

# Leveraging Digital Cultural Memories

Panos Constantopoulos

Athens University of Economics and Business, and Athena Research Centre

**Abstract.** The penetration of ICT in the management and study of material culture and the emergence of digital cultural repositories and linked cultural data in particular are expected to enable new paths in humanities research and new approaches to cultural heritage. Success is contingent upon securing information trustworthiness, long-term preservation, and the ability to re-use, re-combine and re-interpret digital content. In this perspective, we review the use in the cultural heritage domain of digital curation and curation-aware repository systems; achieving semantic interoperability through ontologies; explicitly addressing contextual issues of cultural heritage and humanities information; and the services of digital research infrastructures.

The last two decades have witnessed an increasing penetration of ICT in the management and study of material culture, as well as in the Humanities at large. From collections management, to object documentation and domain modelling, to supporting the creative synthesis and re-interpretation of data, significant progress has been achieved in the development of relevant knowledge structures and software tools. As a consequence of this progress, digital repositories are being created that aim at serving as digital cultural memories, while a process of convergence among the different kinds of memory institutions, i.e., museums, archives, and libraries, in what concerns their information functions is already evolving.

Yet the advantages offered by information management technology, mass storage, copying, and the ease of searching and quantitative analysis, are not enough to ensure the usefulness of those digital cultural memories unless information trustworthiness, long-term preservation, and the ability to re-use, re-combine and re-interpret digital content are ensured. Furthermore, the widely encountered need for integrating heterogeneous information becomes all the more pressing in the case of cultural heritage due to the specific traits of information in this domain.

In view of the above fundamental requirements, in this presentation we briefly review the leveraging power of certain practices and approaches in realizing the potential of digital cultural memories. In particular, we review the use of digital curation and curation-aware repository systems; achieving semantic interoperability through ontologies; explicitly addressing contextual issues of cultural heritage and humanities information; and the services of digital research infrastructures.

Digital curation is an interdisciplinary field of enquiry and practice, which brings together disciplinary traditions and practices from computer science, information science, and disciplines practicing collections-based or data-intensive research, such as history of art, archaeology, biology, space and earth sciences, and application areas

such as e-science repositories, organizational records management, and memory institutions (Constantopoulos and Dallas 2008). Digital curation aims at ensuring adequate representation of and long-term access to digital information as its context of use changes, and at mitigating the risk of repositories becoming “data mortuaries”. To this end a lifecycle approach to the representation of curated information objects is adopted; event-centric representations are used to capture information “life events”; the class of agents involved is extended to include knowledge producers and communicators in addition to information custodians; and context-specificity is explicitly addressed.

Cultural heritage information comprises representations of actual cultural objects (texts, artefacts, historical records, etc.), their histories, agents (persons and organizations) operating on such objects, and their relationships. It also includes interpretations of and opinions about such objects. The recording of this knowledge is characterized by disciplinary diversity, representational complexity and heterogeneity, historical orientation, and textual bias. These characteristics of information are in line with the character of humanities research: hermeneutic and intertextual, rather than experimental; narrative, rather than formal; idiographic rather than nomothetic; and conformant to a realist rather than positivist account of episteme (Dallas 1999). The primary use of this information has been to support knowledge-based access, while now it is gradually also being targeted at various synthetic and creative uses. A rich semantic structure, including subsumption, meronymic, temporal, spatial, and various other semantic relations, is inherent to cultural information. Complexity is compounded by terminological inconsistency, subjectivity, multiplicity of interpretation and missing information.

From an information lifecycle perspective, digital curation involves a number of distinct processes: appraisal; ingesting; classification, indexing and cataloguing; knowledge enhancement; presentation, publication and dissemination; user experience; repository management; and preservation. These processes rely on three supporting processes, namely, goal and usage modelling, domain modelling and authority management. These processes effectively capture the context of digital curation and produce valuable resources which can themselves be seen as curated digital assets (Constantopoulos and Dallas 2008; Constantopoulos et al. 2009).

The field of cultural information presents itself as a privileged domain for digital curation. There is a relatively long history of developing library systems and museum systems, along with recent intense activity on interoperable, semantically rich cultural information systems, boosted by two important developments: the emergence of the CIDOC CRM (ISO 21127)<sup>1</sup> standard ontology for cultural documentation; and the movement for convergence of museum, library and archive systems, one manifestation of which is the CIDOC CRM compatible FRBR-oo model<sup>2</sup>.

Advances such as those outlined above allow addressing old research questions in new ways, as well as putting new questions that were very hard or impossible to tackle without the means of digital technologies. Significant enablers towards this direc-

---

<sup>1</sup> <http://www.cidoc-crm.org/>

<sup>2</sup> [http://www.cidoc-crm.org/frbr\\_inro.html](http://www.cidoc-crm.org/frbr_inro.html)

tion are the so-called digital research infrastructures, which bear the promise of facilitating research through sharing tools and data. Several trends can be identified in the development of research infrastructures, which follow two main approaches:

- a) The normative approach, whereby normalized collections of data and tools are developed as common resources and managed centrally by the infrastructure.
- b) The regulative approach, whereby resources reside with individual organizations willing to contribute them, under specific terms, to the community. A set of interoperability conditions and mechanisms provide a regulatory function that lies at the heart of the infrastructure.

Both approaches are being pursued in all disciplines, but the mix differs: in hard sciences building common normalized infrastructures appears to be a necessity, with a complementary, yet significant role to be played by a network of interoperable, disparate sources. In the humanities, on the other hand, long scholarly traditions have produced a formidable variety of information collections and formats, mostly offering interpreted, rather than raw material for publication and sharing. These conditions favour the development of regulated networks of interoperable sources, with centralized, normative infrastructures in a complementary capacity.

By way of example, a recent such infrastructure is DARIAH- GR / ΔΥΑΣ<sup>3</sup>, one of the national constituents of DARIAH-EU<sup>4</sup>, the Europe-wide digital infrastructure in the arts and humanities. DARIAH- GR / ΔΥΑΣ is a hybrid-virtual distributed infrastructure, bringing together the strengths and capacities of leading research, academic, and collection custodian institutions through a carefully defined, lightweight layer of services, tools and activities complementing, rather than attempting to replicate, prior investments and capabilities. Arts and humanities data and content resources are as a rule thematically organized, widely distributed, under the custodianship and curation of diverse institutions, including government agencies and departments, public and private museums, archives and special libraries, as well as academic and research units, associations, research projects, and other actors, and displaying a diverse degree of digitization. The mission of the infrastructure is then to provide the research communities with effective, comprehensive and sustainable capability to discover, access, integrate, analyze, process, curate and disseminate arts and humanities data and information resources, through a concerted plan of virtual services and tools, and hybrid (combined virtual and physical) activities, integrating and running on top of existing primary information systems and leveraging integration and synergies with DARIAH-EU and other related infrastructures and aggregators (e.g. ARIADNE<sup>5</sup>, CARARE<sup>6</sup>, LoCloud<sup>7</sup>). In its first stage of development, the DARIAH- GR / ΔΥΑΣ Research Infrastructure has offered the following groups of services:

---

<sup>3</sup> <http://www.dyas-net.gr/>

<sup>4</sup> <http://www.dariah.eu/>

<sup>5</sup> <http://www.ariadne-infrastructure.eu/>

<sup>6</sup> <http://www.carare.eu/>

<sup>7</sup> <http://www.locloud.eu/>

- *Data sharing*: comprehensive registries of digital resources;
- *Supporting the development of digital resources*: tools and best practice guidelines for the development of digital resources;
- *Capacity building*: workshops and training activities; and
- *Digital Humanities Observatory*: evidence-based research on digital humanities, monitoring, outreach and dissemination activities.

Key factor in the development of DARIAH- GR / ΔΥΑΣ, ARIADNE, CARARE and LoCloud resources alike has been a curation-oriented aggregator, the Metadata and Object Repository - MORE<sup>8</sup> (Gavrilis, Angelis & Dallas 2013; Gavrilis et al. 2013). This system supports the aggregation of metadata from multiple sources (OAI-PMH, Archive, SIP, Omeka, MINT) and heterogeneous systems in a single repository, the creation of unified indexes of normalized and enriched metadata, the creation of RDF databases, and the publication of aggregated records to multiple recipients (OAI-PMH, Archive, Elastic Search, RDF Stores). It enables the dynamic definition of validation and enrichment plans, supported by a number of micro-services, as well as the measurement of metadata quality. MORE can incorporate any XML/RDF metadata schema and can support several intermediate schemas in parallel. Its architecture is based on micro-services, a software development model according to which a complex application is composed of small, independent services communicating via a language-agnostic API, thus being highly reusable. MORE currently maintains access to 30 SKOS-encoded thesauri, totaling several hundred thousands of terms, as well as to copies of the Geo-names and Perio.do services, thus offering information enrichment on the basis of a wide array of sources. Metadata enrichment is a process of automatic generation of metadata through the linking of metadata elements with data sources and/or vocabularies. The enrichment process increases the volume of metadata, but it also considerably enhances their precision, therefore their quality. Performing metadata aggregation and enrichment carries several benefits: increase of repository / site traffic, better retrieval precision, concentration of indexes in one system, better performance of user services. To date MORE is used by 110 content provider institutions, and accommodates 23 different metadata schemas and about 20,800,000 records.

The advent of digital infrastructures for arts and humanities research calls for a deeper understanding of how humanists work with digital resources, tools and services as they engage with different aspects of research activity: from capturing, encoding, and publishing scholarly data to analyzing, visualizing, interpreting and communicating data and research argumentation to co-workers and readers. Digitally enabled scholarly work and the integration of digital content, tools and methods present not only commonalities but also differences across disciplines, methodological traditions, and communities of researchers. A significant challenge in providing integrated access to disparate digital humanities resources and, more broadly, in supporting digitally-enabled humanities research, lies in empirically capturing the context of use of digital content, methods and tools.

---

<sup>8</sup> <http://more.dcu.gr/>

Several attempts have been made to develop a conceptual framework for DH in practice. In 2008, the AHRC ICT Methods Network<sup>9</sup> developed a taxonomy of digital methods in the arts and humanities. This was the basis for the classification of over 200 digital humanities projects funded by the U.K. Arts and Humanities Research Council in the online resource arts-humanities.net, as well as for the subsequent Digital Humanities at Oxford<sup>10</sup> taxonomy. Other initiatives to build a taxonomy of Digital Humanities include TADIRAH<sup>11</sup> and DH Commons<sup>12</sup>. From 2011 to 2015 the Network for Digital Methods in the Arts and Humanities<sup>13</sup> (NeDiMAH) ran over 40 activities structured around key methodological areas in the humanities (digital representations of space and time; visualisation; linked data; creating and using large scale corpora; and creating editions). Through these activities, NeDiMAH gathered a snapshot of the practice of digital humanities in Europe, and the impact of digital methods on research.

A key output of NeDiMAH is NeMO<sup>14</sup>: the *NeDiMAH Ontology of Digital Methods in the Arts and Humanities*. This ontology of digital methods in the humanities has been built as a framework for understanding not just the use of digital methods, but also their relationship to digital content and tools. The development of an ontology, rather than a taxonomy, stands in recognition of the complexity of the digital humanities landscape, the interdisciplinarity of the field, and the dependencies that impact the use of digital methods in research. NeMO provides a conceptual framework capable of representing scholarly work in the humanities, addressing aspects of intentionality and capturing the diverse associations between research actors and their goals, activities undertaken, methods employed, resources and tools used, and outputs produced, with the aim of obtaining semantically rich structured representations of scholarly work (Angelis et al 2015; Hughes, Constantopoulos & Dallas 2016). It is grounded on earlier empirical research through semi-structured interviews with scholars from across Europe, which focused on analysing their research practices and capturing the resulting information requirements for research infrastructures (Benardou, Constantopoulos & Dallas 2013).

The relevance of NeMO to the DH community was validated in a series of workshops through use cases contributed by researchers. A variety of complex associative queries articulated by researchers and encoded in SPARQL, demonstrated the potential of NeMO as an effective mechanism for information extraction and reasoning with regard to the use of digital resources in scholarly work and as a knowledge base schema for documenting scholarly practices. In a recent workshop in DH2016, researchers created their own NeMO-based descriptions of projects with an easy to use tool (Constantopoulos et al 2016).

---

<sup>9</sup> <http://www.methodsnetwork.ac.uk/index.html>

<sup>10</sup> <https://digital.humanities.ox.ac.uk/people-projects>

<sup>11</sup> <http://tadirah.dariah.eu/vocab/index.php>

<sup>12</sup> <http://dhcommons.org/>

<sup>13</sup> <http://nedimah.eu/>

<sup>14</sup> <http://nemo.dcu.gr/>

Knowledge bases documenting scholarly practice through NeMO can be useful to researchers by (a) helping them find information on earlier work relevant for *their own* research; (b) supporting goal-oriented organization of research work; (c) facilitating the discovery of new paths with regard to resources, tools and methods; and, (d) promoting networking among researchers with common interests. In addition research groups can get support for better project planning by explicitly exposing links between goals, actors, activities, methods, resources and tools, as well as assistance for discovering methodological trends, future directions and promising research ideas. Funding agencies, on the other hand, could benefit from the kind of systematic documentation and comparative overview of project work enabled by the ontology.

## References

- C. Dallas. "Humanistic research, information resources and electronic communication". Electronic Communication and Research in Europe, J. Meadows and H.-D. Boecker, Eds. Luxembourg: European Commission, 1999, 209-239.
- P. Constantopoulos and C. Dallas. "Aspects of a digital curation agenda for cultural heritage". IEEE 2008 Int'l Conference on Distributed Human-Machine Systems, Athens, March 2008.
- P. Constantopoulos, C. Dallas, I. Androutopoulos, S. Angelis, A. Deligiannakis, D. Gavrilis, Y. Kotidis, & C. Papatheodorou. "DCC&U: An Extended Digital Curation Lifecycle Model". The International Journal of Digital Curation, 4, Issue 1, 34-45, 2009.
- D. Gavrilis, S. Angelis, C. Dallas. "A Curation-Oriented Thematic Aggregator". Proc. 17th International Conference on Theory and Practice of Digital Libraries, TPD 2013, Valetta, Malta, September 2013, 132-137.
- D. Gavrilis, S. Angelis, C. Papatheodorou, C. Dallas, and P. Constantopoulos. "Preservation Aspects of a Curation-Oriented Thematic Aggregator". Proc. 10th International Conference on Preservation of Digital Objects, iPRES2013, Lisbon, Portugal, September 2013, 246-251.
- A. Benardou, P. Constantopoulos, C. Dallas. "An approach to analyzing working practices of research communities in the humanities". International Journal of Humanities and Arts Computing, 7, 2013, 105-127.
- S. Angelis, A. Benardou, N. Chatzidiakou, P. Constantopoulos, C. Dallas, L. M. Hughes, L. Papachristopoulos, E. Papaki, V. Pertsas. "Documenting and reasoning about research on ancient Corinthia using the NeDiMAH Methods Ontology (NeMO)". 43rd Computer Applications and Quantitative Methods in Archaeology Conference CAA 2015, Siena, March 2015.
- L. M. Hughes, P. Constantopoulos, and C. Dallas. "Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines". In S. Schreibman, R. Siemens, J. Unsworth (eds.), A New Companion to Digital Humanities, Wiley-Blackwell, 2016.
- P. Constantopoulos, L. M. Hughes, C. Dallas, V. Pertsas, L. Papachristopoulos and T. Christodoulou. "Contextualized Integration of Digital Humanities Research: Using the NeMO Ontology of Digital Humanities Methods". Digital Humanities 2016, Kraków, July 2016.
- P. Constantopoulos, C. Dallas, L. M. Hughes, and S. Ross (organizers). "Ontology-Based Recording and Discovery of Research Patterns in the Humanities", Pre-Conference Workshop, Digital Humanities 2016, Kraków, July 2016.