

---

# The Battleground of Humanities Data: Handling Multidisciplinarity and Diversity

Athanasios Karasimos\*<sup>1</sup>, Yorgos Tzedopoulos\*<sup>1</sup>, Helen Goulis<sup>1</sup>, Eirini Mergoupi-Savaidou<sup>1</sup>, Gerasimos Chrysovitsanos<sup>1</sup>, Athina Iakovidou<sup>1</sup>, Anastasia Falierou<sup>1</sup>, Patritsia Kalafata<sup>1</sup>, Eleni Vernardaki<sup>1</sup>, Maria Spiliotopoulou<sup>1</sup>, and Helen Katsiadakis<sup>1</sup>

<sup>1</sup>Academy of Athens, DARIAH-GR/DYAS – Greece

## Abstract

Humanities and Arts data vary significantly as to their degree of formalization. Whereas language studies and archaeology have developed formal vocabularies and/or more or less strict typologies, historical data present a particular challenge as to their spatio-temporal accuracy (fuzzy data) and as to their interpretation. Historical concepts and terminologies tend to porousness and relativism in relation to social and cultural contexts over time and space. Besides data records of the past in physical form usually contain errors and gaps, if they are not entirely missing or destroyed. The continuous effort to digitize collections of cultural heritage and Humanities data has increased the availability of and access to diverse, multidisciplinary, significant and rare materials; however, this overproduction of digital and digitized data can be easily lost when misidentified or when not meta-identified at all. Unfortunately, 80% of the data are accompanied by zero or false metadata. As digital collections grow, it is imperative that every digitized object is tagged with accurate metadata information; more specifically, all Digital Humanities data should be described by a dominant and widely accepted metadata schema to be easily searched, categorized and processed. At the end of the day, metadata are simply the vehicle to help researchers find, access and reuse data. The Arts and Humanities landscape, a record of human condition and its social, cultural and intellectual evolution, covers diverse disciplines, themselves with a long history over the ages. The common ground among disciplines is an intersecting area of new knowledge, since one of the emerging trends of DH is to address, analyze and process data through the prism of multidisciplinary. According to Richardson (2013), the visualization of citation links in the Arts and Humanities publications shows a cluster of interrelated topics: different facets of investigation that build a complex but straightforward network of multidisciplinary, which cannot be ignored by any research approach and analysis.

The team of the Greek national infrastructure for the Arts and Humanities (DYAS) faced the multidisciplinary challenge when compiling a registry of all Greek collections, whether physical or digital, that are of interest to humanities and social sciences. The registry includes collections of interest to more than 15 disciplines such as Anthropology and Ethnology, Archaeology, Classics, Numismatics, Epigraphy, History, History and Philosophy of Science, Literature, Linguistics, Philosophy, Theatre Studies, Musicology, Byzantine Studies, Ottoman and Turkish Studies, Arts and Art History. The majority of information on collections is recorded in cooperation with the curating institutions. This has often led them

---

\*Speaker

to reassess their method of documentation in the light of interaction with the DH infrastructure. The DYAS Collections Registry enables researchers to detect information about their research in numerous types of collections, varying from collections of objects regarding different branches of Humanities (museum collections, costume collections, carte-postales, maps, etc.) to collections of digitized texts (collections of books, manuscripts, textbooks, correspondence, written and spoken data etc). The vast amount of data and the variety of disciplines required a highly elaborated metadata schema accommodating the description of both physical and digital collections.

The DYAS multidisciplinary team has also developed the DYAS Humanities Thesaurus comprising most of the Arts and Humanities subject fields. As every discipline uses a different terminology, one of the main challenges of this effort was to agree on a common, top level coherent and consistent "backbone" Thesaurus containing a limited number of concepts from which all the thematic vocabularies and terminologies would inherit their attributes. The compilation of this discipline-agnostic Backbone Thesaurus was a bottom-up process; top-level concepts were developed by adequate abstraction from existing terminologies, meeting the demands for intersubjective and interdisciplinary validity. Besides, the DYAS Collections Registry has also been a source for some of the terms for the Humanities Thesaurus. The terms used for the thematic description of collections, in other words a flat vocabulary, are being modelled in a structured way under the top-level concepts of the Backbone Thesaurus.

Handling humanities data is not a simple task, even within the framework of one discipline. Important issues arise concerning the diversity of the relevant sources, the institutions that curate and preserve them, and the users' requirements. Data integration in major pools creates a demand for interoperability. For example, we are currently developing a digital platform for the integration of metadata related to digitized documents, photographs, artefacts, and moving images from the 1940s in Greece. The project aims at devising an umbrella of "meta-metadata" that can bridge the varied specific metadata created by the holding institutions. This interoperable platform will enable the public and the researchers to find, access, and reuse effectively the relevant data, as well as to discover underlying connections between them.

The three aforementioned services are inscribed within an Open Science strategy, which aims at enabling new lines of inquiry by combining available sources on data. Following the necessary steps of processing for easy use (documented, machine-readable, understandable and readable in convenient format), the connection between open data and proper standardized metadata is required since this is the European Community goal to optimize the future impact of research in Europe (and worldwide). To this end, we try to reconcile the necessary standardization requirements with the complexity and diversity of Humanities data.

## Bibliography

Doerr, M., Daskalaki, M., Bekiari, Ch. (2014). Thesaurus Building. DYAS, FORTH. Retrieved from

[https://www.backbonethesaurus.eu/sites/default/files/ppt\\_thesuarus%20building.pdf](https://www.backbonethesaurus.eu/sites/default/files/ppt_thesuarus%20building.pdf)

Doerr, M., Katsiadakis, H., Daskalaki, M., Goulis, H. (2015). Backbone thesaurus: Methodological outlines. DYAS, Academy of Athens & FORTH. Retrieved from

[https://www.backbonethesaurus.eu/sites/default/files/ppt\\_methodological%20outlines.pdf](https://www.backbonethesaurus.eu/sites/default/files/ppt_methodological%20outlines.pdf)

Karasimos, ., Goulis, H., Kalafata, P., Lipke, A., Papadakis, N., Mergoupi-Savaidou, E., Tzedopoulos, Y., Falierou, A., Chrysovitsanos, G., Vernardaki, E., Spiliotopoulou, M. & H. Katsiadakis (in press). From DARIAH-GR/DYAS to PARTHENOS-EU Project: the digital infrastructures, the Collections Registries and the Thesauri in Humanities (in Greek). In Proceedings for Digital Humanities in Greece: issues and challenges. Athens: Research for the Humanities (RCH).

Richardson, M. (2013). Mapping the Multidisciplinarity of the Arts & Humanities. *Research Trends* 32.Elsevier Publications. Retrieved from <https://www.researchtrends.com/issue-32-march-2013/mapping-the-multidisciplinarity-of-the-arts-humanities-2/>.

Wineburg, Samuel (1991). On the Reading of Historical Texts: Note on the Breach Between School and Academy. *American Educational Research Journal* 28 (1991), pp. 495-519. Retrieved from

[https://pdfs.semanticscholar.org/6499/13b351565c07d5d53f852b954eac09a1b0da.pdf?\\_ga=2.124481870.511527096.174633602.1550749431](https://pdfs.semanticscholar.org/6499/13b351565c07d5d53f852b954eac09a1b0da.pdf?_ga=2.124481870.511527096.174633602.1550749431)

**Keywords:** Digital Humanities, Registries, Thesaurus, Ontologies, Multidisciplinarity, Metadata, Infrastructure, Open Data