. In the second edition, about a dozen of these student-written case studies and the four from the first edition were collected in a new chapter.

6. CONTENT ARCHITECTURE AND THE READING EXPERIENCE

The basic contrast between core and supplemental content is a very old one, and readers somehow decide how much of the latter to read when they encounter it (or references to it) in a book. However, the emergence of digital documents has enabled new ways for users to experience supplemental content.

6.1 Hypertext Links to Transport or Transclude

Selectable links that "transport" the reader to the linked content or that "transclude" the content into the core text stream were

foundational concepts of hypertext proposed by Ted Nelson [41]. The "transport" behavior for hypertext links dominated the earliest implementations of linking capability, presumably because it was usually provided by integrating third-party applications or treating them as loosely-coupled services ([9],[44]). The now familiar idea of web browser "plug-ins" for enabling the integration of new format types was anticipated by Phelps and Wilensky [46], who developed "multivalent documents" and an extensible reading application in which new layers of content and their specialized interactions and behaviors could be overlaid on the "base" layer. Contemporary examples include the Hypothes.is open annotation platform [27], the Lens viewer for scientific publications that allow readers to rearrange and focus on different parts of the article [24], and "semantic publishing" efforts inspired by [49].

 "Fluid Documents" were the first "transclusion" implementation of linking, in which a document containing a link to supplemental content "opens up" and alters its layout and typography to present the content in the context of the core text [54]. We had initially hoped to employ transclusion or overlaying as the mechanism for incorporating supplemental content; readers would be alerted to its presence with a disciplinary-specific symbol in the page margin, and selecting the symbol would seamlessly insert the content into the core text stream, perhaps subtly altering its text formatting or font to remind the reader of its supplemental role.

However, transclusion isn't supported in any existing book reader without custom programming, so we initially chose to rely on more traditional content inclusion mechanisms of link following (and return) and pop-up notes.

6.2 Placement of Supplemental Content: Footnotes, Endnotes, or Pop-ups?

Whether to arrange supplemental content in a printed book as footnotes or endnotes is a contentious design issue. It is shaped by usability considerations, like the relative potential for distracting readers and by marketing ones, like the potential for scaring readers away from books as "too academic" if notes are displayed on the same page as the referring text [23]. In ebooks, implementation considerations bias the design toward endnotes, because it is impossible to preserve the location of footnotes when pagination is recomputed in responses to changes in font size or switching between portrait and landscape modes ([12],[51]).

Displaying supplemental content in a pop-up window would seem to be the inclusion mechanism that would best maintain the reader's context. However, like transclusion, pop-up windows are either not supported by popular ebook readers or the implementation is deficient in some respect. For example, Apple's iBooks reader supports pop-ups, but doesn't allow link traversal from a pop-up note. Since most TDO endnotes contain citations, pop-up notes would become dead ends, leaving endnotes as the only option for links in ebooks that would not require additional software development.

6.3 Selective Inclusion of Supplemental Content

Because the distinction between core and supplemental content is generally not based on disciplinary specificity, we could not predict how it would affect the TDO reading experience. The novelty of this distinction made us fail to see the opportunity to make it explicit in the print edition by appending discipline labels to the superscripts marking the endnote, but we were able to do so in the ebook editions by changing a stylesheet.

An informal survey of students in the fall of 2013 using the print edition of TDO revealed that relatively few of them read the core and supplemental content in one linear pass. Many reported they found flipping back and forth to the end of the chapter to be highly disruptive.

A more systematic survey was conducted in the fall of 2014 to better understand reading preferences and usability issues around core and supplemental content. All of the 35 survey respondents were students using TDO in ebook formats. The survey revealed greater likelihood to read supplemental content in ebooks than with print versions, However, it was somewhat disappointing to learn that about a third of the students never or rarely read endnotes, and even more disappointing to learn that they overwhelmingly said they would have be more likely to read footnotes than endnotes (83% to 6% preference). It appeared that students dealt with their dislike of endnotes by not following links to read them as they encountered them in the running text. Instead, a slight majority of the students who said they generally read the endnotes (55%) said they read the text of a chapter first, and only then read the notes.

It logically seems easier for readers to be selective about supplemental content when it is unnecessary to flip to the end of the chapter to learn the disciplinary focus of the endnote, and the survey confirmed this prediction. A majority of the students (57%) said that the discipline of the supplemental content affected whether they would read it. But if 75% of them said they were more likely to read notes in familiar disciplines than in unfamiliar ones, this meant that 25% of them were more likely to read notes in unfamiliar disciplines.

6.4 Inclusion vs. Exclusion

Framing the discussion of content architecture in terms of core and supplemental content assumes that readers are selectively incorporating additional content to a book. It is interesting to consider starting with the complete book and enabling readers to selectively *exclude* rather than *include* content. A critical difference, if content is included by default, is that the reader can more readily understand the nature of the supplemental content in the book. In Section 8.1 we discuss how visualizations of information about the number and disciplinary-specific endnotes might help readers make better choices about what to read.

7. "DESIGN-TIME" CUSTOMIZATION
7.1 Single-Source Publishing

Many publishers have adopted single-source technologies and methods that are motivated by the idea that different output programs can select, transform, and assemble a set of formats and editions by exploiting markup in a book's source files [48]. Typically, the number of formats and editions produced this way from a content repository can be counted on one hand: a print edition or two, and maybe a few ebook formats.

TDO's collaborating authors were unanimous in wanting to publish the work in both print and ebook formats because they expected it would need frequent revision to stay current. We were fortunate to become beta testers for O'Reilly Media's Atlas single-source publishing environment [3], which enabled us to deliver print-ready copy, epub, and mobi versions of TDO from the same XML source files.

These first editions were essentially identical in content and functionality except for the interactions of search and hyperlinking that are intrinsic to the digital formats. This was expedient, but at times it felt disappointing to have not taken much advantage of the "e" in the TDO ebooks. In addition, many instructors were using only parts of TDO and asked for a simpler and shorter version more suitable for undergraduate courses, which means we needed to produce different editions.

7.2 The Production Line Analogy

It was essential not to lose the advantages of single-sourcing while finding ways to enhance the TDO ebooks and produce multiple editions. Doing this required substantial investments in enriching the markup in the book source files and making the programs that transform them more general and configurable [40].

Consider how an automobile production line can support the assembly of thousands of customized variations of a car model, or how software product line engineering and conditional compilation of source code can produce an appropriate version for any target computing platform and operating system [10].

There are some common challenges in the design and operation of production lines, and for the purposes of this case study it is sufficient to describe them in a simplified and qualitative manner. For serious quantitative treatment of assembly line design, balancing, sequencing, and other optimizations see (16),[47]).

The first challenge is to distinguish the components that are contained in every product or output, typically called the core, base, or platform, from those that vary across products, typically called the features, options, or supplements.

A second common challenge in production lines is organizing the variable components to specify the different products that can be built by selectively combining optional components with the required ones. These different outputs are typically called customizations, versions, or editions For example, a base model automobile can be configured with options to make it more sporty, luxurious, economical, or better adapted for a specialized use like towing a boat. With a large number of optional components, the combinatorial possibilities grow very quickly, so it is essential to organize them in ways that reduce both their actual and perceived complexity.

A third common challenge for production lines is recognizing out of all the possible combinations (the second challenge) those that are most attractive or viable. It is necessary to optimize the assembly process around a mix of the most frequently built products while still maintaining the ability to assemble any combination. The pre-built cars in an auto showroom are the most popular configurations, so if a buyer wants to personalize a car with an unusual set of options, it must be "built-to-order" for him or her.

7.3 The TDO Production Line: Static Configurations of Disciplines

For TDO we met the first production line challenge by taking the conventional and largely implicit distinction between core and supplemental content and making it explicit. We defined a set of discipline categories and incorporated these distinctions in the source markup. We then revised the production and presentation programs to filter and process the text according to these disciplinary attributes in the markup. The selected text is transformed into HTML, arranged and styled as required,

and then zipped into the EPUB archive format used by ebook reading devices and applications.

TDO's supplemental content is classified in eleven disciplines. The TDO "family of books" created by selecting any number from zero to eleven of them thus contains 2048 members. We addressed the second challenge of simplifying the set of configurations by analyzing the course descriptions and syllabi of "Information Organization," "Knowledge Management," and other courses at "Information Schools" where TDO was being used (or could potentially be used) [52]. This work identified four sets of related disciplines to which we could assign meaningful names to make the book a tighter fit to the content requirements for courses. The edition names and the disciplines they contained were as follows:

- *Memory Institutions* Edition (Library & Information Science, Museums, Archives)

- *Informatics* Edition (Computing, Information Architecture, Web, Business, Law)

- *Information Architecture* Edition (Information Architecture, Linguistics, Web)

- *Sensemaking* Edition (Cognitive Science, Linguistics, Philosophy)

Instructors using TDO were receptive to the idea of these tailored editions. Nevertheless, business and marketing considerations overruled this heuristic analysis. Instead, the 2014 revision of TDO was published with just two combinations that define the endpoints of possible disciplinary customization:

- *Professional Edition* [22] (contains supplemental content from all disciplines)

- *Core Concepts Edition* [21] (no discipline-specific supplemental content)

Each edition has found a niche; the Professional Edition is typically used in graduate-level courses, and the Core Concepts is usually chosen in undergraduate courses and by lay readers. We expect to revisit the issue of which TDO editions to publish as pre-selected combinations of discipline-specific supplemental content as more schools adopt the book for a wider range of courses and student populations.

8. "READING-TIME" PERSONALIZATION

It made no business sense to publish over two thousand different editions of a book when many of them would be nearly identical, having all but a few paragraphs in common. Nevertheless, it was dissatisfying to constrain TDO's powerful publishing production line to produce just two fixed-configuration editions because this did not fully exploit the multidisciplinary contributions of the authoring team, nor did it align well with the diversity of contexts in which TDO was being used as a textbook. Even more frustrating was that limiting TDO to just two editions ignored some critical facts that we had learned from teaching experiences and surveys:

- Not all students in a particular course have the same disciplinary backgrounds and interests

- Some students prefer to focus on core content and read little or none of the supplementary content

- These preferences are not fixed; a student reading a book for the first time might focus on the core

content, but might read both core and supplemental content more closely while studying for an exam (or vice versa)

- Most students prefer to read notes in familiar disciplines
- Some students prefer to read notes in unfamiliar disciplines
- TDO's chapters differ in the nature and extent of discipline-specific content.

Taken together, this set of facts argues for some way to give readers a mix of core and supplemental content that is personalized to their disciplinary preferences. We are investigating some complementary ways to meet this goal. The first method is designed and implemented, while the last two are exploratory.

8.1 Active Personalization

The most straightforward way to enable readers to personalize TDO's disciplinary mix is to run essentially the same configuration and transformation machinery in the publishing production line as in the design time case, but to defer the disciplinary filtering step as a choice made by the reader.

However, unless the reader knows what the impact of including or excluding content might be, having the power to do so is not that helpful. To enable readers to make informed choices, we modified the build process to record, for each section of the book, the number of notes and total word count of the notes for each discipline. We use JavaScript to insert a list of check boxes before each section, annotated with this information about the note distribution, and have also developed a variety of bar chart visualizations that convey the same information more efficiently and elegantly. After the reader makes their selections, the ebook dynamically reformats itself by modifying CSS properties for the affected paragraphs.

This active personalization mechanism allows the TDO Professional Edition, that contains all the supplemental content, to dynamically morph itself into any of the 2048 members of the "book family." Unfortunately, because active personalization depends on the reader platform's support for JavaScript and other capabilities, we cannot deploy it to every TDO reading context.

Of the ten ePub readers that we test on, three do not run JavaScript at all and two have an option to disable it. At least two of the readers do not support the presence of HTML Forms elements, such as select menus and radio buttons. Finally, the form elements and visualization occupy significant screen real estate and thus are not well suited for small displays. As a result, the personalization mechanism only works as intended in iBooks, Calibre, Readium, and Adobe Digital Editions, and only on devices with sufficient screen size. On other readers, the ebook adapts itself to the reduced capability of the platform, presenting all of the book content and as much of the personalization functionality as will work and fit.

We acknowledge some inspiration about visualizing the supplemental content space from KnowledgeSea [19], where the saturation of the blue color was used to convey the "depth" of supplemental resources. FeatureCommander [17] used color to distinguish the different software products that could be generated from the same code base using conditional compilation.

8.2 Adaptive Personalization

The active personalization mechanisms in TDO ebooks only work if a person uses them. Because of the rich discipline-specific markup and detailed structure of the multi-level table of contents, it seems promising to apply techniques of adaptive personalization to enable the ebook to dynamically change the content in ways that help the reader ([8],[11],[14],[50]).

Adaptive personalization of a textbook like TDO would be based on an explicit user model that contains information about the reader's background knowledge or disciplinary training, goals, and interests that is incrementally refined by tracking the reader's behavior. The former type of information could usually be obtained directly or indirectly from students when they sign up for a course, and the latter from straightforward logging of events as the reader interacts with the ebook. The user model would know where the reader is in the book, how long the reader has been there, and whether the reader has been there before.

TDO ebooks also contain end-of-chapter quizzes to help students learn, and if these assessments were distributed throughout the ebook, their results would provide additional information about how well readers understood the important concepts. An adaptive or "smart" ebook could make suggestions to visit or revisit sections of the book using these quiz results and could generate new questions as needed [4].

The discipline-specific components in the user model could be the basis for removing links to supplemental content that aggregate tracking revealed were rarely viewed or that didn't appear to help readers understand the core context to which it was linked. Alternatively, to improve the understanding of supplemental content, an adaptive ebook might substitute text using simpler vocabulary or more concrete examples

8.3 Content Discovery and Linked Data

Restructuring TDO's content into core and supplemental components was challenging (Section 5.2), but the process yielded guidelines to enable any instructor or institution to create supplemental content. In particular, the precise anchoring of citations and notes often seen in academic research literature would make selection inclusion of the target content awkward or even impossible, and would similarly make it difficult to interpolate new supplemental content. Instead, TDO relies on coarser anchoring of endnotes, always at the end of sentences, most often at the end of paragraphs, and generally at the end of a section.

Making TDO's supplemental content less tightly coupled than that that in more typical texts has encouraged instructors whose perspectives on organizing complement those of the initial authoring group to write notes and sidebars for new disciplines and for under-developed topics. Notes for data science, bioinformatics, geographic information systems, and digital curation are being developed.

It has been suggested that discipline-specific endnotes in TDO can be viewed as pre-emptive annotation by the authors. Endnotes published in some specific print or ebook edition are obviously different in implementation and presumption of authority, but not different in many other respects from notes created by instructors or readers as sticky notes or margin scribbling in print books, or using analogous ebook capabilities. Several TDO instructors have proposed that their notes and those of their students might usefully be shared with those created by other courses.

The current implementation of our book source repository does not allow content to be added by anyone outside of the core authoring team. However, the potential value in creating a larger and more open community of contributors has inspired us to think about how to implement a distributed authoring and publishing system in which new content could be dynamically discovered and logically included in the family of books. An instructor should be able to teach from a customized edition with local supplemental content, but this local repository would be part of a federated "network textbook" in which content marked as discoverable could be incorporated in any other local edition.

This network textbook would also be an excellent vehicle for sharing the growing collection of case studies being written by students at various universities, including about fifty a year just at Berkeley. Many of the case studies are excellent illustrations of some specific design challenge or method in TDO, and someday a reader will be able to ask "are there any case studies that go here?" and have the relevant ones appear from some distant repository.

Finally, in most of TDO's core content, the source files are marked up with semantic elements that identify people, organizations, locations, products, applications, abbreviations, foreign phrases, and other potentially useful semantic "nuggets" that were mixed into the text. We invested in this semantic markup because we imagined being able to interconnect digital versions of TDO with other semantically described web resources by exposing it according to the conventions for "linked data" ([1], [5]) even if the current technology for ebooks was incapable of enabling it. We are just beginning experiments to determine if this investment pays off.

9. THE INTERRELATIONSHIPS OF BOOK DESIGN, AUTHORING, AND EVOLUTION

9.1 Implications of Collaboration on Book Architecture and Evolution

Open or consensus-driven collaborative authoring approaches impose relatively weak constraints on a book's content architecture. Content often develops organically and separately and is then grafted together into ad hoc structures. Any content is potentially revisable. This method requires little overhead to manage, but autonomous and modular authoring doesn't capture the conceptual dependencies that are essential in a tightly integrated textbook.

An author with a clear vision of a transdisciplinary book will only invite collaborators who will accept this vision of the book's content and organization. The book starts with a strong hierarchical scaffold that resists revision. Content is recursively added, and evolution takes place via extension and annotation "at the leaves of the tree" rather than by aggressive pruning and grafting of branches.

This has the effect of making new editions of the book more backwards compatible with previous editions than books with more unconstrained revision.

9.2 Implications of Book Architecture on Collaboration and Evolution

We restructured TDO into core and supplemental content relatively late in the authoring process as a response to multidisciplinary bloat, and it was often necessary to rearrange and revise paragraphs to preserve syntactic and conceptual continuity.

We concluded that it would have been much easier to write a book with this core + supplement architecture if we had started with this architecture in mind. This of course is conventional wisdom in software engineering; re-factoring is harder than building in modularity from the outset on a more generic platform that is designed to be extended with plug-in components.

The implications are obvious: we need to build a collaborative authoring environment that is designed for modular authoring, extensible semantic classification of supplemental content, and visualization of alternative configuration and sequencing of the core and supplemental components. Some people are already trying to do that [26].

10. ACKNOWLEGMENTS

11. REFERENCES
[1] Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. Elsevier.

[2] ASIST 2014 Award Winners (https://www.asis.org/awards/2014_winners.html)

[3] Atlas. (https://atlas.oreilly.com/)

[4] Beer, W., & Wagner, A. "Smart books: adding context-awareness and interaction to electronic books." *Proceedings of the 9th International Conference on Advances in Mobile Computing and Multimedia*. ACM, 2011

[5] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International journal on semantic web and information systems*, 5(3), 1-22.

[6] Booksprints. (http://www.booksprints.net/about/)

[7] Brusilovsky, P. Adaptive hypermedia for education and training. *Adaptive technologies for training and education* 46 (2012).

[8] Brusilovsky, P., & Millán, E. User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*. Springer-Verlag, 2007.

[9] Carr, L., Hall, W., & De Roure, D. (1999). The evolution of hypertext link services. *ACM Computing Surveys (CSUR)*, *31*(4es), 9.

[10] Couto, M. V., Valente, M. T., & Figueiredo, E. (2011, March). Extracting software product lines: A case study using conditional compilation. In *Software Maintenance and Reengineering (CSMR)*, 2011 15th European Conference on (pp. 191-200). IEEE.

[11] Dahn, I., & Schwabe, G. (2002, January). Personalizing textbooks with slicing technologies-concept, tools, architecture, collaborative use. In *System Sciences,*

2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on (pp. 10-pp). IEEE.

[12] Dawson, A., & Wallis, J. (2005). Twenty issues in e-book creation. *Against the Grain*, *17*(1), 18-26.

[13] Donofrio, N., Sanchez, C., & Spohrer, J. Collaborative Innovation and Service Systems: Implications for Institutions and Disciplines (2010)., in D. Grasso (Ed.), *Holistic Engineering Education*, Springer.

[14] Durlach, P., & Lesgold, A. (Eds.). *Adaptive technologies for training and education*. Cambridge University Press, 2012

[15] Emigh, W., & Herring, S. C. (2005, January). Collaborative authoring on the web: A genre analysis of online encyclopedias. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 99a-99a). IEEE.

[16] Erel, E, & Sarin, S. "A survey of the assembly line balancing procedures." *Production Planning & Control* 9, no. 5 (1998): 414-434.

[17] Feigenspan, J., Papendieck, P., Kästner, C., Frisch, M., & Dachselt, R. "FeatureCommander: colorful# ifdef world." In *Proceedings of the 15th International Software Product Line Conference*, Volume 2, p. 48. ACM, 2011.

[18] Forte, A., & Lampe, C. (2013). Defining, understanding, and supporting open collaboration lessons from the literature. *American Behavioral Scientist*, *57*(5), 535-547.

[19] Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., & Coyle, M. (2007, January). Collecting community wisdom: integrating social search & social navigation. In *Proceedings of the 12th international conference on Intelligent user interfaces* (pp. 52-61). ACM.

[20] Glushko, R. (Editor). 2013. *The Discipline of Organizing*, MIT Press.

[21] Glushko, R. (Editor). 2014. *The Discipline of Organizing: Core Concepts Edition*, O'Reilly Media.

[22] Glushko, R. (Editor). 2014. *The Discipline of Organizing: Professional Edition*, O'Reilly Media.

[23] Grafton, A. (1999). *The footnote: A curious history*. Harvard University Press.

[24] Grubisoc, I., et al. eLife Lens. (http://lens.elifesciences.org/about/#info/all)

[25] Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 0002764212469365.

[26] Hitchcock, M., MacFarland, I., & Son, P. (2015) *ReBook*. Unpublished Master's Degree Final Project Report. School of Information, University of California, Berkeley. (http://www.ischool.berkeley.edu/files/student_projects/rebook_finalreport.pdf)

[27] Hypothes.is - Annotate with Anyone, Anywhere. (https://hypothes.is/)

[28] iSchools - Leading and Promoting the Information Field. (http://ischools.org/)

[29] Kästner, C., Apel, S., & Kuhlemann, M. (2008, May). Granularity in software product lines. In *Proceedings of the 30th international conference on Software engineering* (pp. 311-320). ACM

[30] Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American psychologist*, *39*(10), 1123.

[31] Kittur, A. & Kraut, R. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *Proc.CSCW08,* ACM Press (2008).

[32] Klein, J. T. (2010). A taxonomy of interdisciplinarity. In R. Frodeman, J. T. Klein & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 15-30). Oxford: Oxford University Press.

[33] Kolfschoten, G. L., Lowry, P. B., Dean, D. L., de Vreede, G. J., & Briggs, R. O. (2013). Patterns in Collaboration. *Collaboration Systems: Concept, Value, and Use*, 83.

[34] Korhonen, A., Naps, T., Boisvert, C., Crescenzi, P., Karavirta, V., Mannila, L., ... & Shaffer, C. A. (2013). Requirements and design strategies for open source interactive computer science eBooks. In *Proceedings of the ITiCSE working group reports conference on Innovation and technology in computer science education-working group reports* (pp. 53-72). ACM.

[35] Krueger, C. (2002). Easing the transition to software mass customization. In *Software Product-Family Engineering* (pp. 282-293). Springer Berlin Heidelberg

[36] Laniado, D., & Tasso, R. (2011, June). Co-authorship 2.0: Patterns of collaboration in Wikipedia. In *Proceedings of the 22nd ACM conference on Hypertext and Hypermedia* (pp. 201-210). ACM.

[37] Levacher, K., Lawless, S., & Wade, V. (2012). Slicepedia: providing customized reuse of open-web resources for adaptive hypermedia. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 23-32). ACM.

[38] Lowry, P. B., Curtis, A., & Lowry, M. R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *Journal of Business Communication*, 41(1), 66-99.

[39] Lovink, G., Tkacz, N., Reagle, J. M., O'Sullivan, D., Liang, L., Salah, A. A., ... & Chen, S. L. (2012). *Critical point of view: a Wikipedia reader.*

[40] Maloney, M., Glushko, R.,& Milowski, R.A. (2015) Using DocBook to Produce a Polyvalent Academic Work", *XML Prague.* (http://archive.xmlprague.cz/2015/files/xmlprague-2015-proceedings.pdf)

[41] Nelson, T. H. (1965). Complex information processing: a file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th national conference* (pp. 84-100). ACM.

[42] Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200-5205.

[43] Newman, J., & Newman, R. "Three modes of collaborative authoring." In *Computers and Writing*, pp. 20-28. Springer Netherlands, 1992.

[44] Pearl, A. (1989). Sun's Link Service: a protocol for open linking. In *Proceedings of the second annual ACM conference on Hypertext* (pp. 137-146). ACM.

[45] Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1), 88-113.

[46] Phelps, T. A., & Wilensky, R. 2000. Multivalent documents. *Communications of the ACM*, *43*(6), 82-90.

[47] Rekiek, B, Dolgui, A., Delchambre, A., & Bratcu, A.. "State of art of optimization methods for assembly line design." *Annual Reviews in Control* 26, no. 2 (2002): 163-174

[48] Rockley, A. (2001). The impact of single sourcing and technology. *Technical communication*, *48*(2), 189-193.

[49] Shotton, D., et al. "Adventures in semantic publishing: exemplar semantic enhancements of a research article." *PLoS computational biology* 5.4 (2009): e1000361.

[50] Steichen, B., Ashman, H., & Wade, V. (2012). A comparative survey of Personalised Information Retrieval and Adaptive Hypermedia techniques. *Information Processing & Management*, *48*(4), 698-724.

[51] Stiff, P. (1997). 'A footnote kicks him': How books make readers work. *Journal of Scholarly Publishing*, *28*, 65-73.

[52] The Discipline of Organizing - Who's Using the Book? (http://disciplineoforganizing.org/whos-using-the-book/)

[53] Turoff, M., & Hiltz, S. (1982). Computer support for group versus individual decisions. *Communications, IEEE Transactions on* 30, no. 1 (1982): 82-91.

[54] Zellweger, P. T., Chang, B. W., & Mackinlay, J. D. (1998, May). Fluid links for informed and incremental link transitions. In *Proceedings of the ninth ACM conference on Hypertext and Hypermedia* (pp. 50-57). ACM.