Deploying an on-demand VRE for linking data sets in the Humanities

Tobias Blanke, Mark Hedges, Mike Priddy (KCL); Fabio Simeoni (Strathclyde); Leonardo Candela (ISTI-CNR)
Overview

• Aim: to demonstrate a general purpose data-centric VRE for humanities research
• D4Science and gCube
• gMan Experiences
• Demonstrating the Demonstrator
• Our questions for you!
On-Demand VRE Infrastructure

• Why are VREs important for Digital A&H research?
  • Data-driven research
  • Diverse data sources: heterogeneous, rich & complex
  • A move towards international collaborations
  • An identified need to manage data from research on the scholarly process
Aims for a general VRE

• Support the whole of the research life-cycle for A&H
• Provide access to collections, resources and tools
• Researcher centred
• Mechanisms for collaboration on a team as well as community level
D4Science & gCube

• D4Science (& D4Science-II)
• EU e-Infrastructure projects, co-funded by Seventh Framework Programme
• Aim of interoperation of data e-Infrastructures to create e-Infrastructure Ecosystems
• **gCube:**
  • A testbed e-Infrastructure built upon grid middleware.
  • Allows access to heterogeneous, distributed technologies, services and content and guarantees interoperability of these resources.
  • VREs: based on shared local computation, storage and generic service and from EGEE.
• A Virtual Research Environment (VRE) Management Framework

• A Virtual Organisation (VO) models sets of users and resources belonging to a e-Infrastructure.
  • What is shared
  • Who is allowed to share
  • The conditions under which sharing occurs
gCube data model primitives

information object
typed relationships between IOs
Data model derived entities

Can define other concepts in terms of these, e.g.:

Collections
Metadata
Annotations
Associations
Indexes
Importing data into gCube

• Importing resources into gCube involves:
  • Describing the resources and their relationships in terms of the gCube information model
  • Importing the resources based on this description
• Custom scripting language used
• Content does not need to be imported
It is not only the data but it is the tools that make it.
Researchers & Use Cases
Data Sources

• 3 Heterogeneous Data Sources:
  • Heidelberger Gesamtverseichnis der Griechischen Papyrusurkunden Ägyptens (HGV) metadata for Greek papyri (in TEI EpiDoc)
  • Project Volterra: database of basic texts of Roman imperial legal pronouncements (XML dump from MySQL)
  • IAph: corpus of inscriptions from Aphrodisias (in TEI EpiDoc)
Data Sources

• Why these data sets:
  • Overlapping the same historical time period of 1st to 4th Century AD
  • Location and people may be mentioned across the texts.
  • Data sources used previously in LaQuAT demonstrator
Other data

• Catalogues of ”things” 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 ”thing”.

• Not imported: used as links from main datasets
Use cases

• What do our demo researchers do?
  • Assemble virtual collections of documents (research objects (ROs))
  • Search across collections (text/date/...)
  • Add annotations to (parts of) ROs
  • Add (annotated/typed) links between (parts of) ROs
  • Search across annotations/links
  • Share material with selected colleagues
gMan data import

• Document entities correspond to the intellectual objects with which we are dealing: inscriptions, papyri, texts.

• Compound documents built up from lower-level IOs: transcription files, images, metadata objects.
Demonstrator LIVE!
Experiences

• Data import procedures an issue
  • Carried out by specialist D4Science staff
  • Required extensive contact between D4Science staff and researchers staff
• UI requires further development
Next steps

• Integrate more data resources (including those managed externally to gCube)
• Import collections with geo-spatial co-ordinates
• Extend humanities VO to other groups of humanities researchers
• Publication – links to journal articles, linked data
Contacts

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

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

technical.wiki.d4science.research-infrastructures.eu/documentation/index.php/
GCube_Wiki
www.d4science.eu/about
laquat.cerch.kcl.ac.uk/