



# From research data repositories to VREs: a case study from the humanities

Mark Hedges, Tobias Blanke, Mike Priddy (King's College London)

Fabio Simeoni (University of Strathclyde)

Leonardo Candela (ISTI-CNR)



# A message for our sponsors

Project gMan

Funded by JISC (Joint Information Systems Committee) in UK



Part of their VRE (Virtual Research Environment) programme

A “rapid innovation” project

6 month duration, runs until end July 2010

Builds on initial work from DARIAH



# Overview

Aim: to demonstrate a general purpose  
data-centric VRE for humanities research

Context: scholars, data, research

D4Science and gCube

Demonstrator

Experiences

# Context

# Libraries and Archives



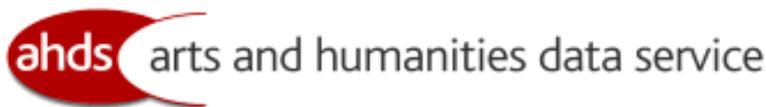
## East London Theatre Archive

Nederlands Instituut  
voor  
Oorlogsdokumentatie  
NETHERLANDS INSTITUTE FOR WAR DOCUMENTATION



## PERSEUS DIGITAL LIBRARY

GREGORY R. CRANE, EDITOR-IN-CHIEF  
TUFTS UNIVERSITY



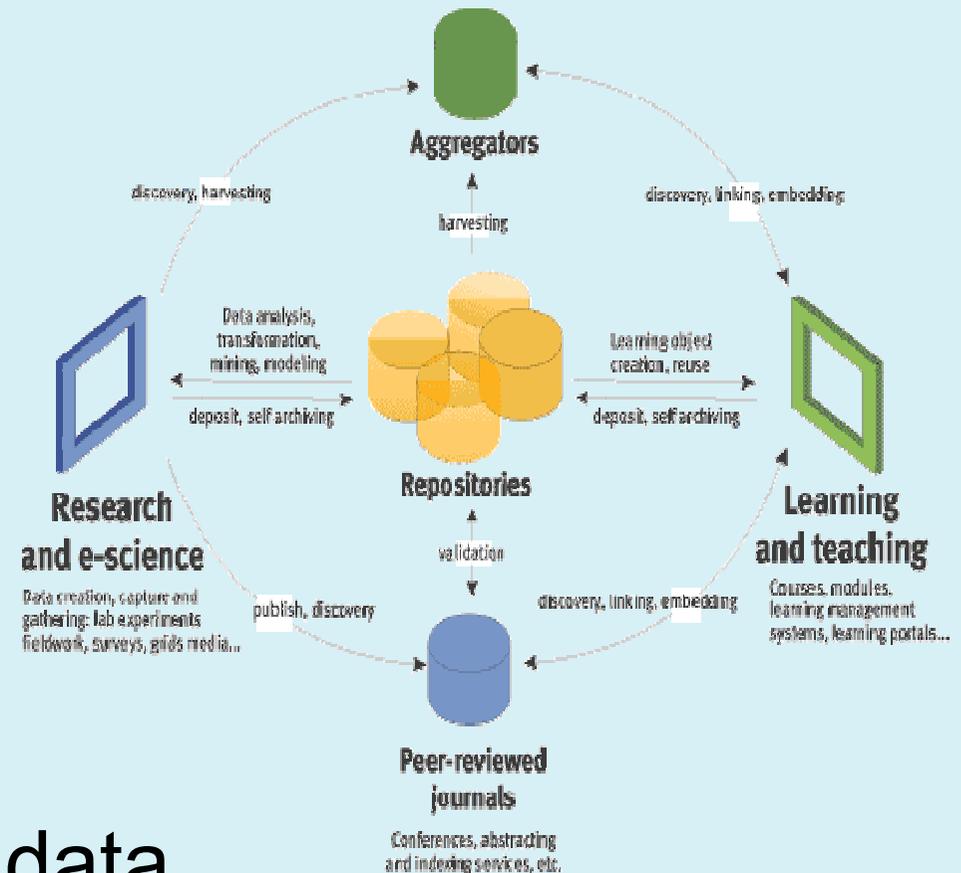
# 'Data'-driven Humanities

Data Centric Research: large, rich, and complex

Design interaction around data  
Scholarly lifecycle perspective

Dave de Roure: New  
e-Science Keynote

## Enable working with data



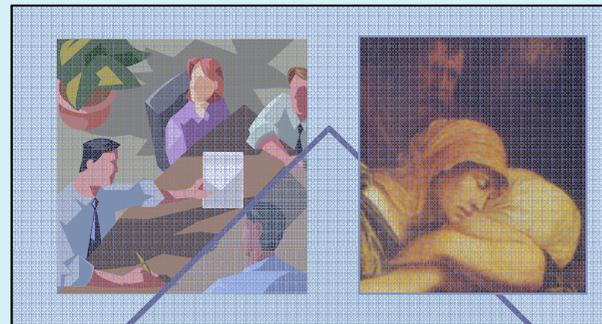
# Virtual Research Environments



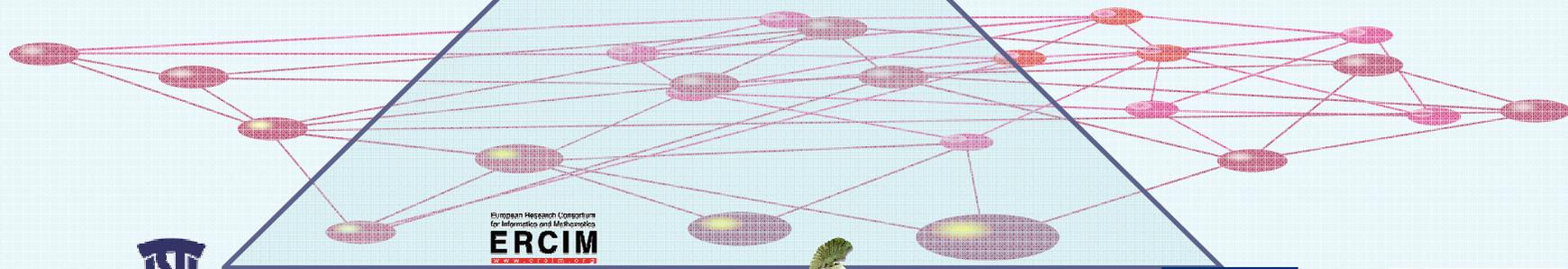
# D4Science and gCube

# DILIGENT and D4Science: From Digital Libraries to VREs

Community  
cooperation



VREs: based on  
**shared** local  
computation, storage  
and generic service  
and from EGEE



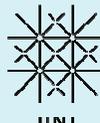
ISTITUTO DI SCIENZA E TECNOLOGIE  
DELL'INFORMAZIONE "A. FAEDO"



European Research Consortium  
for Informatics and Mathematics  
**ERCIM**



NATIONAL & KAPODISTRIAN  
UNIVERSITY OF ATHENS



# A **Virtual Research Environment** (VRE)

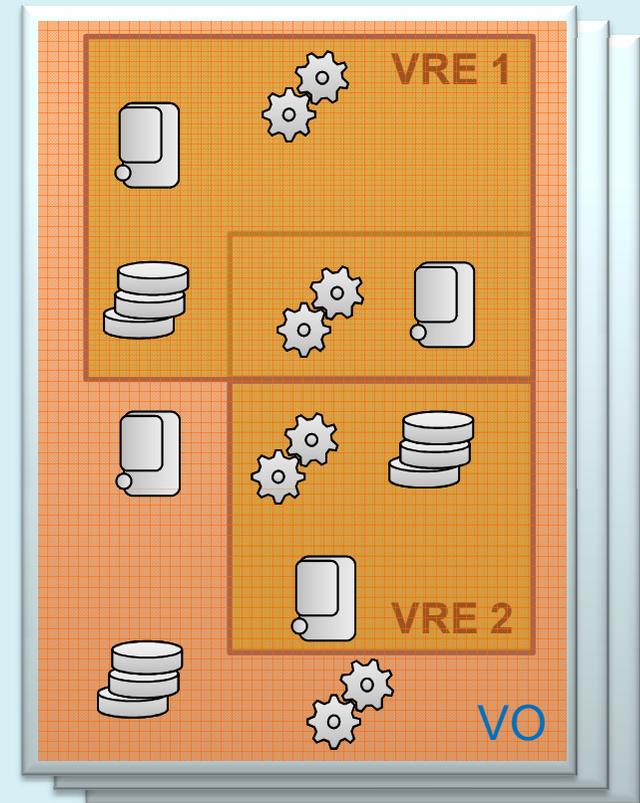
Management Framework

A **Virtual Organization** (VO) models sets of users and resources belonging to a e-Infrastructure.

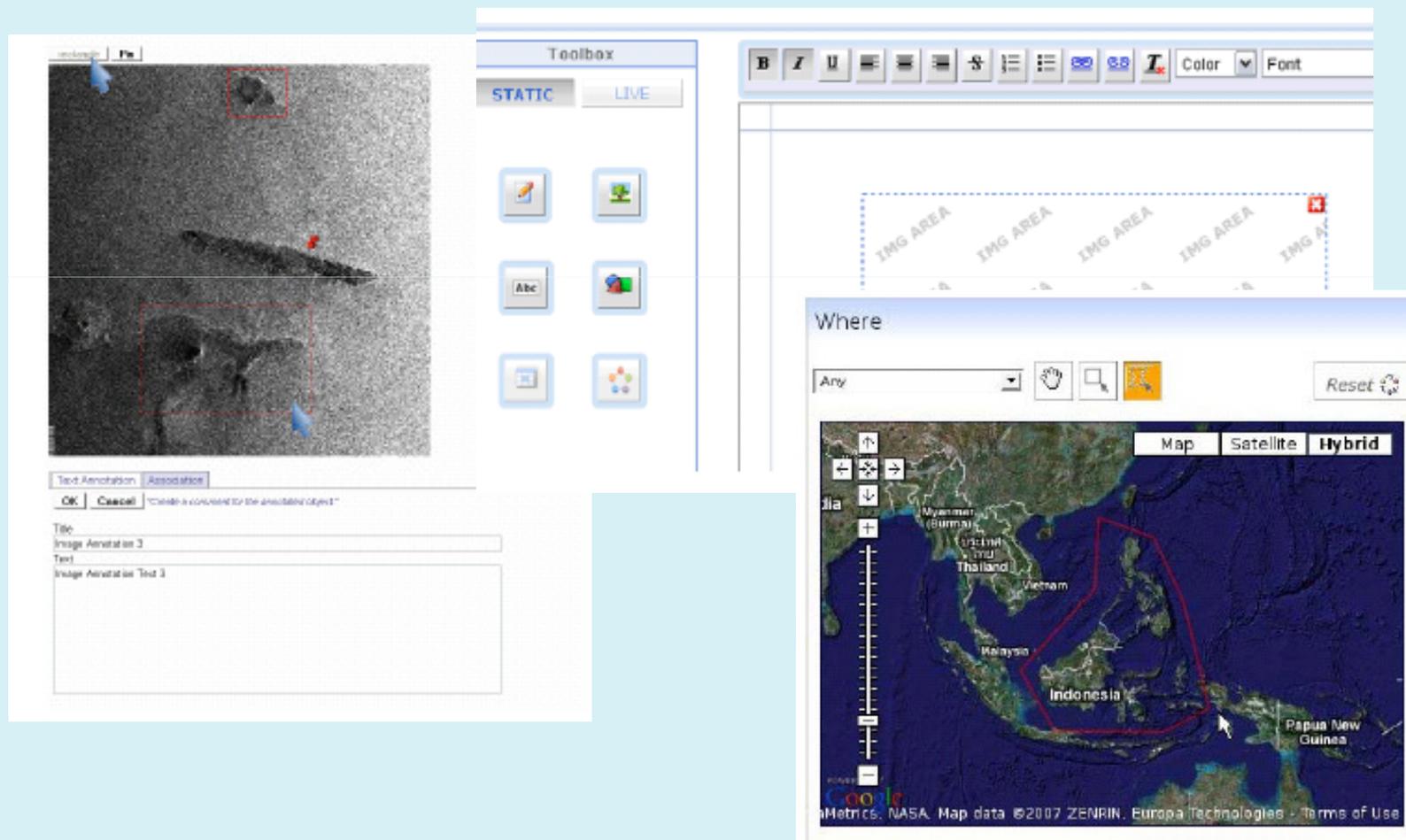
What is shared

Who is allowed to share

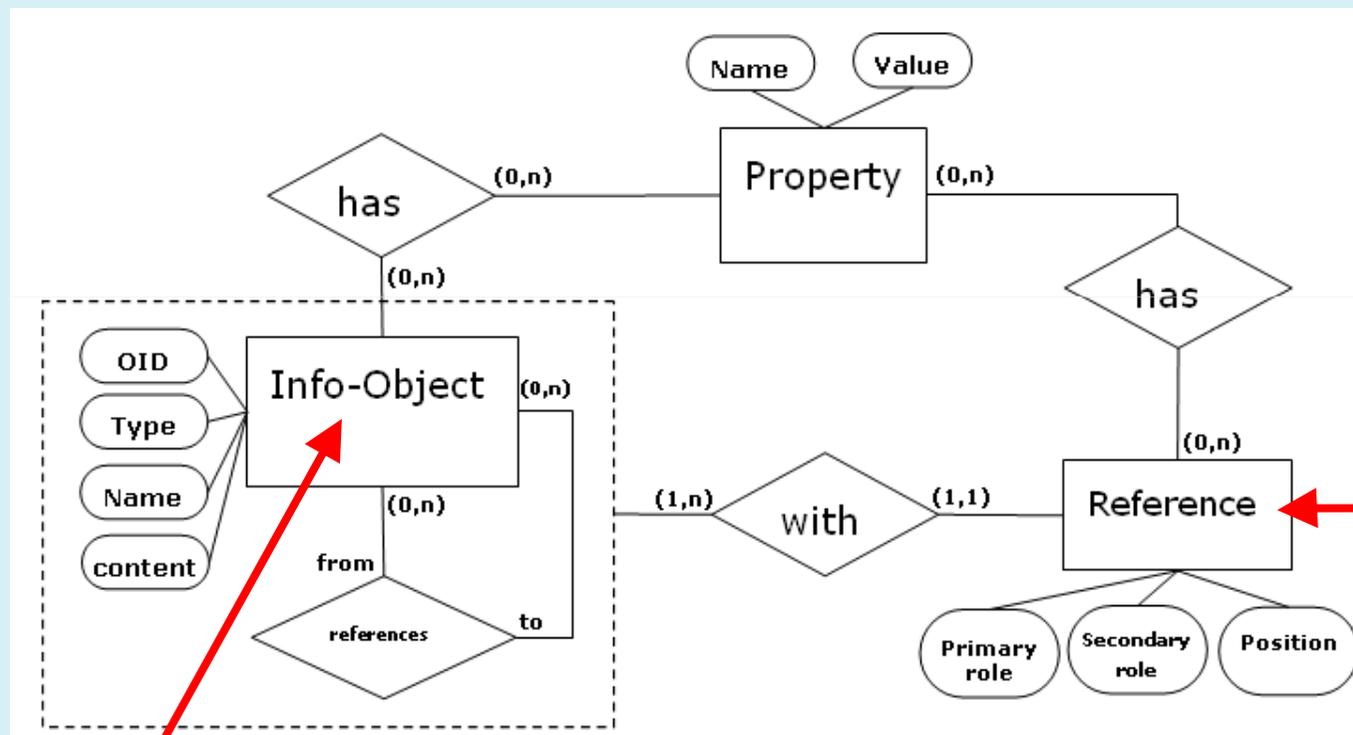
The conditions under which sharing occurs



# It is not only the data but it is the tools that make it



# gCube data model primitives



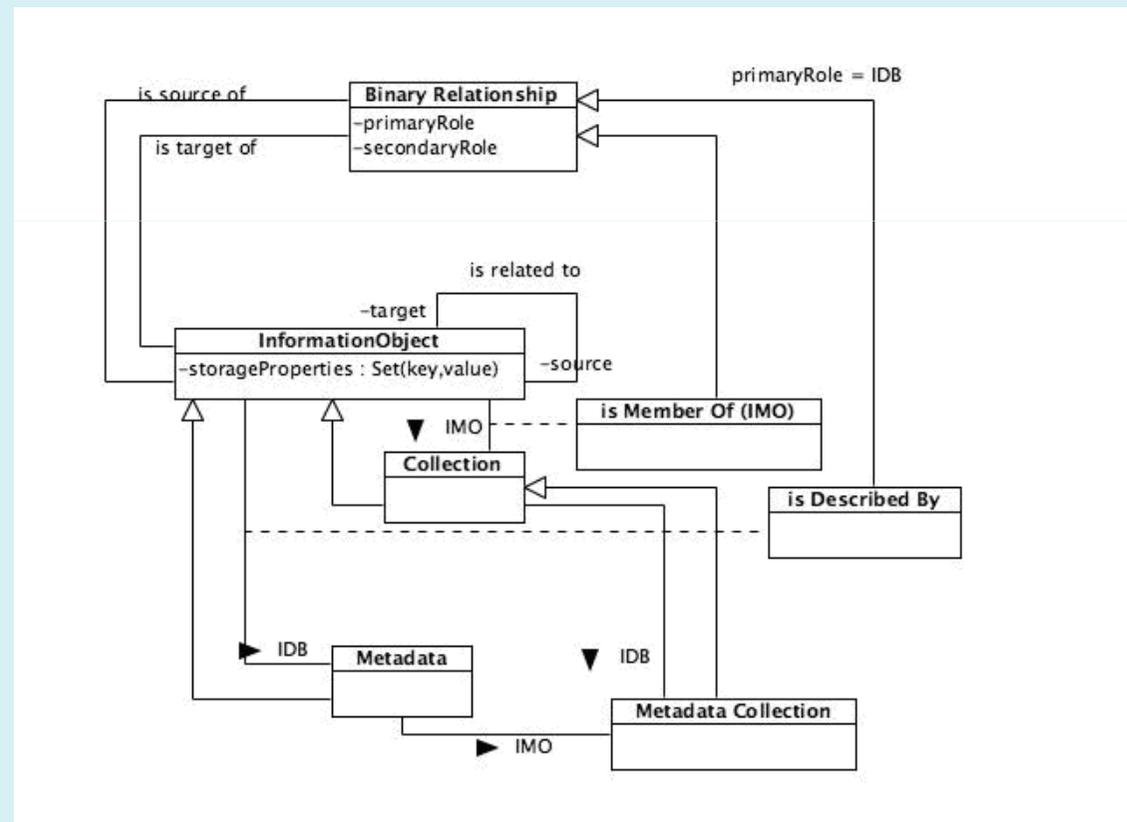
information object

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

Home Collections Management

Information Space Editor FT Index Geo Index Fwd Index Data Transformation IR Boot Strapper **ArchiveImport** IR Bootstrapper

Archive Import Service Interface

**Scripts**

- Inscriptions\_of\_Aphrodisias
- IOA\_DEMO\_import
- ScriptDemo

**Edit Script** Monitor Script Execution

**title:**  
Inscriptions\_of\_Aphrodisias

**description:**  
Inscriptions\_of\_Aphrodisias import script

**script content:**

```

collection contentcollection = collection::content["Inscriptions_of_Aphrodisias_content"]{
  collectionName = "Inscriptions of Aphrodisias",
  isUser=true
};

collection metadatacollection = collection::metadata["Inscriptions_of_Aphrodisias_metadata"]{
  collectionName = "Inscriptions of Aphrodisias",
  collectionDescription = "EA collection of 1500 inscriptions and associated metadata found in
the city of Aphrodisias (modern South West Turkey), or in its civic territory, up to the end of 1994. The
inscriptions are written in Greek, occur mostly on stone, and are mostly dated between the 1st and the
5th century AD.",
  isUser=true,
  isIndexable=true,
  metadataName="Epidoc",
  metadataLanguage="en",
  metadataSchemaURI="http://insaph.kcl.ac.uk/iaph2007/xml/tei-epidoc.dtd"
};

relationship rel= relationship::metadata(metadatacollection, contentcollection)
["metadatarel"+"Inscriptions_of_Aphrodisias_content"+"Inscriptions_of_Aphrodisias_metadata"]{};

print("getting the file list");

```

input is syntactically valid

Save

New Script Delete Script

Data Import script



```

file xmlFile=getLocalFile(getFile(fileUrl));

//create the Information Object, passing as content URL the file URL
resource content = resource::content[fileUrl] in contentcollection{
    isVirtualImport=false,
    contentSourceLocator=fileUrl,
    documentName= title,
    hasContent=true,
    hasMaterializedContent = false,
    isLargeFile=false,
    mimeType="text/xml"
};

print(" Created content object "+count+" with content "+fileUrl);

string metadataContent = toString(xmlFile);
metadataContent = replace(metadataContent, "<!DOCTYPE TEI.2 SYSTEM \"tei-epidoc.dtd\" >","");

//create the Metadata Object, passing as content the file
resource metadata = resource::metadata["metadata_"+fileUrl] in metadatacollection{
    content = metadataContent
};

//make a relationship between the content object and the metadata object
relationship rel= relationship::metadata(metadata, content)["metadatarel_"+fileUrl]{
};

print(" Created metadata object "+count);

//parse the content file looking for all tags containing the href property
list figures=xpath(xmlFile,"//figure[@href]/@href");

foreach j in [0 to listsize(figures)-1]{

    //extract the href value
    string figure=toString(figures[j]);
    print(" Figure: "+figure);
    //create the image url
    string imageUrl = imagesprefix + figure + ".jpg";
    print(imageUrl);

    //create the object part setting as content URL the image URL
    resource part = resource::content[imageUrl] {
        isVirtualImport=false,
        contentSourceLocator=imageUrl,
        documentName=figure,
        hasContent=true,
        hasMaterializedContent = false,
        isLargeFile=true.
    }
}

```

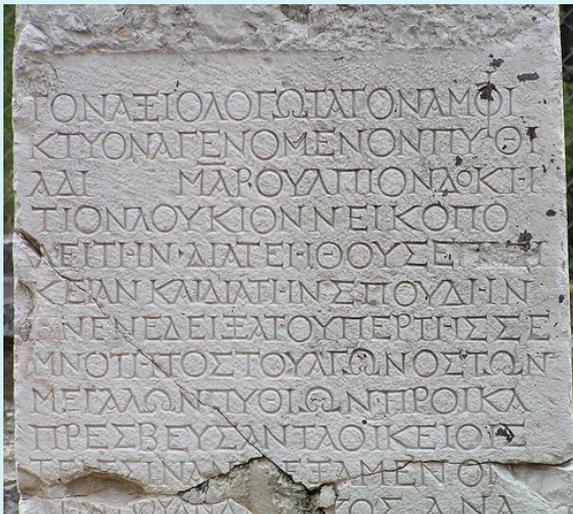
# Researchers and Use Cases

# Data

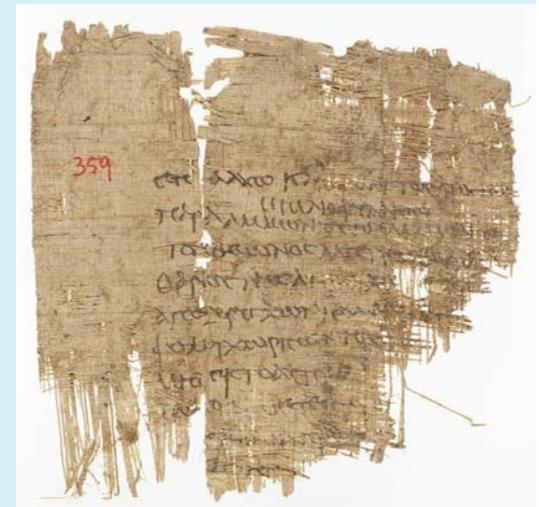
HGV, a database of metadata for Greek papyri

Projet Volterra, a database of Roman legal texts

Inscriptions of Aphrodisias (InsAph), a corpus of about 2,000 ancient Greek inscriptions



```
--<div type="description" n="date">
<head>Date</head>
--<p>
<date notAfter="0014" notBefore="-0050" exact="none">
(
<rs type="criteria">lettering</rs>
<rs type="criteria">content</rs>
)
</p>
</div>
--<div type="edition" lang="grc">
<head lang="en">Edition</head>
--<ab>
<lb n="1">
--<w lemma="οὔτος">
<supplied reason="lost" cert="low">οὔτος</supplied>
</w>
--<w lemma="ὁ">
<supplied reason="lost">ὁ</supplied>
</w>
--<w lemma="τόπος">
<supplied reason="lost">τόπος</supplied>
<unclear reason="damage">ς</unclear>
--</ab>
```



# 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 researchers do?

- 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

# Demonstrator



Search Results **Annotation**

Annotations

Title: Boundary marker, set up by C. Julius Zoilos

Content Type: text/xml

```

- <titleStmt>
  - <title level="m">
    <rs type="textType">Boundary marker</rs>
    , set up by C. Julius Zoilos
  </title>
</titleStmt>
- <publicationStmt>
  - <p>
    - <bibl>
      - <editor>
        Joyce M.
        <name type="surname">Reynolds</name>
      </editor>
      <date>1982</date>
    </bibl>
  </p>
  - <p>
    Creative Commons licence Attribution 2.5 (
    <xref>http://creativecommons.org/licenses/by/2.5/</xref>
    )
  </p>
</n>

```

Text Relate Upload

Title

Object

- Honours for [Julius] Aurelius Charidemos
- Honours for [Julius] Aurelius Charidemos
- Julius to Cyriacus
- Honours for C. Julius Philippos,
- Career of Tiberius Julius Alexander
- Building dedication by C. Julius Zoilos**
- Edikt des Tiberius Julius Alexander vom Jahre 68 n. Chr.

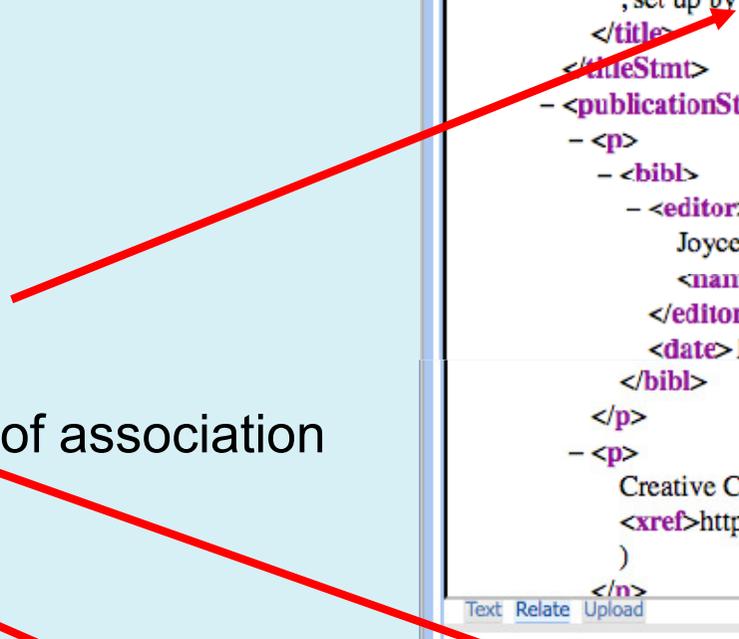
Home | D4Science 2008-2011 | Privacy

01 July 2010

source

nature of association

target



user's workspace

The screenshot shows a web interface for a digital repository. At the top, there are navigation tabs: Home, Search, **Workspace**, Reports, and Profile. Below the tabs, the page title is "Workspace" with a sub-tab "Annotation". The address bar shows the path: /root/test set/Honours for C. Julius Philippos,.

On the left, a "Workspace" sidebar shows a tree view with folders: root, My Default Basket, and test set. The "test set" folder is expanded, showing a list of documents. A red arrow points from the "test set" folder in the sidebar to the "collection within workspace" label at the bottom.

The main area displays a table of documents:

Name	Type	Owner	Creation Time
<input type="checkbox"/> Honours for [Julius] Aurelius Charidemos	Document	Mike.Priddy	15/06/2010 11.21
<input type="checkbox"/> Julius to Cyriacus	Document	Mike.Priddy	15/06/2010 11.21
<input type="checkbox"/> Honours for C. Julius Philippos,	Document	Mike.Priddy	15/06/2010 11.21
<input type="checkbox"/> Career of Tiberius Julius Alexander	Document	Mike.Priddy	15/06/2010 11.21
<input type="checkbox"/> Building dedication by C. Julius Zoilos	Document	Mike.Priddy	15/06/2010 11.21
<input type="checkbox"/> Edikt des Tiberius Julius Alexander vom Ja	Document	Mike.Priddy	15/06/2010 11.21

At the bottom, a "Details" panel is visible, showing "Item details" and "Document info".

**Item details**

Name:	Honours for C. Julius Philippos,	Owner:	Mike.Priddy
Description:		Creation Time:	15/06/2010 11.21
		Dimension:	5 kb

**Document info**

OID:	55070ce0-7344-11df-8b1e-c4e1e46ec70
Collection:	Inscriptions of Aphrodisias
Schemas:	Epidoc_en

Buttons: Show Metadata (1), Show Annotation

At the very bottom of the interface, it says "User workspace area loaded."

collection within workspace

collection

query

URL of object

The screenshot displays the d4science portal interface. On the left, a search results list is visible under the heading 'query result set'. The main area shows a detailed view of an 'Object' (image) with a 'ContentViewer' window open. The viewer displays a handwritten document with Greek text and a list of metadata. A red arrow points to the URL in the browser's address bar, labeled 'URL of object'. Another red arrow points to the 'Object' metadata section, labeled 'image object in viewer'. A third red arrow points to the search results list, labeled 'query result set'. A fourth red arrow points to the search bar area, labeled 'collection'. A fifth red arrow points to the search input field, labeled 'query'.

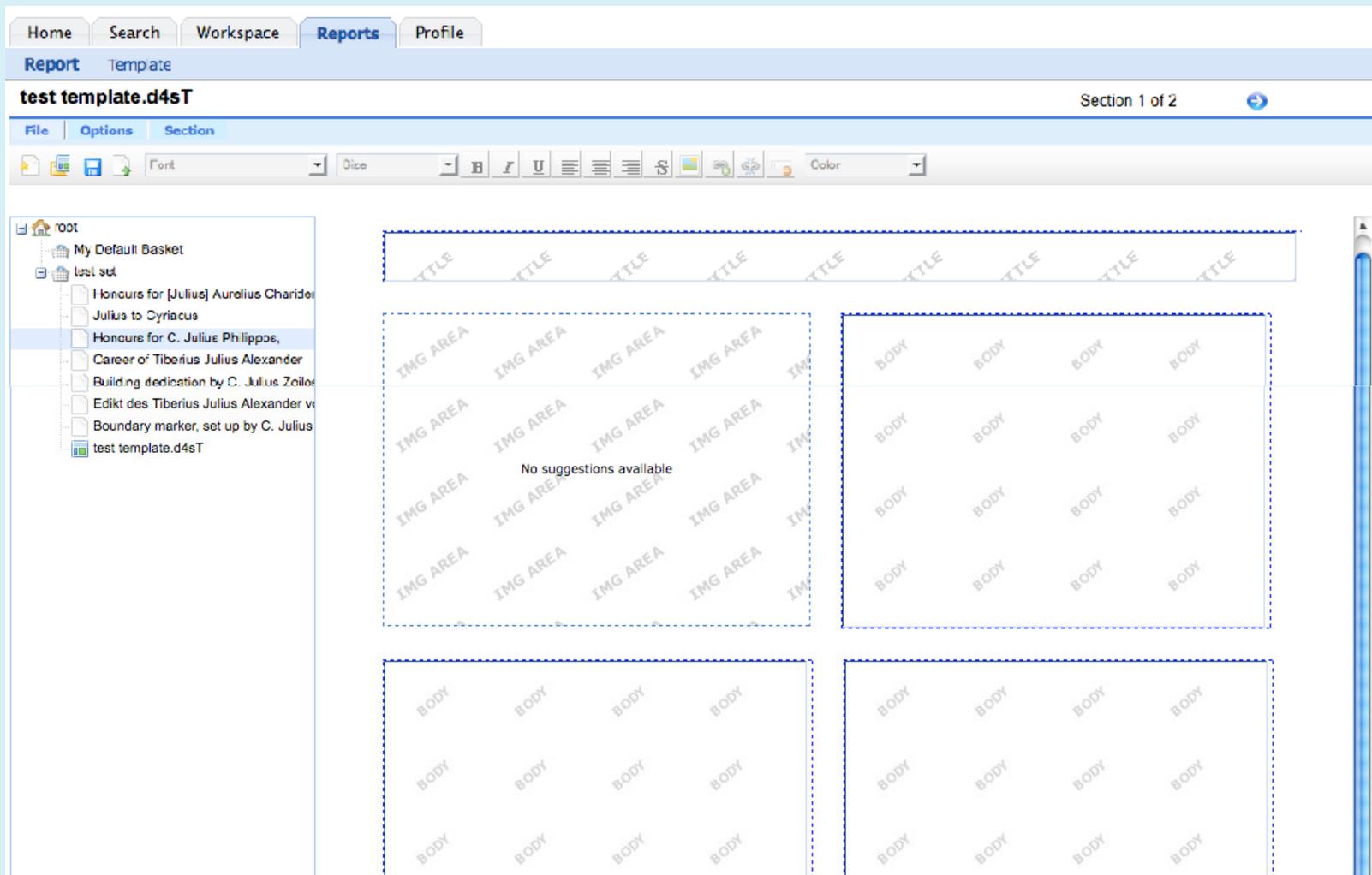
ContentViewer (PEG Image, 1000x609 pixels) - Scaled (57%)  
<http://portal.d4science.research-infrastructures.eu/application/SupportLayerHttp/ContentViewer?OID=1>

Handwritten text in the image object viewer:  
 Fund- und Standort:  
 nicht mit von der Hand  
 gehen  
 Material: W.K.  
 Breite: 19/2  
 Dicke: --  
 Buchstabenhöhe: 2'6  
 Form und Erhaltung:  
 Fragment eines Briefes;  
 oben für ganz rechts ganz  
 Form skizzieren: Notiz über 2'6  
 99  
 Wann und von wem copiert:  
 18/6

Object Metadata:  
 Title Building dedication by C. Julius Zoilos  
 Editor Joyce M. Reynolds  
 SourceDesc Originally published in Reynolds (1982).  
 English,  
 French,  
 ProfileDesc German,  
 Ancient Greek,  
 ...

query result set

image object in viewer



# Researcher Experiences

Data import procedures an issue

Carried out by dedicated D4Science staff

Required extensive contact between

D4Science staff and researchers staff

Not sustainable – need to make it easier

for gCube non-specialists to import data

“How to” guides

# Next steps

Integrate more data resources (including those managed externally to gCube)

Extend humanities VO to other groups of humanities researchers

Publication of data from VRE

# Contacts

mark.hedges@kcl.ac.uk

tobias.blanke@kcl.ac.uk

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