Building a Digital Library the Commons-based Peer Production Way

[Aaron Krowne](#)
Virginia Tech
<akrowne@vt.edu>

Abstract

This article defines the digital library setting as it relates to commons-based peer production (CBPP) [1]. Motivations for selecting the CBPP method in this setting will be discussed, and the challenges of CBPP will be described. The Noosphere system will be presented as a case study to demonstrate CBPP digital library system design. Specific aspects addressed include: how an "economy of ideas" is the basis for productive activity in Noosphere, how logical integration of content is performed, how opportunistic updating is attained, what services Noosphere provides to foster community and provide for social integration, and what could be done to improve the system. Also discussed are different ways to benefit from commons-based peer production in digital libraries.

Introduction

In this article I will be presenting a new way to build digital libraries using commons-based peer production (CBPP), including (and emphasizing) the creation of content to fill the digital library. In CBPP, the ultimate goal is to produce an intellectual work [Benkler]. For us, the "intellectual work" (alternately called an "information" or "knowledge" work [2]) is a digital library.

This is not without precedent. Wikipedia is essentially a very successful CBPP digital library that has not generally been recognized as such [Wikipedia]. My own project, PlanetMath [Krowne et al., 2001], is another example, and I will use it (and the Noosphere system that it runs) as a case study here.

In addition to being a presentation of proof of concept for applying commons-based peer production methodology to digital libraries, this article also demonstrates an approach to digital library sustainability—by fostering participatory sustainability through CBPP. In addition, this article argues that commons-based peer production fits

into a broader range of situations than were previously identified or thought possible. Finally, in this article I suggest that the Noosphere system helps expand the applicability of CBPP.

Digital Library Goals

Before we can understand how to apply commons-based peer production to the digital library setting, we must know what is meant by the "digital library setting". Thus, we must think about what our unique intents and goals are in constructing a digital library. The basic, universal goals of digital libraries are to provide a logically organized, conveniently accessible, and (if possible) easily *actionable* collection of digitized *knowledge* in some field or fields for an audience of *learners*.

Breaking this statement down, by "actionable" I mean usable or applicable. Expert support (from a community) and software tools come into play here. By "learners" I mean learners in the widest possible sense, including all age groups and inside or outside a professional or formal setting. My use of the specific phrase "knowledge in some field" means that the records of the collection are *interrelated* and trace out the "known world" of the field of concern.

The goals in the digital library setting are thus distinct from those of other CBPP projects, such as the Linux kernel, which exists to produce a free UNIX-like operating system core, or Kuro5hin, which provides a forum for article-writing and debate on technology and culture issues [Kuro5hin]. The reader should keep these differences in mind as we proceed.

Motivation

There has been something of a "holy war" in the past few years between the advocates of open source vs. corporate modes of software production. This is fundamentally a backlash to the emergence and prominence of the software-sector brand of CBPP (which has produced Linux, for example). The claims leveled by those suspicious of the CBPP movement are that it is chaotic and irresponsible, and therefore untrustworthy and apt to output works of low quality. Transferring these concerns to the digital library setting, the operative question is this: why should we build a digital library in a seemingly unchecked and unregulated community-driven fashion? Beyond simply asserting that much of this sentiment is a result of "fear, uncertainty, and doubt", I will provide a number of motivations for CBPP in digital libraries that answer the question posed.

There are quite a few motivations that may come into play to nudge us toward a CBPP-based solution for building a digital library. One is that the digital library may be for a niche field for which it is hard to enlist critical mass to produce content. That is, all of the experts are too busy to give up large portions of their time to write about their field in the desired formats. The quantity of existing content, produced through means such as academic publishing, may be quite small and beg complementation, may be out of

date, or may simply be very esoteric.

On the other hand, the problem may be that there is very little money available to hire talent, despite there being a willing pool of contributors. Another motivation may be the desire for freedom from oppressive intellectual property regimes, something hard to attain when dealing with bureaucratic organizations. A spirit of democratic camaraderie may also come into play, coupled with a general wariness of "moneyed" situations when it comes to the intangible and "noble" ideals of knowledge, teaching and learning.

Finally, and perhaps most importantly, the motivation may be the desire to tap a willing base of knowledgeable experts who *could not be completely reached* by a highly structured, centralized effort. The reasons for this may be cultural, geographical, or philosophical. Adding to the difficulty in coordinating a productive effort, the target community's knowledge and expertise may not be evenly distributed. There may be no single person or small set of persons to hire to produce content that provides complete coverage. In this case, it makes sense to solicit bits and pieces from the widest possible subset of the community, and assemble and integrate the result into a single collection. Table 1 sums up the motivations discussed in this section. I have grouped them into philosophical, logistical, fiscal, and optimal categories. The first three categories concern serving needs or solving problems. The last category is concerned with improving on traditional forms of production—using CBPP to go "above and beyond" the status quo.

Table 1: Motivations for employing commons-based peer production to build a digital library. Only a few of these need apply for CBPP to be beneficial.

Category	Motivations
Philosophical	Spirit of camaraderie, democracy, altruism; aversion to hierarchy/command; rejection of IP regimes; dislike of mixing money and learning.
Logistical	Knowledge distributed unevenly and/or widely; inflexibility of centralized effort; experts too busy doing work to commit entirely; DL builders not experts
Fiscal	No or low funding; inability to provide major coverage using works for hire.
Optimal	More material; more peer review; more up-to-date content; greater involvement of readers/learners; new treatments.

Most of these factors came into play in my own situation with the PlanetMath.org site.

well into the second year of the project. In the field of concern of the library (mathematics), the expertise and knowledge was indeed distributed unevenly and widely. An additional factor was, even as the lead in building the library system itself, I knew I was not the best qualified person to produce or appraise all of the content (or might not even be qualified to determine who is the best person). In sum, the project could not have been organized any way other than through commons-based peer production.

Challenges

The challenges of commons-based peer production and how these challenges must influence digital library system design for a successful result are described next.

Logical integration

When building a digital library through commons-based peer production, the first challenge involves "logical" integration. That is, how does one integrate into a whole the disparate contributions taken in? These contributions could be of varying size and contain a varying number of "links" to the rest of the content base in the digital library (at the extreme, containing no links). Benkler identified the important problem of logical integration [Benkler], and it is a problem that must be solved in order for a cohesive work to be produced by a particular community.

Social or political integration

In addition to logical integration, I'd like to add what I call "social" (or "political") integration [3]. That is, how does one integrate contributions from contributors with a diversity of motivations, experiences, opinions, and values? These human differences can surface as disagreements regarding content, particularly in terms of methodology, organization, selection among alternative conventions and, at the extreme, philosophical disagreement in contentious areas of a field [4]. In the firm-based (or "cathedral") setting, these problems are typically solved by corporate policies, procedures, and hierarchy—in a word, authority. This method is not necessarily a bad way to organize, but CBPP requires an appropriate alternative.

Preserving continuity of content

Another challenge is preserving the continuity of content despite the voluntary nature of the contributors. When contributors' efforts are voluntary, this tends to remove the ability to rely on them to "stick with" the project, or even just to maintain their own contributions. In a setting where content is always evolving in response to critical feedback from the community itself, enough absent contributors can cause breakdowns. In other words, we want to avoid having "stale" items present in the digital library due to absentee authors or maintainers.

Updating the collection

The CBPP digital library must be attuned to the work schedules of tens or hundreds or even thousands of contributors simultaneously. What is the best way to update the state of such a collection, modified continuously and unpredictably? This problem also relates to the voluntary nature of the contributions; when contributors are "doing the project a favor" by participating, the project must be able to accommodate them whenever they are motivated and have free time. The general strategy for achieving this is to operate in an asynchronous manner, putting low temporal demands on participants, and opportunistically updating the "whole product" to which all are contributing. However, the details of how to do this can be nontrivial, as we will see.

Minimal administrative load

The final challenge is doing all of the above with minimal administrative load. This is particularly important when there is low or no spare fiscal capital for staff, as administrators are staff, and staff costs money. If there can be no administrators, who will integrate the contributions, mediate disputes between contributors, re-assign stale portions of the collection, and handle collection updates? The answer lies in smart system design.

The Noosphere System

Noosphere (/no-oh-sfeer/) [Krowne et al., 2002] is a system that addresses the above challenges of commons-based peer production of digital libraries. The Noosphere system grew out of the PlanetMath project and serves as the project's software platform. Thus, Noosphere is geared towards some of the particulars of the digital library niche to which PlanetMath belongs.

Noosphere chiefly features online, full-text content in relatively small-sized units. The basic unit of content in Noosphere is the entry, which any registered user can create. The entries comprise the main section of the system, which is called the "encyclopedia". This reflects the general orientation and pedagogical style of the system. In addition to the encyclopedia, Noosphere supports papers, expositions, and e-books.

Noosphere entries consist of title, content (text discussion/explanation), a type, a classification, a list of synonyms of the title, a list of additional concepts defined, and various other metadata. The entries are interlinked, which means that the text of each entry contains hyperlinks pointing to other entries where appropriate. The general intent of this is to provide definitions for each concept utilized, in an easily navigable fashion. Entries are written in L^AT_EX [Lamport et al.], which serves as the basis for Noosphere's mathematics support in addition to allowing for the expression of general document formatting. Displayed in *rendered* form, the mathematical portions of each entry "look right" with a standard browser (with no plug-ins), a considerable improvement over most other attempts to publish mathematics to the web to date. This mathematics support makes Noosphere a good candidate for use in all of the mathematical sciences.

A key feature of Noosphere is the *corrections* system. If any registered user determines there is a problem with an entry, he or she can voice concern by filing a correction to that entry. Until addressed, this correction is displayed when the entry is shown, ensuring that the critique is "out in the open". Finally, each entry in Noosphere has an *owner*, who is initially the person who created the entry. The owner is "in charge" of the entry's maintenance. In more concrete terms, owners are the absolute and final authorities over changes to their entries. This seemingly rigid arrangement is subject to important modifications, exceptions, and extensions, as we shall see.

Economy of Ideas

The name for the Noosphere system (as I use it) comes from the word "noosphere", as employed by Eric Raymond in "Homesteading the Noosphere" [Raymond, 1998b]. Raymond used the word "noosphere" to mean something akin to a "space" of knowledge and ideas. However, it is important to note that this is a *socially shared* space, in which a Lockean notion of property rights has considerable import. Noosphere has drawn heavily upon such concepts, as well as the concepts of democracy, anarchy, capital, currency, common law, and natural law to create a self-regulating "economy of ideas". In the following discussion, I will point out where these notions from economic and political theory apply.

Ownership is only the beginning of how authority manifests in the Noosphere system to address the challenges of commons-based peer production. While ownership is first assigned to the creator of an entry (strongly reminiscent of the "homesteading" method of acquiring property), the owner must adequately maintain the "property" of his or her entry to retain ownership (reminiscent of common law property rules, i.e., "use it or lose it"). This rule manifests in one way through the corrections system: if a correction filed to an entry is pending for too long, it becomes an *outstanding* correction. At this point, ownership can be taken up by any other interested user (called *adopting*). If the entry is not adopted, then at a later point, the entry reverts to being owned by no one (this is called *orphaning*) and can be adopted by anyone except the previous owner. After orphaning, the entry resides in an *orphanage* (naturally), where attention can be drawn to the fact that the entry needs better stewardship.

These are the basics of Noosphere's authority system and how it provides for continuity of content maintenance. It works because the entry, through pride and/or concern for reputation and/or altruistic affect, has value to its owner. In this sense the entry behaves as private property, which in general is maintained because it is valued by its owner. In Noosphere, when this motivation lapses for whatever reason, so does ownership, and therefore control.

There are many more important elements of Noosphere's authority system that are part of the "economy of ideas". For instance, entries can be *transferred* between parties at any time, as long as both parties agree to the exchange (reminiscent of fungibility of capital). In addition, owners can delegate control over their entries through the Access Control List (ACL) system. ACLs can be used to add co-authors, create authoring

groups, or even make an entry world-editable (in full Wiki fashion). This ability to fine-tune authority gives users of Noosphere greater self-determination. (More details about authority models in Noosphere, as well as an empirical study, can be found in "Authority Models for Collaborative Authoring" [Krowne and Bazaz, 2003].)

The final element of the picture is the *scoring* system. Noosphere users receive points for creating entries, revising entries, filing corrections, adopting entries, and more. The number of points awarded for various actions is configurable by the digital library deployer and is intended to track the level of value contributed to the community by its members (reminiscent of monetary currency). Through the scoring system, users can build reputations, since users' scores serve as easily comprehensible summaries of their accomplishments and expertise. The concept of scoring in a productive virtual community has also been discussed in an article by Kelly, Sung, and Farnham [Kelly et al., 2002].

Absent from the above presentation is any discussion of how the creative process is regulated by administrative staff. This is because there is no such regulation in Noosphere, an ideal I call *zero content administration*. (This "administration" is distinct from administration in the systems-support sense.) The design of Noosphere is meant to foster emergent coordination in an *anarchistic* fashion. Here I use "anarchistic" not in the sense meaning "chaos", but rather the more formal sense in which the meaning is "absence of governance". No "outside" governing of the content creation, vetting, or revising process is needed, as Lockean property notions and common-law conventions are the "natural law" of the Noosphere system. This is not to say that the anarchistic route is the only way to have emergent coordination (kuro5hin.org is a more directly democratic example), but the anarchistic route has worked well for PlanetMath.

Automatic Linking

As previously mentioned, Noosphere entries are interlinked, providing a means for the reader to find expositions of (ideally) all concepts utilized in any entry. This is not in itself a new idea—all Wiki software supports hyperlinkage between entries, and the Wikipedia community has done an excellent job of approaching the ideal of linking to all concepts within the collection.

However, there are some drawbacks with the Wiki approach. In Wiki, the author manually creates each link. The first problem with this is in linking from a new entry to existing entries in the corpus. That is, how does one know which concepts utilized in the entry are present in the collection and need to be cited? A search must be conducted to find the answer to this question. The second problem is that of linking from the corpus to a new entry. That is, *already-written* entries may have to be updated with a link if they cite a concept in the new entry. These tasks, which can be somewhat mitigated by search systems and the distributed authorship model of Wiki, are still on the order of the size of the corpus for each entry added.

In Noosphere I sought to eliminate this logical integration task entirely, thus lowering

the barriers to successful content production even further. This has been largely realized in Noosphere's automatic linking system. Whereas Wiki solves the logical integration problem by distributed labor, Noosphere solves it by automation, reserving the distributed labor for the generation of the core content. This labor savings allows one to participate in building the digital library with less of an investment of time.

Noosphere's automatic linking system works as follows. First it scans the text of each entry upon rendering, looking for *concept labels* (titles, synonyms, and "defines" metadata) from other entries. When found, these labels are automatically turned into hyperlinks to the source entries. Aiding this process in the face of semantic ambiguity (i.e., homonyms in the collection) is the use of entry subject classifications. When classifications are not present, a novel method of citation graph-walking is used to infer a most appropriate class. Further, the central chained-hash concept index gives the automatic linking process a time complexity based only on the size of the entry being hyperlinked, not on the size of the collection. This means the algorithm is scalable with collection growth.

How a new entry is processed by the automatic linking system is only half of the story, however. We still have not seen how the problem of updating the *rest* of the corpus to link *to* the new entry is addressed. To accomplish this, an inverted index of the words in each entry is maintained. When a new entry is added or the concept label metadata for an existing entry is modified, instances of those concept labels in other entries are found in the inverted index and are used to mark these entries for link analysis later. (This is done through the cache invalidation system described in the next section.) Further details about the automatic linking system can be found in *An Architecture for Collaborative Math and Science Digital Libraries* [Krowne, 2003].

Evaluation of the effectiveness of this system has shown that it has 100% recall (that is, all terms that should be linked are linked), and 85% precision (that is, some terms are senses of homonyms that are not in the collection and thus should not be linked). An extension to this system, based on the addition of *linking policies* is expected to increase linking precision to about 95% [Krowne, 2003]. These linking policies will not undermine the original goal of not requiring attention to linking in the vast majority of cases. Rather, only a minority of "trouble" entries are expected to require linking policies for the entire collection to benefit. In the event of troublesome cases requiring manual attention, Noosphere sites can still benefit from the same distribution of labor as is found in Wiki, due to Noosphere's collaborative nature.

Opportunistic Updating

I have mentioned rendering of Noosphere entries above, but only as a black box. In fact, rendering of an entry consumes major resources and had to be addressed within the core architecture of Noosphere. Preprocessing, plus automatic linking, plus L^{A}T$_E$X compilation, plus output method processing [5], plus post-processing, all add up to a significant delay in rendering an entry (from a few seconds to tens of seconds for a large entry).

Yet despite the delays inherent in rendering, we must handle uncoordinated updates to the collection. Rather than collect updates and then compile the entire corpus "offline" periodically, Noosphere provides an "opportunistic" cached-rendering system that takes in all committed changes and "instantly" integrates them with the entire corpus. This provides immediate feedback for both writers and readers.

The cached-rendering system is built around a database of entry status, which records whether or not an entry is "valid". An invalid entry must be re-rendered before it is next viewed. When rendered, the entry in rendered state is cached to disk. When requested, it is then served up instantaneously in "static" fashion. When someone commits changes to the entry or when the automatic linking system invalidates the cached entry, the next request to view it will trigger re-rendering. Thus, occasionally, a reader will have to wait for an entry to render before they can see it. However, this architecture guarantees that no entry can ever be viewed in an out-of-date state.

Helping avoid numerous re-rendering delays when readers browse the collection is a background rendering thread. This triggers re-rendering of invalid entries continuously and allows us to put the server to work constantly, using every "spare" CPU cycle to make entry loading faster.

Community

Although Noosphere's core anarchistic and "propertied" model solves some of the social integration problems, others demand interaction *outside* the core content development features. Therefore, Noosphere has a number of other services that provide direct community support.

1. **The requests service.** This service provides a global "to-do" list of content addition for the Noosphere site. Besides simply being a bookkeeping convenience, requests help to alleviate "writer's block" in those who would like to contribute but aren't sure what to add. Requests also make it easy for those who are mostly readers to contribute in a valuable way, simply by giving feedback about what they want to see. This helps to keep the content growing with fresh ideas while automatically being useful to many people.

2. **The discussion service.** This service provides threaded asynchronous messaging. Noosphere's discussions are applied in a manner that I call *ubiquitous forums*. This means that "every object is a forum" [6]. In other words, a discussion can be attached to most of the core objects of Noosphere. This includes encyclopedia entries, corrections, requests, site news items, and the cataloged items (papers, expositions, and e-books). Ubiquitous forums reflect the realization that the best place to discuss an artifact is "at" that artifact. Thus, the discussion is *anchored* to the object. With ubiquitous forums, freeform discussion about an entry or topical questions intended for its owner (the expert) can simply be posted to the entry. Clarifications or debate on the worthiness of a correction can be posted to that correction. Commentary on references or

additional ideas for a request can be posted to that request, and so forth. One can see from these examples how ubiquitous forums can help aid the social integration of the community, by providing a convenient means for communication-in-context.

3. **The notification service.** Noosphere's notification system keeps members of the community aware of activity *relevant to them*, which it does by way of integration with the corrections and the discussions systems. Corrections to an entry result in a notice to the entry's owner. Complementing this, a resolved correction results in a notice to the filer, indicating what action was taken and why. Similarly, replies to a message posted result in a notice that makes the initial poster aware of the reply.

   An important part of the notification system is the ability to create configurable watches. Watches placed on any object by any user result in notices about events to that object being sent to the user. These objects could be encyclopedia entries, requests, or even individual discussion threads. Some examples of events are "message posted", "correction filed" (or closed), or "request fulfilled". Watches are, in fact, the basis for most forms of notification on Noosphere, with watches, by default, being placed on each object a user creates. In addition to the ability to deactivate any individual watch, this default behavior can be changed by the user in their Noosphere preferences.

   Noosphere notices come in both e-mail and Noosphere system "inbox" (web) flavors. Currently, correction notices default to coming via email, but users can change this in their preferences. All Noosphere notices triggered by watches currently come through the Noosphere system inbox, though generalization is planned for the future.

Looking Ahead

As evidenced by the collection growth chart in Figure 1, PlanetMath (and hence Noosphere) has been a successful implementation of CBPP.
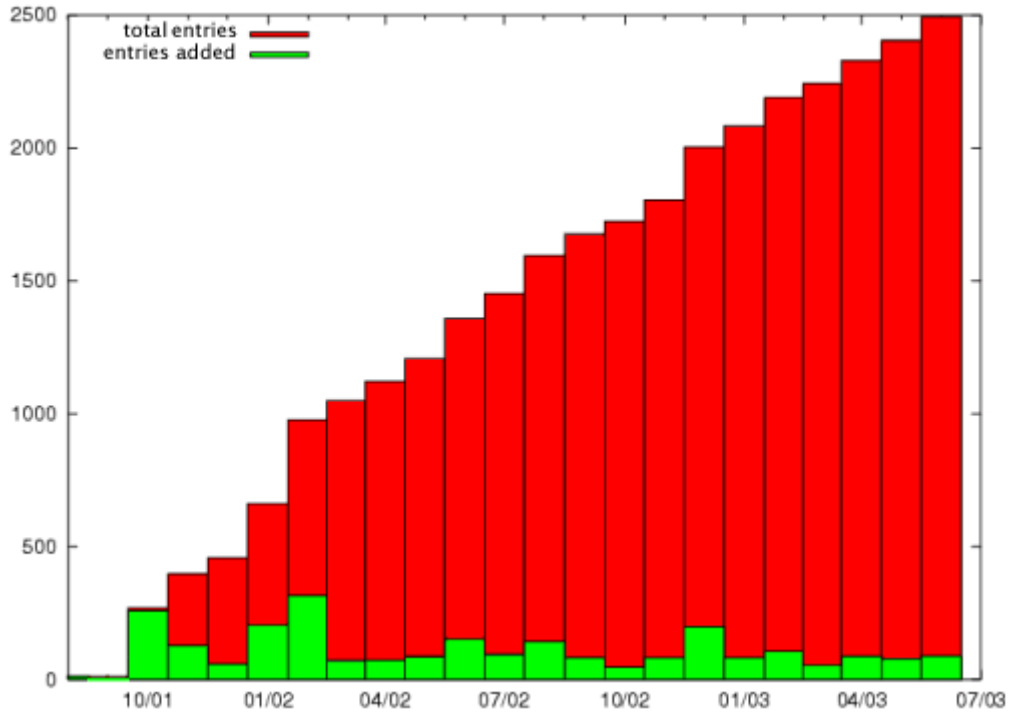
Figure 1: Increase in numbers of items in the PlanetMath collection.

However, in addition to achieving a successful implementation, a legitimate concern for any digital library effort is its sustainability. Aside from fiscal and organizational sustainability, this includes *participatory* sustainability [Krowne, 2003]. Participatory sustainability is the element of sustainability that comes from continued and sufficient participation in the digital library effort by its patrons. The need to focus on this element is particularly acute in a commons-based peer production effort, where the community makes or breaks the effort.

When the challenges of commons-based peer production are addressed well by a system employing it (as I believe Noosphere does), participatory sustainability comes naturally. However, Noosphere is not perfect, and there are still things that could be done to make it more sustainable. For example, there is nothing stopping the PlanetMath community from providing "competing" alternate entries on the same topic. This could arise in situations where two contributors cannot come to an agreement on the methodology or content of an entry. In such a situation, it may even be best to present both alternatives. The problem, then, is what the learner should do when confronted with the choice between two different entries.

A system for representing content quality would help solve this problem. This could take the form of a ratings system, whereby votes on a scale (say one to five) would be averaged to form the overall quality value. Prominently displayed, such a value would help steer the learner toward the best place to start. A challenge for any such system in a setting where the content is always evolving is how to incorporate change into quality metrics. In other words, ratings are somewhat (or possibly entirely) undermined by

subsequent changes to the entry being rated. One possible solution to this problem would be to employ a quality algorithm that weights outdated ratings less than new ones (possibly taking into account the extent of changes).

Another way to expand Noosphere that would greatly assist the community and increase the active user base would be to support syntaxes simpler than full L^AT_EX for authoring. Both Wikipedia and MathWiki employ a L^AT_EX -Wiki hybrid syntax for authoring, which utilizes Wiki syntax for document formatting, and L^AT_EX syntax just for equations and mathematical expressions. Support for such a syntax would make both users and content from these systems more "portable" to Noosphere.

Ways to Use CBPP

Besides simply building a new digital library from scratch, there are other ways to benefit from commons-based peer production in the digital library setting. CBPP could be used to augment an existing digital library, for instance, by adding a glossary or encyclopedia to it. In addition, users could be permitted to add, critique, correct, augment or translate metadata records. Commons-based correction of metadata has been explored in the CiteSeer system [Lawrence et al.] (though, in a way that requires significant administrator burden). To address concerns about mixing "official" and "unofficial" content, commons-produced areas can be "sandboxed" or distinguished via appropriate additions to the metadata.

Federation is another way to benefit from CBPP digital libraries. For example, the computer science subset of PlanetMath's material is harvested by the Computing and Information Technology Interactive Digital Educational Library (CITIDEL) [Fox et al.] using Open Archives [Lagoze et al.]. This is enabled through the use of Open Archives sets, which are in turn implemented in PlanetMath using record categorization metadata provided through the commons-based process. PlanetMath's inclusion in the National Science Digital Library [NSDL], a federation of over one hundred digital libraries, illustrates an openness to including CBPP-type projects into digital library federation efforts, and is an acknowledgement of the worth and quality possible in CBPP digital libraries.

It is also likely that CBPP can even augment digital libraries where a large percentage of the content is produced as works-for-hire. CBPP simply allows us to extend the "content-capturing" net further, picking up contributors who are self-motivated. For example, Howstuffworks.com [Brain et al.] contains a significant percentage (about 10%) of entries that are donated, despite the lack of solicitation of such donations [Brain, 2003]. This percentage could perhaps be grown or better supported through the establishment of a CBPP infrastructure.

Conclusion

Using Noosphere is not the only way to coordinate a commons-based peer production effort for digital libraries. Whereas Noosphere is "anarchistic", "democratic"

alternatives could also be built. However, Noosphere's model of coordination *is* novel for collaborative systems and could be applied more widely.

I hope the reader has enjoyed this introduction to commons-based peer production for digital libraries and is perhaps motivated to join, begin, or adapt an existing project to include some of the ideas and methods presented here.

Acknowledgements

Notes

[1] The term "commons-based peer production" (CBPP) was introduced by Yochai Benkler in his theoretically grounded explanation [Benkler] of a recently recognized Internet-based phenomenon. This phenomenon includes the production of the Linux kernel by a worldwide and shifting team of volunteers, as well as web sites like Kuro5hin [Kuro5hin] and Wikipedia [Wikipedia]. The defining characteristic of CBPP is the voluntary and community-regulated production of an intellectual work. Benkler notes that the rise in CBPP is because the Internet lowers certain communication and collaboration barriers, allowing CBPP to flourish and to serve as a viable alternative to produce a large and complex intellectual work. Eric Raymond also observed (and participated in) CBPP at an earlier time, but called it the "bazaar model", which he contrasted with the "cathedral model" of traditional production [Raymond, 1998a]. In Benkler's economically grounded exposition, the latter would be called "firm-based production". Benkler also discusses the open market as another vehicle of production, contrasting it with firm-based and ultimately commons-based production.

[2] Benkler uses the term "information work", which is more or less necessitated by the generality of his treatment. I prefer the term "knowledge" or "intellectual" work to "information work" for the digital library setting, but I consider all three the same for

the purposes of this article.

[3] Note that the social/logical distinction is mine, not Benkler's. Benkler generally uses "integration" to mean what I identify here as the logical sort, that is, the combination of parts of the intellectual work into a whole.

[4] One can see why I suggest "political" as a synonym for this type of integration. The sense of integration I discuss here has to do more with contention and disagreement *within* a single society than addressing differences *between* societies.

[5] Entries can be rendered either as HTML-and-images (default, fastest), page images (highest quality), or syntax-highlighted TEX source (a key openness and learning feature).

[6] In fact, in Noosphere dedicated forums are just "empty" objects that have a long-running topical discussion attached.

Bibliography

[Benkler] Benkler, Y. (2002) Coase's Penguin, or Linux and the Nature of the Firm. *Yale Law Journal*, vol. 112. <http://www.benkler.org/CoasesPenguin.html>.

[Brain 2003] Brain, Marshall (creator, Howstuffworks.com). Private conversation, Sep. 10, 2003.

[Brain et al.] Brain, M. et al. Howstuffworks.com, <http://www.howstuffworks.com>.

[Fox et al.] Fox, E. A., et al. CITIDEL: Computing and Information Technology Interactive Digital Educational Library <http://www.citidel.org/>.

[Kuro5hin] Kuro5hin.org: Technology and culture, from the trenches. <http://kuro5hin.org>.

[Kelly et al.] Kelly, S. U., Sung, C., Farnham, S. Design for Improved Social Responsibility, User Participation and Content in On-Line Communities. *ACM SIG CHI*, 2002.

[Krowne, 2003] Krowne, A. P. (2003) *An Architecture for Collaborative Math and Science Digital Libraries* (MS thesis) Virginia Tech Department of Computer Science, Blacksburg, VA.

[Krowne and Bazaz] Krowne, A., Bazaz, Anil. (2003) Authority Models for Collaborative Authoring. (in review)

[Krowne et al., 2001] Krowne, A. P., et al. (2001) PlanetMath: math for the people, by the people. <http://planetmath.org>.

[Krowne et al., 2002] Krowne, A. P., et al. (2002) the Noosphere Commons-Based Digital Library System, <http://planetmath.org/noosphere/>.

[Lagoze et al.] Lagoze, C., et al. the Open Archives Initiative <http://www.openarchives.org/>.

[Lamport et al.] Lamport, L., et al. (1994) L^AT_EX: A document preparation system. <http://www.latex-project.org>.

[Lawrence et al.] Lawrence, S., Bollacker, K., Giles, C. L. CiteSeer: Scientific Literature Digital Library. <http://citeseer.nj.nec.com>.

[NSDL] The National Science Digital Library (NSDL), <http://www.nsdl.org>.

[Raymond, 1998a] Eric S. Raymond. The Cathedral and the Bazaar. *First Monday online journal*. Volume 3 Number 3. 1998. <http://www.firstmonday.dk/issues/issue3_3/raymond/index.html>.

[Raymond, 1998b] Eric S. Raymond. Homesteading the Noosphere. *First Monday online journal*. Volume 3 Number 10. 1998. <http://www.firstmonday.dk/issues/issue3_10/raymond/index.html>.

[Wikipedia] Wikipedia: The Free Encyclopedia. <http://www.wikipedia.org/wiki/Main_Page>.