

WORKGROUP REPORT

AHRC ICT Methods Network Workgroup on Digital Tools Development for the Arts and Humanities.

Lorna Hughes, Manager, AHRC ICT Methods Network

Background

On 15 June 2006, the Methods Network convened a group of senior academics, all experienced in the development, support and implementation of projects to develop digital tools in the arts and humanities. Representatives came from a variety of disciplines - library and information studies; musicology; practice-led arts; literature; linguistics; archaeology, and history.

Strategic Issues

Participants identified some key strategic issues that must underpin any programme of digital tools development in the arts and humanities. These include:

- The importance of a **focus on methodology**, and on the arts and humanities methods that are enabled by digital tools. Often there is a focus on 'procedures' rather than a more thoughtful focus on methodologies. The emerging research infrastructure for the arts and humanities needs to be informed by methodologies of usage of digital resources, and tools must be developed that support research practice. The AHDS 'methods taxonomy' was identified as a particularly helpful resource in this respect.
- **Sustainability** of tools over the long term is of critical importance, and must be considered at the outset of a project. The questions concern not just the migrating/upgrading of tools over time and operating system, but also maintaining some kind of support for the end user. If tools development is successful, there will be a community of practitioners, and this community should be sustained over the long term.
- **Interoperability** is equally crucial. There is a need for tools to work across a variety of machines and systems, but also for tools that will work across multiple data formats. This pre-supposes a strategic approach to tools development, one that will support and provide longer viability of the digital knowledge that has been created so far.
- It is important that any project **focus on end users**. Their needs should be evaluated at the outset, to ensure that the tools developed are useful to the user community. Similarly, user interfaces should always be tested by the target user community.
- Tools development is often **collaborative** by nature, across disciplines and subject areas, and often involving arts and humanities academics and computer scientists, engineers and HCI experts, for example. How are such collaborations fostered and supported? How are such partnerships brokered in the first instance? How is this work rewarded and evaluated by the different communities represented?
- **Commercial** relationships may also be essential to this sort of work - again, how are such partnerships to be managed and approached?
- How is the **infrastructure and staffing** required for digital tools development to be provided? How can the arts and humanities creatively break a 'culture of dependency' where we have no support for 'blue skies' thinking and research?

- How can humanities and arts research questions, and the tools that support them, be of benefit for **other subject areas**? Humanities data is complex and non-standard, and tools to analyse this sort of data may well be of interest to other disciplines. How do we begin this conversation?
- There is also a need to find out what **tools already exist**. Often, tools developed by other academic disciplines and subject areas may be applicable to the arts and humanities. A list of existing tools should be built up by the community.

An Arts and Humanities Tools 'Wishlist'

In the final session, participants were asked to come up quickly with one 'wishlist' item - one digital tool that they would like to see developed, assuming that money and resources were no option! The list that emerged is wide ranging, and cross-disciplinary:

a) Tools to automate/assist with tagging and annotation

1. Framework tool for general annotation of research data and information, either on-line or off (i.e. referring to library documents as well as accessing on-line documents directly). Should allow but not be restricted to ontological labelling, and should admit user-defined ontologies with support for consistency checking, etc. Should also allow meta-level annotation of annotations and relationships between them, eg noting dialectic and pragmatic relations, such as rebuttal, support etc, and grouping of relations into arguments. This opens the possibility for visualization and assisted analysis of very complex arguments. In the longer term, it allows comparison between large corpora of data deposited by different researchers using the same approach. Such a tool could be illustrated by a demonstrator tool from any aspect from the humanities.

2. A machine-learning environment that does automatic tagging of predefined document types incorporating several levels of linguistic and semantic annotations in multiple languages. The output should be conformant to generally accepted schemes in the community like TEI and apart from the tagged documents also allow for the generation of several authority lists.

b) Knowledge mining and organization tools

1. Knowledge Crawler for the web: Humanists and other researchers studying culture can use the web as evidence itself, not only as a way to find information. A Knowledge Crawler adapted to cultural study would be a tool that given a set of keywords (like 'Canadian Fiction' or 'Computer Games and Narrative') would crawl the web finding and indexing appropriate web pages. Such a tool would need to have an interface so that humanists could train it by indicating what sorts of pages are appropriate and what pages are not. The crawler would also have an interface for the researcher to study the indexed text database. This might have the following features:

- Latest high probability pages summary that shows recent finds
- Summary diachronic information about the number of pages found per day
- Summary information about location of pages and connections between pages
- Search interface for search the crawled index
- Textual statistics like high frequency words that co-occur with the keywords, names, dates and statistically significant phrases

2. Social Network Analysis: Humanists study networks of people and ideas. They need access to accessible tools for modeling social and intellectual networks. Such a toolset would allow a researcher to enter information about agents and relationships and then be able to study the network.

3. Seriously serendipitous search tool. Would add value not just by replicating, but by going further than the chance finds of traditional stack-browsing by thinking laterally and obliquely, suggesting leads and needs that could not be foreseen at the outset by the searcher

4. Tools to deep mine and explore non-textual materials (moving image, image and sound) to discover patterns, connections etc.

5. Tools to join up and integrate highly distributed and small scale collections of data - this could be based around an extension of OGSA-DAI middleware.

6. Linking tools and resources: tools to assist with creation of marked up documents and semi-automating markup:

i. A general-purpose facility where the user can click on word (or whatever). This would open an entry for that word in a kind of dictionary which would provide more information of various kinds about that word (cf: lexical strategy project, Glasgow; Perseus). What happens after they have clicked on the link depends on the purpose of the project, but it could be added to more intelligent markup, or to help students read the documents, or almost anything else.

ii. Tools to assist with the creation of marked up documents.

These tools would help the document creator mark up things like names of persons and places and time-related information. This would work in a similar way to linguistic markup systems where the creation tool has access to a dictionary of names etc. and also uses some probabilities to attempt to resolve ambiguities.

These would greatly speed up the process of creating useful digital resources, and provide a better foundation for analysis and interpretation of such resources.

c. Visualization Tools

1. Text Visualization: Humanists work primarily with textual evidence and often with large corpora of texts. Visualization promises to let humanists navigate large collections and see patterns. Textual scholars need a toolbox of basic visualization tools (3D, 2D, Centroid, Tree Graph) which can have multilingual text labels for plotted points. These visualization tools should be designed to connect with standard search tools and word frequency tools.

2. Augmented reality toolkit linked to appropriate wearable technologies to enhance experiential research and interpretative methodologies across the arts and humanities.

d) Temporal and spatial exploration and representation

1. Geo-data browser: a placename-finder, which could track the historical change and evolution of placenames, identify spelling and linguistic variants and put placenames in context. This tool would have to be capable of mining data from disparate source materials (maps, charters, secondary sources, etc). It would also have to be able to be linked to mapping co-ordinates so that GIS data could be used.

2. Lexical analysis tools for historical data/placenames, in particular something that is able to do intelligent searches for linguistic variants (e.g., finding the different ways a word has been spelled over a period of 500 years) by using algorithmic searches. Versions of these tools exist in various places, but mostly for modern text. It would be extremely valuable to try to do something which is more general and is applicable to different resources, also perhaps to concentrate a bit on the historical dimension. This would have implications for history, linguistics, language studies, and a host of other subject areas.

3. Large scale artificial life-engines for modeling data/prediction based on existing datasets with visual front ends

e) Mining non-textual resources

1. MSS transcription tools that allow paleographic analysis/forensic graphology - OCR linking of

transcriptions in a standard. Such tools would facilitate the translation process of manuscripts, leaving more time for the scholar to understand their meaning and any cross-relationships between them, or portions of, by the same scribe.

2. Tool which links visual/ graphic media (video) to text resources. A design tool for an integrated interface that can link resources in multiple media types to reveal creative process, especially in relation to 'live' or ephemeral forms such as dance and body-based performance practices. Find ways to develop a suite of tools for this sort of access to existing performance archives. This will enable new possibilities for documentation, analysis and dissemination, particularly for scholars involved with practice-based research in the arts.

3. Better distributed polygonization and decimation engines for surface analysis incorporating the use of fuzzy data sets to improve topology

f) Grid tools

1. Enhanced support for the access grid so that it can be useable, at a generic level, for performance, masterclasses, etc. Grid-type and multicast performances have the potential to allow new ways of thinking about performance, incorporating the limitations of the technology to the composition, production and study of performances. However, there is also tremendous scope to simply improve existing access grid technologies at the desktop level.

2. Improved visualization application links for Access grid technologies to enable enhanced collaborative research networks that incorporate real time analysis across the Access grid

g) Capturing processes

1. Tools for documenting workflows for tools and resources

2. Tools to capture and document the creative and research processes, and that are able to incorporate the interpretive and critical processes that are the essence of research in the arts and humanities. This should include tools that can capture the process of making a work (e.g. of dance, or art).