

**Thick Viewing: Integrated Visualization Environments for Humanities
Research on Complex Corpora**

Timothy R. Tangherlini
Todd Presner
Zoe Borovsky

Visualizing digital assets—as well as the relationships between those assets—in humanities collections is a significant challenge. Properly addressed, the visualization of such assets (from hand-written manuscripts, to historical maps, to paintings, to buildings, to archaeological artifacts and beyond) and their interrelationships promises to allow scholars in both the humanities and the humanistic social sciences an opportunity to address long unanswered questions and to begin posing new questions to materials that have otherwise been well-described in traditional scholarship. Further developing various publication models derived from these well-structured systems holds the promise of bridging the divide between traditional print scholarship and more complex models of representation.

One of the main advantages of computing is the ability of machines to discover in a highly automated fashion patterns and relationships between assets in enormous corpora (many tens of thousands to millions of discrete assets) present those relationships in a visually meaningful manner, and allow end-users to make use of intuitive visual navigation to move within the collection. Beyond expanding the research horizon, the visual presentation of complex assets—or corpora of assets—also allows for wider access to otherwise inaccessible materials. Taken together—the ability of computers to discover, visualize and analyze patterned relationships, and the ability of the digital realm to increase access to such corpora with all of the implications of such access—presages a sea-change in scholarship in the humanities and humanistic social sciences. Our presentation here today focuses on two projects—Danish Folklore and Hypermedia Berlin—both of which present challenges related to computing and visualization because of the heterogeneous nature of the sizable corpora related to each project. Our closing remarks will highlight other UCLA projects, including the Digital Encyclopedia of Egyptology, that are equally engaged with these problems.

Danish Folklore

Danish Folklore is based on the enormous nineteenth century archival collections of Evald Tang Kristensen (1843-1929), the most prolific collector of folklore in Europe. The main component of his collections are 24,000 hand written field diary pages containing stories, songs, games, and descriptions of everyday life collected from nearly 7,000 named individuals. Over the course of his collecting career from 1867 until his death—a period that saw the move toward a democratic Denmark, the development of the railways, electricity and the telephone, as well as the motorcar, urbanization and the beginnings of the social welfare state—he also amassed a sizable collection of material items from rural life, corresponded voluminously with well-known intellectual figures including Grundtvig and Ibsen, and took hundreds of photographs. He encouraged others to collect folklore and descriptions of daily life and send them to him, thereby amassing thousands of pages of hand written manuscripts from ministers, school teachers and university students. Apart from collecting, he edited and published editions of his collected stories and songs, after making fair copy of excerpts of the field diaries. At the end of his career, he had published over forty-five volumes of folklore—some indexed, some not. Fortunately, he produced a four volume memoir of all his travels that included vignettes of most of the people he met, all based on his voluminous correspondence with his wife. Unfortunately, these memoirs were not indexed either. At the same time as he was undertaking this massive collecting enterprise—an enterprise in no small part conditioned by a burgeoning Romantic nationalism—the Danish state was deeply engaged in developing elaborate census data, taxation and probate records, and mapping the landscape, while institutions such as the Lutheran church and insurance companies were busy detailing the minutia of people's lives. All of these materials exist in various Danish archives, some in digital form and others not.

In short, the collection is an intriguing example of a remarkably complex humanities corpus—it not only includes the creative and scholarly output of a single individual (in this case Tang Kristensen), but it also includes the creative output of thousands of other individuals. Seen in this context, the collection is far more than simply a bunch of old stories. While Tang Kristensen's correspondence provides an intriguing window into intellectual and political debates of the time, his storyteller's narratives offer a fascinating lens onto changes in the social, political and economic organization of the countryside. The remarkably detailed historical maps produced at the time allow one to trace changes or discover phenomena in the physical environment that often lie at the root of a particular story. Ancillary materials such as census, insurance and church records contribute to the ethnographically thick description that suddenly begins to take shape when these records are placed in proper relation to one another. The biggest challenge of the project is making sense of this vast amount of data, and then structuring it so that computational techniques can help discover meaningful patterns. These patterns in turn can help us discern the complexities not only of traditional expressions and the

politics of their collection, but also the nuances of everyday life in late nineteenth century rural Denmark.

Most folklore collections paint a remarkably one dimensional view of tradition focusing either on “typical” stories organized around themes and genre, or on the endeavors of a single collector. Scant if any attention is paid to the individual storytellers. The result of these standard presentations of folklore is that the complexity of the interrelationships between the collector, the storytellers, the social/political and physical environment, and their stories all disappear. The goal of the Danish Folklore project is to present a series of tools that allows one to visualize these interrelationships, easily navigate and access the underlying archival materials and, ultimately, understand the entire archive in an ethnographically “thick” manner. By connecting the materials to each other so that the original relationships between storyteller, story, collector, and environment are reestablished, the archive comes alive. Similarly, by connecting the collector to his collaborators, interlocutors, critics and family, and by connecting the storytellers to their social and physical environments through maps and state archives, the richness of these individuals’ lives is much easier to comprehend. Tools that allow one to search across storytellers’ repertoires reveal the interconnectedness of both of the storytelling tradition and the storytellers themselves. Finally, by incorporating visualization tools such as mapping, clustering and other word study tools, the archive opens up to new vistas for interpreting the archive.

The current project incorporates several main “views” onto this remarkably complex corpus. As more of the archival assets are digitized, such as Tang Kristensen’s correspondence, more “views” will be added. The three main views onto the archive in the current project are the fieldtrip view, the informant view, and the story view. The fieldtrip view provides a map over one or more of the collecting trips taken by Tang Kristensen—the user can choose to focus on a single trip, or a series of trips, or all of the trips, by selecting map layers that describe the routes. Informants who he visited on these trips appear as icons along the mapped route. Clicking on an icon allows one to explore in greater detail an individual’s life and her folklore repertoire. A description of the fieldtrip from Tang Kristensen’s memoirs is accessible from this view and, ultimately, the correspondence that lies behind the memoir entry will also be accessible.

The informant view is the second main view onto the archive. This view maps not only the informant’s biographical data into the local landscape using historical maps, but also maps all of the stories that he or she told into that landscape as well. From this view, one has access to a list of fieldtrips on which the informant was visited (bringing one back to the fieldtrip view), a biography of the informant and all of the archival material that relates to that informant (including census, church records, insurance rolls, enlistment rolls, and probate records). A link brings one to photographs of the informant and, in a very small number of cases, audio recordings of the informant originally made on wax cylinders. An index of stories, organized by fieldtrip and the order in

which they were told leads to individual story views.

The story view includes the original hand written recording of the story, along with a Danish diplomatic transcription and English translation. The view also includes an image of the published version of the story, along with a Danish transcription and English translation. Finally, the view includes a scholarly annotation that in turn includes standard folkloric indices for the discovery of other story variants, as well as pointers to other variants in the collection. The underlying structure for incorporation and processing of the stories is still being developed. Ultimately, this view will incorporate various discovery tools including clustering, Sammon and Dendro-visualizations and the ability to perform lemmatized searches across texts. A final view into the archive will allow for discovery of informants or stories by place name.

Hypermedia Berlin

Hypermedia Berlin provides a far greater historical sweep than Danish folklore, and incorporates a greater diversity of assets, while it constrains its geographic scope to a particular city. The project (<http://www.berlin.ucla.edu>) is an interactive, web-based research platform and collaborative authoring environment for analyzing the cultural, architectural, and urban history of a city space. Through a multiplicity of richly detailed, fully annotated digital maps connected together by interlinking “hotspots” at hundreds of key regions, structures, and streets over Berlin’s nearly 800 year history, the project brings the study of cultural and urban history together with the spatial analyses and modeling tools used by geographers. While all the historical maps are georeferenced with latitude and longitude in order to perform spatial queries (such as mapping census data or performing *longue durée* comparisons), every map is preserved in its integrity as an epistemological record of the way in which Berlin was perceived, organized, and represented at a given time. The result is that the window or screen never becomes a portal of clarity, realism, or truth. Through the graphical user interface for Hypermedia Berlin, “the data dandy” (Manovich, 270) explores Berlin by zooming in and out of the maps, scrolling—in any order—through some 800 years of time, and clicking on various regions, neighborhoods, blocks, buildings, streets, and addresses. As the navigation is refined—both spatially and temporally—the database populates the search results with relevant media objects, which can, then, be viewed, selected, sorted, and recombined.

Analogous to the process of archaeological coring, then, data searches are bound by place (proximity) and time (duration), not simply keyword: A user might encircle a region extending, for example, fifteen city blocks south of Potsdamerplatz over the years 1920 to 1962. The data objects displayed in the results field are a function of the time-space coordinates determined by the user, what essentially amounts to a contingent narrative told from the database of possible elements. In this regard, Hypermedia Berlin responds to Manovich’s challenge to consider the recursivity of database and narrative: “How can a narrative take into account the fact that its elements are organized in a database?”

How can our new abilities to store vast amounts of data, to automatically classify, index, link, search, and instantly retrieve it, lead to new kinds of narratives?" (Manovich, 237). For the new media flaneur navigating through Berlin, a unique, hypermedia narrative is produced with each iteration, track, or pathway through the time-space database of the city.

But far from an information silo or "read-only" site, Hypermedia Berlin is constructed as a participatory platform with an elaborate tiered authorship component and a community annotation feature for generating content and data sets. Authenticated users—generally, those from the academic community—are able to add any sort of media object as well as select out material for courses and individual research projects. They also author and publish vetted multimedia articles using the resources of Hypermedia Berlin. General users of Hypermedia Berlin are able to add micro-annotations by geo-tagging points, lines, and polygons. The rationale is that micro-annotations contribute to the creation of a "people's history" of the city, leveraging the democratizing possibilities of the web to create, display, and distribute information. These annotations function as "folksonomies," which complement—but do not displace—the academically generated taxonomies or content, which is peer-reviewed and authenticated. Finally, Hypermedia Berlin leverages some of the new possibilities of the geo-spatial web by interfacing between the digital world and the physical environment. Because every object within Hypermedia Berlin is geo-referenced, a person equipped with a hand-held GPS device or even a GPS-enabled phone can both download and upload geo-specific historical information about their precise location. Through such location awareness technologies, a user standing in front of the Brandenburg Gate today will be able to automatically query Hypermedia Berlin for a 1962 picture of the Brandenburg Gate behind the Berlin Wall or view a map of the same location from 1811. The objective is to endow the Berlin of the present with its missing (or invisible) historical dimension. In this regard, the modern metropolis and new media begin to re-interface through a deep-linking dialectic: The metropolis changes new media, and new media changes the metropolis. As a kind of "augmented reality" (or, depending on what "side" one privileges, an "augmented virtuality"), the line separating media and the metropolis becomes blurred as Hypermedia Berlin is built on top, out of, inside of, and throughout the physical space of the city. In the present age of new media, the digital representational platform cannot be separated from the physical, geographic referent. This "new" new media thus moves significantly beyond first-generation web applications and content providers by combining a geo-temporal database with locative technologies, a participatory platform for community generated data, an interface between the digital and the built environment, and robust content created by extending and remixing publicly available interfaces (APIs).

Concluding Remarks: Visual Analytics: Connecting the Dots

Both Hypermedia Berlin and Danish Folklore are projects that attempt to create environments where large and disparate data sets can be presented in an integrated environment that allows users to perform a visual analysis of these materials. Both projects employ geo-spatial technologies--mapping their data onto representations of the material world. The advantages of adopting these tools as a way of exploiting the human eye's "broad bandwidth pathway into the mind" (Rhyme, 2006) for exploring and understanding large amounts of data simultaneously, seem remarkable. These projects support the promise that digital technology will provide new insights into materials that, by their sheer bulk and disparate nature, have not been presented in a way that promotes interdisciplinary scholarship and synthetic analysis.

The challenges facing humanities scholars who want to work in this fashion are strikingly similar to the agenda put forth by the National Visualization and Analytics Center that is funded by the Department of Homeland Security. Their materials cite the attack on the World Trade Center and Hurricane Katrina, as a "wake-up call" for scientists and technologists to formulate and carry out a research agenda for developing what they call "Visual Analytics": defined as "the ability to analyze large amounts of disparate data to make sense of complex situations and save lives." They are developing tools for visualization that will perform the following functions:

1. *facilitate understanding of massive and continually growing collections of data of multiple types;*
2. *provide frameworks for analyzing spatial and temporal data;*
3. *support the understanding of uncertain, incomplete and often misleading information;*
4. *provide user- and task-adaptable guided representations that enable full situation awareness while supporting development of detailed actions; and*
5. *support multiple levels of data and information abstraction, including the integration of different types of information into a single representation.¹*

The reports from the National Visualization and Analytics Center contain a description of challenges similar to the ones these humanities projects face: short or long textual documents in many languages; numeric sensor data; structured data from relational databases; and audio, video and image data.

When we approach our deans to ask for funding for digital humanities, we are sometimes told that the funding must go to disciplines that can promise to save lives. Will taxpayers sleep safer knowing that stories told by Danish peasants are being scrutinized by folklorists? Are our borders more secure for having understood the political, social and cultural consequences of the Berlin Wall?

We may hesitate, for political reasons, to connect the dots between digital humanities and a science that presents itself as a defense against terrorism. Drawing such a comparison, given the sophistication of the tools the NVAC is developing and what we are using, is quite a stretch. However, I

encourage humanists to study the reports, representations and tools being developed by this group and others. It is clear that decision-makers of the information age will rely on these tools and representations to inform their decisions. As humanists we can test similar tools and methods on materials we are familiar with, comparing the results with the outcome of more traditional analysis. In this way we can gain an understanding of how the tools may shape the outcome, leading to a critical assessment of the tools and the representations they produce. We can train ourselves and our students to engage with these tools and become familiar enough with the medium that we can make significant contributions that will shape the discourse of visual analytics in ways that will allow us all to sleep safer.

Endnotes

¹ A visual analytics agenda Thomas, J.J. Cook, K.A. Pacific Northwest Nat. Lab., Richland, WA, USA.

This paper appears in: Computer Graphics and Applications, IEEE

Publication Date: Jan.-Feb. 2006

Volume: 26, Issue: 1

On page(s): 10- 13

ISSN: 0272-1716

INSPEC Accession Number: 8735471

Digital Object Identifier: 10.1109/MCG.2006.5

Posted online: 2006-01-10 20:09:53.0

ELECTRONIC TECHTONICS: THINKING AT THE INTERFACE

Proceedings of the

First International HASTAC Conference
Duke University, Durham, North Carolina

April 19-21, 2007

EDITORS:

Erin Ennis

Zoë Marie Jones

Paolo Mangiafico

Mark Olson

Jennifer Rhee

Mitali Routh

Jonathan E. Tarr

Brett Walters

^{This} conference was organized by the "Interface" Seminar of the
John Hope Franklin Humanities Institute at Duke University as its
contribution to HASTAC's In|Formation Year.