A Knowledge model for supporting digital exploration of historical industrial archives

Abstract

Business Archives (BA) represent a specific sector of archival science concerning the management of historical archives relative to the complex documentation produced by entities (creators) carrying out an economic activity. A particular area within BA is the industrial archives that handle documentation of the world of industry and labour. The substantial differences that characterize BA compared to an historical archive in general are the particular archive structure, its characteristics and the type of materials that it retains (e.g. photos, video and audio files, 3D objects, text documents). These elements arise from the documentation production context, from the creator type and from the different business activities.

In recent years several initiatives have been undertaken by the international research community to provide common digital archive description and access solutions (http://www.apex-project.eu/index.php/about-apex/related-projects).   However, although BA are of utmost importance for the cultural development of a country and represent valued sources for historical research (Deserno, 2011), their preservation, enhancement and access remain a challenge to be faced. Because of their complexity in terms of heterogeneity of content, BA need specific management and appropriate metadata to describe their structure and contents (Carucci & Messina, 1998); the content description has to take into account that diverse users' communities with large spectrum of competences are interested in archives exploration and thus content understanding might be difficult for users with different cultural background and skill levels (e.g. historian, student, citizen).

The paper proposes a replicable methodology for the definition of a knowledge model to support the different users' communities in the exploration of digital industrial archives. The research activity is performed within the project R.I.C.E.R.C.A. (Business Cultural Resources Preservation and Accessing), an Italian project funded by the European Community, aimed at improving the BA management and access. The project focuses on the historical industrial archives preserved at Ansaldo Foundation, which represents one of the most relevant Italian archive institution (De Martino, Monti, Mosca, Pasini, Pastorino, Rosati & Viano, 2013).

The adopted methodology is characterized by a two-step process: (i) analysis of the state of play of existing standard for metadata and of available domain terminologies

and (ii) definition of a knowledge model for industrial archive including categories of descriptive concepts of the archive content, categories of archive series and semantic relationships among concepts and series.

The first process step investigates the state of play of both the existing standards for metadata (http://www.ica.org/10206/standards/standards-list.html) and the available Knowledge Organization Systems such as thesauri, code lists, ontologies and other sematic repositories, than could be exploited for the definition of a knowledge model enabling a common understanding of meanings and terms ("UKAT" 2014; "Nuovo Soggettario", 2014).

A significant set of representative use cases has been identified and the way users' access Ansaldo Foundation archives has been analysed to identify user's requirements for the archive exploration. The selection criterion of the use cases is based on the users' skill and their different research purposes. Among the wide range of potential users, it is possible to identify two main categories according to different cultural background: the professional user (historian, archivist) with specific knowledge background, scientific approach and expertise in archive consultation and the common user (architect, company employee, journalist) with various education levels and lower competences in archive consultation. Anyway both the categories require the support of an archivist to satisfy their requirements.

The results of the first process step confirm that a shared terminology specific for BA domain is not yet available and needs to be defined. Moreover the emerged user requirements highlight that the standard metadata applied to the description of BA are insufficient to provide complete and proper content description and to capture the complexity of the archive resources. The use case analysis reveals not just the importance of advanced digital tools for resource searching by keywords but also the need of exploiting semantics to allow digital archive exploration through domain descriptive concepts.

The second process step aims at defining a knowledge model to support the user exploration of the archive content, by identifying categories of descriptive concepts for industrial archive and their relationships with different types of archive series. It grounds on the cultural background and expertise of archivists specialized on BA management; starting from available taxonomies and subject headings compiled by institutions with archival holdings dedicated to the BA management (www.gla.ac.uk/services/archives; http://id.loc.gov/authorities/subjects.html), the categories of descriptive concepts, recognized as significant in the description of the contents of the BA domain (e.g. Historiography, Company's organizational structure, Industrial production) are identified, in order to classify the archive contents and to facilitate the exploration. Afterwards, analysing the content of Ansaldo Foundation archives with the support of the available finding aids (Cerioli & Rosati, 2011; Vesco,

2012), the initial sets of categories of descriptive concepts are refined to adequately cover the content description.

Moreover a set of categories of series (e.g. Financial, Personnel, Property management, patents), recognized as common and relevant in the industrial archives, has been identified. Afterwards, analysing the existing series in the Ansaldo Foundation archives with the aim of mapping them with the identified categories, issues are emerged to be faced: the *Model of the levels of arrangement of a fond* according to the standard ISAD(G) ("ISAD(G)", 2014) provides the structure of a generic archive and does not reflect the complex structure which might exist in the industrial archives. This has required a refinement of initial series' categories to comply with the industrial archive complexity.

Finally, the relationships between descriptive concepts and series' categories have been identified.

The defined knowledge model has been formalized in an ontology and it has been validated via simulation of the archive exploration according to the use cases. The ontology will be exploited to facilitate data searching and archives exploration for the Ansaldo Foundation digital archive portal that is under development.

The defined model refer to industrial archives and more specifically to Ansaldo Foundation archives, however the adopted methodology is applicable to any other BA types.

References

Carucci, P., & Messina, M. (1998). *Manuale di archivistica per l'impresa*. Roma, Italy: Carocci Editore.

Cerioli, C., & Rosati, R. (Eds.). (2011). *Archivio Perrone 1871–1945. Inventario*. Genova, Italy: Fondazione Ansaldo Editore.

De Martino, M., Monti, M., Mosca, G., Pasini, R., Pastorino, S., Rosati, C., & Viano, G. (2013). *Semantics for the exploration of historical Business archives. Challenges and Perspectives in the R.I.C.E.R.C.A. project*. In A. C. Addison, L. De Luca, G. Guidi & S. Pescarin (Eds.), *Proceedings of the 2013 Digital Heritage International Congress (DigitalHeritage) vol. 2* (pp. 401-404). New York, USA: Institute of electrical and electronics engineers.

Deserno, I. (2011). *The value of international business archives: the importance of the archives of multinational companies in shaping cultural identity*. *Archival Science*, 9 (3), 215-225.

ISAD(G), General International Standard Archival Description - Second edition (2011). Retrieved April 16, 2014,from http://www.ica.org/10206/standards/standards-list.html

UKAT (UK Archival Thesaurus) (2014). Retrieved April 16, 2014, from http://www.ukat.org.uk/

Vesco, A. (Ed.). (2012). *Archivio Oto Melara 1902-1972. Inventario*. Genova, Italy: Fondazione Ansaldo Editore.