The Mathematics Survey Project

Jim Pitman and David Aldous

Departments of Mathematics and Statistics
University of California, Berkeley

mathsurv@stat.berkeley.edu

Revised, August 24, 2005

Electronic journals, arXiv, Google Scholar, MathSciNet and Wikipedia are just some of the existing venues for electronic publishing and searching of mathematics. But the level between graduate textbooks and research papers is not well served at present. We propose two parallel initiatives: creation of a network of open access survey journals across all areas of mathematics, and creation of a network of web sites on particular topics (e.g. bottom-level MSC topics), each site maintained by an expert editor using the ease of modern software tools, providing both an "encyclopedia entry" for the topic and updated links to electronic resources. An initial journal *Probability Surveys* is already in operation.

Mathematical publication today

Textbooks and monographs on one side, and peer-reviewed research journals on the other side, are the most familiar categories of mathematical publication. They have not changed in essence for 50 or 100 years, and likely will not change much in the near future – the transition of journals from paper to electronic format facilitates physical access without changing the roles of authors, referees and editors and (as yet) without resolving contentious issues of price. But cyberspace provides opportunity for a much broader spectrum of types of publication. One can already find online, for instance

- unreviewed preprints
- peer-reviewed research papers
- peer-reviewed survey papers
- monographs and polished lecture notes
- lectures recorded by slide presentations, scribe notes, and videos
- literature databases
- retroactive digitization of old print literature
- descriptions by individuals or groups of their ongoing research activities
- encyclopedias at an elementary level.

We applaud this variety of content, but find three unsatisfactory features of its structure.

Cost of journals. Commercial publishers impose ever-increasing subscription costs on their ejournals, thereby restricting access, with negligible compensating advantages.

Fragmentation. The totality of mathematical material in cyberspace is at present neither well linked together nor intelligently searchable. Seeking a readable account of Topic X, one could use a search engine like Google or MathSciNet. But Google treats a mathematics page as just another page on the Web, having no conception of the logical interrelationships of mathematics. The new Google Scholar service restricts the search to the scholarly literature, accessing the content of many copyrighted books and journals as well arXiv and other open access sources. But it is not easy to restrict such a search to expository material. MathSciNet and Zentralblatt MATH enable basic searches like "find papers by author A in subject S". But there is no resource currently available for a search like "find a survey on topic X accessible to a first year graduate student". Designing a system which can respond to such queries seems to require more human intervention. As another instance, when you post your lecture notes on subject S, you currently have no systematic way of providing links *to* your material which make it easily accessible to someone searching for material on subject S.

Compartmentalization. Research progress continually increases the gap between research frontiers and first-year-graduate-textbook level material, and the gaps between different disciplines. Monographs help fill these gaps, but we see an increasing need for survey papers. At present, writing of expository survey papers carries insufficient prestige, and such papers are often scattered in hard-to-find conference proceedings and expensive handbooks. Writing high-quality surveys should be encouraged, to help organize mathematical knowledge in accessible form, and to facilitate interdisciplinary work.

The project

Instead of tackling each of these three difficulties separately, we have a bold proposal which attempts directly to solve the problems of fragmentation and compartmentalization, and indirectly to reduce the cost of commercial journals, by promoting the value of openly accessible content. We propose the formation of a large collection of open access electronic survey journals in mathematics, with articles indexed by subject for ease of access. We expect the main organization of survey journals to correspond to the different branches of mathematics, but we hope that national mathematics societies may contribute to the effort by supporting open access publishing of high quality survey papers in all fields of mathematics.

The authors work in the field of probability, which (as measured by papers in *Math Reviews*) is about 1/25 of mathematics. We describe our plan to survey the field of probability and stochastic processes, with the idea that its structure may be copied (and tinkered with) about 25 times to cover all of Mathematics. The *Mathematics Survey* is our vision of such a system of surveys, one for each major branch of mathematics.

As the foundation for a survey of probability and stochastic processes, we have started the new open access electronic journal *Probability Surveys*, with support from the Institute of Mathematical Statistics and the Bernoulli Society. This is a peer-reviewed electronic journal, with a basic user interface similar to that of existing ejournals such as *Geometry and Topology* and *The Electronic Journal of Probability.* Survey articles can be of various size and scope, ranging say from a five page write up of a conference talk on recent developments (or a five page account of some unjustly neglected classical topic) to a several hundred page monograph. They are posted on the web as accepted, with bundling into volumes of convenient size for web display. Papers are to be kept in a format like PDF or its successors, intended to be printed out for reading rather than read on-screen. We note that if high quality survey papers are provided an open access electronic outlet, they should be widely cited, hence appear near the top of a citation ranked search of scholarly articles such as now provided by Google Scholar. An open access survey journal which already exists in another branch of mathematics is *Living Reviews in Relativity*, supported by the Max Planck Institute. We plan to promote the development of such journals in all branches of mathematics, with the help of various organizations.

As a second *encyclopedia layer* of a survey of probability and stochastic processes, we plan to exploit the structure of the 2000 Math Subject Classification (MSC), much like David Rusin's Mathematical Atlas, but eventually may allow finer subclassifications and new overlapping classifications. For instance topics such as

- 60J: Markov processes
- 60J80: branching processes
- 60J80-brw: branching random walk

would conceptually be nodes of the encyclopedia layer. Initially, we imagine this to be just a tree structured index, like that provided by MPRESS for preprints, which would allow the reader to easily browse lists of survey articles and other open access material classified by subject. The structure of such an index is technically quite light: a web-page for each node of the MSC can easily be generated by a script, with some human-specified links, and the rest done by automated pointers to e.g. MathSciNet, Zentralblatt MATH, and Google Scholar. A prototype for such a distributed system of scripted web pages, with authors instead of subjects, and navigation by the author collaboration graph instead of a more complex graph describing interrelations between subjects, is provided by the BibServer system. A typical subject node would initially be generated by scripted links to existing resources. As the content at a subject node expands, we intend that control over its arrangement be provided by one or more associate editors who should develop a web site devoted to that subject. For a relatively small fraction of subjects, such sites already exist, and their maintainers should for the most part be willing to maintain content consistently with requirements of the navigation system. The value of such sites should be obvious enough that they will be created in areas where there is need. These sites and their maintainers would serve three interrelated purposes. First, the site would contain original content designed to be read on-screen – minimally a one page "encyclopedia entry" describing the subject, but this could be expanded arbitrarily according to the energy of contributors. Second, the site would assemble links to related content available on the web, including relevant papers in the survey journal and subject bibliographies. Third, the maintainers of subject specific sites would typically be willing to serve as associate editors of the survey journal.

Once this structure is set up, we expect it to quickly and automatically become the canonical place to look for links to graduate and research level mathematics: people who post material on the web are *ipso facto* wanting others to be able to look at their material, and will be happy to take one minute to transmit the link to the associate editor of the relevant node. The kind of material on the existing Probability Web – links to personal home pages of probabilists, journals, conferences etc – would become part of the material associated with the top-level (60: Probability) node. Along with the link structure, it should be straightforward to search the collection of all sites linked to the Survey.

Is the project feasible?

Consider e-mail, TEX/LATEX, and the WWW. Each started with individuals yet became indispensable, because their usefulness was obvious and because enough people were motivated to help implement them. Similarly, the usefulness of a *Mathematics Survey* is (we hope) obvious. But why do we expect people to contribute to it?

1. By emphasizing survey papers, for which few publication venues exist, we expect that *Probability Surveys* will quickly become the definitive place for authors to publish survey papers in probability.

2. Joining the project doesn't require a huge commitment of time or effort. If you are an active researcher then you typically are an expert on some subject node. To get started as an associate editor maintaining a subject site, all that's needed is to write a one page description of that subject, insert it into a suitable template and insert links to, and brief descriptions of, other online material on that subject. But these are all things you already know – it's just one afternoon's work.

3. Continuing that theme, most people are happy to write about their research speciality, so we hope that eventually a large proportion of active researchers will participate as subject node associate editors, and will contribute occasional survey articles. Indeed, provided the quality of the survey journals is well maintained, as is in the obvious interest of the profession, being invited to edit a subject node should convey the prestige of being "an established research mathematician" akin to receiving tenure. We envisage dynamic interaction with one's professional work, in that on the occasions when one needs to write research overviews – as part of organizing a workshop, planning a monograph, assembling a research group, making a grant proposal, giving a talk – one takes the opportunity to make the intellectual content be openly available on the web rather than hidden in private documents.

Why this particular approach?

Let us imagine three different projects;
(i) a survey ejournal of mathematics
(ii) an online encyclopedia of mathematics
(iii) a site which indexes and searches online mathematics.

In our opinion, each project is in one sense "too big" – it's too difficult to cover all of mathematics under any centralized scheme – while being in another sense "too small", in that it would just add an extra category to the existing categories of mathematical publication in cyberspace. We are ambitious in that we are proposing all three projects at once. But we hope that the obvious synergy between these projects will sustain self-reinforcing growth into a new feature in the landscape of online mathematics. We start with probability as demonstration, because there is a reasonably small, tightly knit community of probabilists, with a strong sense of the identity and importance of their subject in the larger scheme of mathematics and science (exemplified by specialist societies, the Bernoulli Society and the Institute of Mathematical Statistics.

We think it essential that the project be perceived as being run by the mathematical community as a whole, so we expect that individuals' involvement in the project should be largely self-organizing, with only a small degree of hierarchical structure. Perhaps controversially, we regard it as undesirable for the project to be controlled by any single scholarly society, for three reasons. Existing mathematical societies (AMS, SIAM, ...) comprise geographically- and subdiscipline-bounded subsets of the very broadly defined mathematical sciences community, and such boundaries are anachronistic in cyberspace. Societies have bureaucratic structures, which make them slow to innovate or create. And most of them derive revenue from existing publications, causing a perceived conflict of interest with the principle of free access underlying our concept. We do appreciate however that AMS has encouraged authors and editors to use its extremely useful MR Lookup and MRef facilities by pledging to maintain these services on an open access basis. This and other developments, such as the general support for open access provided by the European Mathematical Society through EMIS, and by the International Mathematical Union through its CEIC and Math-Net, offer hope that within a few years time a significant fraction of the mathematical literature may be navigable on the web without gates or tolls. The mathematics survey project could only ever represent a tiny proportion of all journal publication, so it would not directly ameliorate the systemic problem of journal costs. But every successful open access project is progress toward the tipping point when expensive journal subscriptions become unsustainable.

More about the project

1. The survey journal is intended to be non-competitive; any submission reaching the required standard of scholarship will be accepted. Refereeing is intended to improve quality of exposition and to ensure that the paper does a reasonably complete job of surveying the subject (whether broad or narrow) that it claims to survey.

2. The encyclopedia layer is not enslaved to the MSC classification. If an individual perceives some topic as an interesting research area and can articulate that perception clearly, then they can create a new node for that topic in the encyclopedia layer. Indeed, as one of many barely-foreseeable side benefits of the project once established, a listing of recently-created nodes may become the best list of "hot topics" in mathematics.

3. Obviously it will be necessary to provide some technical organization of format for web pages and the survey journal and their cross-links, but that sort of thing is becoming easier and easier to automate on a wide scale. Moreover, web crawlers such as CiteSeer and Google Scholar do much to compensate for lack of uniformity of various sites. We seek to minimize requirements for administration of people. Being an editor of *Probability Surveys* or a sibling survey journal may entail effort and responsibility comparable to being editor of a major research journal. We expect these sibling editors to communicate, but a formal structure seems unnecessary.

4. There is little hope for any human endeavor predicated upon 100% altruism and 0% self-interest. But with regard to self-interest, we have already mentioned prestige, added to which there is the opportunity to publicize one's own view of a mathematical area. Another aspect (addressed to U.S. readers but surely with analogs elsewhere) is that N.S.F. funding programs increasingly seek a "contribution to infrastructure", for individual, group, interdisciplinary and VIGRE-type grants. Involving postdocs and advanced graduate students in the writing of encyclopedia entries and survey papers can perhaps be counted as contributing to the "informational infrastructure" of the Mathematics Survey, as well as to "human infrastructure" in that we are training them to write well.

How to help

The first volume of *Probability Surveys* was posted online in 2004. We hope to launch the encyclopedia layer soon. We would appreciate technical assistance with design and content management for a generic subject site. We encourage mathematicians in other fields to take the initiative of starting up sister survey journals in their fields, and we will do what we can to facilitate this process. Please contact us at `mathsurv@stat.berkeley.edu`, and see `http://mathsurvey.org/` for more information about the project.

Note A talk based on this article was given by Jim Pitman at the Special Session on Electronic Publications at the Joint AMS-SMM International Meeting, Houston, May 13, 2004. The article appears also in the volume *New Developments in Electronic Publishing of Mathematics* edited by Hans Becker, Kari Strange and Bernd Wegner, and published by Fiz Karlsruhe, 2005.