

Sharing ICH Archives: Integration of Online Inventories and Definition of Common Metadata

Maria Teresa Artese, Isabella Gagliardi

A CNR IMATI – Milano, Via A. Corti 12, 20133 Milan, Italy
artese@mi.imati.cnr.it, gagliardi@mi.imati.cnr.it

Abstract. We present here an ongoing study and experimentation of a system (QueryLab) able to query different inventories at the same time, providing a larger virtual collection that can be explored through intuitive and attractive search tools. The aim is twofold: overcome data silos archives and develop a common metadata able to collect lists of intangible cultural heritage that is not yet included in any online database.

Keywords: Global and Local Archives, Multimodal Search Criteria, Open Archives, Storytelling, Common Metadata.

1 Introduction

Numerous inventories regarding tangible and intangible heritage have been published on the web thanks also to the UNESCO Convention of 2003, which has given a great impulse to spread of cultural heritage among ordinary people (<https://ich.unesco.org/>). Based on their organization of data, they can be classified into different categories:

- Data silos web sites, in which data are strictly closed and structured with owner metadata.
- Web sites who collect data from different sources and store them into a common structure.
- Web sites who provide web services to open their data to researchers.
- Web pages without any underlying database or storing structure, in which intangible cultural heritage (ICH) items are listed and described as plain text on static web pages.

From the analysis of a number of different websites concerning cultural heritage inventories, we observed that often they require specific know-how to understand how to search data and frequently they provide users with a single blank field for searching, in front of which the user asks: “what can I look for?”, “what’s in here?” (Clough, Hill, Paramita, & Goodale, 2017). We also realized that particular search criteria developed for a specific ICH archive could be useful and interesting if applied to another or to different archives at the same time. Finally, we studied how to involve and share the collections published on static web pages, developing a common metadata structure

able to collect and store data in a local database, which can be integrated and opened by building dedicated web services.

These are the starting points for the ongoing study and experimentation to develop a prototype, named *QueryLab*, able to query different inventories simultaneously - providing the user with a larger virtual collection to navigate - through intuitive and attractive search tools able to reach a wider audience and overcome two limitations: data silos archives and rigid query criteria.

2 Inventories and Tools

The main difficulty is to collect websites related to intangible cultural heritage, indeed there are a lot of websites related to museums, photographs, arts and music, but living traditions inventories or oral traditions, practices, skills collections are not so easy to find (Artese & Gagliardi, UNESCO Intangible Cultural Heritage management on the web, 2015). Selection criteria used for the choice are:

- Availability of web services giving access to data, able to extract data from the online database and make queries.
- Access to a console where search criteria can be tested using the functions provided by the web service.
- Access to a full documentation of the protocol functions and fields available, as for the search phase and retrieved documents and how to manage them.

2.1 Online Inventories

The analysis of web sites collections has led to a list of inventories which meet the criteria and have been successfully integrated:

- **Europeana.** Is the EU digital platform for cultural heritage. Contributed by more than 3,000 institutions across Europe, their assembled collections let users explore Europe's cultural and scientific heritage from prehistoric to the modern day (<http://www.europeana.eu>).
- **Victoria&Albert Museum.** Is the world's leading museum of art and design (<https://collections.vam.ac.uk/>), holding many of the UK's national collections and housing some of the greatest resources for the study of architecture, furniture, fashion, textiles, photography, sculpture, painting, jewellery, glass, ceramics, book arts, Asian art and design, theatre and performance.
- **Cooper-Hewitt, Smithsonian Design Museum.** Is the only museum in the United States focused on historical and contemporary design (<https://www.cooperhewitt.org/>) and is the curator of one of the most various and exhaustive design collections in existence - more than 210,000 design objects spanning 30 centuries.
- **Lombardy Digital Archive.** Is the online Archive of Ethnography and Social History of Lombardy Region that, since 1972, preserves, studies and enhances documents and images related to life and social transformations, to literature

and oral history, to material culture, to the anthropic landscapes of the Lombard territory (<http://www.aess.regione.lombardia.it/ricerca>).

- **Museum of Contemporary Photography in C. Balsamo (Milano).** Contains a subset of over one million and 800 thousand photographic works - prints in black and white and color images, slides, negatives, videos, installations - taken by about five hundred Italian and foreign authors. It is a significant example of contemporary photographs and a cross-section of photography after World War II to the present days (<http://mfc.itc.cnr.it/home.htm>).

2.2 Used Tools: Web Services

Today an increasingly number of online inventories give access to their data through web services, by which users can query and retrieve information from the outside. Normally there is a specific web site, called endpoint, where a RESTful API web service (<https://restfulapi.net/>), based on HTTP protocol, provide communication with data (REpresentational State Transfer).

The integration of both local and online archives to query them all together can be performed using such tools: it is necessary to have a service available for each online inventory selected and to develop a specific web service for local archives. The development of a web service for local archives has a twofold benefit because gives also the opportunity to open local archives for public machine query.

As regards online inventories, each REST API method provided to query data has been studied and encapsulated into a software layer used to access records and single fields returned from data extraction. Key aspects to pay attention to are:

- A free API key or token would be easily available to query data,
- A method to query data is available, not only a simple extraction of all records of the collection,
- The web service provides methods to query data giving specific destination fields, such as location, author, date. This is useful to refine queries or implement different ways to interact with data,
- Sometimes web services offer the “faceting fields”, useful to extract indexes on specific fields, such as keywords, period of time or location.

3 The Prototype Developed: *QueryLab*

The ongoing experimentation has produced a prototype, named *QueryLab* (<http://arm.mi.imati.cnr.it/querylab/index.html>), with the purpose to collect local and global archives together and to apply them innovative query criteria and storytelling tools (Artese, Ciocca, & Gagliardi, 2017), easy to understand and accessible for users. The structure of *QueryLab* system is shown in Fig. 1.

The query functions software layer contains all the procedures dedicated to interaction to and from different web services, in a transparent way, even for any archive available locally. The core is the design of new query methods, easy to understand and accessible for users, where all inventories are accessible simultaneously with the same

query string. Tools developed so far concern the combination between keywords, temporal data and spatial information, which are common to all examined inventories.

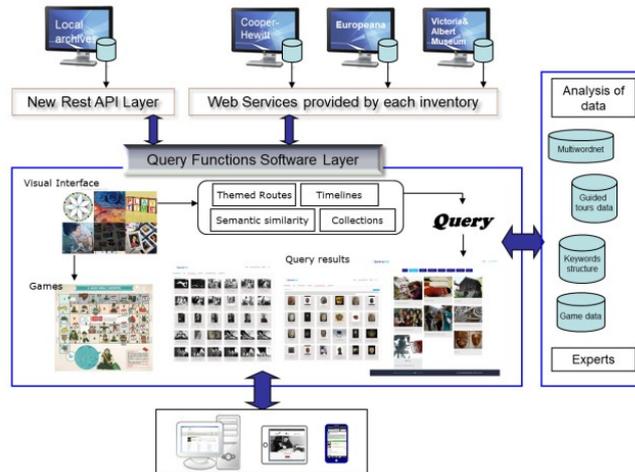


Fig. 1. The QueryLab system structure.

3.1 Thematic Routes

Thematic routes come from the analysis of the most relevant terms and keywords used to query local ICH archives. Terms often recurring have been analyzed by experts and divided into categories and sub-categories, in a tree organization, which are gradually updated and increased as new online inventories are added. Each category is finally equipped by images and translations to allow the creation of multilingual predefined paths, among which users can browse, exploring and retrieving semantically similar documents.

The procedure has given satisfying results when applied to the local archives used at first, but it is interesting to see how these category trees are useful to search other inventories, allowing a quick and easy document extraction from huge archives such as Europeana where it is not easy to understand at first glance what it can offer (Gordea, Haskiya, & Marriot, Europeana DSI 2–Access to Digital Resources of European Heritage. D6.1: Advanced image discovery report., 2017), (Gordea & Simon, Europeana DSI 2–Access to Digital Resources of European Heritage. D6.4: Pilot for time-and-place discovery., 2017).

3.2 Search and Show

When the user types a keyword to make a query, or chooses one from predefined lists of tags extracted from data, the keyword is filtered through the Multiwordnet system, a structured dictionary available online (<http://multiwordnet.fbk.eu/english/home.php>)

(Artese & Gagliardi, Multilingual specialist glossaries in a framework for intangible cultural heritage, 2014), by which it is possible:

- To search both English and Italian terms to get relative translations,
- To get all different meanings, called synsets, of a keyword, and use all terms belonging to a specific synset to make the queries,
- To link each keyword with the tree structure taken from the Multiwordnet hierarchical structure and use it to enlarge or refine the query.

Given a keyword the system retrieves the translations and the tree structure from Multiwordnet and show all available information to the user to make the query. For example if the key is “Magic” the upper-level (hyperonyms) keys retrieved are: “Supernaturalism”, “Performance” and “Supernatural” and all coordinates terms for each hyperonym are listed both in English and Italian language. Now the user can select one of these synsets and retrieve all documents related to it. If the one chosen is “Magic trick” the query is made by using all terms taken from the synset, in both languages: “gioco di prestigio, trucco magico, magia, magic trick, conjuring trick, trick, magic, legerdemain, illusion, deception”, as found from Multiwordnet search.

3.3 Collections and Games

This section is dedicated to new ways to navigate in the contents, oriented to communicate with a wider audience and common people. The creation and management of users' personal collections originates from the need to avoid the loss of retrieved documents that occurs when the browser is closed. Personal collections are fed by the results of queries made over time and provide a valuable tool to improve knowledge of heritage materials. The designed procedure allows the user to store the queries performed and the resulting documents, or a part of them, label them, organize them in topics of interest and tag them with personalized keywords. It is also an effective tool because these small virtual archives can be shared with others, to spread and promote contents that are of great interest for the user.

A game section is under construction. Starting from the analysis of local inventories combined with a work of post-editing done with specialized experts it is possible to develop new tools enriching the archives with a play section. The user can save his game and, by exploring the archives to find the correct answers, can improve his score and his knowledge.

4 Common Metadata for ICH Inventories

From the analysis of data from different sources, we identified common aspects useful to determine which metadata could best collect and describe them, focusing on these two aspects:

- How to organize information needed to collect and preserve intangible assets from different parts of the world: current situation, deepening and refining to adapt to the various types of ICH objects that can be collected,

- How to structure metadata to ensure adaptability to different requirements of collected data and to be able to follow its evolution and changes over time.

The structure obtained, called "ICH light", is intended to collect and store the subset of minimal information needed to perform queries, together with the link to the original online resource, when available. If not available, the system can produce a web page showing all the information collected from the metadata, acting as a safeguard platform for those collections that do not have one. A back-end data access module to ensure updating and adding of new items has been planned and will be developed shortly.

Two collections have been used for the first test of the metadata: IntangibleSearch and the Asia-Pacific Database on ICH. The first is the online database of Lombardy Region containing oral traditions, languages, performing arts, technical knowledge, social practices, rituals and festive events (<http://www.intangiblesearch.eu>) while the second is the "Data Bank on Traditional/Folk Performing Arts in Asia and the Pacific", collected by the Asia/Pacific Cultural Centre for Unesco (ACCU) and provided with simple HTML pages (<http://www.accu.or.jp/ich/en/about/introduction.html>). Items have been migrated into the new metadata structure and the resulting local database has been integrated together with the others available online.

5 Conclusions

We presented here a prototype, *QueryLab*, able to query different inventories at the same time. The use of web services to query inventories allows new archives to be added at any time and the developing of a common metadata structure allows to collect and share "living goods" which are not in a database. Future developments include the completion of the back-end module, the games and collections sections and the transition to the MONGODB database for the common ICH database, which facilitates the multilingual approach. It is planned to add a "near to me" function, capable of extracting all that is available in a given place of interest for the period indicated in relation to a given theme, obtaining data from all the inventories involved.

References

- Artese, M. T., Ciocca, G., & Gagliardi, I. (2017). Evaluating perceptual visual attributes in social and cultural heritage web sites. *Journal of Cultural Heritage*, 26, 91-100.
- Artese, M., & Gagliardi, I. (2014). Multilingual specialist glossaries in a framework for intangible cultural heritage. *International Journal of Heritage in the Digital Era*, 3(4), 657-668.
- Artese, M., & Gagliardi, I. (2015). UNESCO Intangible Cultural Heritage management on the web. In *Encyclopedia of Information Science and Technology, Third Edition* (pp. 5334-5347). IGI Global.
- Clough, P., Hill, T., Paramita, M. L., & Goodale, P. (2017). Europeana: What users search for and why. *International Conference on Theory and Practice of Digital Libraries* (pp. 207-219). Springer, Cham.

- Gordea, S., & Simon, R. (2017). *Europeana DSI 2–Access to Digital Resources of European Heritage. D6.4: Pilot for time-and-place discovery.*
- Gordea, S., Haskiya, D., & Marriot, A. (2017). *Europeana DSI 2–Access to Digital Resources of European Heritage. D6.1: Advanced image discovery report.*
- REpresentational State Transfer*. (n.d.). Retrieved May 28, 2019, from Wikipedia: REpresentational State Transfer (n.d.). In Wikipedia. Retrieved May 28, 2019, from https://en.wikipedia.org/wiki/Representational_state_transfer

Received: May 25, 2019
Reviewed: June 26, 2019
Finally Accepted: July 16, 2019

