

## INTEGRATION OF MULTIMEDIA COLLECTIONS AND TOOLS FOR INTERACTION WITH DIGITAL CONTENT.

### The case study of the *Archaia Kypriaki Grammateia* Digital Corpus

Uros Damnjanovic<sup>1</sup>, Valentina Vassallo<sup>1</sup>, Sorin Hermon<sup>1</sup>

<sup>1</sup>The Cyprus Institute - STARC

(u.damnjanovic, v.vassallo, s.hermon)@cyi.ac.cy

### Abstract

Supporting a discovery, use, and navigation of digital collections is a fundamental part of providing access and encouraging inquiry, interpretation, and knowledge. In this paper we present our efforts to store and explore multimedia collections of archaeological data. Particularly, the case study of the *Archaia Kypriaki Grammateia* epigraphic collection is presented. Our work can be seen twofold. One aspect of our work is to provide a place where the data coming from various sources can be stored and accessed. Another aspect is to provide users with means to explore this data. We argue that currently digital libraries are constrained by their webpage-based paradigm, thus not providing the means for utilizing the full potential of the heritage data.

### Keywords

Digital libraries, Data repository, Epigraphy, User interfaces, Data visualization, Interaction.

## 1. The world going digital

Nowadays digital information format found its ways to all the spheres of our lives. Cultural heritage is no exception to this. Gigabytes of data are being created on an every-day basis. For some time now, digital libraries are

used to store and provide access to digital heritage data. However, existing digital libraries often fail to provide the ability to iteratively explore items, compare data trends, and engender the wisdom that comes from exploring data in new ways. Current digital libraries are limited by their inadequate webpage-based paradigm, and it is easy for even the most experienced scholar to get lost [Bergstorm and Atkinson, 2009]. To overcome this issue, new generation of digital libraries should emerge that will serve as an interactive knowledge environment in which knowledge is created by interacting with the data. Digital collections often lack features for deeper quantitative and qualitative analysis, and even very useful functions, such as the ability to annotate or bookmark content, are often not supported. After digitisation, these collections are typically monolithic, difficult to navigate, and can contain text which is of variable quality in terms of language, spelling, punctuation and consistency of terminology.

According to [Fast and Sedig, 2010] digital libraries are seen as a store of epistemic potential with interaction having the role to reveal much of the hidden complexity. In [Buchel and Sedig, 2014] the authors demonstrated the role that a set of interactions can play in supporting users' understanding of spaces and their formation of cognitive maps when working with map-based visualisations. According to [Algee et al., 2012] interactive visualizations empower users to discover meaning and patterns within digital collections using dynamic, interactive displays. One example of utilizing the interactive graphics in a real world digital library is presented in [Hienert et al., 2012]. Authors developed interactive visualization tools that support exploration of search queries and search results and help users in formulating new queries. Another example, the application called DaisyViz shown in [Ren et al., 2009], tries to help users acquire better insights of the data by enabling users to rapidly develop domain-specific information visualizations without traditional programming. In [Nasar et al., 2011] authors proposed a conceptual framework that uses interactive visualizations for managing personal collections of images and videos with focus on data re-finding and improved filtering. INVISQUE [Wong et al., 2012] is a novel system designed for interactive information exploration. Instead of a conventional list-style arrangement, in INVISQUE information is

represented by a two-dimensional spatial canvas, with each dimension representing user-defined semantics. Search results are presented as index cards, ordered in both dimensions. Intuitive interactions are used to perform tasks such as keyword searching, results browsing, categorizing, and linking to online resources such as Google and Twitter. An approach for overview-first exploration of data collections based on user-selected metadata properties is presented in [Bernard et al., 2012]. In a 2D layout representing entities of the selected property are laid out based on their similarity with respect to the underlying data content. The display is enhanced by compact summarizations of underlying data elements, and forms the basis for exploratory navigation of users in the data space. In [Nan et al., 2012] authors propose a novel visual design termed “Whisper” to fulfil the need for tracing information diffusion processes in social media, in a real time manner. BirdVis [Ferreira et al., 2011] leverages visualization techniques and uses them in a novel way to better assist users in the exploration of interdependencies among model parameters. Furthermore, the system allows for comparative visualization through coordinated views, providing an intuitive interface to identify relevant correlations and patterns.

## 2. “A place where the past meets the future”: the STARC repository

STARC repository was initially developed to store the data coming from the daily work of the research group and as content aggregator for various digital libraries projects: data from archaeological excavations, museums artefacts, epigraphic *corpora*, etc. Data stored in the repository ranges from high resolution images, 3D models, texts, maps, videos and audio resulting from various data acquisition procedures such as photogrammetry, laser scanning, 3D modeling, photography, sketching, drawing and so on. Every item in the repository is described using a metadata schema that depends on the type of data and according to the needs of the project. For most of the items, we are using our own metadata schema (STARC metadata schemas) [Ronzino et al., 2012; Vassallo et al., 2013], that was developed to

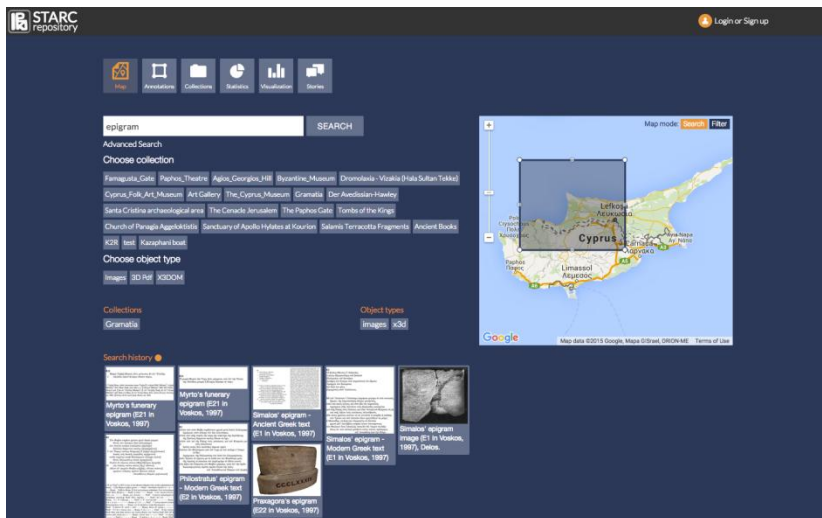
thoroughly describe various aspects of data creation process. Once data is uploaded and available, it can also be used in different systems. For example, the repository serves as source of data for aggregation procedures by which our data feed the Europeana portal (<http://www.europeana.eu/portal/>) through different digital libraries projects (e.g. Athena, CARARE, Linked Heritage, AthenaPlus, EAGLE). In order to facilitate the migration of data from heterogeneous sources, the repository provides ingestion capabilities, that enable easy transfer of data from any available data end-point. Another aspect of the back-end is the user management. It deals with providing different access levels for different users' groups. It also manages personal user space, where users can add their own information to the repository and have access to it when exploring the data. Data annotations and personal collections are some of the examples of how users can add information to the repository and use it for data exploration.

## 2.1. Tools for accessing data

We wanted to provide users with number of tools that they can use while exploring the repository. In traditional web based system, navigating from one tool to another usually means that we click on a link and then move to another page where the tool is available. Once the user is on a page s/he can start interacting with the tool by adding or modifying data or by performing any action relevant to the selected functionality. The process of exploring the repository starts by running the initial query (**Fig.1**).

Once the results are returned and displayed, the exploration process can start. By default the map functionality is available in the view. When a new tool is added to the screen, it automatically gets an access to all the data available in the current search. It also gets notified on any other tools that might be active in the view, so the necessary communication channels are established. Once the results are ready, there are number of filters available to filter out the data, either by data type, collection. Search form can be used to filter out the data by typing the text used to filter the data. As the user

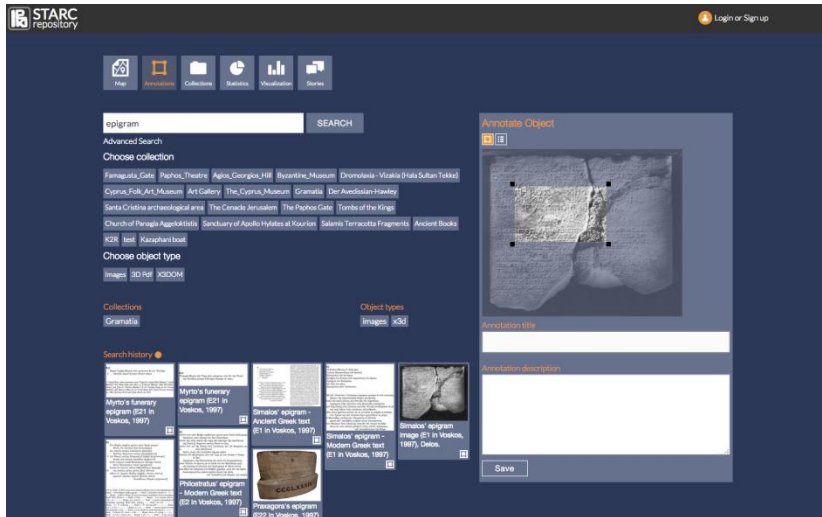
types, the data is automatically updated to show only those items that are related to the text in the search field. The user can then run a new search repeatedly, and the search history functionality will store all those searches for later reference. The map tool (**Fig.1**) is used to show information about the geo-referenced data. It provides two main functions. One is that it accompanies the search tool. Whenever the new search is ran, and the results become available, the map tool updates itself with new data, showing the



**Fig. 1.** The exploration process starts by running the search. Once the results are available the user can add or remove exploration tools. By default map tool is present in the view. The map tool is used to mark the geo-referenced data on the map, and also to perform geo-search by defining the region of interest.

geographical distribution of the search results. The user can then use the map to filter out the data by clicking on the markers on the map. Also the map can also be used as a search tool. By selecting the search map mode, the rectangle mask appears on the map, that is used to define the region of interest on the map. This region is then used as an additional parameter for the search.

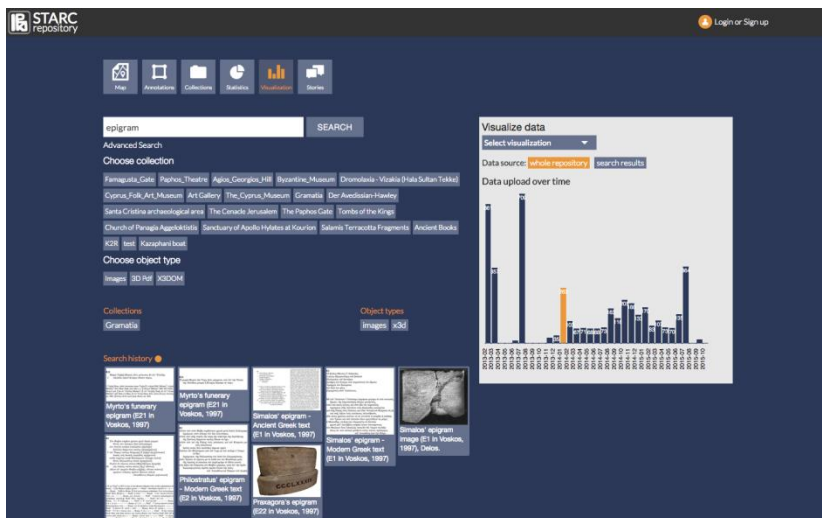
Annotation tool is used to attach the information to the objects in the repository and share it with other users, or use it to facilitate more efficient search. The tool is used by selecting object from search results, then either by selecting a region in the image files (**Fig.2**), or by selecting a point for 3D models, and by adding textual description of the selected part.



**Fig. 2.** The annotation tool. The user first selects the object in the search results, then selects the region of the object and then adds the annotation. The annotation can be accessed when the annotated object is accessed, and by specifying in the search that the search should also include annotations.

Once the search results are available, and the user selects the collection tool, the user can start adding and removing the objects from the collection. Finally the collection is described and saved. Same as for the annotations, the collection information is saved in the user's personal space, and can be also used in search, by specifying to the search that it should include the personal collections in the search. This means that the search will look not only for objects descriptions and metadata, but will also look inside collection descriptions to try and match the query.

In order to help in exploring the datasets, there are statistics and visualization tools. Statistics tool shows the basic properties of the repository. A visualisation tool is used to accompany the search tool (**Fig. 3**). Story creation tool is another tool that supports collaborative user generated content. It provides users with an online document editor, where users can write their own ideas, notes, insights about the data. The document created in this way is stored in the repository and can be accessed in the personal workspace, and shared with the others.



**Fig. 3.** Visualization tool. Visualization tool is used to accompany the search tool by showing live visualizations as the search results change. When the search is ran and the search results are available the visualization tools automatically update itself and shows the visualisation related to the current search results.

3. *Archaia Kypriaki Grammateia*, the STARC Repository and the EAGLE infrastructure

One of the most interesting collection stored in the repository is the *Archaia Kypriaki Grammateia* Digital Corpus [Pitzalis et al., 2012]. The collection is composed of Ancient Greek and Latin epigraphic texts produced within a time span of circa 13 centuries (from the 7th century BC to the 6th century AD). The texts are attributed to Cypriot authors or were produced in Cyprus. The *corpus* consists mainly of funerary or dedicatory epigrams published with their translation in Modern Greek, *critical apparatus* and philological comments [Voskos, 1997].

The peculiarity of the digital collection consists in the fact that, beyond the epigraphic texts, it is composed of a series of multimedia digital resources that describe the content in a multidisciplinary way: digital texts, images related to the epigraphic supports, 3D representation of the inscriptions, video, audio files, and so forth (Fig. 5).

Object				
Format	Text	Image	3D	Audio

Fig. 5. Multimedia data constituting the AKG digital collection stored in the STARC repository.

Such a material needs to be described in the right way, in order to provide detailed information about all the elements that constitute the collection. For this reason a specific metadata schema has been developed: it covers the multidisciplinary research carried out on this material in a manner



that users comprehend the complexity of such an approach through access to heterogeneous types of information [Vassallo et al., 2013].

The metadata schema for Ancient Cypriot inscriptions integrates multidisciplinary information regarding the objects and their multiple digital resources. It is the result of a research line developed within the group [Vassallo et al., 2013] and of the assessment and comparisons of schemas, models and ontologies already in use in the digital epigraphy field (e.g. TEI Epidoc, Dublin Core, CIDOC-CRM).

The metadata schema is organized in groups corresponding to different research areas clustered in wrappers and sub-wrappers, in order to fully describe all information that the 'asset inscription' contains. This schema, distinguished by its multidisciplinary structure, include all disciplines relevant for description and representation of epigraphies: archaeology (investigating for example the context of the finds), philology (analyzing the text, the writing style, the scripture), chemistry and geology (providing details on the material upon which inscriptions were carved), conservation (giving information about the state of the artefact), visualization and museology (about the museums and places of conservation).

Besides the inscriptions, since the Cypriot collection consists of their digital representations (pictures, 3D models of inscriptions, videos, etc.), the metadata schema takes into consideration their descriptive features, their digital provenance and other related information. For example, through the metadata it is possible to describe the digital provenance of the 3D model of an inscription and to give information about the acquisition phase (the technique, the tool used, the specification of the tool, the specification of the output, etc.) the post-processing (the operative info, the file specifications, etc.) and the digital output (data format, software), according to the digital resource obtained (**Fig. 6**).

The metadata schema for ancient Cypriot inscriptions aims at the following goals:

- to describe in detail the digital resources and its digital provenance,

- to provide related information about the context of the inscription,
- to enable harvesting to larger initiatives (e.g. Europeana, Ariadne).

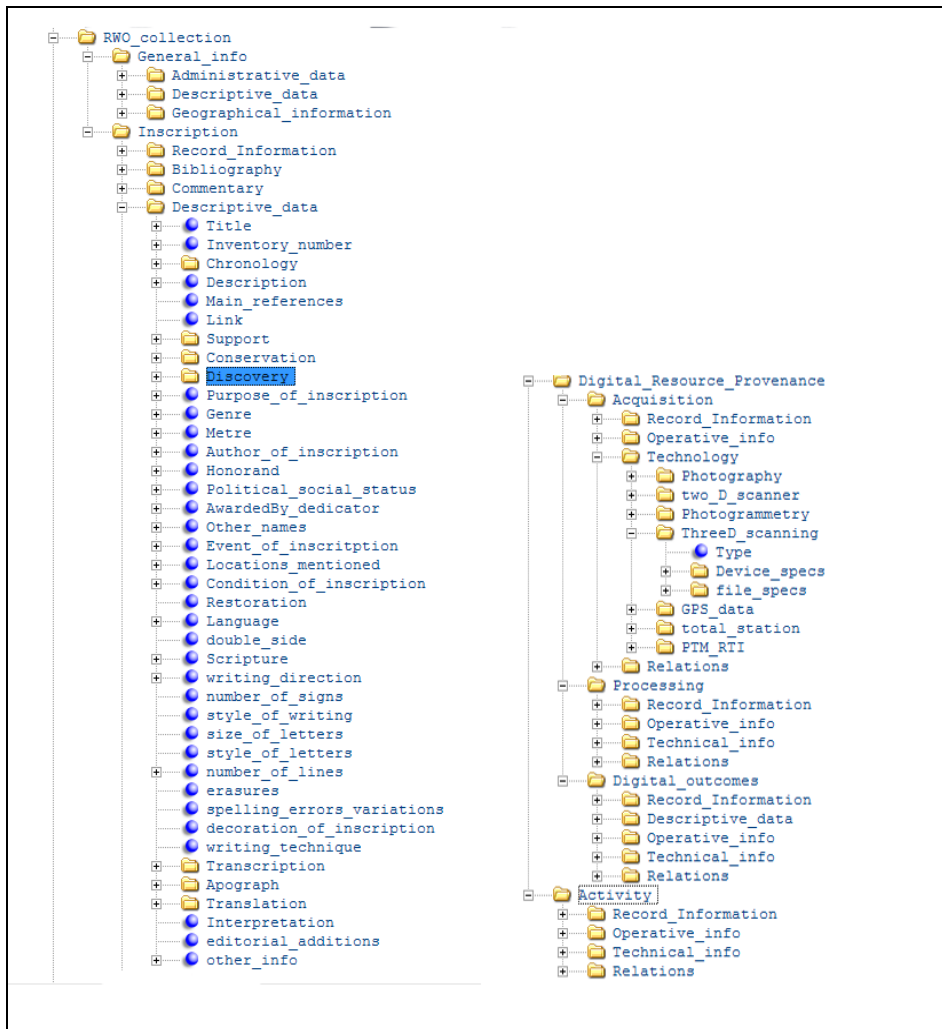


Fig. 6. Diagram of the metadata schema components and relationships.

### 3.1. From STARC Repository to EAGLE infrastructure

Beyond the functionalities developed for an enhanced users' access to the content, the STARC repository plays as aggregator for different digital libraries projects that in turn provide content to Europeana portal. The epigraphic content of the *Archaia Kypriaki Grammateia* digital corpus is ingested through the EAGLE Project.

The “Europeana network of Ancient Greek and Latin Epigraphy”, is an EU funded project under the umbrella of the CIP-Best Practice Network and whose aim is to bring together some of the most prominent European institutions and cultural archives in the field of Classical Latin and Greek epigraphy. One of the aim is to provide Europeana with a comprehensive collection of unique and curated online edition of historical and archaeological sources.

To encode inscriptions from different epigraphic databases, EAGLE developed a metadata format that assessed the provider's metadata structures and considered two sets of standards: TEI EpiDoc and CIDOC CRM. EpiDoc allows a full description of the text of inscriptions while CIDOC CRM enables a further full description object-oriented, reflecting at the same time the different souls of Epigraphy, the philological and the archaeological one. Content providers use different models according to the fact their databases are more oriented towards the text of an inscription or towards the archaeological object that bears the text. According to the epigraphers' community needs, EAGLE consortium developed the data model on the base of the two sets of standards [Liuzzo et al., 2014]. Finally, the EAGLE data model is mapped to the Europeana Data Model (EDM), in order to be ingested into the Europeana portal.

The content provided by the EAGLE consortium, due to the transformation into the EAGLE metadata model, produce three different groups of metadata according to the digital objects they are connected to: artefacts, documental manifestation, visual representation. The data that are sent to Europeana belong to the category of the artefacts (the inscriptions represented as texts, called by Europeana “Cultural Heritage Object” - CHO)

and of the visual representation (the images related to the inscriptions, called by Europeana “WebResource”). The distinction between CHO and WebResource has been introduced in Europeana as a result of the introduction of the new EDM schema to provide the users with a better navigation and data retrieval and to aggregate under the same umbrella all the resources available for a Cultural Heritage Object, avoiding data duplication in the portal.

Most of the consortium data, as well as the majority of the existent epigraphic databases worldwide, are compliant to Epidoc. This allows to perform a smooth mapping to the EAGLE data model, to have results that are aligned to the project aims (e.g. the possibility to create dedicated groups of metadata associated with the category of the object) and to simplify the involvement of other databases in EAGLE through the enlargement of the consortium.

As previously mentioned, the *Archaia Kypriaki Grammateia* digital corpus data integrated in the STARC Repository are different from other epigraphic databases data. The study of the epigraphy is not text-centric. The text is one of the representations of the artefact: the archaeological object, the support, the 2D image or the 3D of the object are all resources that, even if connected, have their own identity and their own set of information. Moreover, the AKG is a collection of texts accompanied by their translations, notes and commentaries as published in its last edition by Voskos [Voskos, 1997], therefore it is a digital form of this important editorial printed work, to which are connected all a series of digital resources that help to enrich the *corpus*.

Such a material and the metadata schema used for the description of this content converges in a difficulty to map our metadata schema to EAGLE data model. This implies a further effort to integrate the content into the EAGLE infrastructure for a compliant visualization with the other content and retrieval in the EAGLE and Europeana portals. Concerning the issues connected to the mapping and integration of the AKG data into EAGLE

portal, possible solutions have been investigated and are currently under tests:

- the creation and use of relations that are able to aggregate all the resources under a specific resource identified as reference item.
- the creation of rules for the development of a script that will be able to map to the Eagle data model every time according to the type of object we are dealing with. This implies also an editing of the metadata, creating single items that have as head the information about the inscription (e.g. the Ancient Greek text) and to which are attached all the other metadata sets concerning the related objects (e.g. the Modern Greek text, the image support of the inscription, the commentary, the 3D of the support-inscription, etc.).

Even if the first solution would be much faster from a technical point of view, at the moment is under preparation a test for the second solution. In fact the latter, even if it is time-consuming in terms of elaboration, seems to be the best solution for avoiding data duplication and for having more cohesive data.

## **4. Conclusions**

We presented in this paper a digital data repository and showed number of innovative tools used to access and explore its collections. One of the most interesting collection stored in the repository is the Archaia Kypriaki Grammateia Digital Corpus composed of Ancient Greek and Latin epigraphic texts. We argued that by interacting with data in an innovative ways can help users better understand data and stimulate knowledge creation. With the set of proposed functionalities we also enabled exploration of the epigraphic text in a way that supports sense making, understanding and collaboration. We also showed how the collection of epigraphic texts can be used within other projects by mapping our data to the appropriate data formats. Next step is to set up an evaluation framework that will try to measure and evaluate the contribution of each tool to the users

daily tasks. We want to measure the benefits of using such a tools by performing various evaluation tasks and measuring the user's performance.

## Bibliography

ALGEE, L., BAILEY, J., OWENS, T., *Viewshare and the Kress Collection: Creating, Sharing, and Rapidly Prototyping Visual Interfaces to Cultural Heritage Collection Data*, D-lib (2012).

BERGSTORM, P., ATKINSON, D.C., *Augmenting the exploration of digital libraries with web-based visualizations*, ICDIM, (2009), 1-7.

BERNARD, J., RUPPERT, T., SCHREER, M., KOHLHAMMER, J., SCHRECK, T., *Content-based Layouts for Exploratory Metadata Search in Scientific Research Data*, JCDL, (2012), 139-148.

BUCHEL, O., SEDIG, K., *Making sense of document collections with map-based visualisations: the role of interaction with representations*, Information Research (2014) 19(3), paper 63.

FAST, K., V., SEDIG, K., *Interaction and the Epistemic Potential of Digital Libraries*, Int. J. Digit. Libr., (2010), vol. 11, n. 3, 169-207.

FERREIRA, N., LINS, L., FINK, D., KELLING, S., WOOD, C., FRIERE, J., SILVA, C., *BirdVis: Visualizing and Understanding Bird Populations*, Visualization and Computer Graphics, IEEE Transactions on, (2011), 2374-2383.

HIENERT, D., SAWITZKI, F., SCHAER, P., MAYR, P., *Integrating Interactive Visualizations in the Search Process of Digital Libraries and IR Systems*, Advances in Information Retrieval Lecture Notes in Computer Science, (2012), 447-450.

LIUZZO, P.M., RIVERO RUIZ, E., VASSALLO, V., *Networking EAGLE with CIDOC and TEI*. CIDOC 2014: Access and Understanding – Networking in the Digital Era, Dresden, Germany; 10/2014.

NAN, C., YU-RU, L., XIAOHUA, S., LAZER, D., SHIXIA, L., HUAMIN, Q., *Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time*, Visualization and Computer Graphics, IEEE Transactions on, (2012), 2649-2658.

NASAR, M., R., A., MOHD, M., ALI, N., M., *A conceptual framework for an interactive personal information management system, i-USer*, (2011), 100-105.

PITZALIS, D., CHRISTOPHOROU E., KYRIAKOU N., GEORGIADOU A., NICCOLUCCI F. (2012). *Building scholar e-communities using a semantically aware framework: Archaia Kypriaki Grammateia Digital Corpus*. In: VAST12 (2012) UK, 89–95.

REN, L., TIAN, F., ZHANG, L., DAI, G., *DaisyViz: A Model-based User Interfaces Toolkit for Development of Interactive Information Visualization*, Visual Information Communication, (2010), 209-229.

RONZINO, P., NICCOLUCCI, F., HERMON, S., *A Metadata schema for Cultural Heritage documentation*, in EVA 2012 Florence, Firenze University Press.

VASSALLO, V., CHRISTOPHOROU, E., HERMON, S., NICCOLUCCI, F., 2013, *Revealing cross-disciplinary information through formal knowledge representation – a proposed Metadata for ancient Cypriot inscriptions*. In Conference Proceeding of Digital Heritage Heritage International Congress (Digital Heritage), IEEE. Ed. by A. Addison, L. De Luca, G. Guidi, and S. Pescarin. Vol. 2, pp. 79–82.

VASSALLO, V., RONZINO, P., DAMNJANOVIC, U., HERMON, S., *Un processo aperto di comunicazione della conoscenza*. ArcheoFoss 2013, 18-19 June 2013, Catania, Italy. In print on Archeopress.

VOSKOS, A. (1997). *Ἀρχαία Κυπριακή Γραμματεία*. Nicosia: Epigramma.

WONG, W., CHEN, R., KODAGADA, N., ROONEY, C., XU, K., *INVISQUE: Intuitive Information Exploration Through Interactive Visualization*, CHI EA, (2012), 311-316.