HAL open science

ARQUIGRAFIA: digital images in a Web collaborative environment

Artur Simões Rozestraten, Vânia Mara Alves Lima, Cibele de Araújo Camargo Marques dos Santos

To cite this version:

Artur Simões Rozestraten, Vânia Mara Alves Lima, Cibele de Araújo Camargo Marques dos Santos. ARQUIGRAFIA: digital images in a Web collaborative environment. DH. Opportunities and Risks. Connecting Libraries and Research, Aug 2017, Berlin, Germany. hal-01665830

HAL Id: hal-01665830

https://inria.hal.science/hal-01665830v1

Submitted on 17 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

ARQUIGRAFIA: digital images in a Web collaborative environment

Artur Simões Rozestraten[1]
artur.rozestraten@usp.br

Vânia Mara Alves Lima[2]
vamal@usp.br

Cibele de Araújo Camargo Marques dos Santos[3]
cibeleac@usp.br

Abstract
Over the last seven years a multidisciplinary research team of the University of São Paulo (USP) with the support of São Paulo Research Foundation - FAPESP (2012/24409-2) has been dedicated to designing a collaborative environment of architectural images on the Web. This collaborative environment denominated ARQUIGRAFIA allows institutional users – such as GLAMs (Galleries, Libraries, Archives and Museums), NGOs, Universities and Research Centers – to collaborate with private users – such as students, teachers, photographers and amateurs – in order to build an open, public and free web digital iconographic collaborative environment on Brazilian architectures and urban spaces. The ARQUIGRAFIA research team has been analyzing the digital repository and the website databases; undertaking analytical-critical studies of the indexing languages for Brazilian architecture images as well as of the tags proposed by the users; raising problems for the metadata accordance between systems; as well as consolidating standards which respond both to international interoperability requirements and to local needs for organization and information access. On the website or on the Android App, users can create an account and share their visual collections (photographs, drawings and videos) by uploading files, as well as indexing, geo referencing and licensing images (Creative Commons). Meanwhile, all users can organize albums, post comments and download high-resolution images with metadata. Institutional users can also ask for an institutional login that allows them to have a collective identity online with a particular name and avatar. Those users can choose, whenever they log in, if they will log as a private user or part of an institutional team. Nowadays, the main images of ARQUIGRAFIA come from the collection of the library of School of Architecture and Urbanism of the University of Sao Paulo (FAUUSP). On a technological perspective, ARQUIGRAFIA plays the role of a pilot program for a template called +GRAFIA that can offer free help for other areas of knowledge to build their own visual collaborative environments, such as a BOTANYGRAFIA dedicated to the *flora*, or an ARTGRAFIA, dedicated to visual arts. Conceptual and technological challenges concerning the design and the operation of ARQUIGRAFIA allows us to characterize it as an online experimental laboratory and a case study on the opportunities and the risks of digital projects on humanities based on

[1] Professor of School of Architecture and Urbanism of the University of Sao Paulo. Rua do Lago, 876 – Cidade Universitária, São Paulo, SP, 05508-020 - Brasil.
[2] Professor of School of Communications and Arts of the University of Sao Paulo. Av. Prof. Lúcio Martins Rodrigues, 443 –Cidade Universitária, São Paulo - SP, 05508-020 – Brasil.
[3] Professor of School of Communications and Arts of the University of Sao Paulo. Av. Prof. Lúcio Martins Rodrigues, 443 –Cidade Universitária, São Paulo - SP, 05508-020 – Brasil.

images constellations.

Keywords: Digital heritage, Architectural images, Image databases, Collaborative environment, Web apps.

Introduction

ARQUIGRAFIA (http://www.arquigrafia.org.br) is an open, public and nonprofit, continuous growth collaborative environment on the Web dedicated to visual culture and critical approaches to architectural images, with a special focus on Brazilian architecture and urban spaces. The main goal of this project is to gather institutions and private users in the same internet environment, dedicated to the collaboration for the research, study and collective knowledge building on Architecture and Urbanism in Brazil and along with the Portuguese speaking communities around the world.

So far, ARQUIGRAFIA's main source of images has been the collection of the library of the School of Architecture and Urbanism of the University of Sao Paulo (FAUUSP), but the collaborative nature of the project distinguishes it from institutional image databases on the internet, precisely because it involves a heterogeneous network of collaborators, both institutional and private, who jointly contribute to the construction and continuous growth of the digital images collection.

Such nature presents several desirable characteristics for the constitution of what Tim O'Reilly (2007) calls Web 2.0, based on parameters developed by Christopher Alexander (1977). Among these parameters, ARQUIGRAFIA emphasizes the experimental meaning of an active and collaborative participation of all users in the continuous construction of the system as an "image constellation" and also as a perpetual beta software, constituting a Web community that constructs knowledge collectively, respecting copyright, and using distinct and complementary interfaces such as a website and an Android application for smartphones.

On the Web 2.0 context, as a thematic environment dedicated to Architecture and Urbanism, ARQUIGRAFIA is one of the rare current systems with free login, free upload and download, which still allows its users to tag, georeference and assign specific license (Creative Commons) to every image from their collections.

All these characteristics are relevant when comparing ARQUIGRAFIA to other digital images databases that still operate fundamentally as Web 1.0, namely: Built Works Register (http://builtworksregistry.org); Artstor (http://www.artstor.org); GettyImages (http://www.gettyimages.com.br)

The multidisciplinary challenges of the ARQUIGRAFIAproject led to the creation of the *Web Collaborative Environments Research Center (NaWeb)* bringing together researchers from the School of Architecture and Urbanism of the University of São Paulo (FAU/USP), in partnership with the Institute of Mathematics and Statistics (IME/USP) and the School of de Communications and Arts (ECA/USP).

The NaWeb research group investigates aspects related to representation and architectural imagination, conservation and digitization of photographic collections, and issues related to the development of free software, dynamic social networking, as well as topics relevant to information science and ontologies.

Actually, NaWeb is a partner of the INCT (National Institute of Science and Technology) of the Future Internet for Smart Cities.

Since 2012 the ARQUIGRAFIA project has been developed by NaWeb in close cooperation with the Iconographic Material Sector of the Library of the School of Architecture and Urbanism (FAU), and this collaborative e-work is based on 3 viewpoints: the architectural critical perspectives, the library and information science and the computer science and web design approaches.

Under the architectural critical perspectives, the ARQUIGRAFIA contributes to expand and popularize the debate on architecture, cultural heritage and documentation, thus stimulating critical judgment over the iconographic representation of the Brazilian architectural and urban production. Doing so, the project proposes a public debate on the nature of collective knowledge building supported by images, related to time (present, past and future) and space (georeferenced images), that can be integrated in a real urban experience with the support of smartphones and Apps. In order to encourage users to closely observe every image and formulate esthetic judgments about buildings and urban spaces represented in photographs, ARQUIGRAFIA proposes the registration of individual impressions based on pairs of opposing qualities organized as semantic differentials such as: open/closed; internal/external; translucent/opaque; complex/simple; symmetrical/asymmetrical; horizontal/vertical.

The approach under the point of view of the Library and Information Science makes possible a closer look at some of its contemporary issues, specifically regarding the representation of images (Dell Valle Gastaminza, 1999; Moreiro Gonzalez and Robledano Arillo, 2003; Smit, 1996), controlled vocabularies and user's participation. Here, we have discussed the knowledge organization in the Brazilian architectural domain, by analyzing the digital repository, the website databases and the analytical-critical studies of the indexing languages and the tags proposed by the users; as well as the consolidation of the standards, which respond both to international interoperability requirements and to local needs for organization and information access.

The Computer Science allowed ARQUIGRAFIA to develop its own software based on PHP Laravel, which can be freely shared as a +GRAFIA template to support other similar initiatives for building georeferenced iconographic collaborative environments on the Internet. Since 2015 the research team has been working on an Android App now available for free download in Brazil at the Google Play Store. Since the first ARQUIGRAFIA Web version in 2010, the project has assumed its beta perpetual nature, and has defined an User-Centered Design (UCD) methodological process that still runs the web design development. At the moment ARQUIGRAFIA faces challenges related to the sustainability of a continuous growth system and explores the collaborative resources that a social network can provide such as moderators.

Methodology

The institutional images inserted in ARQUIGRAFIA weretaken from the FAUUSP Library photographic collection started back in the late 1960's with the systematic donation of students and professors' private images, aiming to build an 'Imaginary Museum' like proposed by André Malraux in 1947, on Architecture, Urbanism and Design. In order to guarantee the preservation, organization, representation, scanning and retrieval of information from the collection,it was made available on the web by the

ARQUIGRAFIA team, which has promoted at a much larger scale the convergence of private and public collections, and also the continuous expansion of the collection, thanks to the web contribution of photographers, students, teachers, researchers, and people interested in architecture. Furthermore, the images register improves public based full access to stimulate the construction of an architectural visual culture.

This work front involves the processesof management, organization and preservation of information, taking into account the interdisciplinary practices within the project. Some of the subjects covered in each stage were the physical photography material processing, the slide or negative scanning; the adequacy of metadata and digital resources formats and the logical treatment system for modeling data collection (Cunha and Lima, 2007; International Federation of Library Associations and Institutions, 1987). Besides that, it was necessary to develop the ARQUIGRAFIA controlled vocabulary, which must include the architecture terminology, the user's natural language and the institutional indexing language (Austin and Dale, 1993; International Standard Organization, 2000; 2009; 2011; Royan and Cremer, 2006; Cabre, 1995). A survey of user's new terms descriptors was undertaken and studies about the adequacy of the design content representation for computational ontologies and information-system-based knowledge were developed. Finally, the project has implemented the infrastructure and repositories; the archiving and backup policies; the data management policies and information retrieval.

To enable a better understanding of the collection features in quantitative and qualitative terms, for a more precise identification of treatment needs, control of costs and work schedule, a diagnosis has been made, considering a critical analysis of the amount of photographic documents, the type of media used, the format (single photograph, slide), chromia and standard dimensions of the photographic documents. We have made the identification of the principal problems for conservation such as abrasion, discoloration, deforming, wrinkling, mirroring, evidence of fungi and stains. The cleaning of the 42,000 images of the institutional collection was done mechanically with air sprayer and a non-abrasive tissue. The use of solvents was necessary only in a few slides stained or damaged by fungi (Braga, 2003).

After the cleaning,a selection of relevant documents for the ARQUIGRAFIA has been made. We have selected 34,000 original slides, 8,000 images on paper, including color and black-and-white photographs on Brazilian buildings and urban spaces, the majority of them being unpublished. About 6,724 images from the library collection have already been catalogued and are already online. In addition to that, the ARQUIGRAFIA has 2,660 images from private users and 1,765 images from other institutional collections totalizing 11,149 images online.

The public profile of ARQUIGRAFIA included architecture students (38.6%); architects (23.4%); graduate students in other areas (10.3%); architecture professors (4.3%); photographers (4.1%). Most users are between 20 and 30 years old, have an undergraduate degree or attend the graduation courses.

Since 2013, the scanning has been made by an outsourced company using a Plustek Optic film 120 and a scanning software, SilverFast Ai Studio 8 (64-bit), which helps in the removal of dust and scratches. The generated files are about 5 MB to 4,000 dpi in TIFF, JPEG and PDF, recorded on DVDs and external hard drives. No computation colors correctionis made in the scanned images, in order to preserve the original appearance of the photographs. Thus, the marks of time (color changes, spots,

saturation, etc) are respected and maintained, aiming to maintain the collection with its historical aspects.

Every single image is identified by a registration number, which enables the cross-checking and comparison between spreadsheet data and images. During the read out, each spreadsheet row is transformed into business objects used by the system. To do this transformation, *The Apache ODF Toolkit* (http://incubator.apache.org/odftoolkit) is used in order to offer a communication interface between the spreadsheet and the system for information mining and transformation. Then, the information storage activity lies in creating associations between the business objects and their storage in database. Once the object association is done (an author is associated to an image and this image to an address) the system uses the persistence library *Hibernate* (http://www.hibernate.org) to store database under *Mysql* (http://www.mysql.com).

It was necessary to clarify the procedures for the representation and retrieval of digital heritage, which is being made in order to ensure access to and sharing of information between all users, whether institutional or individual. We have analyzed the cataloguing standards (AACR; ISBD(NBM); CCO; CDWA), and developed a spreadsheet to integrate the required fields for the data administration in virtual environment. Then we established the set of metadata with the following elements:

• **Description**: Title, Classification number, Name, Country, State, City, District, Street, Image author, Picture date, Building author, Building date, Notes, Registration number date, Cataloguing date.

• **Technical:** Dimensions, Width, Height, Resolution, Bit depth, Color representation.

•**Administrative**: License (Creative Commons), Collection, Donors, Authorization term for web broadcasting.

Bearing in mind that at the ARQUIGRAFIA both personal and institutional users can insert images of collections of photos, it was necessary to do some terminology standardization work between the lists of subjects used by the library for the indexing of photographs and slides; the Controlled Vocabulary of the Integrated Library System of the University of Sao Paulo (VOCAUSP), and the list of tags developed by the team, based on dictionaries and thesauri of architecture. This terminology standardization has resulted in a list of terms used for the indexing of images, which was enriched with tags assigned by private users to their images. A survey was conducted on the keywords employed by the users in order to collect new terms that can be added to the controlled vocabulary.

Users must fill in at least the title and author of the image, country and tags to upload an image to ARQUIGRAFIA, a procedure that can be performed on the website or the App. The system suggests terms selected from the list, but the users can insert new tags in natural language. This inclusion of tags without vocabulary control is called folksonomy, social indexing or social tagging. This indexing process can contribute greatly to the creation and management of digital collections, as it is carried out in a collaborative way, distributing resources, activities, and reducing costs in a way that becomes important for the organization, retrieval and access to digital information. These tags that belong to the semantic universe of the users can enrich the vocabulary, show how the users think and retrieve the information. This could constitute another front of studies in this project. The terminological work contributes to the elaboration of a taxonomy from the tags inserted in the ARQUIGRAFIA, which are compared with the

VOCAUSP, conceptualized and categorized. The expected outcome of this phase will be the ARQUIGRAFIA controlled vocabulary.

Another aspect to be addressed is the process of the indexing of institutional images that should be considered under three different aspects: the denotation, connotation and context as we can see in the example in the Figure 1.

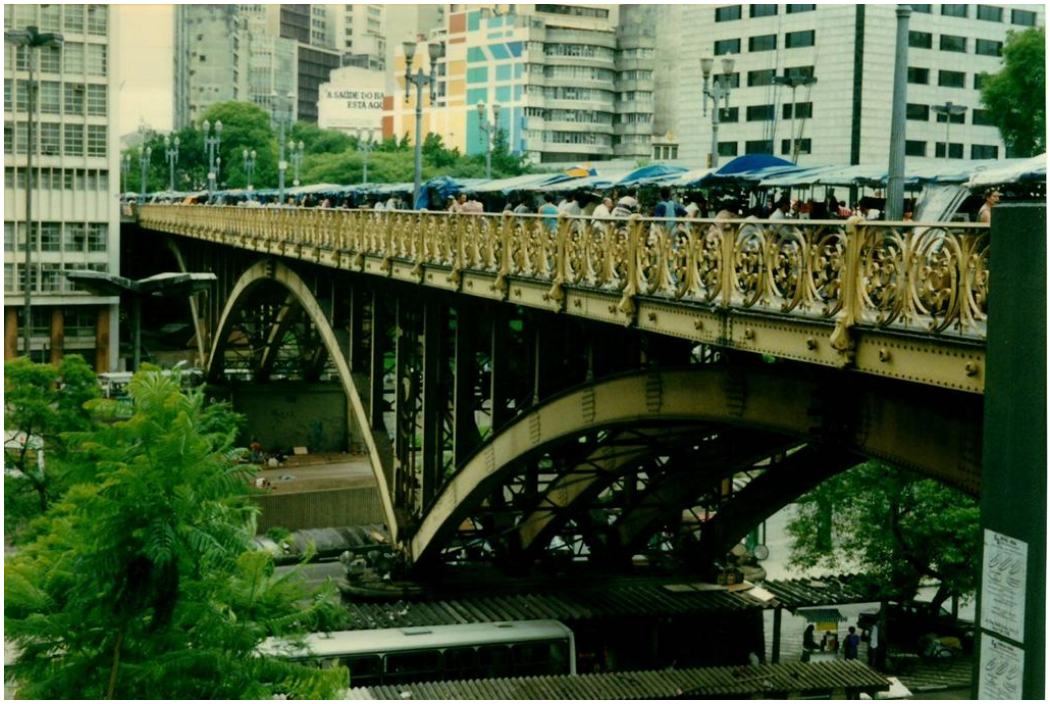

Figure 1 Santa Efigênia Bridge, São Paulo, Brazil (ARQUIGRAFIA, 1997).

The denotation refers to what appears in the picture, being a descriptive reading, where the meaning denotes what will be recognized unequivocally by both the indexer and the user:

Denotation =Santa Efigenia, bridge, São Paulo

The connotation is what the image suggests, which refers to the emotional content of the message, giving meaning to the whole, translating religious, ideological and cultural aspects. It is when the image reading passes through the collective memory:

Connotation = Bridges (any)

The context refers to the image production conditions and involve issues such as how, when, where and why. Therefore, to determine the subject of a picture, one should not only identify and relate what you see, but the ideas behind it, its production conditions, and if possible the author's purpose:

Context = Fair Street at Santa Efigenia bridge

From this analysis we can select the terms that best represent the concepts and translate them under what best fits the information contained in the document.

Recent results

Below, we have list the main achievements and results obtained by the project in the last two years:

- The improvement of the interaction of user impressions records regarding visible aspects of the architectures represented in photographic images based on the binomials of spatial-spatial qualities integrated as semantic differentials, expressed in synthetic infographics, still integrated to the advanced system search;
- The conclusion of the digitalization of a set of 42 thousand images, 34 thousand slides (updated number of the set actually referring to Brazilian buildings and urban spaces of the FAUUSP collection) and 8 thousand physical photographs;
- The cataloguing and uploading of more than 6,724 images from the Library Collection of FAUUSP into the ARQUIGRAFIA system with the authorizations of copyright holders, the respective Creative Commons licenses, metadata, classification and indexing;
- Improvement of image upload mechanisms that were previously performed indirectly based on Excel spreadsheets for direct upload to the online system with differentiation between cataloging / indexing forms for ordinary users (private users) and institutional users (collections) with the reuse of data inserted in an image for the next upload;
- Improvement of an imprecise date insertion mechanism for cataloging images; as well as the inclusion of a mechanism that allows the insertion of data of images of the collection, which still do not have the license of the author to be privately exhibited to the institutions that registered them, without public viewing;
- The design of autocomplete controlled lists for: complete and correct names of Brazilian architects, as well as area specific tags, based on the Controlled Vocabulary of the Integrated Library System of USP (VOCAUSP) in the areas of Architecture, Urbanism, Arts and Engineering;
- The implementation of improvements in system interaction functionalities through UCD procedures, based on gamification elements aimed at greater user engagement especially with interface elements related to collaboration, such as: notifications, news, images and comments, “follow” and “be followed” by other users. And also design for gamification, such as: leadership chart, badges, week highlights, and scores;
- The design and implementation of a bilingual landing page integrated with ARQUIGRAFIA's communication enhancement project with users, potential users and public relations;
- Full development and consolidation of the ARQUIGRAFIA software in PHP (Laravel) with proper integration with the front end with a white background;
- Code refactoring and its fluid interactions, both with the development of the Android App and the design of the template + GRAFIA;
- The development of a first version of the ARQUIGRAFIA Android App available for smartphones in the Google Play Store (since 11/11/2016), which intensified the interaction of the system with new users and the direct experience of urban

spaces in Brazilian cities;

- The enlargement of the ARQUIGRAFIA iconographic base with the experimental upload of drawings and videos as a pilot for a wider digital images platform.

Current Objectives

Considering the recent results, the ARQUIGRAFIA research team deals with new short term objectives:

- Implementation of a first version of the following functionalities: system moderation integrated with game elements (gamification); Instant communication of users logged into the system; Quiz as an instrument for intensifying architectural learning from questionnaires answered by users. Such functionalities have as main objective to promote an active community around the images and its information, seeking long-term sustainability for the system;
- Follow-up for a further 6 months the online version system and its consequent critical evaluation by users according to User-Centered Design with A / B Tests and formal usability tests on the system with the listed functionalities. Statistical analysis of users' engagement metrics with the system will allow the evaluation of the impact of recent changes made on the project;
- Monitoring and analysis of the integration of new institutions into ARQUIGRAFIA, based on the dissemination of the "Manual of ARQUIGRAFIA's Technical Project Procedures";
- The full development and free public dissemination of the +GRAFIA template integrated to the improvement of the Android and IOS App and the open source diffusion of the App + GRAFIA template.
- Deepening of the research on new pertinent themes, future developments and enlargements of ARQUIGRAFIA for the next two years.
- Conduct studies relating to binomials plastic-spatial qualities arranged as semantic differential seeking to define an evaluation model for images of ARQUIGRAFIA.
- Set metadata from the adequacy of descriptive fields of AACR2 cataloging for updated standards as the RDA set (http://www.rda-jsc.org/rda.html); FRBR (http://www.ifla.org/publications/functional-requirements-for-bibliographic-records) or RDA/CCO (http://cco.vrafoundation.org), which are interoperable with the web environment and allow, for example, the application of conceptual ontologies.
- Conduct usability studies to evaluate the current beta version and system redesign from the critical observations made by users
- Expand the iconographic base of ARQUIGRAFIA including experimental digital video, scanned drawings from the original collection (plans, sections, elevations, perspectives and details) of the FAU Library, with due procedure for metadata record
- Improve the indexing and implement search tools in the system data.

Conclusion

The ARQUIGRAFIA provides the convergence of institutional (public and private) and personal collections to one single open and free access website/Android App. For the scope of knowledge organization, this specific nature provides a singular opportunity to look accurately at its contemporary issues, specifically regarding the representation of images, vocabularies, interoperability among systems and user active participation. Furthermore, the template of ARQUIGRAFIA free software system, named +GRAFIA can be replicated and adapted to other areas of knowledge beyond architecture. In order to preserve and disseminate a digital collection of Brazilian architecture, the ARQUIGRAFIA project has been building an unprecedented network - promoting interaction between many researchers - and succeeded in constituting a fruitful scientific research laboratory for the continuous construction of interdisciplinary knowledge.

Bibliography

Alexander, Christopher, Sara Ishikawa, and Murray Silverstein. 1977. A pattern language: towns, buildings, construction. New York.

Assis, César. 1997. Viaduto Santa Efigênia: ARQUIGRAFIA.

Austin, Derek, and Peter Dale. 1993. Diretrizes para a elaboração de tesauros monolíngues. 2. ed. Brasília: IBICT/Senai.

Braga, Gedley Belchior. 2003. Conservação preventiva: acondicionamento e armazenamento de acervos complexos em reserva técnica: o caso do MAE/USP. Master, University of Sao Paulo. School of Communications and Arts.

Cabré, Maria Teresa. 1995. La terminologia hoy: concepciones, tendências y aplicaciones. *Ciência da Informação* 24 (3):289-298. doi: http://dx.doi.org/10.18225/ci.inf..v24i3.567.

Cunha, Jacqueline de Araújo, and Lima, Marcos Galindo. 2007. Preservação digital o estado da arte. Encontro Nacional de Pesquisa em Ciência da Informação 8, Salvador, Oct 28-31 2007.

Del Valle Gastaminza, F. 1999. Indización y representación de documentos audiovisuales. In Manual de documentación fotográfica, 467-488. Madrid: Editorial Síntesis.

Edmondson, Ray. 1998. Uma filosofia de arquivos audiovisuais. Paris: UNESCO/UNISIST.

Institutions, International Federation of Library Associations and. 1987. ISBD (NBM) International Standard Bibliographic Description for Non-Book Materials. Rev. ed ed. London: IFLA.

Institutions, International Federation of Library Associations and. 1998. Supports photographiqueset films. In IFLA Principes de conservation, 53-75. IFLA.

Moreiro Gonzalez, Jose Antonio, and Jesús Robledano Arillo. 2003. O conteúdo da imagem. Curitiba: Ed. UFPR.

O'Reilly, Tim 2007. What is Web 2.0: design patterns and business models for the next generation of software. *Communications & strategies* (65):17-37.

Organization, International Standard. 2000. ISO 1087: Terminology work - vocabulary. Geneva: ISO.

Organization, International Standard. 2009. ISO 704: Terminology work - principles and methods. Geneva: ISO.

Organizations, International Standards. 2011. ISO 25964-1: Information and documentation: thesauri and interoperability with other vocabularies , part 1 Thesauri for information retrieval. Geneva: ISO.

Royan, Bruce, and Monika Cremer. 2006. Directrizes para materiais audiovisuais e multimedia em bibliotecas e outras instituições. IFLA Professsional Reports, 80. London: IFLA.

Smit, Johanna Wilhelmina. 1996. A representação da imagem. *Informare: Cadernos do Programa de Pós-Graduação em Ciência da Informação* 2 (2):28-36.