

The Stanford Encyclopedia of Philosophy: A Developed Dynamic Reference Work

Colin Allen*
Philosophy Department
Texas A&M University

Uri Nodelman†
Computer Science Department
Stanford University

Edward N. Zalta‡
Center for the Study of Language and Information
Stanford University

1 Introduction

A fundamental problem faced by the general public and the members of an academic discipline in the information age is how to find the most authoritative, comprehensive, and up-to-date information about an important topic. The present information explosion is the source of this problem—more ideas than ever before are being published in print, on CD-ROM, and in a variety of forms on the Internet. One can nowadays use library search engines and web-indexing engines to generate lists of publications and websites about a topic and then access them immediately if they are

*Professor of Philosophy at Texas A&M University, Principal Programmer and Associate Editor of the Stanford Encyclopedia of Philosophy

†Graduate Student in Computer Science at Stanford University, Associate Programmer and Assistant Editor of the Stanford Encyclopedia of Philosophy

‡Senior Research Scholar at Stanford University, Project Director and Principal Editor of the Stanford Encyclopedia of Philosophy

online. But even limited area search engines can produce thousands of matches to keywords and even with new interface tools to narrow the search, one is typically confronted with a list that is not informed by human judgment. If one wants an introduction to a topic that is organized by an expert, if one wants a summary of the current state of research, or if one wants a bibliography of print and online works that has been filtered on the basis of informed human judgment, there are few places to turn. One might try a standard reference work, but the main problem with reference works is that they quickly go out of date (even before they are published) and don't reflect the latest advances in research. So the following questions arise: How can an academic discipline maintain a reference work which introduces the significant topics in the field (for those who wish to learn the basics), but which tracks, evaluates, and changes in response to new publications and new research being presented in a variety of media (for those with advanced knowledge on a given topic)? How can this be done so that access to the reference work is low-cost, if not free?

Members of our project started thinking about these questions in 1995, and in order to answer them, we developed and implemented the concept of a 'dynamic reference work' (DRW). A DRW is much more than a web-based encyclopedia. The most important features of a DRW are that: (1) it provides the authors (who may be scattered in universities all over the world) with electronic access to their entries, so that they can update those entries at any time to reflect advances in research, (2) it provides the subject editors (wherever they are located) with administrative access to those entries and updates, by which they can referee them prior to publication (and by which they can add new topics, commission new authors, etc.), and (3) it provides automated tools by which a principal editor can oversee administrative control of (1) and (2) with only a small staff. Thus, on our conception, a DRW includes a highly customized work-flow system by which the members of an entire discipline are empowered to collaboratively write *and maintain* a refereed resource. Such a resource would not only introduce traditional topics in the discipline, but would also track the (new) ideas that are constantly being published on those topics in a variety of media. With this concept of a DRW, all sorts of new and interesting questions arise concerning how to best design, program, and administer such a resource and work-flow system.

No electronic journal or preprint exchange in the sciences or human-

ities approaches this concept in scope. Electronic journals: (1) typically do not update the articles they publish, (2) do not aim to publish articles on a comprehensive set of topics, but rather, for the most part, publish articles that are arbitrarily submitted by the members of the profession, (3) typically serve a narrow audience of specialists, and (4) do not have to deal with the *asynchronous* activity of updating, refereeing, and tracking separate deadlines for entries, since they are published on a *synchronized* schedule. Preprint exchanges not only exhibit features (1), (2), and (3), but also do not referee their publications and so need not incorporate a work-flow system that handles the asynchronous referee process that occurs between upload and publication in a DRW. None of this is to say that electronic journals and preprint exchanges have a faulty design, but rather that a DRW is a distinctive new kind of publication that represents a new digital library concept.

Although commercial publishers have built web-based reference works and claim that they are dynamic, they lack some of the principal design features of a DRW, namely, (1) that authors should have electronic access to copies of their entries and be able to modify them, and (2) that subject editors and the principal editor should have electronic access to the encyclopedia databases and unrefereed entries, so that they can directly carry on the task of adding and commissioning new entries, refereeing entries and updates, etc. These commercial publishers typically don't give academics accounts on their computers, or access to their databases. Instead, the authors and editors must provide/referee content by first interacting with the staff of the publishing house (managing editors, copy editors, computer web specialists, computer markup specialists and others) before changes to the encyclopedia can be made public. On our model, however, the *publishing house* becomes inessential to the process of maintaining a DRW. Academics have direct electronic access to the entries, and can engage and manage the process of writing, refereeing, and updating entries without intermediaries.

Our implementation of a DRW is embodied by the Stanford Encyclopedia of Philosophy (SEP) <<http://plato.stanford.edu/>>. In the remainder of this paper, we document this particular DRW and then discuss some of the outstanding questions and problems it faces.

2 The Implementation of a Dynamic Reference Work

The SEP first came online in September 1995 with 2 entries! Since then, we have designed a workflow system which attempts to maximize efficiency among those involved in its production. The most important parts of this system are the password-protected web interfaces to the central server, which can be accessed by any author, subject editor, or the principal editors from any where in the world there is a computer with an internet connection.¹

The web interface for authors allows them to download our HTML templates, to upload their new entries into a private area of our web server, and to remotely edit copies of their entries stored in this private area. So if an author is lecturing outside her university and encounters a reader of her entry who points out an error or omission, she can sit down at the next net-connected computer (possibly at an Internet cafe), contact the Stanford server using the machine's web browser and, after supplying her ID and password, remotely edit the content of her piece and submit it for editorial review. The web interface for subject editors allows them to enter new topics, commission authors for those topics, referee and comment on entries and updates submitted for review, and communicate their decisions to the editor. So, for example, if a subject editor is visiting another university and learns by email that an entry has

¹These web interfaces, and the file download and file upload capacities which they enable, are the principal enhancements we've made to the SEP since the publication of the paper 'A Solution to the Problem of Updating Encyclopedias', by E. Hammer and E. Zalta, in *Computers and the Humanities*, **31**/1 (1997): 47-60. When that paper was published, the SEP still used an ftp-based file-upload system. We gave authors system accounts on our Unix server, linked their home directories into webspace, and allowed authors to transfer their files by ftp to our server. However, subsequent to the 1997 paper, when browser-based file upload had become a widely adopted and supported standard, we switched to the new technology. Authors and subject editors no longer needed system accounts on our Unix server, and indeed we determined that maintaining Unix accounts for all participants would introduce problems of scale when dealing with hundreds of accounts. Furthermore, we improved security on our machine by deleting those accounts. Instead, authors were given passwords for the browser-based file-uploads. Moreover, subsequent to the 1997 paper, we distinguished a private 'upload-space' (which includes 'revision space') from our public 'web-space'. The former contains private copies of the entries accessible only to authenticated users so that newly uploaded entries, and newly revised entries, do not become publicly viewable until after they have passed through the referee process.

been revised and submitted for review (see the discussion of our tracking and reminder system below), she can use a web browser to log onto the subject editors web interface, display the original and revised versions of the entry side-by-side *with the differences highlighted*, easily determine where the changes are located, referee them, and then accept or reject revised version.

The principal editor also has a special, secure web interface, by which this collaborative process is administered. The principal editor can easily add people to the project, add entries to the database, assign editorial control for entries to the subject editors, issue invitations, track deadlines (for new entries and for updates), and publish entries and updates when they are ready. Many of these things can be done with just the press of a few electronic buttons. For example, when a subject editor submits (through her web interface) a suggestion to commission an author on a particular topic, the suggestion gets entered into a database, and the principal editor is notified and prompted to log into his web interface. He simply hits the New Invitation button, selects the entry in question, and is then prompted to invite the person listed in the database for that entry by hitting the Invite button.

Finally, we should mention that we have designed and implemented a web interface for *prospective* authors. When a prospective author receives an invitation, they are directed to log in to a special web interface to obtain information about the project, to set up an account with us if they plan to accept, and to set a deadline of up to a year for completing the entry (or else write to us with a counterproposal).

These 'front-end' web interfaces supply data to the 'back-end' processing programs and databases in our system. In particular, actions taken, and information entered, by authors, editors, and prospects are communicated to our tracking and logging system. This system can identify the state of any given entry, recognize who now owes work on an entry and which deadlines have or haven't been met, and pass this information to our automated email reminder system, which has recently been developed, initialized, and put into continuous operation. When an entry changes state and another person must now act to continue the publication process, the reminder system will prompt this person about what needs to be done and by when. It will continue to send reminders (on a fixed, inoffensive schedule) until the work is done (or notify the principal editor that that all reminders have been ignored and that human

intervention needs to take place). Finally, when any entry or substantive revision is published, the entry is scheduled for revision within 3-5 years (depending on how swiftly the field moves). Actually, some authors update once a year, but all authors are notified by our reminder system well in advance of any scheduled revision.

The use of these web and computer technologies offers considerable savings over more traditional publishing methods, and has enabled us to develop the Encyclopedia with a small staff and budget. The importance of this project for our profession cannot be overstated. As new ideas in logic, ethics, political philosophy, philosophy of science, philosophy of cognitive science, etc., are published in books and journals of philosophy, both in print and on the web, the SEP provides a rational and efficient system by which the new information is assimilated, digested, and disseminated in entries which are *responsive* to new research.

Here is a basic quantitative analysis of the effectiveness of our design which is justified by the above. Consider first the fact that there was a 30-year gap between philosophy encyclopedias (the Macmillan Encyclopedia was published in 1967, the Routledge Encyclopedia in 1998). So there was no up-to-date encyclopedia for at least 25 years (9125 days). By contrast, a typical Encyclopedia author is regularly visiting the library to read journals or receiving new journal issues at her office. As soon as she realizes that a recently published article advances the topic of her Encyclopedia entry (and, in principle, this could be the day that an article is published, and in some cases, she might even have advance knowledge of the publication if the author has sent her a preprint), she can use her computer to call up the Encyclopedia server and modify her entry accordingly (maybe by adding a paragraph and a Bibliographic item). The next day, assuming that the change is a substantive one, the relevant subject editor(s) will be notified that the revision must be refereed. Suppose it takes a week to referee the minor changes to her entry. Then we have reduced the length of time required for the update process from about 9,000 days to about 9 days, or by 3 orders of magnitude. Even if it were to take up to a year for a new idea (in a book, say) to become reflected in the Encyclopedia after the new idea is published, that would still constitute a 25-fold decrease in the length of time it takes for a philosophy reference work to reflect the advance.

We should mention two other features of the SEP which should be part of any DRW. The first is the fact that authors are encouraged to

write nested, as opposed to linear, documents. That is, we encourage our authors to put highly technical, scholarly, or highly detailed information into supplementary documents and to link these into the main part of the entry. These supplementary documents can have supplements as well, and so forth, and so the reader can then choose the level of detail they wish to explore. Such nested entries become useful to a wider range of readers — intelligent undergraduates should be able to get through the main entry by skipping the links to the supplements, while graduate students and colleagues may skip the basics and follow the links to the supplementary documents, to find the cutting edge material.

The second noteworthy feature concerns archiving. For purposes of citation, a DRW is a moving target, since the entries are always being corrected, updated, improved, etc. It is difficult to cite such a moving target. For example, a reader might quote a passage from a DRW entry in a research article, and after publishing the research article, discover that the author of the DRW entry has altered the passage in question. To solve this problem, we make quarterly archives of the SEP. On the equinoxes and solstices, we make an electronic copy or ‘snapshot’ of the entire encyclopedia as it exists on that day and link that complete copy into our special Archives page. We explain to users that the proper way to cite an SEP entry is to cite the most recent archived version. These archived versions will not be updated or changed in any way, and so scholars can rest assured that the passages they quote will be available for scholarly purposes. Note that every entry in the SEP contains a section called Other Internet Resources which contains links to offsite web-based material and these links may eventually break in the archived entries (especially if the links do not point to similarly archived material). That is a danger of the web. But we do attempt to minimize the problem, however. We have designed and programmed a ‘link-rot’ detection system which automatically notifies the authors anytime links break in the dynamic versions of their SEP entries. The authors are asked to revise or delete the link.

3 Statistics about the SEP as a Dynamic Reference Work

As of September 21, 2001, the SEP had 213 entries online. We had 69 subject editors overseeing 513 authors currently working on a total of 600

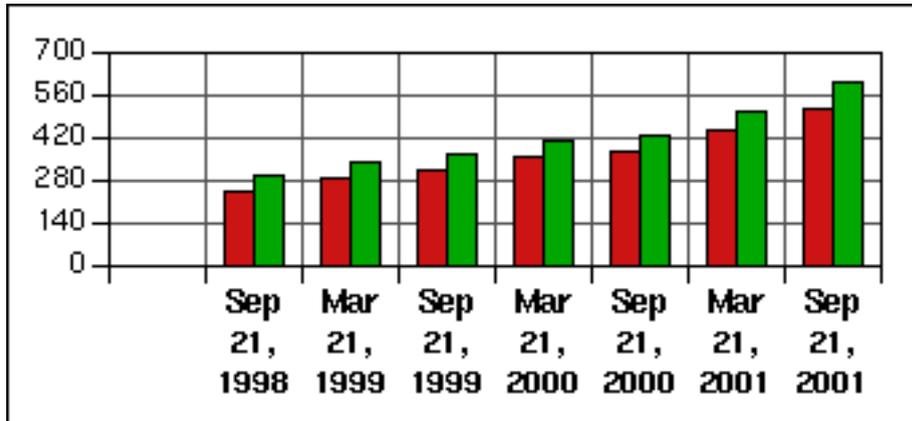


Figure 1: Number of Authors/Commissioned Entries

commissioned entries.

Over 10% of our entries make use of some hierarchical document structure (i.e., they involve more documents than simply a main text and a footnotes page) and just under half of our entries have been updated since they were first published.

The rate at which we commission entries has increased by a factor of 3, from about 5 per month in 1999 and 2000 to about 15 per month in 2001. See Figure 1. During the same period, our publishing rate has increased by a factor of 6, from about 1.5 entries per month in 1999 to 9 entries per month in 2001. See Figure 2. The average length of our entries has also increased from approximately 6800 words per entry in September 1998 to 8900 words per entry currently.² We estimate that, in print, the current version of the SEP would fill over 3000 pages.³

Between September 1997 and September 2001, the content of the SEP has grown from about 3 megabytes to about 26 megabytes. See Figure 3. During that same period our average accesses per month has increased by an order of magnitude, from about 5000 to 57000. See Figure 4.

²These word counts are slightly inflated due to the presence of HTML tags in the text. We estimate that the tags add about 300 to 500 words per entry.

³This estimate is based on an assumption of 600 words per page.

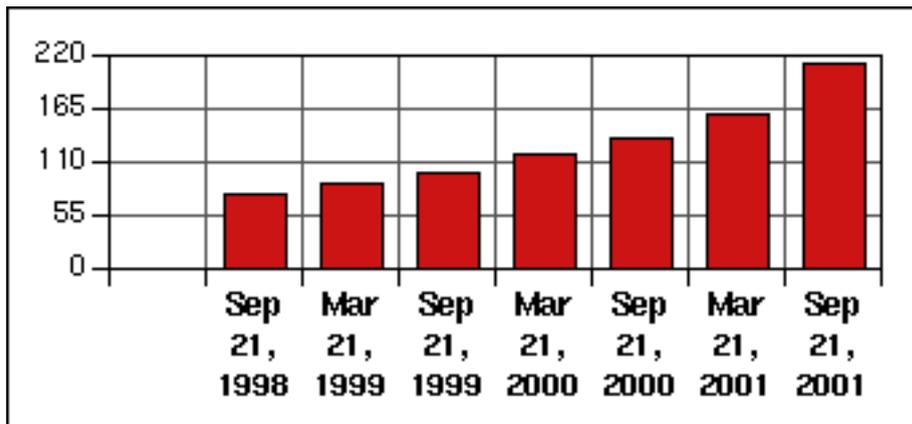


Figure 2: Number of Entries Online

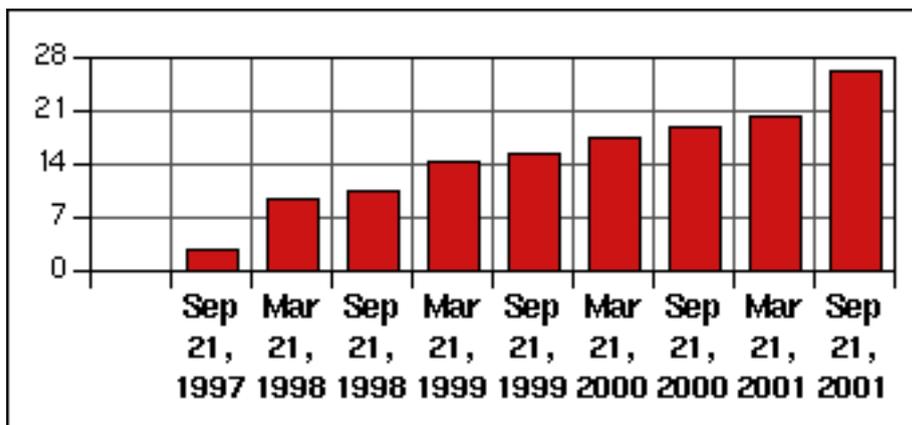


Figure 3: Content in Megabytes

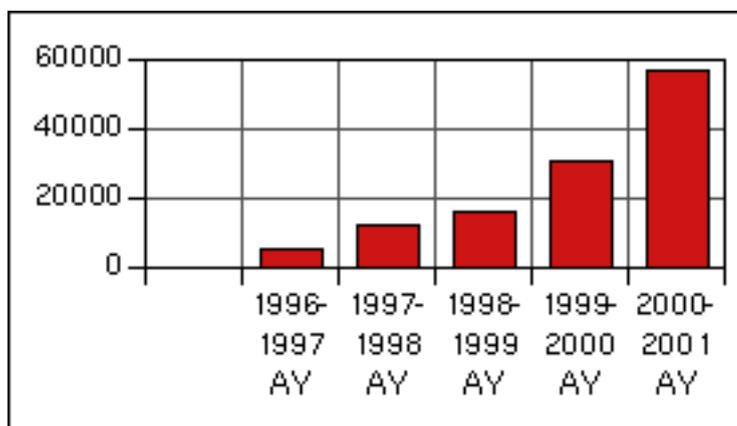


Figure 4: Weekly Average Number of Users
(Excluding Mirror Sites)

4 Why the SEP (and other DRWs) Should Be Free

After 6 years of operation and numerous discussions with authors and subject editors involved with our project, with faculty members and colleagues around the world, with publishers, and with university librarians, we have come to the conclusion that if the means are possible, the SEP should try to remain a free resource. One may think of this as an idealistic goal, but there are several problems associated with a subscription-based, access-restricted models for funding the SEP. We discuss some of these problems in what follows.

The first problem arises from the fact that academics who author entries in scholarly reference works traditionally receive a fee for their efforts. They are, after all, providing a service to the publisher. However, authors of SEP entries are volunteering their time. There are various reasons why they do this. One might be the fact that they will reach a large number of readers. (As long as the SEP remains free or low-cost, it will have a large readership.) Another might be the intellectual obligation academics might feel they have to contribute to the profession and world at large by playing a role in maintaining an up-to-date resource in philosophy. Another might be the prestige they might acquire should they become known widely as an expert on a certain topic, or by becoming

associated with a Stanford University project. And, of course, authors may be motivated by the fact that they can list these entries as invited publications on their curriculum vitae, for the latter play an important role in promotion, hiring, and tenure.

This volunteer arrangement, however, might become compromised if the SEP were required to charge subscriptions in the attempt to make a profit. Even if the SEP charged subscriptions at a rate that simply recovers costs, authors might suggest both that we should recognize their efforts as part of that cost and that we should therefore increase prices a little so as to collect enough revenue to distribute royalties to the authors. The situation is complicated, moreover, by the fact that authors might *additionally* argue that maintaining an entry is also a service. In any year they update their entry, they might be owed remuneration. As we scale up to 1000+ authors, this problem becomes even more acute. And a final wrinkle is the fact that our authors currently provide us with entries formatted in HTML and nearly ‘web-ready’. If we charged a subscription, if only to recover costs, authors might claim that we are *offloading our typesetting costs* onto them. It would be best to sidestep all of these questions by finding a model on which the SEP could remain a free resource and our authors never feel that they are being exploited.

As second, somewhat similar, problem concerns our subject editors. These subject editors are constantly suggesting new entries, commissioning entries, and refereeing entries and updates. If we required subscriptions for access to the SEP, if only to recover costs, the subject editors might argue that their services should be recognized as part of those costs, and that we should increase our subscription rates to pay them accordingly. This problem is magnified by the fact that the work involved in maintaining a dynamic reference work continues from year to year, with no fixed endpoint, and it is complicated by the fact that some subject editors have higher workloads than others.

A third problem concerns the difficulty of finding a subscription model that everyone can live with. University libraries and other institutions are reluctant to pay for something the rest of the world can get for free. Consequently, it would be difficult to implement a model in which these institutions would pay a subscription fee while everyone else would be allowed free access. So a subscription model might eventually force us to require *everyone* to subscribe (albeit, at appropriately proportional rates).

A fourth problem is the fact that many deserving groups of people would be disenfranchised by a subscription model. Small colleges, public libraries, and K-12 school libraries usually can not afford even a modest subscription fee. Moreover, people who are accessing the web from home over an ISP seem reluctant to pay for content, much less for a subscription to a *philosophy* encyclopedia. And last, but not least, our users in the developing parts of the world would become disenfranchised; this includes university students and colleagues in developing countries, as well as lay persons in those countries. Since the SEP has been accessed by users in over 150 countries around the world, academics associated with the project as authors would lose a significant base of readers.

A fifth problem concerns search engines. When people search Google, Yahoo, etc., for philosophical topics, our pages figure prominently in the results. That is because our pages are free to everyone, and web-indexing spiders can access and index our pages. But on a subscription model, our server would restrict access to those who pay. This would make it nearly impossible for search engines to index our site, and consequently people wouldn't find the SEP pages when they conduct web searches. Even if some arrangement could be made to allow indexing spiders to index our site, access restrictions would make the links returned by a search engine useless to the majority of web users, and make it difficult and pointless for non-paying users and institutions to create links to our site on their own pages. But without those thousands of links to the entries on our site, the prominence of our pages in the results of certain web search engines (e.g., Google) would greatly diminish, since those search engines often prioritize the websites which are returned as matches to keyword searches by the number of cross-referencing links to that website which exist on the web.

A sixth problem concerns our mirror sites. Currently, the SEP has mirror sites (U. of Sydney, U. of Amsterdam, and, soon, U. of Leeds) which perform the following functions. (1) They guarantee our users access to the Encyclopedia pages when the Stanford server temporarily goes down. (If anything happens to our server, our readers use a mirror site until we fix the problem and reboot.) (2) These mirror sites give our users in other parts of the world faster access to our pages. (3) The mirror sites provide very important layers of digital preservation for the SEP pages by keeping complete copies of our data. These institutions completely underwrite the costs of the mirror sites and provide this service for free. But this arrangement would come to an end if we started charging

a subscription for our site. On a subscription model, either we would have to pay the mirror sites for their costs in manpower, equipment, and overhead (assuming that some monetary arrangement could be made), or we would have to face the much more expensive proposition of running ‘mission critical’ servers that provide fast, world-wide access 24 hours of every day. (If institutions have to pay for the SEP, they would expect a level of service at least comparable that which they now enjoy.)

A seventh problem is the fact that if the SEP were forced to rely on cost-recovery to survive, the decisions which are now made *in-part* on the basis of the interests of the profession may have to be made instead on the basis of cost recovery *alone*. This might have a negative impact on the quality of the Encyclopedia, such as placing strict word limitations on the length of entries and bibliographies (to save disk space or stay within bandwidth limitations), banner advertising, links to online booksellers, etc. The latter, for example, would compromise the integrity of the Encyclopedia, since a user might wonder whether the link to an online bookseller is present because of the merits of the book or because the online bookseller was kicking-back some of the profits to the publisher.

In addition to the above seven problems that would arise if the SEP adopted a access-restricted, subscription-based cost-recovery model, we believe that there is a positive reason for remaining free, namely, that it would be an outstanding legacy for the SEP and profession as a whole if it could provide both academics and non-academics around the world with a free resource by which they could satisfy their intellectual curiosity from an authoritative source on philosophical questions of all kinds and, in particular, those concerning the human condition.

5 Costs of the SEP

Of course, it does cost money to produce the SEP and if it is to remain free, those costs will somehow have to be underwritten. Currently, the SEP costs are underwritten by a grant from the NSF, which funds the SEP from October 2000 to September 2003.⁴

⁴The NSF grant (#IIS-9981549) was made possible by a significant financial contribution from the NEH. The NEH previously funded the Encyclopedia project from September 1998 to August 2000 (#PA-23167-98). From September 1995 to August 1998, the SEP was funded by seed money from the Center for the Study of Language and Information (Stanford University), where the project was conceived and developed.

The NSF grant pays for a Principal Editor (working 50% time), a Consultant Perl Programmer (working 33% time), an Associate Programmer (working 20% time), an Assistant Programmer (working 10% time) and an Administrative Assistant (working 10% time).⁵ Clearly, this adds up to a tiny staff, since the percentages total to one person at 100% employment and one at 23% employment. We believe our accomplishments are significant when viewed in light of our tiny staff, though it should be mentioned that these accomplishments were made possible by the fact that our personnel regularly put in longer hours than official records indicate. We are currently seeking grant money to hire a business consultant to determine exactly what costs are required to run/staff the SEP and what funding models are available to ensure that the SEP's long-term survival.

We believe that the SEP could evolve into an even greater publication (i.e., with even higher quality entries, with fewer typographical and other errors, etc.) if it had more adequate staffing. It does not strike us as unreasonable to think that the SEP should, for the long term, require a Principal Editor at 50% time, an administrative assistant at 50% time, a programmer at 50% time, and an HTML (XML) copy-editor at 50% time. With an administrative assistant and HTML copy-editor working at these levels, the Principal Editor would be relieved of certain routine tasks and could concentrate more on content issues, on supporting subject editors for those subjects of the SEP which still aren't very far along, etc. Similarly, with a programmer working half-time, we could re-engineer many of the compromises we made in designing and programming our workflow system, we could adapt the SEP to new technologies that improve web-based publication, etc. Although our estimate of these long-term staffing requirements for the SEP will need to be subjected to a strict business analysis, it should give the reader some idea of the money for salary income that the SEP will need. Once money for overhead costs (including hardware and software) are included in the equation, it should be clear that the SEP will have to find sources of income if it hopes to remain a free publication after the NSF grant expires.

CSLI still contributes some cost-sharing funds under the terms of the NSF grant.

⁵After we complete the task of programming the essential workflow-systems underlying the SEP, there will be less programming time required. However, if we are to keep up with changes in technology, programming time will always be necessary.

6 Future Challenges

Clearly then, the principal future challenge for the SEP is to find a source of income by which it can continue to be universally accessible. In this concluding section, we discuss this challenge, along with two others.

6.1 Funding Models

We will, of course, attempt to raise an endowment. Indeed, we plan to prepare grant proposals to foundations, asking them to give us money to hire a fund-raiser for a year, at full or half-time. But we anticipate that it could take up to 10 years to raise an endowment large enough to cover our operating costs.⁶

Our present focus, however, is to explore funding models intermediate between universal free access and full-fledged subscription models. For example, we plan to study the following two models, both of which might be adopted without incurring all of the problems outlined in Section 4.

1. *Voluntary Archive-Acquisition Program.* On this income-producing plan, the SEP would remain free to everyone, but institutions which subscribe to this program would be entitled to download and store our 4 quarterly archives each year they subscribe. (Those who choose not to pay would still be able to access all of our pages.) There would be 2 advantages to subscribing: (a) the institutions which subscribe would be entitled to serve their copies of the archives whenever our server and its mirror sites are down, and (b) should the SEP project ever cease to exist, these institutions would own, and be able to locally circulate, copies of our archives even though our servers are no longer active. For an extra fee, we could enhance this service by burning and distributing yearly CDs of the archives for that year.
2. *Archive-Access Program.* On this income-producing plan, *only* those institutions/users who pay a yearly subscription fee would be able to access our archives for that year. The SEP's dynamic entries (which are always changing) would still be available free to everyone. However, since citation can take place only to fixed, archived versions of

⁶Remember that if endowments are managed properly and make 8-10% a year, then 5 of the 8-10% would be returned to the SEP and the remaining 3-5 of the 8-10% would be reinvested.

the entries, the motivation for joining this program would be clear to scholars and librarians. On this plan, we could again create the following levels of service: (a) for a basic yearly fee, users at subscribing institutions would receive web-based access to the archives on our server, (b) for a higher fee, the institution would obtain the right to download and store the archives, and (c) for an even higher, but still moderate, fee, we would burn and distribute CDs to the subscribers.

We plan to investigate these and other models by which income could be raised. They strike us as offering the best hope of raising money in a way that addresses the problems discussed in Section 4.

6.2 Institutional Home

The question of funding models is connected to the question, where is the best institutional home for the Encyclopedia? We believe that the SEP is better off in an academic environment than in the hands of a commercial or even non-profit publisher/publishing house. Normally there would be three good reasons for joining forces with a non-profit, university-based publisher, namely, that such a publisher would: (1) offer expertise in the business of producing publishable material, (2) offer a more stable institutional home, and (3) have the mechanism for marketing and collecting revenues for the materials it produces. We consider these in turn.

Expertise?

We are not aware of any non-profit publishers which have the kind of expertise which already resides with the SEP project. Since 1995, we have been perfecting our model for dynamic reference works which can be run on a low-cost basis. Grants from the NEH and NSF for 1998-2003 have given us the financial resources to program a workflow system which would compare, on the open market, to \$200,000 – \$300,000 off-the-shelf ‘web content management’ software systems. The Stanford Encyclopedia of Philosophy has been successful to date as a project designed, programmed, and run by academics (not specialists) who have acquired the tools necessary to make use of the power that computers, the Internet, and web-based technologies provide. In particular, the most highly technical positions required by our project, Unix system administration and

Perl/CGI programming, are filled by *academics* working part-time.

A More Stable Institutional Home?

Currently, the SEP is published at Stanford University's Center for the Study of Language and Information (CSLI), and the Stanford Philosophy Department serves as its Advisory Board. Our analysis suggests that the best institutional home for the SEP is an academic setting of this kind. Numerous reasons for this are readily apparent. An institution such as CSLI or an academic philosophy department would have a more intimate and direct concern for the academic excellence of the project and for safeguarding a resource for the profession. Such a concern might not be shared by a non-profit, university-based publisher, since traditionally they let titles go out of print. Moreover, academic setting of a philosophy department or research institute has a concern for educating its graduate students and graduate students could play various and vital roles in the SEP project.

Many graduate students in philosophy have a background in mathematics and computer science. They arrive at graduate school with enough knowledge about Unix, web servers, etc., to work on or consult for our project. Here are two ways in which these students can provide a steady stream of innovative ideas for the future of the project:

1. *Paid, part-time positions on the SEP staff.* Graduate students would make excellent part-time staff. As such, they could (a) help subject editors plan and commission entries (thereby becoming known to distinguished members of the profession outside their home institution), (b) acquire and use HTML or XML skills and help the Encyclopedia with content mark-up, (c) acquire and use their knowledge of Unix and web-servers to help administer the project (thereby preparing them to use those skills in their later academic life), and (d) work as office staff, handling correspondence with the authors and relieving the Principal Editor of routine tasks in the administration of the Encyclopedia.
2. *Part of the graduate curriculum.* A philosophy department, without exploiting the graduate students in any way, could create a one-hour/week proseminar for all first year graduate students. Each week, the students would be required to read and report on one

article from the SEP in their field or in a related field. In this way, graduate students would enhance their breadth and analytic skills as philosophers, improve their writing skills by focusing on whether entries are well-written from a pedagogical point of view, and suggest ways to improve and/or update entries. They could bring their talents with web-based searching to identify whether any related material on the web should be linked into the entries they consider.

Finally, it is important to note that a research institute such as CSLI can be an important collaborator in this project. For instance, CSLI has researchers with expertise in linguistics, computation, data-mining, etc. It also has a highly-skilled technical staff, which can deal with any technical issues (such as those connected with server operation, backup, etc.) that might go beyond the expertise of a typical academic department.

Marketing and Income Collection?

Of course, if we can find a way to keep the SEP a free resource, then the fact that a publisher offers marketing and income collection becomes moot. However, as we have seen, we may be forced to adopt an operating model intermediate between that of keeping the Encyclopedia free and requiring universal subscription. Such models would require some marketing and income collection.

Basic marketing would not be problematic for the Philosophy Department and CSLI. When our NSF grant runs out in 2003, the SEP will have been on the Web as a *free* resource for 8 years. During our first 6 years, we have become well-known throughout the philosophical world (both academic and non-academic) and many thousands of individuals know about, read, and rely upon our pages. If our service were to be shut-off or restricted in some way, libraries and other institutions would certainly hear about it from their constituents. Moreover, most of our 500+ authors and 70 subject editors would notify the libraries at their own institutions. In addition, the fact that the SEP is a free resource means that there are thousands (if not tens of thousands) of links to our web pages. If we were to start charging subscriptions for access, these links would all end up in an *advertisement to subscribe*, as soon as anyone followed the link.

Income collection could also be easily accomplished, and could be done by using a university-based eCommerce center, such as the one at Stanford

University.⁷ These eCommerce centers can provide subscription services via the Web and can handle subscription payments for departments that create journals or other publications for sale. These eCommerce centers are relatively inexpensive to use.

Finally, we should mention that if a more sophisticated system for marketing and income collection is required, an alliance with the *non-profit* Philosophy Documentation Center might be possible. Discussions to this effect have already taken place.

6.3 A Move to Newer Technologies?

The SEP was designed to run on proven, free technologies. Of course, since we began the project in 1995, some of these technologies are now “legacies”. But, in many cases, we chose to use certain technologies over others because they made the most sense given our budgetary constraints. We did have to make compromises in some cases.

We use HTML rather than XML for entry markup. We don’t use a heavy-duty ‘application server technology’, since performance is excellent with our Apache server and Perl/CGI scripts. We don’t rely on a heavy-duty database technology (e.g., Oracle), since our main database has/will have only 1,000 to 2,000 records, as opposed to hundreds of thousands of records. We don’t rely on Java or Javascript, though our authors are free to use it in their entries if needed. We avoid frames on the main pages of the Encyclopedia, though we use them in the web interfaces.

Those who follow trends on the web, however, will know that this implementation is relatively ‘low-tech’. But this low-tech approach does have several advantages. One is that the system can run on any PC running a free Unix-based operating system such as Linux or FreeBSD. Our low-tech approach also does not require that we purchase any licenses or require that we become dependent on any commercial software. Just as important is the advantage that both philosophers with limited computer savvy and expertise, and philosophers in other parts of the world where access to hi-tech or up-to-date computer systems may be limited, can participate in the collaborative production of our DRW.

In addition, the newer technologies always seem to increase the costs of production (they often require specialized personnel, for example), and until we are satisfied that the benefits outweigh the costs, we will exercise

⁷See <<http://www.stanford.edu/group/itss-ccs/project/ecommerce/>>.

caution when considering the latest technologies. But eventually some of these newer technologies may supercede the older ones and the SEP will have to make the needed adjustments, assuming it is in a financial position to do so.

The question we are asked most frequently is, when do we plan to adopt XML as a markup standard? XML is now highly touted as the markup language to use in web publications. As a markup language, XML offers some serious advantages over HTML. For one thing, the tags are constructed on the basis of the kinds of content that appear in a document. For example, in HTML, one would format book titles using the italicizing tags, such as `<it>...</it>` or `...`. But in XML, one could format book titles with the tag `<booktitle>...</booktitle>`. This would allow one to search of the SEP in more sophisticated ways. One could tell the search engine to search only keywords in the `<booktitle>` environment, whereas in HTML, there is no way for a search engine to distinguish a keyword found in a book title from one found in other italicized environments, such as emphasized text or foreign words and phrases.

Another advantage of XML is the promise of more sophisticated mathematical and logical formatting. HTML has only weak resources for formatting sophisticated mathematical and logical notation. There is some promise that MathML (a formatting language which is a special instance of XML), when supported by MathML-aware browsers, will give web publishers the ability to publish professional-looking mathematical and logical notation.

However, these virtues of XML come at a cost. The first and foremost of these is the fact that since XML is a more sophisticated markup language, the costs of production rise significantly when taking proper advantage of XML's extended capabilities. Currently, our authors provide us with nearly 'web ready' HTML documents, which they produce with freely available HTML-editing software.⁸ Indeed, they are free to use any HTML-editor to compose their entries—we do not want to force all authors to have to learn and/or use the same composition software. Until XML-editing software tools become widely available and easily configurable, our authors will not be willing to spend extra time using all

⁸It must be mentioned, however, that we always have to spend time to bring their documents into compliance with international standards (e.g., removing proprietary HTML formatting codes that the HTML-creation software introduces) and to make them consistent with our other entries.

the new tags provided by XML to format their entries. (For example, authors will understandably be reluctant to familiarize themselves with all the special tags such as <booktitle> and use them in their entries.)

There may be some ways, however, to ameliorate these costs. Suppose, for example, that there was a free software application which the authors could use to help them to graphically and automatically format entries in XML without learning the new tags. For example, any time an author wanted to insert a new item into the Bibliography, such a piece of software would, in response to a click on an ‘Add Bibliography Citation’ menu item, pop-up a window containing all the relevant fields (book title, article title, author name, date, city, publisher, etc.) of a typical citation. When an author inserts the information into these fields, the software would then mark the information with the appropriate XML tags in the sourcefile. There is very little extra cost to the author, since they have to type in the information in the Bibliography citations anyway. Of course, such a piece of software would be expensive to design, produce, and support on the major computer platforms. But until such software is widely and freely available or the SEP has the financial resources to hire XML-markup specialists, the move to XML will be problematic.

There is also a second problem with XML, which has to do with the fact that our authors now have electronic access to their entries and can keep them up to date. As we mentioned much earlier, authors can use their browser to contact the SEP’s server through a web interface. When they activate our ‘Make Changes’ function, their browser will divide up a private copy of the entry into sections, and for each section, display both the rendered HTML and an editing box to the HTML sourcefile. The author can then edit the HTML sourcefile and redraw the screen to see that the HTML is rendered correctly. It is not difficult for authors to read past the HTML formatting tags to edit the text they wish to change, or even add basic HTML formatting to their updated text.

But this procedure becomes more difficult in XML, and especially, MathML. XML sourcefiles are much more highly formatted than HTML sourcefiles. A simple equation such as $2^2 + x = 8$ requires numerous MathML formatting tags and it is much harder to read past these tags to find and edit the text. (MathML was designed with the idea that authors would never actually edit the sourcefile, but always edit the file through a graphical interface.) So a move to XML would make it more difficult for our authors to update their entries since the sourcefiles would become

much more difficult to edit. Again, there may be a way to get around this through a Java-based applet/servlet system which presented authors, no matter where they are located in the world, with a graphical editing interface to their XML sourcefiles. But such a Java applet/servlet combination is extremely difficult to program so that it works with all combinations of computer architectures, operating systems (and their different versions), web browsers (and their different versions), etc. The costs would be exorbitant for a project that hopes to keep costs to a minimum. It is doubtful that a single full-time programmer could design, produce, maintain, and support such an application.

Consequently, a premature move to XML would interfere with the ease of scholarly communication which the SEP now enjoys. Still, our NSF grant includes a budget line for organizing a workshop (in the 2002-2003 academic year) on the use of, and standards for, XML in Philosophy. We plan to invite members of the profession, the web-publishing community, and the XML design team to discuss these issues.

As one can see, then, there are many challenges facing the SEP. We have a system that works reasonably well now, and we are working now to put ourselves secure the SEP's future for the long-term.