

# Classical Studies facing Digital Research Infrastructures: From Practice to Requirements

Dr. Agiatis Benardou

[a.benardou@dcu.gr](mailto:a.benardou@dcu.gr)

Digital Classicist & Institute of Classical Studies Seminar  
London, July 1<sup>st</sup>, 2011



# Preparing DARIAH (2008-2010)

- This paper reports on work conducted in the context of *Preparing DARIAH: Preparing for the construction of the Digital Research Infrastructure for the Arts and Humanities*
  - collaborative European project co-funded by the ESFRI e-Infrastructures programme,
  - aimed at providing the foundations (for the timely design and construction of the digital infrastructure requisite for scholarly research in the arts, humanities and cultural heritage in Europe)
- Major initiatives to develop appropriate, timely and interoperable infrastructures, services and tools to serve the current and emerging needs of humanists; at a time of large-scale digitization programmes in the field of cultural heritage, producing Web-accessible digital resources of primary interest to arts and humanities research

# Motivation and conceptual framework

- Provide mechanisms: **strategic, legal, financial, organizational, technological and conceptual**
- Ensure the **long-term availability and access to resources** needed by European researchers in the arts and humanities
- Encompass **all stages of the scholarly research process**, by means of an appropriate digital infrastructure spanning across Europe
- Emergence of digital humanities as a methodological current that is radically **interdisciplinary** in approach, often formal or quantitative in method, and making **intensive use of the data management and visualisation capabilities of information technology**
- **Shifts in research interests and methodological changes** in the arts and humanities

# DARIAH Transition Phase

After the completion this preparatory project in February 2011, the DARIAH infrastructure is now moving through a transition phase aiming at entering the construction phase in the autumn of 2011.

# DARIAH Conceptual work

- Conceptual modelling work-package of DARIAH consisting:
  - of the formulation of a conceptual model for scholarly research activity suitable for the representation of actual information practice in scholarly work
  - of an empirical study of scholarly research activity, based on the
    - elicitation,
    - transcription,
    - encoding,
    - analysis and interpretation of open-question interviews with humanities scholars across Europe

**User-centric / Need-centric approach**

# Background work

- 1980s:
  - Earlier studies of scholarly work range from ethnographically-based research and epistemologically informed monographs from the field of **science, technology and society studies (STS)**, (Latour & Woolgar, 1986; Knorr-Cetina, 1999), to information behaviour studies stemming from an information science perspective - solitary nature of humanistic research
- 1990s – early 2000:
  - Investigation of information work by humanities scholars visiting the Getty Research Institute, noting in their results the major importance of **named entities (proper names, places, titles of works)** as entry points to resource discovery, but also the fact that **the introduction of digital tools did not supplant traditional methods, but co-existed synergistically with them** ( Bates, Wilde, & Siegfried, 1995).
  - Exploration of the reliance of humanities scholars, as well as artists, on **multimedia resources** (Bates, 2001).
  - While these studies dealt primarily with the needs of humanists *vis-à-vis* scholarly objects (i.e. outcomes of scholarly research), such as in a library setting, further studies accommodated also information work on primary and secondary sources, focusing on disciplines such as **history** (Dalton & Charnigo, 2004; Delgadillo & Lynch, 1999; Duff, Craig, & Cherry, 2004; Tibbo, 2003) and **art history** (Beeman, 1994; Hemmig, 2008; Odum, 1998).
  - Short reports, based on the consultations and research concluded by the AHRC Methods Network in the UK, summarise needs and likely scenarios for ICT use in humanities research in particular fields, including **history, art history, archaeology, and museums and heritage**; a conceptual overview of ICT research methods employed by researchers has been developed by the Methods Network in the form of a thesaurus (“**methods taxonomy**”).

# Background work #2

- Further **empirical** research pointed to:
  - 1980s: the increasingly **interdisciplinary** nature of humanistic research, bearing significantly on the information service characteristics needed to cater for them.
  - the increased reliance of humanists on an **accommodative process of assembling information resources** relevant to the task at hand though **browsing**, rather than on **comprehensive searching**
  - the importance of **strategies** for **finding** and **collocating** relevant information other than searching, mainly through **berry-picking**, or **chaining**, illustrated, in the case of scholarly objects, through strategies of **footnote chasing**, **citation searching**, **journal run**, **area scanning**, **subject and author searches**, but also relevant to primary source information use
  - the importance of **annotation** for scholarly work
  - the frequent reliance on **serendipity** and the possibility of **non-deterministic contextual “discovery”** by scholars

# Scholarly primitives

- **Processes employed by literary scholars as a concept to identify common, low-level scholarly activities**
- *Starting, chaining, browsing, differentiating, monitoring and extracting* (Ellis, 1993)
- *Discovering, annotating, comparing, referring, sampling, illustrating, representing* (Unsworth, 2000)
- *Accessing, networking and verifying* (Meho & Tibbo, 2003)
- *Searching, collecting, reading, writing, collaborating, browsing, collecting, re-reading, assembling, consulting, notetaking* were found to be particularly common in the humanities (Palmer, Tefteau, & Pirmann, 2009)

# Motivation and conceptual framework

## #2

- The specification of digital infrastructure for the arts and humanities needs obviously to address the **historical practices**, needs and perceptions of **actual researchers**, rather than be merely driven by the interests and priorities of technology and service providers
- Understanding the **nature and information requirements of scholarly research**, notwithstanding **differences** between disciplines, research fields and methodological approaches
- Requirements need to be substantiated **by actual evidence**, and encompassing not only which digital resources, services and tools researchers use, how they interact with the whole spectrum of information and conceptual entities – **digital as well as non-digital**
- **Evidence-based rationale** of why and how DARIAH tools and services will correspond, and address, real, recurrent and important work patterns and needs of arts and humanities research
- **Conceptual perspective** for the identification of pertinent categories and properties representing scholarly research

# A&H conceptual modeling

## – Why ?

- Establish an operational framework for the actual representation of the information
- Analyse data
- Understand evidence related to scholarly information practice in the Arts and Humanities

## – Where ?

- Literature
- Interviews with experts
- Experts forum

## – How ?

- Activity systems theory
- Application of the CIDOC Conceptual Reference Model

## – Benefits

- Represent actual information on a structured set of events (what, where, when was done ?)
- Encompass notion of subject (who did it ?)
- Encompass notion of method (how was it done ?)

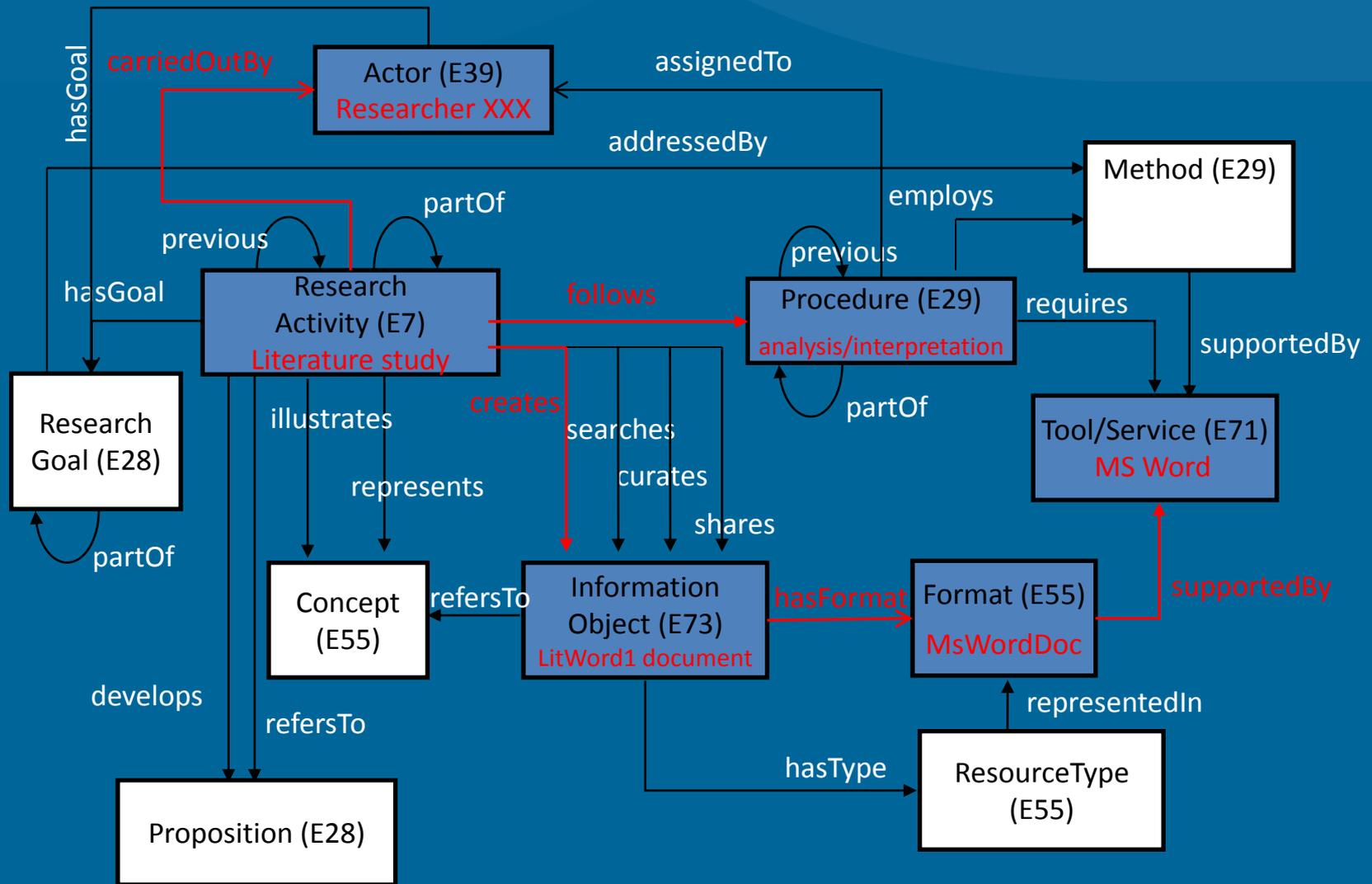
# Scholarly research activity model

- Entity examples:
  - Actor
    - Researcher, curator
  - Research activity
    - Study literature, Search Internet, Collaborate, Consult
  - Research goal
    - Determine chronology (e.g. of pottery collection), Provenance of a ceramics collection
  - Information object
    - Pottery collection
  - Tool / Service
    - Google maps, OPAC Catalog, MS Access
  - Format
    - PDF, Word, JPEG
  - Resource type
    - Electronic document, Map, Image
- Properties examples
  - hasType
  - Creates
  - part of
  - Searches
  - refersTo

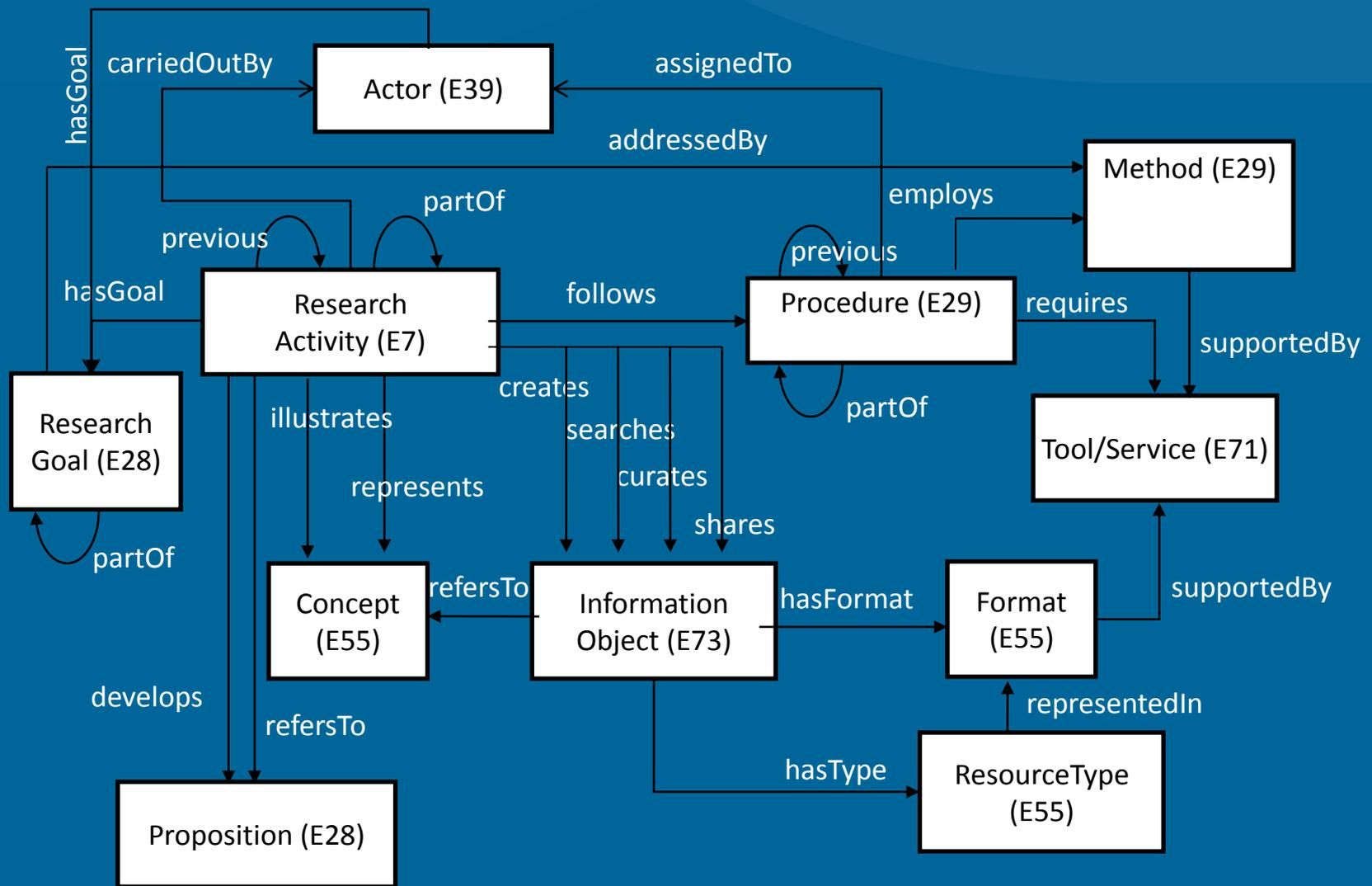
# Scholarly research activity model

- Example
  - Researcher XXX **carries out** Literature study
  - Literature study **follows** standard scholarly procedures of analysis and interpretation
  - Literature study **creates** Lit\_Word1 document
  - Lit\_Word1 document **hasType** MS Word Document
  - MS Word Document **supportedBy** MS Word

# Scholarly activity model



# Scholarly activity model



# Participants

- Mainstream humanists producing valid research
  - Not limited to, nor focusing on, niche research fields such as archaeological computing
- Focus on classics and archaeology
- Different career stages (ranging from PhD candidates to senior scholars)
- Diverse familiarity and intensity of use of ICT tools
  - Need to cover both the ‘average’ and the ‘early adopter’ scenarios
- Number: 24

# Empirical research

- Objective
  - enable a detailed documentation and deeper understanding of the research processes enacted by particular scholars
  - **not** meant for statistical evaluation or modeling
- Interviews with open question, free / conversational structure
- Both practice-related and conceptual questions
  - **practice-related**: annotation, terminology, raw data – sampling, work saving, workplace, availability and usability of digital resources
  - **conceptual-methodological**: conception of research topic and point of departure, motivation, notions of comparison and collaboration
- Content analysis
  - Identification of scholarly activities
  - Identification of data / scholarly objects

# From Practice to Requirements:

- User requirements established by DARIAH on the basis of evidence from scholarly practice in the Classical Studies and the Humanities fall within five key themes: scope of information objects, access conditions, semantic interoperability, collaborative work, and information use versus analytical tools.

# Scope of Information Objects

- Scholarly activity engages simultaneously with different kinds of information objects, from those typically defined as “primary data” to scholarly objects that are the outcome of scholarship (such as publications); in fact, the clear distinction between primary data and scholarly objects is blurred in actual practice. Thus, it is imperative that arts and humanities infrastructures provide effective means for identification, reference, access, representation and management mechanisms for this continuum of information objects, ranging from primary data and resources to complex scholarly objects.

# Access conditions

- In humanities research, environment is characterized by a complex layering of different access regimes for different kinds of resources. These range from open access for some, to heavily restricted access for other kinds of resources due to either intellectual property or heritage protection legislation. It is important that this broad diversity of access regimes cuts across different kinds of information objects, from primary resources to the outcomes of scholarly work. A credible system that allows consensual sharing and reuse of such socially generated information (as in annotations on particular information objects) needs to provide adequate mechanisms of sharing and trust.

# Semantic interoperability

- Semantic interoperability emerges as a key need for scholarly research, especially when such research is cross-disciplinary, multilingual, or based on distributed resources. DARIAH will need to provide access to resources that are located in different primary repositories, expressed in different languages, and providing content organised on the basis of different disciplines.

# Collaborative work

- While research in the arts and humanities still is typically solitary, findings in our empirical work and further evidence suggests that there are important social dimensions in the way research is supported by specific information-laden activities. A second trend is the emergence of tightly-knit, active communities of practice between researchers in particular sub-disciplines and research areas, that is built around a collective blog, an online archival resource, or other digital tool or service. As digital classicists become participants of multiple communities, they share, interact and co-create on the basis of their research interests, scholarly competence, and activities in which they participate.

# Tool-supported use of information

- Understanding the nature and information requirements of Classical studies scholarly research, notwithstanding differences within the discipline, research fields and methodological approaches, is an important motivation and a prerequisite for the definition of infrastructures, services and tools fit for the purpose of current and future scholarly research.

# Thank you for your attention