A Typology of Electronic Projects with Implications for Project Management

John A. Lehman,
University of Alaska Fairbanks, Fairbanks, AK 99775, USA

Abstract

This paper reviews the use of Information Systems Technology in the context of the humanities. It first presents a classification of humanities computer projects based both on the target user and on the type of project. The classification closely follows the historical development of such projects in North America, in China and Taiwan, and in Europe. Electronic projects are thus placed in a historical context and the decisions which must be made in developing such projects are shown to be the same type of decisions as must be made for all other computer-based projects.

The first set of decisions is the target audience; the electronic cultural atlas projects described here are aimed at the world scholarly community. The next set of decisions involve standards for data definition. The third set involves technical issues for computer encoding. The last set of issues is the linking of geographical data to supporting texts and images.

Context

The context of a project depends on the target audience for the project and the type of project. Target audiences for scholarly computer-based projects may the individual scholars, groups of scholars or disciplines, or else the world scholarly community (or at least a sizable subset thereof). Project types may include catalogs, collections, analyses of collections, or maps. Figure 1 shows the current dominant pattern of such projects in North America and Western Europe; historically the dominant band originated in the lower left hand corner with individual catalogs, and has been moving "North-east" for the last thirty years or so.

Figure 1:

Audience	Catalogs	Collections	Analyses
World	XXXXX		
Group		XXXXX	
Individual			XXXXX

Audience

The audience is the group of target users for whom a project is developed. Historically, computer-based projects began as attempts by individual scholars to automate portions of their own work. This sometimes expanded to projects carried out either by research groups, or in a few cases, common to the entire discipline. Finally, over the last decade we have started to see projects targeted at scholars in multiple disciplines.

Individual Scholars

Computer-based projects by individual scholars to automate portions of their own work have historically focuses on either cataloging or on analysis and processing. Since these projects are for use by a single individual, their design has generally been idiosyncratic, and often undocumented. This has meant, among other things, that the projects have not been amenable to use by others, and that they become useless on the retirement or shift in focus of the original author. Each scholar generally solves technical issues for him or her-self, with the resulting lack of standardization. A concomitant problem is that the time, energy, and expertise of the author is diverted from his/her discipline to computer problems. A final characteristic of such projects is that data and metadata are usually closely held and communicated selectively if at all. Usually they are treated as the owner's personal intellectual property.

Group or discipline

Probably the best known group or disciplinary project is the Thesaurus Linguae Graecae which attempted to put all classical Greek literature in computer-readable form. Such projects generally have a mixed focus on processing versus data. There is a group adoption of common technical standards and a development of common software tools sets, which means that individual scholars do not need to spend time becoming computer experts, but are able simply to add study of the appropriate tools to the required methodological knowledge for the discipline. There is also a sharing of data, and especially of metadata, throughout the community of users.

World Scholarly community

Projects with an audience of a significant subset of the world scholarly community (such as TEI, the Text Encoding Initiative and ECAI, the Electronic Cultural Atlas Initiative) transcend disciplinary boundaries. They focus on data rather than processing. Metadata is public and common in adoption, and there are common database designs. Generally such projects adapt international technical standards where possible, either those promulgated by the ISO (International Standards

Organization) or those widely used in industry. Where such standards do not exist, they are developed as part of inter-disciplinary cooperation. This common technical and metadata framework serves as a means to integrate research projects so that they complement and build on one another. Users' day to day use of the tools are facilitated by common interfaces and tool-sets

Project types

Project types are the part of context which describe what sorts of activities are supported by computing. Most computer-based projects fall into one of four classes: catalogs, collections, analyses, and maps.

Catalogs

Catalogs are lists of resources. They were historically one of the first types of projects which were computerized. Catalogs may be of texts (whether published or archives), or of objects. Under the leadership of the international library community, a tremendous amount of international standardization has been done in this area.

Collections

Once a collection has been cataloged, the next logical step is to digitize the collection itself. There are currently two basic classes of such projects: collections of texts and collections of objects (images).

The software facilitating collections of texts was originally developed for government use. With the adoption of TEI (or more commonly TEI-lite) as the standard DTD for humanities text collections, such projects have multiplied internationally. Examples include the Oxford Text Project, Projects at the University of Michigan and Virginia, the text collections at Academia Sinica in Taiwan, the ongoing project in mainland China to digitize the Si-ku Quan-shu, as well as the international (sometimes multi-lingual) projects such as the Electronic Buddhist Text Initiative (EBTI). In addition to the entry of the texts, attention in such projects must be devoted to markup of structure (both text and commentary), and of names and places to facilitate electronic searching. Similarly, indexing is a major issue, both internally and more importantly, linking to library catalogs and other search entry points.

Collections of objects are usually digitized by providing a database of images. Examples are the British Library Dun-huang project and the Huntington Archives at Ohio State University. Because of the historical lack of appropriate industry standards in this area, standardization has been lacking,

with the result that most of these projects are in the individual scholar or small group stages.

Analyses

Analysis is the use of collections to produce scholarly results. Techniques include content analysis, indices, dictionaries and encyclopedias, statistical analysis, and maps. Most of these areas, representing individual scholarly productivity, are idiosyncratic. The exception is map-based analysis.

Map-based analysis

The importance of Geographical Information Systems to governments (especially their military branches) and industry has meant that much effort has gone into developing and standardizing the technology for geographical information systems; declining computing costs have meant that costs are now affordable for humanists and social scientists. However, it is only within the last year that the technology has advanced to the point where "drop-in" GIS templates are technically feasible.

Implications for project management

While scholars often lament the lack of funding for research projects in the humanities, the resulting tardiness in applying technology has meant that other disciplines, primarily the sciences, engineering, and management, have had several decades to pioneer; given the common definition of a pioneer as someone with an arrow in his back, this means that many of the problems with using the technology have been discovered the hard way by other users, and techniques worked out to deal with them. The decisions which must be made in developing such projects in the humanities are the same type of decisions as must be made for all other computer-based projects, and so we can rely on previous experience in other disciplines. The remainder of the paper comments on some of the issues which managers of business and scientific projects have dealt with relevant to each of the types of projects discussed above.

Projects carried out by Individual Scholars are analogous to the private information systems used by managers and their staffs. A common example in a University environment is the use of personal soft ledgers to keep track of expenditures independently of the (often untrusted) official accounting system. The first section discussed the major problems with such systems; the most important are the lack of usability/incompatible results and the diversion of professional expertise (resulting in highly paid managers and professionals spending there time on unproductive computer support tasks). The managerial approach to this problem has been to standardize on appropriate applications development packages, and to provide professional technical support both to save the

experts' time and to enforce common technical solutions.

Projects carried out by groups or disciplines avoid the problems with standardization and inappropriate use of time, at the expense of significant management overhead and coordination. This is particularly an issue with university research, since the participants are often split among different institutions. The major danger here is the diversion of professional expertise to administrative overhead. Recognition of the administrative needs and explicit provision for it is the appropriate response.

Any project which involves multiple participants requires substantial administrative overhead in terms of project planning and tracking. Experience with large projects since the 1950's has shown that organizations are reluctant to "waste" time and effort on detailed project design and planning, and as a result waste orders of magnitude more time in the implementation phase. Depending on the project type, detailed methodologies for software and database design have been developed, and should be used for academic projects. Similarly, project planning and tracking methodologies originally developed for military and aerospace use such as GANTT charts, PERT and CPM are automated and available for personal computers (e.g. Microsoft PROJECT). There use for any project involving more than a few people is highly recommended.

Projects carried out for the World Scholarly community have significantly higher administrative overhead, and require substantially longer to carry out due to the need of consensus building among a wide constituency, many of whom are quite sensitive to imagined slights from not being involved sufficiently in the consultative process. Recognition of this problem and provision of the appropriate administrative expertise is the standard solution.

In terms of project types, the major addition is that scientific users of geographical information systems (GIS) have found that for large projects, the problems which arise are often not GIS-related, but arise from problems with database design (often done on an ad hoc basis). Proper application of modern database design methodologies such as entity-relationship or object based design will substantially improve the effectiveness of such projects.

Conclusion

Software projects have suffered from focus on technology rather than content or management for fifty years. Analysis of the different types of projects helps identify the usability and organizational problems which are likely to arise. Most of these problems have solutions developed by trial and error within the software industry over the last forty years. Since many of the usability and organizational problems are best solved by moving to project types which involve wider

participation, awareness of the techniques for managing such projects may have a substantial impact on their success.