How Digital Cultural Heritage Resources can Lead to New Understandings in the Humanities: Future Challenges for Digital Libraries and Archives

Maristella Agosti

Department of Information Engineering, University of Padua, Italy
maristella.agosti@unipd.it

Abstract. This paper reports on the presentation made during the panel on "Digital Libraries and Digital Archives: Problems and Challenges for AI Approaches" of the 1st Workshop on Intelligent Techniques At LIbraries and Archives (ITALIA 2015) co-located with the XIV Conference of the Italian Association for Artificial Intelligence, 22 September 2015, Ferrara, Italy[1].

Keywords: digital cultural heritage, digital cultural heritage resources, digital libraries systems, digital archives systems, interoperable systems, digital humanities

1 Introduction

Throughout their history, cultural heritage institutions such as libraries and archives have had two central purposes. They have been charged firstly with preserving artefacts of cultural significance, and secondly with describing and cataloguing these artefacts in a way that makes them accessible to a variety of audiences, from experienced researchers to members of the general public [10]. While the evolution of these roles is not without its challenges, and while the requirements of preservation and access often stand in direct opposition to each other, the stable and long-established role of traditional cultural institutions means that there is general agreement on the responsibilities of institutions, the processes and procedures by which their work is carried out, and an established statutory and legislative basis for their work.

The advent of the widespread digitization of cultural heritage collections has significant implications for institutions that hold these types of collections. The imperative to preserve unique and delicate resources by producing digital surrogates has been, and continues to be, an important driving force for digitization projects that see the participation of cultural heritage institutions together with research and industrial organisations experts in computer science.

---

[1] URL: http://italia2015.dei.unipd.it/

Important though preservation undoubtedly is, access is more important still. As such, computer science can play an important role in contributing to the creation of the necessary methods for inventing and designing systems able to manage and preserve digital resources in a way that the systems themselves can be enjoyed by both non-specialist and specialist users. The different types of systems of this sort are often named "digital library system" [2] even though they are mostly focused on digital resources of archives and museums.

2   Creation of New Models and Systems

A new generation of systems that manage and preserve digital resources of cultural heritage is based on methods of active involvement and interaction of different categories of users [1]. Indeed, the active involvement of computer science experts together with domain experts will contribute to the creation of new methods useful for the realization of systems and services providing tools that allow different types of users to interact with these new systems; these new systems could also provide customizable services that can be adapted to the specific and diversified needs of users.

The conception and realization of this new generation of management systems will concern in particular the need to create new models for automating processes of representing and processing specific cultural heritage resources that we want to represent and manage in digital form. Depending on the type of cultural resources, which from time to time can be of specific interest, the design process of a new model of information management can result from effective collaboration between experts in the specific domain of cultural heritage and experts in information and computer science. In fact, the experts in the specific domain of cultural heritage – such as archives, art history, library science, archeology, linguistics or history – know the history and methods of their specific domain, while the experts in information and computer science know the story and methods of computing. In a synergistic relationship for the development of new methodological solutions, experts in the two sectors may consider the functional and informative requirements as well as the tasks users need to carry out and come up with new methods and solutions.

The design process then is not limited to the analysis of user requirements conducted in isolation by the computer science experts. Instead it requires a synergistic cooperation of the computer scientists together with specialists in the specific field of cultural heritage of interest. Once the new model has been created, a corresponding new information management system can be devised. Computer science then becomes one of the cultures needed to design the new system, which, through an innovative approach to management, will produce new knowledge that could not previously be represented or processed.

3   Why New Models and Systems Are Important

We could ask us why we need to devise new models and systems. The answer is because we want to consider aspects of reality that are more complex than those that were previously addressed. As we increase the complexity of the aspects of reality that we want to address and manage, we need new methods and systems capable of dealing with them and managing them. Bearing in mind the greater complexity of the aspects of reality that we want to address, we need to devise methods to match them and systems to manage them.

The first step towards tackling a new aspect of reality, which will be managed with the help of computer methods, is to highlight the different levels of possible intervention and divide the overall problem into sub-problems, as is common when devising a scientific solution for solving a complex problem. The sub-problems, in the context of information processing systems, are identified and highlighted by considering the interaction that the system we want to build is going to have with its users. In fact, the system has to be used by different categories of users through an interface that has the features and functions that users need. The interface capabilities are based on internal processing features, functions and data management information that are usable and manageable.

Consequently, a system of this type is usually represented on three levels:

- the *external level*, where the interaction takes place between the users and the presentation layer services provided by the information processing model designed and manufactured by the system;
- the *intermediate level*, or logic level of services, where the processing of information of specific interest to the application that is made available to the users is carried out; and
- the *innermost level*, or logic level of the data, where the information of interest is represented and stored to be made available to the user, possibly with the aid of permanent storage devices.

One thing to keep in mind when designing and creating a new system of information management is interoperability. Any system for the management and preservation of digital resources of cultural heritage does not operate in full "isolation" from other management systems, rather it needs to be "interoperable" with other management systems of information that are of interest and that can be related or connected to the new system.

There are several aspects to be taken into account if we want to build interoperable systems and provide qualified services to the final user [5, 6]. The different aspects that have to be taken into account include some that can be considered different dimensions of the problem [4]:

- *institutions* that work together to provide services to users,
- *information objects* that are represented and that can be enjoyed by the user,
- *functions* that are made available,

- *interactions* between users and systems,
- *technologies* that are put into play.

Examples of such systems have been made available in recent years in the context of the SIAR[2] and the CULTURA[3] projects [3, 6].

4 An Outlook to the Future

If we have correctly applied the new method to create systems suitable for representing and managing the various information resources that are of interest to diversified cultural heritage institutions – such as libraries, archives and museums both general and specialized – we now have at our disposal a number of systems that manage the representations of digital resources together with the information resources themselves.

While these systems become a practice that provide final and experienced users with those digital collections of information, we have available an amount of data that is professionally managed and that increases over time. This type of data is of the sort that is maintained in the so-called "curated databases", which require a great deal of human effort to populate and update. The value of curated databases lies in the organization and the quality of the data they contain [8].

This increasing quantity of data and related information provides us with a new way of exploring previously available digital resources together with the newly related and newly inserted resources. Although these systems have been designed and developed in line with the previously introduced new vision of creating digital libraries and archives systems, with the aim of better exploiting the digital resources managed by the different systems, practices would be needed that address and manage interoperability among systems at a higher level of abstraction than what is presently done and in line with what has been proposed in [7].

When the interoperability at a higher level of abstraction between different digital libraries and archives systems will be a reality, since most of the systems would be conceived using the combination of knowledge of the domain specialist together with knowledge of the computer science specialist, the resulting information resources would be richer and similar to those that at present are available in data-intensive science practice [9]. A hypothesis is that it will be possible to consider using those sources of information in an innovative way, because those sources contain data that have been selected and assembled in a way that can contain verified knowledge on the domain of interest. The study and correlation of those data will contribute to the discovery of new facts, the correction of facts previously considered correct, and in general will lead to new understandings in the area of humanities.

---

[2] URL: `http://www.regione.veneto.it/web/cultura/progetti`
[3] URL: `http://www.cultura-strep.eu/`

References

1. Agosti, M.: Information Access using the Guide of User Requirements. In: Agosti, M. (ed.) Access through Search Engines and Digital Libraries. pp. 1–12. Springer-Verlag, Heidelberg, Germany (2008)
2. Agosti, M.: Digital libraries. In: Melucci, M., Baeza-Yates, R. (eds.) Advanced Topics in Information Retrieval, The Information Retrieval Series, vol. 33, pp. 1–26. Springer Berlin Heidelberg (2011)
3. Agosti, M.: Una nuova generazione di sistemi di gestione e conservazione di risorse digitali del patrimonio culturale grazie al coinvolgimento di diverse comunità di utenti. Atti e Memorie dell'Accademia Galileiana di Scienze, Lettere ed Arti già dei Ricovrati e Patavina CXXIV (2011-2012), 11–22 (2012)
4. Agosti, M., Ferro, N.: Interoperabilità tra sistemi di biblioteche digitali. DigItalia: Rivista del digitale nei beni culturali 5, 95–112 (2010)
5. Agosti, M., Ferro, N., Silvello, G.: Handling Hierarchically Structured Resources Addressing Interoperability Issues in Digital Libraries. In: Biba, M., Xhafa, F. (eds.) Learning Structure and Schemas from Documents, Studies in Computational Intelligence, vol. 375, pp. 17–49. Springer Berlin Heidelberg (2011), `http://dx.doi.org/10.1007/978-3-642-22913-8_2`
6. Agosti, M., Ferro, N., Silvello, G.: Biblioteche digitali tra modellazione, gestione e valutazione. In: Ciotti, F. (ed.) Digital Humanities: progetti italiani ed esperienze di convergenza multidisciplinare. Atti del convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale, pp. 33–50. AIUCD 2012, Sapienza Università Editrice – Roma, Italy (2014)
7. Agosti, M., Ferro, N., Silvello, G.: Digital library interoperability at high level of abstraction. Future Generation Computer Systems 55, 129–146 (2016), `http://www.sciencedirect.com/science/article/pii/S0167739X15003003`
8. Buneman, P., Cheney, J., Tan, W.C., Vansummeren, S.: Curated databases. In: Lenzerini, M., Lembo, D. (eds.) Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2008). pp. 1–12. ACM (2008), `http://doi.acm.org/10.1145/1376916.1376918`
9. Hey, A.J., Tansley, S., Tolle, K.M. (eds.): The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, Wash. (2009)
10. Koolen, M., Kamps, J., de Keijzer, V.: Information Retrieval in Cultural Heritage. Interdisciplinary Science Reviews 34(2-3), 268–284 (Sep 2009)