Wandering Lonely as a Cloud....
Arts and Humanities, Clouds,
Crowds and Seamless Infrastructures

Sheila Anderson
Centre for e-Research,
King’s College London
ISGC 2011, Taiwan
Clouds and Crowds - Wordsworth got there first.....

I wandered lonely as a cloud
That floats on high o'er vales and hills,
When all at once I saw a crowd,
A host, of golden daffodils;
Beside the lake, beneath the trees,
Fluttering and dancing in the breeze.
Humanities Research Practices

• Hermeneutic rather than experimental
• Narrative, textual, rhetorical
• Not seeking formal laws and explanations
• Recursive, not linear – a constantly questioning process
• Deep reading / reasoning of sources
• Individualistic, even when collaborating
• Rooted in disciplinary understandings and approaches
Humanities ‘Data’

- Sources rather than ‘data’ – the human record
- Texts of all kinds, images of all kinds
- Objects, artefacts
- Video, film, audio (speech, music)
- Publications, grey literature
- Complex, varied, multi-faceted – mix of digital and analogue
- Highly distributed in multiple locations
Digital Humanities - evolution

• Same practices, different environment
• A gradual move into the digital
• Responding to digital content and opportunity to do ‘stuff’ quicker
• Understanding scholarly primitives
  – Unsworth: discovering, annotating, comparing, referring, sampling, illustrating, representing
  – DARIAH: discovering, collecting, comparing, delivering, collaborating
Digital Humanities - transformation

• Changing practices; converging practices
• Dealing with scale
• Space, place and time
• Language and understanding
• Social media, social networking

• But remember humanities research context – individualistic, complex, hermeneutic
DARIAH: Digital Research Infrastructure for the Arts and Humanities

• EU FP7 Infrastructure project
• Preparatory phase funded from October 2008 to March 2011
• Construction phase starts now with support from around 10 European countries with more joining
• Applying for an ERIC soon
• Building systems and services that support evolutionary and transformative activities
Some examples

• General purpose VRE: D4Science and gCube for the humanities, the gMan project, King’s College London
• Spatial Humanities: GIS in Literary Studies, University of Lancashire, Indiana University Purdue University, Indianapolis
• FP7 Infrastructure project - European Holocaust Research Infrastructure: collaboration with archives and researchers
• Collaborative across disciplines and commercial : 1641 Depositions, University of Aberdeen, Trinity College Dublin, IBM
• Storage clouds – go listen to Jens Jensen talk about Kindura.....
General Purpose VRE

• Support research processes using digital content, tools and services
• Support the whole of the research life-cycle
• Provide access to collections, resources and tools
• Researcher centred: allow for interpretation, recursive thinking and annotations
• Mechanisms for managed collaboration and sharing
Diligent, D4Science and gCube

• Funded by FP7 – aims to create e-Infrastructure Ecosystems
• gCube: A Virtual Research Environment (VRE) Management Framework
• A testbed e-Infrastructure built upon grid middleware.
• Allows access to heterogeneous, distributed technologies, services and content and guarantees interoperability of these resources
• Based on shared local computation, storage and generic service and from EGI
Use Case for Classics

• HGV, a database of metadata for Greek papyri
• Project Volterra, a database of Roman legal texts
• Inscriptions of Aphrodisias (InsAph), a corpus of about 2,000 ancient Greek inscriptions
Plus additional content

• Supplemented with additional resources with stable URLs:
  • Pleiades catalogue of ancient place names
  • Lexicon of Greek Personal Names
  • American Numismatic Society’s coin collection

• Each URL that resolves to a systematic representation of the corresponding resource.

• Not imported: used as links from main datasets

• Mirrors humanities scholar’s use of multiple sources
Researchers were able to:

- Assemble virtual collections of documents (research objects)
- Search across collections (text/date/...)
- Add annotations to (parts of) objects
- Add (annotated) links between objects
- Search across annotations/links
- Share material with selected colleagues
user’s workspace

collection within workspace
Next steps

• Improve interface for use by non-experts – import particularly difficult
• Explore how access to and use of tools for further processing of content could be included e.g. to normalise spellings
• Integrate more data resources (including those managed externally to gCube)
• Extend humanities VO to other groups of humanities researchers
• Publication of data from VRE
Spatial Humanities

classified space on culture and society.

• Embraces all spatial technologies but bends them toward the humanities.

GIScience

(David Bodenhamer, Indiana University Purdue University, Indianapolis)

(David Bodenhamer, Indiana University Purdue University, Indianapolis)
Spatial Humanities

GIS in literary studies

- Early tours of the Lake District:
  - T. Gray (1769): Proto-Picturesque
  - S.T. Coleridge (1802): Romantic

"Distant reading" with density smoothing

RA funded project with D. Copper and S. Bushell
Technology and Tradition

Close Reading with Google Earth

http://www.lancs.ac.uk/mappingthelakes
Multiple Data Sources

Integrating other data

Altitude of mentions

Population density
Volunteer Information

To Flickr...

http://www.flickr.com
Back to the texts…
Immersion

Display medium +

Spatial Story Board #

Users

Qualitative GIS

Notes:
* Versioning; Topological network
  Expert system data mining
# Natural Language, Object oriented
  Scale-time, ESDA
+ Brushing and linking enabled
  Deep mapping and deep contingency

Geospatial semantic web integrator

GIS
SOM
Tree maps
Animation

Tag clouds
Game engines

User generated content/Spatial Databases
Data:
Text, oral history, spatial data, photographs
primary sources, sound, music, video, sketches,
maps, narratives – UGC/Sp
European Holocaust Research Infrastructure

• To build an infrastructure that supports next generation Holocaust research

• A large European consortium including experts from Holocaust research, Archives, Research Infrastructures

• A domain that is in many ways most advanced in terms of digitisation and online presence

• How can we enhance traditional archives-based research in the Humanities using research infrastructures
About the Project

Language and Linguistic Evidence in the 1641 Depositions is an AHRC-funded, multidisciplinary project that aims to develop new ways of interacting with a digitized corpus of Early Modern English witness testimonies. The 1641 Depositions comprises approximately 4,000 depositions or 20,000 pages of newly-digitized and transcribed witness testimony, originally collected by government-appointed commissioners, regarding the rebellion by the Catholic Irish in 1641. These statements constitute the chief evidence for the sharply contested allegation that the rebellion began with a general massacre of Protestant settlers, and they have been central to protracted and bitter historical dispute.

Using a suite of innovative software designed for both linguistic analysis and visualisation of results we are interrogating the corpus about a range of linguistic issues. In cooperation with IBM we have adapted IBM's
The 1641 Depositions project aims to explore how a computer environment can be created in which scholars interested in historical and corpus linguistics can work collaboratively with historians and other specialists to interrogate a key historical source, the 1641 Depositions, in ways not currently possible, by exploiting effective language technology developed by IBM LanguageWare.

images © James Cranford, *The Teares of Ireland* (London, 1642)
Corpus Analytics Tools

Use Voyeur Tools to explore the entire Depositions corpus.

Click here for a guide to using Voyeur Tools

Voyeur Tools

Learn more at the Voyeur website

Use an interactive Word Tree visualisation to explore the content of each of the Deposition collections, and discover all the different contexts in which the word or phrase appears.
Summary

• We are on the way to data-driven Humanities
• We have the tradition of the Digital Humanities
• We need
  – Infrastructures
  – But mainly ways of working with the data
• We need successful collaboration between CS and Humanities, and between humanities disciplines and others
• European (ESFRI) and international collaboration offers many opportunities
Further Information

- **Sheila.anderson@kcl.ac.uk**
- **Spatial Humanities:**
  - Ian Gregory, University of Lancashire
  - David Bodenhammer, Indiana University Purdue University, Indianapolis
- **gCube VRE:**
  - tobias.blanke@kcl.ac.uk
  - Mark.hedges@kcl.ac.uk
- **CLRLE 1641 Depositions Project:**
  - http://kdeg-vm-15.cs.tcd.ie/omeka-1.2.1/about