An Innovative Web Application for Advanced Library Services

Maurizio Calderamo[1], Simona Ibba[2], Filippo Eros Pani[2], Francesco Piras[1]
and Simone Porru[2]

[1] SoSeBi Srl, Via dell' Artigianato, 9, 09122 Cagliari, Italy
{calderamo, piras}@sosebi.it
[2] Department of Electrical and Electronic Engineering, University of Cagliari,
Piazza d'Armi, 09123 Cagliari, Italy
{simona.ibba, filippo.pani, simone.porru}@diee.unica.it

Abstract. The future scenario in which libraries will work will see an increasing centrality of the Internet and of services provided through it, in order to meet the needs of physical and virtual users, which remotely access the library. SoSeBi srl, an innovative company located in Sardinia (Italy), and the Department of Electrical and Electronic Engineering (Dipartimento di Ingegneria Elettrica ed Elettronica, DIEE) of the University of Cagliari (Sardinia, Italy) intend to create a product at the forefront of the evolution of an ever-expanding sector, Information Retrieval, where general-purpose products like search engines, digital archives and social networks are competing. These products are, and will continue to be, a significant presence in our digital lives. Therefore, the objective of this research project is to create a software that responds to the users' tastes, that encompasses the Web 2.0 paradigms, and that is in close contact with technology innovation.

1 Introduction

In the last decade, libraries have gone from a heritage mainly made of paper, to a hybrid one, combining printed resources to multimedia materials of different kinds: music, audiovisuals, databases, electronic books, audiobooks, websites.

The evolution towards digital libraries, where reference services and loan of digital content can take place, is one of the means needed to meet the current and future needs of the citizens, more and more used to peruse digital contents of various types [1]; [2]. An especially important branch of the digital content area is eBooks, which have become increasingly widespread in Italy since the end of 2010. Despite the still small percentages compared to the paper book market, it is going to become a new reading support in the future. eBook is meant to represent the textual work, while the device used to read it is a different, separate topic. Currently, technology offers several types of eBook readers (called e-readers), from new-generation mobile phones (smartphones) to tablets (for example, the iPad), to readers specifically thought and designed with this purpose, among which Amazon Kindle is the most prominent example. An increasing number of Italian libraries is developing the need to provide reference and loan services for eBooks.

4

The presented project, financed by the Autonomous Region of Sardinia with European funds (Single Programming Document 2007-2013 - P.O. FESR 2007-2013 – Line of Activity 6.2.2.d – Interventions to support competitiveness and innovation, under the Regional Committee Resolution no. 33/41 of 08/08/2013), has the aim of implementing a Web-based application geared towards bibliographic cataloging and reference services in libraries, with innovative functions of semantic search, management of digital contents and the creation of a social network. The proponents also intend to cover the study of future interaction methods among libraries, eBooks, and eReaders.

The creation of the Web application will stem from the strategic partnership between SoSeBi Srl[1] and the Department of Electrical and Electronic Engineering (DIEE)[2] of the University of Cagliari. The purpose of this choice to use the results of fundamental research as well as of industrial research to elaborate an innovative prototype, unique in the domestic market for its features.

The software prototype is designed to have a high-flexibility modular architecture. In this way, the different functions can be independent, and not hinder the essential core of the management program in their development. The new software product aims to be a reference point in the Italian market of management software for libraries, in terms of innovation and completeness of integrated functions. A new-generation application will be created, with an innovative vision on the very concept of "automation program for libraries". The application will become a precursor to a general development trend for this type of software, anticipating and seizing the opportunities in the market in the coming years.

This paper is structured as follows: the second section describes the context in which the proposed project lies, while the section following it describes activities and objectives in detail. In the fourth section, the schedule of the project is outlined, especially concerning the activity of digital content management. The last section hosts our final observations about the project.

2 Context of Research Proposal

Hundreds of thousands of libraries exist throughout the world, of varied sizes and types. The libraries of Italy are estimated to be between 25,000 and 30,000, and the main ones (around 16,000) have a profile in the Italian Libraries Database (Anagrafe delle Biblioteche Italiane, AIB)[3] of the Central Institute for the Union Catalogue of Italian Libraries (Istituto Centrale per il Catalogo Unico, ICCU)[4]. The collection and verification of the data is continuous, also thanks to contributions by Regional governments, Universities, organizations and cultural institutions that work with ICCU.

Currently, those 16,000 public libraries (of which 52% belong to local organizations) provide basic services, also called "public reading", that have gradually grown and become a well-received reference point for thousands of citizens, together with other established public services. Thanks to innovation and automation, many of

---

[1] SoSeBi. srl, http://www.sosebi.it/.

[2] DIEE, University of Cagliari, http://dipartimenti.unica.it/ingegneriaelettricaedelettronica/

[3] AIB, http://anagrafe.iccu.sbn.it

[4] ICCU, http://www.iccu.sbn.it

those libraries now offer their services through the Internet, for example as remote support or online, digitized versions of part of their collections. The most widespread and important of those services is the free browsing of the Online Public Access Catalog (OPAC)[5].

The Italian IT companies in this sector that market products and services, and that also produce and supply certified applications to be used in the library network of the National Library Service (Servizio Bibliotecario Nazionale, SBN)[6], managed by ICCU with agreements with the Regional governments, are only 15 in number.

As regards the investments in the library sector, according to a nationwide analysis, it is noteworthy to mention that investments in general, thus including those pertaining to automation, have dwindled lately due to the economic-financial crisis.

In Italy, the presence of SBN strongly affects figures and features of the market. The adoption of the SBNMARC[7] protocol has, in fact, forced the companies that wanted to keep being competitive to pursue that direction in their development. The tendency of many public institutions (especially at the regional level, subsequently involving local organizations) was to adhere to SBN, since the communication function is becoming a vital aspect for any software aimed to succeed in Italy

The technological evolution of the current software systems suffered from a slower "innovation pace", due to the low margins offered by the market. Currently, the market is in a phase of "almost-maturity" in technology.

The most important aspect addressed recently is the integration of the information from catalogs in libraries with the information directly available on the Internet. The development now is focused on OPAC, and, more generally speaking, on the search tools available now, in comparison with all-purpose search engines used on the Internet (starting, obviously, from Google).

The solutions are many, but they all tend to take into account several needs. On the one hand, the need to expand and widen the information that can be searched (enrichment, federated search, access to databases); on the other hand, the presentation, that borrows typical systems of the Internet (tag cloud, conceptual associations, users' comments, and so on). They follow two tendencies, sometimes coexisting: the creation of portals and collections (usually customizable) of resources that crowd one page, and the proposal of a very lean and simple, Google-like, interface. As it had begun to happen in the early years of the World Wide Web, some international producers became specialized in building tools that complement or overlap with the actual OPAC, and promote those tools as solutions that can be adopted even by users of other systems, as third-party products. Furthermore, the high specialization and specificity of library standards are a significant technology barrier to the entry of new competitors.

3 Description of Research Project

The proposed project is a software devised for the net, entirely web-based, with the

---

[5] OPAC, http://opac.sbn.it

[6] SBN, http://www.sbn.it

[7] SBNMARC, http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/evoluz_indice_sbn/pagina_147.html

6

management of all data outsourced in the SoSeBi server infrastructure. It will be supplied to users as a service (Software as a Service, SaaS), according to the increasingly widespread method among IT services that lets the customer be in the position of using the software without having to deal with what is "behind" it. The customer will access the platform in the same way they access a website, using a simple browsing program, and all the data they will enter and the related elaboration received will be handled by SoSeBi through an advanced company technological infrastructure.

The technology used for the software part of the server will be Microsoft's .NET framework, the ASP.NET language on Windows Server platform. Hardware-wise, a cutting-edge server will be used as the base technology, with an Intel multiprocessor and a large quantity of dedicated RAM, combined with a data redundancy system (discs in RAID 1 configuration and backup on external unit). Using this technology, it will be possible to switch to a more powerful infrastructure at a specialized national supplier. Internet connectivity will be of a 2 Mbit/second guaranteed minimum band, which could scale up to 100 Mbit/second on demand.

Regarding product innovation, the objective is to merge four fundamental aspects in one software. These aspects are expected to have a large development in the future.

1. Management of documents according to library science;
2. Management of digital contents and related access rights;
3. Social network;
4. Semantic search into electronic documents.

Some of those aspects, like the digital content management and semantics, have just become a part of citizens' habits. This can pave the way for new scenarios in terms of behavior and use of those technological tools.

The first distinctive feature will be the homogeneous integration of the four aspects stated above, called macro-functions. The dialogue method among the four macro-functions will be innovative, and allow, for example, to integrate the traditional library accuracy with Web 2.0 trends, namely reviews and tagging of books, and social relations between users and libraries.

In particular, each macro-function will have the following functions.

1. Management of documents according to library science. Careful management of bibliographic information according to library science standards and national industry rules. Certification of the application at Level 4 of the SBN dialogue protocol released by ICCU.

2. Management of digital content and related access rights. Integration with multimedia content management platform, with eBooks, audiobooks, digital audiovisuals, images, digitized content. Their management includes a description of their content, search, and viewing, also in streaming for audiovisuals. Access to contents is regulated at the source by a copyright and Digital Rights Management (DRM) control system, so that the legal and commercial boundaries established by the publisher will be respected.

3. Social Network. Creation of a network of potential social relations among the various actors allowed the use of the software: libraries and readers. The main reason for the social network is the sharing of an interest in books, expressed in different ways by the participants, where libraries showcase their catalogs, users

comment and vote books. A functional ecosystem centered on books, where social relations can be built and expressed, giving added value to the traditional catalog search of a library.

4. Semantic search in electronic documents. Integration of a language analysis engine (morphology, grammar, logic), and semantic disambiguation of texts. Use of a semantic network made primarily of a thesaurus of concepts and relations that would provide a conceptual representation of the language. Automatic categorization of documents, interpreting their content, in certain categories. Extraction of data from indexed texts, with normalization and transformation in metadata.

The activities covered in the project were anticipated by a feasibility study in order to analyze the information needs related to the project, which was generally defined during its planning stage. Among the objectives of the study were the following.

1. Find one or more architecture solutions related to the applications, the technologies and organizational solutions;

2. Propose technical-organizational solutions;

3. Provide the management with enough evaluation tools to decide on the operative completion of the project.

3.1 Project Subdivision

The project covers a number of operation stages. Every stage of the working plan is organized in Work Packages (WP), parallel phases in which operation objects are reached with work group activity, through the production of expected Results and Products and the application of a specific Methodology. The WP included in the projects are six:

1. System Architecture
2. Content Management
3. Social Network
4. Semantic Engine
5. Fundamental Research
6. Experimentation

Below is a brief description of each phase in the development of the project.

3.1.1 System Architecture

The future scenario in which libraries will have to work will be analyzed. That scenario will see an increasing centrality of the Internet and its digital services to meet the needs of both physical users and virtual users, who access the library remotely.

Although several software development models exist, we chose to perform a SWOT analysis limited to the Spiral and Agile [3] models, as major representatives – in software engineering – of iterative methodologies and, especially in the latter case, modern ones. The methodology will therefore be the Agile development model, since it is more in tune with the new demands of the market/customers.

The architecture will have to possess a high flexibility, and will have to be able to

8

incorporate advanced functions, resulted from research, in a modular and incremental way (for example, an evaluation of the adequacy of federated or plug-in based solutions could be useful).

3.1.2 Content Management

An interesting topic to study in this field is certainly the management of the metadata associated with multimedia objects. In this field, a significant role is given to ontologies, which define a formal, shared, and explicit representation of a conceptualization of any knowledge area. They are powerful tools to describe the entities of any domain, and the relations among them. Consistently with this view, it is necessary to guarantee an appropriate management of the semantics of metadata, so that the same entities, properties, and relations of the domain with which the system interacts are denoted.

The guiding purpose of this study lies in the definition of the semantics of metadata associated to multimedia contents by using an ontology. In particular, it is necessary to consider the possibility of defining specific ontologies for the domain of interest in an automatic or semi-automatic way. Alternatively, given an existing ontology, for example manually written by domain experts, the possibility of expanding that ontology with new metadata automatically extracted from a known set of multimedia documents can be considered.

3.1.3 Social Network

The functionality of the Social Network module must be defined taking into account the dialogue between readers and libraries, introducing the well-established tools of library tradition into the Web 2.0: OPAC, loans, reference. It is necessary to improve the dialogue between libraries and their users, providing an innovative service, which will bring them into closer contact with the libraries, thanks to the Web 2.0. The idea is to target Web users that not necessarily are also users of one of the libraries in the network. They would be able to sign up to use the social function and communicate on the Internet, sharing their taste and preferences; they could eventually become users of a library, after having come into contact with its services through the social network.

This macro-function is meant to bring forth a social evolution of the traditional online catalogs, shifting the focus on the customers' views and their expectation in terms of Web 2.0.

3.1.4 Semantic Engine

Semantic search is a field where many studies are focusing, considering the enormous and ever-increasing volume of documents currently available (especially digital documents). Academic research in this field will be fundamental, because semantics is still a young science, with a strong connection to Information Technology, and which takes part to a heated debate with semantic Web. It will thus be necessary to develop a precise outlook to understand what will become of the dialogue between semantics of libraries and semantics of the Web. These two aspects, in fact, are going to become

more and more interconnected in the future, given the increasingly faint line between owned documents and accessible documents on the Web.

3.1.5 Fundamental Research

This phase will encompass all the activities at the DIEE of the University of Cagliari oriented to fundamental research:

1. the definition and study of a metadata ontology for multimedia objects;
2. the definition and study of graph analysis techniques applicable to Social Network;
3. the definition and study of recommendation techniques on Social Network;
4. the analysis and study of techniques and models of semantic analysis;
5. semantic classification from available information coming from thesauri, subject indexes, summaries, and classifications.

The applied research will be based on the methodological results of this research, guaranteeing innovative functions of the product..

3.1.6 Experimentation

The technical activities related to the creation of the software prototype will be verified and validated appropriately, regarding both code and execution. The verification plan will be connected to the definition of software requirements according to the Agile Methodologies. The draft of the validation plan will aim to allow the verification of product requirements along the entire design phase. Once the whole system is broken down into different modules, three test phases will be performed:

1. module test, to test the components of each module and verify the communication of components belonging to the same module;
2. integration test, to test the integration and interaction mechanisms between each module;
3. system test, to test the whole system and come to an acceptance test of the final prototype.

4 Project Schedule

The project officially began on March 6, 2013, and its conclusion is estimated to be on March 5, 2016. Currently, the prototype covered in the project is in its implementation phase.

Among the activities performed up to now, special attention was paid to the the management of metadata associated to multimedia objects, leading to the definition of a taxonomy of metadata for the representation of multimedia objects managed by a digital library.

The approach we followed included the application of 3 distinct phases. The first phase consisted of the analysis of metadata standards used as a reference to represent information associated to multimedia content. In order to cover the reference domain,

10

made of multimedia objects of interest in a library sector, metadata standards that allow the best and most complete description are used [4][5][6]: the choice fell on the Dublin Core [7][8], XMP [9] e MAG [10] standards. The first provides a general description that can refer to content of any kind (a feature that helped its large scale usage); XMP focuses on more accurate and peculiar characteristics to multimedia content, such as audio and video content, other than images, for which it uses Exif metadata (specialized in digital images) [11]; MAG, instead, is an application profile that, included in our taxonomy, allows to communicate with the Internet Culturale portal without interoperability problems. The portal, in fact, uses MAG for the representation of its contents. The data sheets of the Italian Central Institute for Cataloguing and Documentation (Istituto Centrale per il Catalogo e la Documentazione, ICCD[8]) offer the opportunity to perform a mapping between the metadata in our taxonomy and the PICO application profile used by the portal Cultura Italia, thus increasing the interoperability of the system [12].

We also chose to consider some tags for the management of User-Generated Content (UGC) [13]. Those standard are often complementary, covering information that they could not represent singularly, and sometimes redundant: in that case, a direct mapping was performed.

As regards the second phase, a study on the resources to represent was carried out, using the following sites as reference: Europeana[9], Internet Culturale[10], Cultura Italia[11], Internet Archive[12], Open Library[13], and Project Gutenberg[14].

The third phase involved comparing metadata taken from the standards analyzed during the first phase with the data collected during the second phase. The purpose of the comparison was to verify whether all the characteristics studied during the second phase were represented by the metadata retrieved during the first phase.

Moreover, we studied how to enrich the available metadata in order to manage some often neglected aspects, through a new study on the objects available at the Mediateca del Mediterraneo (MEM)[15] of Cagliari, and through the study of the copyleft licenses Creative Commons[16] and Copyzero X[17]. Topics such as the completeness of information on eBooks, the identification of documents as belonging to gray literature, and the management of rights on the cataloged resources were addressed.

Thanks to the adoption of the MAG application profile, the taxonomy guarantees the interoperability with the portal Internet Culturale, that grants access to the wealth of public libraries and prestigious Italian cultural institutions. The presence of the DC standard allows also to be Open Archive Initiative (OAI)-compliant, with the opportunity to exploit the interoperability protocol Open Archive Initiative Protocol for

---

[8] ICCD, http://www.iccd.beniculturali.it

[9] Europeana, http://labs.europeana.eu

[10] Internet Culturale, http://www.internetculturale.it

[11] Cultura Italia, http://www.culturaitalia.it

[12] Internet Archive, https://archive.org

[13] Open Library, https://openlibrary.org

[14] Project Gutenberg, http://www.gutenberg.org

[15] MEM, http://www.comune.cagliari.it/portale/it/studisardi.wp

[16] Creative Commons, http://creativecommons.org/

[17] CopyZero X, http://www.costozero.org/wai/copy0.html

Metadata Harvesting (OAI-PMH) [14], and ensuring the communication with another important portal, Cultura Italia.
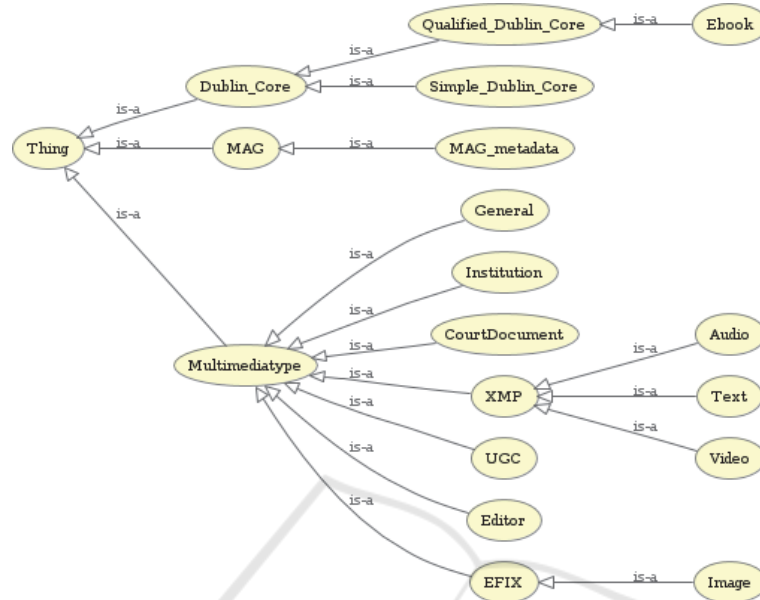

Fig. 1. The resulting metadata taxonomy.

The information network with which the system adopting the taxonomy will be able to interface, and the set of metadata studied to cover the general, main information and the specific information of multimedia content, will make its use suited to a digital library aspiring to modernity. Such a library would become able to manage often neglected pieces of information (gray literature, UGC, management and representation of rights on the resource), and could become part of an information and cooperation network with the most important cultural portals of Italy.

5 Conclusion and Prospects

This project is coherent with the strategic objective of the regional planning in Sardinia, since it aims to implement innovative methods of the ICT sector in the library industry, and it complies with the objectives described in the Regional Strategic Document (Documento Strategico Regionale, DSR) 2007-2013[18] for Sardinia, which states the necessity of promoting the adoption of ICT in order to fill the gap that makes Sardinia less competitive in innovation, although the region stood out in Italy for having started innovative initiatives ahead of its time, marking a sharp passage from a traditional economy to an innovative economy, based on knowledge and inno-

---

[18] Regione Autonoma della Sardegna, Documento Strategico Regionale (DSR) 2007-2013, http://www.regione.sardegna.it/j/v/17?&s=1&v=9&c=4756&na=1&n=10

12

vation. The project is meant to become a part, albeit small, of the context of the competitiveness of Sardinian companies, with its purpose to create a software for the automation of cultural services.

The Italian production system is characterized by a high volume of very small-sized companies in every sector, which work outside cooperation agreements or networks. This structural characteristic does not allow these companies to compete on the markets of goods where price, which comes from production costs, is the main factor to acquire and maintain market shares.

The limited availability of funding will strongly affect any new investments made by libraries in goods and services, and will force them to question the worth of maintaining current services, including computer-based services and the automation of bibliographic catalogs. The market prospects regarding this particular aspect, that is the reduction of management costs of technical services, could be favorable for those companies that are committed to creating and supplying more effective software, more competitive in terms of price, and technical support services that would help institutions to save on service cost keeping the same or a higher quality level.

As regards the demand, the product to be created is clearly geared towards a target of customers that have already completed the first steps in service automation, thus it has to meet the needs of a well informed user base with high expectations in terms of performance of management systems.

A new product with the features described in the previous sections has great market prospects, especially in Northern Italy, where the demand for innovative products is stronger and where more libraries, more resources, and more well-established cooperation networks among organizations are located, to provide advanced services to the citizen.

References

1. Candela, L., Castelli, D., Pagano, P.: History, Evolution, and Impact of Digital Libraries. E-Publishing and Digital Libraries: Legal and Organizational Issues, E-Publishing and Digital Libraries: Legal and Organizational Issues, IGI Global (2011) 1-30.
2. Fox, E. A., Gonçalves, M. A., Shen, R.: Theoretical foundations for digital libraries: The 5S (societies, scenarios, spaces, structures, streams) approach. Synthesis Lectures on Information Concepts, Retrieval, and Services, 4(2) (2012) 1-180.
3. Poppendieck, M.P.: Lean Software Development: An Agile Toolkit. Addison-Wesley Professional (2003).
4. Lagoze, C., Van de Sompel, H.: The Making of the Open Archives Initiative Protocol for Metadata Harvesting. Library Hi Tech Volume 21 (2) (2003) 118-128.
5. Scherp, A., Eißing, D., Saathoff, C.: A Method for Integrating Multimedia Metadata Standards and Metadata Formats with the Multimedia Metadata Ontology. International Journal of Semantic Computing, 6(1) (2012) 25-49.
6. Zengenene, D.: Global interoperability and linked data in libraries. New Library World,

114(1/2) (2013) 84-87.

7. Hillmann, D. I.: Using Dublin Core. Dublin Core Metadata Initiative Recommendation (2005). Retrieved from: http://dublincore.org/documents/usageguide/

8. Hutt, A., Riley, J.: Semantics and Syntax of Dublin Core Usage in Open Archives Initiative Data. Joint Conference on Digital Libraries (2005).

9. Adobe Systems Incorporated: Adobe XMP Specifications, Additional Properties (2010). Retrieved November 13, 2014, from http://www.adobe.com/content/dam/Adobe/en/devnet/XMP/pdfs/XMPSpecificationPart2.pdf

10. MAG, Comitato: Metadati Amministrativi e Gestionali: Manuale Utente, Pierazzo, E. (ed.), version 2.0.1. Roma: ICCU (2006).

11. Technical Standardization Committee on AV IT Storage Systems and Equipment:. Exchangeable image file format for digital still cameras: Exif version 2.2. Published by Standard of Japan Electronics and Information Technology Industries Association (2002). Retrieved from: http://www.exif.org/Exif2-2.pdf

12. Buonazia, I., Masci, M. E., Merlitti, D.: The Project of the Italian Culture Portal and its Development. A Case Study: Designing a Dublin Core Application Profile for Interoperability and Open Distribution of Cultural Contents. In Proceedings ELPUB 2007 Conference on Electronic Publishing (2007) 393- 404. Retrieved from: http://elpub.scix.net/data/works/att/114_elpub2007.content.pdf

13. Pani F. E., Concas G., Porru S.: An Approach to Multimedia Content Management. In: Proceedings of The 6th International Conference on Knowledge Engineering and Ontology Development, KEOD (2014) 264-271.

14. Lagoze, C., Van de Sompel, H.: The Making of the Open Archives Initiative Protocol for Metadata Harvesting. In Library Hi Tech, 21(2) (2003) 118-128.