TEXTvre: embedding a VRE in an institutional context

Mark Hedges, Tobias Blanke
Centre for e-Research,
King’s College London, UK

ISGC 2010, Taipei, Taiwan – 10th March 2010
A message for our funders

• Funded by JISC (Joint Information Systems Committee)
• Part of their VRE (Virtual Research Environment) programme
• Runs for 2 years 1st April 2009 – 31st March 2011
Starting Point

Collaborative environment for textual scholarship, using German Grid infrastructure

Research practices (to be) supported by institutional infrastructure

ISGC 2010, Taipei, Taiwan – 10th March 2010
Institutional Background

• Research context
  – Digital Humanities
  – Textual scholarship, online editions, TEI XML
  – Collaborations between scholars and technology experts (and other institutions)

• Infrastructural context:
  – Institutional repositories
  – Preservation environment
  – Research management (REF/CRIS)
Integration and embedding

- These contexts are quite separate
- Integrate with institutional infrastructure
- Embed within the day-to-day work practices of the researchers
Institutional infrastructure

- Researcher work area, UI, tools, services
- TextGrid middleware
- Institutional repositories
- Data curation and preservation
Researcher practice

• Scholars do their own thing, so ...
• Explore current and potential research practices and use of tools
  – Semi-structured interviews
  – Tool walk-through (what could they do?)
• Integrate tools to support and enhance these practices (current and potential)
Research Partners

- Early English Laws
  - Heterogeneous sources
  - Distributed researchers

- Inscriptions of Roman Tripolitania
  - Established editing tradition
  - Material sources

- Gascon Rolls
  - Records editing tradition
  - Entity analysis
New Horizons ...

- **GATE**
  - Information Extraction
  - Text mining
  - Annotation

- **OCRopodium**
  - OCR (Optical Character Recognition)
  - Text extraction from digitised images
  - Editing integration
GATE
OCRopodium

OCRopodium is a tool for Optical Character Recognition (OCR) developed by the Centre for e-Research (CeRCh) at King's College London. It allows users to upload images and receive the text content extracted from those images.

ISGC 2010, Taipei, Taiwan – 10th March 2010
More institutional integration

archivists

ocropus

TEXTvreme

repository

scholars

ISGC 2010, Taipei, Taiwan – 10th March 2010
To summarise

- Dispersed scholars working on diverse (digital) humanities projects
- Develop a VRE that is embedded in the day-to-day research activities of scholars
- Integrate VRE with institutional infrastructure(s): repositories, preservation, archives/library
- Provide integrated framework for dealing with (text-based) historical and archival material
Contacts

http://textvre.cerch.kcl.ac.uk/

mark.hedges@kcl.ac.uk
tobias.blanke@kcl.ac.uk