

Probability Digital Library

Bob Dobrow and Jim Pitman

Proposal submitted to the NSF's NSDL program

<http://www.stat.berkeley.edu/users/pitman/mathsurvey/probdl/>
April 22, 2003

Contents

1	Project Summary	1
2	Background	2
2.1	Probability, Mathematics and Science	2
2.2	The Probability Community	2
2.3	The Probability Web	3
2.4	NSDL Resources in Probability	3
2.5	Statement of Need	3
3	Project Goals	4
3.1	“What are the chances?” online help-desk	4
3.2	Encyclopedia of Probability	4
3.3	Probability Survey Journal	4
3.4	Success criteria	5
3.5	Vertical Integration	5
4	Project Design	6
4.1	Upgrading the Probability Web	6
4.2	What are the chances?	6
4.3	Encyclopedia of Probability	7
4.4	The editorial network	8
4.5	Peer review	9
4.6	Probability Survey Journal	9

5	Technical Implementation	9
5.1	Open Journal System	9
5.2	Systems and Infrastructure	10
5.3	Metadata	10
5.4	File Formats and Web Display Technologies	10
6	Personnel	11
6.1	Management	11
6.2	Collaborations and Support	11
6.3	Student Support	12
7	Institutional Partnerships	13
7.1	Societies	13
7.2	Journals	13
7.3	Research Institutes	13
7.4	University Departments	13
7.5	Library Organizations	14
8	Timeline	14
9	Dissemination	15
10	Evaluation	15
11	Sustainability	16
12	Broader impacts	16
13	References	17

1 Project Summary

It is proposed to aggregate, expand, and manage a significant branch of the NSDL in the field of probability, by technically upgrading and developing an existing metasite, The Probability Web [21], It is proposed to use the upgraded Probability Web as a platform for the following three specific developments:

- to create a new online help-desk, to be called “*What are the chances?*” as an outreach service from the probability community to high school students and the members of the general public with an interest in chance phenomena;

- to create a new online *Encyclopedia of Probability*, to serve as a highly organized interface to brief articles on topics in the theory and applications of probability with links to the online literature, written for the benefit of students and researchers at all levels, in all of the many subjects where ideas of probability play a fundamental part;
- to create a new online *Probability Survey Journal*, to serve as a repository for longer survey papers of value to more advanced students and researchers in probability and related fields.

Such a platform will provide a portal into the world of probability that is relevant, accessible, and useful to students at all levels. As such it may serve as a model throughout mathematics and the sciences for building a discipline-specific digital library that is inviting, dynamic, that will foster scholarly communication, and that will make current knowledge and research accessible for both scholarly and public access. Funds are requested on the NSDL Collections Track to promote the rapid growth of NSDL collections in the field of probability. Funds would be used to organize and catalog the collection, and to facilitate the contribution of new content by authors, the organization and quality control of that content by editors and reviewers, and the Web display of that content in formats convenient for users. While the collection activity is specific to the field of probability, the proposed means of developing the Encyclopedia and Survey Journal tests a novel organizational strategy, which if successful in probability could be replicated in other branches of mathematics and science to dramatically increase the rate of growth of high quality NSDL content.

2 Background

2.1 Probability, Mathematics and Science

Among mathematical disciplines, probability plays a unique role in the breadth of its reach throughout the educational spectrum. It is critical to quantitative literacy of the general population, and the concepts of randomness and chance are of wide interest. According to the National Council of Teachers of Mathematics (NCTM) Standards [12], “Instructional programs from prekindergarten through grade 12 should enable all students to . . . understand and apply basic concepts of probability.” To highlight the current importance of probability in science, we quote the following assessment from a recent NSF sponsored Workshop in Current and Emerging Research Opportunities in Probability [25]:

Probability is both a fundamental way of viewing the world, and a core mathematical discipline, alongside geometry, algebra, and analysis. In recent years, the evident power and utility of probabilistic reasoning as a distinctive method of scientific inquiry has led to an explosive growth in the importance of probability theory in scientific research. Central to statistics and commonplace in physics, genetics, and information theory for many decades, the probabilistic approach to science has more recently become indispensable in many other disciplines, including finance, geosciences, neuroscience, artificial intelligence and communication networks.

2.2 The Probability Community

As observed in [25], the development and application of probability is scattered in universities and in professional organizations:

“Professional probabilists,” that is, individuals who develop the mathematical foundations of probability and who discover and promote novel applications, can be found in departments of

mathematics and statistics and in colleges of engineering and business as well as in many other university settings. The AMS and SIAM (mathematics organizations), IMS and the Bernoulli Society (statistics organizations), IEEE and INFORMS (engineering organizations) all sponsor journals and research conferences in which probability plays a central role. However, it is all too easy for the subject that is of interest to everyone to become the responsibility of no one, and the diffuse settings in which probabilists work may lead to isolation and lack of communication. Nevertheless coherence of this scattered probability community is essential to the exchange of new ideas and methods that benefit fundamental research and inspire new applications. Innovations can be motivated by problems in one sector which modify techniques from another and then find application in a third. The mathematical objects created in probability often find application far beyond the original intent. . . .

University faculty and administrators will no doubt recognize the problems of this dispersed intellectual community as belonging to a wider collection of challenges currently forcing a revolution in academic organization. Motivated by the dramatically increased importance of interdisciplinary research and education, new academic structures are being sought and implemented on many university campuses.

2.3 The Probability Web

The Probability Web is a community based service which developed in response to the needs of this “scattered probability community.” The Probability Web is a metasite consisting of specialized pages of links devoted to abstracts, books, conferences, jobs, journals, newsgroups and listservers, organizations, people, publishers, quotations, software, and teaching resources.

The Probability Web was initiated in 1995 by Phil Pollett of the University of Queensland. Since February 2001 it has been maintained and developed by the Principal Investigator, Robert Dobrow. The site resides on a server at Carleton College and is supported, in part, by technical personnel in the Math and Computer Science Department. When first conceived, The Probability Web was intended mainly as a resource for academic researchers. In the last two years, the PI has redesigned the site, improved the overall appearance, added functionality, added new pages on teaching resources and quotations, added a “what’s new” page, and has popularized the site more vigorously among undergraduate educators and by communicating with other mathematics, probability and statistics sites on the Web.

Today, The Probability Web is the first site retrieved by a Google search for “probability.” The site’s directory of People in Probability lists over 800 links to personal homepages. The Probability Web has received numerous citations and recognitions, for instance the recent review [29].

2.4 NSDL Resources in Probability

A search for “probability” on the NSDL site www.nsd1.org yields a pointer to the Electronic Journal of Probability and Electronic Communications in Probability (EJP/ECP). These research journals, produced since 1996, were among the first free electronic journals. They have grown in influence and prestige in the past few years. They are now affiliated with the Institute of Mathematical Statistics (IMS) and supported by a strong community of researchers. But NSDL does not yet provide any indexing or searching of the approximately 200 individual articles in EJP/ECP. These journals are currently in the process of upgrading their archives in such a way that their content could quite easily be made part of the NSDL, but that has not yet been arranged. Other research level content in probability is provided by about 1,000 eprints on the arXiv [27] eprint server, some large fraction of which have been published in research journals. There is

also a limited amount of undergraduate level exposition in probability from the general online mathematics encyclopedia PlanetMath and the Mathematical Sciences Digital Library (MathDL) [8] hosted by the Math Forum [7].

2.5 Statement of Need

We refer to [35, 17, 16] regarding the general need for government organizations to promote free and open access to research and educational materials. To justify our special focus on probability, we quote the following recommendation from [25]:

Funding agencies and university administrations need to provide the resources necessary to respond to the increased demand for probabilistic methodology.

The demand for online expository material in probability and more generally in mathematics is strong, as evidenced by the number of commercially published encyclopedias which are available or becoming available online, such as Kluwer's Encyclopedia of Mathematics, and Wiley's Encyclopedia of Statistical Science. But these encyclopedias are unaffordable for many institutions, and essentially unavailable to individuals. And while there are a number of free mathematical encyclopedias available online [14, 24, 4, 6] these have little high quality content in probability. We note that current content of NSDL in probability is both limited and in need of organization and integration. A goal of the Probability Digital Library is to make an organized digital record of the intellectual activity of the probability community. We envision such an achievement as having the potential for reaching out and influencing the development of NSDL in other branches of mathematics and science. The foundations for such a development in probability are all in place. They are the organizations mentioned in the previous section, that is The Probability Web, the electronic journals EJP/ECP, the arXiv, and the IMS which sponsors two of the leading research journals in probability, *The Annals of Probability* and *The Annals of Applied Probability*. Also in place is the human network, the *probability community* which supports all of these organizations and is conscious of the importance of electronic communication to maintain its identity.

We further believe there is a need for a portal that works "equally well for both scholarly and public access" to this collective intellectual activity. In [36], the Principal Investigator of the Public Knowledge Project argues for the design of such a "knowledge exchange portal. . . . The goal is to situate the research in ways that increases its intelligibility, contribution and value, whether for the public, professionals, students, or researchers." Our proposal addresses this vision by offering an entrypoint into an important corner of the scientific, technological, engineering, and mathematical (STEM) world, in a way that makes the breadth and depth of that discipline accessible to all.

3 Project Goals

Starting from the list of 800 people currently on the People Page of the Probability Web, and our own professional contacts, we aim to develop a sense of commitment in the probability community to the development and the maintenance of the Probability Digital Library as a living network of information which reflects the knowledge and expertise of the probability community. We propose to set up a credible combination of technical and human infrastructure to give authors, editors and reviewers the confidence that if they contribute their time and energy to the creation of the Probability Digital Library, their contributions will be properly organized and preserved for the long term benefit of both the probability community and the broader intellectual community which it serves. In particular, we plan to provide the following resources and services as a means of expanding the content of NSDL in the field of probability.

3.1 “What are the chances?” online help-desk

Primary users of this service will be students and teachers from K-12 and the general public.

3.2 Encyclopedia of Probability

The primary users of content in the encyclopedia will be school and university students studying probability. A secondary audience will be the large number of students and researchers in other fields of mathematics and science who need to learn about probability for applications in those fields.

3.3 Probability Survey Journal

The purpose of the Survey Journal would be to provide an account of the current state of development of probability in the form of lucid, broad, expository papers, both on established subjects, and on subjects which are developing fast and hold great promise. The content of the journal will not be research articles and thus the journal is not meant to compete with, but rather will complement, existing on-line journals such as EJP/ECP. Primary users of this service will be graduate students and researchers in probability and related fields.

3.4 Success criteria

To be successful, each of these three projects demands

- on the technical side: a Web interface which is functional for users as well as for members of the probability community contributing the content or service; for the encyclopedia and survey, this would mean the provision of authoring and editorial management tools which facilitate the work of authors, editors and reviewers.
- in the probability community: recruitment of an adequate number of people to serve as answerers, authors, editors and reviewers, and thereby provide a steady growth of high quality content.

3.5 Vertical Integration

An important goal in our construction of the Probability Digital Library is vertical integration, meaning connection between materials at various levels, and especially the integration of teaching and scholarship. The proposal expands on the historic focus of NSDL to undergraduate education and will provide a discipline-specific collection for learners and teachers at all levels: K-12, undergraduate and graduate students, researchers and professionals, as well as the general public. We envision the construction of a portal into the world of probability that will be relevant, accessible, and useful to students at all levels, that will bring the ideas of probability to the lay public, that will foster scholarly communication, and that will make current research accessible for both scholarly and public access.

Our vision is to build an integrated environment where a student’s inquiry on, for instance, the Birthday Problem (a classic probability chestnut), will lead them to:

- teaching modules and Java applet demonstrations,
- articles and references,
- an opportunity to “ask an expert,”

- survey articles to more advanced material on combinatorial probability, and
- pointers to research-level papers.

How far a student might progress along that path is, of course, beyond our control. At each level, we would strive to provide content which might serve as an invitation to the student to progress to the next level: to encourage a high school student to attend college, or an undergraduate to go to graduate school, or a graduate student to select a topic for research.

To illustrate how the Encyclopedia and Survey Journal might facilitate just one of these difficult vertical transitions in the career path of a student, we mention the example of a student who has just finished a first year graduate course in probability. Once at this stage, there are a large number of diverse areas of more advanced probability and stochastic processes which the student might consider studying, and which might lead to research problems. At this stage, an easily accessible list of such areas, and some brief expositions of them, would be of great value. Formulation of such a list could be facilitated by editors of the Survey Journal. The student could then consult Encyclopedia articles about these areas which would point in turn to surveys of the areas from various perspectives.

This sort of orientation, which may be gained by a student with the help of a good faculty adviser, is otherwise very difficult to obtain. It is a form of knowledge which is currently widely distributed among members of the probability community, and is not well recorded. We intend to provide a place for it to be recorded, through the Encyclopedia and Survey Journal media, and to encourage authors and editors to consider the needs of students when organizing their material.

4 Project Design

4.1 Upgrading the Probability Web

As the technical foundation of all following projects, we plan to overhaul The Probability Web to bring it up to the standards necessary for the goals of our project and for inclusion into NSDL. Specifically, on the technical side, we intend to

- Rewrite existing scripts using MySQL and PHP tools.
- Develop and expand annotation of existing records.
- Where necessary, create metadata to catalogue existing resources based on Dublin Core standards.
- Vigorously seek out new probability-related resources for inclusion in our metasite collection.
- convert the existing directory of over 800 homepage URLs on the People page of the Probability Web into a database for use by editors of the Encyclopedia and Survey Journal, including people's e-mail addresses, their areas of interest, and their roles in development of the Probability Digital Library (e.g. user, author, editor, referee).

4.2 What are the chances?

The Probability Web currently receives about ten communications per week. These are now sent to the maintainer (Dobrow) via e-mail from one of the forms provided on the site's homepage. Many letters come from outside academia, from people who ask questions and are curious about probability-related issues. Here is a small sample of some of the more recent ones:

- a college scout with the Green Bay Packers wants to discuss probability issues with regard to the football draft;
- a mother whose friend is having twins asks what are the chances that one baby will get Downs Syndrome;
- a high school student asks how to program his calculator to solve the birthday problem;
- a father of a Middle School student seeks help on his son's project "comparing experimental probability with theoretical probability";
- a seventh grader in El Cerrito, Calif., searches for possible ideas and books for a probability project in school.

News events often trigger inquiries. The coincidence of New York State's Pick Three Lottery coming up with the winning combination 9-1-1 on the anniversary of Sept. 11, resulted in several e-mails. After the shuttle Columbia disaster a news reporter contacted the site seeking help on a story concerning the probability of being hit by falling debris. Presently the PI has been answering all inquiries, but the number of communications is growing.

We raise these examples because they suggest an opportunity to reach out to students and the public at large with probability and mathematical ideas. We propose enhancing the site with an on-line "What are the chances?" question-and-answer interface, modeled after the "AskA" digital reference services, such as the "Ask Dr. Math" service of the Math Forum [7] based at Drexel University. We intend to collaborate with the Virtual Reference Desk [23]. This project, supported by the Department of Education, is dedicated to the advancement of digital reference and the creation and operation of human-mediated, Internet-based information services. The "AskA" services that they support are Internet-based question-and-answer services that connect users with experts and subject expertise.

Questions received through our interface will be answered by volunteers who will respond by e-mail. Questions and answers will be archived and, as a long-term goal, we will work toward the development of a searchable Frequently Asked Questions page. We will vigorously promote the new service through press releases and e-mailings to educators, schools and the mass media. We will coordinate our activities with the existing AskNSDL [2] initiative and seek partnerships and shared experiences with the network of AskA services focusing on STEM topics.

4.3 Encyclopedia of Probability

We plan to create an online encyclopedia in probability containing brief presentations of subjects, at various levels, as well as indexed integration of online resources in probability. We aim for quality of content comparable to such existing print encyclopedias as the Encyclopedic Dictionary of Mathematics [32] and the Encyclopedia of Statistical Sciences [33]. Initially, the content of the encyclopedia would consist almost entirely of pointers to remotely archived content, specifically elementary material available on other freely accessible websites like the more broadly based online encyclopedias, such as PlanetMath, Wikipedia, Distributed Encyclopedias of Science, Eric Weisstein's World of Mathematics, and openly archived content in related fields such as that assembled by the Net Advance of Physics. We intend to point readers to these sources whenever they provide adequate coverage of material related to probability. But compared to these sources, we intend to focus on the field of probability and its applications, and within that field provide greater depth and vertical integration of material. The online format will allow easy linking between theory and applications.

We also intend where appropriate to link to content through other online portals and accumulations of reference and expository material such as [19, 5, 8, 7, 10]. Pointers would also be given to more advanced material in the online literature, especially the arXiv and EJP/ECP, so the reader can jump from the encyclopedia to the latest research in an area, or move from a research survey article to an encyclopedia article in search of background.

We intend to provide interfaces for browsing beside the usual alphabetic one. One would be a tree structured display roughly following the 2000 Mathematical Subject Classification Section 60-xx (Probability Theory and Stochastic Processes) [1] and enhanced by simple ways of navigating around the tree. As a prototype for a typical encyclopedia entry, written at an advanced undergraduate level, we point to an entry on the topic of Branching Processes at <http://mathsurvey.org/branchfiles/branch/Overview.html>. Such a page offers a brief overview of the subject. From there, links might be provided to companion pages: History, Surveys, Generalizations, Specializations, Related subjects, Applications, and Bibliography.

The Encyclopedia would also link up to other resources, through a system of OAI type indexing and metadata. This can be implemented through the Research Support Tool (RST) that comes bundled with the OJS software discussed in the next section. When the reader of an article clicks on the RST, a pop-up window is displayed in the margin of the article offering a menu of choices.

Research Support Tool Menu

For this Encyclopedia article:

View Metadata

Capture Citation

Printer Version

See related

Probability Abstracts

e-Journals

Java applets

Background

Author's homepage

Other works

Define terms

Other sites

NSDL

WWW search (Google)

CiteSeer

Action

E-mail author

Add comment

About this tool

The tool works directly with the document's metadata to find related material. For example, clicking on "Probability Abstracts," automatically launches a search in a second window of Probability Abstracts. Clicking on "Other works" launches a second window with a list of links to the author's other papers.

An innovative feature of our implementation, achieved through the RTS, will be the ability to create a common indexing of formal and informal, cutting edge and background elements, which can then be

brought into alignment with each other. This will create a rich context supporting the understanding and interpretation of both novice and expert users.

4.4 The editorial network

We will develop a network of editors which more or less matches the tree structure of the 2000 Math Subject Classification in probability. About 10 Core Editors reporting to a Chief Editor will each be responsible for several related subject areas. Each core editor can delegate responsibility to up about 10 Associate Editors. The Chief Editor, who will be subject to approval by one or more of the sponsoring probability societies or organizations, will determine editorial policy in conjunction with the Core Editors. Editors will be responsible, in part, for recruiting authors to write material for both the journal and encyclopedia. A large part of the core of this editorial network is already in place. The core editors include a number of previous editors of the major research journals in probability. See [20], which includes also a list of about 85 prominent and well-respected names in the probability community who have been supportive of the Encyclopedia and Survey Journal proposals as first presented in [34].

4.5 Peer review

We plan to set a high standard for the quality of exposition for all Encyclopedia and Journal articles, initially using the traditional peer review mechanism of most mathematics journals, including EJP/ECP. We note, however, that a number of alternatives to traditional peer review are made possible by the online medium, and we anticipate experimenting with these in the future if authors, editors, and referees are willing.

4.6 Probability Survey Journal

The Survey Journal is intended to be a repository of high quality survey articles of all kind in the field of probability and its applications. Each such survey will be written as a document available free online with hyperlinks to other such documents, to the current journal literature, to relevant books and papers, and other electronic resources. We plan to initially approach a large number of experts and get each to record a brief survey of that part of the subject they know best. We intend the Survey Journal to be structured much like a standard electronic journal, such as EJP/ECP. The Encyclopedia will then serve as an intelligently structured overlay of the entire openly archived electronic literature of probability, including what is laid down in the Survey Journal. The network of editors will initially solicit material for the Journal and Encyclopedia from those who have pledged support for the project. Once the infrastructure of the Journal is in place, more attention will be focused on broadening the domain of the Encyclopedia, and articles for both outlets will be solicited from the entire probability community.

We see the Survey Journal as an outlet for expository material, of the kind that typically appears in print volumes devoted to proceedings of conferences and summer schools, festschrifts, and memorials. We intend to seek the participation of editors of such volumes to contribute their material to the Survey Journal, possibly in conjunction with a publisher who might receive print but not digital copyright. We note that the Mathematical Sciences Research Institute Book Series [11] sets an excellent precedent for such free online publication combined with traditional distribution of paper volumes by a commercial publisher.

5 Technical Implementation

5.1 Open Journal System

We plan to base the technical architecture for the Encyclopedia and Survey Journal on the Open Journal System (OJS) [13] software developed by the Public Knowledge Project [15]. The OJS is open source software made freely available to journals worldwide to make open access publishing a viable option. It is written using standard PHP and MySQL tools. It features: online submission of articles and reviews, online management for each stage of the publishing, indexing of published articles, and e-mail notification and commentary for readers. It will allow editors to set up and manage the Survey Journal website and publishing process from article submission, to assignment of editor and reviewers, organization of layout, copyediting and proofreading, and archiving. The software also provides an impressive Research Support Tool, which will be customized for probability, which accompanies each article and will provide links from the article to related resources.

We emphasise that we do not intend to engage in general software system development. Rather we expect the developers of OJS to cooperate with us to adapt that system to the needs of our project, and to make minor improvements over time with their continued support.

The proposers are in close contact with Chris Burdzy, one of the founding editors of the Electronic Journal of Probability (EJP). Burdzy has recently been working with the support staff of the Public Knowledge Project to convert EJP to the OJS system. He is well satisfied with the results, and we expect to build on that success to lay the foundation for the Survey Journal. We note also that Peter Suber, editor of the Free Online Scholarship Newsletter, lists the release of OJS in November 2002 as a landmark event in his timeline of the free online scholarship movement.

5.2 Systems and Infrastructure

The site will reside on a new dedicated server in the Mathematics and Computer Science Department of Carleton College linked to the college's overall network. The College has an extensive, stable, and well-maintained network. The OJS/RTS software will be set up and customized to drive the Survey Journal and Encyclopedia. Existing Probability Web pages, database, and scripts will be rewritten, where necessary, using PHP and MySQL. Management of the digital repository will include regular backups, at the department and college level, mirroring by other sites, refreshing media, and disaster recovery. The project will also seek cooperation with a major digital repository, such as those currently under development at Cornell, MIT and the University of California, to ensure perpetual access.

5.3 Metadata

Metadata for the Encyclopedia and Survey Journal will be developed using Dublin Core standards and be made available for harvesting via the OAI protocol. The OJS system will provide authors with a template and terms (abstract, discipline, subject, method, etc.), as well as examples to guide authors, all based on OAI protocols. The OJS system provides comprehensive indexing that adheres to the OAI protocol, as well as a metadata harvester, OAIster, for all OAI registered sites. We note the benefits of making metadata OAI compliant: broader global access to the content, the implications for identification, searching and retrieval, and the advantages metadata will afford the manager of the Probability Web in maintaining and administering content.

5.4 File Formats and Web Display Technologies

We expect that most authors will prefer to write contributions in $\text{T}_{\text{E}}\text{X}$ or $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$. Following the arXiv and EJP/ECP we intend to support a variety of file formats, including various forms of $\text{T}_{\text{E}}\text{X}$, html, pdf, and postscript. We intend to adapt and provide existing software, templates and style files for $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and $\text{BibT}_{\text{E}}\text{X}$, linking protocols, search engines and the like, and to exploit conversion programs such as $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}2\text{HTML}$ [28], TTH [31], $\text{T}_{\text{E}}\text{X}4\text{ht}$ [30] and $\text{pdfL}^{\text{A}}\text{T}_{\text{E}}\text{X}$, which support hyperlinks. There is an educational component here, to teach the community how to use these authoring tools effectively by demonstrating examples of their use.

We are aware of recent developments in the display of mathematics on the Web, in particular the development of MathML [9]. Browsers have changed dramatically over the past few years, and MathML seems to be gaining an increasing number of followers [26]. Still, we expect $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ to remain the dominant tool for authors of mathematical material for some time to come, with translation to other formats provided by programs like $\text{T}_{\text{E}}\text{X}4\text{ht}$. We intend to influence authors to write their mathematics in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ in a way that could be translated more-or-less automatically into MathML when the time comes. We expect that at some stage it will become necessary to redo Web pages in XML subject to an appropriate refined XHTML DTD. Such standardization would facilitate automated processing of the files for immediate and future applications and adaptations. But rather than immediately face the challenges of cutting edge technology for display of mathematics on the Web, we prefer to start more modestly with well established tools in common use, and presume that as the new technologies become more widely established, it should become easier to adopt them.

6 Personnel

6.1 Management

Dobrow will manage the technical and design aspects of the project and be responsible for the overall project direction. Pitman will supervise the Survey Journal and the initial content of the Encyclopedia. He and David Aldous, both at the University of California, Berkeley, will assume overall editorial responsibility, and oversee the article submission and review process. We attach letters from Jean Bertoin, David Brillinger, Steven Evans, Yuval Peres, Ruth Williams, and Ofer Zeitouni, who have agreed to be **core editors** and collectively organize the structure of some major part of the Encyclopedia and the Survey Journal. Other letters are from Endre Csáki, Jim Dai, and Zhan Shi, who have agreed to be **associate editors** responsible for recruiting authors for Encyclopedia articles and surveys in smaller branches of the subject, and coordinating with editors of neighboring branches.

We expect the prestige of the editorial network to solve a problem that has plagued on-line electronic journals in many disciplines. They are sometimes seen as less prestigious than main-line, traditional paper journals. Authors, particularly young scholars seeking tenure, are hesitant to publish in such “alternative” mediums. Based on the initial pledges of support we have received so far we do not think it unrealistic to expect that once the Survey Journal infrastructure is in place we can get 50 to 100 high-level encyclopedia entries right away and thus obtain substantial content quickly. Once we have content and visibility we believe there are several factors that will propel the Journal and Encyclopedia forward: researchers will want their work, and their research areas, linked to the Encyclopedia, and the prestige of the editors and authors should lend prestige to publications in the Encyclopedia.

Dobrow and Pitman will work closely together and be responsible for the top-level content and interactions with supporters and consultants. A full-time technical associate will be hired who will be responsible

for Web page creation and software development. He/she will work closely with the PI and student workers on all technical aspects of the proposal. We will also draw on the expertise of the Carleton College Technical Associate Mike Tie, who is responsible for the computer services in the Math and Computer Science Department, and who has implemented PHP-based software tools and created online databases and Web-based resources using PHP and MySQL.

6.2 Collaborations and Support

Carleton College is the home of the Science Education Resource Center (SERC). SERC, supported by NSDL and other NSF grants, works to improve undergraduate science education through projects that focus on supporting faculty. SERC Director Cathy Manduca and Technical Director Sean Fox have a wealth of expertise in STEM education, community organization, workshop leadership, digital libraries, and website development. They have both expressed enthusiasm for the Probability Digital Library and we believe they will provide invaluable assistance throughout the life cycle of this project.

We expect to coordinate various aspects of the project with other NSF and NSDL projects involving on-line distribution of mathematics, including the Mathematical Sciences Digital Library (MathDL) [8] hosted by the Math Forum [7], the Cornell Digital Mathematics Library [3], and the arXiv [27]. Such coordination would involve metadata standards, file and database formats, taxonomies for Web-based mathematical objects, and avoiding duplication of activity. We include a letter of support from Lawrence Moore, Editor-in-Chief of MathDL.

John Willinsky, Principal Investigator of the Public Knowledge Project, which provides the OJS software, has agreed to consult on aspects of the OJS development phase of the project.

Chris Burdzy has agreed to share his experience in adapting OJS to the needs of EJP/ECP.

Eitan Gurari, author of one of the \LaTeX to html converters \TeX4ht [30], has agreed to advise the project on technical issues of file formats and developing Web display technologies.

Ann Jensen, librarian at the U.C. Berkeley Mathematics Statistics Library has indicated her support for the project and her willingness to advise on organization of the collection, publicity to other mathematics librarians, and input related to trends in scholarly communication.

Sam Demas, College Librarian and Senior Lecturer at Carleton College's Gould Library, has conveyed a strong level of interest and support for the project, particularly in working with the PI and students on metadata standards, the virtual reference desk, and bringing relevant issues of scholarly communication forward for discussion in the Carleton community.

6.3 Student Support

A novel component of our project implementation is the involvement of student workers. Part-time computer science students will be included in some of the software development and Web-page design. At the same time we hope to draw math majors who are interested in probability into aspects of the project such as cataloguing probability resources, writing items for the Encyclopedia, and answering questions from the public interface. Both the PI and Co-PI are regular teachers of undergraduate probability courses and actively involved in issues of pedagogy and undergraduate education. Last term, the PI used The Probability Web as a resource for teaching an introductory probability class. Students were made aware of the page, and questions that were received at The Probability Web were discussed in class.

The PI plans to use the Encyclopedia directly in his undergraduate probability course at Carleton. For example, a term project assignment will be developed to write an encyclopedia entry, complete with links,

references, and examples. Topics will taken from the course syllabus, based on the needs of the encyclopedia (e.g., coupon collectors's problem, Monty Hall, St. Petersburg paradox, conditional probability, etc.).

Carleton has consistently ranked among the top five national liberal arts colleges in academic reputation. Its math and science faculties have been models for other colleges and have provided leadership in undergraduate education. Carleton students enjoy a student-faculty ratio of 10-to-1, which provides abundant opportunity for student-faculty research and collaborative scholarship. Among first-year students entering in the fall of 2001, 29% indicated that they plan to major in a science or math. And, while nationally only one in four entering students who plans to major in math or science does so, at Carleton the rate is about two out of three. We thus believe that it will be easy to recruit students to this project who are reliable, mature, motivated and sufficiently skilled for our needs. And we believe that this aspect of our implementation is an exciting opportunity to involve undergraduate students in a multi-faceted project that enhances interactions among students and faculty and broadens the educational experiences of our students.

In accompanying letters of support, both Yuval Peres and John Rice, in the Statistics Department at U.C. Berkeley, have indicated they intend to use grant support administered by other branches of NSF for the purpose of engaging students in research projects which would generate potential content for the Probability Digital Library. We intend to promote and encourage such activity at other universities.

7 Institutional Partnerships

As the Survey Journal and Encyclopedia begin to take shape, we will actively seek out partnerships between the Probability Digital Library and various on-line journals, professional societies, research institutes, university libraries and departments, and commercial publishers. Following is a list of such partnerships that have already been agreed on, as indicated in letters of support which accompany this proposal.

7.1 Societies

In an accompanying letter of support, S.R.S. Varadhan, President of the Institute of Mathematical Statistics (IMS) indicates that the IMS would welcome an affiliation with the Probability Digital Library along lines of the current affiliation of IMS with the electronic journals EJP/ECP.

In another letter, Don Dawson, president-elect of the Bernoulli Society, writes

In particular I would like to underline the need for and importance of the Survey Journal in Probability and the Encyclopedia of Probability. The most successful contribution to date in this direction is the set of volumes of the Springer Lecture Notes in Mathematics coming from the *École d'Été de Probabilités de Saint-Flour* 1971–present. Imagining this set freely available on-line with forward links and integrated in the type of structure you envision gives one the sense of the immense potential in this enterprise. As such it fully deserves the support of the international probability community and in particular the professional societies (Bernoulli and IMS), universities, research institutes and funding agencies. One obvious role for the societies (either individually or jointly) is to assume responsibility for the selection of the editorial board and this can certainly be explored in the coming months.

7.2 Journals

The electronic journals EJP/ECP and the Probability Abstracts service have expressed interest in collaborating with the Probability Digital Library to develop linking and referencing standards that would bring all

free probability literature on the Web into a single, well organized structure.

7.3 Research Institutes

Another partner is the Mathematical Sciences Research Institute (MSRI), which strongly supports the project and is interested in expanding it to other areas in the mathematical sciences. MSRI has broad experience in using the Web for disseminating mathematics. It has pioneered the use of Streaming Video to make over 1,610 lectures, seminars and conferences conducted at the Institute available at the click of a mouse, without charge, to scientists around the world. In addition, it makes available through the Web its series of books edited in-house but published by Cambridge University Press. MSRI is interested in hosting all or part of the project on its Web site and is willing to commit some staff time (roughly equivalent to the amount of time involved in editing one volume of its book series) to the project. MSRI will share its experience with the PI and Co-PI and plans on using the experience gained to further a digital library in mathematics.

7.4 University Departments

A letter from Jean Bertoin, director of the Laboratoire de Probabilités et Modèles Aleatoires, Université de Paris VI, expresses the support of that department, and makes a commitment to host a mirror site for the Probability Digital Library. Pitman has close ties with this department, as he has collaborated since 1980 with researchers there including Jean Bertoin, Jacques Neveu, Marc Yor and Loic Chaumont.

A letter from John Rice, Chairman of the Department of Statistics at U.C. Berkeley, indicates the support of that department, which employs several of the core editors involved in the Encyclopedia and Survey Journal projects.

7.5 Library Organizations

A letter from Rick Johnson, Enterprise Director of SPARC [22], the Scholarly Publishing and Academic Resources Coalition, indicates that SPARC is eager to explore a formal SPARC Publishing Partnership [18] with the project when we have obtained the requisite development support. Such a partnership should aid in promoting awareness and use of resources created by the project. SPARC is an organization of some 250 academic and research libraries that seeks to incubate viable open-access publishing initiatives.

8 Timeline

We demarcate four fluid phases of the project.

Phase I (Months 1-4)

- Purchase and set up new server with backup capability; hire Technical Associate.
- Set up OJS software for Journal and Encyclopedia.
- Design and Test “What are the chances?” interface.
- Begin software changeover of the existing Probability Web pages.
- Finish conversion of the People Directory database.

- Begin to solicit copy for the Journal and Encyclopedia

Goal for the end of Phase I: Having the new homepage of the Probability Web online with working links to the Journal, Encyclopedia, search capability, and “What are the chances?”

Phase II (Months 4 - 12)

- Articles begin to become available
- Refinement of the Journal and Encyclopedia architecture.
- Metadata cataloguing of existing resources.
- Outreach to new sites and organizations
- Monitor usage

Goal for end of Phase II: Proof of concept. Journal and Encyclopedia developing content.

Phase III (Months 12-21)

- Continuing development of the Survey Journal, while Encyclopedia begins to solicit lower-level content. Greater work at vertical integration, e.g., developing the RTS interface and setting up links to resources
- Metadata cataloguing and annotation of metasite resources.
- Outreach to new sites and organizations

Goal for end of Phase III: Full system functionality,

Phase IV (Months 21-24)

- Project evaluation

Goal for end of Phase IV: Full system functionality with project evaluation based on assessment of all components of the project.

9 Dissemination

We intend to actively promote use of the Probability Web, and bring visibility to the Probability Digital Library, e.g., by e-mail communication with the over 800 individuals on the People in Probability page, and with all mathematics and statistics departments in four-year-colleges and universities, by articles in newsletters of professional societies, presentations at the annual Joint Statistical Meetings and the Joint Mathematical Meetings, and press release mailings to news media and STEM educators. We also intend to spread the word similarly in fields where probability is applied such as economics/finance, computer science, and biology. We will also arrange links to the Probability Digital Library from the Web sites of various supporting organizations such as IMS, MSRI, and numerous college and university departments.

10 Evaluation

At all times, the site will contain a “Comments or suggestions? E-mail the Webmaster” link. These suggestions will be acted upon as necessary on an ongoing basis. In addition, an evaluation process will be designed to measure the overall usage and quality of the site. E-mailed questionnaires and survey forms on the site will be the chief tools. All surveys will be kept short and straightforward to encourage response. The process will have three phases.

Phase 1 is part of the formative evaluation which will take place as soon as the People in Probability database is installed. It will focus on the performance of the dissemination effort. E-mailed questionnaires asking simply whether the recipient has heard of The Probability Web (and if so, from where), will be sent to individuals in the database and to mathematical sciences departments in colleges and universities. Website usage statistics will be captured and maintained, and users of the site will be asked to say where they heard about the Probability Web. The response will be used to aim the dissemination effort appropriately.

Phase 2 is also formative and will take place about one year after the project begins, at which time the site will be considerably richer than it was in its initial state. Phase 2 will consist of getting feedback from users and contributors. The focus will be on:

- the ease of navigation within the site, the smoothness of the technical aspects, and the quality of the visual presentation;
- the scope, level, and quality of the contents;
- a comparison of the contents of the site with what is easily available from other sources.

Contributors will also be asked to comment on the process of publishing on the site, on specific issues such as the time-frame given to prepare contributions, the work involved in getting their document into submittable format, and so on. Questionnaires will be posted on the Web site, and their URLs will be e-mailed to all those contacted in Phase 1, as well as to all contributors and identified users such as those who use the online help-desk. The responses will thus be anonymous (in contrast with e-mail surveys), which will encourage response and reduce bias. Responses will be summarized and used to make adjustments as necessary.

Phase 3 is longer term performance evaluation, to be measured two years after project begins. Website usage statistics, maintained since the inception of the project, will provide a numerical outcome measure. A brief survey will ask for comments (such as “Things I like about the Probability Web” and “Ways in which the Probability Web should change”) instead of answers to specific questions. The URL for this survey will be e-mailed to all those contacted in Phases 1 and 2, and to all other known users and contributors. The summarized responses will help to provide the focus for future design and development efforts.

11 Sustainability

We emphasise that we are starting this project on a foundation established by the experiences of The Probability Web and EJP/ECP. Both have been functioning for the last seven years with minimal support from host universities, zero cash flow, and zero staff support. Both of these organizations have survived the critical first transfer of executive authority from their originators to their current maintainers. We propose to

technically enhance and extend these systems in a way which should improve their visibility, prestige, and chances for long-term existence. On the technical side, the OJS software is open source, and its highly automated management system offers great advantages for sustainability. On the human side, we plan to develop an extensive network of scholars who will take pride in the Probability Digital Library as an important collective creation of the probability community. The Encyclopedia and Survey Journal will have very modest long term operating costs, which should easily be absorbed by any one of a number of organizations which have already indicated their support of the project: for instance a large university, a research institute, or a professional society. We expect that as the Probability Digital Library grows in content, the duty to host and perpetuate this collection will be accepted as a matter of pride and prestige by one or more of these institutions.

12 Broader impacts

Probability is so essential to other STEM disciplines that we envision a long-term process of linking up, sharing, and collaborating with digital libraries in numerous other disciplines. While the proposal involves a collection component that is specific to probability, the proposed means of developing the Encyclopedia and Survey Journal tests a novel organization of human and digital resources. If successful in probability that organization could be replicated in other branches of mathematics and science to dramatically increase the rate of growth of high quality NSDL content. That would promote the paradigm of free online scholarship, with great benefit for the academic community and the general public.

13 References

- [1] 2000 Mathematical Subject Classification 2000 Section 60-xx Probability Theory and Stochastic Processes. <http://www.ams.org/msc/60-xx.html>.
- [2] AskNSDL. <http://asknsdl.askvrd.org/index.asp>.
- [3] Digital Mathematics Library. <http://http://www.library.cornell.edu/dmlib/>.
- [4] Distributed Encyclopedias of Science. <http://wims.unice.fr/wims/>.
- [5] EEVL: Internet Guide to Engineering, Mathematics, and Computing. <http://www.eevl.ac.uk/>.
- [6] Eric Weisstein's World of Mathematics. <http://mathworld.wolfram.com/>.
- [7] Math Forum. <http://mathforum.org/>.
- [8] Mathematical Sciences Digital Library. <http://www.mathdl.org/>.
- [9] MathML. <http://www.w3.org/Math/>.
- [10] MERLOT. <http://www.merlot.org/>.
- [11] MSRI Book Series. <http://www.msri.org/publications/books/>.

- [12] NCTM Standards. <http://www.nctm.org/standards/>.
- [13] Open Journal System (OJS). <http://www.pkp.ubc.ca/ojs/>.
- [14] PlanetMath. <http://planetmath.org/>.
- [15] Public Knowledge Project. <http://www.pkp.ubc.ca/>.
- [16] Reading list on scholarly communication issues. <http://www.lib.uconn.edu/ris/scholcommreadinglist.htm>.
- [17] Scholarly communication crisis. <http://www.lib.uconn.edu/ris/scholarlycommunication.html>.
- [18] SPARC Publishing Partners. http://www.arl.org/sparc/resources/sparc_partners.pdf.
- [19] The Atlas of Mathematics. <http://www.math-atlas.org/welcome.html>.
- [20] The Mathematics Survey: Supporters Page. <http://stat-www.berkeley.edu/users/pitman/mathsurvey/support.html>.
- [21] The Probability Web. <http://www.mathcs.carleton.edu/probweb/probweb.html>.
- [22] The Scholarly Publishing and Academic Resources Coalition (SPARC). <http://www.arl.org/sparc/>.
- [23] Virtual Reference Desk Organization. <http://www.vrd.org>.
- [24] Wikipedia. <http://www.wikipedia.org>.
- [25] Workshop in Current and Emerging Research Opportunities in Probability. <http://www.math.cornell.edu/~durrett/probrep/probrep.html>, 2002.
- [26] Math on the Web: A Status Report. http://www.dessci.com/en/reference/webmath/status/status_Jan_03.htm, January, 2003.
- [27] arXiv.org. Automated E-Print Archives. <http://arxiv.org/>.
- [28] Nikos Drakos. The \LaTeX 2HTML Translator. <http://www-texdev.mpce.mq.edu.au/l2h/docs/manual/>.
- [29] Mitch Leslie (editor). Netwatch: Take no chances. *Science*, 296, May 31, 2002.
- [30] Eitan M. Gurari. \TeX 4ht: \LaTeX and \TeX for Hypertext. <http://www.cis.ohio-state.edu/~gurari/TeX4ht/>.
- [31] Ian Hutchinson. TTH: The \TeX to HTML Translator. <http://hutchinson.belmont.ma.us/tth/>.
- [32] Shōkichi Iyanaga and Yukiyoji Kawada, editors. *Encyclopedic Dictionary of Mathematics. Vols. I and II*. MIT Press, Cambridge, Mass., Japanese edition, 1977. Translation reviewed by Kenneth O. May.

- [33] S. Kotz and N.L. Johnson, editors. *Encyclopedia of Statistical Sciences*. Wiley, 1982-1988.
- [34] J. Pitman. The Mathematics Survey Proposal. Submitted to Notices of the American Mathematical Society. <http://www.stat.berkeley.edu/users/pitman/mathsurvey/proposal.html>, 2002.
- [35] J. Pitman. Two rules of scholarly communication: publish for the public, and keep the journals. Submitted to Notices of the American Mathematical Society. <http://www.stat.berkeley.edu/users/pitman/mathsurvey/tworules.html>, 2002.
- [36] J. Willinsky. Proposing a Knowledge Exchange Model for Scholarly Publishing. *Current Issues in Education*, 3(6), Sept. 13, 2000. Available online at <http://cie.ed.asu.edu/volume3/number6/>.