Building a web-based bibliography for interdisciplinary sciences

Deborah Day
Scripps Institution of Oceanography Library

Abstract

The author describes the effort to build the cumulative bibliography on the history of oceanography and the bibliography of biographies of ocean scientists compiled by the IUHPS Commission of Oceanography http://scilib.ucsd.edu/sio/indexes/cbho.html The author stresses the importance of determining the audience, objectives, and scope of the bibliography, addressing the problem of multiple languages and alphabets, and determining the method of collecting citations before compilation work begins. The author discusses the importance of addressing technical infrastructure before work begins. The project will need a web host institution with sophisticated infrastructure including secure, reliable, mirrored servers and the technical expertise to maintain hardware, advise on software, and troubleshoot problems. The author finds that preliminary work setting goals, technical specifications, prioritizing tasks and allocating resources are essential to building bibliographies. The plan of work should emerge from goals, with a set of specific objectives, tasks and deadlines that can be assigned to workers in the field. The author suggests various means to build a bibliography in stages so that the task does not overwhelm the resources and scholars available to do the work.

Introduction

In 1999, I was Secretary of the IUHPS Commission on Oceanography and met with Eric Mills, President of the Commission. The Secretary of the Commission is responsible for compiling the annual bibliography on the history of oceanography which is published in the History of Oceanography Newsletter. In 1999, Eric Mills was looking forward to publishing the newsletter in electronic form. I was looking for means to make the work of compiling the bibliography easier, and I wanted to cumulate the annual bibliographies. We decided we could serve scholars better by compiling a web bibliography on the history of oceanography.

The Decision to Undertake an On-Line Bibliography

The Commission on Oceanography publishes the bibliography to facilitate the work of scholars interested in the history of oceanography. A bibliography may also draw attention to sources, encouraging their acquisition, retention and preservation in libraries. Oceanography is an international science, and texts on its history appear in many languages and many countries, in both the science and history of science literature.
These sources are not always drawn together bibliographically in catalogs. Cataloging is still done to national rather than international standards, and the application of subject terms to sources can be uneven.

While a great deal has been written on the history of oceanography, citations can be difficult to find, and papers difficult to obtain for a variety of reasons.

> There were no cumulative bibliographies of publications in the history of oceanography before our effort began, although there were some bibliographies of subfields, such as physical oceanography, Arctic science, navigation, history of scientific instruments.
>
> Publications in the history of oceanography are often hidden in the scientific literature. For instance, many scientific reports of expeditions are prefaced by excellent histories.
>
> Publications in the history of oceanography were often issued in small editions and were not disseminated worldwide to libraries.

The poor state of the bibliography is an impediment to scholars who wish to study the history of oceanography, especially scholars who wish to undertake broad studies, for instance, studies of colonial science, technology transfer, field sciences, etc. Even experienced historians must make strenuous and repetitive efforts to identify, evaluate, and locate sources.

Many scholars build and maintain their own bibliographies and libraries. These are often lost when a scholar retires or dies.

Eric Mills and I agreed that the quality of scholarship in the field could be improved and the time of scholars could be saved if a comprehensive cumulative bibliography of the history of oceanography was available to everyone online. The Librarian of the Scripps Institution of Oceanography, Peter Brueggeman, offered to host and post the bibliography.

Goals and Definitions

We began with lists of sources. Eric Mills maintained a personal card file of sources and agreed that I could have access to it. We also had ten years of bibliographies on the history of oceanography compiled by Mme. Jacqueline Carpine-Lancre, my predecessor as secretary. We had access to citations which were published in the papers of the International Conferences on the History of Oceanography. We had access to the online catalogs of a number of important libraries with strength in oceanography. We had published bibliographies relevant to oceanography, such as part B of Deep-Sea Research, (Section F.330: History of Science) and Section 522 of the Bulletin Signaletique: (Histoire des Sciences et des Techniques: Geophysique, Oceanographie), and the Isis bibliographies. So we already had many citations.

We began our work by developing a statement of the goals and scope of the bibliography. We limited our work to citations published after 1922, the date of the publication of Herdman's *Founders of Oceanography.* We decided that the bibliography would not

include newspaper articles or government documents, but would consist largely of books and scholarly journals. We concentrated on the secondary literature and did not include manuscript collections and museum collections.

We defined our audience as teachers and students of the history of science from around the world interested in oceanography. We defined oceanography as consisting of all the sciences of the sea: marine biology, biological, geological, physical and chemical oceanography, fisheries, the history of oceanographic ships, instruments, institutions and expeditions. We included some collateral fields of great interest, such as the history of navigation and communications at sea. We included publications on the subject of tides, but we excluded ship building. We included biography as a category, encompassing all substantial autobiographies, biographies, festschrifts and obituaries of prominent oceanographers, marine biologists, fisheries scientists, and other scientists who worked in the marine environment. We excluded juvenile literature – for instance, there are numerous of biographies of famous individuals, such as Fridjof Nansen and Rachel Carson written for a juvenile audience.

We decided that citations should appear in their original language, with translations of titles into English. The *Chicago Manual of Style* (14th ed., 1993) was selected as the authority for citations. These were compromises made to simplify my work as a bibliographer. I recognize now that they limited the usefulness of the bibliography. I was not able to reproduce non-Roman alphabets, so some titles were transliterated, making them less useful to scholars. While the Chicago style manual is concise, well written and is used widely in the United States, it is not an international style manual. We would have been better advised to select an international bibliographic authority.

The choice of a style manual is less important if the bibliography is based on a relational Database. A database provides flexibility in the style of display, but if a database approach is selected, it is very important to define all the fields in the citation before Compilation begins.

Text-based Sources

When we began the bibliography in 1999, we expected all our sources to be text-based sources – books, journal articles, and other published material. We did not think about digitial sources, audio-visual sources, image or specimen collections. These sources are important and should be considered when the scope of the bibliography is designed.

The web has transformed document delivery, and now scholars want not only the citation to a promising paper, but an abstract or even digital version of the whole. The bibliographer must consider how the bibliography may grow at the time the bibliography is designed. For instance, if abstracts are to be displayed, they should be acquired at the time a citation is added to the bibliography. It is a great deal more trouble to add abstracts after a bibliography is built.

It is most important to decide at the beginning if information on unpublished materials will eventually be included in the bibliography. The form of citation for manuscript material differs from that of published material, and appropriate fields must be designed into the database to accommodate description. It is most efficient to do this in the design phase of the bibliography.

Many of the works cited in our bibliography are now available in digital form. The copyright status for many citations is in the public domain, so books and articles could be digitized. The copyright holders for material under copyright protection may license digitization. New bibliographies must consider enlarging the scope of their work to encompass born-digital sources and to include plans to make full text available in searchable formats when means are available to do so. Again, it is not necessary to offer digitized full text when a bibliography is first presented, but it is wise to design in this option should the resources and technology become available to present full text, as instant access to full text is what most scholars want.

Organizational Considerations

One of our implicit decisions when we began working on the cumulative bibliography on the history of oceanography is that I would be the compiler and editor of the bibliography. There are good reasons to centralize a bibliography and have one editor with overall charge. One person is responsible for the overall quality of the bibliography and can enforce consistency of the work. The best bibliographies are compiled when the editor has access to the original item when composing the citation. A bibliographer with access to a very large library will find this work easier. It is very important to decide if the bibliography will be compiled with each piece in hand, or if citations will be accepted from contributors without consulting the original item.

But there are also disadvantages for putting the work in the hands of a single bibliographer. Chief among them is that no one person has language skills broad enough to handle citations from around the world. I was assisted by officers of the Commission of Oceanography, scholars in the history of oceanography and bibliographers from my own library. But even with all of this help, our bibliography is strongest in western citations and weakest in Asian citations.

With hindsight, I would recommend a single editor-in-chief with ready access to computers and programmers located in a center with a library. This person could be assisted by bibliographers with complementary language strengths, or by a national committee structure who would feed citations to the central editor.

Technical Considerations

We organized the work of the bibliography to fit our resources. While that was practical, it was also somewhat shortsighted. We made one important omission. We should have begun the work with the assistance of a programmer. For in the long run, the problem was not creating or posting a bibliography, the problems is maintaining and expanding it.

One is constantly adding, correcting, deleting citations, updating the bibliography and identifying incomplete or duplicate citations, locating and evaluating translations. In order to do this work methodically, the citations need to be entered in a relational database to facilitate searches and display options.

The software we chose, WAIS database software, was locally supported and easy to use, and it handled thousands of citations with ease. But it became cumbersome when the bibliography grew to tens of thousands of citations. It is English-language based and cannot handle international alphabets and diacritical marks. It is proprietary software, and it is not easy to upload new citations or download the entire bibliography to new software. We needed a big relational database, and we did not have that, nor did we have programmers who might have helped us make better choices in the beginning or migrate to more pliable databases as the bibliography grew.

It is also important to remember that not all users of the bibliography will have similar computer workstations and computer skills. Even when more sophisticated display and search options are available, some researchers just want pdf. All they want to do is print out the bibliography and read it on paper. This is such a prevalent attitude that my library now contracts with vendors to give us xml and pdf for every document we digitize.

The most important message I will deliver in this paper is the recommendation that bibliographers work in a close partnership with programmers in an environment with excellent IT support before they design a new bibliography.

Budgetary Considerations

A bibliography is expensive because it is a continuous effort, a work that is never done. A bibliography that is not up-to-date will gradually become useless. For that reason, it is best to begin the work by setting a series of reachable goals matched to the resources available for the work. Begin with a definition of the work as a whole and divide it into components with deadlines.

We began by limiting the work to citations dates 1922+. We began with the resources available to us – the labor of the Commission on Oceanography and the IT resources of the Library of the Scripps Institution of Oceanography. We know that our bibliography is not comprehensive for literature in non-Roman alphabets. We need to secure additional funds to obtain sources in Arabic, Russian, Chinese, Japanese. While it is difficult to obtain the resources to begin a bibliography, once the bibliography is online and one can demonstrate its usefulness, it is easier to obtain funding to expand and improve it.

Conclusion

I have found it enormously rewarding to build the bibliography on the history of oceanography. Scholars have been very generous with their help. I get emails almost daily offering corrections and new citations. It has sharpened my scholarly skills and broadened my knowledge of a field I thought I already understood thoroughly. It has

introduced me to an international community of people interested in this field, including students who are making important decisions about what fields they invest their lives. I have learned how new technology can be used to facilitate scholarship, and I have gained great respect for computer scientists and programmers. And finally, working on a bibliography gives one an opportunity to think about how knowledge is acquired, taught and passed from one generation to another across national and cultural boundaries. Ultimately, that is what all scholars are trying to do, to pass knowledge that has been gained through hard work to those who succeed us and to safeguard important sources of scholarship.