

## **A Model for Virtual Archiving**

**Lewis Lancaster**  
**University of California, Berkeley, USA**

The PNC has been organized in response to the problems faced by the scholars in terms of the next generation of research. The digital age has arrived but it is not yet being exploited to its fullest extent by scholars in field such as humanities and some social sciences. In particular, the use of Asian language databases offers an example of how limited our current state of research is, in relationship to the possibilities. The goals of the PNC are to investigate and document the standards and practices that are necessary for the use of East Asian language data. Beyond the role of working with the data, the consortium will give support to the scholarly groups that have traditionally formed around thematically coherent bodies of scholarship. At the present, the growing collections of digital materials in Asian languages have not been edited, peer-reviewed, do not contain metadata for search and retrieval and lack analytical tools beyond browsers. To address issues such as these, the PNC must provide prototypes for international standards. The group will work toward editorial apparatus and consistent metadata. In these ways, the members hope to provide for a historical record that can migrate from platform to platform. There is a need to develop new tools for analysis of East Asian texts in distributed collections. At the same time it is important to test tools for electronic editions and for document management systems such as Geographic Information Systems (GIS), Standard General Markup Language (SGML), Text Encoding Initiative (TEI), and Web publishing tools. As a final part of this process, the PNC must provide documentation tools for those in the thematic collaboration groups.

At the 1997 meeting of the Digital Library Workshop on Distributed Knowledge Work Environments, it was stressed that there needs to be a focus on the problem of providing the “connective tissues to bring users and collections together.” Nowhere is this more important than in the field of Asian language materials. There is a need to have a basic academic framework in place as well as the physical links through high performance cables. The next-generation libraries and digital centers, must provide for the persistence of the data, persistence that ensures existence across time and platforms. Only when scholars can be reassured that the data will survive and flourish in future decades will there be the necessary trust for their involvement in these enterprises. As the PNC was first conceived and subsequently planned, one of the themes of discussion was the development of an intellectual community within the technology of our time.

At the moment, institutions give little weight to the electronic projects of individuals when awarding tenure and promotion. The PNC members recognize the problem and have developed plans to help with this matter. Only when major institutions around the world provide the framework for

scholarly projects within the computer technology can there be assurance that scholars will receive rewards for their labors. The creation of information within the academic environment has a distinctive pattern. There is first the creation of a work by a scholar, this is followed by review and publication. At that point the commercial and academic presses provide distribution and libraries step in with purchase, cataloging and archiving. From this map of the scholarly approach to information, it is apparent that there are distinct groups and individuals who assume responsibility for the process, author, reviewer, editor, typesetter, printer, distributor, librarians as purchase agent and archiver. The electronic text has its own pattern of development and we have yet to define exactly how this can be accomplished in a way that provides rewards and persistence of the archive over time. For these reason the cooperation of the consortium must be centered on content as well as the machines and fiberoptic cables. There is a recognition that the creation of large databases requires the collaboration of a group of scholars, as well as technicians. This activity cannot be achieved by one individual as is the case with the creation of a printed monograph. Review of such work is also difficult for one individual, databases have both text and technical apparatus which must be considered and just as the creation requires a group, so too the review must be done in concert with others. The same is true for the editorial process. Databases must be put into a format that can be used for search and retrieval, software is a central issue and the adaptation to these technical protocols add to the complexity of the editing. The development of metadata for the material is an essential element of editing. Once again, there appears the need for a number of people to be involved in setting up the final form of the data. The use of the digital form raises a host of problems in terms of the medium which will be used for distribution, whether CD ROM or WEB. As the product of these labors moves into the sphere of the librarian, older patterns of acquisition and archiving do not suffice. The state of the database at the time of purchase will not be a form that can be used without modification for the future hardware and software.

The PNC recognizes that there must be serious research and development work dealing with the process of creating an electronic text through each stage to see if every participant is supported. Just as authorship, review, distribution and archiving moves from individual efforts to collaboration and multiple memberships, so too the work of providing the infrastructure must be joint efforts. The PNC provides an important step in the direction of having a plan for dealing with the next-generation problems that will arise in the interactions between institutions and between the phases of the creation of a digital text or database. It is a social problem not just a technological one, because it requires cooperation between groups that are not yet in close contact with one another.

The size of the databases brings certain problems to the surface, not the least of them being the administrative and intellectual property rights. As images and texts are brought together, copyright issues are often a crucial step in the construction of the database.

All of these challenges require a large number of people working together, even though the fields and range of expertise varies. The PNC with staff members, research scholars, technicians can provide the framework for dealing with them.