BEAM Repository: A Proposal for Family and Personal Repository

Rachel Cristina Vesu Alves São Paulo State University, Brazil rachel@marilia.unesp.br

Paula Regina Ventura Amorim Gonçalez São Paulo State University, Brazil paulagoncalez@marilia.unesp.br Ana Carolina Simionato Federal University of São Carlos, Brazil acsimionato@ufscar.br

Ana Paula Grisoto São Paulo State University, Brazil grisotoana@reitoria.unesp.br Felipe Augusto Arakaki São Paulo State University, Brazil fe.arakaki@marilia.unesp.br

Plácida Leopoldina
Ventura Amorim da Costa
Santos
São Paulo State
University, Brazil
placida@marilia.unesp.br

Abstract

Preservation of cultural heritage has been widely discussed in the last decades. Different groups of people contribute to the production and preservation of cultural heritage through personal and family performance. However, there is a lack of environments specifically prepared to store and organize the resources produced by these groups, resulting in difficulties to access and preserve these materials along the time. The hypothesis is that the digital repository and the structured metadata standards are relevant tools to provide the suitable environment to store, describe, access and preserve family and personal resources. The study herein has a theoretical and applied basis, for it aims to investigate and confirm the hypothesis using theories and applying them. It aims at demonstrating that the digital repositories are relevant for the storage, description, access and preservation of personal and family information. During implementation of the digital repository, DSpace software and Dublin Core standard were used. As a result, the implemented repository showed itself as a viable alternative for storing this information. It is possible to conclude that such a digital repository constitutes a tool that guarantees the preservation, access and sharing of archives, resources and data produced by families and individuals in the digital environment.

Keywords: BEAM Repository; DSpace; Dublin Core; Family and personal repository.

1. Introduction

Preservation of cultural heritage has been widely discussed in the last decades. Many institutions are providing their preserved cultural patrimony through their digital collection. In this scenario, the big challenge is to provide the suitable representation of digital information resources, guaranteeing the integration of different communities and the interoperability of data. Thus, the use of metadata and metadata standards has become a common practice among the several areas that seek to preserve and provide cultural heritage in digital collection.

Cultural patrimony consists of several categories like the tangible cultural patrimony (paintings, sculptures, manuscripts, monuments, cities, shipwrecks, ruins, etc.), immaterial cultural patrimony (oral traditions, arts, music, etc.) and the natural patrimony (natural reservation, archeological or geological sites, etc.) (UNESCO, 2009). Thus, different groups of people, especially individuals and families, contribute to the production and preservation of cultural heritage, building and perpetuating the tangible cultural patrimony, immaterial cultural patrimony and the natural patrimony.

Ordinary people and their family can find on the Internet environments such as Facebook, Flickr, Blogs, Instagram, which provide access to some of their personal information. However, people provide a diversity of family and personal content such as pictures, documents, videos, intellectual and artistic productions, material related to trips among others, and they do not have a specific environment to store and organize these resources options, resulting in difficulties to access and preserve these materials along the time.

The personal and family contents, denominated herein as personal and family information resources, in many cases held by one or more members of the family, are sometimes discarded or do not receive informational treatment that guarantees the access and makes them easy to be located. Thus, it is important to provide storage, adequate description of resources, preservation and, at the same time, extend the access and the sharing of personal or family production, reducing physical and temporal spaces among people from the same family core. Such actions contribute to preserve cultural heritage produced inside a family environment or by the personal interaction with the existing cultural heritage.

In the last years, a growing number of digital repositories were developed by different types of organizations such as digital libraries, universities, public archives, and research centers among others. The Open Archives Initiative and the creation of open source softwares's made it easier to provide digital contents of these organizations. However, it is possible to see that there are not many initiatives of digital repository implementation with familiar and personal purpose.

Digital repositories are considered environments that provide storage, description, organization and preservation of information resources, guaranteeing that their access and the family or personal history is not lost along the time.

The hypothesis for this study is that the digital repositories and the structured metadata standard are relevant tools to provide an environment able to store, preserve and provide the access to family and personal information resources in a more organized way.

This study has an applied basis, for it aims to investigate and confirm the hypothesis established and solve practical and immediate application problems. It is also considered a qualitative and exploratory study because it seeks information in order to clarify the subject investigated taking into account several aspects (Cervo & Bervian, 2003; Gil, 2002).

The aim of this study is to demonstrate, by means of an implementation, that the digital repositories constitute themselves as relevant environment to store, describe, access and preserve the personal and family information resources. It is possible to conclude that the implementation of such a digital repository constitutes a tool to guarantee the preservation, access and sharing of collection of resources archives and data produced by family cores and individuals in the digital environment.

2. Digital repository and metadata

According to Pollak (1992, p. 204), memory contributes to build individual and collective identity. The family and personal information resources represent the bond between individuals and their lives, performing an important role in registering and perpetuating memory, for they bring memories of previous experiences, places they visited, their ancestry and their life history.

With the advance of technologies, several kinds of family and personal information resources have started to be produced, requesting organization to be available in digital environment.

In doing so, the implementation of digital repositories is presented as a relevant initiative to store, share and access resources. Repositories can be described as

systems available in the web that provide, mainly, facilities to add and access digital objects . . . repositories aggregate a great variety of facilities, most of them related to management of digital objects added . . . besides managing digital documents, they have facilities related to their preservation and they are flexible systems that can adequate themselves to fit several purposes (Shintaku & Meireles, 2010, p. 17).

There are different repositories such as academic, administrative, technical and hybrid. According to a more general classification, there are also institutional and thematic repositories. Regarding the thematic repositories, the object of this study, it is possible to highlight that the origin of the information resources provided is diverse and that the theme is the main point of aggregation of these resources, making their access easier (Shintaku & Meireles, 2010). In personal repositories from Brazilian initiatives, information resources are organized by grouping the kinds of resources.

In this study, the purpose is to keep and manage resources for a long period and provide the appropriate sharing and the access to those interested in a thematic structure. The software chosen for implementing the repository was DSpace¹, because according to the Registry of Open Access Repository (ROAR)² it is the most used software for implementing open source digital repositories. This software meets most requirements listed in open-source³ software analysis in the literature and it is widely used.

DSpace is open-source software to store, manage and distribute collections in digital format. The software meets the necessary requirements for the implementation of the repository, which is the purpose of this study: open source software. It does not have costs to be acquired and has simple and intuitive web interface; stores different kinds of information resources. It allows the inclusion of more than one format of archive per work described and the creation of distinct collections. It manages collections of items, communities and sub communities with more than one collection; establishes relations among resources, collections and their preservation; stores, imports and exports resources and their metadata, according to Dublin Core standard or other standard if necessary.

It supports many idioms in the metadata and digital content field; uses unique identifiers (Handle System); provides digital preservation compatible with Open Archive Information System model (OAIS); shares metadata through protocol Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), exports archives in Metadata Encoding and Transmission Standard eXtensible Markup Language (METS XML) and also works with Open Uniform Resource Locator (URL) protocol; it presents levels of customization in both user interface and its structure; offers access control to communities, sub communities, collections and resources by means of determining activities. Interaction with user is performed by email and with available information in the repository; the interface with the user allows browsing among communities, sub communities and collections. Browsing and searching can be performed by creator, title, subject, date and key-words found in metadata; the submission can be performed by the creator of resources; it provides a satisfactory documentation for its implementation (Pirounakis & Nikolaidou, 2009; Sayão & Marcondes, 2009; Romani, Fusco & Santos, 2010).

DSpace suits better in cases when it is necessary to establish communities, sub communities and collections, manage information resources and submit these resources. These are determinant characteristics for the repository at issue, which is in process of implementation inside the Library of Study and Application of Metadata (Biblioteca de Estudos e Aplicação de Metadados - BEAM). Another fundamental characteristic was the possibility of using Dublin Core standard for the description of information resources, for this is a metadata standard, which can be used by experts or those who are not experts. The aim is that any individual, expert or not, can organize their family and personal information resources in a digital environment. The idea is using a system in which the following functionalities are observed: a) easy collection and insertion of information resources, including their metadata; b) easy access to information resources either by list of communities, sub communities, collections and items or also by the search interface; c)

¹ Retrieved March 25, 2015, from http://dspace.org/.

² Retrieved March 25, 2015, from http://roar.eprints.org/view/software/.

³ "software package whose distribution follows its source code, allowing the user to modify and adequate the software according to his necessities" (Toutain, 2006, p. 20).

promotion of long-term preservation of information resources stored in this system (Lewis & Yates, 2008).

The metadata standard adopted to implement and implant the repository herein is the Dublin Core developed from characteristics such as simplicity; semantic interoperability; international agreement; extensibility and modularity of metadata in the Web. Such characteristics are understood as follow: the proposal of simplicity makes it possible for those who are not experts to describe a resource in the Web; the semantic interoperability comprehending several metadata standards enabling the interoperability; the agreement in accepting internationality the Dublin Core Metadata Element Set; the extensibility and flexibility of Dublin Core when widening and adding descriptive elements, which are presented as elective and repetitive; the modularity of metadata in the Web, metadata can be combined with other schemes, even if they contemplate different semantic and syntactic structures (Alves & Santos, 2013).

3. Development and Implementation of BEAM Repository

BEAM Repository⁴ is a project created by the BEAM, of the Group of Research in New Technologies in Information (Grupo de Pesquisa Novas Tecnologias em Informação – GPNTI), in Philosophy and Science University at São Paulo State University (UNESP), Marília/SP. BEAM aims at providing the students linked to the library of studies and application an environment for the development of researches related to the creation and manipulation of digital objects metadata. In doing so, the goals in constructing the repository are: a) to provide an environment for studies and practical applications over metadata and metadata standards; b) to manage digital collections; c) to make possible the study and practices relating to the interoperability, harvesting, digital ownership among others.

The first initiative using the implementation of BEAM repository was the creation of family and personal archives to organize travel material, in order to make the access to information resources easier, preserving the history and memory of the individuals in the family.

The implementation of BEAM Repository was performed based on the following phases:

- Phase 1 Planning and defining the repository scope: it started after identifying the necessity of organizing, in a digital way, travelling materials produced or acquired by families and people. Later, actions to solve the problem were taken, contributing to the establishment of the repository scope to be implemented and to the adoption of Dublin Core metadata standard. These actions were distributed in a more detailed way in the following phases. The software was also chosen according to what was mentioned in item 2 herein.
- Phase 2 Implementation and personalization of software: this phase embraced the installation of DSpace software in the research group server, the personalization and configuration of its visual interface (model, layout, sources, colors, BEAM logo insertion etc.).
- Phase 3 Definition of metadata in the repository: the simple scheme of Dublin Core metadata standard was chosen in the planning phase so that the template used to describe the information resources at the moment of their insertion in the repository could be built. This phase is related to the previous phase, because it is also related to the personalization of DSpace software.
- Phase 4 Definition of communities and sub communities: in order to insert information resources in the repository, it was necessary to establish first the communities and sub communities that would group the collection of information resources. The collection was classified according to the primary and secondary needs. The primary needs were defined based on the family composition, that is, the family is composed of two or more people related by birth, marriage, adoption, civil union or some other similar legal way that groups a

⁴ http://beam.marilia.unesp.br

family (International Federation of Library Associations, 2009). The secondary necessities are related to people belonging to a family. Therefore, person is defined as an individual or identity established individually or in group (International Federation of Library Associations, 2009). This way, the superior hierarchic position corresponds to the necessities of family grouping and the subordinated positions correspond to people belonging to the family. Thus, the communities will be families and sub communities, and the people related to these families, with the possibility of having different hierarchic structures.

- Phase 5 Definition of collections: the definition of the collections in communities or sub communities was established from the idea of events that represent an action or occurrence with these people and families (International Federation of Library Associations, 2009). This way, families and people can organize their digital resources in a diversity of collections that correspond to several kinds of events or occurrences (travels, birthdays, weddings, vacation etc.). The collections can be established individually or they can be related to other people or even to the family.
- Phase 6 Definition of search system: the DSpace software provides three kinds of search to discover and recover resources: surfing the communities, sub communities and collections; the simple and advanced search, which can be refined by title, author, subject, date of publication (of the information resource in the repository); and date.
- Phase 7 Definition of a Use guide: the definition of the user guide was established considering the lay user who can access the metadata scheme for inserting information resources in his collection. It is a case of defining metadata, guidelines for entering the value and defining which metadata is considered obligatory for the scope of this repository. Although all Dublin Core standard metadata is optional, it was necessary to establish some obligatory metadata so that the repository and its search system could work minimally. The information in the user guide was inserted in the repository description template so that the users could access the information at the moment of the description, without the necessity of accessing other resources.
- Phase 8 Insertion and description of resources in the system: the information resources of the fictitious family were inserted in communities, sub communities and collections, based on the template description. Some resources were already in digital form and the printed resources were previously converted to digital form to be inserted.

The BEAM repository has already gone through all the phases of its development and implementation and currently it is in a phase of evaluation and performance of some adjustments considered necessary to a good operation of the repository. One of these adjustments is related to the advanced search system settings, which are being improved in order to include the type filter (kind of resource) and correct the date filter (dc. date). It would make it even easier for people from the same family to locate and access the resources.

After performing all the phases, some considerations about the practical application of the repository are presented as follow. First, in relation to the BEAM repository, it is possible to observe that it corresponding to the DSpace hierarchic structure through which the information resources are released: Communities, Sub communities, Collections and finally the Items. The Communities are composed from the union of two people, forming a family that conducts the relationship of the other members.

This family can have joint events called Community Collection, which have joint information resources. Sub communities are members of this family and incorporate resources and individual occurrence called Events (Collection).

In doing so, figure 1 hierarchically shows the family, the person and the event, composing respectively the communities, sub communities and collections structured in BEAM Repository.

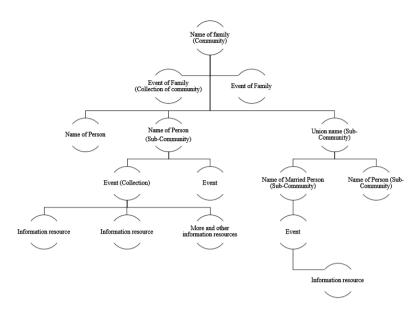


Fig. 1. Hierarchy built for family and personal archives

From this hierarchy, a fictitious family community was created, the Silva's family. The patriarch José Silva is a Portuguese immigrant, from Porto city and arrived in Brazil in 1950. He met Maria Souza in Santos, São Paulo, they got married in 1959 and had three children: José Silva Filho, Paula Silva and Carolina Silva. One of the daughters, Paula, got married to Luiz Oliveira, creating a new family related to the first one. The names of the family members are communities and sub communities and the collections can be private or with other members of the family.

Figure 2 shows the hierarchy built. The fictitious names appear in alphabetic order on the repository page.



Fig. 2. Silva's family

The information resources which can be inserted in the family and personal repository, were defined as follow: images, maps, slides presentations, music recordings, sound recordings, travel brochures and leaflets, touristic guides, museum and subway tickets, air tickets, videos, travel cards, stamps and postcards, personal documents such as certificates, letters among others. Other types of information resources can be inserted. The repository holder defines the criteria from the considerations that take into account the level of importance of resources for preservation, memory, guard and access with the community to be attended.

Resources description is performed with simple Dublin Core, with its 15 description elements, because it is a suitable standard for users who are not experts in generate metadata.

Considering that the user of the system is not an expert, there was the necessity of adapting the representation of information (metadata values). To do so, metadata guidelines and definitions were proposed and obligatory and optional metadata was indicated for this repository, as follow:

- Title: (obligatory element) main title of the resource or title in which it can be known. Insert preferably the title in the resource. In case you do not find it, attribute a title;
- Creator: (optional element) insert the resource creator. Assign the responsibility for creating the resource to the person or group that has more intellectual or artistic responsibility. Some possibilities are the author, editor, photographer, producer among others;
- Subject: (obligatory element) introduce the subject or keyword, which represents the resource. It is recommended to insert at least three subjects in order to make the recovery easier;
- Description: (optional element) insert any suitable description or comments to represent characteristics in relation to the resource: characteristics of the place visited, characteristics about the document stored or personal comments. It is recommended to insert information about time and day of visit or the contact information about the place visited; permission for registering images or videos in the place; permission to use flash; information about people related to the information resource or any other information considered relevant to the resource described herein;
- Publisher: (optional element) insert the name of who published the resource described. The publisher is the person or group of people responsible for publishing and distributing the material:
- Contributor: (optional element) insert the name of those who contribute with the resource. It can be a person or group of people that contributed intellectually or artistically for the creation of information resource. However, this person is not the main responsible for the resource;
- Date: (optional element) insert a date or period of time referring to the resource. It can be a
 date when the resource was acquired, the date when the image was obtained, date when the
 photo was taken, date when a video was recorded, etc. It is recommended that the dates are
 standardized according to what was established by W3C based on ISO 8601 (for example
 YYYY-MM-DD). In case there is no accuracy, use approximate month and year or just the
 year.
- Type: (optional element) insert the type (kind or nature) of the resource. Select the type of resource according to the values in the list (to select more than one value, keep the "Ctrl" or "Shift" keys pressed);
- Format: (optional element) insert if the format is physical or digital. Insert the file format, physical environment or resource dimensions;
- Identifier: (optional element) insert an identifier or a unique reference for the resource (item) in a context;

- Source: (optional element) insert the source or origin in which the resource was derivated. It is a case of indicating the original source that derivated the resource, or the relationship between the parts of a resource. It is recommended to insert the name of the person who has the original resource;
- Language: (optional element) select in the list the main idiom of the resource. In case it does not appear in the list, select "Other" and if the resource does not present an idiom, like in the case of the images and photographs, select "N/A";
- Relation: (optional element) insert the resource relationships. The relationship indicates if a resource is a physical or a logical part of another resource, if it is a version of another resource, if it presents a transformation, if it is a reproduction, etc.,
- Coverage: (optional element) insert a coverage: spatial location, temporal period or a jurisdiction referring to the resource. For the spatial location, insert the name of the city and country; for the time coverage, insert a period of time or dates interval; for the jurisdiction, insert the name of the jurisdiction;
- Rights: (optional element) insert information concerning the resource rights. It includes the
 declaration of rights about the access and availability of resource, the indication of
 intellectual property rights, copyrights, etc.

BEAM repository has defined the control of access to the communities, sub communities, collections and resources, by means of determining permissions. Communities, sub communities, distinct collections and management of collections and items already implanted.

Silva's Family community, which is the example herein, is already available to be used in 45 resources Web, in different ways: travel brochures, photographs, videos, maps, tickets, visit guide, among others. In BEAM repository, the definition of relations between resources and collections is in process of validation.

The proposal is that in the case of the family repository, the interaction is performed through information available in the repository, with an interface that allows the browsing between communities, sub communities and collections.

The main concern in developing the repository is to guarantee facility in collecting, inserting information resources and determining metadata values; and in accessing information resources, either through list of communities, sub communities and items or through search interface. The idea is offering, in digital format, the leisure possibilities provided by family albums in gathering resources that tell a little about the characteristics of a family history or of a person in a pleasant and easy access.

Dublin Core standard contributes to this purpose because it is easy to understand and it is constituted of few metadata, generating a disposition to feed descriptive values. In doing so, the description of "São Paulo arts map" resource (Figure 3), the type of map, belonging to José Silva Filho sub community, is presented as a way to illustrate what was developed.

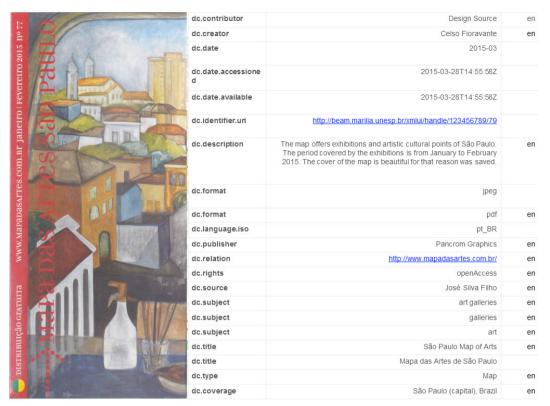


FIG. 3. Register of São Paulo Arts Map

It is interesting to observe that a proposal of family and personal repository requires that the description of events and resources go beyond the descriptions defined formally by schemes, codes and standards of the area. It is necessary, in such work, an informal action in describing the resource stored, a description that refers to the proposal of preserving the family history in which there are notes of recommendation, expression of feelings about moments experienced in life, description about a special family moment. The way to show this was a concern in constructing the repository.

The alternative was using the dc.description option, because it satisfactorily receives values that are not controlled in the description of a resource whose purpose is, somehow, to preserve an affective and emotional bond in the preservation of personal history.

5. Final Considerations

Information resources are gathered during travels and individual or collective experiences and many times they do not have an easy access and organization. In order that these resources are not lost in space and time, it is important to create an environment that allows the access, recovery, sharing, use and reuse of these resources and that also allows the preservation of a community memory.

The aim of this study was to present an environment that provides an organization of family and personal digital information resource.

The repository implementation using DSpace had good results, although its installation and personalization demands specific knowledge. After implemented, DSpace software efficiently meets the structuring and representation requirements of the family and personal digital information resource.

In relation to the storage and description of information resources, it was possible to see that simple Dublin Core standard sufficiently meets the expectations and necessities of representing the information resources inserted. As it is a metadata standard dedicated to experts and to those

who are not experts, the description of resources can be made by a professional or by the members of the community attended.

The proposal is considered innovative not only from the point of view of the librarian, but also the kind of public attended. A family and personal repository is believed to be a viable alternative for storing information resources, guaranteeing memory preservation and family history construction, besides organizing the representation and preservation of information resources and family and personal data.

In BEAM, the development of manuals and tutorials for those who are not experts is in development as a continuation of this study, besides the continuation of studies referring to personalization of personal and family repository.

Acknowledgements

The authors would like to thank Jorge Janaite Neto and Luiz Felipe Galeffi contributions during the implementation and implantation of BEAM Repository.

References

- Alves, Rachel Cristina Vesu, and Plácida Leopoldina Ventura Amorim da Costa Santos. (2013). Metadados no domínio bibliográfico [Metadata in the Library Domain]. Rio de Janeiro: Intertexto.
- Cervo, Amado L., and Pedro A. Bervian. (2003). Metodologia científica [Scientific Methodology]. São Paulo: Prentice Hall
- Gil, Antônio Carlos. (2002). Como elaborar projetos de pesquisa [How to develop research projects]. São Paulo: Atlas.
- International Federation of Library Associations. (2009). Functional Requirements for Authority Data: a conceptual model. Washington: IFLA. Retrieved January 30, 2015, from http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.
- Lewis, Stuart, and Chris Yates. (2008). The DSpace Course Introduction to DSpace.Retrieved March 22, 2015, from http://cadair.aber.ac.uk/dspace/bitstream/handle/2160/617/Module%20-%20An%20introduction%20to%20DSpace.pdf?sequence=8&isAllowed=y.
- Pirounakis, George, and Mara Nikolaidou. (2009). Comparing Open Source Digital Library Software. Retrieved March 22, 2015, from http://galaxy.hua.gr/~mara/publications/ideaDL09a.pdf.
- Pollak, Michael (1992). Memória e identidade social [Memory and social identity]. Estudos Históricos, 5(10), 200-212.
- Romani, Lucas Salviano, Elvis Fusco, and Plácida Leopoldina Ventura Amorim da Costa. Santos. (2010). Análise e implantação de Repositório Digital utilizando Software Livre DSPACE [Digital Repository analysis and deployment using Free Software DSPACE]. Proceedings of the Simpósio Brasileiro de Sistemas de Informação SBSI, 2010, 01-16. Retrieved March 25, 2015, from http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2010/0019.pdf.
- Sayão, Luis Fernando, and Carlos Henrique Marcondes. (2009). Software livres para repositórios institucionais: alguns subsídios para a seleção [Free software for institutional repositories: some subsidies for selection]. In: Sayão, Luis Fernando, Lídia Brandão Toutain, Flavia Garcia Rosa, Carlos Henrique Marcondes. (Eds.). Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação (pp. 9-22). Salvador, Brasil: EDUFBA. Retrieved March 25, 2015, from https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf.
- Shintaku, Milton, and Rodrigo Meirelles. (2010). Manual do DSPACE: administração de repositórios [DSPACE guide: repositories of administration]. Salvador: EDUFBA, 2010. Retrieved March 30, 2015, from http://www.repositorio.ufba.br/ri/handle/ri/769.
- Toutain, Lídia Maria Batista Brandão. (2006). Biblioteca digital: definição de termos [Digital Library: Definition of Terms]. In: Marcondes, Carlos Henrique, Hélio Kuramoto, Lídia Brandão Toutain, and Carlos Marcondes. (Eds.). Bibliotecas digitais: saberes e práticas (pp. 15-24). Salvador, Brasil: EDUFBA.
- UNESCO. (2009) What is meant by "cultural heritage"? Retrieved April 1, 2015, from http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-cultural-property/unesco-database-of-national-cultural-heritage-laws/frequently-asked-questions/definition-of-the-cultural-heritage/.