The post-repository era: scholarly practice, information and systems in the digital continuum

Costis Dallas

Athena Research Center - IMIS

Abstract

Research in the arts and humanities is often associated with the world of the solitary scholar, surrounded by dusty books, manuscripts, or artefacts. As early as 1959, C.P. Snow lamented the "gulf of mutual incomprehension" separating humanities scholars from scientists. Yet, the wide-ranging changes in scholarly work in the arts and humanities associated with the inception of digital technology, the crisis of disciplinarity, the rise of new methodological and theoretical frameworks, the increased challenges to the longevity of cultural heritage resources, and the emergence of new fields of contestation around their interpretation and value, casts a different light on Snow's notion of the "two cultures", introducing new issues and opportunities.

For information researchers and computer scientists engaged with the conceptualization, design and development of digital infrastructures, tools and services in the domain of the arts, humanities and cultural heritage, understanding the nature and direction of these changes is of paramount importance. The distinct field of digital humanities, "a genuinely intellectual endeavor with its own professional practices, rigorous standards, and exciting theoretical explorations" in the words of Katherine Hayles, is only part of the story. In fact, digital humanists working with big data, crafting their own schemas and encoding formalisms, engaging in ontological modeling, and scripting their own analytical and representational tools are but a small minority among an increasing number of scholars producing influential, highly cited research merely facilitated by digital technology – what may be called digitally-enabled humanists. As indicated by a major new survey conducted by the Digital Methods and Practices Observatory Working Group of DARIAH-EU, the Digital Research Infrastructure for the Arts and Humanities in Europe, digitally-enabled humanists use frequently applications such as word processors and spreadsheets besides repositories to organize and curate research resources, controlled vocabularies and classification systems that are more often homegrown than standard, and a variety of readily available online services and social media for information discovery, collaboration and dissemination. And, in tandem with changes in scholarly practice, the rise of computational intelligence and social and participatory media, as well as the increasing availability of humanities and heritage resources in the networked and mobile digital environment at a time of globalization, bring about new important stakeholders in their representation and interpretation, such as descendant and source communities, amateurs engaging in citizen science, and culture and heritage publics.

Established wisdom on digital infrastructures for the arts and humanities is shaped by a notion of centralized custodial control, replicating the traditional structures of the physical archive, library and museum: in other words, on the notion that research resources can be curated and preserved in the future in large-scale, centralized digital

repositories. This becomes problematic as financial means grow increasingly scarce, and as the cultural record broadens to include a proliferation of born digital resources, grey literature, outcomes of independent and commercial research, fruits of self-publication, remix and social media interaction, and manifestations of community and personal memory. In fact, research on emerging digital research practices in a discipline such as archaeology shows how the availability of multimodal, interactive, real time recording and documentation technologies, and the plurality of research actors, interpretations and uses of archaeological knowledge give rise to multiple kinds of densely interconnected digital resources (including GIS, LiDAR, formatted data, 3D models, annotations, interpretive narratives, video documentation, blogs, and social media interactions) and interweave the ostensibly distinct processes of data recording and interpretation. The shift towards a digital infrastructure for humanities and heritage resources, in tandem with these changes, brings about a rising "curation crisis" which calls for a radical reconsideration of priorities in the specification and design of digital infrastructures.

Central to this reconsideration is the concept of the records continuum, originally advanced by Australian archival scholars to indicate the limits of a lifecycle approach in dealing with the capabilities and challenges of digital information. Criticizing the custodial notion of archives as data mortuaries, continuum thinking calls for a unified approach to recordkeeping capable of attending to records from the point of creation to their "pluralizing" interpretation and use by diverse communities. It resonates with a call for a radical re-examination of the theory and practice of digital curation, based on the recognition that curation of research resources facilitated by ubiquitous pervasive digital technology takes place increasingly "in the wild", involves multiple stakeholders "exercising the archive" beyond data custodians, concerns not merely information resources qua digital objects but also their evolving epistemic content and context, and therefore requires a rethink of the requirements, affordances and priorities of digital research infrastructures.

The promise of going beyond traditional repositories to deploy a digital infrastructures which explicitly focuses on the provision of curation capabilities is demonstrated by the Metadata and Object Repository (MORe), a system deployed by the Digital Curation Unit, IMIS-Athena Research Centre to support the dynamic evolution and continuous semantic enrichment of heterogeneous metadata and registry descriptions of arts and humanities resources and collections. MORe has been used extensively for Europeana metadata aggregation in the Connecting Archaeology and Architecture with Europeana (CARARE) and Local Institutions in a Europeana Cloud (LoCloud) projects, providing a combination of semi-automated and manual digital curation activities by diverse curation actors, and leveraging workflows of external services such as historic names gazetteers and SKOS vocabularies. It provides for the curation of resources "in the wild", such as Wikimedia assets and Flickr images with their associated metadata. It also connects with client systems such as Omeka-based LoCloud Collections, and uses a Metadata Entry Tool to support "sheer curation" on a digital tablet at the moment of data creation or capture.

Systems such as MORe herald the first steps towards a new approach to digital infrastructures, beyond the architecture and functionalities of traditional repositories such as Fedora or DSpace. Yet, a key challenge remains how to address the fact that scholars in the arts and humanities, amateur researchers, holding institutions,

collectors, and users of digital information of cultural value will continue to employ a bricolage of digital tools, services and resources available "at hand", some of which will be increasingly imprisoned within technical or commercially-controlled silos. An overarching vision for scholarly digital infrastructures in the post-repository era might thus call for a combination of distributed cloud storage of open, intelligent, self-documenting, dynamic information objects with curation-enabled information systems and orchestrated services accessible to a native interface of end-user tools and applications in the continuum.