

VII. POLICY FORMULATION

A. COLLECTION POLICY

As a consequence of forty years of rapid growth in the computer industry and in the associated scientific disciplines, there is substantially more information available regarding the history of software than any institution is in a position to review. Soliciting materials which cannot be properly reviewed or, if appraised cannot be stored or serviced, is professionally reckless. Therefore, the software archive needs to define explicitly what its collection policies are, and how it expects to fulfill its obligations to access the materials it solicits and to take the appropriate destruction or retention action with respect to them. It needs to implement these policies in procedures which can be carried out within the limits of its resources. And it needs to establish mechanisms to review and evaluate the policy and make changes when appropriate ³³

Because the collection policy is intended both to inform potential donors and professional colleagues, as well as to serve the staff as a yardstick by which to evaluate immediate acquisition decisions and longer term tactics and plans, it will be broad with respect to particular contents of the archives being collected, but explicit with respect to the purposes which the collection is to serve. The collection policy should address the geographical scope of the collection (worldwide, or U.S.?); the chronological scope (all recorded history, post 1940, post 1980?); the natural languages of the holdings (English only or all languages?); and the media collected (printed, sound recording, film, photographs, machine-readable records- all?). It should explicitly identify any materials not collected.

The collection policies should address the relationship between the software archives and other activities of the repository, including its other collections, and its exhibits, publications, and outreach programs. While stating that the collection is open to all researchers, it should define the clientele which are anticipated to have the greatest interest in the collection. It should address cooperative collection development, resource sharing and information exchange policies. And it should define any external advisory

³³ Phillips, Faye; "Developing Collecting Policies for Manuscript Collections", *American Archivist*, 47(1), Winter 1984, p.30-42

structures which assist in setting collection policy.

In formulating such a statement, the repository should define a core collection, for which it maintains an on-going collection focus. For example, a dedicated software archive, such as that envisioned by the Computer Museum, might aim to collect U.S., post 1940, developments of non-proprietary operating systems and languages and defacto and de jure standards, and be continually seeking additional accessions to complete its documentation of these bases of modern computing. A software archive formed as part of a regional history collection might define its core collections as applications related software developed for in-house use by firms in its region.

It is important that historical repositories consider non-commercial, and even non-proprietary, software within their collection scope. Software systems and concepts which have become part of the public domain of computing, whether initially conceived in academia, in industry or in government, have been very influential in shaping subsequent software development. Non-proprietary software and the concepts embodied in it, are a community product, and have no other "natural" home, while the proper responsibility for documentation of the history of proprietary endeavors rests with corporate archives. Software systems which begin their lives as proprietary, may enter the mainstream of software discussion once they are successful, and the ideas they embody become public, not in the sense of ownership, but of discourse. Widely imitated operating environments, such as the IBM 370, are thus public in the sense that the ideas which are embodied in them are part of the background of every computer scientist, even if technically the operating system is still licensed and owned by IBM. Also, by documenting standards and widespread operating functions, software archives preserve a record of the fundamental structures of the software environment which will contribute to future understanding of more specialized software.

B. ACCESS POLICY

An organization dedicated to public education and the advancement of historical research has a clear obligation to make its holdings available to all on an equal basis³⁴. This will have direct implications on the kinds of

³⁴ Smart, John; "The Professional Archivists' Responsibility as an Advocate of Public Research", *Archivaria*, 16, Summer 1983, p.,139-149

restrictions which it is acceptable to permit donors to place on materials and it is one of the reasons why a program of documentation of holdings is an essential component of an archive. The policy of the archive needs to be clear about when it departs from open collections. It should prohibit the acceptance of materials which will be closed indefinitely - specific dates and/or conditions should be attached to all restrictions. It should prohibit individual exceptions to collection restrictions and should make explicit the process whereby closed collections may be opened.

Implicit in what has been said about equality of access to the research facilities is the assumption that members, if there is such a category of users, would not have any special benefits in using the archives except that if the archives is located with the repository, and their membership entitled them to free admission or other benefits, this would apply to their use of member benefits on days when they were using the archives as well. Insofar as the archives charged incidental fees, these fees (reproduction charges, prices of publications, etc.) could be discounted to members without violating the spirit of open access. Policies which charge differential rates for essential services or access should be avoided.

C. COOPERATION

In exploring the concept of a software archive, we found that there were no organizations in the United States presently dedicated to creating such a research facility (with the possible exception of the Charles Babbage Institute whose mission is broader), and that general purpose archival repositories are either unaware of the need, paralyzed by the prospect, or committed to collecting machine-readable records in software independent formats. If the task of documenting the history of software is indeed larger than the capabilities of one institution, a substantial program of profession building and cooperative enterprises will be required to build the capability to adequately document the history of software. Organizations forming software archives need to adopt such an orientation as a matter of policy. The policy must explicitly be to encourage the preservation of materials of significance to the history of software, not just to collect them. Among the efforts it must support will be to publicize the reasons why collecting software is both desirable and possible, and to define the "universe of documentation" in terms of the issues and aspects of software history which consti-

tute a complete picture of the field. If the best way to encourage their preservation is through having others collect them (and this is frequently the case) then institutions need to try to adopt the longer term view and encourage the building of collections elsewhere.

A decade ago, the professional organizations most concerned about the documentation of science and technology in the United States (the History of Science Society, the Society for the History of Technology and the Society of American Archivists), formed a Joint Committee on Archives of Science and Technology (JCAST). One of the premises of the findings of JCAST was that archival materials have "natural homes", settings which contribute best to their interpretation, and that insofar as is possible, agencies should encourage documentation to be preserved in these "natural homes" rather than removing it from such contexts for the convenience of subject matter researchers ³⁵

Following the advice of JCAST requires that institutions cooperate in building collections. Cooperation does not end with cooperative collecting however; for to be successful it must spill over into the development of standards for description of materials, including, in this situation the development of authority languages for the description of software, and to the construction of information sharing mechanisms, newsletters and union lists, which can serve to alert the other members of the community of professionals concerned with software history to development in the field.

D. ACQUISITION POLICY

Acquisition policy governs a pro-active function, whereas collection policy passively demarcates a scope for acquisitions ³⁶. As such, an acquisition policy answers such questions as how a particular focus of collecting will be determined, what types of collecting tactics will be sanctioned, how emerging collection strengths are to be built, and what criteria are to be used in

³⁵ The position of JCAST, which reflects the authors views on archives of science and technology, is presented in: Elliott, Clark ed., Understanding Progress as Process: Documentation of the History of Post-War Science and Technology in the United States, Chicago, IL, Society of American Archivists, 1983.

³⁶ Anderson, R. Joseph; "Managing Change and Chance: Collecting Policies in Social History Archives", American Archivist, 48(3), Summer 1985, p.296-303

appraisal of accessions. As policy, it stops short of such procedural guidelines as how to approach potential donors and what kinds of agreements are within bounds.

The acquisition policy must address unsolicited offers and systematic collection building. It must establish how on-going collecting priorities are to be treated with respect to focussed research efforts or exhibit plans. And it must establish a framework for evaluating or appraising records which is cognizant not only of some "inherent" values of the material, but of their utility for the repository and its clientele and the likelihood that they will be used in this century ³⁷

First, an acquisition policy should identify the mechanisms sanctioned for active acquisitions efforts. For instance, it could identify the following:

- a) solicitation of donations of materials which are related to the permanent collection interests of the repository
- b) targeted acquisition by borrowing or purchase of documentation to be used in conjunction with an exhibit; and
- c) surveying of materials associated with specific, funded initiatives.

Second, an acquisition policy must provide criteria for appraisal. For instance, it might establish that:

- a) documentation is appraised with respect to its relevance to other holdings of the repository, its importance as a software industry standard, or its relevance to known research or planned exhibits. Such a statement severely restricts collecting, by providing explicit foci for the staff, but it leaves open virtually any arena for future collecting as the collection grows, research interests shift and exhibit plans reach out for new perspectives on the field.

A repository could decide to collect documentation regarding the development of software concepts and the impact of software products on society, but not collect documentation of software products in the form in which they were made commercially available or accession those records of software firms which reveal only the financial and managerial process. Such a policy both excludes the prospect of becoming a software library or a business history archive, and focuses the attention of the staff on intellectual

³⁷ Thompson, Gloria; "From Profile to Policy: A Minnesota Historical Society Study in Collections Management", Midwestern Archivist, 8(2), 1983, p.29-39

and social change agents, yet it is broad enough to permit collections incorporating some materials which will not be retained to be appraised.

Third, an acquisition policy needs to define a scope. For instance, it might include documentation of software which is used in any civilian contexts, or in civilian contexts other than financial or manufacturing industries except when software from those spheres illustrates a concept of important to an exhibit or developed in such away as to influence more general computing. Such a policy almost could exclude computing in government contexts, with similar qualifications to those above, on the grounds that public archives should take responsibility for documentation of governmental computing which is part of the public record ³⁸

The acquisition policy should also address scope with respect to the forms of material, physical formats, and volumes of records which are to be acquired because in accepting records the institution is making a commitment to provide access to them. As we have noted, this commitment can be a considerable obligation if the format is electronic. Providing the means to study the documentation of software will not always be easy, and collecting policy must reflect this consideration ³⁹.

As an aspect of scope, the acquisition policy must address the extent to which the software archives will consist of published materials, including published software, and how such materials are to be acquired if they are collected (ie. if systematically, then their needs to be an adequate acquisition budget to support subscriptions and a mechanism for maintaining a reference library). If published documentation is not to be acquired except as part of an archival collection, the staff and researchers will still need access to a good library of reference tools and historical studies of software.

Fourth, it should address the openness of holdings, specifically whether any materials of a confidential nature, whether protected as corporate or government secrets or by confidentiality relationships should be acquired ⁴⁰.

³⁸ Levine, David; "The Appraisal Policy of Ohio State Archives", American Archivist, 47(3), Summer 1984, p.291-293

³⁹ DeWhitt, Benjamin; "Long-Term Preservation of Data on Computer Magnetic Media", Part 1, CAN #29, Part 2, CAN #30, 1987. see also, Schenck, Thomas; "Magnetic Tape Care, Storage & Error Recovery", Library High Tech, 2(4), 1984, p.51-54

⁴⁰ Whyte, Doug; "The Acquisition of Lawyers' Private Papers", Archivaria, 18, Summer

Fifth, the acquisition policy must establish methods of acquisition; whether collections will be acquired only by gift, or if they will also be purchased, and if purchased, what kinds of materials would be considered for purchase and what proportion of the archive budget might be dedicated to purchase. The kinds of gifts which will be accepted, including bequests and donations of out-of-scope materials which might have commercial value, needs to be determined. The policy should address whether donations are acceptable under such special terms as requiring that all parts be kept, or that the collection be inventoried or exhibited. While the acquisition policy can always leave room for consideration of the exceptional case, it should provide sufficient guidance to staff on a day to day basis to permit them to act and to give potential donors an appreciation for the integrity of the organization and the objectivity of the appraisal process.

In this context, the policy should, finally, also address de-accessioning. Ideally, the policy would state that de-accessioning is an on-going function for the archive that holdings undergo regular review, and it would establish a mechanism for deciding what materials should be de-accessioned. Although it is a topic of hot dispute within the archival profession⁴¹, there seems to be reason for a newly established collecting area, like software archives, to be more open about de-accessioning than others might be, since we have very little understanding yet about what kinds of materials will prove most important to research. It seems better to over acquire at this early stage, setting a time for review of the holdings which will permit us to acquire a greater perspective.

E. DOCUMENTATION POLICY

Some special practices will need to be adopted by a software archive to accommodate the requirements posed by cataloging and classifying the software documentation. These practices should, so far as possible, be built upon a base of standard archival practice, now best exemplified by the MARC-AMC format and associated rules for cataloging described by Steven

1984, p.142-53

⁴¹ Dowler, Larry; "Deaccessioning Collections: A New Perspective on a Continuing Controversy", in Nancy Pease, ed., Archival Choices (Boston, 1984).

Hensen and Elizabeth Betz.⁴² These will be well suited to the needs of a software archive in all but two respects:

1) The software archive will need to determine how to catalog the software itself, insofar as it has software products among its holdings.

2) The software archive will need a classification scheme to index its holdings so as to support access by researchers with a variety of interests since the categories provided by the Library of Congress subject headings is hopelessly inadequate to discriminate between topics in this collection.

Conventions for the cataloging of software remain in dispute despite several years of library community discourse. Several cataloging traditions have met (head on) around the issue of how to catalog software. The community of data archivists, who developed the MARC-Machine Readable Data Files (MRDF) format in the late 1970's and the community of audio-visual librarians, who see software as simply another "special" audiovisual format are finding themselves badly outmaneuvered by the traditional book catalogers in the library community for whom commercially published software packages bear a great resemblance to books. A recent revision of the AACRII, Chapter 9, guidelines for cataloging software has unified the published item position but doesn't resolve the underlying issue. Staff of a software archive can follow this debate (hopefully to its resolution) in the pages of Cataloging & Classification Quarterly.

These library cataloging discussions are taking place independently from those in the publishing community, which is trying to develop a unique identifier for commercially published software equivalent to the International Standard Serial Number (ISSN) or International Standard Book Number (ISBN). A National Information Standards Organization (NISO) Committee has been wrangling with this problem for several years and should report by the end of 1987.

Since the dedicated software archive will have no particular vested interest, its work could be greatly simplified by accepting whatever the

⁴² Hensen, Steven; Archives, Personal Papers and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies and Manuscript Libraries (Washington DC, Library of Congress, 1983).

Betz, Elizabeth W.; Graphic Materials: Rules for Describing Original Items and Historical Collections. (Washington DC, Library of Congress, 1982).

library community ultimately adopts, assuming (and this is supported by great historical weight) that the library community will immediately incorporate into its scheme any further data required or desired by publishers or vendors. This will, of course, not solve the problem for those software archives which are part of a larger organization and wish to see their holdings incorporated into institution-wide systems or databases.

The basis for a policy on indexing is even more difficult to resolve. Even though only Library of Congress subject terminology is recorded in the "subject" headings of bibliographic records, there are lots of other places indexing terminology unique to software collections could reside within library formats and library networks and perform its purposes for retrieval, if such a language existed. Unfortunately, no existing classification of software and software issues seems very useful. One reason for this is the rapid change which has been underway in software since its inception. No scheme developed to depict software at one time, or in one environment, seems very intuitive once we are removed from that moment or context. Since the absence of firm and widely held concepts is going to limit the transportability of any scheme, or its relevance over time, the repository should avoid fixed schemes.

Instead, terms from a number of facets of description should be assigned to each collection and related to a scheme which works at the present time. When that scheme requires updating, terminology embedded in the records can be left unchanged while external authority files are restructured to make the types of connections which seem best suited to contemporaries.

Three "perspectives" appear to represent software adequately and can guide us in assigning terms. The first perspective focusses on the documentation itself and employs form of material terminology to describe the documentation. The second perspective described the activity which the software supported and employs function terminology for authority control.

The third perspective focuses on the software which is the subject of the documentation and the activity and describes that it is "about." This perspective employs terms from the "lattice-work" for a software thesaurus sketched out in Appendix B. The schematic framework of that thesaurus is not fully elaborated, both because it is intended to be illustrative and because the use of the terminology from such a descriptive language leaves the typology open both to elaboration form within, by the addition of new facets

and terms within facets, and reconfiguration from above, when and if required. Staff of a software archive could use the "latticework" thesaurus, adding to it terms they for which they find "warrant" in the literature, to assure themselves that it can grow and be useful. Once a reasonable complete thesaurus has been constructed, the policy decision to adopt it can be made.

Note that the schematic representation is intended to only provide an hierarchical structure for the terms. Numerous terms may be assigned to the same item. The vocabulary does not locate any given collection or item in one place in the structure. The terms should be used rather than facet numbers, since the scheme itself can then be revised over time.

F. COLLECTIONS MANAGEMENT POLICY

Collections management policies govern the storage, retrieval, reproduction and preservation of the collections and types of records which will be maintained concerning actions taken in the management of collections.

Decisions need to be made in the context of the staffing and facilities of the software archive. Will documentation be stored in a separate facility? When and by whom will it be retrieved? What units of documentation will be retrieved (items, containers or entire collections)? What facilities will be made available for reproduction of archival materials and under what conditions? What facilities will be supported for the preservation of these materials and when will materials be considered for active conservation?⁴³

Because the software archive will consist of both machine readable materials and more traditional paper, film and sound recordings, facilities will need to be provided for the storage of a variety of media. Acquisition decisions, already made, considered the most accessible media before deciding what to retain ⁴⁴. Retrieval of machine readable media at least (and other media if desired using optical scanning) will be from workstations attached in some fashion to data storage media on which the machine-read-

⁴³ National Archives and Records Service; Evaluating a Vital Records Program: A NARS Self-Inspection Guide for Federal Agencies (Washington DC, NARS, 1983) 14pp.

⁴⁴ Day, Deborah Cozort; "Appraisal Guidelines for Reprint Collections", American Archivist 48(1), Winter 1985, p.56-63

able information is stored.

Original materials deposited or donated to the archive must be safeguarded. This requires a supervised reading area within the museum itself and a stack area which is staffed on a regular, if not constant, basis. Researchers frequently wish to make copies of materials consulted in libraries and archives for future reference. Normally this desire is accommodated by research facilities within the boundaries of copyright and on a cost reimbursable basis. The software archive needs to establish a policy governing the copying of materials (not otherwise restricted from being copied) and determine what, if any, basis is to be used to charge patrons for this service. The variety of formats of the projected archive create logistical problems which will need to be considered in developing procedures.

Most of the materials which will be received by the archives will have been written on paper. Until very recently, the paper stock used in computer facilities was as highly acidic as any ever made. Together with the extremely short half-life of information recorded on magnetic tape (stored under the best of conditions but not regularly rewound), the software archive can expect to inherit a large conservation and preservation problem. While these issues will be addressed further in development of procedures, the policy issue will need to be confronted at the time of accessioning (can the archives afford to accession materials which require conservation or will soon?).

Since the entire concept of collecting software is so new, it might be desirable for a software archive to establish a retention period for all accessions and re-appraise the holdings at that time. A reasonable retention period for material considered to be of archival significance today might be about twenty-five years, allowing the archivist to return to the material in 25 years with a fuller appreciation both of the evolution of the software industry since that time and of the holdings of the archive.

Another important aspect of a collections management policy relating to storage of the collection is to have a disaster recovery plan, based on the particular facility and staff and tested by at least a structured walk through on an annual basis.⁴⁵

⁴⁵ Hendricks, Klaus B. & Brian Lesser; "Disaster Preparedness and Recovery: Photographic Materials". *American Archivist*. 46(1), Winter 1983, p.52-68

G. DISSEMINATION & LOAN POLICY

In the ideal future, software history will be well documented, and materials about most important developments will be cared for by widely distributed archives. Individual researchers will need to know what materials are available and how to obtain them, and will want to be able to view, borrow or acquire materials locally to avoid the expense of travel. Policies which will realize a system of inter-institutional cooperation and sharing of information will benefit researchers as will carefully crafted policies on remote access.

Archives in the United States have rarely established policies on the dissemination and loan of their holdings except to prohibit either. This is an unnecessary restriction, born of a confusion between the values which lead archives to accession documentation and the market concept of value. Since most archival materials retained for the informational or evidential value, that is for the data content, not for their artifactual or "intrinsic" value, there is no reason why they, or facsimiles of them, should not be distributed, with reasonable care. Because the users and staff of a software archive are likely to be able to exploit the potential of automation and communication technologies, the software archive is in an ideal position to experiment with facsimile transmission and on-line, even full-text, databases. Indeed, a sufficiently open policy on dissemination and loan of materials may be a route by which a software archive could avoid maintaining a reading room at all.

H. POLICY ON RESEARCH USE

Earlier, we observed that the distinction between a software archive and a software library was in part based on the purpose to which users put the software. In the software library, the software was used by patrons for the same purposes it was original sold to perform while in the archive, the software was studied to understand its origins or appreciate its effect on an application arena. In order to preserve its relations with vendors, and stand a good chance of receiving most of its software and related materials as donations, the software archive must rigorously enforce the use of its holdings for research purposes only. If the software is being used for the purposes for which it was intended (or is being copied by researchers and then pirated or used without a license) the owners will, with reason, be antagon-

istic to the archives and its purposes and might even resort to legal action. EDUCOM has developed a model policy for use by educational institutions, which could serve as a framework for a policy statement by a software archive, and the basis for an agreement with researchers.⁴⁶

I. POLICY ON SPONSORSHIP

As long as corporate sponsorship for the archives is not tied to differential access to its holdings or special treatment for its records, there is no reason to establish a prohibition against any kind of corporate sponsorship or financial support. However, since the criteria for inclusion of materials within an historically oriented software archives are not the significance of the software product alone, and are definitely not an "endorsement" of the product, policies against permitting vendors to advertise that their products and associated documentation are deposited in the software archive would seem appropriate.

Finally, a structural feature of the marketplace for software might be employed by archives as a means to obtain some additional support from corporations. Software which is developed on contract, and licensed to users in a "custom" implementation, is often covered by a clause in the contract which requires the developer to deposit the source code in "escrow" in order to protect the buyer in the event the company which developed the software should cease to do business. In addition, the nature of copyright law and patent protection is such that developers of software need to retain large bodies of documentation in order to uphold their rights to products. While somewhat unorthodox, it might be possible to mix these two functions while assuring the software archive of an early deposit of desired records. The software archive could agree to serve as escrow agent, keeping the source code closed except under the provisions of that agreement for a fixed period of time. In such an arrangement, the archive might also expect to receive payment for its services as the escrow agent.

⁴⁶ EDUCOM's statement on intellectual and property rights has been widely published and is being incorporated into the policies of many academic institutions alongside their rules about plagiarism and falsification of data. For copies, write to: Steven W. Gilbert, Managing Director, EDUCOM Software Initiative, P.O. Box 364, Princeton, NJ. 08540.