Puget Sound Access:
The Case History of a Local Database.

Tom Moritz
California Academy of Sciences
Golden Gate Park
San Francisco, California

Abstract

The problem of providing best possible access to "local", (geographically defined) grey litera-
ture (of limited distribution or unpublished) is discussed. A related problem is also discussed
concerning how appropriate indexing can be developed which provides adequate access to infor-
mation with local - geographic & taxonomic, or general- subject specificity?

The Puget Sound Access project was developed as an effort to resolve these problems. The
design and implementation of the project is described and a tentative evaluation is offered.

The Problem

The Puget Sound Access project developed from a practical recognition -- based on several
years experience in the Fisheries-Oceanography Library and Friday Harbor Library at the
University of Washington -- of problems of access to the literature of Puget Sound. It should be
noted that these problems are not unique to the Puget Sound region and that, correspondingly,
various solutions have been proposed and developed by librarians, information specialists and
others, worldwide.

One aspect of the problem concerns the treatment of "grey literature", the mass of informa-
tion which is produced in a variety of ways but which is characteristically:

1) unpublished (in any formal sense) or of very limited distribution;
2) locally available, fugitive and/or ephemeral, that is, published in forms or formats
   which are not intended or well suited to conventional library control;
3) not "mainstream" scientific literature (i.e. not subject to the normal peer review process).

Although this literature is frequently cited and requested by library users, it is typically not
indexed by the major abstracting/indexing services and is difficult to control and to supply.

A second aspect of the problem concerned the perceived inadequacy of existing indexing sys-
tems (including Library of Congress and ASFIS) to provide adequate treatment of locally specific
literature with particular reference to geographic and taxonomic indexing.

The SOUND ACCESS project was conceived as a way of applying state-of-the-art computer
technology to the problem of compiling, organizing, indexing and providing convenient access to
the extensive body of Puget Sound literature. The project was a logical continuation of a tradi-
tion of efforts by previous generations of University of Washington staff to contribute to effective access and control of Puget Sound literature which extends back at least 35 years.

Project Conception & Design

Several premises controlled project design: IBM-genre hardware was assumed to be standard; low cost; distribution of bibliographic data together with search software (as opposed to holding the database centrally and providing remote access); effective local geographic and taxonomic indexing was held to be essential as was effective authority control; friendliness to non-expert, end users was also considered to be critical. It was estimated that contracting with established database producers for completion of the project would have doubled project costs. It was moreover, intended that independent, critical peer review could be brought to bear on the project at several points, this proved to be only partially possible.

1) Reconnaissance -- This was viewed as an effort to systematically and comprehensively survey the existing body of Puget Sound bibliographic data and to obtain data in print or machine readable form. Bibliographic data was sought in machine readable form whenever possible and it was hoped that optical character recognition technology could be used for conversion of print to machine readable form (this proved impractical in that OCR'd records showed only 80-95% verisimilitude requiring editing which proved more costly than original keyboarding).

It soon became obvious that clearly defined editorial parameters were essential. The data collected was to be geographically specific to Puget Sound (for example, a "technique study" which was not specifically revealing of the Sound as a geographic locale would be excluded). In general, the hydrologic cycle was taken to be a parameter of content meaning that depositional studies, and ground water studies were included whereas deep geology and literature pertaining to large-scale meteorology was excluded. Political/policy related literature was included insofar as it fell within the other parameters. An inclusive definition of "literature" was applied with the general exclusion of newspaper articles, maps, personal communications and data sets.

2) Software evaluation -- Resisting the initial impulse to make a prejudicial decision on software, it was decided to make a selection of software based on a systematic review and comparison -- in the context of the specific needs of the project. Extensive programming was to be avoided (as both costly and probably redundant); it was perceived that a suitable off-the-shelf options for "bibliographic" software was commercially available. (This latter decision was consistent with recent rules promulgated by the U.S. Office of Management and Budget.)

A general needs assessment for the project was made and criteria were prepared. A systematic review of currently available software was made using a broad definition of "bibliographic software". This allowed us to take a "snapshot" of the existing market and to select the best available software package in terms of the project's needs assessment. Selection was to be based on use of a matrix for comparison as well as by consulting the existing literature. "Finalists" were to be evaluated on the basis of actual tests of sample data (this latter form of testing proved impractical given unwillingness of software distributors to cooperate and time/resource constraints of the decision process).

The evaluation process looked initially at 68 different software packages. The results of the review were published in January, 1987. INMAGIC software employed for production aspects of the project and GENCAT a Revelation-based software system produced in Canada was selected for distribution.
3) **Database design and production** -- these elements were conceived of as separate from the software evaluation phase although in practice it turned out to be virtually inextricably linked. The selected system permitted the establishment of hierarchic taxonomic and geographic authorities which were regionally specific and which became resident in the software once established.

NODC codes were used for the taxonomic system which also supported scientific and common names and extensive synonyms. Geographic coding was based on prevalent standards for identifying oceanographic regions within marine areas of the Puget Sound basin and on the existing river drainages within the basin. Washington State Water Resources Inventory Areas (WRIA's) and Indian Fishing Treaty Areas are also searchable as were habitat types (based on the U.S. FWS standard classification of aquatic habitats). The selected system also supported a resident version of the ASFIS Thesaurus.

Custom editing and revision of help screens and on screen aids was essential in view of our pursued goal of end-user friendliness.

4) **Distribution** -- it was felt that the database could be distributed as a self-contained package, together with the selected search software, philosophically, it was the Project's intent to make the package as inexpensive as possible so that it could be distributed to (i.e. affordable by) as broad an audience as possible. Cost of the final product was to reflect the direct cost of the software together with an overhead factor which would support ongoing database maintenance and updating.

5) **Core document set** -- recognizing the enormous economies of microforms, it was felt that a core set of Puget Sound documents could be assembled which would correspond to selected parts of the database and thus provide a complete document access service. Preliminary conversations were carried out with the National Technical Information Service (NTIS) and the addition to NTIS of a core document set from the Puget Sound Access database was discussed. Many Puget Sound documents are already a part of the NTIS file. (It was also felt that there was real promise in the possibility of developing a CD-ROM based, full-text delivery module which could complement the search system).

6) **Database maintenance and updating** -- it was hoped that enough interest could be generated in the developed product to provide ongoing support perhaps on a subscription basis.

**EVALUATION**

**Project Development:** The project plan was developed over a period of several months through conversations with a core group of library professionals experienced in science librarianship and/or computing but was most fundamentally based on several years of experience working in the primary aquatic science library in the Sound region. This latter factor proved invaluable to project development and underscores the importance of creative use of the insight gained from daily contact with the full range of users in a actively used library.

**Project support:** The project was funded on the basis of a proposal submitted to the newly-formed Puget Sound Institute of the University of Washington. A private company was incorporated specifically to manage the project and funding was directed through the company. Costing of services was based on an estimate that approximately 8,000 records would be processed. (The final file included 12,000 records and so our working estimate was 50% low. This
seems in part to have been the result of including the literature of Canadian inland waters but
was also a function of the thoroughness of the reconnaissance effort which seems to have far ex-
ceeded any previous efforts.) Completion of the project and continued maintenance is depend-
ent on as yet undetermined sources of support.

**Project recruitment and staffing:** Staff was recruited from a local pool of people with library
and/or regional aquatic sciences experience. Most staff agreed to work at rates considerably
below professional standards or to donate time extensively.

**Project management:** In that the project director relocated to San Francisco shortly after the
inception of the project, management proved to be one of the most problematic issues. Almost
all staff were dividing there energies between the project and other responsibilities and holding
together a coherent team of workers proved to be exceptionally difficult. Extensive use was made
of telephone and electronic mail communications and in addition frequent commuting between
San Francisco & Seattle by the project director proved essential. A major consequence of staff-
ing/management problems was a significant delay (six months) in actual completion of the first
phase of the project.

Since the delivery of this paper, several the PSAC project has come to the completion of its
first phase. The database in its "beta version" is currently available at EPA Region X in Seattle at
the Fisheries-Oceanography Library and at Western Washington State University. It is currently
available for purchase.

**Reference**

Jewell, Timothy D. (with contributions by Randall H. Erickson). An analysis of bibliographic database management