Summit on Digital Tools for the Humanities

Report on Summit Accomplishments

Introduction

Humanities scholars met at a Summit on Digital Tools for the Humanities, convened at the University of Virginia in Charlottesville, Virginia, on September 28-30, 2005. Participants from a wide variety of disciplines, such as history, literature, archeology, linguistics, classics and philosophy as well as computer science and several social sciences attended. The common characteristic shared by the participants was the use of digital resources to enable and support scholarly activities. The Summit's objectives were to explore how digital tools, digital resources, and the underlying cyber-infrastructure could be used to accomplish several goals:

- Enable new and innovative approaches to humanistic scholarship.

- Provide scholars and students deeper and more sophisticated access to cultural materials.

- Bring innovative approaches to education in the humanities, thus enriching how material can be taught and experienced.

- Facilitate new forms of collaboration among all those who touch the digital representation of the human record.

The Summit did not follow the usual format of papers, posters, and panel discussions, but was conducted as a dialogue. Participants did not present their own research to the other participants, but instead they discussed and vigorously debated selected key issues related to advancing digitally enabled humanistic scholarship. Participants were chosen by the Organizing Committee based on the submission of a brief paper that described at least one crucial issue related to digital support for humanities scholarship and education. The original intent of the Organizing Committee was to restrict Summit attendance to 35-40 participants. The response was stronger than expected and in the end more than 65 individuals were invited to participate.

The broad availability of digital tools is a major development that has grown over the last several decades, but the use of digital tools in the humanities is, for the most part, still in its infancy. For example, Geographic Information Systems were developed to perform spatial representation and analysis. Geographic Information Systems are robust and powerful and have clear applications to many humanities fields, but these systems have not been developed with these applications in mind. The question is not whether to use these and similar tools but how we can adapt them to our purposes (or vice versa). The development of strategies to aid scholars in the use or re-use of existing tools is considered by some to be more important than the creation of wholly new tools that are specially designed for the humanities.

Many, if not the majority of scholars use general purpose information technology, including tools such as document editors, teleconferencing, spreadsheets, and slide presentation formatters (i.e. tools that are broadly used in academia as well as in

business). It was the consensus of participants that only about six percent of humanist scholars go beyond general purpose information technology and use digital resources and more complex digital tools in their scholarship. And, it is this deeper use of information technology in humanities scholarship that was the basis of discussion for this Summit.

Summit participants discussed tools tailored to many purposes: analysis, creative development of new scholarly material, curation of digitally represented artifacts, as well as productivity enhancement. The Summit addressed text as well as non-textual media (audio, video, 3-D, and 4-D visualization). Many tools and collections of resources are shared and are even interoperable to some degree. Hence, it is important for the community to consider a tool's effectiveness not just for the individual but for an interactive community. Significant activity in the tool-building and tool-using community has raised new possibilities and new problems, and it is that activity that gave rise to this Summit.

Enormous efforts have been made over the past few years to digitize large amounts of text and images, establish digital libraries, share information using the Web, and communicate in new ways. Some scholars now wield information technology to help them work with greater ease and more speed. Nonetheless, there has not been a major shift in the definition of the scholarly process that is comparable to the revolutionary changes that have occurred in business and in scientific research.

When information technology is introduced into a discipline or some social activity there seem to be two stages. First, the technology is used to automate what users were already doing, but now doing it better, faster and possibly cheaper. In the second stage (which does not always occur), a revolution takes place. Entirely new phenomena arise. Two examples illustrate this sort of revolutionary change.

The first example involves inventory-based businesses. When information technology was first applied, it was used to track merchandise automatically, rather than manually. At that time the merchandise was stored in the same warehouses, shipped in the same way, depending upon the same relations among producers and retailers as before the application of information technology. Today, a revolution has taken place. There is a whole new concept of just-in-time inventory delivery. Some companies have eliminated warehouses altogether, and the inventory can be found at any instant in the trucks, planes, trains and ships delivering sufficient inventory to re-supply the consumer or vendor-just in time. The result of this is a new, tightly interdependent relationship between suppliers and consumers, greatly reduced capital investment in "idle" merchandise, and dramatically more responsive service to the final consumer.

The second example involves scientific research. For centuries there were three modes of performing scientific research: observation, experimentation and theory. Early applications of information technology merely automated what scientists and engineers were already doing. The first computers performed repetitive arithmetic computations, often more accurately than humans. Then, a revolution occurred; it

was the rise of computational science, which is now accepted as the fourth and new mode of conducting research. Computational simulations permit astronomers who cannot perform experiments-say with galaxies-to embody their hypotheses in a computer simulation and then compare simulation results with actual astronomic observation to test the validity of their hypotheses. An astronomer can simulate the collision of two galaxies with particular properties, hypothesize the outcome, and look to the sky for observable data to support or reject a hypothesis.

These two examples illustrate the kind of revolutionary change that can be enabled by information technology, specifically by digital tools and the use of an information infrastructure. The hallmark of the second stage is that the basic processes change-what the people engaged in the area actually do changes.

It is the belief of the Organizing Committee that humanists are on the verge of such a revolutionary change in their scholarship, enabled by information technology. The objective of the Summit was to test that hypothesis and challenge some leading humanist scholars to enunciate what that revolutionary change might be and how to achieve it.

Early in the Summit, participants confirmed that such a revolutionary change has not yet occurred. What emerged from Summit discussions was an identification of four processes of humanistic scholarship where innovative change was occurring, and that, taken collectively, advancements in those areas could possibly lead to a new stage in humanistic scholarship. Whether in the future this will be viewed as a revolution remains to be seen. Participants strongly believe that changes in these four fundamental processes of humanities scholarship were visible to them. This report focuses on these processes, which if dramatically changed by the use of information technology, will enable a material advancement in humanistic scholarship. The processes are:

- Interpretation

- Exploration of Resources

- Collaboration

- Visualization of Time, Space, & Uncertainty

While initial sessions (see Appendix A) focused on "tools", gradually discussion came to focus on the "processes" of scholarship. We believe that this is a hallmark of a maturing field.

The Organizing Committee authored this document in order to record for others the future as we see it unfolding. It melds together the observations and conclusions that emerged from the vibrant discussions among Summit participants. Each chapter records the consensus among participants, but there remains diversity of opinion. We offer this document as a record of the dialog, contention, and consensus that occurred. We do assume that the reader is somewhat knowledgeable about terms

and names commonly used by humanities scholars. We believe that digitally enabled research in the humanities and in humanistic education will substantively change how the human race understands and interacts with the human record, because the technology-made properly useful-can aid the human mind in doing what it uniquely does, and that is to generate creative insight and new knowledge.

The Summit commenced on Wednesday evening, September 28, 2005 with a keynote speech by Brian Cantwell-Smith from the University of Toronto in which he asked how the computer might bridge the division between body and mind, how as a mechanism it processes information, but fails to fuse meaning and matter as people do.

On the morning of the first full day of the Summit we convened two sessions of four parallel panel meetings. Panel topics were derived by the Organizing Committee based on the short issue papers that participants had submitted. These parallel sessions served to let participants meet each other and to firmly establish that the mode of interaction at the Summit was free-form discussion, as opposed to scholars reporting on their work. These panel discussions became a spring-board from which the focus on four specific processes of humanistic scholarship emerged. Appendix A describes the topics of the eight panel discussions.

We wish to thank the National Science Foundation for their support of this Summit as well as the University of Virginia Office of the Vice Provost for Research and the Institute for Advanced Technologies in the Humanities.

Organizing Committee

Bernie Frischer, Director, Institute for Advanced Technologies in the Humanities (IATH), University of Virginia (Summit Co-chair)

John Unsworth, Dean and Professor, Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign (Summit Co-chair)

Arienne Dwyer, Professor of Anthropology, University of Kansas

Anita Jones, University Professor and Professor of Computer Science, University of Virginia

Lew Lancaster, Director, Electronic Cultural Atlas Initiative (ECAI), University of California, Berkeley, and also President, University of the West

Geoffrey Rockwell, Director, Text Analysis Portal for Research (TAPoR), McMaster University

Roy Rosenzweig, Director, George Mason University Center for History and New Media, George Mason University

Interpretation

Interpretation develops out of an encounter with material or experience, and out of a reaction to some provocation-in the form of ambiguity, contradiction, suggestion, aporia, uncertainty, etc. Literary interpretation begins with reading. When the reader encounters an ambiguity, he decides if it is an interesting ambiguity, possibly a meaningful one, possibly an intended one. The reader may ask what opportunities for interpretation are offered by this ambiguity. In the next phase, interpretation may move from private to public, from informal to formal, as the interpreter rehearses and performs it, intending to persuade other readers to share an interest and some conclusions.

Commentary is one way to convey interpretation, and it can be embodied as annotation. Such annotation needs one or more points of attachment within the corpus of material under study. Multiple classes of commentary may be needed. Annotations may address only the author, or may be written to be shared. An annotation may have provenance, for example it may have been peer-reviewed itself or have been the subject of commentary. Provenance of the annotation may need to be preserved. And, annotations should attach to any type of media and should be able to contain any kind of media.

The discussion group on tools for interpretation identified the following abstract sub-components of annotation, as an interpretation-building process, grouped here by phases:

Phase 0
0.1   Identify the environment (discipline, media)
0.2   Encounter a resource (search, retrieval)

Phase 1
1.1.   Explore a resource
1.2   Vary the scope/context of attention

Phase 2
2.1   Tokenize, segment the resource (automatically or manually)
2.2   Name and rename parts
2.3   Align annotation with parts (including time-based material)
2.4   Vary or match the notation of the original content

Phase 3
3.1   Sort and rearrange the resource (perhaps in something as formal as a
        semantic concordance, perhaps just in some unspecified relationship)
3.2   Identify and analyze patterns that arise out of relationships
3.3   Code relationships, perhaps in a way that encourages the emergence of an
        ontology of relationships (Allow formalizations to emerge, or to be brought to
        bear from the outset, or to be absent)

We considered that phases 0 and 1 were probably outside the scope of our charge of interpretation per se, though we hoped that other groups, like the group focusing on exploration, might help with some of these phases. We thought that phases 2 and 3 were directly relevant to tools for interpretation.

Further, we thought that tools for interpretation should ultimately allow a user to perform actions in any of the phases in arbitrary order, and on or off the Web. Though actually publishing annotations/interpretations/commentary is probably out of scope for a tool for interpretation, narrowly defined, we agreed that there's no question that one would want to disseminate interpretation at some point in the process, and that those annotations should ideally be connected to networked resources and to other interpretations.

We spent some time discussing the audience for the kind of tools we were imagining: Developers? Power users? All humanists? High school students? With respect to users, we agreed that it was best to develop for an actual use, not a hypothetical one, but that it was also salutary to build tools for more than one use, if possible. This raised the question of whether we envisioned tools for more than one (concurrent) user: in other words, are we talking about seminar-ware? How collaborative should these tools be, and how collaborative must they be? Should they have an offline mode (for some, the answer to this question was clearly yes)? Should they allow, support, or require serial collaboration? In the end, we decided that the best compromise was a single-user tool designed with an awareness of a collaborative architecture (and we hoped to get some more information about what such an architecture might look like, from the collaboration group).

We also discussed some more specific technical matters, for example:

- Should the tools be general and broad or deep and specific? Probably the latter, but... Could we imagine an architecture that supports both? Perhaps.

- Could we specify some minimal general modalities for recognizing a unit/token? Yes: for example, a) unit is a file, b) unit is defined with a separator, c) unit is drawn by hand.

- What should be the data input options for this tool, or toolkit? Certainly, at a minimum, text, image, and time-dependent media (that might leave out geospatial information, though; is the distinction proprietary/non-proprietary formats?)

- Should these tools have a common data-modeling language? For example, UML, Topic Maps, something else? We decided that this would be necessary, but it could be kept under the hood, as it were-optionally available for direct access by the end-user.

- Could this tool be a browser plug-in (e.g. a Firefox extension)? Might this be (in its seminar-ware version) an extension to Sakai?

At this point, in an effort to bring our discussion to bear on a particular tool, and to cut short an abstract discussion of general tools for interpretation, we focused on a very specific kind of tool for annotation, namely a "highlighter's tool." We supposed that this tool would:

- Work with texts, images, time-dependent media, annotated resources, or no primary source at all.

- Work on-line or off-line.

- Allow the user to demarcate multiple parts of interest (visually or by hand) according to a user-specified rule or an existing markup language specified in a standard grammar.

- Allow the user to classify the highlighting according to that user's own evolving organizational structure, or according to an existing ontology/taxonomy specified in some standard grammar.

- Allow the user to cluster, merge, link, and overlap the demarcated parts.

- Allow the user to attach annotation to the demarcated parts.

- Allow the user to attach annotations to clusters and/or links.

- Allow the user to search at least categories of annotations, e.g. the user's own annotations.

- Produce output in some standard grammar.

- Accept its own output as input.

- Allow the user to do these things in arbitrary order.

- Allow the user to zoom to or select arbitrary levels of detail.

To further illuminate the tool (or tools) that we envision, we indicated some specific examples of uses and users. The following list suggests the range of topics, sources, and goals that we hope such a tool (or toolkit) might support:

- Chopin Variorum Project: http://www.ocve.org.uk/ This project represents a number of different editions of a work of music. The user wants to be able to study facsimiles of the documents, talk about the way they were printed, their differences, their coordinated parts, and comment on those parts and their relationships.

- A scholar currently writing a book on Anglo-American relations, who is studying propaganda films produced by US and UK governments and needs

to compare these with text documents from on-line archives, coordinate different film clips, etc.

- Dobbs Project: http://ils.unc.edu/annotation/publication/ruvane_aag_2005, serves a geographer's needs to reconstruct a history of 18th-century land settlement in the North Carolina Piedmont region, coordinating 6,000 historical documents with geospatial data.

- The Society of Professional Indexers or others creating indexes based on textual or other data.

- Open Journal Systems might use this as an add-on tool for readers (or reviewers) of journal articles.

- A UBC project looking at medical literature in electronic form, such as how autistic children interact with psychiatrists or how psychiatrists communicate with one another, could use this tool to comment on those interactions and communications in multiple media.

- Anthropologists studying the Migmaq language and culture, an ongoing scholarly enterprise that involves coordination of thousands of pages of texts, could use it for documents that are hard-to-parse for orthographic reasons.

- Variations3: http://newsinfo.iu.edu/news/page/normal/2453.html This is a large database of online music, needing annotation tools. Users will be able to play back music, view and annotate the score, and will have a tool for drawing score pages and a thematic annotation tool for audio resources. Work on such tools is already underway, in the context of this project and its data types.

- The Salar-Monguor Project: an endangered language documentation project that deals with language variation and language contact. It involves multilingual resources and multimedia, with collaborative annotation by scholars at various skill levels.

At the end of the discussion, a straw poll showed that half of the eighteen people in the room wanted to build this kind of tool, and all wanted to use it. We closed the discussion by affirming, once again, that we should build for particular applications and users but also with consideration for the agreed-upon standards for interfacing software tools. The building process should include communication, if not collaboration, with other developers. We hope that follow-up from this event will result in people in this discussion realizing a framework for collaboration, and building tools for interpretation such as the ones imagined in this discussion.

Exploration of Resources

Joanna is interested in notions of "presence" in 18th-century French and English philosophers. She calls up her Scholar's Aide (Schaide) utility to find the texts she wants to study. By clicking and dragging texts that meet her needs into Gatherings, she creates a personal study collection that she can examine. An on-line thesaurus helps her put together a list of words in French and English that indicate presence (such as near and proche), and she searches for texts containing those words. She then launches a Schaide search that only looks in her Gathering, even though the texts are in different formats and at different sites. When she checks in after teaching her Ethics of Play class, she finds a concordance has been gathered that she can sort in different ways and begin to study. She saves her concordance as a View to the public area on the Schaide Site so her research assistant can help her eliminate the false leads. Maybe she'll use the View in her presentation at a conference next week once she's found a way to visualize the results according to genre.

How can humanists ask questions of scholarly evidence on the Web? Humanists face a paradox of abundance and scarcity when confronting the digital realm. On the one hand, there has been an incredible growth in the number and types of digital documents reflecting on our cultural heritage that are now available. Projects like Google Print will in the coming years dramatically add to that abundance. Tools for discovering, exploring, and analyzing those resources remain limited or primitive, however. Only commercial tools, such as Google, search across multiple repositories and across different formats. Such commercial tools are shaped and defined by the dictates of the commercial market, rather than the more complex needs of scholars. The challenges faced by scholars using commercial search tools include:

- It is difficult to ask questions across intellectually coherent collections. What the inquirer considers a collection is usually spread across different on-line archives and databases, each of which will have a different search interface.

- Many resources are inaccessible except with local search facilities and many are gated to prevent free access.

- A user cannot ask questions that take advantage of the metadata in many electronic texts indexed by commercial tools.

- A user cannot ask questions that take advantage of structure within electronic scholarly texts (such as those encoded in TEI XML.)

- Where there is structure, it is rarely compatible from one collection to another.

- Collections of evidence are in different formats, from PDF to XML.

What kinds of tools would foster the discovery and exploration of digital resources in the humanities? More specifically, how can we easily locate documents (in multiple formats and multiple media), find specific information and patterns in across large numbers of differently formatted documents, and share our results with others in a range of scholarly disciplines and social networks? These tasks are made more difficult by the current state of resources and tools in the humanities. For example, many materials are not freely available to be crawled through or discovered because they are in databases that are not indexed by conventional search engines or because they are behind subscription-based gates. In addition, the most commonly used interfaces for search and discovery are difficult to build upon. And, the current pattern of saving search results (e.g., bookmarks) and annotations (e.g., local databases such as EndNote) on local hard drives inhibits a shared scholarly infrastructure of exploration, discovery, and collaboration.

The tasks are large, and many types of tools are needed to meet these goals. Among other things, our group saw the need for tools and standards that would facilitate:

- Multi-resource access that provides the ability to gather and reassemble resources in diverse formats and to convert and translate across those resources.

- A scholarly gift economy in which no one is a spectator and everyone can readily share the fruits of their discovery efforts.

- Serendipitous discovery and playful exploration.

- Visual forms of search and presentation.

But the group had a strong consensus, concluding that the most important effort would be one that focused on developing sophisticated discovery tools that would enable new forms of search and make resources accessible and open to discovering unexpected patterns and results. We described this as a "Google Aide for Scholars" (or Schaide in the story above)-something much broader than the bibliographic tool Google Scholar-that would be built on top of an existing search engine like Google but would allow for much more sophisticated searches than Google. Our talk of "Google" was not, however, meant to limit ourselves to a particular commercial product but rather to signal that we were interested in building on top of the existing infrastructure created by the multi-billion dollar search-industry giants such as Yahoo, MSN, and Google. Some of the features of the Google Aide for Scholars would be:

- It would take advantage of commercial search utilities rather than replace them.

- It would allow scholars to create gatherings of resources that fit their research rather than be restricted by resources. These gatherings could be shared.

- It would allow scholars to formulate search questions in different ways that could be asked of the gatherings.

- It would allow scholars to ask questions that take advantage of metadata, ontologies and structure.

- It would negotiate across different formats and different forms of structure.

- It would allow researchers to save results for further study or sharing.

- It would allow researchers to view results in different ways.

Just as Google and the other search engine companies have created an essential search infrastructure that a tool-building effort like ours needs to leverage, there are also specific tool-creation efforts underway that we should at least examine closely and perhaps even embrace. Several were mentioned and discussed as part of the brainstorming process: Pandora (a search tool for music); Content Sphere (a personal search engine developed by Michael Jensen); Meldex (another music search tool); Syllabus Finder and H-Bot (tools that make use of Google application program interface developed by Dan Cohen at CHNM); Firefox Scholar (a scholarly organization and annotation tool, also from CHNM); I Spheres (middleware that sits on top of digital collections); TAPoR (an online portal and gateway to tools for sophisticated analysis and retrieval based at McMaster University); Antartica (commercial data mining by Tim Bray); Citeseer; Proximity (a tool for finding patterns in databases developed by Jensen); personal search from commercial search engines (Google personal search and Yahoo Mindset); Amazon's A9; Cluty; and data-mining packages (NORA, D2K, and T2K from NCSA).

We developed several key specifications for this new Google Aide for Scholars. It would be extensible through Web services and, hence, might work as a plug in to Firefox or some other open client. It would be transparent in the sense that it would show the user how it behaves, rather than simply hiding its magic behind the scenes. It would also offer customizable utilities like a "query builder" that would allow one to write her own regular expressions and ontology. Most important, it would be able to plug in any ontology; filter results in complex ways and save those filters; classify and tag results; as well as to display, aggregate, and share search results.

But the success of such a tool also rests on the formatting of the resources that it seeks to access for the scholar. Scholarly resources-whether commercial aggregations (such as ProQuest Historical Newspapers), digital libraries (such as American Memory and Making of America), gated repositories of scholarly articles (such as JSTOR), and especially the emerging mega-resource promised by Google Print-need to be visible and open. Achieving that goal is more of a social and political problem than a technical challenge. But we can facilitate that goal by offering guidelines for how to make a site visible through existing and emerging standards, such as OAI and the XML approach followed by Google.

In general, then, we see on the one hand the need for a lobbying group that will promote making resources openly available and discoverable. On the other hand, we believe that the actual tools development can proceed in an incremental and decentralized fashion through three different development groups: (1) a group developing a client-based tool (perhaps built into the browser) that can access multiple resources but using Google or its counterpart; (2) a group developing a server-side repository that would aggregate information from searches and annotations; and (3) a decentralized group (or set of groups) that would write widgets, Web services, and ontologies that would operate in the extensible client software as well as off the server.

Collaboration

The humanities, as scholarly disciplines, prize individual scholarship. There is a long history of collaboration in producing joint works and analysis, but traditionally the greatest rewards go to those who have worked in isolation. The books and papers that emerge from that isolation make unique and personal contributions to scholarly fields. The book format provides a flexible medium for arguing, explaining, and demonstrating. The relatively long production period allows for repeated interaction on a specific topic by a limited and known set of authors.

Digital tools enable a new kind of collaboration, grounded in a shared, rich representation (perhaps evolving with the collaboration activity) of textual, audio, and visual material. Digital representations of material can be searched, analyzed, and altered at electronic speed. More dramatically, they lead to orderly cooperation by many, perhaps hundreds or thousands, of individuals. And, all of these collaborators can access and edit the same representation of data from geographically distant sites.

Digital scholarly projects, especially if they use custom software for presentation and processing, demand a level of technical, managerial, and design expertise that content providers often do not possess. These projects require a team of scholars to handle information that is based on a well-defined methodology and technicians to design, build, and/or implement software and hardware configurations. The end-products of such collaborations are passed on to academic and research libraries and archives that must be able to collect, disseminate, and preserve machine-readable data in useful and accessible forms. At one level, collaboration diverges from a basic principle of individual scholarship. For better or worse, when a scholar decides to create or make use of shared digital resources, he loses the option of working solo. This change is not merely the formation of a team of experts, but a basic shift in academic culture. By comparison, research in the sciences has long recognized team efforts. Research reports and papers are often the product of coordinated efforts by many researchers over a long period and multiple members of the team will be credited as authors. A similar emphasis on collaborative research and writing has not yet made its way into the thinking of humanists, so it is not surprising that the movement toward complex digital tools-which the individual scholar often does not master and use on her own-has been slow.

Summit participants explored aspects of digitally enabled collaboration, focusing on the tools that facilitate collaboration. We believe that they enable the creation of new scholarly processes as well as dramatic increases in productivity. Such tools:

- Provide novel forms of expression (visual, virtual reality, 3D, and time compression).

- Translate between alternative forms of representation (lossless and lossy).

- Collect data with mechanical sensors and recorders.

- Track the details of collaborative activity and coordination.

- Facilitate sharing of data and complex, timely interactions.

- Enforce standards and user-defined ontologies, data dictionaries, etc.

- Analyze material based on domain knowledge that is built into the tool.

By consensus, the group divided such tools into two categories. The first category is not specific to humanistic research, but includes tools that ride the technology wave and appear to serve a very general audience. They include:

- Grid technologies, in particular data grids that will provide transparent access to disparate, heterogeneous data resources.

- Very high speed communication, such as Internet3 and the National LightRail.

- Optical character recognition tools that can scan-in text and image content, even content that is obscured or otherwise difficult to see.

- "Gisting" tools that translate portions of text from one language to another so that the researcher can get a simple notion of the content of a text.

- Data mining tools that find complex patterns within poorly structured data.

- General purpose teleconference.

These tools are certainly valuable and are heavily used by scholars as well as business and personal users. Academic or research applications may be secondary to commercial applications, however, so the scholarly community cannot expect tool designers to cater to its particular needs. Indeed, the ebb and flow of development of these tools is more likely to be driven by market forces and popular interests.

The second category encompasses tools that are tailored to the humanist and to scholarly collaboration. These are what the community itself needs to design and build. They include:

- Wikis expanded for humanists. Some should support greater structure over content as well as additional editorial or quality control functionality.

- Tools to find and evaluate material in user-specific ways.

- Tools to define narrative pathways through a three-dimensional environment, attaching content to locations along the path, and with the ability to explain at each juncture why particular paths were chosen.

- Tools to translate between representation-rich formats at a sophisticated level, with annotation and toleration for limited ambiguity.

A scholar operating in an environment populated with digital data resources channels her scholarship in her choices and uses of her tools, such as mark-up, database structure, interfaces, and search engine. The scholarly environment has become more technical, and the quality of the work produced is influenced by decisions that are of a technical nature. This requires that the scholar understand the function as well as the limitations of the tools selected. Many tools must be tailored to particular uses. That is, the user determines and sets parameters to control what results are returned. A simple example is the search engine. Advanced search parameters permit the user to specify what a search returns, so that the user is not overwhelmed with irrelevant or superfluous items. However, the user also needs to be able to predict what will not be returned as results, so that blind spots do not emerge.

Because these tools are just emerging and are evolving rapidly, participants concluded that the most valuable aid for collaboration would be a clearinghouse to inform and educate digital scholars about useful tools. Specifically, they would like to see a forum for evaluating when and how a tool is usefully wielded and what undocumented bugs the tool might exhibit in its functioning. This would include a careful description of each tool and a URL link to the tool's host site; the creator's vision of the tool; related software tools, resources, and a list of projects which use that tool; and rich indexing and categorization of tools listed in the clearinghouse or repository.

To be even more useful, site content could be subject to a quality control process, which might include peer review of tools that are listed, as well as in-depth objective analyses of the pros and cons of the tool for various applications and necessary level of maintenance. Competitive tools in a category should be rated fairly against one another on several dimensions. The site should also provide relevant facts, such as dissemination history, pointers to full documentation, objective descriptions of uses by other scholars across multiple disciplines, statistics on adoption, and published results that cite the tool's use.

This kind of a clearinghouse would require long-term funding support. However, it would be a useful mechanism for funding agencies to learn what tools are more productive, how they are being used, who uses them, and what tool functionality is still needed.

Participants also felt that there should be follow-up conferences and outreach meetings to discuss the issues related to digitally enabled collaboration. These conferences should include both those who fund tools and those who use them (e.g., graduate students, post doctoral fellows, and rising and established scholars). Outreach meetings could expand the digital scholarly community and target those who are reluctant or ill-equipped to take advantage of new technologies.

Collaboration that uses computer technology requires connectivity. Scholars around the world can collaborate in this virtual environment only where there is ease of communication and transfer of information. Connections between institutions have expanded greatly over the last decade, and many of the bigger and richer institutions use terrestrial fiber optic cables. As a result, many scholars routinely exchange

gigabytes of data per second and humanities scholarship has become "data rich." However, not everyone has access to high speed connections. Not all institutions provide access to sufficient bandwidth and there is a growing divide between have and have-not institutions and individuals. Scholarly teamwork that can rely on the transfer of gigabytes of memory every second will be quite different from that which is limited to 50 kilobytes per second.

In order for the evolving use of digital tools in the humanities to be explored and to blossom fully, institutions need to make certain that scholars who pioneer digital methods are rewarded and encouraged. Universities should work together to design a clear method of evaluating and appraising collaborative creation of digital data resources and other digital works. Our faculties and administrators cannot have assurance that such computer-aided research is a legitimate part of the academic enterprise until there is a process for dealing with digital scholarly works.

In the end, scholarship is the result of human thought. The tools are simply devices to aid in applying the human mind. Collaboration tools are meant to help groups of people excel in concert. With digital support, these efforts can incorporate the thought of many more individuals than those without automated support.

Visualization of Time, Space, and Uncertainty

Any interpretation of the past is an attempt to piece together fragmentary data, dispersed in space and time. Even so-called "intact" cultural historical sites brought to light after centuries, or even millennia of deposition are rarely perfect fossils or freeze-frames of human activity prior to their abandonment. In this sense, all work with material remains constitutes an archaeology of fragments. However, the degree and nature of such fragmentation can vary significantly. Historical landscapes coexisting with modern cityscapes produce far more fragmented records compared to open-air sites. It is both the spatial and temporal configuration of excavations in the city that causes the archaeological record to become further fragmented at the post-depositional stage. The sequence and duration of excavations is frequently determined by non-archaeological considerations (building activity, public works, etc.) and the sites are dispersed over large areas. Thus, it becomes problematic to keep track of hundreds of excavations (past and present) and to maintain a relational understanding of individual sites. On the other hand, it is clear that historical habitation itself can also contribute to the fragmentation of the record. Places with a rich occupation history do not simply consist of a past and a present. Rather, they comprise temporal sequences intertwined in three-dimensional space; the destruction of archaeological remains that is caused by overlapping habitation levels can be considerable.

Existing Software Tools for Solving the Problem, and their Limits

To help us manage this "archaeology of fragments," there are two sets of existing software tools: Geographic Information Systems (hereafter GIS) and 3D visualization software (a.k.a. "virtual reality"). GIS provides the tools to deal with the fragmentation of historical landscapes in contemporary urban settings, as it enables the integration and management of highly complex, disparate and voluminous data sets. Spatial data can be integrated in a GIS platform with non-spatial data; thus, topological and architectural features represented by co-ordinates can be integrated with descriptive or attribute data in textual or numeric format. Vector data can be integrated with raster data; that is, drawings produced by computer-aided design imagery. GIS helps produce geometrically described thematic maps, which are underpinned by a wealth of diverse data. Through visualization, implicit spatial patterns in the data become explicit, a task that can be quite tedious outside a GIS environment. In this respect, GIS is a means to convert data into information. Visualization can be thought-provoking in itself. However, the strength of GIS is not limited to cartographic functionality. It mainly lies in its analytical potential. Unlike conventional spatial analysis, GIS is equipped to analyze multiple features over space and time. The variability of features can be assessed in terms of distance, connectivity, juxtaposition, contiguity, visibility, use, clustering etc. In addition, the overlay and integration of existing spatial and non-spatial data can generate new spatial entities and create new data. Such capabilities render GIS an ideal toolbox for grappling with the fragmented archaeological record of an archaeological site from a variety of perspectives.

3D visualization software can add the third dimension to a scholar's data sets, enabling a viewer not only to visualize the distribution and spacing of features on a flat time-map, but also to understand how the data cohere to form a picture of a lost world of the past. Once this world has been reconstructed as accurately as we can make it, we can do things of great use for historical research and instruction: re-experience what it was like to see and move about in the world as it was at an earlier stage of human history; and run experiments on how well buildings and urban infrastructure functioned in terms of illumination, ventilation, circulation, statistics, etc.

While proprietary GIS and 3D visualization software packages exist, they were designed with the needs and interests of contemporary practitioners and problem-solvers in mind: geologists, sociologists, urban planners, architects, etc. Standard packages do not, for example, have something as simple as a time bar that shows changes over time in 2D (GIS) or 3D. Moreover, GIS and 3D software are typically distinct packages, whereas in historical research we would ideally like to see them integrated into one software suite.

Finally, there is the matter of uncertainty and related issues, such as the ambiguity, imprecision, indeterminacy, and contingency of historical data. In the contemporary world, if an analyst needs to take a measurement or collect information about a feature, this is generally possible without great exertion. In contrast, analysts of historical data must often "make do" with what happens to survive or be recoverable from the historical record. But, this means that if historians (broadly defined) utilize GIS or 3D software to represent the lost world of the past, the software makes that world appear more complete than the underlying data may justify. Moreover, since the very quality of the data gathered by a scholar or professional working on a site in the contemporary world is, as noted, generally not at issue, the existing software tools do not have functions that allow for the display and secondary analysis of data quality-something that is at the heart of historical research.

What is Needed: an Integrated Suite of Software Tools (i.e., a Software "Machine")

Participants in this group of the Tools Summit have been actively engaged with understanding and attempting to solve specific pieces of the overall problem. The problem of how to do justice to the complexities of time and its representation has been confronted by B. Robertson (HEML: Software facilitating the combination of temporal data and underlying evidence); J. Drucker (Temporal Modeling: Software for modeling not so much time per se as complex temporal relations); and S. Rab (Modeling the "future of the past," i.e., how cultural heritage sites should be integrated into designs for future urban development). The problem of the quality of the data underlying a 2D or 3D representation has been studied by D. Luebke (Stylized Rendering and "Pointworks" [Hui Xu, et al.]: software based on aesthetic conventions for representing incomplete 3D datasets); G. Guidi (handling uncertainty in laser scanning of existing objects); and S. Hermon (using fuzzy logic to calculate the overall quality of a 3D reconstruction based on the probabilities of the individual components of the structure being recreated).

Discussion of these various approaches to this cluster of related issues led us to think of the solution to the problem as entailing not so much a new software tool as a software "machine," i.e., an integrated suite of tools. This machine would allow us to collect data, evaluate them as to their reliability/probability, set them into a time frame, define their temporal relationships with other features of interest in our study, and, finally, represent their degree of (un)certainty by visual conventions of stylized rendering and by the mathematical expressions stated in fuzzy logic or some alternative representation.

THRESHOLD is the proposed name for this software machine. It stands for "Temporal-historical research environment for scholarship." The purpose of THRESHOLD is to provide humanists with an integrated working and display environment for historical research that allows relationships between different kinds of data to be visualized. There are seven goals that this machine needs to fill, listed below. The first five are based on Colin Ware's work on visualization; we added the last two items.

- Facilitating understanding of large amounts of data.

- Perception of unanticipated emergent properties in the data.

- Illumination of problems in the quality of the data.

- Promoting understanding of large- and small-scale features of the data.

- Facilitating hypothesis formation.

- Promoting interpretation of the data.

- Understanding different values and perspectives on the data.

THRESHOLD would function in two modes: authoring and exploring. Two pilot projects that would be suitable as testbeds for THRESHOLD were identified: C. Shifflett's Jamestown and the Atlantic World; and S. Rab's study of the future of cultural heritage sites in Sharjah (U.A.E.).

Conclusion

The Organizing Committee believes that humanities scholarship is expanding and changing both dramatically and rapidly as scholars harness information technology in creative ways. This Summit provided an opportunity for a few of the individuals who are driving that change to discuss it, and to chart advancement for the future. We hope that this report promulgates our observations and conclusions to a wider audience with interest in this profound change.

One premise of the Summit was that revolutionary change in digitally-enabled, humanities scholarship is possible because the "right" computer and communications technology aids permit new kinds of analysis and profoundly interactive collaboration that was not possible before. A second premise of the Summit is that revolutionary change is only possible when the foundational processes of scholarship change. The Summit participants identified four of the processes of humanities scholarship that are being altered and expanded through the use of information technology. These four processes are employed by scholars with or without the support of information technology, that is, they are fundamental to the performance of scholarly study in the humanities

Interpretation

Participants emphasized the centrality of the individual interpreter, while recognizing a need for collaboration. Structured annotations in digital form should be able to incorporate any medium (text, audio and video) and the provenance of the annotation should be recorded with it. The group went on to describe a highlighter's tool that would permit simple, efficient inclusion of annotation represented in any media, and that would simplify the effort to record provenance. This tool needs to encourage clear delineation of the many relationships among elements and the expression of categories of materials. As envisioned at the Summit, information technology can provide the interpreter with powerful tools for complex expression of their interpretations.

Exploration of Resources

"How can scholars ask questions of scholarly evidence?" Because it is possible to query and explore data regardless of its location on the Web, scholars now do so. This expansion of the materials immediately at hand is empowering. But, while the variety and amount of digital representation of materials of interest to scholars is increasing, scholars have concern about their sustained ability to access all such materials. This is not a technical issue, but a social and political one. The proliferation of resources that can usefully be explored gives rise to concerns about how to deal with the heterogeneity of formats of the resources, and how to access the metadata and annotations that enrich the resources. Information tools broaden the reach of the scholar and in some cases increase the speed of access and the selectivity of materials to a degree that the scholar can perform at a higher level.

Collaboration

Historically, most contributions to humanities' scholarship have individual authors. While there is a long history of collaboration, the greatest rewards have gone to those who have worked in isolation. Digital tools enable a new kind of collaboration in which the result of both individual work and collaborative interaction is captured in digital form for others to access. Intermediate and piecemeal thoughts are captured. Collaboration across the nation or globe is equally cost-effective. However, there is strong reliance on digital tools to track the details of collaborative activity and to translate between alternative forms of representation. Standards, data dictionaries and domain specific ontologies are needed to provide a context in which multiple contributors can add information in a coherent way. Participants decided that the most useful single activity to advance such collaboration would be to create a long-lived clearinghouse containing descriptions and objective evaluations of the many domain specific tools that scholars are using. This clearinghouse would inform and educate digital scholars about useful tools. This recommendation is based on recognition that the scholarly environment has become more technical and the quality of work produced is influenced by decisions that are of a technical nature. Scholars need to make wise judgments, and a clearinghouse would aid in those decisions.

Visualization of Time, Space, and Uncertainty

Much of humanities' scholarship involves human habitation of time and space. Information technology offers new ways to represent both. And, even more tantalizing is the question of whether uncertainty and alternatives can be represented as well. Participants coined the phrase "archaeology of fragments" to refer to the record of human habitation grounded in time and space. Juxtaposing realistic and even imagined fragments in a visual way offers a new basis for consideration to most humanists. This group proposes an integrated suite of software tools that go beyond classic and general-purpose Geospatial Information Systems. The suite should support domain specific contexts and should use visualization to facilitate understanding, perception, and hypothesis formation. It should aid the scholar in dealing with the data, highlighting data problems, understanding of large-and small-scale data features and understanding different perspectives on the data.

Research in Scholarship

The development of tools for the interpretation of digital evidence is itself research in the arts and humanities. Tools can be evidence of transformative research as tools can encode innovative interpretative theories and practices. The development of quality, transformative tools should be encouraged, recognized, and rewarded as research.

It is timely for an international effort to develop new tools and transformative interpretative practices. If this effort is inclusive, is reflective, and is supported adequately it will be transformative, enhancing the way the arts and humanities are studied and taught.

Appendix A: Summit Panel Discussions

The Summit addressed issues related to the state of tool design and development. This included the proliferation of new data formats; effective markup language annotation; integration of multiple modes of media; tool interoperability, especially when tools are shared across multiple disciplines; open source for shared and evolving tools; tools with low (easily mastered by an untrained end user) and high (usable only by expert personnel) thresholds of usability; data mining; representation and visualization of data in the geo-spatial framework; measurement; game technology; and simulation.

The Organizing Committee integrated across the accepted issue papers and structured the first morning of the Summit as two sessions of four parallel sessions, each on a different topic. A short description of each topic follows:

Session 1

a)   *Text Analysis, Corpora and Data Mining.*   This discussion focused on the needs for tools for the publication, searching, and study of electronic text collections. It considered the special needs for making corpora available to researchers and how data mining techniques could be applied to electronic texts. Participants also discussed the development of text tools, their documentation for others, and their interoperability.

b)   *Authoring and Teaching Tools.*   Digital technology has profoundly altered writing and teaching in the humanities over the past two decades. Yet, the digital tools used by humanists for authoring and teaching remain surprisingly limited and generic; they are mostly limited to the commercial products offered by Microsoft (e.g., Word, PowerPoint), Macromedia (e.g., Dreamweaver), and Courseware vendors (e.g., Blackboard, WebCT). Participants discussed the kinds of teaching and authoring tools (including ones involved in organizing materials for teaching and writing) that might be created to specifically serve the humanities scholar.

c)   *Interface Design.* Good interface design requires a clear understanding of just what service an information system will provide and how users will want to use it. Information systems to support scholarly activities in the humanities are just now emerging. Participants discussed how to design interfaces that serve both the expert and novice users.

d)   *Visualization and Virtual Reality.* For some years, visualization and especially virtual reality survived and flourished because of the "wow" factor: people were impressed by the astonishing things they saw. But more recently, the wow factor has not been enough. People want to understand what they are seeing; and scholars want to use visualizations to gain new insights, not simply to illustrate what they already know. Hence, tools are critical. We need tools that analyze as well as illustrate. Building a 3D environment is one thing;

using it as the place where experiments can be run to gauge the functionality of a space in terms of heating, ventilation, lighting, and acoustics is more challenging.

Session 2

e)   *Metadata, Ontologies, and Mark up.* These at least three critical topics lurk in the background of every humanities computing project. The panel discussed tools that facilitate the creation and harvesting of structured data about a resource (i.e., metadata), annotation tools (predominantly those for content) to aid in the systematic definition, application, and querying of researcher tag sets (i.e. markup), and ontology tools, to facilitate the specification of entities and their hierarchical relations, and the linking of these entities to idiosyncratic tag sets (i.e. ontologies).

f)   *Research Methods.*   Research methods constitute a core functional requirement for principled tool development. Participants discussed research methods that may (or should) guide the development of future tools for humanities scholarship, having set the context by considering existing tools and the research methods they afford. Of course, available implementation techniques, technological and conceptual, greatly impact the fidelity of implementation, even occasionally keying a substantial evolution of the fundamental research method desired. Finally, the panel discussed possible commonalities of research methods across humanities disciplines that might broaden the audience of scholars for the resulting tools.

g)   *Geospatial Information Systems.*   Disciplines within the humanities are discovering the value of using spatial and temporal systems for recording and displaying data. While Geospatial Information Systems are a well developed technology in other fields, the special tools for mapping and capturing information needed in the humanities are still under development. Tools must be designed for: cultural heritage study; changing boundaries for empires and political divisions; map display of types of metadata categories; archiving and retrieval of images and other data linked to latitude and longitude; and biographical information mapped by place as well as by their relationship to networks of contact between individuals. Spatial tools will be needed for digital libraries and the cataloging and retrieval of large data sets.

h)   *Collaborative Software Development.*   In other disciplines information technology has the greatest impact when the discipline has the capacity to build its own tools. In this regard, the humanities have far to go, but we are beginning to see emergent communities of tool-builders. This session focused on the resources, opportunities, and challenges for collaborative software development in the humanities.