

SOCIAL SHAPING OF DIGITAL PUBLISHING:  
EXPLORING THE INTERPLAY BETWEEN CULTURE  
AND TECHNOLOGY

This page intentionally left blank

# Social Shaping of Digital Publishing: Exploring the Interplay Between Culture and Technology

Proceedings of the 16th International Conference  
on Electronic Publishing

Edited by

**Ana Alice Baptista**

*University of Minho, Portugal*

**Peter Linde**

*Blekinge Institute of Technology, Sweden*

**Niklas Lavesson**

*Blekinge Institute of Technology, Sweden*

and

**Miguel Abrunhosa de Brito**

*University of Minho, Portugal*

**IOS**  
Press

Amsterdam • Berlin • Tokyo • Washington, DC

© 2012 The authors and IOS Press.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior written permission from the publisher.

ISBN 978-1-61499-064-2 (print)

ISBN 978-1-61499-065-9 (online)

Library of Congress Control Number: 2012940104

Cover picture: Rui Vale de Sousa (ruivaledesousa.com)

*Publisher*

IOS Press BV

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: [order@iospress.nl](mailto:order@iospress.nl)

*Distributor in the USA and Canada*

IOS Press, Inc.

4502 Rachael Manor Drive

Fairfax, VA 22032

USA

fax: +1 703 323 3668

e-mail: [iosbooks@iospress.com](mailto:iosbooks@iospress.com)

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

## Preface

Since the advent of the Web, the processes and forms of electronic publishing have been changing. The open access movement has been a major driver of change in recent years with regard to scholarly communication; however, changes are also evident in other fields of application such as e-government and e-learning. In most cases these changes are driven by technological advances, but there are also cases where a change in social reality pushes technological development. Both the social and mobile web and linked data are currently shaping the edge of research in digital publishing. Liquid publishing is on the more daring agendas. Digital preservation is an issue that poses great challenges which are still far from being solved. The legal issues, security and trust continue to deserve our full attention. We need new visualization techniques and innovative interfaces that will keep pace with the global dimension of information. This is the current scenario, but what will follow? What are the technologies and social and communication paradigms that we will be discussing in ten or twenty years?

ELPUB 2012 focuses on the social shaping of digital publishing, exploring the interplay between culture and technology. This makes the fact that it is being held in the European Capital of Culture for 2012, Guimarães, Portugal, all the more appropriate.

52 submissions were received for ELPUB 2012, from which 23 articles and 10 posters were accepted after peer review. Of the accepted articles, 11 were submitted as full articles and 12 as extended abstracts. These articles have been grouped into sessions on the following topics: Sessions 1 and 4 – Digital Scholarship & Publishing; Session 2 – Special Archives; Session 3 – Libraries & Repositories, Session 5 – Digital Texts & Readings, and Session 6 – Future Solutions & Innovations.

The programme features two keynote speeches. Kathleen Fitzpatrick's speech is entitled "Planned Obsolescence: Publishing, Technology, and the Future of the Academy", that of Antonio Câmara is entitled "Publishing in 2021". Finally we call your attention to the panel on e-books, which is entitled "Academic e-books – Technological hostage or cultural redeemer?".

We believe this is another great edition of the ELPUB conference. We would like to take this opportunity to thank both the members of the ELPUB executive committee and the members of the local advisory committee, for making it happen. Together they provided valuable advice and assistance during the entire organization process. Secondly we would like to mention our colleagues on the program committee, who assured the quality of the conference through the peer review process. Last but not least, we wish to thank the local organization team for ensuring that all this effort culminates in a very interesting scientific event on the 14th and 15th of June. Thank you all for helping us to maintain the quality of ELPUB and merit the trust of our authors and attendees. We wish you all a good conference and we say farewell, hoping to see you again in Sweden in 2013!

The editors,

Ana Baptista

Peter Linde

Niklas Lavesson

Miguel A. Brito

This page intentionally left blank

## Organizing Chairs

**General Chair:** Ana Alice Baptista – University of Minho (Portugal)  
**Programme Chair:** Peter Linde – Blekinge Institute of Technology (Sweden)  
**Organizing Chair:** Miguel Abrunhosa de Brito – University of Minho (Portugal)

## Executive Committee

John Smith – University of Kent at Canterbury (UK)  
 Peter Linde – Blekinge Institute of Technology (Sweden)  
 Karel Jezek – University of West Bohemia in Pilsen (Czech Republic)  
 Ana Alice Baptista – University of Minho (Portugal)  
 Sely Costa – University of Brasília (Brazil)  
 Jan Engelen – Catholic University of Leuven (Belgium)  
 Mícheál Mac an Airchinnigh – Trinity College Dublin (Ireland)  
 Bob Martens – Vienna University of Technology (Austria)  
 Leslie Chan – University of Toronto (Canada)  
 Susanna Mornati – CILEA (Italy)  
 Turid Hedlund – Swedish School of Economics and BA, Helsinki (Finland)  
 Yasar Tonta – Hacettepe University (Turkey)

## Local Organization Team

Ana Alice Baptista – University of Minho (Portugal)  
 Miguel Abrunhosa de Brito – University of Minho (Portugal)  
 Mariana Curado Malta – ISCAP / University of Minho (Portugal)  
 Cristina Marques Gomes – University of São Paulo (Brazil) / University of Minho (Portugal)  
 Luís Miguel Costa – University of Porto / University of Minho (Portugal)  
 Anabela Serrano – University of Aveiro (Portugal)

### **Local Advisory Committee:**

Gabriel David – University of Porto (Portugal)  
 Cristina Ribeiro – University of Porto (Portugal)  
 Fernanda Ribeiro – University of Porto (Portugal)  
 Joaquim Arnaldo Martins – University of Aveiro (Portugal)  
 Joaquim Sousa Pinto – University of Aveiro (Portugal)  
 Maria Manuel Borges – University of Coimbra (Portugal)  
 José Luís Borbinha – Technical University of Lisbon (Portugal)  
 Paulo Quaresma – University of Évora (Portugal)

# Programme Committee

Agenjo, Xavier – Ignacio Larramendi Foundation (Spain)  
 Al, Umut – Hacettepe University (Turkey)  
 Alvim, Luísa – University of Évora (Portugal)  
 Baptista, Ana Alice – University of Minho (Portugal)  
 Barreto, Aldo – University of Grande Rio (Brazil)  
 Bernardino, Teresa – Politechnical Institute of Viana do Castelo (Portugal)  
 Bia, Alejandro – Universidad Miguel Hernández (Spain)  
 Björnshauge, Lars – SPARC Europe  
 Borbinha, José – Technical University of Lisbon (Portugal)  
 Borges, Maria Manuel – University of Coimbra (Portugal)  
 Can, Fazlı – Bilkent University (Turkey)  
 Catarino, Maria Elisabete – University of Londrina (Brazil)  
 Chan, Leslie – University of Toronto Scarborough (Canada)  
 Costa, Sely M.S. – University of Brasilia (Brazil)  
 David, Gabriel – University of Porto (Portugal)  
 Delgado, Jaime – Politechnical University of Catalunya (Spain)  
 Dias de Figueiredo, António – University of Coimbra (Portugal)  
 Engelen, Jan – Catholic University of Leuven (Belgium)  
 Ferreira, Miguel – Keep Solutions (Portugal)  
 Frías, José Antonio – University of Salamanca (Spain)  
 Gathegi, John – University of South Florida (USA)  
 Griffin Speer, Julie – Georgia Institute of Technology (USA)  
 Hedlund, Turid – Swedish School of Economics and BA, Helsinki (Finland)  
 Hernández, Francisca – DIGIBIS Producciones Digitales (Spain)  
 Horstmann, Wolfram – University of Bielefeld (Germany)  
 Iyengar, Arun – IBM Research (USA)  
 Jezek, Karel – University of West Bohemia in Pilsen (Czech Republic)  
 Johnson, Henric – Blekinge Institute of Technology (Sweden)  
 Johnson, Ian M. – The Robert Gordon University (Scotland)  
 Kuramoto, Hélio – IBICT (Brazil)  
 Lavesson, Niklas – Blekinge Institute of Technology (Sweden)  
 Linde, Peter – Blekinge Institute of Technology (Sweden)  
 Mac An Airchinnigh, Micheal – Trinity College Dublin (Ireland)  
 Martens, Bob – Vienna University of Technology (Austria)  
 Martins, Joaquim Arnaldo – University of Aveiro (Portugal)  
 Mele, Salvatore – CERN (Switzerland)  
 Mendéz, Eva – Universidad Carlos III, Madrid (Spain)  
 Moreira, Fernando – Portucalense University (Portugal)  
 Mornati, Susanna – CILEA (Italy)  
 Morrison, Heather – British Columbia Electronic Library Network (Canada)  
 Nicolaisen, Jeppe – Royal School of Library and Information Science (Denmark)  
 Nisheva-Pavlova, Maria – Sofia University (Bulgaria)  
 Quaresma, Paulo – University of Évora (Portugal)  
 Ramalho Correia, Ana Maria – New University of Lisbon (Portugal)  
 Ribeiro, Cristina – University of Porto (Portugal)



Ribeiro, Fernanda – University of Porto (Portugal)  
Rodrigues, Eloy – University of Minho (Portugal)  
Saraiva, Ricardo – University of Minho (Portugal)  
Savenije, Bas – Utrecht University Library (The Netherlands)  
Schranz, Markus – Presstext Austria (Austria)  
Serrano, Anabela – Politechnical Institute of Porto (Portugal)  
Smith, John – University of Kent at Canterbury (UK)  
Sousa Pinto, Carlos – University of Minho (Portugal)  
Sousa Pinto, Joaquim – University of Aveiro (Portugal)  
Soydal, İrem – Hacettepe University (Turkey)  
Tenopir, Carol – University of Tennessee (USA)  
Tonta, Yasar – Hacettepe University (Turkey)  
Van Helvoort, Jos – The Hague University of Applied Sciences (The Netherlands)  
Vaz de Carvalho, Carlos – Politechnical Institute of Porto (Portugal)  
Virkus, Sirje – Tallinn University (Estonia)  
Ünal, Yurdagül – Hacettepe University (Turkey)

This page intentionally left blank

# Contents

Preface	v
<i>Ana Baptista, Peter Linde, Niklas Lavesson and Miguel A. Brito</i>	

Committees	vii
------------	-----

## Full Articles

### Digital Scholarship & Publishing

Creative Commons: A Convergence Model Between the Ideal of Commons and the Possibilities of Creation in Contemporary Times, Opposed to Copyright Impediments	3
<i>Elizabeth Roxana Mass Araya and Silvana Aparecida Borsetti Gregorio Vidotti</i>	

### Repositories & Libraries

Open Access in Developing Countries: African Open Archives	12
<i>Lydia Chalabi and Madjid Dahmane</i>	
The University and Its Libraries: Reactions and Resistance to Scientific Publishers	24
<i>Flavia Bastos, Silvana Vidotti and Nanci Oddone</i>	
Information Retrieval Educational Goals in Library and Information Science and in Health Sciences	34
<i>Anabela Serrano</i>	

### Digital Texts & Readings

Evolving Definitions of Authorship in Ebook Design	43
<i>Celeste Martin and Jonathan Aitken</i>	

### Special Archives

Archives Information Publishing New Design in Post-Custodial Regime: The National Archives Experience Digital Vaults	52
<i>Maria José Vicentini Jorente</i>	
Ontology-Based Search and Document Retrieval in a Digital Library with Folk Songs	62
<i>Maria Nisheva-Pavlova and Pavel Pavlov</i>	

## **Future Solutions & Innovations**

- Knowledge Network of Scientific Claims Derived from a Semantic Publication System 72  
*Carlos Henrique Marcondes*
- WIKI::SCORE – A Collaborative Environment for Music Transcription and Publishing 82  
*José João Almeida, Nuno Ramos Carvalho and José Nuno Oliveira*
- Book Design Program: A Transition to a Hybrid Publishing Context 94  
*Ana Catarina Silva and Maria Manuel Borges*
- “Tangible Culture” – Designing Virtual Exhibitions on Multi-Touch Devices 104  
*M. Hecher, R. Möstl, E. Eggeling, C. Derler and D.W. Fellner*

## **Extended Abstracts**

### **Digital Scholarship & Publishing**

- A Comparison of an Open Access University Press with Traditional Presses: Two Years Later 117  
*Rory McGreal, Nian-Shing Chen and Tim McNamara*
- Huddersfield Open Access Publishing 118  
*Graham Stone*
- Freemium as a Sustainable Economic Model for Open Access Electronic Publishing in Humanities and Social Sciences 119  
*Pierre Mounier*
- Some Trends in Electronic Publication and Open Access in Portuguese History Journals 120  
*Maria Cristina Guardado and Maria Manuel Borges*
- Development of a Publishing Framework for Living Open Access Textbooks 121  
*Anita Eppelin and Richarda Böttcher*
- OAPEN-UK: An Open Access Business Model for Scholarly Monographs in the Humanities and Social Sciences 122  
*Caren Milloy, Graham Stone and Ellen Collins*

### **Repositories & Libraries**

- Accessibility and Self Archiving of Conference Articles: A Study on a Selection of Swedish Institutional Repositories 123  
*Peter Linde, Jörgen Eriksson, Lars Kullman, Margareta Fathli, Klemens Karlsson, Marjatta Sikström, Ylva Sköld and Ingela Tång*

**Digital Texts & Readings**

- Text vs Visual Metaphor in Mobile Interfaces for Novice User Interaction 125  
*Chrysoula Gatsou, Anastasios Politis and Dimitrios Zevgolis*
- Digital Reading: The Transformation of Reading Practices 126  
*Gustavo Cardoso, Carla Ganito and Cátia Ferreira*
- Are Turkish Universities Ready for E-Learning: A Case of Hacettepe University  
Faculty of Letters 127  
*İrem Soydal, Gülten Alir and Yurdagül Ünal*

**Special Archives**

- The Biographical Portal of the Netherlands – Experiences in Collecting and  
Curating Data from the Web and Elsewhere 128  
*Jelle Gerbrandy and Els Kloek*

**Future Solutions & Innovations**

- The Future of Digital Magazine Publishing 129  
*Dora Santos Silva*
- Subject Index 131
- Author Index 133

This page intentionally left blank

## Full Articles

This page intentionally left blank



# Creative Commons: a Convergence Model Between the Ideal of Commons and the Possibilities of Creation in Contemporary Times, Opposed to Copyright Impediments

Elizabeth Roxana MASS ARAYA<sup>a,1</sup> Silvana Aparecida BORSETTI GREGORIO  
VIDOTTI<sup>a</sup>

<sup>a</sup>*Universidade Estadual Paulista Júlio de Mesquita Filho*

**Abstract.** The contemporary individual finds on the Internet and especially on the Web facilitating conditions to build a basic infrastructure based on the concept of commons. He also finds favorable conditions which allow him to collaborate and share resources for the creation, use, reuse, access and dissemination of information. However, he also faces obstacles such as Copyright (Law 9610/98 in Brazil). An alternative is Creative Commons which not only allows the elaboration, use and dissemination of information under legal conditions but also function as a facilitator for the development of informational commons. This paper deals with this scenario.

**Keywords.** copyright, commons, creative commons, information and technology, Web 2.0, information science

## Introduction

Information and Communication Technologies (ICT), the result from the fusion of analog telecommunications and computing in the 1970s, allow information contents in text, image and sound formats to be conveyed by a single support (the computer). They also enhance the progressive passing of information in a hierarchy in which communication is from one to all (mass media) to individualized forms of production, diffusion and storage of information based on rhizomatic multiplicity, that is, communication from all to all: Internet allows hypertextual communication between groups and individuals and it decentralizes dissemination of information whereas the Web encourages interactivity and collaborative action [1].

Information and Communication Technologies are present in the daily life of a significant portion of global society and they promote important changes in the way information is developed and accessed in terms of time and space: “*Information spaces store*

---

<sup>1</sup> Corresponding Author: Elizabeth Roxana Mass Araya, Av. Hygino Muzzi Filho, 737, Campus Universitário Marília-SP, Brazil. E-mail: elizabethmass@ymail.com.

information items, memories, means of transfer, and the interaction reality of information recipients all in the same communication environment" [2].

Therefore, individuals are now experiencing an ever growing social context in which ICT, especially the Internet and the Web on the Internet, allows them to produce information and not only be information consumers. Interaction and collaboration practices favored by ICT fill information spaces in which these individuals' creativity modifies the flow of information, expanding it and giving it voices not too long absent. Article 19 of the Universal Declaration of Human Rights has currently a greater chance of becoming reality: *"Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers"* [3].

However, despite finding on the Internet and especially on the Web facilitating conditions to build an infrastructure based on the concept of commons as well as favorable conditions to collaborate and share resources for the creation, use, reuse, access and dissemination of information, individuals also find obstacles. The copyright (Law 9610/98 in Brazil), is one of them.

This paper focus on Creative Commons licenses as an alternative system to copyright that ensure creation and dissemination of information within legal patterns and promote informational commons

## 1. Commons in the Web 2.0

Web 2.0 is an information space characterized by the use of collective intelligence [4] as a booster of network connections growth: users add value to the network. In the Web 2.0 complex context, collaboration is understood as a practice people interested in achieving some common goal use. Each individual's motivation is what enhances collaborative action, as it works well for complex situations:

*"Complexity is unpredictable. So is collaboration, but it is also adaptable. Complexity is disordered and diverse. Collaboration calls for the diversity and the creativity of people who improvise and innovate."* [5].

The individual's creativity is also favored in this Web. Remix, which Lemos [06] defines as set of social and communicational practices that combine collage and information content cut-ups from digital technology are more present in informational processes. But such practices are not new, as each individual naturally creates from previous creations. The new, therefore, would be precisely in the technology and in the ease with which the product of this technology can be shared.

Cyberspace is an information environment in which individuals are free to play a much more active role than it was possible in the industrial information economy of the 20th century.

*"This new freedom has a great practical promise: as a dimension of individual freedom, as a platform for better democratic participation as a means to promote a culture more critical and self-reflective and, in a global economy increasingly dependent on information, as a mechanism to achieve improvements in human development everywhere."* [7].

In this information environment, creativity, collaboration and sharing are strongly associated with a phenomenon that has become stronger in the digital culture scenario:

*“A series of changes in the technologies, economic organization, and social practices of production in this environment has created new opportunities for how we make and exchange information, knowledge, and culture. These changes have increased the role of nonmarket and nonproprietary production, both by individuals alone and by cooperative efforts in a wide range of loosely or tightly woven collaborations.” [7].*

This phenomenon is called commons, a word whose meaning can be what is common or spaces and things that are public. In some cases it can have the meaning of community or of a shared production between pairs [8]. In the present context, commons refer to:

*“A particular type of institutional arrangement that governs the use and the disposition of resources. Their main characteristic, which defines them differently from property, is that no individual has the exclusive control of use and disposition of any resource in particular. On the contrary, resources governed by the community may be used and disposed by anyone among a given number of people (more or less well defined) according to rules which can vary from “everything is valid” to clear rules formally articulated and effectively imposed”. [9]*

Digital commons and their undeniable power are present, according to Bollier [10], for example, in the success of Linux and other open code software or in Wikipedia and Flickr growth. Internet itself, Lessig writes, is a common:

*“The space that anyone can enter, and take what she finds without the permission of a librarian, or a promise to pay. The net is built on a commons — the code of the world wide web, html, is a Computer language that lays itself open for anyone to see — to see, and to steal, and to use as one wants. If you like a web page, then all major browsers permit you to reveal its source, download it, and change it as you wish. It’s out there for the taking; and what you take leaves as much for me as there was before”. [11].*

It is through the internet that projects such as Move Commons (MC), created in 2011, run. MC was meant to be a tool for initiatives, collectives, non-governmental organizations and social movements to state the basic principles they use in their projects, as well as to make their identification on the web easier. Therefore, MC will be able to help others to understand and find other people’s work and approach in an easier manner, to find other collectives with common interests, and to attract volunteers to these collectives.

On the MC website (<http://movecommons.org/preview/ject>), the individual writes the name of his initiative, the web address, the key words related to his project and chooses four labels among the following ones to identify it:

**Non-Profit**- Your initiative does not seek economic benefit or **For-Profit**-Your initiative seeks economic benefit; **Exclusive**-Your initiative prefers to protect their work and not disclose it to others or **Reproducible**-Your initiative is visible, open and transparent, which allows it to be adapted by others; **Reinforcing the Commons**-Your initiative protects/expands the Commons, resources that are collectively owned or **Reinforcing Other Aims**-Your initiative focuses on other areas different than The Commons and **Representative**-Your initiative is driven by representatives or **Grassroots**-Your initiative is driven by horizontal collective decisions. The Move Commons logo with characteristics of the project is both human and machine readable.

MC arises from the initiative of Comunes (<http://comunes.org/>), a non-profit collective dedicated to making the work of other collectives and activists easier by developing web tools and free resources. Projects such as Ourproject.org and Kune are some Comunes projects.

Ourproject (<http://ourproject.org/>) active since 2002, now hosts 1,000 projects and its services receive around 1,000,000 monthly visits. It acts as a central location that offers a web space and tools for projects that focus on free knowledge of any topic. It provides multiple web services (hosting, mailing lists, wiki, ftp, forums, etc.) to

social/cultural/artistic projects as long as they share their contents with Creative Commons licenses or other free licenses.

Kune (<http://comunes.org/>) is a platform that encourages collaboration, content sharing and free culture. It aims to improve, modernize and replicate the work ourproject.org does, but in an easier manner and expanding its features for community-building. It allows the creation of online spaces of collaborative work, where organizations and individuals can build projects online, coordinate common agendas, set up virtual meetings and join people and/or organizations with similar interests.

Humankind experiences a unique moment in history. With the development of ICT a paradigm, in which new habits are being incorporated, arises. Collaboration and sharing are common practices in the digital culture which modify the way information is elaborated and disseminated.

In this context, views and theories indicate the transformation of a society of passive consumers who used what few others produced into a society in which everyone can participate actively in its social, political and cultural construction. The ideal of Commons is seen as a practice that makes this transformation possible. Copyright, however, has to be reconsidered as technological development confronts laws that protect intellectual work and create new conflicts and contradictions such as piracy, a common practice among the great majority of producers and intellectual content consumers.

## **2. Brazilian Legislation: Law 9610/98**

In Brazil every creator of a literary intellectual, artistic or scientific work is protected. This protection constitutes a moral right (creation) and a property right (pecuniary). It is regulated by Law No. 9610 of 02/19/98 ([http://www.planalto.gov.br/ccivil\\_03/leis/L9610.htm](http://www.planalto.gov.br/ccivil_03/leis/L9610.htm)) and it focuses on the author, unlike copyright law which focuses on the work and on the prerogative of being able to be copied.

The work protected by Brazilian law is the one which represents an intellectual manifestation expressed through any medium or fixed in any tangible or intangible support, whether known or to be invented in the future. Originality is the basic requirement for a work to be protected legally.

Brazilian law protects any work regardless if it is registered or not, for the time the author lives, and for 70 years following his/her death. After that, the work will be of public domain and may be used without authorization, provided they do not hurt the author's moral rights.

Law 9610/98 as well as main intellectual property right institutions, Lemos writes, [6] were created in the nineteenth century based on a completely distinct social reality and have not been changed ever since.

Nowadays, therefore, copyright acts as a big "no". One must obtain prior authorization to use any work due to "all rights reserved", and as Lemos [12] warns, the consequence is that transaction costs to obtain a previous authorization will restrict immensely the amount of culture that society can access at any given time.

Brazil has been interested in creating discussion environments to find solutions and suggest changes for copyright laws. From 2007 until November 2009, in the so called National Forum of Authors' rights, almost all categories in question (authors, artists, publishers, record companies, users, consumers, among others) were able to criticize and suggest a review for the current law. In June 2010, the Ministry of Culture made

Public Consultation available. On the Ministry of Culture website citizens could make suggestions and review the current authors' rights law with the suggested changes.

*"The online consultation was divided into two periods, each of them spanning roughly 45 days. The first period of the consultation involved a debate about general principles, which then served as reference for the writing of the text of draft Bill. These principles were divided into three groups: (1) individual and collective rights (privacy, freedom of speech, and access rights), (2) principles related to intermediaries (net neutrality and civil liability), and (3) governmental directives (openness, infrastructure, and capacity building). The draft text for the Bill, reflecting the comments received on its first phase, was then put under consultation for the second period". [13].*

This project, an initiative of the Brazilian Minister of Justice in partnership with the Center for Technology and Society, Getulio Vargas Foundation (CTS / FGV), called the "Civil [Regulatory] framework," as opposed to the "Criminal framework" is currently being discussed in Congress as the Bill of Law no. 2126/2011 (<http://www.a2kbrasil.org.br/wordpress/wp-content/uploads/2011/09/Marco-Civil-Ingles-CC-82s-pm.pdf>).





*"The Civil Regulatory Framework contains 25 articles divided into 5 chapters, concerning: (1) Preliminary Provisions; (2) User Rights and Guarantees; (3) Provision of Connection Services and Internet Services; (4) The Role of Public Authorities; (5) Final Provisions. The bill begins by advancing users' rights and some general principles for the regulation of the Internet, before dealing with the issues of the preservation of connection logs, secondary liability for ISPs, and net neutrality, and then wraps with directives aimed at the public sector". [13].*

However, although society is concerned with reviewing authors' rights, Brazilian scenario is not very different from the rest of the world. Individuals once predominantly information consumers are or may become producers, but they must find alternatives to develop their information processes according to what is foreseen in the law. Therefore, society segments concerned with possible restrictions seek alternatives to ensure common citizens their basic right to information as they believe it is essential for the development of a global culture with fewer inequalities.

### **3. An alternative: Creative Commons**

Creative Commons licenses are an alternative that can contribute to the flowing of information processes, even though the legislation in effect does not benefit them. Creative Commons (<http://creativecommons.org>) is a non-profit project with voluntary membership which was designed by North American Lawrence Lessig and is headquartered at Stanford University in the United States. It is already present in seventy-two countries.

Since 2002, Creative Commons has offered individual creators, large companies and institutions, a set of audio, image, video, text, and education licenses. These licenses offer intellectual content authors such as musicians, filmmakers, writers, photographers, bloggers, journalists, scientists, educators and others a simple and standardized alternative to grant permission to copy rights and get credit for their creative work. It allows, therefore, others to copy, distribute and make specific uses of works. In other words, it allows these works to have "Some rights reserved" - according to the uses the author wishes to grant these creations - as opposed to the traditional "all rights reserved" of Copyright.

Creative Commons offer six license models by combining the licenses  (Attribution)  (NonCommercial),  (NoDerivs) and  (ShareAlike).

**Attribution** (give credit) of original work to the author is a common requirement of all Creative Commons licenses. **NonCommercial** license prohibits any person who has access to the work to sell or use it for profit. **NoDerivs** license grants the author free copy, distribution and use of the work. The author, however, does not authorize changes or its use to create new works. **ShareAlike** means the author authorizes free copy, distribution and use of the work but requires that in case the work is used to create other works, the resulting work must be shared by the same license.

**Attribution** license lets others distribute, remix, tweak, and build upon other individuals' work, even commercially, as long as they credit the creator for the original. **Attribution-ShareAlike** license lets others remix, tweak, and build upon other individuals' work even for commercial purposes, as long as they credit the creator and license their new creations under the identical terms. **Attribution-NoDerivs** license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the author. **Attribution-NonCommercial** license lets others remix, tweak, and build upon other individuals' work non-commercially, and although their new works must also acknowledge the creator and be non-commercial, they don't have to license their derivative works on the same terms. **Attribution-NonCommercial-ShareAlike** license lets others remix, tweak, and build upon other individuals' work non-commercially, as long as they credit the creator and license their new creations under the identical terms. **Attribution-NonCommercial-NoDerivs** license lets others to download other individuals' work and share them with others as long as they credit the creator, but they can't change them in any way or use them commercially.

Creative Commons licenses have three different links: a summary of the license in plain language (Commons deed) containing relevant items for the user's understanding; the detailed license (Legal Code) with legal terms that guarantee the validity before the court, and the digital code of the license with a language that helps computers and search engines as well as other applications to identify the work and its terms of use.

Creative Commons licenses are already present in a great number of society segments. The number of works licensed under Creative Commons was less than 1 million at the end of its first year. By the end of 2010, it had reached over 400 million only in Yahoo! SiteExplorer and Flickr. After the first year, only about 20% of the licensed works allowed commercial use and remix, that is, "free" or "open" use. After eight years, this number has almost doubled [14].

It is worth mentioning that Creative Commons licenses have become important tools for Open Access Movement to scientific literature. They are considered ideal to protect authors' rights and at the same time grant access and use of scientific production, as the author may hold the right to prohibit distribution of modified, non-authorized copies as well as copies intended for commercial use. *"Basically, it legally prevents plagiarism, disfiguration and commercial use of the work. It also allows all kinds of uses requested by lawful academic use, including research and consultation easiness".* (15).

In 2003, Brazil joined Creative Commons licenses coordinated by Ronaldo Lemos, a lawyer and law professor at Getulio Vargas Foundation in Rio de Janeiro. Brazil was the third country after Japan and Finland to join the project. Lemos argues that Creative Commons licenses foster creative freedom, access, dialogue and cultural transformation, and thus enhances one of Brazilian's characteristic: to transform cultural elements of their own and other cultures creatively [12].

In May 2010, CCMonitor (<http://monitor.creativecommons.org/Brazil>) had 1,796.052 works registered under Creative Commons licenses from different society segments in Brazil.

In August 2006, the movie *Cafuné* directed by Bruno Vianna was released simultaneously in theaters around the country and on the Internet under a Creative Commons license which allowed anyone to make their own cut of the film, create different interferences in the story and release it subsequently, as long as they used the same Creative Commons license.

Bruno Viana's initiative opened new possibilities to help overcome the limited number of distribution channels. Despite the growing number of movies produced in Brazil, only approximately 50 of them are released in Brazilian movie theaters every year, creating a significant distribution funnel. With Viana's idea, movies are not restricted to a few movie theaters that can show them only for a short period. Internet can increase the number of viewers and favor a public that rarely has access to different types of knowledge.

The first film festival under Creative Commons, CC MAD, Madrid Creative Commons Film Festival (<http://ccmad.cc/>) organized by BccN, Barcelona Creative Commons Film Festival, was held in Spain in January 2012. It addressed new forms of filmmaking in the digital culture under Creative Commons licenses as they connect the author directly to viewers, boost shared creation and production and allow a larger number of viewers to see the work.

These initiatives have the same interests for musicians worldwide, but especially in developing countries. In fact, music was the first field to benefit greatly from alternative licenses.

The Brazilian band *Projeto Axial* that explores different sounds from electronic elements and makes its music available under a Creative Commons license has created a new free music media format called *Bagagem* (<http://www.axialvirtual.com/Axial/Axial/Bagagem.html>). *Bagagem*, also available under a Creative Commons license, retrieves and updates the concept of visual artwork lost in the digital network. It produces, packages, and returns the visual interface of physical means lost in this virtuality. It brings musicians, artists, and the social network itself together, disseminates works and personal projects of users and spreads music freely.

Creative Commons licenses are also used by several Brazilian government agencies such the National Center for Folklore and Popular Culture (<http://www.cnfcp.gov.br/>) that develops programs and projects for studies, research, documentation, diffusion and promotion of Brazilian culture; the Research Program in Biodiversity (<http://ppbio.inpa.gov.br/>); the Department of Airspace Control (<http://www.decea.gov.br/>), and the Digital Library of the Superior Court of Justice (<http://bdjur.stj.gov.br/xmlui/handle/2011/17968>) among others.

Among the several other highly relevant initiatives benefiting a society that calls for changes in the way public affairs are considered is SERPRO (<https://www.serpro.gov.br/>), a public enterprise under the Ministry of Finance which provides Information Technology and Communications services for the public sector in order to facilitate the access to educational resources. It offers 14 complete courses developed by Serpro's Corporate University under a Creative Commons license.

The courses offered are Time Management, Water for All, BrOffice 3.1, Express Mail, Understandings Computing, Fundamentals of Java PHP5, Introduction to Project Management, Introduction to Java Programming Language, Introduction to Object

Oriented Programming, First Job, Awareness to use Free Software, Ubuntu, Using UML, and Mozilla Firefox. These courses, available for distance education, range from environmental issues to info digital inclusion, and provide the necessary knowledge to ensure the inclusion of citizens in their community.

Another initiative has been the publication of every teaching material produced by the Education Department of São Paulo state under a Creative Commons license since June 2011. In September of the same year, this led to the Decree No. 52681<sup>2</sup> enacted by São Paulo City Hall which determines that all educational material produced by the Education Department of São Paulo state is released on the Internet for free download. This allows school systems of other cities and states, as well as teachers, NGOs and others to use the content produced by the Municipal School Network of São Paulo.

São Paulo City Hall's proposal goes beyond a merely political interest and promotes knowledge socialization which is built primarily with resources society itself has and provides.

Creative Commons project is an alternative for the imbalance between what Information and Communication Technologies provides and what traditional laws (still related to the press culture) establish for the production of informational content.

Commons are environments in which non-proprietary production does not belong to the market system. They result from individual and collaborative participation of different individuals in different parts of the world. They carry the promise of a free and democratic dimension as a concept of promoting a conscious and critical culture that can be practiced in a global economy based on information and knowledge.

#### **4. Considerations**

We are experiencing the third generation of Information Science – an era of interactive knowledge marked not only by the possibility of transforming every information input provided by ICT into a digital database, but also by significant changes in the way we produce and disseminate information.

Individuals in the digital cultural scenario not only consume information but also produce it. Therefore, they have their concepts as to authorship and creation reformulated in their information processes.

In modern times, originality and authorship concepts referred to a privileged individual able to create works from a spontaneous inspiration. In contemporary times, these concepts are more related to a shared, collaborative and interactive creation.

Views / theories show a society once predominantly of passive consumers changing into a society of active consumers who participate in their social, political and cultural construction. In this context, the ideal of Commons is seen as an agent of this transformation.

However, new habits require renewed looks. One of them should deal with established Copyright models.

Creative Commons project is one of the alternatives that promote information flow in the contemporary context as opposed to those that generate excessive censorship and

---

2



control such as the Sinde law in Spain (which determines the closure of file-sharing sites), the French Hadopi law (which seeks to punish people who download copyrighted content) or the Stop Online Piracy Act (an act the United States intends to approve to legalize the closure of any site containing material that violates copyright or intellectual property rights). Creative Commons project offers technical and legal infrastructure for the collaboration and effective sharing of information contents. These practices enrich digital commons – environments which grow and convey the ideal that information may be an effective universal human right.

Creative Commons licenses do not meet all needs digital culture requires, but they certainly minimize the current paradoxical situation – even in scientific production – and promote information flow, access to knowledge and the development of a fairer global culture.

## References

- [1] A. Lemos, *Cibercultura, tecnologia e vida social na cultura contemporânea*. Sulina, 2<sup>a</sup> ed., Porto Alegre, 2004.p.68.
- [2] A.d A. Barreto, A condição da informação. *São Paulo Perspec.* [online]. 2002, vol.16, n.3 [cited 2010-02-25], pp. 67-74. <<http://www.scielo.br/pdf/spp/v16n3/13563.pdf>> viewed Dec.2011.
- [3] United Nations. *Universal Declaration of Human Right*. <[http://www.un.org/events/humanrights/2007/hrphotos/declaration%20\\_eng.pdf](http://www.un.org/events/humanrights/2007/hrphotos/declaration%20_eng.pdf)> viewed Dec.2011.
- [4] P. Levy, *A inteligência coletiva*. Loyola, São Paulo, 1998.
- [5] A. Schuschny, *Compartir, coordenar, cooperar, colaborar*. 27 jul.2009. <<http://humanismoyconectividad.wordpress.com/2009/07/27/compartir-coordenar-cooperar-colaborar/>> viewed Dec. 2011.
- [6] A. Lemos, *Ciber-Cultura-Remix - Seminário “Sentidos e Processos” dentro da mostra “Cinético Digital”*. Itaú Cultural, São Paulo, agosto de 2005. <<http://www.facom.ufba.br/ciberpesquisa/andrelemos/remix.pdf>> viewed Dec.2011.
- [7] Y. Benkler, *Wealth of networks: How Social Production Transforms Markets and Freedom*. Yale University Pres, E.U.A.2006. p.2.
- [8] S.A d. Silveira, *Comunicação digital e a construção dos common: redes virais, espectro aberto e as novas possibilidades de regulação*. Fundação Perseu Abramo, São Paulo, 2007.p.7.
- [9] Y. Benkler, A economia política dos commons. In *Comunicação digital e a construção dos common: redes virais, espectro aberto e as novas possibilidades de regulação*. Fundação Perseu Abramo, São Paulo, 2007.p.12.
- [10] D. Bollier, *Elinor Ostrom and The Digital Commons*, 2009. <<http://www.forbes.com/2009/10/13/open-source-net-neutrality-elinor-ostrom-nobel-opinions-contributors-david-bollier.html>> viewed Dec.2011.
- [11] L. Lessig, *Code and the Commons Draft*, Keynote, given at a conference on Media Convergence Fordham Law School, New York, NY, February 9, 1999. P.3. <<http://cyber.law.harvard.edu/works/lessig/Fordham.pdf>> viewed Dec.2011.
- [12] R. Lemos, Creative Commons, mídia e as transformações recentes do direito da propriedade intelectual. *RevistaDIREITOGV*, v.1, n.1, p181-187, maio 2005. <<http://www.inovacao.unicamp.br/report/Ronaldo-Lemos.pdf>> viewed Dec.2011.
- [13] J. Varon, *English translation of the Draft Bill Proposition on Civil Rights-Based Framework for Internet in Brazil*. <<http://www.a2kbrasil.org.br/wordpress/lang/pt-br/2011/09/latest-version-of-civil-rights-based-framework-for-internet-in-brazil/>> viewed Dec.2011.
- [14] Creative Commons. *The power of open*. <<http://thepowerofopen.org/>> viewed Dec.2011.
- [15] Acesso Aberto Brasil. *Sobre o movimento Acesso Aberto* <<http://www.acessoaberto.or/>> viewed Dec.2011.

# Open Access in developing countries: African Open Archives

Lydia CHALABI<sup>a</sup>, Madjid DAHMANE<sup>b</sup>

<sup>a</sup>DRDSI, CERIST research center, Algiers, Algeria; Information Science Department,  
University of Algeirs2, Algeria

*lchalabi@mail.cerist.dz, net\_lydia@yahoo.fr*

<sup>b</sup>DRDSI, CERIST research center, Algiers, Algeria  
*mdahmane@mail.cersit.dz*

**Abstract.** This paper presents the results of a study on the open archives in developing countries. It provides the elements of type, size and contents of open archives. The methodology is based on information collected from open repository websites. The survey is based on an almost exhaustive sample list of developing countries websites retrieved from directories and a list of open repositories. The purpose of this study, carried out from 2011 up to now, is to measure, at a second level, the impact of open access on the Algerian researchers by analyzing their practices related to open access, through the identification of their scientific publications at a second level in the open archives, in which they can deposit.

**Keywords.** Open Access, developing countries, Open repository, Open archives, academic publication, Africa.

**Acknowledgements:** I am grateful to Chérifa Boukacem, for her invaluable support and guidance throughout this study.

## Introduction

Open Access movement has become an increasingly strong and significant part of research landscape and a major political issue [1]. The aim is to meet the dissatisfaction of the traditional publishing model and make academic research results freely available for anyone, anywhere through the World Wide Web, via two ways Open Access scientific journals, or Open Access archives.

In this context, open access is the prime alternative to publish in traditional journals, whose subscription, being based on business model, inhibits the distribution of scientific knowledge [2].

Moreover, the studies and efforts expended to create open archives and to measure their impact have demonstrated the strength of open access to increase the impact of publications that have led publishers to move towards greater acceptance of self-archiving, so the number of open archives increased through making deposit compulsory for universities, institutions and research foundations [3].

We do mean by open access repository or open archives; a server / platform to deposit (authors or laboratories) and consult publications, documents, texts, data and tools... of the research world and higher education. This must be accessible to through Internet without any restriction, using common protocols that facilitate the accessibility to contents from multiple servers which are maintained by different data providers. In addition to that, it should allow "...fast and wide dissemination of results; the preservation of the record..."[4].

We mean by institutional repository; a comprehensive archive of the whole institution outputs (research, heritage, educational, administrative, ...), preservation and management in order to serve interests of faculty researchers and teachers.[5 It proposes a wide variety of materials in digital form, such as research journal articles, preprints and post prints, digital versions of theses & dissertations, administrative documents, course notes, or learning objects."[6].

However, while the open access movement is becoming a trend in the developed countries, and both the scientific community and professionals have started analyzing the results of its existence through studies on open archives use, studies on this movement and on its impact over the scientific production of developing countries are still lacking.

In this way, the objective of this part of study is to identify the developing countries open archives, by mapping out its characteristics, such as the volume of deposits of scientific researchers, their disciplinary belonging, fields and type. Once this unique mapping defined, we can have an analytical view, so that the results will help building further works and analysis, such as studying the researchers practices and the uses of these open archives (own researchers of the repository or foreign researchers who may file in this repository).

It should be noted that the developing countries repositories (Africans open repositories in this paper) is just a part of the overall sample of the global study.

## 1. Methodology

To have clear and reliable information on our questioning, our methodology is as follows:

*State of the art:* This part was elaborated to study the existing operations and analyze the available literature in our study.

*Directories and lists of reference websites:* The second step is to identify the open archives directory web sites to obtain a comprehensive sample of open archives. The list of directories listing of open archives is appended as Appendix 1.

*Criteria analysis & data collection:* After analyzing the open archives directories lists, we have compiled a list of open archives in all subject fields. We opted for open archives whose interface language is at least one of the following:

- Arabic (mother tongue of the Arab-Muslim African developing countries)

- French (French-speaking region: common colonial history of North Africa)
- English (openness to the market economy).

This selection criterion is the specificity of our study. For each site, we checked the URL, the location in Africa and the presence of recent deposits. Each site was characterized on the basis of the following criteria:

- General information: (name, logo, URL, institution, institution type ...)
- Open archives characteristics (year of establishment, Archives type, subject, content, software, interface language, number of items, the possibility of filing by author, ...)

It should be noted that the development of the analytical framework was based on several surveys: Schöpfel [1], Bath [6], Chérifa Boukacem & all [8] and Ben-Romadhane [16].

Data collection and results obtained are incorporated into a word processing statistical data (Sphinx), verified, validated and / or modified.

The study has been conducted as follows:

- Identification of reference directories: September 2011
- Choice of open archives: September-November 2011
- Data collection of open archives: September-December 2011
- Validation and operating data: November to January, 2012

## 2. Results

### 2.1. Size, institutions and languages

During the analysis of the directories and reference lists of open archives we have listed more than 50 web sites under the name of an open access repository (digital libraries, open access journals articles sites.

Through the selection criteria we have defined above and in order to limit the scope of the study and make it significant and homogeneous we added another criterion; open archives containing over 100 documents. And as a result, we obtained 27 Open Archives in different subject fields, visited and thoroughly tested.

Most of the different institutions that are responsible for the management of open archives belong to the public research, and are distributed as the table above.

**Table 1.** Number open archives by institutions type

Institutions type	Sites on 2011
University	26
Research center	1
Academic library	0
Others	0

The majority of the archives are presented with an English interface, which reveals and demonstrates the understanding and awareness of developing countries research institutions, that English is the language of exchange, sharing and research.

It is also symptomatic of the history of some of these countries, which have been colonized by UK in the past.

**Table 2.** Interface language

Language	Sites on 2011
Arabic	1
French	2
English	25
Others	4

## 2.2. Typology

Another analysis step based on Armbruster & Romary (2009) typology of open archives focuses in our study on three types by adding the document type archive:

- Subject-based or disciplinary repository which records documents relating to a common topic. We identified two Africans repositories; South Africa “UCT Computer Science Research Document Archive” and “UCT Lawspac”
- National or Central repository that gather up scholarly output of different communities (research centers, institutes and libraries). Among these sites to national character, there are for example Uganda Scholarly Digital Library and that of Addis Ababa University Libraries
- Institutional repository contains various outputs of the institution
- Open archive by kind of document (such as thesis)

**Table 3.** Open Repositories types

Types	Number of site in 2011
Institutional repository	27
Subject-based or disciplinary repository	2
National or central repository	5
Document type repository	4

All the open repositories collected belong to Institutional type. However, it should be noted that these institutional repositories can include one or two other type of open repositories, as presented in the table below. University of Cape Town Lawspace South Africa repository is an example. It is an institutional and subject repository at the same time.

It is sort of reactions, symptomatic of a lack of Scientific and technical information (STI) policy and open access policy.

By combining the two criteria - type of institution and type of archives - we obtain the following table:

**Table 4.** Number of sites by archives type and institutions type (2011)

	University	Research center	Others
Institutional repository	26	1	0
Subject-based or disciplinary repository	2	0	0
National or central repository	9	0	0
Document type repository	4	0	0

The results show that open archives in developing countries are institutional type and belong to the public sector.

### 2.3. Discipline

According to their discipline, we have classified the open repositories into three categories:

- Multidisciplinary that covers the humanities social sciences, applied sciences, and science, technology and medicine
- Humanities and Social Science (HSS)
- Science, technology and medicine, covering applied science ( STM)

**Table 5.** Discipline of open archives in developing countries 2011

Disciplinary	Number of site in 2011
Multidisciplinary	23
HSS	2
STM	2

The results show that most of the sites are multidisciplinary which means that there are several institutional open archives of that kind which is logical regarding the fact that most of them are institutional and are linked to university libraries.

In the other disciplinary type, the number is equal and not significant compared to the multidisciplinary sites.

If we compare disciplinary with institutions and the type of archives, we obtain the following results:

**Table 6.** Type of archives and disciplinary archives 2011

	Institutional	Subject	National	Document type
Multidisciplinary	24	0	5	5
HSS	1	1	0	0
STM	2	1	0	0

Almost institutional sites are multidisciplinary. According to their discipline, there are only two repositories; one related to humanities & social science (SHS), the other to Science, technology & medicine (STM).

**Table 7.** Type of institutions and disciplinary archives 2011

	University	Research center	Academic library	Others
Multidisciplinary	21	1	2	0
HSS	1	0	0	0
STM	2	0	0	0

The most developing countries are multidisciplinary repositories

In the area where the subjects are clearly defined (SHS & STM) in our case disciplinary open archives consist of universities archives.

#### 2.4. Content growth and document type

**Table 8.** Number of items the first year of the survey 2011

	Number of site in 2011
Unanswered	1
Less than 2000	15
From 2000 to 4000	3
From 4000 to 6000	5
From 6000 to 8000	2
From 8000 to 10000	1
From 10000 to 12000	0
12000 and more	1

**Minimum= 133, Maximum=16002, Sum=79339, Mean=2938, 48, Standard deviation=3550.36**

The majority of repositories have less than 2000 items: 144 items for Kenya SU-Portal, 597 items for Tunisia UVT e-doc, etc.

The only repository of more than 12000 items, is SUNScholar Stellenbosch University in South Africa which is also an online academic research information service supplier of the library we can cite SUNJournals, SUNProxy and SUNSearch<sup>1</sup>.

This content is split up into the following types of documents. Most is for these, articles and conferences respectively

<sup>1</sup> Stellenbosch University Research Repository, from <http://scholar.sun.ac.za/>

We can see also a great number of content for other documents which include maps, letters and dataset (records with abstracts)<sup>2</sup>, speeches, books chapters, manuscripts, inaugurates lectures etc.

**Table 9.** Number of items by type of documents

<b>Number of site in 2011</b>	Books	Articles	Conferences	Thesis	Reports	Teaching and learning	Multimedia	Others	Total
Unanswered	0	0	0	0	1	0	0	0	<b>1</b>
Less than 2000	2	12	11	12	5	4	5	7	<b>58</b>
From2000to 4000	0	2	2	3	1	0	0	1	<b>9</b>
From4000to 6000	0	3	3	4	0	1	0	4	<b>15</b>
From6000to 8000	0	1	0	2	0	1	1	1	<b>6</b>
From8000 to 10000	0	0	0	0	0	0	0	0	<b>0</b>
From10000 to12000	0	0	0	0	0	0	0	0	<b>0</b>
12000 and more	1	1	1	1	1	0	1	1	<b>7</b>
<b>TOTAL</b>	<b>3</b>	<b>19</b>	<b>17</b>	<b>22</b>	<b>8</b>	<b>6</b>	<b>7</b>	<b>14</b>	<b>96</b>

## 2.5. Software & statistics

As to the African developing countries open repositories software, DSpace takes the Lion's share.

Nevertheless there are a few attempts to use free software innovation for African Higher Education Research Online in South Africa and Digital Assets Repository in Egypt.

**Table 10.** Number of sites by categories of software

<b>Software</b>	<b>Number of sites</b>
Dspace	20
Eprints	3
ETD-db	1
Free Software innovation	2
Greenstone	1
Digital Commons	1

Otherwise, a few numbers of open archives which offer use statistics

**Table 11.** Statistics data offering

<b>Statistics Data</b>	<b>Number of sites</b>
Yes	03
No	23

<sup>2</sup> Dataset: references.



### 2.6. Submission

Most archives allow authors to submit a document Most of them do not define any Deposit policy

**Table 12.** Authors Ability submission

Submission	Number of sites
Yes	23
No	1
Not mentioned	4

### 3. Findings and Discussion

The study reveals results which give a better vision of the OA repositories reality in this area of the world:

University appears clearly as the most active and more involved in open access projects.

Like the other developed countries, English language – Science dominates and we can hypothesize that the developing open archives present windows to abroad.

As was clearly indicated by Schöpfel (2010) in his report, the non-commercial distribution has found a means of communication, which are open institutional archives in particular. In our case of study, the first Position (non-commercial) publication items returns to thesis.

However, few are open archives whose purpose is the communication of pre-publication and post publication. : the only archives that proposes these types of documents are la Wits institutional repository of Dspace of preprint publication and SUNScholar Reseach Repository for post print, both in South Africa, both are from south Africa

Directories and repositories references list reveal some heterogeneity. Each site listing applies its own criteria. Directories referencing identify under the heading of open archives, any kind of site that the most common is free access to content.

Although free access means unrestricted access, unlimited and free, our results reveal another reality.53.6% of open archives in developing countries do not open completely access to their content, in the same area the access is limited and reserved only for the institution users.

Virtually no site informs archive creation date as well as of its evolution by document type. It is difficult to have the number of items per type. The Open Archives such as DSpace software rarely provides those statistics, only Eprint offers this possibility.

The heterogeneity observed is probably the result of a lack of political leadership, and that initiatives are actually quite isolated and experimental. For proof, the critical mass of open archive remains quite modest.

It clearly transpires that the open archives, particularly institutional archives, dominating our results are the way the academic libraries have embarked in, in the absence of a national policy. These are places to archive and display their scientific production results. But it seems they are not integrated to the enlarged process of the scientific communication. They play the role of 'a second library', in adding in digital items to the traditional library (theses, limited access to local users,...) but they are not engaged in an open scientific publication. This can explain the "confidential" character of some of them. Finally, we can say that OA have come along to assist libraries traditional functions however they do not yet allow taking part in the open scientific communication adventure in these countries

## References

- [1] Joachim Schöpfel, Hélène Prost, Développement et Usage des Archives Ouvertes en France. 1e partie: Développement, report, 2010, from [http://archivesic.ccsd.cnrs.fr/docs/00/49/73/89/PDF/DUAO\\_Partie1.pdf](http://archivesic.ccsd.cnrs.fr/docs/00/49/73/89/PDF/DUAO_Partie1.pdf)
- [2] S. Bernius, M. Hanauske, Open access to scientific literature: Increasing citations as an incentive for authors to make their publications freely accessible, *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences*, 2009.
- [3] Hélène Bosc, Le droit des chercheurs à mettre leurs résultats de recherche en libre accès : appropriation des archives ouvertes par différentes communautés dans le monde, *N° spécial « Propriété Intellectuelle »*, 2008.
- [4] C. Armbruster & L. Romary (2009), Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication, *Social Science Research Network Working Paper Serie* (2009).
- [5] Carr Leslie, Brody Tim, Size isn't everything: Sustainable repositories as evidenced by sustainable deposit profiles, *D-lib Magazine*, 13(7/8). (2007) from <http://www.dlib.org/dlib/july07/carr/07carr.html>
- [6] M. H. Bath, Open access repositories in computer science and information technology: an evaluation'. *IFLA Journal* 35(3) (2009), 243-257.
- [7] Smith MacKenzie, Communicating with data: new roles for scientists, publishers and librarians, *Learned Publishing*, Vol. 24, No. 3. (2011), pp. 203-205.
- [8] C. Boukacem-Zeghmouri, M. Ben Romdhane, A. Abdi, Le libre accès l'information scientifique dans le pays en voie de développement : étude comparative de ses potentialités et réalités en Algérie et en Tunisie, *Colloque international franco-tunisien (SFSIC/ISD/IPSI) – (2008)*.
- [9] Laurent Romary, Communication Scientifique : pour le meilleur et pour le PEER, *Hermès*, Vol. 57 (2010), pp. 131-136.
- [10] Kathlin Smith, Institutional Repositories and e-Journal Archiving: What Are We Learning? *JEP: the journal of electronic publishing, Volume 11, Issue 1* (2008) from <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0011.107>
- [11] Linda Hulbert, Dani Roach, Gail Julian, Integrating Usage Statistics into Collection Development Decisions, *The Serials Librarian*, Vol. 60, No. 1. (2011), pp. 158-163.

- [12] Tammy Sugarman, Stephanie Krueger, Louise Kelly, Evaluating Usage of Non-text Resources: What the COUNTER Statistics Don't Tell You, *The Serials Librarian*, Vol. 60, No. 1. (2011), pp. 83-97.
- [13] Rieger, Ova Y, Select for success: key principles in assessing repository models, *D-lib Magazine*, 13(7/8) (2007) from <http://www.dlib.org/dlib/july07/rieger/>
- [14] Jeff Matlak, What Drives Usage, *Journal of Electronic Resources Librarianship*, Vol. 22, No. 3. (2010), pp. 144-165.
- [15] Christopher Stewart, Keeping Track of it all: The Challenge of Measuring Digital Resource Usage, *The Journal of Academic Librarianship*, Vol. 37, No. 2. (18 March 2011), pp. 174-176.
- [16] M. Ben Romdhane, Lobna Ouertani, Le libre accès à l'information scientifique et technique à travers l'implication du chercheur tunisien en sciences de l'information et de la communication dans les archives ouvertes, *Colloque : "Les nouveaux medias : nouvelles technologies pour un monde nouveau"*, Bahrein, (2009)

## Appendix 1: Reference Directories

**Dspace:** DSpace Registry.

<http://www.dspace.org/whos-using-dspace>

**OpenDOAR:** Directory of Open Access Repositories.

<http://www.openoar.org/>

**ROAR:** Registry of Open Access Repositories.

<http://roar.eprints.org/view/software/>

**APBHATOO:** Documentary research portal.

<http://www.abhato.net.ma/index.php/fre/Webographie>

**Webometrics:** Ranking Web of World Repositories.

<http://repositories.webometrics.info/>

**AIOA:** Arab Initiative of Open Access

<http://aioa.blogspot.com/>

## Appendix 2: African Developing Countries Open Archives list

Archives	URL	Countries
Addis Ababa University Libraries Electronic Thesis and Dissertations Database	<a href="http://etd.aau.edu.et/dspace/">http://etd.aau.edu.et/dspace/</a>	Ethiopia

Institutional repository for KNUST	<a href="http://dspace.knust.edu.gh:8080/jspui/">http://dspace.knust.edu.gh:8080/jspui/</a>	Ghana
Mahider	<a href="http://mahider.ilri.org/handle/10568/1">http://mahider.ilri.org/handle/10568/1</a>	Kenya
SU-Portal	<a href="http://www.digital.library.strathmore.edu/">http://www.digital.library.strathmore.edu/</a>	Kenya
Biens Culturels Africains	<a href="http://bca.ucad.sn/jspui/">http://bca.ucad.sn/jspui/</a>	Senegal
African Higher Education Research Online	<a href="http://ahero.uwc.ac.za/">http://ahero.uwc.ac.za/</a>	South Africa
Research Repository North-West University	<a href="http://dspace.nwu.ac.za/community-list">http://dspace.nwu.ac.za/community-list</a>	South Africa
CSIR Research Space	<a href="http://researchspace.csir.co.za/dspace/">http://researchspace.csir.co.za/dspace/</a>	South Africa
Digital Knowledge	<a href="http://dk.cput.ac.za/">http://dk.cput.ac.za/</a>	South Africa
Durban university of technology-institutional repository	<a href="http://ir.dut.ac.za/handle/10321/210">http://ir.dut.ac.za/handle/10321/210</a>	South Africa
Rhodes eResearch Repository	<a href="http://eprints.ru.ac.za/">http://eprints.ru.ac.za/</a>	South Africa
SUNScholar Research Repository	<a href="http://scholar.sun.ac.za/">http://scholar.sun.ac.za/</a>	South Africa
UCT Computer Science Research Document Archive	<a href="http://pubs.cs.uct.ac.za/">http://pubs.cs.uct.ac.za/</a>	South Africa
University of Cape Town Lawspace	<a href="http://lawspace2.lib.uct.ac.za/dspace/">http://lawspace2.lib.uct.ac.za/dspace/</a>	South Africa
UJDigispace	<a href="https://ujdigispace.uj.ac.za/">https://ujdigispace.uj.ac.za/</a>	South Africa

UKZN ResearchSpace	<a href="http://researchspace.ukzn.ac.za/xmlui">http://researchspace.ukzn.ac.za/xmlui</a>	South Africa
Unisa Institutional Repository	<a href="http://uir.unisa.ac.za/">http://uir.unisa.ac.za/</a>	South Africa
University of Fort Hare Institutional Repository	<a href="http://ufh.netd.ac.za/js">http://ufh.netd.ac.za/js</a>	South Africa
University of Limpopo Institutional Repository	<a href="http://ul.netd.ac.za/jspui/">http://ul.netd.ac.za/jspui/</a>	South Africa
University of Pretoria Electronic Theses and Dissertations	<a href="http://upetd.up.ac.za/UPeTD.htm">http://upetd.up.ac.za/UPeTD.htm</a>	South Africa
UWC Research Repository	<a href="http://repository.uwc.ac.za/xmlui/">http://repository.uwc.ac.za/xmlui/</a>	South Africa
University of Zululand Repository	<a href="http://uzspace.uzulu.ac.za/">http://uzspace.uzulu.ac.za/</a>	South Africa
Wits Institutional Repository on DSpace	<a href="http://wiredspace.wits.ac.za/">http://wiredspace.wits.ac.za/</a>	South Africa
Sudan Open Archive	<a href="http://www.riftvalley.net/index.php?view=soahome">http://www.riftvalley.net/index.php?view=soahome</a>	Sudan
DSpace@ScienceUofK	<a href="http://oascir.uofk.edu/">http://oascir.uofk.edu/</a>	Sudan
UVT e-doc	<a href="http://pf-mh.uvt.rnu.tn/">http://pf-mh.uvt.rnu.tn/</a>	Tunisia
Uganda Scholarly Digital Library at Makerere	<a href="http://dspace.mak.ac.ug/">http://dspace.mak.ac.ug/</a>	Uganda

# The University and its Libraries: Reactions and Resistance to Scientific Publishers

Flavia BASTOS<sup>a,1</sup>, Silvana VIDOTTI<sup>a</sup> and Nanci ODDONE<sup>b</sup>

<sup>a</sup>*Universidade Estadual Paulista*

<sup>b</sup>*Universidade Federal da Bahia*

**Abstract.** This paper addresses the relationship of copyright and the right of universities on scientific production. Information and Communication Technologies (ICTs) are causing many changes in the system of scientific communication, such as the creation of Institutional Repositories that aim to gather scientific production in digital format. The University needs quicker ways of spreading academic production and many questions are emerging due to contexts such as the Open Access movement. Thus, this paper questions the positioning of Universities, especially Public Universities, which despite having policies related to intellectual property to protect the transferring forms of research results to society; many times do not have a positioning or a mechanism that regulates the self-deposit of scientific production in these Institutional Repositories. In order to develop this paper, the following issues are addressed: lack of interest of the University in storing scientific production; reports on the relationship of the library with scientific publishing houses; the participation of faculty members and students in supporting the Free Access movement; and initiatives aimed at greater flexibility of copyright to the context of scientific production. In order to follow the development of these issues at international level, it was opted for qualitative research with non-participating direct observation to carry out the identification and description of copyright policy of important publishers from the ROMEO SHERPA site; therefore, it can be observed that there are changes regarding the publishers' flexibility before self-archiving of authors in open access institutional repositories in their universities. Given this scenario, we present reflections and considerations that involve the progress and mainly the integration of the University and its faculty members; the institution should recommend and guide its faculty members not to transfer their copyrights, but to defend their right of copy to Institutional Repositories along with Publishing Houses.

**Keywords.** Institutional Repositories, Copyright, Electronic Scientific Communication, Public University, Open Access Movement

## Introduction

The university, in its mission of disseminating knowledge to the advancement of society, is inserted today in a new scenario in which the communication and information technologies provide changes in the scientific communication system, presenting a new form of storage, retrieval and dissemination of its scientific production by means of Institutional Repositories.

---

<sup>1</sup> Corresponding author: Flavia Bastos, Universidade Estadual Paulista, São Paulo, Brazil; E-mail: fmbastos@hotmail.com

In the context of scientific communication, Institutional Repositories are aimed at gathering the University scientific production in digital format in order to preserve and enhance access to such production and, consequently to reach the desired visibility and recognition.

However, the scientific production developed by faculty members and students of Universities continues to be protected by copyright law, requiring the permission/authorization from the author to deposit it in Institutional Repositories of Universities and, despite the scientific production is carried out in the field and by members of the University itself, it continues without rights regarding the storage of results of their research.

This situation demonstrates apathy of Universities by not requiring some kind of right on this production for storage in their repositories; except when patents are involved.

In this sense, this paper questions the positioning of Universities that, despite presenting policies related to intellectual property to protect the transfer forms of research results, many times do not have a positioning or mechanism that regulates the self-deposit of scientific production in these Institutional Repositories, nor a participation policy that would strengthen the process of negotiation and acquisition between libraries and publishers.

Such a proposition is needed because of some aspects to be addressed, since a significant portion of scientific production can only be accessed by purchasing the materials offered by major publishers at high prices in the context of libraries and universities. Considering that the purchased material is produced within the University and by its own faculty members and students that give their copyrights to the publishers, the other part of scientific production not published by major publishing houses often becomes gray literature due to lack of dissemination and access. Thus, once the author has the right over his creation, the University – especially the Public University – should also have the copyright of this production, strengthening the storage and representation of informational scientific content in their Institutional Repositories.

In order to develop this work, questions related to the lack of interest of the University in storing scientific production are addressed, as well as the relationship of libraries with publishing houses, the participation of faculty members and students in the Open Access movement; and also the identification of the current position of major international publishers and universities in face of open access institutional repositories to obtain subsidies to chart a future outlook on these issues before the current scenario.

## **1. The University and Its Libraries in the Context of Scientific Communication**

The change in the information scenery resized the use and definition of time/space relationship in communication networks which allowed the large flow of interactions that shaped an unprecedented multicultural interaction. Thus, there is a large network of collective knowledge that multiplies and expands itself in every exchange of information between the individuals themselves.

The reflection of this change can also be observed and compared before the traditional model of scientific communication that continues focused on knowledge generation, while the model of electronic scientific communication expands its focus to

dissemination and access, but with distinct pathways that involve the periodical subscriptions and open access.

Guédon [1] mentions the transition movement from the subscription model to the open access model through two reactions of the scientific community that are related to resistance against high prices of scientific magazine subscriptions and the emergence of consortia. According to the author, these are defensive resistance moments that do not change the context. The Budapest Declaration of 2001, which pioneered the Open Access Movement and triggered a series of integrated movements around the world to support the transition to the so-called open access paradigm, has not yet reached part of its propositions.

During these eleven years, The Open Access movement has been broadening its discussions guided by researchers through marking positioning that reinforce it; however, opposing to Open Access mandates we have been observing contrary and resistant actions against the movement by members of the United States Representative Chamber that created the bill [2] in October, 2011 called Stop Online Piracy Act (SOPA) or H. R. 3261 that enlarge the legal ways to fight against the online traffic of protected property and of falsified articles; but, due to the polemics related to the law, users began to boycott companies that did not support the law that was suspended in January 20 this year.

We highlight the protests that motivated the boycott to Elsevier [3], guided by reputable researchers that decided not to send their articles to be published at this publishing house because of Elsevier' support to the bill sent to the White House representatives in the United States in December 16, 2011, denominated as Research Work Act (RWA), also known as H. R. 3699 regarding the prohibition of open access mandates to researches financed by the federal government, with support of the Association of American Publishers (AAP).

Therefore, by means of movements and boycotts promoted by the scientific community, it shows the change in the scientific communication flux that allows greater interaction between the producer and the user of information. In this context, the University, mainly the public University, is inserted in the scope of generation and transmission of knowledge, promoting the circulation of social, technological and scientific conquers. Such advancement offers the knowledge of universal culture in various sciences and the university can fulfill its mission: teaching, research and extension contributing to the country development.

The Open Access in the context of scientific production of Universities is to Guédon [1], a powering system of the use of all minds, that is, system that extends and shares the results already achieved in our Universities, offering support and making them known by the impact of knowledge.

Even with so many differences, in the scope of universities, the Institutional Repository detains itself in some aspects that interfere with its consolidation, for example, the legal aspect that covers copyright and in particular, the right of copy for these environments.

In this context, the libraries also face difficulties not only /in terms of the high cost of subscriptions and renewals of periodicals with the publishing houses, but also in the agility to provide information to its users regarding authorizations to use the materials that are under publishers' right, which impedes the progress of researches as if publishers would lose something by releasing their contents after publication.

To illustrate how publishers are resistant to cede their information we present some case reports that occurred at "Universidade Estadual Paulista" (UNESP) and the first



one is a research group that requests to a certain publisher authorization to translate to Portuguese an specific article published in one of its journals, to which the university pays subscription so that, via library, information is made available to its users.

By making contact with the manager of the publisher headquarters, three questions related to the use of the article were posed: What purpose are you translating the article for, i.e. is it for a new book or dissertation? Which language are you intending to translate it into? What rights would you require; i.e. Print Regional (Brazil only), Print World, Electronic World or Print & Electronic World?

The publisher said that it would be happy for you to translate this article; however, it would incur a permission fee. For the right to translate the article into Portuguese (Electronic World), the fee would be £ 210.00 GBP + Tax, if applicable. If you wish to proceed, please let me know and I will raise the appropriate invoice.

As mentioned previously, the university has a subscription with the publisher and for this reason it was asked for a discount or even exemption from that tax to the representatives of Brazil office because according to the conversion the value would be around R\$ 530,00. However, according to Brazil officials, the headquarter requested further clarification on the possible translation which, in accordance with the considerations of the research group leader would be of great importance aiming the Portuguese publication of the concepts pointed by the author at the time of presenting ideas In the Information Science area and especially, because the author himself would be present in the main event of the area in the same year. Thus, the researcher sent new answers pointing the characteristics of magazines in the area that would receive the translation, for example, the kind of publication, language (original + translation or only translation), title of the journal, if the journal is academic or commercial, if it will be a printed or online version, with public or restrict access, how many readers the journal has.

Only in October, the eve of the event, the researchers received the answer from the publisher that after looking through the journals website (which took a while as citrix was being its usual useless self), then there are 2 options:

1. The article can be translated and given to the 20 or so researchers who will be attending the lecture free of charge providing that the following statement appears on each copy. This article is © Group Publishing Limited and is provided for your own personal use only. It may not be used for resale, reprinting, systematic distribution, emailing, web hosting, including institutional repositories/archives or for any other commercial purpose without the permission of the publisher, this is because it is only a small number of people.

2. If they still wish to proceed with the translation for the journal, then because it is a registered journal and it is open access and electronic (and therefore a potentially wide audience) then the translation fee of £ 155 (+ tax if applicable) for the electronic rights will have to stand.

Because of the importance and necessity of translating the article to the research group, the requested amount by the publisher was paid, showing that in this case it makes no difference to be a subscriber or not to any publisher's product, or even the value of the account that the institution keeps/funds annually, and even less the use for academic purposes, since what really becomes important now is the fact of the publisher not miss any opportunity to do business for profit.

When publishers began marketing electronic periodicals, many libraries considered that somehow they could optimize the university's budget; the units that had the same course could adopt the electronic subscription rather than pay for each printed

subscription of the same title. At the time, the price was much lower compared to the value of the electronic subscription.

Today, publishers, in order not to lose their profit margin, bind or condition the lower value of the electronic periodical only if the institution subscribe to the printed form.

UNESP adopted the policy of not subscribing to printed periodicals, but unfortunately this option is not always available to Latin America, and if the title is part of a basic nucleus of the university periodicals we once more acquire what is offered by the representative in this country, even knowing that by consulting the publisher's website there is the information that the title exists in both electronic format and in printed form.

Another situation that causes certain indignation is to realize that Brazilian institutions still hold in their collection printed and electronic formats increasing three times the institution budget; however, by communicating the publisher about the discontinuance of printed titles, the value of the electronic title increases absurdly, making the agreement to purchase printed titles remain.

All these cases demonstrate how librarians need to be attentive not only regarding management issues and services offered to users, but also in relation to commercial practices involving the complex relationship with scientific publishers.

## **2. The Copyright and the Right of Universities on the Scientific Production**

The scientific production can be understood as an invaluable resource to promote the development of science. The generating source of this production is the University which has the important mission of disseminating knowledge, providing not only access to information, but also transforming it into service to the community. Thus, access to scientific production serves the social interests and needs, becoming a right of society which is funding the Universities. Therefore, the Universities, especially Public Universities, have an important role in the use of knowledge for the social, political, economic and technological development of the country.

Regarding this issue, Law [4] states that Universities have no other purpose except the creation, dissemination, understanding and knowledge development, and it is inevitable that the management of intellectual property be an area of growing concern.

The current legislation that regulates the intellectual property, in turn, points out deficiencies based on a monopoly system that privileges some for others, something that has been implicitly incorporated by society over the years and that can be observed in the reflection of Mark Rose [5]:

The institution of copyright is of course deeply rooted in our economic system, and much of our economy depends, in turn, on intellectual property. But, not less important, the copyrights are deeply rooted in our conception of us as individuals... and it is associated with our sense of privacy and our belief, at least in theory, that it is essential to limit the power of magnificence. I think that we are not prepared to abandon the notion of whom we are.

Complementing the author's reflection, in fact we are not prepared to abandon the notion of who we are. The feeling of privacy, even utopian, is still necessary. Moreover, the laws of copyright protection, emerged in the early eighteen century, are

no longer appropriate to the context of publications and reproductions in the scenery of digital networks of collaboration, which requires open communication among people.

According to Hughes [6], the academic nature is supported by open communication, on the assumption that teaching and research are processes that require sharing for the construction of new knowledge. In addition, one of the hallmarks of the scientific production can be verified at the moment its value increases from the growing number of accesses and from the addition of new ideas for interpretation. The author also points that researchers need to keep the possibility of depositing their work in repositories even when they give up their rights to publishers.

The contractual issue may be considered really important to establish the rules to govern rights and duties of the author, publisher and the University itself regarding copyright. However, the lack of knowledge about the contractual content demonstrates the weakness of the author in relation to his effective right on production.

However, the lack of interest of Universities regarding the domain of scientific production copyright is related to the custom and practice; the academic freedom and the lack of financial gain.

Due to technological advances, today there have appeared new ways to improve the management of copyrights created at research Universities, such as the Open Access movement, which claims a greater flexibility in copyright for the Universities, considering that the authors should not give the rights to publishers; and publishers should not have the power to sell to Universities access to materials protected by copyright at very high prices.

Some high education institutions, aware of this issue, are now proposing a license model for academic use only with the publisher, which demonstrates that this concern starts to affect and influence people involved in this process that continues to be motivated by the desire of establishing Institutional Repositories of electronic publications. However, even with the increased number of proposals submitted to Universities, they still take time to manifest, showing a total lack of concern with the issue.

To Fujino [7], Law [4] and Oppenheim [8], this issue has been addressed in academic research as one of its features because the very freedom of research and the change in performance policy could be seen as a form of censorship, control and ownership of copyright by the University, including when and where they could publish.

Thus, to claim copyright on research publications in the context of Universities becomes something quite complex: since this possibility can interfere with the “academic freedom” or because of the fact of not achieving any return on own rights; or because Universities acquire copyright on these publications and they are forced to license them to third parts.

The researchers are so interested in having their papers approved by publishers that they do not care about the implications of the signed contracts, and in many cases, what happens is an ownership transfer of all copyrights including the digital right to the publisher.

Therefore, Universities bear the costs of acquiring the work of its own faculty members in the form of subscriptions at very high prices that are traded through secret contracts between publishers and, most of the times, with libraries; and for electronic publications, for example, it is paid a kind of lease with restricted access that is tightly controlled by the owner copyright.

In this case, so that Universities effectively have possession of these informational contents, it is required the backfiles purchase, that is, the archiving purchase of retrospective content published by certain publishers.

The following is a critical reflection on the attitude of Universities regarding the storage and preservation of their own scientific production in institutional repositories.

### **3. The Integration Between the University and the Researcher: The Search for Flexible Policies and Initiatives for the Storage of Scientific Production in Institutional Repositories**

Currently, discussions on the short and long term effective benefits of the creation and maintenance of Institutional Repositories in universities, stimulated by the Open Access movement, are usual. They are often led by faculty members with entrepreneurial profile, in a context where the departments have a more administrative and trading posture. In this sense, the university could be a support so that the researcher can extend the recognition and visibility of his production from the moment when the university itself manages, stores and preserves its production.

In contrast, there are the faculty members with a more conservative profile that associate the quality of their scientific production with the subject of their discipline, which makes these faculty members feel as the copyright owners of journal articles and books that they wrote, prior to publication.

In this context, if universities continue to refuse some kind of manifestation in favor of the storage and preservation of their scientific production in institutional repositories, only faculty members will have the ability and decision power to the self-archiving in these environments. And the changes regarding copyright and obligation to deposit in Institutional Repositories should be implanted so that the Universities fulfill their missions and the knowledge circulation is freer and more democratic.

The work of the Zwolle Group [9], bringing together many researchers, was a very important initiative not only for the redefinition of policies and procedures, but also for the formulation of new systems in various parts of the world that are rapidly integrating their work with established channels of scientific communication.

Another important initiative was the Romeo project (Rights Metadata for Open Archiving) founded by the Joint Information System Committee in order to detail the issues related to transfer of copyright to publishers of scientific periodicals and to the self-archiving of researches by the community of United Kingdom in repositories based on the Open Access; currently, the continuity of this project happens through the SHERPA/ROMEO basis that presents the policies of publishers in various knowledge areas in relation to self-archiving.

In Brazil, the motivation for the use and adoption of institutional repositories can be followed by an initiative of Instituto Brasileiro de Informação, Ciência e Tecnologia (IBICT), that aiming to promote the construction and implementation of institutional repositories, has been offering a larger visibility of scientific information produced in Brazil through the use of Dspace software.

According to the IBICT website [10], there are 66 registered repositories benefited by this initiative; among them, 44 belong to national universities and regarding the policies established by these universities, especially public universities, the user is released from deposit of objects that have contractual restrictions of copyright; thus,

demonstrating how universities actually fit the policy established by the publishing industry.

Another initiative worth mentioning is Books SciELO [11] that is aimed at online publishing of academic books in order to maximize visibility, accessibility, use and impact of researches, essays and studies they publish. The books published by SciELO are digital texts formatted according to international patters that allow the control of access and citations and that are readable in eBooks, tablets, smartphones and computer screens. This program is financed by a consortium formed by the publishers of Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Universidade Federal da Bahia (UFBA) and Fundação Oswaldo Cruz (FIOCRUZ).

In order of identify the degree of acceptance of each publisher in relation to self-archiving, colors are used for such identification, according to Figure 1. Thus, we have green for publishers that accept to archive pre-print and post-print. With this permission, there are publishers such as Springer Verlag, Biomed Central, Harvard University/Harvard Law School, Kluwer, IP Publishing, Cambridge University Press; the blue color for editors that accept to archive post-print. With this permission, publishers such as Thompson Reuters, Rockefeller University Press; the yellow color for publishers that accept to archive pre-print. Publishers such as Oxford University Press, Blackwell Publishing, Wiley-Blackwell, Taylor & Francis and Nature Publishing Group have this permission; and the white color for not supporting the filing. American Medical Association has this permission.

### RoMEO colours

We have used different colours to help highlight publisher's archiving policies. These colours are a development from the original RoMEO project list, and differentiate between four categories of archiving rights:

ROME colour	Archiving policy
green	can archive pre-print and post-print or publisher's version/PDF
blue	can archive post-print (ie final draft post-refereeing) or publisher's version/PDF
yellow	can archive pre-print (ie pre-refereeing)
white	archiving not formally supported

Each publisher's entry is coded according to one of these colour categories.

The entry for each publisher also lists conditions or restrictions imposed by the publisher which govern archiving rights or activities. *Conditions* are taken as terms which can be easily accommodated and which do not hinder an author in archiving their work. A typical condition is to acknowledge the publisher's copyright in the work. *Restrictions* are more prohibitive, typically requiring some additional action on behalf of the author. Where a *Restriction* effectively blocks access to the eprint, such as in the case of an embargo on its public release, or requiring password-controlled access, then the partial archiving right is noted but the full colour categorisation does not apply.

Sometimes open access discussions talk about "gold" publishers. This is a later development independent of RoMEO categories, and is used to describe publishers of open access journals. For the purposes of archiving, all open access journals allow archiving and can be taken as RoMEO "green".

Some of the larger publishers have different archiving rights for different journals. This is particularly the case where they publish learned society journals on behalf of the society. A learned society might insist on a more liberal, or more restrictive archiving policy than the general publisher's copyright agreement allows. The RoMEO colour coding relates to the overall permissions given by a publisher. For example, a publisher has to apply the "green" archiving rights across all of their journals for their code to be "green".

**Figure 1.** RoMEO project colours

By analyzing the copyright policy of the publishers in this site, it is observed both the permission for unrestricted deposit of payment or seizure and the permission of publishers to deposit the PDF version or of the article itself in an Institutional

Repository after permission obtained by the publisher; or after payment and seizure period or even, publishers with payment options for open access, considering that these policies do not always permit the deposit in open access repositories.

The information contained in the website demonstrates how some publishers are becoming more flexible in relation to the transfer of copyright to institutional repositories.

#### **4. Final Considerations**

Influences arising from economic, political issues and largely determined by technological changes have interfered in our attitudes regarding access to scientific information. Thus, on one hand, there is the scientific publication divided between the commercial publishers and involving the paid access of subscriptions, the authors, libraries and their universities, besides all the people that have the right to scientific information produced within the university. On the other hand, there are the scientific societies and groups of scientists that cast a growing number of scientific initiatives that provide open access to knowledge.

Given this profile, we still have commercial publishers dominating the flow of scientific information, making it clear that we are still denying equal opportunities of scientific research access to several countries with lower purchasing power, because this is the scientific publication aspect that still moves the world publisher system with very high charges. Today, these costs often need to be shared among consortia of libraries to provide access to updated scientific production.

By submitting scientific articles to paid access magazines we contribute to a knowledge distribution system in which Universities and Research Institutes are required to buy back at a very high price something that many times was produced by their own faculty members or students.

Today, the investment made by universities and institutions that finance researches can be considered negligible compared to the return earned by the financial investment from commercial publishers; therefore, only changes in the current system, in which the academic community is inserted, will make this movement gain strength so that a greater number of universities would join the open access movement by means of incentives geared to their institutional repositories.

Another way to join the academic community may be by the increase of access possibilities as it is possible to deposit in more than one repository; through search of repositories that facilitate survey time and also offer the possibility of being linked to other providers, increasing access, unlike the access achieved only by the publisher page; some repositories also offer value-added services providing links to publishers' pages.

However, despite the number of proposals submitted to the Universities, individually and associatively, they have been slow to respond. To integrate the University and the faculty members, the institution should recommend and guide them not to surrender their copyright, but to struggle for the right to copy for Institutional repositories with publishing houses.

The University, in turn, should claim the right to copy and not the copyright on scientific production. It should also require flexible counterparts for nonprofit use of periodicals and database aiming not only the development of the global educative community, but also stressing the future importance of science in the world.

From the moment we have the diversity of forms and technical/practical resources combined with the involvement of the academic community toward the desired changes, we will find the most appropriate way to global access to scientific information.

## References

- [1] J. C. Guédon, *Toward Optimizing the Distributed Intelligence of Scientists: the Need for Open Access. Seminário Internacional de Bibliotecas Digitais, 2.*, 2004, Campinas: UNICAMP, 2004.
- [2] Stop Online Piracy act. Wikipédia: The Free Encyclopedia. [Internet]. [date unknown] [cited March 05 2012]. Available from: [http://en.wikipedia.org/wiki/Stop\\_Online\\_Piracy\\_Act](http://en.wikipedia.org/wiki/Stop_Online_Piracy_Act)
- [3] R. Poynder, Elsevier Steps Away from Research Works Act. Open and Shut. Reino Unido, 17 Feb. [Internet]. 2012. [cited March 05 2012]. Available from: <http://poynder.blogspot.com.br/2012/02/elsevier-steps-away-from-research-works.html>
- [4] D. G. Law, R.L. Weedon, M. R. Sheen, Universities and article copyright. *Learned Publishing*. 2000; 13(3): 141-150.
- [5] M. Rose, *Authors and Owners: the Invention of Copyright*. Cambridge: Harvard University Press, 1993.
- [6] C. A. Hughes, The Case for Scholars' Management of Author Rights. *Portal: Libraries and the Academy*. [Internet]. 2006. [cited Jan 08 2012]. 6(2). Available from: [muse.jhu.edu/journals/portal\\_libraries\\_and.../6.2hughes.html](http://muse.jhu.edu/journals/portal_libraries_and.../6.2hughes.html)
- [7] A. Fujino, E. Stal, G. A. Plonski, A proteção do conhecimento na universidade. *Revista de Administração*. 1999; 34(4): 46-55.
- [8] C. Oppenheim, Is copyright strangling higher education? *UKOLUG Newsletter*; 2001; 12(5): 26-29.
- [9] Surf. Copyright Management Scholarship. *Zwolle Principles*. Amsterdam. [Internet]. 2002. [cited Feb 10 2010]. Available from: <http://www.surf.nl/copyright/keyissues/scholarlycommunication/principles.php#Principles>
- [10] Repositórios Brasileiros IBICT. [Internet]. [cited March 05 2012]. Available from: [http://dspace.ibict.br/index.php?option=com\\_content&task=view&id=28&Itemid=92](http://dspace.ibict.br/index.php?option=com_content&task=view&id=28&Itemid=92)
- [11] Reunião da Rede SciELO Livros, 1., 2012. [cited March 05 2012]. São Paulo. [Internet]. Available from: <http://eventos.scielo.org/scielolivros>

# Information Retrieval Educational Goals in Library and Information Science and in Health Sciences

Anabela SERRANO <sup>a,1</sup>

<sup>a</sup>*Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro*

**Abstract.** This paper is intended to compare information retrieval (IR) educational goals in different academic and professional areas such as Library and Information Science (LIS) and Health Sciences (HS), analysing and identifying a shift on user seeking goals in the digital era and, eventually, on educational goals as well. It starts with a section on information literacy where several aspects are specified, such as user goals, IR systems, IR skills, information seeking strategies (queries) and user perception of search success. Another section focuses on teaching IR aspects, like IR educational goals, assessment and feedback, and e-resources in LIS and in HS. Teaching in an academic environment for academic audiences is somehow different from teaching for professional audiences even though these are located in an academic environment as well. Those are the issues and particularities that throughout the analysis of information literacy and teaching IR aspects will be explained along the full paper.

**Keywords.** Information retrieval, Educational goals, Library and Information Science, Health Sciences

## Introduction

Information retrieval is a core teaching area in Information Science and is also taught in Computer Science and in Informatics Engineering. It is intimately related with the professional area of Information. However, in an information and knowledge society, information retrieval skills are increasingly needed by professionals and academics in many other areas and by every citizen in his daily life, including professional and civic obligations, to satisfy their multiple and frequent information needs.

This paper is structured in two sections. The first section is on information literacy, specifying several aspects focused in IR: user goals, IR systems, IR skills, information seeking strategies (queries) and user perception of search success. The second section focuses on teaching IR: IR educational goals, assessment and feedback, and e-resources in LIS and in HS.

Teaching in an academic environment for academic audiences is somehow different from teaching for professional audiences even though these are located in an

---

<sup>1</sup> Corresponding author: Anabela Serrano, Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA), Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal; E-mail: [anabelaserrano@ua.pt](mailto:anabelaserrano@ua.pt), [anabelaserrano@hotmail.com](mailto:anabelaserrano@hotmail.com)



academic environment as well, so a distinction between education courses and training courses is established whenever it is needed along the full paper.

## 1. Information Literacy

### 1.1. User Goals

Everyone produces and consumes information. Nowadays, we consume it on a daily basis and in an increasing volume. We use information all the time – not only the one we produce and the one we receive, but also the one we look for. Looking for, i.e. searching, information has become a routine in a global society wirelessly connected. IR techniques, formerly restricted to computer and information scientists (and some others as well) and to information professionals, are coming out to the streets popping up in several gadgets we use every day: cell phones, PDA/smartphones, netbooks, tablet PC, laptops, e-readers, etc. Search engines like Google made information look really accessible – from something that we all have the right to access it turned into the core of our lives and of our society. Vannevar Bush's [1] vision has become reality. We are addicted to information whether it comes through the internet (email, news, webpages, blogs, etc.), cell phones or broadcast companies, and we love to interact communicating and even publishing information all the time. We suffer from information overload, so we need desperately to know how to retrieve the information we need, and only that one – relevant and pertinent information according to our needs. However, fewer of us know how to achieve success in search results. [2] [3]

IR skills are essential in several professional areas like library and information science (LIS), computer science (CS), health sciences (HS), business, criminal investigation, etc. IR specialists are educated in degrees in LIS and CS, but users that work in health – physicians, nurses, etc. – have urgent information needs which satisfaction can have an enormous impact in our society. In 1992 Marshall published the results of the Rochester Study [4], developed at the Faculty of Library and Information Science (University of Toronto, Ontario, Canada) in the *Bulletin of the Medical Library Association* (continued by the *Journal of the Medical Library Association*). From September 1990 to March 1991, hospital librarians of the fifteen participating hospitals, at the geographic area of Rochester, NY, have distributed a survey to 448 physicians and received 208 answers. The results are impressive. The conclusion is that physicians consider that the hospital library has a significant impact on clinical decision making, which is clearly reinforced by statistical data presented. Revealing a few examples may illustrate the real impact that information services and systems have in life in general and even in hospital mortality rates: 80% of the respondents consider that accessing library services caused change somewhere in the clinical decision process, 51% changed the selection of tests, 45% changed drug therapy, 19% avoided patient mortality...

Searching for information is a time consuming task and it can be confusing or less motivating when you don't know how to do it. IR specialists are important pieces in the information society as professionals and as trainers of professionals from several other areas such as health sciences. Evidence based medicine has pushed physicians, since a long time ago, to invest in life-long learning. And it is not by chance that MEDLINE is one of the oldest bibliographic databases. On what concerns HS professionals, this

chapter will focus on physicians' profile as information users and as IR trainees. Due to length restraints, nurses and allied health professionals profiles will not be explored.

### *1.2. IR Systems and Skills and Information Seeking Strategies – user Knowledge and Use*

Education and training somehow differ on what concerns its focus. Education is focused in theory and training is centered on practice [5]. LIS students learn IR through education programs, while physicians acquire IR skills through training programs usually developed for academic or hospital library users.

Teaching models are also different for education and for training courses. Curricular development should be adapted to the students/users learning expectations. User studies about information behavior and surveys about students' impressions on the information seeking process, before enrolment in an IR course, may help teachers and trainers to develop education or training programs.

Education courses are centered at teaching theoretical concepts: IR fundamentals like models and systems, searching techniques/features, etc. Otherwise, training courses are centered at teaching practical applications, like searching in specific systems without comparing the IR systems explored. In this teaching approach, trainees never learn that the searching technique showed at system A is also present at system B under a different symbol/command. However this approach may be useful to professionals overwhelmed by their work duties, like physicians, that don't have the time to learn IR fundamentals and that rush onto information resources to extract pertinent information to support their clinical practice. Mixing both approaches wisely may have encouraging results.

Giving students or trainees a perspective, deep or brief according to their course profile, on IR fundamentals is always a good idea. It makes them aware of all possible choices and they become more autonomous when discovering an IR system. Modeling the IR process may be a good start for a LIS course. Even, though, without using UML, analyzing the process itself as they know it by the time they enter the course could be an interesting challenge and a start point to explore IR history since Bush's vision with the help of Michael Lesk seven ages, for example. [6] Taking a wider view of the entire information access process could help them to understand the importance of IR and its connection with information analysis tasks (modeling information systems, indexing and abstracting, authority control, thesaurus construction) and also with information storage issues. Knowledge of information systems modeling and of mathematics [7] would be important at this stage. LIS students not always have those skills, but it should be considered by curriculum developers when scratching a new LIS course or reformulating an existing one. For further information see section 2 of this chapter.

IR models should be explained and explored from the three classic models, never forgetting the alternative ones, to the structured models. Browsing models should also be mentioned. [8] Query and results are two of the concepts that students better perceive. Information needs is the concept that they less remember. [9] Teaching them search syntax is rather important so they can construct quality search strategies and queries that can extract relevant results from IR systems.

### 1.3. User Perception of Search Success

User perception of search success is one of the most apprehensive problems in IR. Sometimes, due to ignorance of information resources dimension and of search techniques/features it is usual that users don't perceive if they aren't being well succeed in a search when analyzing results. Information exponential growth adds an help to this problem. Even efficient search engines like Google can't assure that a certain query was completely answered. Knowing when to stop may be a harsh task. And time limits are not always the best choice. Even in HS, older information may be just as valuable as recent one. For example, some diseases may have been studied or some drugs may have been tested a long time ago and have never been looked up in recent decades. Probably, that's why MEDLINE has launched information back to 1947. [10] However, most HS users think they have success when performing a search in a IR system. Perhaps, this is motivated by the quality of the information retrieved (usually from MEDLINE/PubMed) and by its impact on clinical practice as it was already referred above.

## 2. Teaching IR

### 2.1. Educational Goals

As mentioned above, teacher educational goals should, somehow, match student learning expectations or, at least, the latter should be considered while developing curricular programs.

Recently, in Europe, during the "Bologna Process" of adequacy of higher education degrees, several countries have done a full revision of their courses curricular plan and have also created other courses in the spirit of the Bologna Declaration signed by 29 European countries on 19<sup>th</sup> June 1999. The adequacy of higher education degrees has implied, in many cases, a change in the selection of the various possible approaches, depending on the scientific-technical domain and on the essence of each curricular unit (CU), to the teaching/learning process. The spotlight that some courses placed on knowledge transmission before Bologna was now turning the to skills development through an active participation of the student inside the teaching/learning process itself. In LIS, the subject area of IR constitutes an exciting challenge in applying the recommendations of Bologna either by belonging to the "core" of LIS either by being an area where there has always been a great research activity.

In Portugal the "Bologna Process" is now concluded. For example, in 2006, at the Department of Information Science of the largest Portuguese polytechnic institute (and one of the best ranked, at a national level, on some world university rankings) – Instituto Politécnico do Porto (IPP) –, the degree in Information and Documentation Sciences and Technologies with the duration of 5 years was restructured to a degree with a smaller length of time of 3 years. [11] The degree adequacy was supported by two important works in LIS: *Euroguide LIS* and *European Curriculum Reflections on Library and Information Science*. [12] [13] [14] In order to, somewhat, cover Information Seeking and Information Retrieval (IS&R) area skills, two courses were created: Information Behavior and Information Retrieval. Student skills were identified according to the *Euroguide in LIS*. For the IR course the following were defined:

Generic skills:

- Understanding and defining information retrieval fundamentals;
- Developing analysis, evaluation and diagnosis skills;
- Applying acquired knowledge on new occasions in order to solve professional problems;
- Implementing projects;

Specific skills:

- Analyzing complex information retrieval queries, constructing search strategies and outputting search results;
- Identifying and selecting information sources;
- Being at ease with every search feature/technique;
- Evaluating IR systems.

The underlying educational goals of this definition, and of the contents defined at the curricular plan, were the knowledge of IR fundamentals and of IR systems, and training in search strategies, just like other authors stated recently. [5] The teaching program that has been carried out, since 2006, puts its accent on seven topics that were found to better suit the educational goals above mentioned:

- Introduction to IR – the IR process
- IR systems
- IR models
- Query operations
- IR on the World Wide Web
- User interfaces and visualization
- Information retrieval systems evaluation

These topics may be developed in many ways and in many levels of depth according to each teaching context. A tutorial style of delivery will also meet the principles stated on Bologna process.

## *2.2. Assessment and Feedback*

Assessment is, to some extent, linked with the teaching methods adopted and with the teaching system selected. In the IR area, an active teaching method is, perhaps, the more obvious option. Mixing multiple teaching methods and combining them in several ways, depending on how each class responds, can create interactivity in very interesting ways inside the classroom. When having an interactive and practice teaching approach, assessment will probably have its focus on practical work - classroom or e-learning exercises, assignments, practice part on final examination etc. All these pretend to recreate real professional life problems to solve inside the teaching/learning environment whether the teaching system selected is the traditional one, e-learning or b-learning.

Developing skills implies that students/trainees have an active role and that teachers/trainers promote interactivity. Communication leads us to interaction, as Shannon showed on his mathematical info-communication theory: when transmitting information in a message, through a transmission medium, the abstract communication channel between sender and receiver allows feedback.

Searching/retrieving information implies, above all, critical thinking. To have success while solving an exercise (what implies the synchronous use of almost every skills above referred to the CU of IR at IPP) a student/trainer shall:

- be able to analyze and translate the content of an information need,
- know several information resources, or know how to access them, and know how to evaluate which ones are more suitable to each case,
- deeply know searching techniques/features that can be used in an IR system,
- know how to construct a query creating quickly several alternatives,
- know how to evaluate results presented by the IR system and how to create a new search strategy if needed.

Having strong conceptual basis in IR, technological aptitude, abstracting capacity, cognitive processing speed, critical thinking and evaluation ability can enable a good performance in IR either at the academic environment either at the professional environment.

Traditional lectures based, exclusively, on a passive teaching method like formal lecture without interaction with the students are less likely to promote skills development or to stimulate reasoning. A mix of several methods like group discussion, forum, interactive lecture supported by audiovisual materials, brainstorming techniques, among others, can easily promote interactivity at the classroom. Mixing traditional teaching system with e-learning, i.e., adopting blended learning or b-learning teaching system can also enable a dynamic and interactive environment accessible all the time.

On digital platforms, communication may be synchronous, just like in the classroom (physical or virtual), or asynchronous. Examples of synchronous communication are Messenger, ICQ, Google Talk, Skype, Second Life (allowing virtual classroom or virtual workplace), etc. Examples of asynchronous communication are email, forums, blogs, Google Sites, etc. [15] Using e-learning systems demands the analysis of multiple aspects that should be considered like type of content according to the last update and inner typology. Stable content, like IR fundamentals, that doesn't need to be often updated is suitable to be delivered in more static formats, while unsettled content, like research topics, that needs to be frequently updated should be delivered in more dynamic formats [16].

Stimulating students and placing challenges to them in order to obtain their adherence, either of those considered "deep" ones either of those considered "superficial" ones, is an important asset. A challenge may be placed by the promotion of competitions where winners get a prize as a bonus reflected at the final classification of the student at the course. Usually, students adhere with enthusiasm to this kind of challenge. Testing velocity when solving problems in IR with a chronometer (at an advanced stage of the course) might improve results in examination and, consequently, final classification at the course. This kind of competition proves to increase levels of confidence and to decrease stress at final examination. At the IR course mentioned above as an example this was confirmed, since 2006, by positive approval rates of the students, positive feedback from students when evaluating the course in pedagogical annual blinded surveys (rated with 4,60 in a 0 to 5 scale, with some aspects rated with 4,83), and by good performance on curricular professional practice in the IR area. Feedback from students can also be taken through specific surveys conducted by the CU teacher in order to continuously improve the teaching and learning process in IR. In spite of the positive results achieved at the example being analyzed, this kind of measure will begin to be applied in that course in 2010. The survey results, depending

on the questions asked, may lighten up new directions or changes wherever and whenever needed.

In training courses it is usual to ask trainees to give feedback of the course by answering a questionnaire at the end of the course. Usually the duration of these courses is shorter, so the questionnaire will have some other kind of questions like possible professional benefits after taking the course. HS trainees will be interested to evaluate the impact of the course on their performance in clinical practice, but this cannot be well known until a period of time after the course. User satisfaction can also be assessed as well as searching utility to clinical questions, for example. MEDLINE and EMBASE lead physicians preferences, being the former really detached from all other information sources [17].

### 2.3. e-Resources in LIS and in HS

e-Resources in LIS (selected list):

- IFLA – International Federation of Library Associations  
<http://www.ifla.org/>
- Internet Library for Librarians – portal since 1994  
<http://www.itcompany.com/inforetriever/>
- Librarian's Index to the Internet – Since 1995  
<http://www.ipl.org/index.html>
- Library of Congress (LC) portal to librarians  
<http://www.loc.gov/library/>
- LISA – Library and Information Science Abstracts  
<http://www.csa.com/factsheets/lisa-set-c.php>
- LISTA  
<http://www.libraryresearch.com/>
- Medical Library Association (MLA)  
<http://www.mlanet.org/>
- National Library of Medicine (NLM)  
<http://www.nlm.nih.gov/>
- ODLIS – Online Dictionary for Library and Information Science  
<http://lu.com/odlis/index.cfm>

e-Resources in HS (selected list):

- Arizona Health Sciences Library – e-Journal Locator  
[http://zp9vv3zm2k.search.serialssolutions.com/?V=1.0&L=ZP9VV3ZM2K&N=100&S=T\\_AZ&C=A](http://zp9vv3zm2k.search.serialssolutions.com/?V=1.0&L=ZP9VV3ZM2K&N=100&S=T_AZ&C=A)
- BioPortal – Ontologies used in biomedical communities  
<http://bioportal.bioontology.org/>
- CiNAHL  
<http://www.ebscohost.com/cinahl/>
- EMBASE - Excerpta Medica Database  
<http://www.embase.com/>
- Environmental Health & Toxicology - portal  
<http://sis.nlm.nih.gov/enviro.html>
- Intute / Medicine – UK universities portal  
<http://www.intute.ac.uk/medicine/>

- Intute / Nursing, midwifery and allied Health - UK universities portal  
<http://www.intute.ac.uk/nmah/>
- MEDLINE/PubMed  
<http://www.ncbi.nlm.nih.gov/pubmed>
- Merriam-Webster's Medical Dictionary  
<http://www2.merriam-webster.com/cgi-bin/mwmedsamp?book=Medical&va=sample>
- National Centers for Biomedical Computing (NCBC)  
<http://www.ncbcs.org/>
- National Institute of Health (NIH)  
<http://www.nih.gov/>
- PDQ – Physician Data Query  
USA National Cancer Institute (NCI) comprehensive cancer database.  
<http://www.cancer.gov/cancertopics/pdq/cancerdatabase>
- PLoS – Public Library of Science  
<http://www.plos.org/>
- PLoS Medicine – open-access journal published by PLoS  
<http://www.plosmedicine.org/home.action>
- TOXNET – Toxicology Data Network  
<http://toxnet.nlm.nih.gov/>

## References

- [1] V. Bush, As we may think, *Atlantic Monthly* **July** (1945), 47-61.
- [2] A.P.S. Serrano, A biblioteca digital ou o acesso global, *Páginas a&b: arquivos & bibliotecas* **6** (2001), 21-39.
- [3] A. Serrano, A avaliação de serviços de informação do ensino superior, *Páginas a&b: arquivos e bibliotecas* **12** (2003), 7-65.
- [4] J.G. Marshall, The impact of the hospital library on clinical decision making: the Rochester study, *Bull Med Libr Assoc* **80(2)** (1992), 169-178.
- [5] J.M. Fernández-Luna, J.F. Huete, A. MacFarlane et al, Teaching and learning in information retrieval, *Information Retrieval* (2009), doi 10.1007/s10791-009-9089-9
- [6] M. Lesk, The seven ages of information retrieval, IFLA, 1996.  
<http://www.ifla.org/VI/5/op/udtop5/udt-op5.pdf>. Accessed 29 March 2010
- [7] A. MacFarlane, Teaching mathematics for search using a tutorial style of delivery, *Information Retrieval* (2009), doi 10.1007/s10791-008-9078-4
- [8] R. Baeza-Yates, B. Ribeiro-Neto, *Modern information retrieval*, ACM Press, New York, 1999.
- [9] E.N. Efthimiadis, D.G. Hendry, Search engines and how students think they work. In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2005, doi 10.1145/1076034.1076145
- [10] A. Serrano, Objectivos educacionais da Recuperação da Informação em Ciências da Saúde, In: *A medicina na era da informação: Atas do Colóquio Internacional MEDINFOR II*, Porto, 2011 [em publicação].  
[http://medinfor.med.up.pt/images/resumos/Finais/serrano\\_2011\\_objectivoseducacionais\\_ri\\_cienciaassaude\\_cready.pdf](http://medinfor.med.up.pt/images/resumos/Finais/serrano_2011_objectivoseducacionais_ri_cienciaassaude_cready.pdf). Accessed 26 March 2012
- [11] A.P.S. Serrano, O b-learning no desenvolvimento de competências em Recuperação da Informação, In: M.M. Borges, E. Sanz Casado (eds) *A ciência da informação criadora de conhecimento: actas do IV Encontro Ibérico EDIBCIC*, Imprensa da Universidade de Coimbra, Coimbra, 2009, vol. 1, 319-331.

- [12] European Council of Information Associations, *Euroguide LIS*: Vol. 1: competencies and aptitudes for European information professionals. ADBS, 2004. <http://www.certidoc.net/en/euref1-english.pdf>. Accessed 29 March 2010
- [13] European Council of Information Associations, *Euroguide LIS*: Vol. 2: levels of qualification for European information professionals. ADBS, 2004. <http://www.certidoc.net/en/euref2-english.pdf>. Accessed 29 March 2010
- [14] J. Bates, D. Bawden, I. Cordeiro et al, Information seeking and research, In: Kajberg L, Lørring L (eds) *European Curriculum Reflections on Library and Information Science Education*. The Royal School of Library and Information Science, Copenhagen, 2005.
- [15] G.J.F. Jones, Teaching information retrieval using research questions to encourage creativity and assess understanding. In: McFarlane A, Fernandez Luna JM, Ounis I et al (eds.) *Proceedings of the First International Conference on Teaching and Learning Information Retrieval*. London, 2007.
- [16] A. Henrich, S. Sieber, Blended learning and pure e-learning concepts for information retrieval: experiences and future directions, *Information Retrieval* (2009). doi 10.1007/s10791-008-9079-3
- [17] W.R. Hersh, D.H. Hickam, How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA* (1998). doi:10.1001/jama.280.15.1347



# Evolving Definitions of Authorship in Ebook Design

Celeste MARTIN<sup>1</sup> and Jonathan AITKEN  
*Emily Carr University of Art + Design, Canada*

**Abstract.** The development of ebooks for tablet devices offers a rich space for collaboration between writers and designers. This paper examines how this emerging media affects ideas around collaboration and authorship. Specifically it considers the changing role of designers in shaping meaning and content and how this may affect existing paradigms of authorship. Using class-based projects as case studies, the paper presents and discusses examples of how designers have shifted their role from historic notions of “crystal goblet” design or expressive design to genuine collaborators and authors.

**Keywords.** enhanced ebook, authorship, designers as authors, collaboration, co-authorship, digital book applications

## Introduction

First generation ebooks are dominated by Kindle, Sony, and the iPad’s iBooks reader. As in many early adaptations of technology, old patterns have been mapped onto new media. In this case, ebooks function largely like traditional print—the concept of discrete pages, a linear narrative, a static interface—all endure. In some cases, one can even turn pages with the flip of a finger, accompanied by a “swish” sound effect!

Second generation technology allows for a rethinking of an application from the ground up. In this instance, the new tablet platform (eg Apple’s iPad, Blackberry’s PlayBook, Motorola’s Xoom) offers an almost dizzying array of possibilities. Full use of a tablet for “book” publication requires rethinking the nature of a book itself. It might include video, animations, kinetic typography, hyperlinks, geo-location, social interaction. It may evolve more as a computer game than a linear narrative. A “story” might be presented in layers, allowing tangential exploration of one topic before proceeding to another. In fact the story itself may be too linear a concept for this new media. Here, traditional notions of authorship tend to be ill suited. If text is but one component of an ebook, then where does authorship reside? If the organization of the reader/user experience informs understanding of content, then does the designer rightfully claim a role as author? This paper discusses new directions in the development of enhanced ebooks for tablets and the collaboration between designers and writers as co-authors.

Defining the word ebook is problematic. The baggage of the word “book” in ebook and its association with the codex form can be restrictive at the moment of imagining

---

<sup>1</sup> Corresponding Author: Celeste Martin, Assistant Professor, Emily Carr University of Art + Design, 1399 Johnston St., Vancouver, BC V6H 3R9, Canada; E-mail: mmartin@ecuad.ca

and producing new forms of books for tablets. “Book” presupposes pages and the linear form of the argument, usually organized in sections, chapters, subsections, and a number of typographical elements like table of contents and indexes that allow the reader to move from one option to another; this is the typical state of first generation ebooks which also tend to incorporate some basic note making and sharing tools. In some of the ebook proposals developed in the course, content is parsed and tagged, and then searched, reconfigured and accessed in unique ways by each reader. The separation lines between ebooks, webpages and tablet applications are also difficult to establish. The incorporation of social media tools to book communities online where readers are able to share discussions has long been implemented in the Web, while touch screen and geo-location technology has been extensively exploited in tablet apps. What defines ebooks for tablets? We argue that their most defining quality is their communicative purpose in a broad sense; this includes a form of re-writing of the text through participation. As opposed to tablet applications, which by definition are meant to help the user realize a specific task, ebooks require a level of processing of textual and image content that form a specific experience. In relation to webpages and books online, the main differentiating aspect comes not from the technological capabilities but the expectations of the users: tablets offer a scale, tactile experience and relationship to the hand and body that encourages a level of contemplative and exploratory experience of textual information which is different from the more impatient, searching and skimming mode of experience of text on a Web browser. Tablet ebooks also offer a more imposing “frame” around content that separates it as demarcated yet essentially porous to the network of textual connections; paradoxically this physical referential of the layout to the tablet screen size recalls the discreet unit of the printed page yet the screen layout is fluid and dynamic rather than static.

## **1. Context**

In communication design, the extent to which the designer participates in shaping a message has varied from the most ascetic “crystal goblet typography” and the modernist aesthetic of clarity to the more dramatically expressive and personal approaches of New Wave and postmodernist design. Contesting the modernist Swiss dogma of universal, simple, and legible messages, Wolfgang Weingart and others experimented with the semantic malleability of textual information and pushed the limits of legibility and comprehension as they experimented with spacing of letters, words, lines and the hierarchy of design. Weingart pushed the rigorous grid while maintaining it as a central aesthetic reference that became complex and layered, bringing in elements of collage and a closer integration of image and text [1]. In the 1980s the writings of French post-structuralist philosophers (Roland Barthes, Michel Foucault, Jacques Derrida) and their ideas about language influenced the way designers approached the visualization and layout of text [2]. With critical theory came the displacement of the figure of the author as the owner of meanings infused into a text and focus was brought to the reader and the intertextual character and internal workings of writings [3]. For Barthes, “The reader is the space on which all the quotations that make up a writing are inscribed without any of them being lost; a text’s unity lies not in its origin, but its destination” [4]. In “From Work to Text” Barthes distinguishes “work” as a complete and affiliated piece of writing which has been fixed in space and bound by typography and printing [5] and is passively consumed by readers, whereas

“Text” is a dense tissue or fabric that connects with other texts and references in the form of a network and is “experienced only in a form of production” which includes “playing the Text as one plays a game, looking for a practice which reproduces it” [6]. Designers were also influenced by Derrida’s concept of deconstruction as expressed in his book *Of Grammatology*; most notably at Cranbrook Academy of Art under the leadership of co-chair Katherine McCoy, graduate students experimented with this post-structuralist mode of inquiry to question the dualities of content/form image/text text/texture inside/outside revealing their inner instability through typographic interpretation [3,7]. In more general terms, deconstruction in graphic design has been used to describe works that use complex grids and layered imagery and text with ambiguous hierarchies that literally show the openness of meaning and the literary notion of text as a network of references. Even though the emphasis is still on the reader as the active creator of meaning, the designer participates more actively in the authoring process.

With the advent of the Web, Barthes’ notion of “Text” became more fully realized; what used to be a static piece of writing had the potential to be dynamically linked to a web of references, obliterating the temporal and spatial distance between writing in the printed page of the codex and the network of references and connections that writing belongs to. In her essay “Text,” Ellen Lupton analyzes how the web has displaced the reader for the “user” as the more significant subject of our time, “a figure conceived as a bundle of needs and impairments—cognitive, physical, emotional” [5]. The significant change is that in the Web’s interactive environment “[h]ow texts are *used* becomes more important than what they mean” [5].

Tablet devices offer a new and permeable space for texts that shares the connectedness to the referential structure of the Web user while preserving some of the contemplative mode of being of the reader of print books. Current scholarly research in ebook design primarily deals with static first generation epubs [8,9]. As enhanced ebooks only began to be available in 2010, with the first iPads, little research exists beyond forecasting and conceptualizing the potential of digital publishing by such groups as IDEO [10].

Enhanced ebooks in tablets offer an experiential space where, we argue, the main figure could be the “participant,” a subject that realizes Barthes’ idea that text is experienced through performance and production: the text is written over. What differentiates these figures is a combination of expectation and medium capabilities. Like the “user” the “participant” can experience the text as an interactive database, which is dramatically different from the linearity of the printed book, yet the participant goes beyond consuming and searching to sharing and producing content, re-writing the text. In Barthes’ words, “the Text (...) decants the work (the work permitting) from its consumption and gathers it up as play, activity, production, practice” [6]. The successful realization of this possibility requires a profound collaboration between writers and designers. Even though the work of designers, through their manipulation of form and typographic treatment, has always contributed to shaping the text’s message, in ebook applications the designer re-conceives the content as an experiential space that becomes part of the content, extending the writing and realizing the “Text.” Enhanced ebook applications allow for a long list of activities of consumption, sharing *and* production of content, and it is through this close collaboration and co-authorship between designers, writers and other content generators that the writing can be re-contextualized for this interactive and participatory space.

## 2. Project Description

In order to test our ideas about authorship in the development of ebooks, we created a project for Aitken's 4th year "Enhanced Ebook Design" class at Emily Carr University of Art + Design. With colleague Alexandra Samuel, director of the Social and Interactive Media Centre at Emily Carr, we found 5 partners willing to collaborate on the development of ebook concepts specifically targeted to tablets. The partners had textual and other content they wanted to repurpose digitally. Each partner was assigned a group of 3-4 design students. The partners and their titles included:

- Alexandra Samuel: *Sane and Social* shows professionals how to make meaningful and intentional use of social media in their working lives and manage key social media tools.
- Reid Shier from Presentation House Gallery: *Glen Lewis* presents a catalogue of Vancouver senior artist Glen Lewis' art practice of the late 1960s and early 1970s and his interest in ephemeral collaborations, interactive experiments and prescient efforts at social networking.
- Ron Burnett, President of Emily Carr: *How Images Think*, published in 2005 by MIT Press. This title explores the new ecology of digital media, including television, film, photography, animation, video games, data visualization, and the Internet, and its implication for how people perceive images.
- Hanna Cho and Viviane Gosselin from the Museum of Vancouver: *Bhangra.me* will reflect the 2011 exhibit of the same name that investigates the dance and music form of Bhangra as a potent manifestation symbol of Vancouver's emerging transnational/diaspora cultural productions rooted in traditional Punjabi folk practices and yet morphing into forms deeply connected and fed by their existence outside of India.
- Glen Lowry from Emily Carr, and Jonathan Dewar from the Aboriginal Healing Foundation: *Truth + Reconciliation* draws on the experience and insights of more than 60 contributors and foregrounds the writing and images of First Nations, Inuit, and Métis along with other minority groups as that work is brought to bear on Canada's Truth and Reconciliation Commission.

## 3. Project Methodology

As the project involved testing ideas about shared, or collaborative, authorship in ebook creation, we constructed a collaborative flat hierarchy of five separate groups of content creators and design students. Our research methodology followed current thinking in participatory design using a co-creation model such as the one outlined by Sanders and Stappers [11]. This process allows design students insight into the tacit knowledge held by external groups; in this case, content developers (writers, photographers and curators). Students began by developing co-creation kits that facilitated conversations and allowed the non-designer participants an easy entry point into possibly unfamiliar brainstorming techniques. These kits varied widely across the groups: post it notes were used to construct mind maps; printed text was cut up to explore navigation and contest the linear nature of a printed narrative; word and icon cards were collaged to consider content organization. Students recorded the flow of ideas through note taking and photo/video recording. From here, students and content

creators were challenged to think divergently about this emerging medium. The groups met weekly to discuss concepts and consider alternatives. The design students gathered input and ideas from the content providers, considered alternatives and brought them to the meetings for feedback and refinement by the full group. All members of the group were, at least in theory, to be involved in the design. Final projects were presented as “proof of concept” presentations, rather than necessarily functional prototypes. This freed group members from constraints around their technical knowledge.

#### 4. Results

Project results varied considerably as one would expect in an undergraduate class. All groups were (at least initially) heavily influenced by our use of the term ebook. The semantics associated with any use of the word “book” brings with it a host of preconceptions, biases and assumptions—none of which were intended by the course instructor. Several books used traditional “pages”, fitting the tablet screen. Most had some form of navigation system that while not essentially linear, encouraged a linear exploration.

What did set apart the ebook concepts from their codex counterparts was an almost universal inclusion of the reader as participant. In some cases the participant was encouraged to upload annotations, or follow others. The *Truth + Reconciliation* concept relied heavily on reader/participants adding their own stories to the content, creating a cooperative form of content generation and expansion. In other projects they were brought out of reading mode and into a more game-like appreciation of content. The *Bhangra.me* concept included a drum that could be used to play and record beats, mimicking the museum experience.

Arguably the most creative solution was for the book *How Images Think*. Here the group imagined breaking apart the text, creating a completely different understanding of content from the original codex. The result encourages exploration “vertically” rather than linearly. One navigates easily by swiping to find an area of interest, then using a 2-finger “pinch and zoom” gesture to explore deeper into that area. Random exploration is encouraged and a map is created tracking connections between ideas. As well, images and text can be uploaded, creating an organic body of content. This results in an intentionally different experience than is offered with the original book.

#### 5. Discussion

As mentioned, results of the project were mixed. Students had the usual range of motivation and creative skills. Content providers varied in their commitment to the project and the degree to which they had preconceived notions of the final design. But several problems and successes emerged.

In most cases, there was a significant imbalance of “power”. Students were mostly in their early 20s and the collaborators were older and established professionals. As well, design students are largely taught the client/service provider model of design and the collaborators may have been accustomed to working with designers *as* such clients. For both these reasons, students tended to follow the direction of their content providers, answering their stated “needs” rather than pushing them into quite different results. One notable exception was in the case of Ron Burnett. As president of the

university, there was certainly a power imbalance; Burnett was not able to meet the group as regularly as other collaborators, though he was certainly engaged with the group and reviewed material via email. Here, ironically, the power imbalance led to favourable results—Burnett’s comfort with saying “no” to several prototype ideas resulted in a much more interesting solution, making the project one of the most successful.

By breaking away from text as the source, we can more effectively consider meaningful methodologies to communicate experience. One of Burnett’s constant demands for *How Images Think* was to explore a non-linear presentation of the work. It is likely that Burnett’s familiarity with the design process and his own critical examination of the ebook media facilitated a situation where the students were forced out of their “booklike” comfort zone into re-imagining it as a completely original reading experience (Figure 1). The requirement forced students to renounce the attachment to the text’s organization in chapters and its linear argument and embrace the idea of the book as a “sandbox.” This changed their understanding of the book from a mass of content that had to be “consumed” and “followed” to the book as the “occasion” of a series of participatory and productive activities, which included, of course, reading but also annotating, searching, sharing with other participants and including the participant’s own content in the form of images from their tablet photo stream and their commentary. The book’s interface design offers layers of content that progress from more general to more involved content, breaking away from the linearity of the codex and linking it to a mode of database. The group could eventually arrive



**Figure 1.** Concept for *How Images Think* by Dulce Ayala, Renee Shen and Karston Smith.

at a level of ownership of the text as to remix it into a new state, pulling away from the traditional roles of designers at the service of client and content needs; this level of co-authoring was also evidenced in their adding to the title the tagline “How Images Think

is Really How You Think.” It is this parsing, tagging, and re-contextualization of the text along with the inclusion of social media tools for sharing and extending the content how designers essentially re-write the text itself into a new mode of experience which is radically different from the experience of the printed text.

In *Bhangra.me*, the group was charged with the idea of recreating the exhibition as an ebook. However, the exhibition itself was highly engaging and interactive. Viewers could play the drums or listen to music, add stories and locations, and explore a culture through objects and sounds and videos (Figure 2). In creating an ebook for this exhibition, a linear approach was simply unsuited. Instead, the group looked at methods of creating a similar experience in a different media. Clues from gaming were used to provide interesting interaction. A “drum” allows for experiments with sound and participants could record their creations and share them with other users. In this ebook particularly, the designers shaped meaning through careful consideration of the user experience. As the exhibition itself, meaning emerged from immersing the participant in the Bhangra culture, not simply describing it textually.



**Figure 2**, Concept for *Bhangra.me* by Pansy Hui, Sophie Lundstöm Halbert and Dawei Yang

As a print book, *Truth + Reconciliation* provides a vehicle for many different contributors. These different authors’ voices reflect the shared nature of their experiences. With the ebook, the designers chose to reflect these multiple perspectives by creating a true online community (Figure 3). This idea of community is critical to Canada’s aboriginal people. While the text was presented in a more-or-less traditional method, participants could add their own stories that they felt were connected to the book and read others’ uploaded stories. The collaborative nature of the original text was expanded further by facilitating the creation of a shared space. Here, the role of the designer was critical in imagining how the concept of “community” might be applied to an ebook.

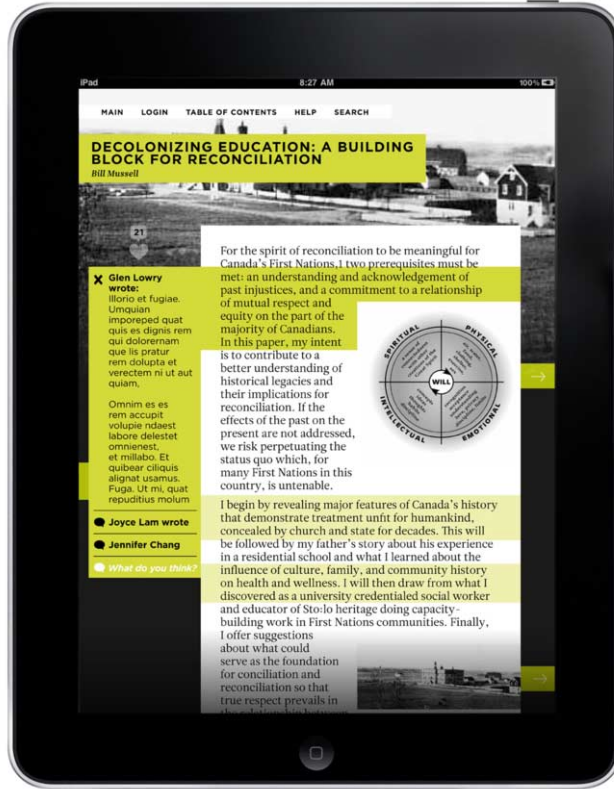


Figure 3. Concept for *Truth + Reconciliation* by Jennifer Cheng, Joyce Tam and Sarah Kozlowski.

## 6. Conclusion

The “proof of concepts” emerging from each group in this project showed the potential and need for designers to intervene as “imagineers” of new modes of being for the book in tablets. The complex ecology of the ebook application creates an experience that is essentially different from that of the codex. The movement between the text and other media “reframes” textuality. In its inclusion of a wide range of participatory elements, the ebook encourages a performance and practice of the text that more fully realizes Barthe’s notion of “Text” as a methodology of production, linking reading to writing in a broad sense. In the most successful examples, the ebook space calls not for “users” but “participants” with the text. The conception and development of this ecology, this malleable space, this re-configuration of the interface of text and its extension and breakage becomes in itself part of the content that defines this ebook experience. We can no longer see the ebook as a copy or translation of the print book to a screen-base realm but a completely new and transformed space the designer has co-authored.



## References

- [1] W. Weingart, *My Way to Typography. Retrospective in Ten Sections*, Lars Müller Publishers, Baden, 2000.
- [2] P. Cramsie, *The Story of Graphic Design: From the Invention of Writing to the Birth of Digital Design*, Abrams, New York, 2010.
- [3] E. Lupton, and J. Abbott Miller, *Design Writing Research: Writing on Graphic Design*, Phaidon Press, London, 2000.
- [4] R. Barthes, The Death of the Author, *Image Music Text*, Heath S (translator), Fontana Press, London, 1977, 142-148.
- [5] E. Lupton, Text, in *Thinking with Type: A Critical Guide for Designers, Writers, Editors & Students*, 2nd ed., Princeton Architectural Press, 2010, 87-100.
- [6] R. Barthes, From Work to Text, *Image Music Text*, Heath S (translator), Fontana Press, London, 1977, 155-164.
- [7] M. Rock, The Designer as Author, *Graphic Design Theory: Readings from the Field*, H. Armstrong, Ed. Princeton Architectural Press, New York, 2009, 108-114.
- [8] C. Mod. A Simpler Page. *A List Apart: For People Who Make Websites*, A List Apart Mag. 11 Jan 2011.
- [9] EPUB Zen Garden: Beautiful eBooks. <http://epubzengarden.com>.
- [10] IDEO. Future of the Book for IDEO. 2010. <http://www.ideo.com/work/future-of-the-book>.
- [11] E. B.-N. Sanders, and P. J. Stappers, Co-creation and the New Landscapes of Design, *Co-Design: International Journal of Co-Creation in Design and the Arts*, 2008, 4:1, 5-18.

# Archives Information Publishing New Design in Post-Custodial Regime: the National Archives Experience Digital Vaults

Maria José Vicentini JORENTE<sup>1</sup>

*Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, Brazil*

**Abstract.** The uses of Information and Communication Technologies (ICT) and Web environments for creation, treatment and availability of information have supported the emergence of new social-cultural patterns represented by convergences in textual, image and audio languages. This paper describes and analyzes the *National Archives Experience Digital Vaults* as a digital publishing web environment and as a cultural heritage. It is a complex system – synthesizer of information design options at information setting, provides new aesthetic aspects, but specially enlarges the cognition of the subjects who interact with the environment. It also enlarges the institutional spaces that guard the collective memory beyond its role of keeping the physical patrimony collected there. Digital Vaults lies as a mix of guide and interactive catalogue to be dealt in a ludic way. The publishing design of the information held on the Archives is meant to facilitate access to knowledge. The documents are organized in a dynamic and not chronological way. They are not divided in *fonds* or distinct categories, but in controlled interaction of documents previously indexed and linked by the software. The software creates information design and view of documental content that can be considered a new paradigm in Information Science and are part of post-custodial regime, independent from physical spaces and institutions. Information professionals must be prepared to understand and work with the paradigmatic changes described and represented by the new hybrid digital environments; hence the importance of this paper. Cyberspace interactivity between user and the content provided by the environment design provide cooperation, collaboration and sharing knowledge actions, all features of networks, transforming culture globally.

**Keywords.** Web Publishing. Historical Archives. Information and Communication Technologies. Post-custodial. Hypertext.

## Introduction

The increasingly intensified uses of Information and Communication Technologies (ICT) and the Web environments for creation, treatment and availability of information have supported the emergence of new social-cultural patterns represented by convergences in textual, image and audio languages used within the media that were created in these environments. Such actions of information creation, storage, retrieval,

---

<sup>1</sup> Corresponding Author: Maria José Vicentini Jorente, Departamento de Ciência da Informação UNESP, Campus de Marília, São Paulo, Brazil; E-mail: mjjorente@yahoo.com.br

use and re-use occur at varied levels of complexity depending on the organization of information in the digital environments, the information context and the subjects involved.

Archivistics, an Information Science subarea, traditionally deals with *fonds* of records according to centenary, if not millenary, principles that have not changed much regarding to patrimony and custody of information collections. The Historical Archives custody spaces are mostly still meant to professional visitors, not to the general public. In this context of Archives, Research Instruments traditionally lead to data, interfacing the relations between the searcher and the information stocks: they are guides, catalogues and inventories destined to mediate searching. With the changes brought by the ICT, this landscape has already been significantly altered and the relations realm of the above mentioned professionals with the information in the Archives collections were facilitated by digitalization.

Furthermore, some software designs and information retrieval models can give access and make general public interaction possible within the digital interfaces of the Archives, provoking relevant changes in the uses of the information stocks preserved by these institutions. These changes implement historical Archives with a broader social meaning, according to a new and post-custodial paradigm, in which any individual is able to access, research in and rebuild virtual collections, creating unique paths to approach historical contents. In such situations, deriving from the way information is published in Web for visualization, a re-contextualization of the archives institution role in society and culture may occur.

The custodial paradigm is therefore to overcome physical constrains by providing access to information which could once only be known by visiting the collections in person.

Considering the analyzes of the interactive experience in a Web environment made possible by the American *National Archives Experience Digital Vaults* [1] as a means and as a cultural message, this paper intends to focus on the results of the representation convergences that are offered as interface of a complex system [2]. This system synthesizes information design options at information setting, provides new aesthetic aspects, but specially enlarges the cognition of the subjects who interact with the means. It also enlarges the institutional spaces that guard the collective memory beyond its role of custodying the physical patrimony kept there.

## 1. Brief Overview

Documents left behind by Paul Otlet [3] show that he had already continuously experienced alternative representation views and design, as well as relationship perceptions among classes of objects in studies carried out at the beginning of the 20th century, experimenting “various forms of knowledge integration and distribution that imply the need for mechanisms currently recognized as similar to various types of interface.” [2]

Designing information storage support that is capable of converge different kind of documents and records to facilitate subsequent retrieval was also carried out by Vannevar Bush [4] in 1945. The Memex [5] project was presented as the solution through a new form of information storage support and retrieval, able to relate and link different documents with each other, creating what is currently called multimodality.

Although the Memex building was never accomplished, remaining as a project, the mere possibility of a multimodal support as Memex inspired other researchers.

Douglas Engelbart [6] expanded the concept of multimodality on one same support and created a collaborative workspace, the oN-Line System (NLS), in the 1960s. The NLS can be considered the first collaboration system in a digital environment, having links and a peripheral tool – the mouse - to point out the wishes of the subject interacting with the machine, as well as the window-like screens to organize information according to its importance in creating the content. Engelbart [6], working at Air Force Office of Scientific Research, wrote a report to the Director of Information Sciences in 1962, called “*Augmenting Human Intellect: a conceptual framework*”[7]. On the document, Engelbart claims that computers augment comprehension due to its complex organization format of information and thought. Also, Engelbart highlights the aspect of discontinuity in relation to the traditional forms of access to knowledge.

Following the same discontinuity idea, Theodor Nelson [8], influenced both by Bush’s Memex and Engelbart’s ideas presented the concept of Hypertext [9] in 1965, which was considered a landmark in the history of multimodality. Nelson developed Xanadu Project [10] hypertext system lying between environmental design and software, being able to view a text and its links on computer screen. Xanadu Project [10] has received various updates and different versions and Nelson had to continuously readapt to new teams working on the project in the forty years from the date the project was first launched to the present days. In Nelson’s hypertext vision, a serious electronic literature, created to teaching and sharing must support two-directional and profuse links and must offer means for a proper reuse; however this reality has not been accomplished yet.

A version of the complex Project Xanadu [10] was the one prevailing at Timothy Bernes-Lee’s [11] World Wide Web [12] in 1990. Yet very similar and dependent on the technological paradigms of writing on paper, and thus considered continuity line, the Web that prevailed is criticized by Eric Drexler [13, 14] for its linearity and mono-directional links. According to Drexler, a complete hypertext must support links in a way they can be followed in two directions; meanwhile the semi-hypertext can be followed in only one direction as in the Web. Even though, the way it is, the Web has revolutionized the means of disseminating information and has fostered the development of an information sharing culture that has brought us to our days.

In this context, another landmark to be considered is the meeting hosted by O’Reilly Media [11] to discuss the prevailing Web, in 2004. At the meeting, a second period in the Web lifetime was identified, a moment broadly disseminated by the term Web 2.0 that, despite controversies, survived: digital environments now add 2.0 to emphasize qualities as interactivity, sharing and collaboration.

Considering what has been discussed about Web 2.0 [12], this paper will especially focus on access and information sharing through views provided by convergences of multimodal language design, with a special emphasis on image, which have been developed since the Graphic User Interface (GUI) creation [15] in the 1990s and have an important role consolidating the concepts involving Web 2.0 [12].

## 2. National Archives Experience Digital Vaults (USA)

The American National Archives is a governmental federal agency for custody and preservation of American Government documents, a function shared by every Historical Archive in the world. However, in the USA context, due to the democratic commitments speech on which the United States are grounded, access, discovery, use and learning of the cultural heritage preserved on the documents guarded by the institution need to be ensured. Thus, according to the institutional web page, the National Archives exists to “support democracy, promote civic education, and facilitate historical understanding of our national experience.”[1]. Apart from keeping historical documents, ideological reasons clearly emerge from the quotes selected in the text disposed in the Web site that can justify the intentions of favoring access, learning and preserving not only the collections, but the immaterial heritage there contained:

Every subject relating to American history is covered in the nine billion records, the millions of photographs, maps and electronic records and the thousands of motion pictures and audio recordings that are available to the public at the National Archives nationwide. The National Archives maintains 20 regional records facilities and 12 Presidential libraries nationwide, as well as the Office of the Federal Register, the National Historical Publications and Records Commission (NHPRC), and the Information Security Oversight Office (ISOO) [1]

*Digital Vaults Experience* [1] is therefore developed in this context. The expression digital vault must be highlighted because in general terms it is related to priceless items which must be well safeguarded so that they are not lost or exposed, but here it is probably strategically used to explore the semantic oppositions between closure and openness of the archives. A “vault” is usually within a bank facility where objects or critical data are kept and accessible only to few privileged ones, not frequently available for the public in general. However, the web Page describing the experience reads:

The Foundation for the National Archives works in partnership with the National Archives to “open the stacks” of the Archives and enable millions of visitors to interact personally with the original records of our democracy through the National Archives Experience, filling the public spaces of its flagship building in Washington, D.C., with exciting exhibitions and varied learning opportunities. [1]

The Digital Vaults [1] Experience initiative thus respond with information new design intending to provide a large number of visitors with access to valuable archives documents, unlike the above mentioned traditional Historical Archives in the whole world, which tend to be very patrimonial, serving specialists and not the general public. To accomplish this goal, in USA the Foundation for the National Archives [16] was founded in 1992 as an independent non-profitable government organization whose mission is to call public attention to actions that focus on access and dissemination of the information kept in custody and to rise funds among civil society to create alternatives of access as the focused one. Together, Government and Foundation are partners of a common effort for the development of a synchronic work with The USA National Archive, whose duties of treatment, organization and information selection for

dissemination are preserved; the Foundation raises the funds to fulfill this mission claimed to be essential to education for democracy, parallel to mere documentary custody and preservation. In this context it is possible to understand that the creation of National Archives Experience as a “national civic literacy initiative which includes permanent exhibits, educational programs, traveling exhibits, special events and screenings, educational literature, and historical/records-related products and media”[1].

The initiative afforded design specialists to build The Digital Vaults [1] Experience database and the interactive hypertext digital space. This database, which conveys the experience, holds information related to the American National Archives contained by approximately 1,200 chosen documents in distinct supports: pictures, drawings, maps, etc. What links them is the design software, based on a digital system of relationships between keywords, translated into a convergence of visual experiences presentations. Such features provide visitors with an opportunity to customize by personalizing information exhibition.

Digital Vaults Experience [1] then seems to be an instigator that follows the shape of traditional physical Archives research tools: it lies between a mix of guide and interactive catalogue to be approached visually in a playful and enjoyable way. The web studio design, Second Story Interactive Studios [17] defined it as a Web 2.0 experience that won a Information Design award and “was selected Honorable Mention in Exhibition at the 2008 Museums in the Web Conference” [1], according to the press release page of the National Archives Digital Vaults Experience.

The presentation design of the archives as described above is thought to facilitate access to knowledge and create opportunities for History learning by providing, in a broad sense, primary sources of information that bring attractive, carefully and aesthetically presented images. The choice of paths to be followed from the relationships among the documents will constantly re-draw a historiography of informational flows by association which is similar to natural human neural connections [18], according to previously thought contents created as collections by teachers and students or other education professionals.

These actions meet the post-modern concepts of historical and collective memory construction: unlike the information kept in the stacks of physical archives, in Digital Vaults Experience the documents are displayed in a dynamic and not chronological manner. The digitalized documents are not divided in *fonds* or distinct categories as the physical ones must be, but through controlled interaction and association made by documents previously indexed and linked by topic mapping software which creates different design approaches for information and viewing of documentary content.

As a result of the partnership between private and public sectors and the consequent possibility to act on dissemination through interactive Web 2.0 generation technology, new perspectives were added to Historical Archives.

Tom Wheeler, President of the Foundation for the National Archives, said, “We are thrilled that the first stage of the Digital Vaults has been launched and is ready for online visitors. The web site has long been considered a critical component of the *National Archives Experience*, one that would go far beyond the ‘granite walls’ of the National Archives building in Washington, to enhance the experience of visitors worldwide. The Digital Vaults will help to ‘unlock’ the stacks and reveal some of the treasures entrusted to the National Archives.” [1].

To embody the educational aspect, the construction of the digital environment searched to simulate the historical research to encourage the discovery urge in visitors, according to Suzanne Isaacs and Lee Ann Potter's analyses in *Teaching with documents* [19], published by *National Council for the Social Studies*.

The authors highlight that when one enters the web site (s)he views the publishing image of eight thumbnail format documents. The screens overlap the documents linked to each other by related topics. A group of five keywords (tags) and four types of filters are placed in a squared field centered as an interactive box on the left side of the screen. Whenever one selects an image, an exploratory path is triggered. As the selected image moves toward the center of the screen, the view is completely remodeled, reloading to every click and so do the keywords and the filters' box. If one filter is chosen and clicked on, another group of relations are also shown in a second level, providing tools to reduce and facilitate the search and the creation of new associations. On the right side of the screen there is another tool meant to reduce/amplify the options related to the distance among the links related by the graph theory that embody the framework. If a document cannot be reached due to the limits of the screen, it can be brought to a better visualization point by the arrows in this box or, on the contrary, documents placed nearer to the centered of the screen can be enlarged through the tool, zooming the search. With a move of the mouse on the tags displayed on the left of the screen, lines pop up spotting the relations links to the topic. When the mouse is hovered on the zoomed document, other details can be viewed by the opening of another box that displays mixed image/text description and briefly contextualizes the document. In addition to these actions, the Digital Vaults Experience [1] also allows the user to execute various tasks within the environment through five buttons at the bottom of the screen that once clicked allow interactive actions in the page: *collect*, *backtracking*, *pathways*, *search*, *create*.

The *collect* button creates an account and saves the settings for every visit if wished by the user, allowing previous visit and reading retrieval at any time. After the settings are collected, they can be transferred to a slide presentation program to build teaching material for further sharing.

The *backtracking* button takes one back to the paths already chosen, displaying documents previously exposed centrally as well as to the settings previously established for such actions.

The option *pathways* bring possibilities for choosing diverse paths to retrieve historical information and "challenges" where some tags can also be erased to retrain options. This tool is meant for teachers who can direct playful and delightful learning.

The *search* button allows a kind of search done by tagging each document and it offers a link to the main catalogue (Archival Research Catalog – ARC) which in turn is linked to regional historical Archives and other Libraries. The results follow thumbnail format images.

There is also an interactive creation section which enables creating podcasts, posters and presentations with the preview option by pushing *create*. The creation button opens two options of publishing: poster and movie. These must be brought from the collection previously built by the chosen subjects and combine again text and images.

It is important to highlight the fact that depending on the way the design professionals build the indexation and tagging of the 1,200 chosen documents, failures can happen and do happen in this process of digital environment building. Declared as an experience, critics may improve and lead the experience to achieve a better accuracy

level avoiding distortions to accomplish its role as a reliable information and knowledge interface. Information professionals hired to work together with the web designers in multidisciplinary creation groups for these publishing environments should be able to bring Information Science expertise techniques and strategies related to ontology, thesaurus, and so on, closely related to the identification of suggested improvements.

Nonetheless, the point to be made in this article regards the fact that because this experience is supported on digital means, the linking elements and possibilities described above create mobility, dynamics and detailing impossible to be otherwise performed in physical environments of Historical Archives: apart from the descriptions above, the links also relate to other articles, online exhibitions, teaching and learning plans and interoperability with other governmental agencies creating a flow, and emphasizing the aspect of knowledge processing as described by Michael Buckland in *Information as a thing* [19].

Second Story Interactive Studios [17] is a commercial studio and also designed other immersive digital experience as the one for the Coca-Cola Company, located in Atlanta. The experience follows the tendency already explored in The National Archives Digital Vaults Experience, of provoking curiosity and exploring the feeling of penetrating a vault compound in which interactive actions can be performed; and which appeals to the same concept of the Vaults, but overcomes it: in Coca-Cola Company situation the immersion invokes an even more intense sense of participation, with other possible various actions. As a well-known wealthy company, the inversions of money destined to the immersive environment were one differential to be considered, but still makes one think on even broader possibilities for enlarging universes of Historical Archives publishing records keeping faithfulness to principles of *fonds* and yet conveying information only possible by simulation and associative relationship of links.

### **3. Design Favoring Publishing and Information Viewing: a Post-Custodial Perspective for Information Science**

The custodial paradigm and its practices have turned Historical Archives into privileged environments for historians and information professionals. Limited by buildings, spaces allocated for archives were changed in function from administrative processes managers (*records*) to historical testimony keepers (*archives*) very similarly to the museums' role in modern society, although the Archives collections remained closed to the public.

The idea of a Post-custodial Archivistics was born in Quebec Canada. Louise Gagnon-Arguin linked the concept to post-modernity [21]. Just after, Carol Couture [22], Jacques Ducharme e Jean-Yves Rousseau also did it.

Fernanda Ribeiro [23] and Armando Malheiro da Silva [24], Portuguese archivists, defend a post-custodial paradigm related to analyses of archives on current days too. Unlike the previous period, known for praising custody, patrimony and technicalities regarding Archives and Libraries, the post-custodial paradigm does not depend on physical spaces and institutions created mainly since 1789 along with the administrative changes brought by the French Revolution: a model of institutional keeping and guarding public administration documents has become available for the sake of democratic ideals. Such paradigmatic model is attached to the building, the



physical place, which according to Ribeiro [23] was a popular idea fuelled by the positivism.

The French model, widely spread in Europe in the 1800s, characterized as historical, positivist and custodial consolidated along the 20<sup>th</sup> century was re-shaped due to technological developments. Since the beginning of 20<sup>th</sup> century, new patrimonial values have been added to historical archives, becoming clear that they are cultural asset, possible to own, acquire, collect or even sell because they are economically, affectively and aesthetically valuable. [23]

The Dutch model, parallel to the French, related to the current administration, on the other hand, focused on document management, more instrumentalist, however more independent from other disciplines. According to Ribeiro [23] both models represent a dichotomy and oppose in forms of treating administrative documentary production. Dichotomy and opposition that became more evident after the period between Wars, with the spread of documentation on paper and the increase of mass media, new technologies then. However, none of the models overcome the issue of a physical place as a limited environment and neither the emphasis on management and autonomy employed by the Dutch line of thought, nor the conceptual crisis triggered by the new technological environments, embodied the Archives practices with the processes scientificity, during the 20<sup>th</sup> century.

Archivistics has developed as an independent discipline, but the historical-technicist paradigm featuring the field in the last two centuries, which led to its technical autonomy ended up as an embarrassment factor. For this reason, there was not a qualitative improvement needed for the development of the discipline when technology "threatened" traditional documents in their static support and the emergence of the Information Society posed new challenges. [23]

The search for an independent identity constituted a new science field in the 1950s - Information Science (IS) - also marked by the parallel increase of informatics which deeply influenced the emergency of the shift theory: to post-custodial and informational scientific paradigm.

The importance of access and information dissemination currently represented by new design and new visualizations made possible by Information and Communication Technologies (ICT) and put into practice via Web environments, thus, must be thought as means of new opportunities for IS to build scientific knowledge. On the other hand, one must also weaver ICT innovations with the millennial Archive practices placed along History as well as with the rupture moment instituted by the French Revolution - that led modern democratic ideologists.

Regarding to considerations that paradigms shifts only happen based upon visible ruptures, and that no rupture was imposed in Archivistics or in IS to be considered a paradigm break, one can argue that paradigms are also historically broken in continuity line as the systems they represent exhaust and go into entropy.

This paper then considers that digital environments of information publishing as the one described here, mediate, as interfaces, a considerable amplified access to information; the design strategies used within the environment build new knowledge in Archivistics - or through Archives - to layers of subjects before ignored; and thus should be thought as indicatives of the paradigmatic changes. Because Information

Science and Computer Science share systems and interdisciplinary fairly closely, it is possible to envision emerging zones of hybrid design not present before in Historical Archives environments. They are not mere research instruments; they are not materialized as collections *fonds* either; they are immaterial, but real. Therefore, the *National Archives Experience Digital Vaults* [2] as an information system currently inserted in Information Science new paradigms of interactivity can be also considered part of a post-custodial information regime. It enables new points of view to perceive the world and its values - not statically, but dynamically, like the converged formats built in hybrid modes of digital environments. The scientific fields emerging from there do not belong to one specific discipline or domain; due to technological revolutions, they are considered inter and cross-disciplinary, composing new systematically inter-related fields. The methodologies approaching these changes will also need changing to meet the systemic values developed within these dynamics.

Professionals working in the fields previously defined will have to be prepared to understand and learn about the paradigmatic changes described and represented by the new hybrid digital environments. As highlighted by Ribeiro [23],

The technician, document keeper who discretely waited for the organic entity, once information provider, to send him documental support which was not longer in administrative use will have to be at the front line in the "post-custodial era". This means he will have to be ahead together with the information production and be manager and structure for the information flow that runs and feeds the function and decision making of the organization. [23]

The new information professional must consider that on cyberspace interactivity between user and the content provided by the environment design occurs in real time and continuous space; that on cyberspace custody and safeguarding ideas are broken because physical distance is irrelevant; that on cyberspace distances only depend on access conditions determined by networks and software; that barriers are broken creating opportunities for pathways through links and actions selected by individual internet users; last, that interactivity and creativity consequently comprehend cooperation, collaboration and sharing knowledge actions and all these peculiarities have their own level of knowledge on the net and hence transform culture globally.

## References

- [1] NATIONAL ARCHIVES. *National Archives Experience "Digital Vaults" Online Exhibit Nominated for Webby!*. p.1. In: <http://www.archives.gov>/  
<<http://www.archives.gov/press/press-releases/2008/nr08-88.html>> . Viewed: Jan.2012.
- [2] HEUVEL,C., RAYWARD, W. B. Facing Interfaces: Paul Otlet's Visualizations of Data Integration. *Journal of the American Society for Information Science and Technology*. P.2314. Volume 62, Issue 12, pages 2313–2326, December 2011. Viewed: Jan 2012. < <http://onlinelibrary.wiley.com/doi/10.1002/asi.21607/full>>
- [3] In: <http://www.archives.gov/press/press-releases/2008/nr08-88.html>> .
- [4] In: [http://en.wikipedia.org/wiki/Vannevar\\_Bush](http://en.wikipedia.org/wiki/Vannevar_Bush).Viewed: Jan. 2012
- [5] In: <http://en.wikipedia.org/wiki/Memex>. Viewed: Jan.2012

- [6] In: [http://en.wikipedia.org/wiki/Douglas\\_Engelbart](http://en.wikipedia.org/wiki/Douglas_Engelbart). Viewed: Jan. 2012
- [7] ENGELBART, D. *Augmenting Human Intellect: A Conceptual Framework*. Stanford Research Institute. 1962. Viewed: Jan. 2012  
[http://www.invisiblerevolution.net/engelbart/full\\_62\\_paper\\_augm\\_hum\\_int.html](http://www.invisiblerevolution.net/engelbart/full_62_paper_augm_hum_int.html)
- [8] In: [http://en.wikipedia.org/wiki/Ted\\_Nelson](http://en.wikipedia.org/wiki/Ted_Nelson). Viewed: Jan. 2012.
- [9] In: <http://en.wikipedia.org/wiki/Hypertext>. Viewed: Jan. 2012.
- [10] In: [http://en.wikipedia.org/wiki/Project\\_Xanadu](http://en.wikipedia.org/wiki/Project_Xanadu). Viewed: Jan. 2012
- [11] In: [http://en.wikipedia.org/wiki/Tim\\_Berners-Lee](http://en.wikipedia.org/wiki/Tim_Berners-Lee). Viewed: Jan. 2012
- [12] In: [http://en.wikipedia.org/wiki/World\\_Wide\\_Web](http://en.wikipedia.org/wiki/World_Wide_Web). Viewed: Jan. 2012
- [13] DREXLER, E. *Hypertext Publishing and the evolution of knowledge*. 1987. In: <http://e-drexler.com/> Viewed: Jan. 2012
- [14] In: [http://en.wikipedia.org/wiki/K.\\_Eric\\_Drexler](http://en.wikipedia.org/wiki/K._Eric_Drexler). Viewed: Jan. 2012
- [15] In: [http://en.wikipedia.org/wiki/Graphical\\_user\\_interface](http://en.wikipedia.org/wiki/Graphical_user_interface). Viewed: Jan. 2012
- [16] In: <http://www.archives.gov/nae/support/>. Viewed: Jan. 2012
- [17] In: <http://www.secondstory.com/>. Viewed: Jan. 2012
- [18] JORENTE, M. J. V. *Tecnologias, Mídias, Criação e Hipertextualidade na transformação da Informação em conhecimento interativo*/ 244 pgs./ Maria José Vicentini Jorente. Marília, UNESP, 2009, p.91-95.
- [19] ISAACS, S., POTTER L.A, Teaching with documents: The Digital Vaults. In: *National Council for the Social Studies*. p.297-300, 2008. In: <http://www.archives.gov/preservation/conferences/2009/presentations/isaacs-social-education-2008.pdf>>. Viewed: Jan.2012
- [20] BUCKLAND, M K. Information as Thing. *Journal of the American Society for Information Science*; Jun 1991; 42, 5; ABI/INFORM Global, pg. 351-360.  
 <<http://people.ischool.berkeley.edu/~buckland/thing.html>> Viewed: Jan.2012
- [21] In: [http://fr.wikipedia.org/wiki/Louise\\_Gagnon-Arguin](http://fr.wikipedia.org/wiki/Louise_Gagnon-Arguin). Viewed: Jan. 2012
- [22] In: [http://fr.wikipedia.org/wiki/Carol\\_Couture](http://fr.wikipedia.org/wiki/Carol_Couture). Viewed: Jan. 2012
- [23] RIBEIRO, F. *Os Arquivos na era pós-custodial: reflexões sobre a mudança que urge operar*. ps.4,5,6,7,8,9.  
 In:< [ler.letras.up.pt/uploads/ficheiros/artigo10091.pdf](http://ler.letras.up.pt/uploads/ficheiros/artigo10091.pdf)> Viewed: Jan.2012.
- [24] SILVA, M.A. *A Informação: da compreensão do fenômeno e construção do objecto científico*. Porto: CETAC. Media; Edições Afrontamento. 2006.

# Ontology-Based Search and Document Retrieval in a Digital Library with Folk Songs

Maria NISHEVA-PAVLOVA<sup>1</sup> and Pavel PAVLOV  
*Faculty of Mathematics and Informatics, Sofia University*

**Abstract.** The paper discusses some aspects of a work in progress aimed at the development of technologies for digitization of Bulgarian folk music and building a digital library with Bulgarian folk songs presented with their music, texts and notes. This library provides digital preservation of the sound recordings, lyrics and notations of more than 1000 Bulgarian folk songs and tools for various types of search and analysis of the available resources. The presentation is focused on the subject ontology especially developed for the occasion and its application in the implementation of a tool for semantics oriented search in the lyrics of songs.

**Keywords.** digital library, metadata, ontology, search engine, document retrieval

## Introduction

Traditional Information Retrieval technology is almost entirely based on the occurrence of words in documents and therefore it may be characterized as keywords-based. Search engines augment this in the context of the Web with information about the hyperlink structure of the Web. In the typical cases, the user provides the search engine with a word or phrase about which he/she is trying to gather or research information. The kinds of queries a keyword-based system can accept are quite limited and semantically poor. Moreover, simple keyword-based search usually returns too many results which have to be additionally filtered somehow.

*Semantic Search* [1] attempts to augment and improve traditional search results. Ontologies play a key role in this kind of search. An ontology, by definition, represents a formal model of the common interpretation of the entities and relationships in a domain of interest. Therefore, ontologies should be widely used in digital library systems. In particular, at least three types of ontologies have been identified as applicable in a specific type of digital libraries – the so-called *Semantic Digital Libraries* [2,3,4]: bibliographic ontologies, ontologies for content structures (or subject ontologies, according to the term used in this paper), community-aware ontologies. Subject ontologies are useful for supporting the semantic annotation of all types of library resources. They also play the role of knowledge sources which define the meaning of most domain concepts, their hierarchy, properties and relationships.

---

<sup>1</sup> Corresponding Author: Maria Nisheva-Pavlova, Faculty of Mathematics and Informatics – Sofia University, 5 James Bourchier blvd., 1164 Sofia, Bulgaria; E-mail: marian@fmi.uni-sofia.bg.

The paper discusses some results of the activities within an ongoing project aimed at the development of technologies for digitization of Bulgarian folk music and building a semantic digital library (named DjDL) with Bulgarian folk songs presented with their notes, text and music. DjDL is intended to serve as a platform for digital preservation of the sound recordings, lyrics and notations of a significant number of Bulgarian folk songs and to provide adequate access to them. The emphasis of the presentation falls on the provided tool for semantic (ontology-based) search in the lyrics of songs.

## 1. An Overview of DjDL

Currently DjDL contains a collection of over 1000 digital objects which represent a part of the unpublished archive manuscripts of Prof. Todor Dzhidzhev including folk songs from the Thracia region of Bulgaria.

DjDL has the typical architecture of an academic digital library with heterogeneous resources. Its functional structure includes six main components:

- a metadata catalogue;
- a repository;
- a subject ontology;
- a search engine;
- a module implementing the library functionality;
- an interface module.

The library catalogue consists of short descriptions (in XML format) of the particular folk songs included in the repository. These descriptions contain various types of metadata, for example: the title of the song, the song genre in accordance with different classification schemes (e.g. according to the typical time and space of performance, the thematic focus(es), the context of performance, etc.), the region of folk dialect, the informant (the person who conveyed the song to folklorists), the folklorist who gathered the song, the singer(s), the date and place of record, etc. More accurately, each catalogue entry contains the text (i.e., the lyrics) of a particular song accompanied with the corresponding metadata.

The repository of DjDL contains heterogeneous resources of four types [5,6]:

- lyrics of songs (in PDF format);
- notations of songs (in LilyPond format [7] and properly visualized in PDF format);
- musical (MP3) files with the authentic performances of the songs – as far as such exist in the archives;
- musical (MIDI) files generated with the use of LilyPond from the notations of the songs.

The subject ontology includes concepts of different areas related to the contents of the lyrics of songs, with description of their properties and different kinds of relationships among them. It plays a significant role in the implementation of the full functionality of the search engine.

The purpose of the search engine is to provide adequate access to the complete palette of resources stored in DjDL.

The library functionality and the user interface of DjDL are designed in accordance with the expected requirements of the typical users of the library. The

interface module provides adequate online access to the library resources and supporting software tools.

## 2. Subject Ontology

The subject ontology describes a proper amount of domain knowledge (with definitions of the main concepts, their properties/relationships and representative instances) that has been used to build the catalogue descriptions and to process the user queries. It consists of several interrelated subontologies needed by the search engine of DjDL and developed especially for the occasion:

- ontology of folk songs – includes various genre classifications of folk songs (by their thematic focus – historical, mythical, etc.; by the context of performance – Christmas folk songs, harvest songs, etc.; by their cultural functions – blessing, oath, wooing, etc.);
- ontology of family and manner of life;
- ontology of impressive events and natural phenomena;
- ontology of social phenomena and relationships;
- ontology of historic events;
- ontology of disasters;
- ontology of feasts;
- ontology of traditions and rites;
- ontology of blessings and curses;
- ontology of mythical creatures and demons;
- ontology of administrative division – combines the current administrative division of Bulgaria with the one from the beginning of XX century.

Most classes of the subject ontology are constructed as defined OWL classes, by means of necessary and sufficient conditions defined in terms of proper restrictions on certain properties (see e.g. Figure 1).

The properties “form” and “synonym” provide the search engine with suitable grammatical forms and synonyms of the terms used as names of ontology classes.

## 3. Patterns and Rules

The folklore lyrics uses lots of similes, metaphors, idioms and other sophisticated or language-dependent stylistic devices. For that reason, it is expedient to accompany the use of proper ontologies with other Artificial Intelligence tools to provide more adequate support for the semantic search.

In this sense we direct our attention to the design and use of proper patterns of typical stylistic or thematic constructs which could be matched with relatively large parts of the texts of folklore songs. We call them *concept search patterns*. For instance, we have already defined a number of search patterns of constructs standing for “unfaithfulness”, “jealousy”, “discontent”, “sedition” etc. as well as the corresponding pattern matching rules and recently make various experiments with them.

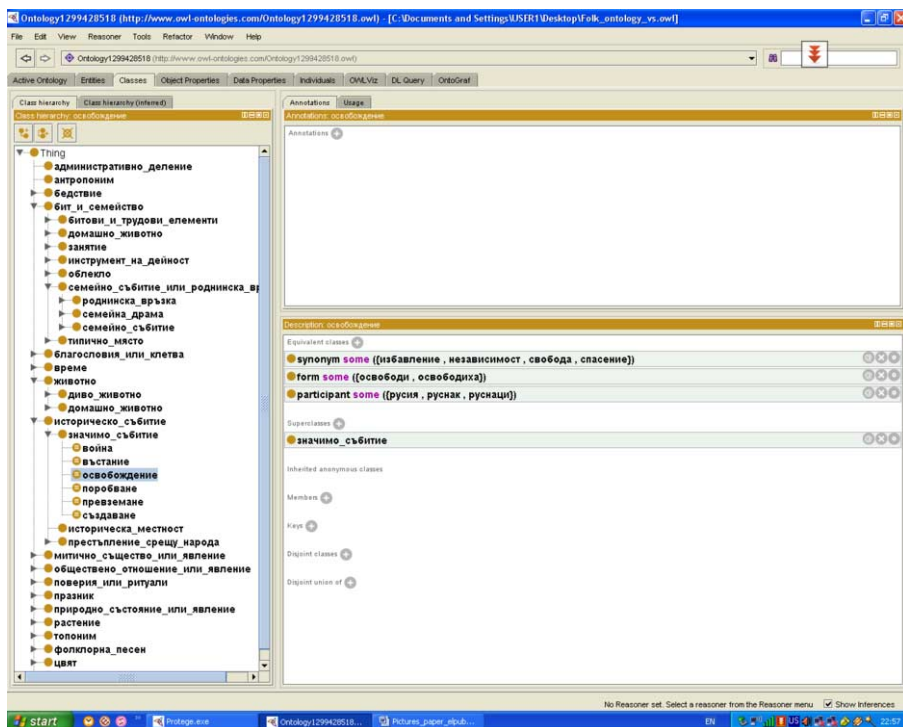


Figure 1. Part of the subject ontology.

Let us consider for example the definition of a pattern for the concept “love infidelity” (“любовна изневяра” in Bulgarian) which is oftentimes used in Bulgarian folk songs. Multiform phrases have been used in the lyrics of songs to express (real or possible or future) love infidelity, e.g. “Та друго либе залиби”, “Друго любе жъ залюбя”, etc.

An essential part of them match the pattern

```
<< друг$? <любим_а $? зал?б? >>
```

Here “?” and “\$?” are used as wildcard symbols (the question mark “?” matches any letter at the corresponding position and the symbol “\$?” matches any corresponding sequence of zero or more letters) and the angle brackets “< >” enclose the name of an ontology concept (the concept “любим\_а” in the subject ontology means “beloved” in English).

#### 4. Functionalities of the Search Engine

The search engine of DjDL supports two main types of search: keywords-based and semantic (ontology-based) search. The design of this search engine is based on some former results of the authors [8,9] and some ideas from [1] and [10]. Its current version realizes some facilities for search in the catalogue metadata and the lyrics of songs only. The functionalities supported at present were specified after a careful study of the

requirements of the typical user groups (specialists and researchers in ethnomusicology and verbal folklore, philologists, etc.).

The user queries define restrictions on the values of certain metadata attributes and/or the texts of the required folk songs. The search procedure consists of some pattern matching activities in which the catalogue descriptions containing metadata and lyrics of songs are examined one by one and those of them having a specific set of element values that match the corresponding components of the user query, are marked in order to form the search result.

The matching process within the keywords-based search consists in testing the appropriate sources for equality.

During the construction of a query for keywords-based search, the user is asked to indicate the search source(s) – search in the lyrics of songs, search in the metadata or combined search in the lyrics of songs and catalogue metadata. One can define a search query consisting of an arbitrary number of words or phrases as well as specify proper logical connectives between them: conjunction (and) or disjunction (or). Negation (not) is also allowed as a unary operator indicating that the negated word or phrase should not appear in the searched text. As a result of the user query processing, a list of links to the discovered catalogue files with metadata and lyrics of songs has been properly displayed. This list may be ordered according to various criteria.

Here are some typical examples of queries for keywords-based search:

- search and retrieval of songs whose lyrics contain specific words or phrases;
- search (and retrieval) of songs with distinct thematic focus or context of performance;
- search of songs performed by a given singer;
- search of songs performed in a particular settlement;
- search of songs performed by singers from a given place.

The semantic (ontology-based) search tool of DjDL is aimed at the provision of a set of additional facilities for augmentation and refinement (automatic reformulation according to the available explicit domain knowledge) of the queries for keywords-based search.

The augmentation of the user query is based on proper utilization of the subject ontology. First of all, an exhaustive breadth-first search in the graph representing the “is-a” concept hierarchy described by the subject ontology is performed, starting from the node which corresponds to the original user query. The names of the visited nodes that are in fact the respective more specific concepts included in the ontology, are added to the one given by the user. The resulting list of concepts if properly visualized and placed at user’s disposal for further refinement (see Figure 2).

Within the next step of query expansion, the search engine adds to the newly constructed set of queries some synonyms and derivatives of the main terms found as their relevant properties in the subject ontology. The corresponding property values from the definitions of all concepts included by that time in the expanded user query and the existing instances of these concepts are added to the query as well.

Thus the user query is augmented as far as possible in terms of the subject ontology and in fact it has the shape of a disjunction of all included forms of concepts and instance names. In this form the resulting query is ready for further refinement (see e.g. Figure 3) and processing.



As example queries for ontology-based search, being of interest for folklorists (according to [11]), that can be executed by the search engine of DjDL, we could indicate the queries for search and retrieval of:

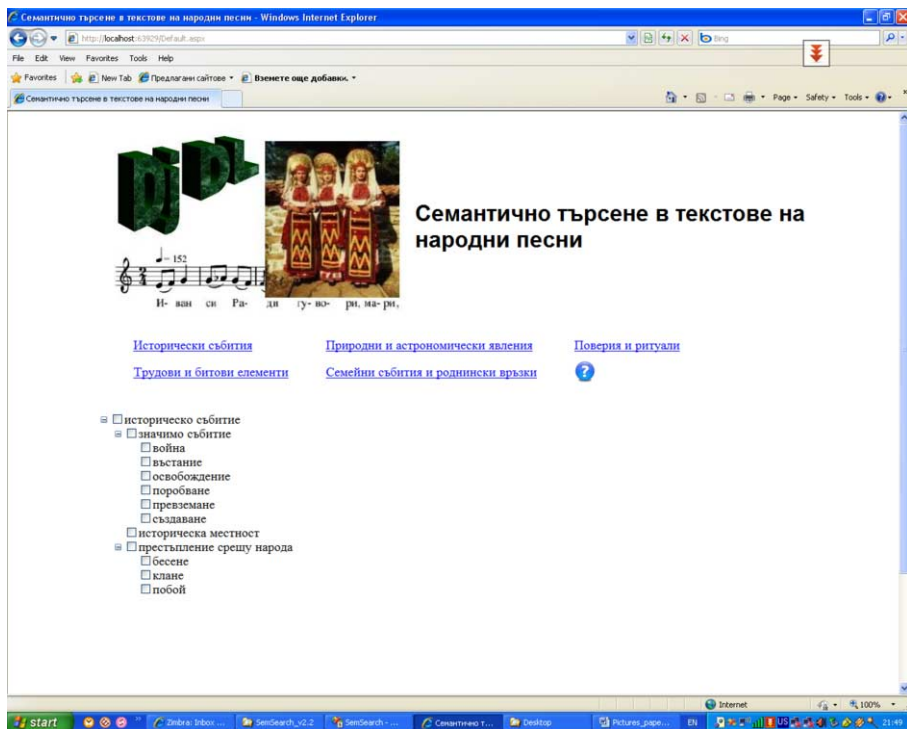


Figure 2. Construction of a user query for ontology-based search (step 1).

- songs devoted to (or mentioning) historic events or important social phenomena;
- songs in which exciting natural or astronomical phenomena are described or mentioned;
- songs in which typical (or typical for a certain region) folk beliefs are described;
- songs in which elements of country work and life are described or mentioned;
- songs in which significant family events or human relations are mentioned.

The search engine provides also some facilities for processing of user queries presuming examination of equality or inequality. For example, it is possible to formulate and execute queries for search of:

- songs performed alone/in a group;
- songs performed by men/women only;
- songs performed by one and the same singer;
- songs performed by singers, born in one and the same settlement or region;
- songs performed in a specific region (grouped by regions of performance);

- songs performed in settlements to the west/east/north/south of a specific settlement/region.

The current version of the search engine of DjDL is provided by a very small number of concept search patterns, therefore their application is still quite limited. When appropriate, the pattern matcher performs the last step of the execution of queries for ontology-based search.

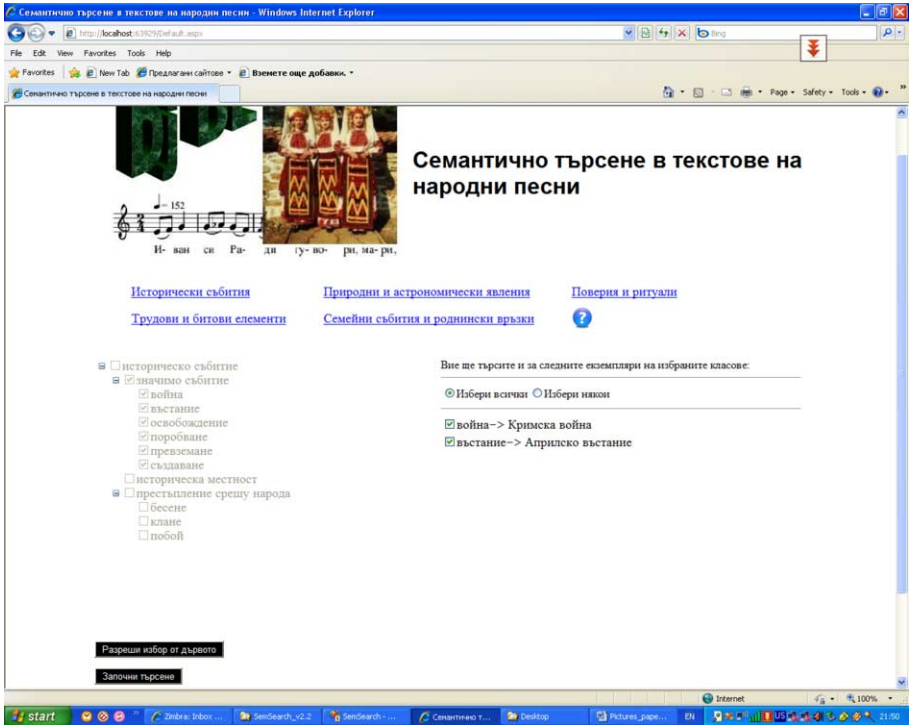


Figure 3. Construction of a user query for ontology-based search (step 2).

## 5. Implementation

For the development of the subject ontology we used one of the most popular ontology editors – Protégé-OWL [12]. The search engine has been implemented using the following technologies and tools:

- ASP.NET 3.5 technology;
- C# programming language;
- IDE Visual Studio 2008.

The implementation of the basic component of the search engine which realizes semantics oriented (ontology-based) search in the lyrics of songs works on two main classes defined for the purpose:

- the *OntologySearcher* class searches the ontology for primitive and defined classes that match the original user query as well as for instances of these classes. By doing so, it implements the augmentation of the user query;

- the XMLSearcher class finds the files with catalogue descriptions in a given folder which have "folk\_song\_text" nodes with values containing at least one word from a given list. It also returns a dictionary of the discovered incorrect catalogue files.

## 6. An Example

Let us suppose for example that the user defines a query for ontology-based search in the lyrics of songs which concerns the concept “historic event” (“значимо историческо събитие” in Bulgarian). During the execution of this query, first of all it is augmented and refined with the assistance of the user as shown on Figure 2 and Figure 3. Then a consecutive search in the catalogue descriptions of songs follows. As a result, all documents with <folk\_song\_text> element values containing phrases that are juxtaposed with at least one element of the augmented user query, are extracted. A list with the titles of the discovered songs is properly visualized on the user screen (see Figure 4).

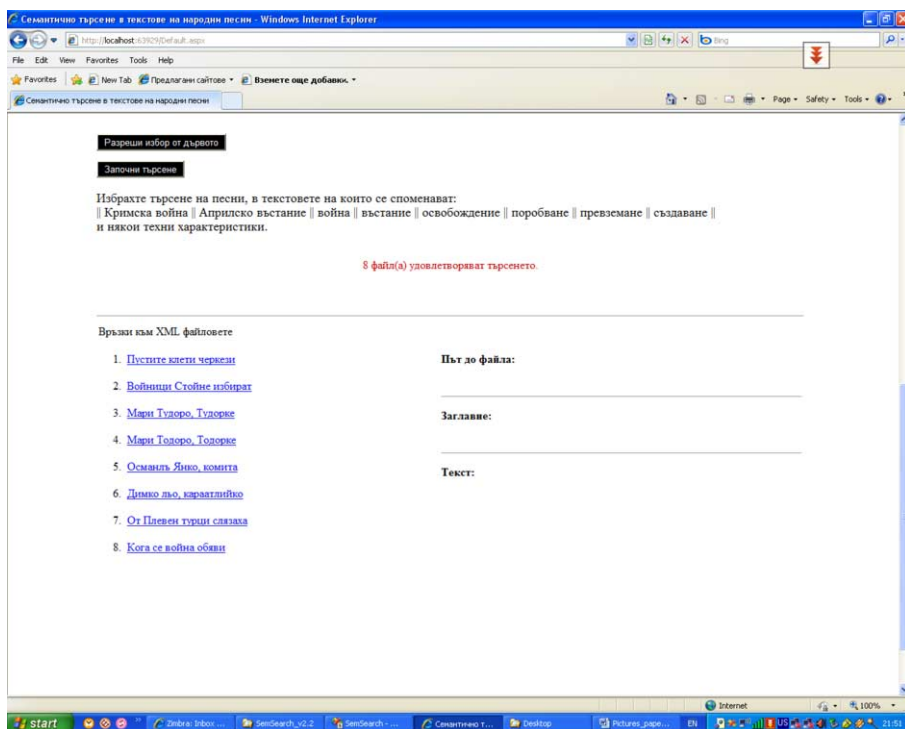


Figure 4. Search results for a user query containing the phrase “historic event” (level 1).

When the user clicks on the name of a particular song satisfying the augmented query, the text of this song is displayed in a new window. The discovered words and

phrases that match (are semantically related to) the query, are highlighted. Figure 5 shows some results for the user query example containing the phrase “historic event”.

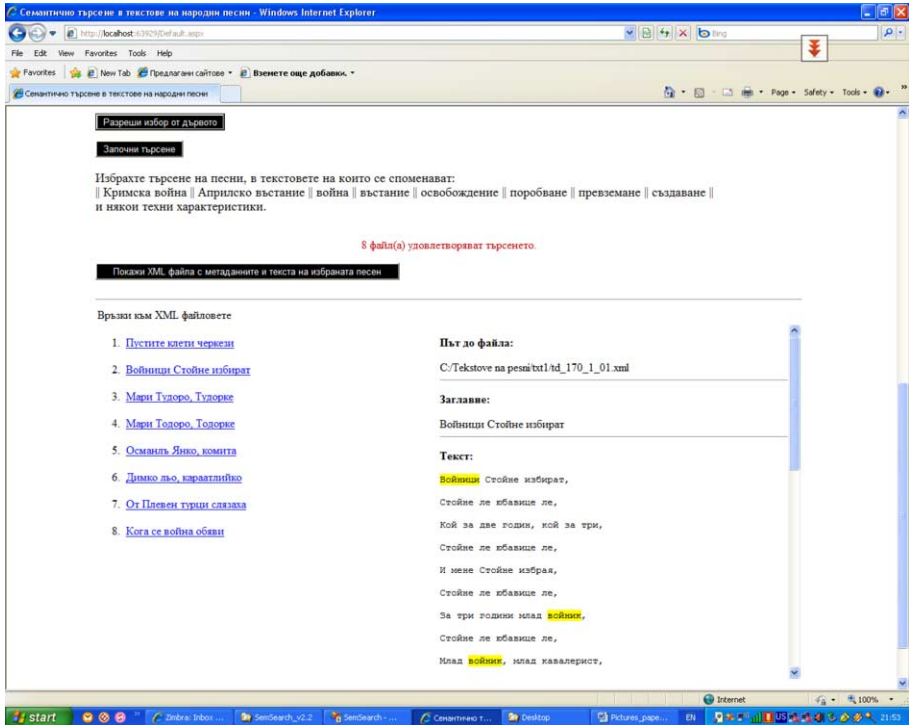


Figure 5. Search results for a user query containing the phrase “historic event” (level 2).

## 7. Conclusion

The working principles of the search engine of DjDL are designed in order to support its flexibility, interoperability and reusability.

Our current activities are directed to:

- evaluating the performance of the search engine (computing a tentative value of its average precision);
- design and implementation of a tool for flexible and convenient (intuitive) construction of complex user queries without using natural language and proper modification of the search algorithm.

The next step will be to extend the functional facilities of the search engine with a proper tool for semantic search and knowledge discovery in the notes of songs. A main goal in this direction will be to automate the further study of some musical characteristics of Bulgarian folk songs (e.g., their melodies and rhythms) with the aim to discover similarities of songs according to various criteria.

In this way the final version of DjDL will be developed with the aim to provide a complete set of tools which will be useful for a series of further studies in folkloristics, philology and musicology.

**Acknowledgements.** This work has been supported by the Bulgarian National Science Fund within a project titled “Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library”, Grant No. DTK 02/54/17.12.2009. The authors are thankful to their student Dicho Shukerov for his contribution to the implementation of the prototype of the search engine of DjDL.

## References

- [1] R. Guha, R. McCool, E. Miller, Semantic Search. In: *Proceedings of the 12th International World Wide Web Conference (WWW 2003)*, Budapest, Hungary, 2003, 700-709.
- [2] S. Kruk et al., JeromeDL – a Semantic Digital Library. In: J. Golbeck, P. Mika (Eds.), *Proceedings of the Semantic Web Challenge Workshop at ISWC2007*, Busan, South Korea, 2007.
- [3] L. Castro, O. Giraldo, A. Castro, Using the Annotation Ontology in Semantic Digital Libraries. In: A. Polleres, H. Chen (Eds.), *Proceedings of the ISWC 2010 Posters & Demonstrations Track*, Shanghai, China, 2010.
- [4] S. Kruk et al., The Role of Ontologies in Semantic Digital Libraries. *10<sup>th</sup> European Conference on Research and Advanced Technology for Digital Libraries*, Alicante, Spain, 2006.
- [5] L. Peycheva, N. Kirov, Bulgarian Folk Songs in a Digital Library. In: R. Pavlov, P. Stanchev (Eds.), *“Digital Preservation and Presentation of Cultural and Scientific Heritage. International Conference, Veliko Tarnovo, Bulgaria, 11-14 September, 2011”*, ISSN 1314-4006, IMI – BAS, Sofia, 2011, 60-68.
- [6] M. Nisheva-Pavlova, P. Pavlov, Semantic Search in a Digital Library with Bulgarian Folk Songs. In: Y. Tonta et al. (Eds.), *“Digital Publishing and Mobile Technologies. 15th International Conference on Electronic Publishing, June 22-24, 2011, Istanbul, Turkey”*, ISBN 978-975-491-320-0, Hacettepe University, Ankara, 2011, 103-109.
- [7] N. Kirov, Digitization of Bulgarian Folk Songs with Music, Notes and Text. *Review of the National Center for Digitization* **18** (2011), 35-41.
- [8] M. Nisheva-Pavlova, Providing and Maintaining Interoperability in Digital Library Systems. In: *Proceedings of the Fourth International Conference on Information Systems and Grid Technologies (Sofia, May 28-29, 2010)*, ISBN 978-954-07-3168-1, St. Kliment Ohridski University Press, 2010, 200-208.
- [9] M. Nisheva-Pavlova, P. Pavlov, Search Engine in a Class of Academic Digital Libraries. In: T. Hedlund, Y. Tonta (Eds.), *“Publishing in the Networked World: Transforming the Nature of Communication. 14<sup>th</sup> International Conference on Electronic Publishing, 16-18 June 2010, Helsinki, Finland”*, ISBN 978-952-232-085-8, Edita Prima Ltd, Helsinki, 2010, 45-56.
- [10] P. de Juan, C. Iglesias, Improving Searchability of a Music Digital Library with Semantic Web Technologies. In: *Proceedings of the 21st International Conference on Software Engineering & Knowledge Engineering (SEKE'2009)*, ISBN 1-891706-24-1, Knowledge Systems Institute Graduate School, Boston, Massachusetts, 2009, 246-251.
- [11] L. Peycheva, G. Grigorov, How to Digitalize the Folk Song Archives? *Review of the National Center for Digitization* **18** (2011), 42-58.
- [12] H. Knublauch, R. Ferguson, N. Noy, M. Musen, The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. *Third International Semantic Web Conference*, Hiroshima, Japan, 2004, 229-243.

# Knowledge Network of Scientific Claims Derived from a Semantic Publication System

Carlos Henrique MARCONDES<sup>a,1</sup>

<sup>a</sup>*University Federal Fluminense, Department of Information Science, R. Lara Vilela, 126,  
24210-590, Niterói, Rio de Janeiro, Brazil, phone 55 21 26299758, marcon@vm.uff.br*

**Abstract.** Currently, the conventional communication channel for reporting scientific results is Web electronic publishing of scientific articles in paper print formats, such as PDFs. The emergence of the Semantic Web and Linked Data environment provides new opportunities for communicating, sharing, and integrating scientific knowledge in digital formats that could overcome the limitations of the current print format, which is only suitable for reading by people. The results of scientific research can be published electronically and shared in structured, interlinked formats. This integrated knowledge network could be crawled by software agents, thereby facilitating semantic retrieval, knowledge reuse, validation of scientific results, identification of traces of scientific discoveries, new scientific insights, and identification of knowledge contradictions or inconsistencies. This paper explores the possibilities of this new environment for scientific publishing and reports the implementation of a prototype semantic publishing system, which publishes scientific articles in a paper print format and publishes the claims made in the conclusions of each article as structured triples using the Resource Description Framework format.

**Keywords:** semantic publishing, semantic web, linked data, knowledge networks, knowledge representation, terminological knowledge bases, e-science

## Introduction

The Web is fast becoming a universal platform for the disposal, exchange, and access of knowledge records. An increasing amount of human cultural records, derived from text, static or motion images, sound, and multimedia, are now being created directly in a digital format. Scientific activity has always involved the intensive processing of data and information. It is also being affected by the emergence of Web facilities for Grid technologies that can process vast amounts of scientific data, as well as the publication and maintenance on the Web of full text, datasets, and algorithms used for scientific data processing.

The current scholarly Web publishing environment is still an electronic metaphor of the paper print publishing environment used throughout the twentieth century. Despite numerous advances in information technology, Web electronic publishing is

---

<sup>1</sup> Corresponding Author: Carlos Henrique Marcondes, Department of Information Science, R. Lara Vilela, 126, 24210-590, Niterói, Rio de Janeiro, Brazil; phone 55 21 26299758; E-mail: marcon@vm.uff.br.

still based on the print text model. Scientific results are traditionally published in articles with a textual format, which limits the possibilities for the reuse and validation of scientific results that are published in the Web environment. Reuse requires the identification of similar patterns and the ability to compare similar parameters and differences in vast amounts of data. Validation requires reproducible methods, which can only be achieved if the same data processes, and algorithms are in the original experiments.

The reuse and validation of scientific results demands tools for information discovery, retrieval, and comparison in a very specific, precise, and meaningful manner. Current information retrieval systems do not contain explicit meaningful relations between elements, the content of documents, or resources they represent. Boolean operators are too general and they lack the semantic expressiveness necessary for content retrieval in specific scientific domains. Relations expressed using Boolean operators are processed as extensive set operations using the keywords included in the bibliographic records, rather than as intensive semantic relations among concepts.

Scientists now have to search through a vast amount of information resources that may be of potential interest, which is available on the Web. These resources range from electronic publications, digital libraries, repositories of full-text papers, algorithms, datasets of scientific data, terminological knowledge bases, and virtual machines that can process these data. Scientists have to devote greater efforts to discovering, examining, comparing, and integrating these resources [1]. The Web provides a potentially integrated and comprehensive environment for scientific activities, including scientific publication, but the textual format used for publishing results prevents computer programs from directly processing the content of scientific articles, which would facilitate scientific knowledge management tasks.

The Web environment now provides an opportunity to publish and integrate all these elements. The challenges posed by this scientific environment demands the development of methods and technologies for the direct processing of knowledge by computers. Scientific terminologies are one of the most essential components of scientific communication, and they are evolving towards more formal knowledge bases [2] that can be processed by computers. Comprehensive biomedical terminologies include the Unified Medical Language System (UMLS) and the Semantic Network (SN)<sup>2</sup>, which is a classification schema of the UMLS Meta-thesaurus that organizes every concept into hierarchical trees, where each has as its root in a top level Semantic Type. The UMLS SN uses 54 Relation Types to express the semantic relations among concepts in the Semantic Type hierarchies used to index Biomedical scientific articles. The UMLS SN contains the permitted relations among Semantic Types. According to National Library of Medicine, USA, UMLS Fact Sheet [3]: *“The purpose of NLM's Unified Medical Language System (UMLS®) is to facilitate the development of computer systems that behave as if they understand the meaning of the language of biomedicine and health”*.

In particular, the Semantic Web and Linked Data technologies [5] provide new opportunities for communicating, sharing, reuse, interlinking, and integrating scientific knowledge published in digital formats that may overcome the limits of the current print format used for the publication of scientific results, which is only suitable for reading and processing by people. We are now beginning to use these technologies for sophisticated tasks such as knowledge discovery, knowledge comparison, and

---

<sup>2</sup>UMLS Semantic Network, <http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html>

integrating multiple sources, which facilitates inference capabilities among different and autonomous information resources.

Scientific publications were first recorded in bibliographic databases and citation analysis, citation models, and citation networks were developed as tools to understand and manage the development of science [5]. The Web offers the possibility of developing a richer and more multifaceted scientific knowledge environment where navigating throughout a citation network will be only one of the many possibilities.

This paper addresses the integration of a scholarly semantic publishing environment with the future e-science environment, which is being designed based on the Semantic Web and Linked Data technologies. Specifically, we address the question of how to identify, extract, and represent the knowledge embedded in the text of Web published scientific articles in a machine processable format in compliance with Semantic Web standards. The paper is structured as follows. The next section outlines the proposed semantic publication model and reports the development of a Web author submission interface to a journal system, which partially implements it. Section 2 presents the proposed semantic record model and discusses its implementation in the Resource Description Framework (RDF). Section 3 discusses how the model proposed may provide the core of a knowledge network, thereby facilitating a range of scientific knowledge management methods, presenting conclusions, and future research developments.

## 1. A Semantic Model for Scholarly Electronic Publishing

In a previous study, we proposed [6] a semantic model for electronic publishing. The model was developed through the analysis of scientific articles in the area of biomedical science, where we identified patterns of reasoning and semantic elements. The aim of the model was to achieve a semantically richer content for the representation of biomedical articles in a computer program “understandable” format. This proposal is based on the following hypotheses: (a) scientific knowledge that appears in the text of scientific articles consists of scientific conclusions made by authors and takes the form of relations between phenomena; (b) these relations are expressed linguistically using propositions and related concepts. Thus, it is feasible to use an authoring/Web publishing tool to ask authors to enter a conclusion and perform natural language processing (NLP) of the conclusion text to identify, extract, and represent this knowledge in a structured format.

This knowledge representation format allows programs to perform “inferences” based on the knowledge content of articles, thereby facilitating more semantically powerful content retrieval and knowledge management compared with current Bibliographic Information Retrieval Systems.

The proposed model has two components: an enhanced record model, i.e., a semantic record, and a Web interface allowing authors to self-publish and self-submit articles to a journal system. The record model *extends* conventional bibliographic record models, which comprise of conventional descriptive elements such as authors, title, abstract, bibliographic source, publication date, and content information, such as keywords, descriptors, and references to cited papers. In addition to these elements, the model includes claims made by authors in the conclusions of their articles. Miller [7] states that “science is a search after internal relations between phenomena”. Scientific knowledge found in the text of scientific articles consists of scientific claims made by



authors throughout the article text, which is synthesized in the article's conclusion. It takes the form of relations between phenomena or between a phenomenon and its characteristics. These relations are expressed linguistically using propositions that relate concepts. Scientific claims are represented as *relations* between two different phenomena or between a phenomenon and its characteristics [8], e.g., (a) “telomere shortening (Phenomenon) causes (Type\_of\_relation) cellular senescence (Phenomenon)” [9], (b) “telomere replication (Phenomenon) involves (Type\_of\_relation) nontemplate addition of telomeric repeats onto the ends of chromosomes(Phenomenon)?” [10], or (c) “tetrahymena extracts (Phenomenon) show (Type\_of\_relation) a specific telomere tranferase activity (Characteristic)” [11]. Such relations could be modelled as triples of <Antecedent><Type\_of\_relation><Consequent>.

Our research also includes the development of a prototype Web author submission interface for a journal system, which partially implements the model [12] where a general framework is proposed for identifying discoveries in scientific papers based on two aspects: their rhetoric elements and patterns and by comparing the content of the article conclusions with terminological knowledge bases [13].

In the proposed model, authors use a journal system Web submission interface to type the article conclusions and other standard metadata during the submission/upload of their article text. The system performs NLP of the text of the conclusion, before formatting it as a relation. Thus, we propose to engage authors in the development of a richer content representation of their own articles. The system interacts with authors and asks them to validate the extracted relations while the system of concepts found in the conclusion is mapped to concepts in a domain terminological knowledge base.

The result of this processing are recorded as a richer semantic content bibliographic record where scientific claims made by authors throughout articles are expressed as relations. In addition to being published in a textual format, each article's claims are also represented as structured relations and recorded in a machine-understandable format using Semantic Web standards such as RDF [14] and OWL (Web Ontology Language) [15], allowing the claims to be formally related by the author, i.e., mapped and annotated to concepts in a domain terminological knowledge base.

The terminological knowledge base used by our prototype system is UMLS. The author is asked to validate the automatic mapping made by the system, selecting other terms from a list displayed by the system and deciding whether satisfactory mapping options are offered. If satisfactory options are not available, the system assigns this specific element of the relation to “no mapping.” The result is that the conclusion terms annotated by the author are related to terms in a terminological knowledge base *at the time of article publication*.

Once the relations are recorded in a database the machine-understandable records resulting from this publishing model can be processed and compared with public knowledge using software agents, e.g., published scientific articles, or with terminological knowledge bases throughout the Web. This provides scientists with new tools for knowledge retrieval, claims comparison, identification of contradictory claims, the use of these claims in different contexts, and the identification and validation of new contributions to science made in specific articles.

The proposed model is described in detail in [6].

A prototype of the submission system was developed to evaluate the dialogue with authors and the extraction routine. The prototype of the interface is in its initial

development phase. It is not yet a final version or a production system. It has been developed with the specific aim of demonstrating the feasibility of the proposed system. It was developed as a Java application using MetaMap<sup>3</sup> program to perform NLP of article conclusions with LingPipe<sup>4</sup>.

The prototype system processes selected parts of the article text uploaded by authors, i.e., the title, abstract, keywords, introduction, methods, and results. The introduction and abstract are used to extract the objectives of the article by the identification of phrases such as the "*objectives of our work...*" and "*the goal of the present work...*" The author is asked to type the *conclusions* of the article that is being submitted. The extraction routine uses a formula, which is based on the frequency of occurrence of a term in the title, abstract, keywords, method, results, and objective, and this method weights the terms in the conclusion in order to format them as a relation. The syntactic components with higher weights in the conclusion are candidates for the Antecedent and Consequent of the relation.

After the author validates the relation, the system records it as a record using the format shown in Figure 4, together with the conventional bibliographic metadata and the article full text. In the future, we plan to integrate this prototype with the PKP Open Journal System<sup>5</sup>, which is an electronic journal system that is largely used in Brazil. In its current implementation, the prototype processes only the article conclusion.

Some of the steps involved in the processing of the conclusion, i.e., "*The results presented herein emphasize the importance to accomplish systematic serological screening during pregnancy in order to prevent the occurrence of elevated number of infants with congenital toxoplasmosis*" [16], are shown in the following Figures.

**Indicate the Conclusion**

Write the conclusion briefly below.

- The conclusion should provide a comprehensive summary (less than 50 words).
- The conclusion should clearly answer the questions posed if applicable.
- The conclusion should not introduce any information or ideas yet described in your article.
- **If it exists several conclusions the main it should be chosen**
- Provide the conclusion which was only directly supported by the results.
- **Avoid speculation, overgeneralization, supposition and don't create a hypothesis.**
- Avoid sentences among commas and parentheses.
- Avoid explanations between commas and parentheses.
- Describe the main finding only. **Ideally, it should be only one sentence in length (less than 50 words).**

the results presented herein emphasize the importance to accomplish systematic serological screening programs during pregnancy in order to prevent the occurrence of elevated number of infants with congenital toxoplasmosis.

**Figure 1.** The author specifies the article conclusion.

<sup>3</sup>MetaMap, <http://mmtx.nlm.nih.gov/>.

<sup>4</sup>Lingpipe home, <http://alias-i.com/lingpipe/>.

<sup>5</sup>PKP Open Journal System, <http://pkp.sfu.ca/>

**Make The Relation**

Fill in the boxes below according to summarized idea based on your paper's conclusion, like as relation e.g. "HPV (Antecedent) causes (Verb) neoplastic cervical lesions (Consequent)"

**Conclusion:** the results presented herein emphasize the importance to accomplish systematic serological screening programs during pregnancy in order to prevent the occurrence of elevated number of infants with congenital toxoplasmosis.

Choose an option for the relationship or type a verb

prevent

happen

Type a verb

Antecedent

systematic serological screening programs during pregnancy

Relation

prevent

Consequent

elevated number of infants with congenital toxoplasmosis

Choose the option for antecedent or type one

systematic serological screening programs during pregnancy

Not the option above - type the antecedent

Choose the option for consequent or type one

elevated number of infants with congenital toxoplasmosis

Not the option above - type the consequent

Continue ...

Figure 2. The article conclusion is formatted as a relation.

**Indicate The Concepts**

Choose, if possible, the concepts related to each part of the relationship. More than one concept can be chosen for each part. Don't mark any of the options in case the concept is not directly related.

**Conclusion:** the results presented herein emphasize the importance to accomplish systematic serological screening programs during pregnancy in order to prevent the occurrence of elevated number of infants with congenital toxoplasmosis.

Choose an option for the relationship

prevent is...

Stops, hinders or eliminates an action or condition.

any previous one

Antecedent

systematic serological screening programs during pregnancy

Relation

prevent

Consequent

elevated number of infants with congenital toxoplasmosis

Choose the concepts related to the Antecedent

systematic - Functional Concept

Serologic - Functional Concept

Aspects of disease screening - Functional Concept

Programs (Publication Type) - Intellectual Product

Screening - procedure intent - Functional Concept

Screening procedure - Health Care Activity

Special screening finding - Finding

Pregnancy - Organism Function

Choose the concepts related to the Consequent

High - Qualitative Concept

Count of entities - Quantitative Concept

MDF Attribute Type - Number - Idea or Concept

Numbers - Quantitative Concept

Infant - Age Group

Toxoplasmosis, Congenital - Disease or Syndrome

Continue ...

Figure 3. Authors are asked to map/annotate concepts in the article's conclusion to UMLS terms.

## 2. A Semantic Record Model in RDF for Scientific Articles

A key component in the scope of the proposed publication model is a *semantic record model* for scientific articles that focuses on the semantic elements of the scientific methodology in terms of scientific reasoning, e.g., questions, hypotheses, experiments, and conclusions. These elements are related to the knowledge content and reasoning patterns of an article, which this study aims to identify and record in a machine-processable format. The *Conclusion* is an essential semantic element that synthesizes the knowledge content of an article. In the scope of a recently published article, this is provisional knowledge, although it is verified by the experiment reported in the article. Semantic elements such as Questions and Hypothesis are also important because they enable researchers to trace the evolution of a research question and its resolution in a paper. Other elements have rhetoric functions, as extensively discussed in [17] and [18], or they may serve to describe methodological options more clearly, such as the experiment performed, its context, or the results obtained.

Thus, relations are the core of the proposed knowledge representation scheme. A relation has the form of an Antecedent (a concept referring to a phenomenon), a Semantic Relation, and a Consequent (a concept referring to another phenomenon or a characteristic of the phenomenon in the Antecedent). A Semantic Relation may be a specific *Type\_of\_relation* such as “causes,” “affects,” or “indicates,” or a “has\_as\_characteristic” relation.

Relations may also appear in different semantic elements throughout the article text, such as within the Problem that the article addresses as a *Question* where either one of the two *relations* or the type of relation is unknown, in the *Hypothesis*, or in the *Conclusion*. Frequently, the Conclusion also poses new Questions.

The implementation of a model described using an electronic journal systems submission interface poses several challenges, e.g., how to represent the model, even partially, in a machine-understandable format, and how to extract and format a relation from the text of the article conclusion. We began to address these challenges during the development of the prototype.

The following figure considered the conclusion “*telomere replication (Antecedent) involves (Type\_of\_relation) a terminal transferase-like activity (Consequent),*” found in [19], which has been formatted in RDF.

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:sa="http://example.org/semarticles/"
  xmlns:umls="http://www.nlm.nih.gov/research/umls/">
  <rdf:Description rdf:about="http://art_id">
    <dc:title>title</dc:title>
    <dc:creator>creator</dc:creator>
    <dc:subject>subject</dc:subject>
    <dc:date.published>date</dc:date.published>
    <sa:conclusion>
      <rdf:Description rdf:about="http://art_id/conclusion">
        <sa:antecedent content="telomere
replication">http://www.nlm.nih.gov/research/umls/CUI01</sa:antecedent>
        <sa:type_rel
content="involves">http://www.nlm.nih.gov/research/umls/CUI02</sa:type_rel>
        <sa:consequent conten="a terminal transferase-like activity">
http://www.nlm.nih.gov/research/umls/CUI03</sa:consequent>
      </rdf:Description>
    </sa:conclusion>
  </rdf:Description>
</rdf:RDF>

```

Figure 4. Conclusion of article, represented in RDF. CUI means UMLS’s concept unique identifier.

### 3. Knowledge Network Derivation from a Scholarly Semantic Publishing Environment

The *Conclusion* is an essential semantic element of the article text that synthesizes its knowledge content. Within the scientific community, the agreed mechanism for validating scientific results is the number of citations an article receives. Thus, the Conclusion represents provisional knowledge in recently submitted/published articles without any citations. However, these results are verified by the experiments they report and by the peer review process performed by the journal editorial board that approved the article's publication.

The enhanced bibliographic records found in the proposed model are nodes in a knowledge network with links to full-text articles, conventional bibliographic metadata, the bibliographic references cited, and the article conclusions formatted as RDF triples. Authors are also asked to annotate terms in the conclusion text using terms in a domain terminological knowledge base. Bermes [23] stressed the important role of terminological knowledge bases in a Linked Data environment, because they function as “hubs” that connect resources from different domains. All of these elements may also be identified using URIs, thereby facilitating semantic navigation by scholars, as suggested in the Linked Data proposal [4].

Figure 5 illustrates this process.

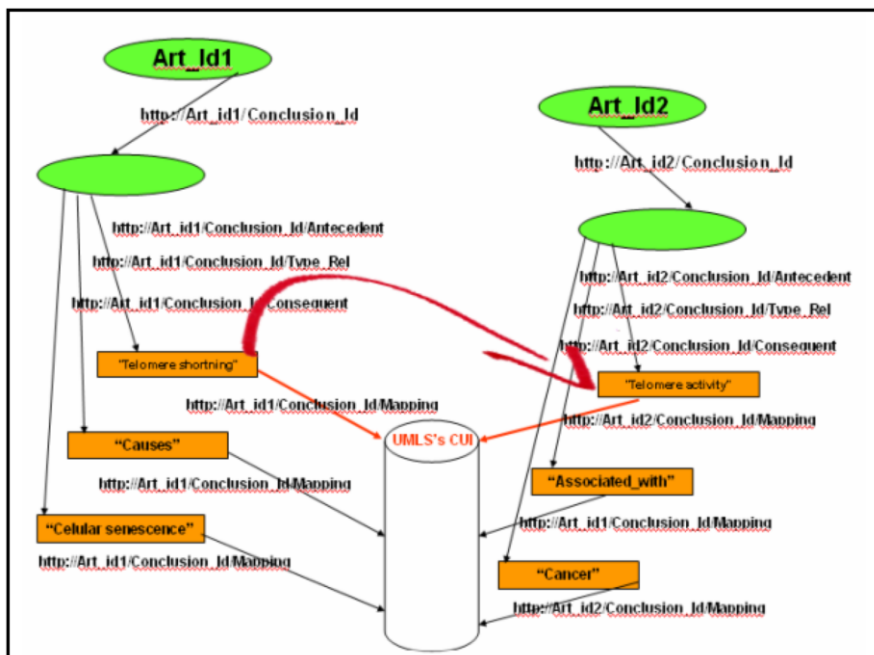


Figure 5. Two articles with related claims connected by a UMLS Semantic Type.

The example shown in Figure 5 is based on a free interpretation of the two claims found in following statement: “It has been proposed that the finite cell division capacity of human somatic cells is limited by telomere length. This is consistent with reports that telomerase activity is often high in cancer and immortalized tissue culture cells” [20].

This example illustrates the situation where two articles have related claims. The first article's claim states that "telomere shortening causes cellular senescence," whereas the second article states that "telomerase activity is associated with cancer." The concepts "telomere shortening" and "telomerase activity" are both mapped, i.e., linked, to the same concept in UMLS, which is identified by its Concept Unique Identifier (CUI) as "telomerase activity," which can function as a URI.

Even a partial implementation of the record model proposed in RDF, where the only semantic element captured is the *conclusion*, will facilitate more expressive semantic retrieval from a knowledge network using SPARQL, as shown in Figure 5 [21].

In this example, a software agent might infer a new claim, i.e., that (maybe) "telomere shortening" "is associated with" "cancer". The claim can only be trusted based on the evidence presented in the experiments described in both articles and by the judgement of journal referees, who certified that both articles had sufficient scientific quality to merit publication. The result is a knowledge network of *claims* extracted from refereed scientific articles, which are coded in machine-readable format, and linked to *terminological knowledge bases* [23] and to *citing/cited articles*. Such a network overcomes the problems of conventional citation networks [5] when analyzing scientific developments and managing scientific knowledge. Semantic Web technologies are increasingly used in electronic publishing environments to enhance citation analysis [22], so it is reasonable to expect that a citation network would merge from the proposed generalized knowledge network.

Software agents can navigate through this richer network and perform "inferences" that facilitate sophisticated tasks such as hypothesis formulation, hypothesis comparison, knowledge discovery, and the identification of traces of scientific discoveries and knowledge misunderstandings.

For example, if CUI01 is the UMLS CUI for the relation "Functionally\_related\_to" and CUI02 is the UMLS CUI for the concept "Telomerase activity." Then a knowledge network, such as that shown in Figure 5, enables SPARQL queries, as follows:

```
@PREFIX sa: <http://example.org/semarticles/>
SELECT ?Consequent
WHERE {type_rel/mapping="CUI01"}
AND    antecedent/mapping="CUI02"}
```

We have also shown elsewhere [13] that our proposed semantic publication model facilitates the identification of traces of scientific discoveries. We hypothesize that there is a correlation between the article content and the fact that these articles report scientific discoveries. Based on this approach, we propose a new scientific discovery indicator that is different from bibliographic/citation indicators. Since there is a structural delay between the publication date of an article and the date of citation, this new indicator stresses the importance of such metadata elements for the proposed schema. The proposed approach is based on a content comparison of the conclusions of scientific articles and biomedical terminological knowledge bases, to verify that article conclusion terms can be mapped to terms in a terminological knowledge base. Evidence was found that articles, where just a few terms in the conclusions or even none were mapped or just mapped at a generic level to terms in biomedical terminologies such as UMLS, indicated new discoveries.

We envisage other applications that could enhance current literature-based discovery methods [24], [25] by comparing indirect indicators such as citations or use

the same content descriptors with the claims made in articles. The proposed semantic publication model may be part of a future wholly integrated e-science environment.

## References

- [1] A.H. Renear, C.L. Palmer, Strategic reading, ontologies and the future of scientific publishing, *Science* 325, (2009), 828-832.
- [2] G.C. Marion, An Expanded Review of Information-System Terminology, Proceedings of AFCEA 's Sixteenth Annual Federal Database Colloquium and Exposition, "Information Dominance and Assurance" San Diego, CA, September (1999), 21-23.
- [3] Unified Medical Language System Fact Sheet, <http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html>
- [4] C. Bizer, T. Heath, T. Berners-Lee, Linked data – the story so far, In: T. Heath, M. Hepp, C. Bizer (eds.), Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS).
- [5] E. Garfield, I.H. Sher, R.J. Torpie, The Use of Citation Data in Writing the History of Science, The Institute for Scientific Information, Philadelphia, (1964).
- [6] C.H. Marcondes. A semantic model for scholarly electronic publishing. In: International Workshop on Semantic Publication - SePublica2011-, 1st, at the Extended Semantic Web Conference (ESWC), 8th, in Hersonissos, Crete, Greece, Proceedings... CEUR Workshop Proceedings, 721, (2011). ISSN: 1613-0073.
- [7] D.L. Miller, Explanation Versus Description, *Philosophical Review* 56 (3) (1947) 306-312.
- [8] Ingetraut Dahlberg. Conceptual structures and systematization. International Forum on Information and Documentation 20, (3), (1995) 9-24.
- [9] L.Y. Hao et al. Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. *Cell* 123 (2005) 1121–1131.
- [10] J. Shampay, J.W. Szostak, E.H. Blackburn. DNA sequences of telomeres maintained in yeast. *Nature* 310, (1984) 154-157.
- [11] C. W. Greider, E. H. Blackburn. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, 43 (1985) 405-413.
- [12] L.C. da Costa, *Um proposta de processo de submissão de artigos científicos à publicações eletrônicas semânticas em Ciências Biomédicas*, Tese (doutorado), Programa de Pós-graduação em Ciência da Informação UFF-IBICT. Niterói, (2010).
- [13] C.H. Marcondes, L.R. Malheiros. Identifying traces scientific discoveries by comparing the content of articles in biomedical sciences with web ontologies. In: 12 ISSI - International Conference on Informetrics and Scientometrics, 2009, UFRJ, Rio de Janeiro, v. 1, *Proceedings...* São Paulo, BIREME/PAHO/WHO, (2009), 173-177.
- [14] RDF, <http://www.w3.org/TR/rdf-primer/>
- [15] OWL Ontology Web Language Overview, <http://www.w3.org/TR/owl-features/>.
- [16] G.R.S. Segundo *et al.* A comparative study of congenital toxoplasmosis between public and private hospitals from Uberlândia, MG, Brazil. *Mem. Inst. Oswaldo Cruz.* [online], 99 (1) (2004) 13-17.
- [17] J. Skelton. Analysis of the structure of original research papers: an aid to writing original papers for publication. *British Journal of General Practice*, 44, (1994), 455-459.
- [18] Kevin Ngozi Nwogu. The Medical Research Paper: Structure and Functions. *English for Specific Purposes*, 16 (2) (1997) 119-138.
- [19] C.W. Greider, E.H. Blackburn. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* 51, (1987) 887--898.
- [20] M.J. McEachern, E.H. Blackburn, Runaway telomere elongation cause by telomerase RNA mutations, *Nature* 376 (1995), 403-409.
- [21] SPARQL Query Language for RDF (2008), <http://www.w3.org/TR/rdf-sparql-query/>
- [22] The citation ontology, CITO, <http://speroni.web.cs.unibo.it/cgi-bin/lode/req.py?req=http://purl.org/spar/cito>
- [23] E. Bernes. Convergence and Interoperability: a Linked Data perspective. In: IFLA World Library and Information Congress, 77<sup>th</sup>. Puerto Rico, 2011, *Proceedings.* (2011)
- [24] D.R. Swason, N. R Smalheiser, V. I. Torvik, Ranking indirect connections in literature based discovery. The role of Medical Subject Headings, *Journal of the American Society for Information Science and Technology* 57 (11) (2006), 1427–1439.
- [25] R.N. Kostoff, M.B. Briggs, J.L. Solka, R.L. Rushenberg, Literature-related discovery (LRD): Methodology, *Technological Forecasting and Social Change*, 75 (2), (2008), 165–185.

# WIKI::SCORE

## A Collaborative Environment For Music Transcription And Publishing

José João ALMEIDA <sup>a</sup>, Nuno Ramos CARVALHO <sup>a</sup> and José Nuno OLIVEIRA <sup>a</sup>

<sup>a</sup> e-mail: {jj,narcarvalho,jno}@di.uminho.pt

*University of Minho, Braga, Portugal*

**Abstract.** Music sources are most commonly shared in music scores scanned or printed on paper sheets. These artifacts are rich in information, but since they are images it is hard to re-use and share their content in today's digital world. There are modern languages that can be used to transcribe music sheets, this is still a time consuming task, because of the complexity involved in the process and the typical huge size of the original documents.

WIKI::SCORE is a collaborative environment where several people work together to transcribe music sheets to a shared medium, using the notation. This eases the process of transcribing huge documents, and stores the document in a well known notation, that can be used later on to publish the whole content in several formats, such as a PDF document, images or audio files for example.

**Keywords.** music transcription, Abc, wiki, collaborative work, music publishing

### Introduction

Music transcriptions are most electronically shared in pictorial formats, as pictures or scans (copyright permitting) of the original music scores. Although this information is semantically rich it is hard to re-use or transform in automatic and systematic ways, without human intervention. Imagine you have a scanned book, content in images, but if you want to count the number of words in the entire book a transcription of the content would prove much more useful and easy to process. This premise is also valid for music representations: a more textual or symbolic transcription of the content, while semantically as rich as the original, is easier to re-use, transform and process. Ultimately this representation is rich enough to produce the common pictorial formats with a high level of quality and accuracy.

The problem with this type of transcription is that it is harder, more complex and time consuming than simply scanning or photographing manuscripts (books or music scores). Distributed proofreading [18] has already proven that a community driven distributed proofreading approach can ease this task. Applying a map-reduce [6] philosophy, the original manuscript is partitioned in small chunks, and the work of transcribing and validating is performed on each chunk individually. Eventually all the processed chunks are compiled together to build the final transcription. From this source many documents



can be produced including images. Moreover, this plain text representation is easier to process and transform for performing any arbitrary task. The works of *Hoos et al* [12] on musical information retrieval, *Nienhuys et al* [20] on a system for automated music engraving for example, emphasize the advantages of having a more processable plain text notation. Such advantages are also defended and advocated in *Simões et al* [24] about automatic music classification using text mining techniques.

The platform described in this article supports a large scale, collaborative transcription of music scores into a plain text symbolic representation. This representation is easy to re-use, process and transform, and can be published in a heterogenous set of different formats, MIDIs, images or PDF documents for example. Section 1 introduces relevant work and concepts. Section 2 describes WIKI::SCORE in some detail. Section 3 describes the experimental usage of the platform in a real case study. Finally, in the last sections some conclusions are driven and some ways to improve this work are described.

## 1. Related Work

### 1.1. Music Writing Applications

Several music score authoring systems are already available, examples of the most popular being:

- **Sibelius** [5]  
This commercial desktop application provides a clear and intuitive interface to write music. It can be used to transcribe music scores, and with the use of plugins, export scores to some known formats. This application is widely used by music professionals.
- **Finale** [16]  
This is another commercial desktop application that can be used to write or transcribe music, also enjoying wide popularity.
- **Denemo** [2]  
Denemo is a free software (GPL) music notation editor available for many different platforms.
- **MuseScore** [23]  
is a free cross-platform music notation program that offers a cost-effective alternative to commercial programs such as some programs described earlier.

All these applications are useful for writing music but they lack direct support for distributed, collaborative work. But nothing prevents anyone for using them to create transcriptions written in well known notations and later sharing and combining them, as some of these tools export music scores to plain text representations.

### 1.2. The ABC Notation

The ABC [26,21,9] notation for music transcription is based on plain text. It is very compact and with many associated tools already available. It relies essentially on a direct representation of musical elements as ASCII characters making it economic and intuitive to use. This language has a simple and clean syntax, and is powerful enough to produce professional and complete music scores. Figure 1 is a small example of a source file



- easy to extend features and syntax via plugins;
- possible to customize via template engines;
- a wiki invites users to contribute to the content;
- wiki pages are subject to the phenomenon of *Darwikinism* [27], which typically provides for higher quality contents [13];
- the edited content is immediately available;
- built-in lock system to prevent content corruption.

In particular, DOKUWIKI [8] was adopted because of the following reasons, implements all the features described in the previous list:

- implements all the features described in the previous list;
- easy to install and maintain;
- the wiki content is stored in plain text files;
- it uses a clean plugin system that allows adding new features and syntax options;
- a plugin that adds ABC syntax to the wiki was already available;
- it is open source.

Wikis have already proven very useful and versatile in many scenarios, a good example being the well known Wikipedia [7], a free encyclopedia with almost 4 000 000 articles which anyone can edit or validate.

#### 1.4. Other Projects

A brief list of other interesting projects that in some way contributed to the development of this work follows:

- **Mutopia** [4]  
Mutopia is a LilyPond based music score archive. All the provided files are free to download, print out, perform and distribute. At the time of writing, Mutopia has 1710 musics for download from Beethoven to Joplin, from Harp to Piano or full Orchestra. This project clearly helps to show how sharing transcriptions in source textual formats is an advantage.
- **Noteflight** [15]  
Noteflight is an online application more in line with the work described in this paper although it lacks the collaborative driven philosophy. It allows to import and export from MusicXML.
- **Distributed Proofreaders** [3]  
Although not related with music this project is a success case of the distributed proof reading paradigm that this work is advocating and could well be adopted in collaborative music transcription.

## 2. The WIKI::SCORE Toolkit

WIKI::SCORE is a toolkit that provides a wiki with the required features and tools to aid in collaboratively transcribing and publishing music scores. The main advantage is to make use of the typical features already provided by a wiki in music transcribing tasks.

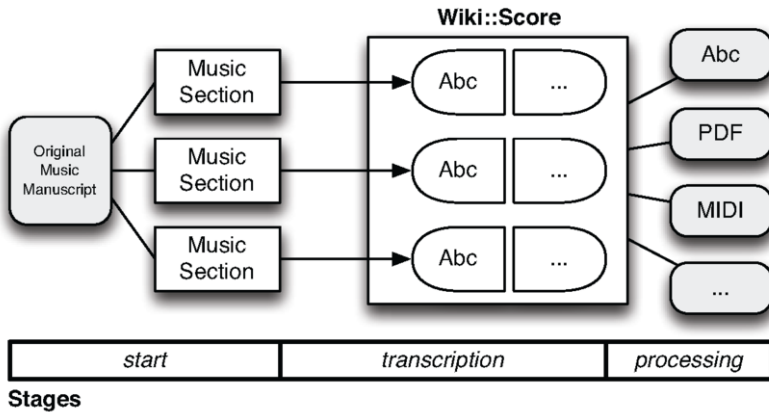


Figure 3. WIKI::SCORE architecture overview.

### 2.1. Workflow

This section describes the workflow used to process a music manuscript in the WIKI::SCORE environment. Figure 3 gives an a simplified overview of the entire process.

The main idea is that the overall process starts with an original manuscript, most commonly available in paper or digitally scanned, and ends with a complete transcription of the music to ABC notation. This transcription is also used to produce a MIDI audio file, and a PDF document which contains a high quality music score. The complete ABC source is also provided for anyone who needs to perform any other transformation, including data mining for analysis in musicology.

*Project Start.* The first stage when aiming to transcribe a music manuscript is to create a new project. The wiki application provides a particular page to perform this task. The process starts by providing some meta information and a name for the project, which can be updated later. The next step is to split the manuscript in sections and parts, letting the system then create all the required wiki pages, based on default templates. Automatically using default templates, the page for every section and part is created and became immediately accessible for users to edit.

This two-dimensional array of units (sections and parts) forms a *matrix* that initially has no ABC associated with which cell. (More details about this in the sequel.)

*Transcription Stage.* In this stage the music manuscript parts are transcribed by humans independently (one per wiki cell). This means that the corresponding ABC will be filled up in a cell-wise manner, starting from an initially empty matrix. Any cell (section or part) can be addressed individually and validated. Contributors are required only to access the cell they wish to transcribe and edit the corresponding wiki page, as every page is linked and easy to find.

During this stage validation tasks can be performed in parallel. These tasks can also take advantage from the wiki system, since comments and feedback can be added to individual sections' pages.

*Processing and Publishing Stage.* In this stage, that can be performed at any time, two functions are available for processing the matrix:

$$\begin{aligned} process &: Matrix \times Cut \times Template \longrightarrow Abc \\ publish &: Abc \times Format \longrightarrow File \end{aligned}$$

The *process* function is responsible for compiling an ABC document for a given project. To build this, the function operates over the project *Matrix*, and a *Cut* matrix that can be used to process a smaller subset of the entire project. This function also uses a *Template* that can change the final ABC generated. These files written in ABC can then be used by the *publish* function to produce a document in a given *Format*, for example an PDF or a MIDI file.

It is possible to publish the entire document or a small part of it, as this proves useful for proof-reading tasks for example. For this purpose, simple matrix operations are used. A row in this matrix is associated with a part (instrument, voice) and a column corresponds to a piece (eg. a movement of a symphony, the overture of an opera etc.) The matrix algebra is such that multiplying by 1 means selection and by 0 means discarding. As explained in Figure 4 the *Cut* matrix can be regarded as a pair of vectors  $v1$  (resp.  $v2$ ) selecting parts (resp. pieces). The matrix additive operation corresponds to concatenation. Equation (1) shows how to publish a single part just by multiplying the matrix with a vector which selects the first part (the only 1 in the vector).

$$(1\ 0\ \dots\ 0) \times \begin{pmatrix} Abc_{1,1} & Abc_{1,2} & \dots & Abc_{1,n} \\ Abc_{2,1} & Abc_{2,2} & \dots & Abc_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ Abc_{p,1} & Abc_{p,2} & \dots & Abc_{p,n} \end{pmatrix} = (Abc_{1,1}\ Abc_{1,2}\ \dots\ Abc_{1,n}) \quad (1)$$

Equation (2) shows the effect of selecting all parts. Each element in the resulting vector joins all the parts for each section, See Equation (3).

$$(1\ 1\ \dots\ 1) \times \begin{pmatrix} Abc_{1,1} & Abc_{1,2} & \dots & Abc_{1,n} \\ Abc_{2,1} & Abc_{2,2} & \dots & Abc_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ Abc_{p,1} & Abc_{m,2} & \dots & Abc_{p,n} \end{pmatrix} = (Section_1\ Section_2\ \dots\ Section_n) \quad (2)$$

$$Section_i = Abc_{1,i} + Abc_{2,i} + \dots + Abc_{p,i} \quad (3)$$

The conductor's score,  $Section_1 + \dots + Section_n$ , is obtained by multiplying  $(Section_1 \dots Section_n) \times (1 \dots 1)$ .

Using this approach, it is simple to process or publish the intended elements of the matrix. Figure 4 illustrates some other examples that can be used to produce typical scores.

## 2.2. Web Interface

A wiki is a web application. So any common browser or equivalent tool is enough for collaborating in WIKI::SCORE transcription tasks. Typically there is a wiki page for each project, which contains some meta information (original manuscript name, author, etc.) and links to other relevant pages. One important section in this index page is the matrix

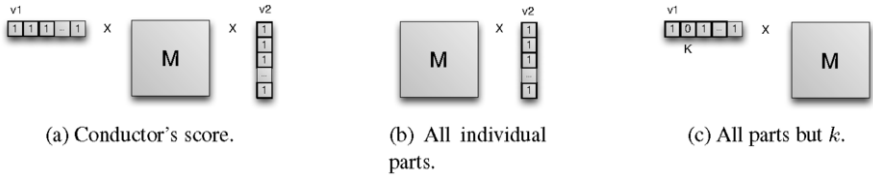


Figure 4. Matrix multiplying examples. Row vector  $v1$  works part-wise; column vector  $v2$  works piece-wise.

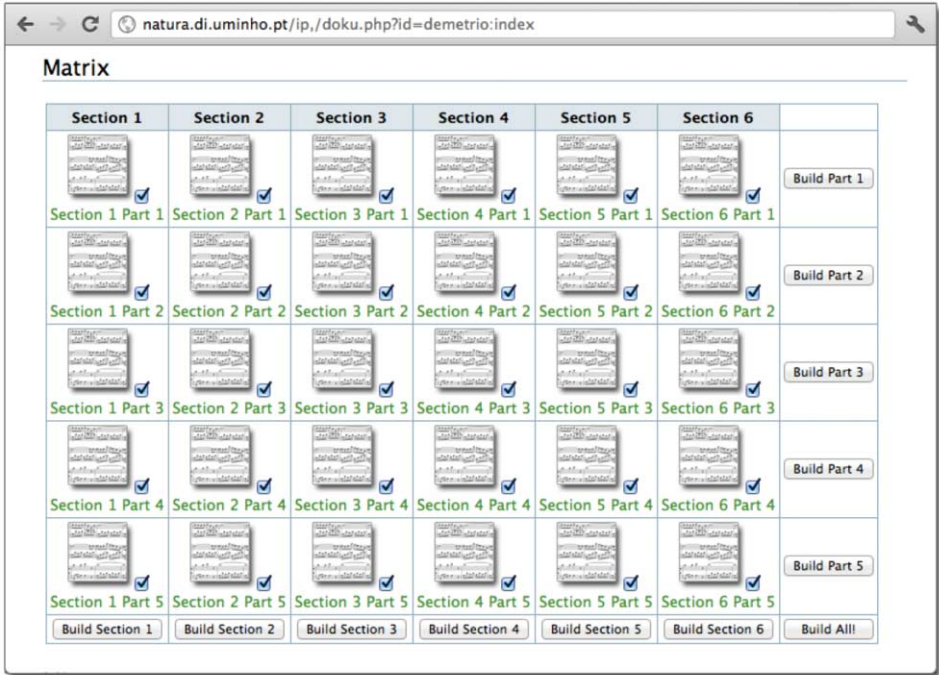


Figure 5. Matrix of sections and parts that make up the entire music. The "Build" buttons implement the matrix selection mechanism.

of links available to access each music section and voice. This matrix is illustrated in Figure 5.

From this matrix any contributor may access each cell in the matrix, corresponding to a particular part (eg. instrument) and section (eg. aria of an opera). There is a specific page for each cell where the ABC transcription can be made, and since this is a wiki much more information can be added, notes about validation and error issues for example. The edition page is illustrated in Figure 6.

The project index page provides links for documents published in several formats, built from the ABC notation:

- complete transcription in ABC format, compiled from all the transcriptions in the matrix;
- PDF document containing a high quality pictorial representation of the music score;
- a MIDI file that can be used to hear the transcribed music.

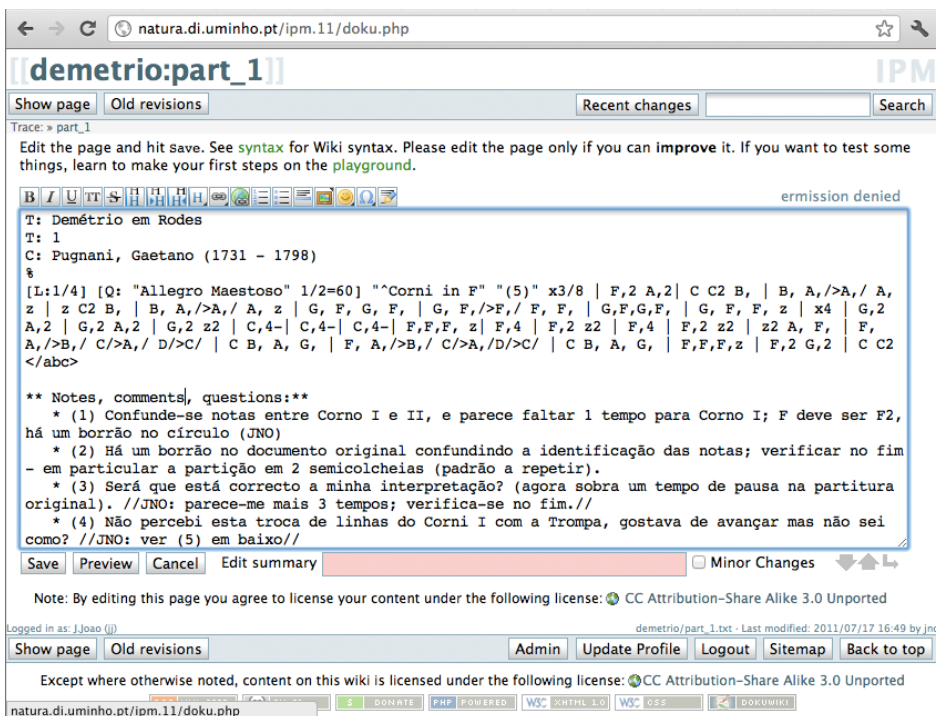


Figure 6. Editing.

These items are also provided for each cell in the matrix, in the corresponding wiki page. This is useful while editing and when performing validation tasks.

### 3. Experimental Validation

The WIKI::SCORE prototype has been used to support the transcription in modern notation of an Italian 18c opera manuscript: *Demetrio a Rodi: Dramma per Musica: Rap-presentado nel Regio Teatro di Torino nella Primavera del 1789*, by Gaetano Pugnani (1731–1798). The libretto by Giandomenico Boggio and Giuseppe Banti was published by Presso O. Derossi, 1789, a digitalized version of which is available from the public Bayerischen Staatsbibliothek [1].

The original music manuscript was found in Fundo do Conde de Redondo and is available online from the *Biblioteca Nacional Digital* [22]. To the best of the authors' knowledge, there is no book edition of the score.

Some properties of this project are:

- the manuscript has 18 parts (4 voice choir + 14 instrumental parts);
- the main divisions – *Overtura* (40 pages), *Atto primo* (149 pages), *Atto secondo* (221 pages) – are very big;
- the manuscript presents some paleographic challenges;
- 18c music notation conventions have to be understood and reproduced in modern notation;

**Demetrio a Rodi (Overtura)**  
Pugnani, Gaetano (1731 - 1798)

The image shows a page of a musical score for the Overture of the opera 'Demetrio a Rodi' by Gaetano Pugnani. The score is arranged in a standard orchestral format with multiple staves. From top to bottom, the staves are: Horns in D (Re) I and II, Trumpets I and II, Flutes I and II, Clarinets I (in A) and II (in A), Oboes I and II (with 'con oboe' and 'con flauti' markings), Violins I and II, and Viola. The music is in common time (C) and the key signature has one sharp (F#). The score includes various musical notations such as notes, rests, and dynamic markings.

Figure 7. Published PDF document excerpt.

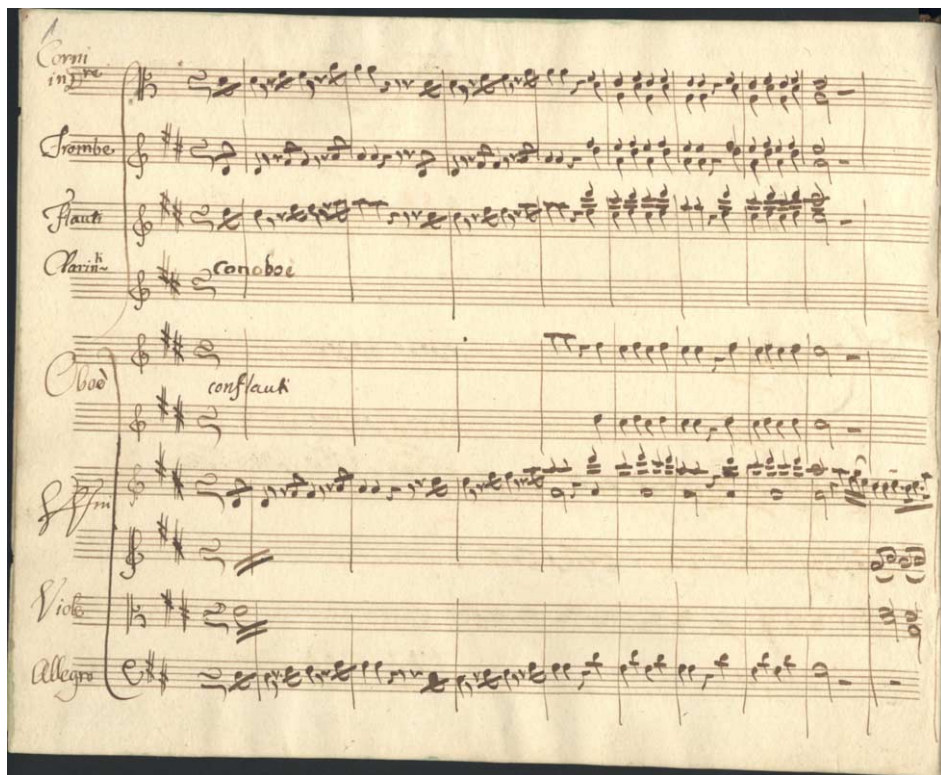
- the extant manuscript is the conductor's score, full of abbreviations, calling for the necessary adaptation in order to produce a document readable by a common musician;
- the final result should be a modern-notation standard edition.

An instance of this environment was made available to the students of a BSc degree in music<sup>1</sup>. The subject of their work in the lab sessions and home work was "the first column of the matrix": the *Overtura* of the opera. All students were used to systems such as Sibelius [5] and no one had worked on a wiki platform before. Their reaction in the beginning was negative, as they thought they would do the job faster using the systems they had used before. Gradually, they became aware of the advantages of cooperative work, in particular when they realized how easy it was for their teachers to correct their mistakes overnight, as if by magic. One student at some stage lost most of her work in a ill-fated edition session. She was about to start all over again when the teacher recalled that there was a version control behind the scene and nothing had been lost. This remains one of the most interesting episodes of the course, showing how laptops force people to work in isolation, ignoring all that has been achieved meanwhile in cooperative contents editing.

Figure 7 illustrates the first page of the published final version in PDF format. This corresponds to the manuscript score depicted in Figure 8.

<sup>1</sup>Computing for Musciology course, 2nd year of the University of Minho's degree in Music, academic year 2010/2011.





**Figure 8.** Folio 1 of the manuscript score of the Overture of the opera Demetrio a Rodi. (Available from [17]).

#### 4. Conclusion

ABC is a simple yet powerful notation for transcribing music sheets which interacts with several tools that later allow to publish contents in different well known formats, as PDF or MIDI files.

As ABC source files are plain text, it is easy to take advantage of a common collaborative work environment – a wiki. This allows for people to get into the system and interact with each other, contributing with ad-hoc work without having to learn to use other tools or buy expensive WYSIWYG tools.

Putting everything together based on a small set of glueing functions proves an excellent way to transcribe huge manuscript scores in small time spans, in a collaborative way. Also aiming at a more symbolic representation has proven useful since the same typical documents could be produced (images or MIDI), but this representation is easier to transform and process in more automated and systematic ways.

Experimental cooperative work in this area introduces a new challenge related with the bi-dimensional structure of multi-part music: synchronization. Thus the horizontal dimension of the matrix had to be introduced, the notion of reference part being crucial for proof-reading. The ABC matrix often suffers from stability issues, mainly due to synchronization problems. A simple error in the duration of a note in a part is enough to create huge synchronization problems. This is why it is important to be able to de-select (problematic) parts until they are corrected.

#### 4.1. Future Work

There are several features that could improve this work, among others:

- allow the edition of default templates for several operations, including generating the final ABC transcript;
- visual editor for cell editing for non-proficient ABC contributors;
- cooperative building of style books, gathering knowledge from music paleontologists;
- produce all-but-one MIDI files which musicians can use at home when training their part (already technically possible);
- development of a Domain Specific Language to allow a more accurate control over the project structure and synchronization.

We believe the interference implicit in the bi-dimensional structure of music contents presents new challenges in collaborative editing. Open Source communities have a lot of experience in this aspect (as computer programs are also made of interfering parts) and learning with them could be interesting in the future.

#### Acknowledgments

The authors would like to thank *Biblioteca Nacional Digital* for the excellence of their work in making music manuscripts available to the general public. They would also like to thank the reviewers for their valuable insight and detailed comments, which aided in improving this paper.

#### References

- [1] Bayerischen staatsbibliothek. Web site: <http://www.bsb-muenchen.de/> [Last accessed: 26-03-2012].
- [2] Denemo. Web site: <http://www.denemo.org/> [Last accessed: 18-01-2012].
- [3] Distributed proofreaders. Web site: <http://www.pgdp.net/> [Last accessed: 18-01-2012].
- [4] The mutopia project. Web site: <http://www.mutopiaproject.org/> [Last accessed: 18-01-2012].
- [5] Inc. Avid Technology. Sibelius. Web site: <http://www.sibelius.com/> [Last accessed: 18-01-2012].
- [6] J. Dean and S. Ghemawat. Mapreduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.
- [7] Wikimedia Foundation. Wikipedia. Web site: <http://www.wikipedia.org/> [Last accessed: 18-01-2012].
- [8] Andreas Gohr. Dokuwiki. Web site: <http://www.dokuwiki.org/> [Last accessed: 18-01-2012].
- [9] Guido Gonzato. *Making Music With Abc Plus*, 2011. <http://abcplus.sourceforge.net/#ABCGuide> [Last accessed: 18-01-2012].
- [10] M. Good. Musicxml for notation and analysis. *The virtual score: representation, retrieval, restoration*, 12:113–124, 2001.
- [11] Michael Good. MusicXML: an internet-friendly format for sheet music. In *XML Conference & Exposition*, Orlando, Florida, 2001.
- [12] H. Hoos, K. Renz, and M. Görg. Guido/mir-an experimental musical information retrieval system based on guido music notation. In *International Symposium on Music Information Retrieval*, 2001.
- [13] B. Lamb. Wide open spaces: Wikis, ready or not. *EDUCAUSE review*, 39:36–49, 2004.
- [14] B. Leuf and W. Cunningham. The wiki way: quick collaboration on the web. 2001.
- [15] Noteflight LLC. Noteflight. Web site: <http://www.noteflight.com/> [Last accessed: 18-01-2012].
- [16] Inc. MakeMusic. Finale. Web site: <http://www.finalemusic.com/> [Last accessed: 18-01-2012].
- [17] Biblioteca Digital Nacional. Demetrio a Rodi. Web site: <http://purl.pt/16734/1/P1.html> [Last accessed: 26-03-2012].

- [18] G.B. Newby and C. Franks. Distributed proofreading. In *Joint Conference on Digital Libraries, 2003. Proceedings.*, pages 361–363. IEEE, 2003.
- [19] Han-Wen Nienhuys. LilyPond, automated music formatting and the art of shipping. In *Forum International Software Livre (FISL7.0)*, 2006.
- [20] H.W. Nienhuys and J. Nieuwenhuizen. Lilypond, a system for automated music engraving. In *Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, pages 167–172. Citeseer, 2003.
- [21] I. Oppenheim. *The ABC Music standard 2.0 (December 2010)*, 2010. <http://abcnotation.com/wiki/abc:standard:v2.0> [Last accessed: 18-01-2012].
- [22] Gaetano Pugnani. Demetrio a Rodi – Dramma per Musica : Rappresentado nel Regio Teatro di Torino nella Primavera del 1789. Digital manuscript music, Biblioteca Nacional Digital. Fundo do Conde de Redondo, available from <http://purl.pt/16734/1/P5.html> [Last accessed: 18-01-2012].
- [23] Werner Schweer. Musescore. Web site: <http://www.musescore.org/> [Last accessed: 18-01-2012].
- [24] Alberto Simões, Anália Lourenço, and José João Almeida. Using text mining techniques for classical music scores analysis. In José Neves, Manuel Filipe Santos, and José Manuel Machado, editors, *New Trends in Artificial Intelligence*, pages 791–799, December 2007.
- [25] Alberto Manuel Simões and J.J. Almeida. Music publishing. In Sely Costa et al., editor, *EIPub 2003 – International conference on electronic publishing*, pages 288–298. Universidade do Minho, June 2003. Guimarães, Jun. 2003.
- [26] Chris Walshaw. Abc notation. Web site: <http://abcnotation.com/> [Last accessed: 18-01-2012].
- [27] Wikimedia. Darwikinism. Web site: <http://meta.wikimedia.org/wiki/Darwikinism> [Last accessed: 16-03-2012].

# Book Design Program: a Transition to a Hybrid Publishing Context

Ana Catarina SILVA<sup>a,1</sup> and Maria Manuel BORGES<sup>b</sup>

<sup>a</sup> *CETAC. Media, Universidade do Porto, Portugal*

<sup>b</sup> *CETAC. Media, Universidade de Coimbra, Portugal*

**Abstract.** This study is developed within the scientific areas of Information Science, Communication Sciences and Editorial Design. It refers to the hybrid publishing context in which there is a coexistence of electronic and print processes, and to the role played by Information and Communication Technologies. This scenario is conditioned by technical, morphological and cultural transformations, in the words of Roger Chartier, and represents a singular moment to observe the book and its construction processes, valuing it in its various embodiments and contexts. Considering this, the present study sought to contribute to an adaptation of the methodological tools; and an optimization of the designer's communication and information flow. In order to achieve these goals we defined two main tasks: (i) comparative analysis of traditional publishing and hybrid publishing, (ii) characterization of the designer's communication and information flows that exist in the context of hybrid publishing. To do so, we adopted qualitative research methods, with literature review and content analysis. This procedure allowed us to conclude that the publishing industry is making major investments (in what concerns organization and capital) in implementing hybrid publishing schemes. Nevertheless, there is still a lack of adequacy of design programs in this scenario. This adaptation involves a redirection of the focus of the design program, not to controlling the final shape of the object (book), but to regulating it. This procedure can only be possible with a dynamic information flow in which the editorial subject is flexible and built inside a participatory scheme by the various agents.

**Keywords.** Hybrid publishing; printed book, e-book, editorial design.

## Introduction

Recent technological developments have facilitated the emergence of a new order of communication. This changes the way we see the book today. The dynamic, decentralized and interactive modes of reading introduce alternative forms of communication and some properties of the printed book (such as the extension of memory, its intertextuality, accessibility and functionality) are enhanced in the new electronic reading devices. Others are constrained.

Roger Chartier [10] believes that this "revolution" is now the most radical to date, since it presents changes that, for the first time, occur simultaneously:

---

<sup>1</sup> Corresponding Author: Ana Catarina Silva, CETAC.MEDIA - Centro de Estudos das Tecnologias e Ciências da Comunicação, Praça Coronel Pacheco, 8, 4045-453 Porto, Portugal; Email: anavieirasilva@gmail.com.

La conséquence de tout cela est de penser que la radicale nouveauté du temps contemporain est le fait que c'est dans la même période de temps que se lient les unes aux autres les révolutions techniques, morphologiques et culturelles qui autrefois n'étaient pas liées les unes aux autres.

For all these reasons, we witness a mutation in the dominance of the printed word as a foundation of culture. Its status is now under discussion, as in several episodes of history.

The matrix of thought is that of “typographic man” and as such the application of principles that refer to print is inevitable, now that new writing and reading technology of digital nature arise. But we also recognize the need to think about new solutions that leverage the electronic format of the new book and publishing processes. It is therefore necessary to consider the specifics of the book, in both its print and electronic nature. The designer must be fully aware of the operational and symbolic potential of the book, without forgetting its heritage as a founder of civilization and the establishment of the modern human. This opens a huge field of theoretical and practical reflection on the processes of communication and information flows associated with the book, in which design plays an important role.

This study has its basis in the scientific field of Information and Communication Sciences and Editorial Design, and refers to the design of the book in a hybrid editorial context. Its relevance stems from empirical verification that the publishing market requires more flexible and interchangeable designs, which, in turn, require different design programs. This means that not only the content but also the structure and morphology of the book are now widely discussed in the traditional circuits and new media [22].

The hybrid editorial context we refer to fits within the universe of the printed book (with traditional publication processes), and the universe of e-book (aimed at electronic reading devices). It concerns the editorial processes that, at times, may alter the direction of its flow and adapt to different modes of production (such as offset or digital) or result in final objects of different natures (such as printed or electronic). For all these reasons, we find this the perfect timing to think about the book and its construction processes, perceiving and appreciating it in its various materializations and contexts, whether printed or electronic.

## **1. Scope and Methodology**

The aim of this study is to contribute to an adaptation of the methodological tools; and an optimization of the designer's communication and information flow. In order to achieve these goals we defined two main tasks for this study: (i) comparative analysis of traditional publishing and hybrid publishing, (ii) characterization of the designer's communication and information flows that exist in the context of hybrid publishing. To do so, we adopted qualitative research methods, with literature review and content analysis. The study of written sources was based in the insights of some prominent authors in the field of book history such as Henry-Jean Martin, Lucien Febvre [14], Robert Darnton [12], Roger Chartier [9,10], in authors who address the mediation and the digital future of the book, as Alistair McCleery and David Finkelstein [16], Bill Cope and Angus Phillips [11], N. Katherine Hayles [21], Johanna Drucker [13], Sven

Birkerts [5] and Jay David Bolter [7, 8], and authors that explore the context of hybrid publishing as José Afonso Furtado [17, 18, 19, 20] and John B. Thompson [29].

## 2. The Transition to a Hybrid Context

The digital mediation essentially altered the *order of books* to the extent of the production, distribution and reading of the book, more than the book in itself. The introduction of new information and communication technologies in publishing contexts implies a new literacy, but also an organizational paradigm shift. The tasks inside a book production workflow change and the outcome is less definitive and increasingly divided, within a cross-media publishing logic.

The digital processes allow a product diversification, without incurring a replacement of old by new. From this stems a hybrid publishing scenario that definitively ends the fundamentals of the dichotomy printed book / e-book, something that we can see graphically represented in the Adoni and Nossek model [1].

This is a dialectical model of media interactions that presents three possible scenarios: functional equivalence, functional differentiation and functional multiplicity. The first leads to a substitution of one media for the other. The second allows the co-existence of both. The third represents the synthesis of two media, resulting in a new media.

The hybrid publishing context we refer to, depending on the degree of functional differentiation of the editorial products (printed book, e-book, multimedia product, etc.), is illustrated at the bottom of the diagram (co-existence, interchangeability). Both the traditional (printed) and the new forms originating from digital fit there.

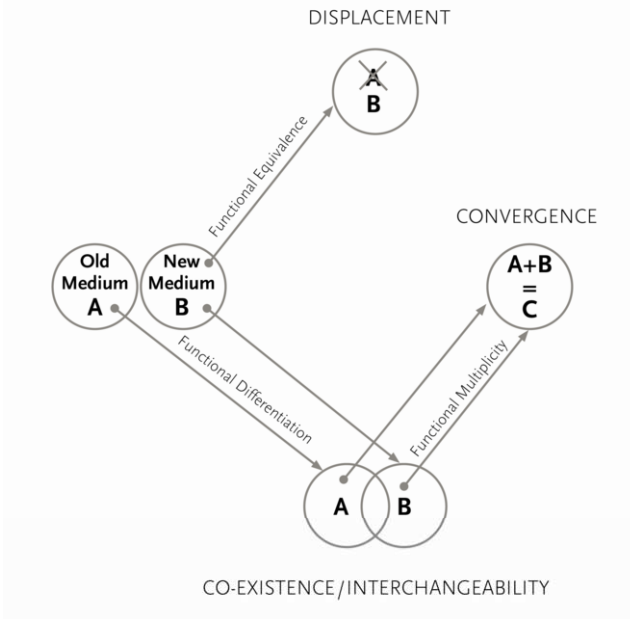


Figure 1. Dialectical model of media interactions [1].

The present and future of the book, in its various forms, is therefore based on this differentiation factor. For this reason, we can say with certainty that the predictions of the death of the book are at least exaggerated:

...the persistence of reading books might be related to the degree of interchangeability or uniqueness of print media use. If reading books fulfills psycho-social needs for the individual reader, then it can be assumed that it will persist as a cultural behavior even in the multi-media environment [1].

We find that the current publishing context isn't based on a fundamental change of the object (from print to electronic, for example), but rather on its repositioning. If before the digital revolution the book was the epitome of content and its representation, while its entire value chain was largely based on its materiality, storage and distribution, currently the impact of new technologies has led to greater awareness of the intangible nature of the contents of books. Today books, before being books, they are digital content. They are "information architecture" [11].

The digital assets become publishers' main raw material in this digital revolution. And the head function of these agents becomes the Digital Content Management, or DCM.

The distribution across multiple platforms, or cross-media publishing, implies a flexible reuse of content and its publication through different channels and in different media, digital or printed. For this, we must develop appropriate workflows, a true digital workflow that includes DCM tools.

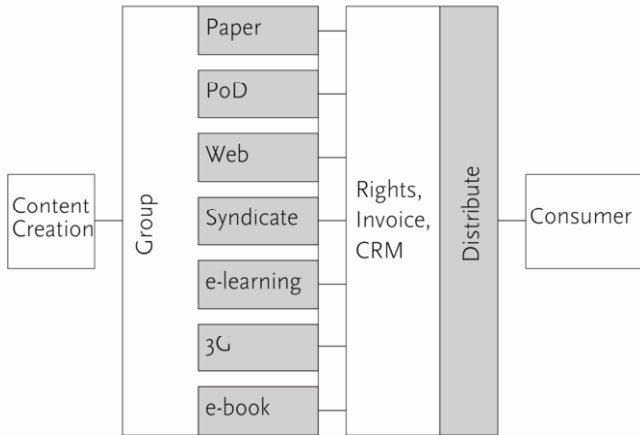


Figure 2. Cross-media publishing [20].

The adoption of a digital workflow is an ongoing process within many industries in the publishing universe. Organizational practices, work and capital flows in the publishing industry are changing, but until it doesn't reconvert, we will see a certain precariousness in the quality of hybrid publishing, as digital files are used in the manufacture of books with different natures (print and electronic), without taking into account their specificities.

Fig.3 presents a digital workflow, as Thompson [29] sees it. We can see that this scheme is planned to have a particularization of the final output, with respect to the production of the book. In the last phase of the scheme proposed by Thompson, we can see that there is a planned *conversion* of the digital file that originally serves print, to an electronic output, i.e., the e-book, in its different formats (when this scheme was published). Yet it is curious that there is a “conversion.” This means that digital assets are still designed for one particular output, only there is the possibility to adapt them to another.

What we believe to be a true digital workflow is somewhat different, in that it includes another mode of production, in which the digital content has no fixed form. Instead, it takes form in an appropriate manner to each of the different distribution channels, as illustrated in Fig.4.

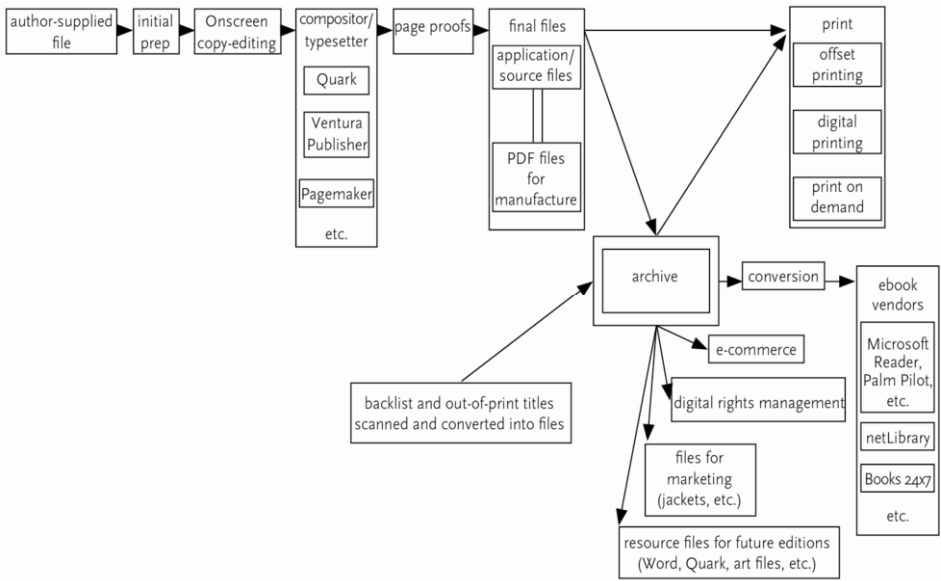


Figure 3. Digital workflow [29].

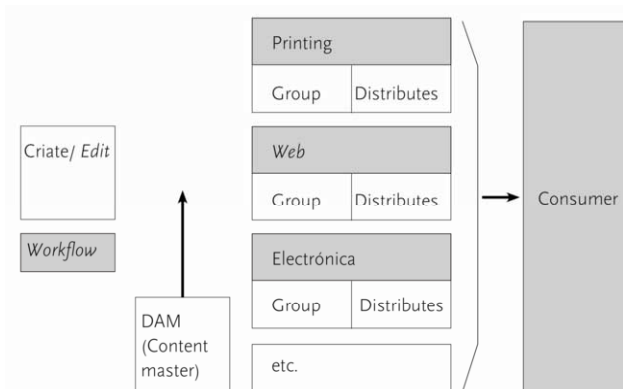


Figure 4. Repurposing workflow [20].



This new framework also enables some significant changes in reading and in its processes. And the main result of the digital revolution, with regard to the book is the definitive transition from what Roger Chartier calls a system of intensive reading to that of extensive reading, a process that begins with the printed book (and its mass distribution). Maria Augusta Babo [3] referring to the concepts of Chartier, defines the first regime (intensive reading) as recurrent and meditative, and the second (extensive reading) as comprehensive and informative. Now, these modalities have been heightened in two types of reading, known as “speed reading” and “deep reading.” But inside this new regime of extensive reading, if not radicalized, we can also recognize some insights of Edgar Morin, about the complexity and different ways of thinking the real, not through a set of unconnected simple units of knowledge, but as a result of their relationship and interconnection. Only this way can we devise an appropriate way of reading, suited to the new requirements and incentives of a “plurisemiotic” [4] and intertextual regime in which we live today. Babo [4], in reference to Bertrand Gervais, describes this scenario in which the text spreads through the media, complemented with images, in different embodiments, hipertextualizing, leading to an accelerated and diversified reading.

The concept of hypertext is, precisely, one of the major reasons for the change in the ways of reading, as we move from mechanical to digital. “Hypertext” is a term introduced by Theodor Nelson in 1965 that defines a non-sequential writing, in which the player chooses the links.

Reading in a digital environment therefore requires different ways of reading and a transformation in the appearance of the text, considering its different operability. The e-book is one support that enables this. And a new problem emerges: the book, that established unity, with five centuries of history, is once again undefined. The digital technology mediated book must recreate the codex functionalities, something that in the last decades has been mistaken for a formal recreation, by many book agents, and in particular, by the designer.

The dream of the electronic distribution of content has led to many experiences since the late eighties of the twentieth century. But it was the internet (and not the floppy-disk or CD-ROM) that paved the way for the digital publishing revolution we live in today. Thompson [29] lists four reasons for the initial limited success of the e-book: (1) hardware, (2) formats, (3) rights, (4) price. Nowadays, within these parameters, an e-book already has technical, industrial and trading features that take it to overcome these hurdles. But as Johanna Drucker [13] identifies, the conceptual obstacles remain: “Functions are not the same as formal features”. And the author highlights the phenomenon of the e-book emulating the codex. Drucker, referring IBM’s researcher Harold Henke, questions the relevance of the “metaphors” of book structure, attempts to seduce the reader with familiar features from print.

Reproductions of: page volume, turn of the page, sound of flipping, spine of the book (referring to the double page spread when it really comes to a single page that is the screen), among other effects have, in Drucker’s point of view, an opposite result: “I would argue that as long as visual cues suggest a literal book, our expectations continue to be constrained by the idea that books are communication devices whose form has a static and formal, rather than active and functional, origin”.

Drucker suggests another approach to the book: rather than working as a formal structure, the book should be made from its program which in turn comes from the needs that the formal structure of the book presents.

### 3. Book Design Program

The definition of the design program of a book is born in the dialogue (direct or otherwise) of the designer with the publisher and the author. It involves an analysis of the contents and comes from the identification of all the requirements and constraints. Drucker describes her approach as follows: (i) analyze “how” the book “works” and not describe what the book “is”, (ii) determine the “program” that comes from the formal structures of the book, (iii) discard any iconic “metaphor” of book structure, focusing on understanding how these forms regulate the performance of the book itself. From here, the literal space<sup>2</sup> of the book becomes a space of intervention for the reader. Bill Cope and Angus Phillips [11] also argue that a new definition of what the book is should be focused on function rather than the object: “A book is not a thing. It is a textual form, a way of communicating. A book is not a product. It is an information architecture” [11] But emphasize the intrinsic characteristic of any book, which is its “bookish” structure: “A book is no longer a physical thing. A book is what a book does” [11].

A book, recognizable in the world of books, presents some general characteristics such as: the fact that it is an extensive text (by “text” we refer to text and/ or images), being organized by some paratexts as covers, production notes; title, author, copyright, ISBN, or DOI (Digital Object Identifier) table of contents, headings, subheadings and body text, indexes, acknowledgments, foreword, author’s biography, introduction, among others.

In summary, this is how the authors [11] identify what the book “does”: presents a characteristic textual or communicative structure, and features the typical functions of “books”, because it is defined, registered and recognized as a book.

This concept of *information architecture* is also present in the insights of Craig Mod [26] about the future of the book. He understands it stems from relationships and not from a (metaphorical and literal) surface. He states that the books are systems: “They emerge from systems. They themselves are systems – the best of which are as complex as necessary and not a bit more. And once complete, new systems develop around their content.”

Electronic text, distributed online, accessed through a computer, smart-phone, tablet, or dedicated reader, brings a new variable to this equation that is the hybrid publishing process. That workflow, only addressed to print, must now be a hybrid one, so that the final product can either be the printed book or the e-book. Both have the same origin in a digital file, although the way this digital information is mediated by the different stakeholders in the chain of production of the book conditions the book and its use.

At this point it also has to be considered the relationships between technologies, markets and types of content [20]. Because the publishing industry is very diverse, it appears that the digital revolution also has a different impact on the various sectors of the publishing universe. One can understand why some book categories (like encyclopedias, for example) have adopted hybrid forms so quickly, if not migrated to electronic form, while others hold to the paper.

---

<sup>2</sup> Drucker defines the literal book, “that familiar icon of bound pages in finite, fixed sequence” and contrasts it with the phenomenal book, “the complex production of meaning and effect that arises from dynamic interaction with the literal work.”

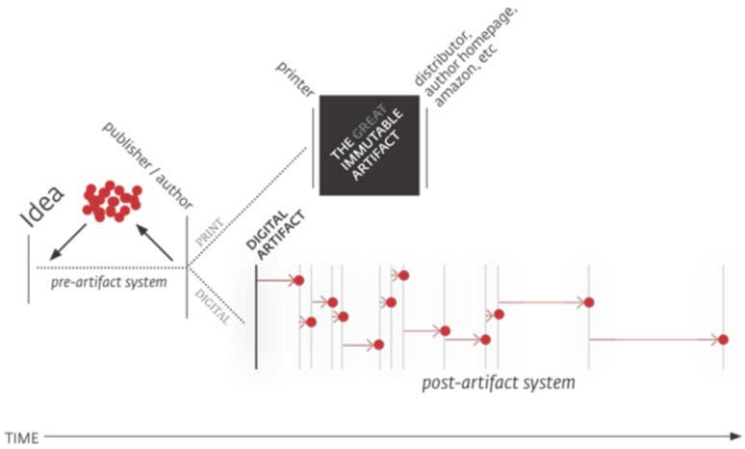


Figure 5. A ridiculously simplified diagram of how digital affects book publishing [26].

The fact that digital media have come to add great possibilities to the methods already confirmed by history and part of the modern way to access and read information is now unavoidable. For this reason it is necessary to rethink the processes and relationships between the various actors so that the shape of the final book, whether printed or electronic, is potentiating of its content. In the present study we seek to do so from the point of view of one player in particular: the designer.

Design is what makes “the Beautiful, the Good and Useful” possible, wrote Giambattista Bodoni<sup>3</sup> in 1818 [6]. It is also the process that leads to the operation of a projectual methodology that determines the solution of a problem or as Bruno Munari [27] puts it, a series of operations, arranged in a logical order dictated by experience.

Within the disciplines of graphic design, editorial design is perhaps the one that remained the most faithful to the functional ideals of the origins of design. Not only because it was after the book (and printing) that a theoretical corpus first started to take form, but also, because this object, in its modern configuration of codex has remained largely unchanged for five centuries, such was the degree of optimal functionality and aesthetic reached. Today, with the introduction of digital processes, this form is not compromised, but the design program is, as Hayles [21] states:

So essential is digitality to contemporary processes of composition, storage, and production that print should properly be considered a particular form of output for digital files rather than a medium separate from digital instantiation. The digital leaves its mark on print in new capabilities for innovative typography, new aesthetics for book design, and in the near future new modes of marketing.

<sup>3</sup> In his work of 1818 [6], Bodoni establishes a set of rules to build a good book. Some of which were more scientific, others more empirical. His analysis begins by distinguishing three aspects “the Beautiful, the Good and the Util.” The utility is measured in how many people are lured into reading, the number of readings, the pleasure of reading and the speed of reading. A good book is one that motivates most from this. As for the beautiful, it consists of two things: “harmony which satisfies the spirit by enabling it to recognize that all individual parts of a work are subordinated to a complete idea; it also consists of proportions which please the eye or rather the imagination, for the imagination contains within itself certain definite ideas or images, and the more that which is perceived coincides with them, the greater the pleasure one takes in it.”

The physical characteristics of the book as the texture, smell, weight, size, proportion, light, color are significant and contribute to an experience of the book as *cultural text*. But the text, seen as a set of words inscribed on the pages (printed or digital) of a book should also be clarified in terms of its operability. Thus, the effectiveness of communication is influenced by dimensions such as readability and intelligibility, but also depends largely on design issues that are based on typography, layout and text body. The *mold* that regulates its meaning.

This relationship, with respect to the electronic environment, is now problematic. Not only did the graphic design has been unable to adapt to new demands of the environment, but also the interfaces and platforms need to become accessible, including principles of user-centered design. In 2009, a report of the Joint Information Systems Committee (JISC) [23], found that “they are far from ideal, and in some cases, barely serviceable.” Today the scenario is still undefined: “What’s happening at the moment is that most publishers are handing their major titles over to app developers who are ruining these titles with rushed, unprofessional layout and design. (...) The complete lack of care and attention paid to the production of digital books is genuinely mystifying” [24].

Defining the future of digital media, and particularly the e-book, is a difficult task. However, there are key concepts in the interaction design field that can be a reference. In the french committee's report on the e-book Bruno Patino [28] lists some that can help improve in creating new uses and experiences: mobility, interoperability, access, connection, interactivity, information optimization.

Now graphic design will have to find the tools to enable the best experience of the book, both as printed and as e-book. This will be possible through dynamic information flows, acknowledging the book as an information architecture [11] and not as a fixed and unchanging structure.

#### 4. Next Chapters

In this scenario of digital revolution the “professions of the book” have been changing. Namely the designer who is required a greater versatility. The publishing industry is now making major investments (both organizational and capital) in order to implement hybrid publishing workflows. However, this study pointed us that there is still a lack of adequacy of design programs to this scenario. This adaptation demands a refocusing of the design program, not to the control of the final shape of the object (book), but towards its regulation. This way of acting must be based on dynamic information flows, in which the editorial subject is flexible and built in a participatory way by various agents.

#### References

- [1] ADONI, Hanna; NOSSEK, Hillel, *The New Media Consumers: Media Convergence and the Displacement Effect*. Communication & Medicine , Volume 26 (1), (2001)
- [2] BABO, Maria Augusta, *A Escrita do Livro*, Vega, Lisboa, 1993.
- [3] BABO, Maria Augusta, *Os desafios do hipertexto*, 2005, [http://www.interact.com.pt/memory/12/html/interact12\\_10.htm](http://www.interact.com.pt/memory/12/html/interact12_10.htm)

- [4] BABO, Maria Augusta, *Do texto como textura heterogénea ao texto como textura híbrida*, 2005, <http://bocc.ubi.pt/pag/babo-maria-augusta-texto-como-textura-heterogenea-texto-como-textura-hibrida.pdf>
- [5] BIRKERTS, Sven, *The Gutenberg Elegies: the Fate of Reading in an Electronic Age*, Faber and Faber, New York, 2006.
- [6] BODONI, Giambattista, *Manuale Tipografico 1818*. In FÜSSEL, Stephan (ed.), *Manual of Typography*, Taschen, Köln, 2010.
- [7] BOLTER, Jay David, *Remediation: Understanding New Media*, The MIT Press, Cambridge, 2000.
- [8] BOLTER, Jay David, *Writing Space: Computers, Hypertext, and the Remediation of Print*. Routledge, New York, 2001.
- [9] CHARTIER, Roger, *A ordem dos livros*, Vega, Lisboa, 1997.
- [10] CHARTIER, Roger *Roger Chartier*, Interview by Eric Ouzounian. October 10, 2010. December 10, 2011. <http://www.laspirale.org/texte.php?id=311>
- [11] COPE, Bill; PHILLIPS, Angus (Eds.), *The Future of the Book in the Digital Age*, Chandos Publishing, Oxford, 2006.
- [12] DARNTON, Robert, *The Case for Books: Past, Present and Future*, PublicAffairs, New York, 2009.
- [13] DRUCKER, Johanna, *The Virtual Codex from Page Space to E-space*, 2003, <http://www.philobiblon.com/drucker/>
- [14] FEBVRE, Lucian; MARTIN, Henri-Jean, *The Coming of the Book – The Impact of Printing 1450 – 1800*. Verso, second ed. London, 1997.
- [15] FINKELSTEIN, David; McCLEERY, Alistair, *An Introduction to Book History*. Routledge, New York, 2005.
- [16] FINKELSTEIN, David; McCLEERY, Alistair, *The Book History Reader*. second ed. Routledge, New York, 2006.
- [17] FURTADO, José Afonso, *O Livro*, Difusão Cultural, Lisboa, 1995.
- [18] FURTADO, José Afonso, *Livro e leitura no novo ambiente digital*. 2002. <http://www.educ.fc.ul.pt/hyper/resources/afurtado/index.htm>
- [19] FURTADO, José Afonso, *O Papel e o Pixel: Do impresso ao digital: continuidades e transformações*, Ariadne, Lisboa, 2007.
- [20] FURTADO, José Afonso, *A Edição de Livros e a Gestão Estratégia*, second ed. Booktailers, Lisboa, 2009.
- [21] HAYLES, N. Katherine, *Electronic Literature: New Horizons for the Literary*, University of Notre Dame Press, 2008.
- [22] HASLAM, Andrew, *Book Design*, Laurence King, London, 2006.
- [23] JISC, *National E-Books Observatory Project: Key findings and recommendations. Final Report*, 2009. [www.jisc-collections.ac.uk](http://www.jisc-collections.ac.uk)
- [24] MEADOWS, Chris, *Are programmers 'ruining' e-books?*, 2012, <http://www.teleread.com/ebooks/are-programmers-ruining-e-books/>
- [25] MOD, Craig, *Books in the Age of the iPad*, 2010, [http://craigmod.com/journal/ipad\\_and\\_books/#starts\\_content](http://craigmod.com/journal/ipad_and_books/#starts_content)
- [26] MOD, Craig, *Post-Artifacts Books & Publishing*, 2011, [http://craigmod.com/journal/post\\_artifact/#section\\_5](http://craigmod.com/journal/post_artifact/#section_5)
- [27] MUNARI, Bruno, *Das Coisas Nascem Coisas*, Edições 70, Lisboa, 2008.
- [28] PATINO, Bruno, *Rapport sur le livre numérique*, 2008, <http://www.culture.gouv.fr/culture/actualites/conferen/albanel/rapportpatino.pdf>
- [29] THOMPSON, John B, *Books in the Digital Age*, Polity Press, Malden, 2008.

# “Tangible Culture” - Designing Virtual Exhibitions on Multi-Touch Devices

M. Hecher<sup>a</sup> R. Möstl<sup>b</sup> E. Eggeling<sup>a</sup> C. Derler<sup>b</sup> D. W. Fellner<sup>a</sup>

<sup>a</sup>*Fraunhofer Austria Research GmbH, Austria*

<sup>b</sup>*JOANNEUM RESEARCH GmbH, Austria*

## **Abstract.**

Cultural heritage institutions such as galleries, museums and libraries increasingly use digital media to present artifacts to their audience and enable them to immerse themselves in a cultural virtual world. With the application eXhibition:editor3D museum curators and editors have a software tool at hand to interactively plan and visualize exhibitions. In this paper we present an extension to the application that enhances the workflow when designing exhibitions. By introducing multi-touch technology to the graphical user interfaces the designing phase of an exhibition is efficiently simplified, especially for non-technical users. Furthermore, multi-touch technology offers a novel way of collaborative work to be integrated into a decision making process. A flexible export system allows to store created exhibitions in various formats to display them on websites, mobile devices or custom viewers. E.g. the widespread 3D scene standard Extensible 3D (X3D) is one of the export formats and we use it to directly incorporate a realtime preview of the exhibition in the authoring process. The combination of the tangible user interfaces with the realtime preview gives curators and exhibition planers a capable tool for efficiently presenting cultural heritage in electronic media.

**Keywords.** multi touch, virtual exhibitions, interactive design, digital preservation, X3D

## **1. Introduction**

In cultural heritage institutions digital content has gained increasing importance over recent years. Digitalized media like photographs, videos, audio and three-dimensional representations of exhibition objects are available to museums and recent projects like 3D-COFORM [1] gather to enhance and simplify the process of acquiring digital representations of cultural heritage assets, as well as to document and persistently store the accumulated information. Therefore, in the future even more assets will be available in digitalized and annotated form.

The availability of scanned 3D objects, photographs and videos of historical artifacts offers many advantages. Exhibits, which would be secured in an archive, can now be presented to a wide audience. Furthermore, virtual artifacts cannot only be used to replace existing objects, but also to complete fragments or to show them in a state never seen before. In digital form the content of an exhibition can be visualized with highlights and enriched with information. Combining those enriched digital artifacts with current digital media technologies and/or modern web browsers capable of displaying 3D con-

tent the cultural content reaches an audience all over the world without replacing the real museum.

One challenging problem is to represent the digitalized cultural heritage assets and bring them in a form applicable to be shown to a "virtual visitor". An interesting project in this area is the "Google Art Project" [7]. It digitalizes museums in taking panorama pictures of the exhibition areas and the exhibits. Interested users can then navigate freely within the museum and explore the exhibits in using an ordinary web browser as a viewer. The application eXhibition:editor3D we present in this paper also concentrates on the goal to allow the distribution of cultural heritage in a digital form. However, its focus is not only the presentation of an exhibition but it provides an authoring tool capable of assembling digital data as part of a virtual exhibition. eXhibition:editor3D is an existing application for authoring 3D virtual exhibitions and exhibition tours. We will not present a detailed description of the tool itself, but we describe the aspect of extending it with a gesture-based touch interface that allows a curator to utilize the advantages of multi-touch technologies to simplify the authoring process of digital exhibitions. The multi-touch paradigm also enriches the application with a new form of collaborative work when designing exhibitions, which is described later on.

The next section gives an overview of other available tools for designing digital exhibitions and shortly describes their concepts. Section 3.1 then introduces the eXhibition:editor3D application. Section 3.2 describes the extension of the tool to support gesture based user interfaces on multi-touch devices. Afterwards - in section 3.3 - the visualization module supporting the authoring process is described. Section 4 concludes the outcome of the project and gives an outlook of future work.

## 2. Related Work

The need for authoring tools in museums to manage the increasing amount of digitalized data gained and to be gained by projects like 3D-COFORM [1] raises the research effort in the field of computer aided planning of virtual exhibitions. Creating this kind of content is a creative process and as such there are different approaches to get the task done, depending on the specific need. Tools currently available typically provide a workflow suite for curators and editors that supports them in their daily work of managing digital exhibits, annotate the data with additional (multi-media) information and organize it in virtual exhibitions to make the cultural heritage information available for a wide audience.

The different tools have the common goal to (a) provide a workflow that allows the management and visualization of cultural heritage and (b) be usable by non-technical users which do not necessarily have experiences in the field of 3D graphics or computer programming. The scale of the applications differs regarding their feature level and coverage of the workflow process, as well as their user interface concept.

In [2] a system is presented to design virtual exhibitions from pre-existing 3D models. The tool provides a 3D interface in first-person view to navigate through the virtual space and arrange exhibits. An artificial intelligence algorithm supports the curator in defining interesting viewpoints inside the exhibition area. Finally the viewpoints are connected to form a virtual tour. For inexperienced users of 3D applications navigating through the virtual space may be complicated, especially when they are not used to this

kind of navigation. The tool encounters that issue in providing the possibility to overlay a 2D map of the virtual space to support navigation. Assistance for the placement procedure of exhibits within the scene is also provided.

Another popular approach is to utilize game-engine technologies and adapt existing frameworks to the field of cultural heritage presentation. In [11] a game engine framework forms the basis of the authoring tool and the visualization module. According to the paper the game engine basis has user and system side advantages, like the good overall performance and the ease of use for curators. From the view of a programmer the framework already provides basic functionality for the shading and the rendering of models, as well as functionality to navigate through the virtual world. In using a game-engine framework such and similar basic necessities do not have to be reimplemented and thus allow for a shorter development cycle of the product.

Besides applications with 3D user interfaces another type exists that - instead of arranging the exhibition in 3D - utilizes a 2D interface system to perform the authoring work. For instance, [15] presents a tool based on floor plans where a curator places exhibits on a flat canvas in a top view user interface. The 2D data generated by the authoring module is then converted in a 3D format and presented by an extensible visualization module.

As stated before it is not trivial for non-technical users to navigate within a three-dimensional world with traditional input devices like a mouse pointer or keyboard on a two-dimensional monitor. Similar to [15] our approach uses a 2D user interface for the authoring tool to provide a simple possibility for navigating inside the exhibition area and for arranging exhibits and virtual tours. However - in contrary to the related work presented in this section - the traditional mouse pointer input method is replaced by touch gesture input interfaces to allow for an authoring process that maps more naturally to our haptic experience of organizing and arranging objects in everyday life situations.

### 3. The eXhibition:editor3D Application

#### 3.1. The Authoring Tool

The eXhibition:editor3D application is developed by the Austrian Institute of Information and Communication Technologies of JOANNEUM RESEARCH [9]. Its purpose is to create virtual 3D exhibitions from existing digitalized exhibition content, like 3D assets, photographs, videos and audio clips. The application's tool chain allows curators and editors to

- design and preview upcoming exhibitions,
- archive temporary exhibitions
- create interactive 3D presentations for multi-media terminals
- create catalogs or websites
- do advertising for their exhibitions.

Museum curators were deeply involved in shaping the look-and-feel as well as the functionality of the tool. As this cooperation has been established right from the beginning of the software development process, the application is specifically tailored to fit their needs.



To manage a virtual exhibition the editor organizes its content in virtual *rooms*, where multiple rooms can be combined together within a project to present different exhibition floors or different variants of a single exhibition. In a first step the user equips a room with the desired exhibits. The authoring of exhibits is managed by two-dimensional graphical user interfaces and concepts based on floor plans. With a room represented as a 2D floor plan the author places exhibition objects virtually on a flat canvas. The media content (e.g. images) of the exhibits is managed by so called *content repositories*. Media assets are either located in the filesystem or on a web server. Exhibits are easily created by just dragging and dropping a media asset onto the 2D canvas. The different objects are represented as icons that determine the type of exhibit, e.g. 3D object, image, video or audio clip are denoted by their own icon. The user interface previews the content represented by the icon if it is selected. This way a user can simply distinguish between the different exhibits. Moreover, a *property view* is provided that summarizes all properties of an exhibit and allows e.g. to define a name for the object or to enter a precise location or rotation value. eXhibition:editor3D uses existing digital content, the user is free to choose the tools for digitalizing the desired cultural heritage assets and to post-process them. In a second step - after the exhibit arrangement - a virtual tour through the room is planned. For this reason so called *viewpoints* are specified by the curator defining the location and view directions of a visitor to direct her or his attention to a certain aspect of the exhibition. The viewpoints are then connected to form a virtual tour. Finally the resulting exhibition and the virtual tour can be exported as a X3D or Adobe Flash SWF file for visualization. Figure 1 shows an overview of the application.

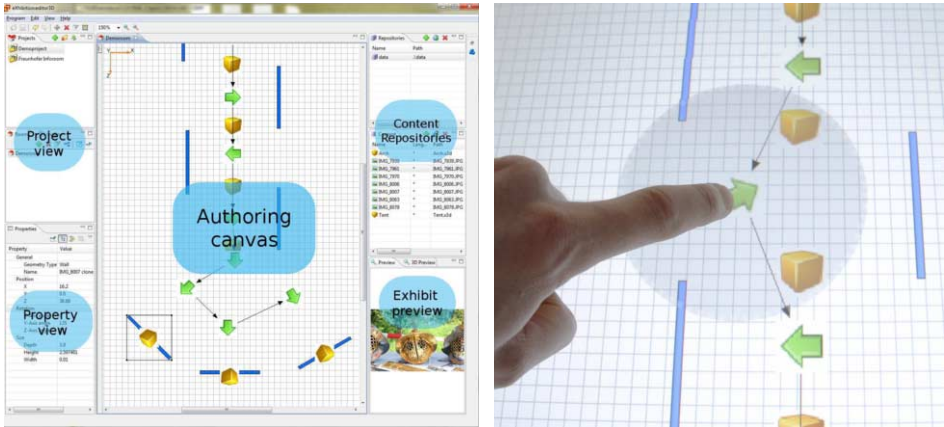
eXhibition:editor3D is based on the open-source and platform independent *Eclipse* environment. More specifically, it uses the *Eclipse Rich Client Platform* (RCP) [4]. RCP allows application developers to reuse components of the Eclipse platform for developing custom applications. Basic window management, different views on content or functionality like an undo/redo system do not have to be re-implemented, they are provided as part of the framework. Another advantage is the presence of a powerful plug-in system for extending the application. So called *extensions* can be attached to integrate additional functionality. Two custom extensions form the foundation of the modules described in the following sections. The first extends the application with gesture recognition on multi-touch devices and integrates gestures into the graphical user interfaces, the second implements a convenient realtime preview for the authoring process.

An in-depth explanation of the eXhibition:editor3D application is given in [12] and [13].

## 3.2. Gesture Interface

### 3.2.1. Introduction

In this project JOANNEUM RESEARCH cooperated with the Fraunhofer Research GmbH, Austria [6] to extend the eXhibition:editor3D application with the capability to handle touch gesture input. Fraunhofer contributed its knowledge in the field of multi-touch technologies. The traditional way of interacting with the authoring application uses the mouse pointer as the input device. This user interaction paradigm has changed with the upcoming of more touch-enabled devices in recent years. Multi-touch technology is the interactive and intuitive means of providing a new user experience and enhances the usability of a range of applications. The success of gesture based applications is im-



**Figure 1.** (left) The GUI interface of eXhibition:editor3D showing the authoring canvas, the content repositories, the exhibit preview windows and the property and project view. (right) Touching an exhibit gives visual feedback in marking the area considered for additional multi-touch input with a gray area. Nearby exhibits within this area are not selected when touching them with another finger. Instead, the touch contributes to the gesture on the currently selected exhibit.

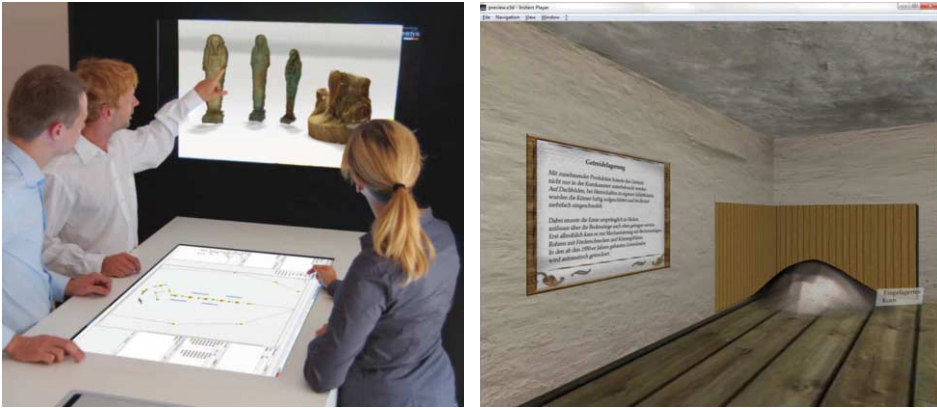
pressively demonstrated by mobile devices, where 2D user interfaces are operated on in solely using multi-touch gesture input without mouse or keyboard input devices.

### 3.2.2. Supported Gestures

The multi-touch input paradigm was applied to the authoring interface of eXhibition:editor3D. Two areas are managed by gestures: The first area is controlling the transformation of an exhibition object, the second area concerns the navigation inside a virtual room. The typical workflow for selecting an item from the content repository, dragging it onto the virtual room canvas and adjusting its position and rotation with the help of mouse and keyboard is replaced by typical multi-touch gestures, allowing for an intuitive exhibition arrangement that maps naturally to the human haptic experience of positioning objects on a canvas. Gestures are provided to translate, rotate and scale an object. Translation is accomplished in selecting an object and moving it around with a single fingertip. For rotation two fingers are placed on or in the near vicinity of an object (e.g. if it is too small to place 2 fingers directly on the icon). When rotating the fingers the rotation angle is mapped on the exhibit. Scaling an object is achieved in first selecting it with a two finger tap and then dragging a third finger forth or back in the vicinity to apply a scale factor. To navigate through the virtual room two distinct gestures are implemented. The first one is a zoom gesture allowing to shrink or enlarge a section of the room. To activate the gesture two fingers are put on the room canvas without hitting an exhibit. When bringing the two fingers nearer together the view is zoomed in, dragging the fingers apart zooms out. Especially for zoomed-in views it is necessary to pan the view to explore other areas. That is accomplished with a single touch on the room canvas and dragging the view until the desired section is reached.

### 3.2.3. Precision Considerations

Touch input brings up a problem that is commonly named as the “fat finger” problem (a more formal explanation of the issue can be found at [5], where touch vs. mouse

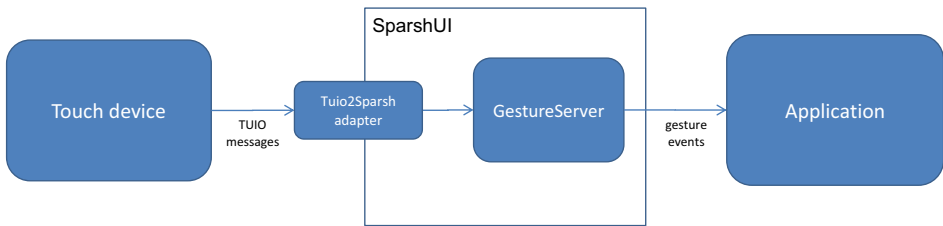


**Figure 2.** (left) Example setup of a monitor touch device with a separate monitor for the visualizer module with multiple persons collaboratively working on an exhibition. (right) Visualizer module previewing an exhibition using 3D elements and an image with descriptive information of the scene.

input is compared). Selecting an object that is surrounded by others can be tedious as the natural size of a finger does not provide the selection accuracy of a pixel-accurate mouse pointer. The problem gets even more complicated when a multi-touch gesture is to be applied on a relatively small object. In this case it is possible that the user wants to e.g. scale an object in putting two fingers on it and then use the third finger for the actual scaling. If the third finger is placed on a nearby object this object will be moved instead of scaling the previously selected one. As a second example, if two users are manipulating the transformation of nearby objects it is possible that one user selects an object with two fingers to apply a rotation. When the second user selects an object in vicinity with another finger to apply a translation it is possible that the third finger is interpreted as scale factor for the first object. To overcome these selection issues we introduce visual feedback for the selected object. When selecting an item the influence area accepting touches considered as part of a gesture on the object is visually highlighted as shown in Figure 1. When a second or a third finger is placed inside this area of influence the touch is always applied to the selected object. Nearby items located within the area are not considered for touch input. The visual indication of the currently selected object and its area of influence solves confusing behavior when selecting objects in a simple but effective way.

### 3.2.4. Collaborative Working

The multi-touch paradigm also enriches the workflow of designing exhibitions with the possibility to collaborate on a single exhibition. Around a multi-touch table setup as shown in Figure 2, curators can jointly discuss ideas respectively position and adjust exhibition artifacts in a comfortable way. The setup shows one vertical touch-monitor device integrated in a table furniture. Within the table's sideboard a second monitor is integrated. The second monitor is used to display a 3D visualization of the exhibition in real-time, a feature described in Section 2. The setup allows multiple persons to stand around the touch-device, using the eXhibition:editor3D collaboratively to directly support the decision making process. It becomes easy to try out different exhibition arrangements as



**Figure 3.** The TUIO input messages from the multi-touch device are converted by the SparshUI framework into gesture events and are then delivered to the touch interface of the application.

each member of the decision making group has the possibility to quickly show ideas in simply arranging the exhibits with a fingertip.

It is possible that multiple users simultaneously add, remove or transform exhibits in the virtual room. For serious exhibition planning this feature may not be very helpful. However, when using the application for educational purposes the feature can be used to provide e.g. school classes with an interesting possibility to create exhibitions and to waken the interest of pupils in packaging historical events into a playful like experience when arranging exhibits with multi-touch gestures.

### 3.2.5. Technical Implementation

The technical implementation of the gesture input is based on the *SparshUI* [14] library. It provides a gesture recognition framework that relieves the developer from manually transforming a sequence of unconnected low-level touch input events into corresponding gesture events. A *gesture server* calculates gestures out of a stream of touch inputs and provides them in form of events to the application. The library supports default built-in gestures like drag, rotate, scale, etc. but can be extended with user defined gestures as well. The low-level touch input is read by an adapter component that converts touch input events from different sources to the internal format of the gesture server. In this project an adapter connects the *tangible user interface objects* (TUIO) protocol [10] used by many touch devices to translate the input into the gesture server's recognized format. Adapter e.g. for the native Windows 7 touch input events are also available. The high-level gesture events are then mapped to transformation values of the exhibit objects. The objects are represented with the *Eclipse Graphical Editing Framework* (GEF) [3], a 2D based framework for creating graphical interfaces. Extending the existing *eXhibition:editor3D* application is achieved in utilizing *extensions*, a plugin mechanism provided by the Eclipse framework. Figure 3 shows the connection between the gesture recognition and the authoring tool.

## 3.3. Visualization Module

### 3.3.1. Realtime Preview

Two-dimensional user interfaces have the advantage to be easily handled by non-technical users, however, their disadvantage in this field of application is the lack of visual feedback on how the end result looks like. *eXhibition:editor3D* thus provides a 3D preview that changes and updates the 3D representation of the scene immediately. The application provides an extensible plugin system to support different output formats.

Currently the exhibition can be exported to the X3D format, as well as to Adobe Flash SWF. However, concerning user-friendliness this approach has the flaw that changes in the 2D view provide no instant visual feedback for the curator. To see the changes made on an exhibition the scene first has to be exported to the 3D format for visualization. This conversion step slows down the authoring process considerably, especially when fine tuning arrangement details.

To make the workflow of (a) adjusting an exhibits transformation, (b) export the scene, (c) visualize and judge the result to then (d) iterate the process until a satisfying result is achieved the existing plugin system for visualization modules was used to extend the editor with a new module. This module is based on the X3D format and uses its dynamic capabilities to enhance the user friendliness of the authoring process in instantly previewing the current exhibition in its 3D representation. Changes are reflected in realtime by the visualizer. This instant visual feedback enhances the efficiency of the workflow and simplifies the fine tuning task considerably. From a curator's perspective the authoring tool and the visualizer component are no longer separated entities of the workflow, but form an integrated part.

The visualization module directly supports the definition of viewpoints and their connection to a virtual tour. Per drag'n'drop the curator positions viewpoints within the exhibition room. In activating a viewpoint with a double tap on its icon the visualizer changes its camera to the selected viewpoint. With translation and rotation gestures the location and viewing direction can be adjusted, the result is directly shown in the visualizer. That allows a very fast way to define interesting points of view of an exhibition and to connect these points to form a virtual tour. Besides the possibility to activate single viewpoints the application provides touchable buttons that allow to skip through the whole tour in the visualizer to give an impression of the result and to evaluate the immersion of the story telling process.

### 3.3.2. Technical Implementation

The visualization module is implemented on top of the *InstantReality* suite developed by Fraunhofer IGD [8]. The suite provides a X3D viewer - the *InstantPlayer* - that visualizes the 3D representation of the exhibition. The *eXhibition:editor3D* and the *InstantPlayer* are connected via the X3D scene format. When a new authoring session is started *eXhibition:editor3D* internally generates a X3D file representing the exhibition scene. This file is loaded by the *InstantPlayer* and shows the current state of the scene. The *InstantPlayer* supports the *External Authoring Interface* (EAI), a protocol to dynamically change the loaded 3D scene. This feature is used to instantly map changes applied in the authoring tool to the visualization. The EAI is also utilized to switch the camera in the *InstantPlayer* when a viewpoint is selected by the curator. The result is a preview of the scene that updates in realtime to give the curator instant visual feedback of the changes and also allows to quickly adjust viewpoint locations and viewing directions.

*InstantReality* is a powerful framework to visualize the exhibition scene in different environments. E.g. it is possible to directly render the result as stereoscopic image on a projection wall or in a CAVE (Cave Automatic Virtual Environment). The EAI implementation of the suite supports remote invocation over the network. Therefore the authoring tool and the visualizer do not necessarily have to run on the same host. This loose coupling allows a flexible combination of authoring and visualization devices, which can be used e.g. for interactive presentations of exhibitions where the authoring tool is

running on a multi-touch monitor device and the visualization output is presented on a stereoscopic projection wall.

#### **4. Conclusion and Future Work**

Changing the input paradigm from a mouse pointer and keyboard oriented interface to a gesture oriented touch interface extends the eXhibition:editor3D with a modern possibility for an effective planning of virtual exhibitions and museum tours. The touch interface maps the natural haptic experience of users and therefore allows an effective workflow to create exhibitions. In combination with the visualization module that is no longer a separated entity, but an integral component during the authoring process, curators can efficiently arrange and position exhibition tours. The independence of external input devices like mouse or keyboard eases collaborative work and helps to accelerate the decision making process.

An additional advantage of the multi-touch technology is its "show" factor. This factor makes the application attractive e.g. for educational purposes to awaken the interest of pupils on the topic of cultural heritage. With the innovative touch interface the young generation can be attracted to delve into the world of cultural heritage in experimenting with the presentation of exhibits, arrange their own tours and learn about the historical background during the process.

The InstantPlayer as viewer component provides a flexible tool to visualize the resulting exhibition scene on ordinary devices, as well as in complex environments like stereoscopic or CAVE setups. This flexibility allows museums to appoint the eXhibition:editor3D not only for the actual planning process of exhibitions, but also for showcases and multi-media presentations of their cultural heritage assets. With its extensible export system the application can also easily store exhibitions in formats which are suitable for different types of electronic media, e.g. to make past exhibitions accessible via a browseable internet catalogue or on mobile devices. The built-in support for the flexible and accepted exchange format X3D ensures not only interchangeability with other systems, but is also a future-proof and open format for long-term archiving of digital exhibits.

However, in some parts it becomes obvious that the application was not designed to exclusively use multi-touch input without a mouse and a keyboard device. E.g. the properties view demands some sort of keyboard input to change the different properties in entering values. Another issue is encountered in the field of collaborative work when more curators are standing around the multi-touch device to discuss the exhibition. The authoring canvas itself is independent from where it is looked at, as it uses icons and no fonts to visualize exhibits. However, for tasks like saving or loading a project, managing the content repositories, etc. it is necessary to stand on a certain side of the multi touch table to get the correct view.

Concerning future work two areas can be improved. Currently the content repositories providing the exhibits either read supported media formats directly from the file system or retrieve files from network locations. It would be beneficial to be able to access the content repository of the 3D-COFORM project. The project is currently in progress, however, supporting the upcoming infrastructure will improve the interconnectivity between cultural heritage institutions.

The second area of improvement is the authoring interface. Currently it provides an orthogonal top-view of the exhibition room. That only allows to directly translate or rotate an exhibit in one coordinate plane when using only gestures and no keyboard device. In introducing additional views all degrees of freedom would be accessible with touch gestures and the keyboard necessity would be obsolete for changing transformation values.

## 5. Acknowledgments

The furniture integrating the multi-touch monitor shown in Figure 2 is provided by Bene AG, Austria. The usage of the 3D content in the same Figure is granted by the Universal Museum Joanneum, Schloss Eggenberg, Graz, Austria and JOANNEUM RESEARCH.

## References

- [1] D. Arnold. Tools and Expertise for 3D Collection Formation. In *Proceedings of EVA 2009 Florence*, pages 94–99, 2009.
- [2] Luca Chittaro, Lucio Ieronutti, Roberto Ranon, Eliana Siotto, and Domenico Visintini. A High-Level Tool for Curators of 3D Virtual Visits and its Application to a Virtual Exhibition of Renaissance Frescoes. In *VAST10: The 11th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, volume Full Papers, pages 147–154, Paris, France, 09/2010 2010. Eurographics Association, Eurographics Association.
- [3] Eclipse Graphical Editing Framework (GEF). <http://www.eclipse.org/gef>.
- [4] Eclipse Rich Client Platform (RCP). [http://wiki.eclipse.org/index.php/Rich\\_Client\\_Platform](http://wiki.eclipse.org/index.php/Rich_Client_Platform).
- [5] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. Direct-Touch vs. Mouse Input for Tabletop Displays. In *PROCEEDINGS OF CHI 2007*, pages 647–656. ACM, 2007.
- [6] Fraunhofer Austria Research GmbH. Austria, [www.fraunhofer.at/vc](http://www.fraunhofer.at/vc).
- [7] Google Art Project. <http://www.googleartproject.com/>.
- [8] InstantReality. <http://www.instantreality.org>.
- [9] JOANNEUM RESEARCH GmbH. Austria, [www.joanneum.at](http://www.joanneum.at).
- [10] Martin Kaltenbrunner, Till Bovermann, Ross Bencina, and Enrico Costanza. TUIO: A Protocol for Table-Top Tangible User Interfaces. In *Proceedings of the 2nd Interactive Sonification Workshop*, 2005.
- [11] Victor Mateevitsi, Michael Sfakianos, George Lepouras, and Costas Vassilakis. A Game-Engine based Virtual Museum Authoring and Presentation System. In *International Conference on Digital Interactive Media in Entertainment and Arts*, pages 451–457, 2008.
- [12] Moritsch, O., et al.. Creating 3D Virtual Exhibitions. In J. Trant and D. Bearman, editors, *International Cultural Heritage Informatics Meeting (ICHIM07): Proceedings*, 2007. Published October 24, 2007 at <http://www.archimuse.com/ichim07/papers/moritsch/moritsch.html>.
- [13] Murg S., et al.. Using Standardized Methods To Present Three-Dimensional Content On The Web In The Context Of Cultural Heritage. In J. Trant and D. Bearman, editors, *Museums and the Web 2008: Proceedings*, 2008. Published March 31, 2008. Consulted June 24, 2011. <http://www.archimuse.com/mw2008/papers/murg/murg.html>.
- [14] Prasad Ramanahally, Stephen Gilbert, Thomas Niedzielski, Desirée Velázquez, and Cole Anagnost. Sparsh UI: A Multi-Touch Framework for Collaboration and Modular Gesture Recognition. *ASME Conference Proceedings*, 2009(43376):137–142, 2009.
- [15] S. Vassallo, E.; Spina. VirtualXibIT: Framework for the Venue Authoring, Population and Visualisation of 3D Virtual Exhibitions. In *VAST2009: VAST-STAR, Short and Project Proceedings*, 2009.

This page intentionally left blank



## Extended Abstracts

This page intentionally left blank

# A comparison of an open access university press with traditional presses: Two years later

Rory McGREAL<sup>a1</sup> Nian-Shing CHEN<sup>b</sup> Tim McNAMARA<sup>a</sup>  
<sup>a</sup>*Athabasca University*  
<sup>b</sup>*National Sun Yat Sen University*

**Abstract.** This study is a comparison of AUPress with three other traditional (non-open access) Canadian university presses. The analysis is based on the rankings that are correlated with book sales on Amazon.com and Amazon.ca. Statistical methods include the sampling of the sales ranking of randomly selected books from each press. The results of one-way ANOVA analyses show that there is no significant difference in the ranking of printed books sold by AUPress in comparison with traditional university presses. However, AUPress, can demonstrate a significantly larger readership for its books as evidenced by the number of downloads of the open electronic versions.

**Keywords.** Open access, OA, open publishing, university press

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=101\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=101_elpub2012)

urn:nbn:se:elpub-101\_elpub2012

---

<sup>1</sup> Corresponding Author: Rory McGreal, Athabasca University, Athabasca Canada; E-mail: rory@athabascau.ca

# Huddersfield Open Access Publishing

Graham STONE<sup>a,1</sup>

<sup>a</sup>*Information Resources Manager, University of Huddersfield*

**Abstract.** This paper presents the findings of the Huddersfield Open Access Publishing Project, a JISC funded project to develop a low cost, sustainable Open Access (OA) journal publishing platform using EPrints Institutional Repository software.

**Keywords.** open access, publishing, University Press, e-journals, culture, technology

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=106\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=106_elpub2012)

urn:nbn:se:elpub-106\_elpub2012

---

<sup>1</sup> Corresponding Author: Information Resources Manager, Computing and Library Services, University of Huddersfield, Queensgate, Huddersfield, UK; E-mail: [g.stone@hud.ac.uk](mailto:g.stone@hud.ac.uk)

# Freemium as a Sustainable Economic Model for Open Access Electronic Publishing in Humanities and Social Sciences

Pierre MOUNIER<sup>a,1</sup>

<sup>a</sup>*EHESS/Cleo*

**Abstract.** Between the two paths of open access - green and gold - the later is the harder to develop and has the less support from the research community. The main difficulty is about finding a sound economic model. Open Access journals usually depend on two funding sources: subsidies and/or donations from institutions and publication fees from research units in the authors-pay model. These two ways of funding open access journals and books proved effective in some cases (Plos), but are not flawless. The Center for Open Electronic Publishing, a french initiative for open access publishing in humanities and social sciences, has recently developed a new economic model based on “freemium” for its full open access journals and books series, in order to address two issues: improve their economical soundness and give them more visibility in libraries. Freemium, the contraction of “free” and “premium”, preserves open access to information together with the marketing of premium services.

**Keywords.** open access, humanities, social sciences, freemium, economic models

Full text available at :

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=107\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=107_elpub2012)

urn:nbn:se:elpub-107\_elpub2012

---

<sup>1</sup> Corresponding Author : Centre for Open Electronic Publishing, EHESS, 190-198 avenue de France, 75244 Paris cedex 13 ; E-mail : pierre.mounier@gmail.com

# Some trends in Electronic Publication and Open Access in Portuguese History Journals

Maria Cristina GUARDADO<sup>a,1</sup> and Maria Manuel BORGES<sup>b</sup>

<sup>a</sup> *Escola Superior de Tecnologia e Gestão de Águeda – Universidade de Aveiro, Portugal*

<sup>b</sup> *Faculdade de Letras da Universidade de Coimbra; CETAC.MEDIA, Portugal*

**Abstract.** In the last decades, ICT development and the use of computer networks led to major changes in the way scientists communicate the results of their researches. One of the most important transformations occurred in scholarly communication through the acceleration in the unbounded disclosure of scientific information (open access). However, the adoption of new channels varies according to the scientific areas. This study is part of an ongoing research that aims to understand the impact of digital media in the mechanisms of production and dissemination of scientific knowledge within a specific scientific community: historians. This paper will look at whether and how the history journals in Portugal are adopting electronic publishing and providing open access to their contents. To operationalise the study, thirteen R&D Units were contacted and asked to provide a list of the journals published and the format used. Furthermore, the Directory of Open Access Journals (DOAJ), Latindex and European Reference Index for the Humanities (ERIH) were researched. Finally, the websites of Portuguese higher education institutions with History Departments were consulted, with the aim of identifying any periodicals that might not have been previously identified. At the end, twenty-two titles were considered. Data seems to show that universities and History R&D Units increasingly value the free electronic access to research results produced by historians.

**Keywords.** Electronic publication, open access, scholarly communication, history, Portugal

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=108\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=108_elpub2012)

urn:nbn:se:elpub-108\_elpub2012

---

<sup>1</sup> Corresponding Author: Maria Cristina Guardado, Escola Superior de Tecnologia e Gestão de Águeda, Apartado 473, 3754 – 909 Águeda, Portugal; E-mail: [cguardado@ua.pt](mailto:cguardado@ua.pt).

# Development of a publishing framework for living open access textbooks

Anita Eppelin<sup>a,1</sup>, Richarda Böttcher<sup>b</sup>

<sup>a</sup>*German National Library of Medicine (ZB MED), Cologne, Germany*

<sup>b</sup>*Trauma Hospital Berlin (UKB), Berlin, Germany*

**Abstract.** We are suggesting a framework for open access publishing of comprehensive, elaborately structured digital textbooks, potentially enriched by extensive non-textual data. The framework will comprise of a software toolbox containing a collaborative authoring platform, manuscript workflow system, tools for editorial work, presentation platform, updating workflow, interfaces for dissemination and long-term preservation. In addition to the software toolbox itself, guidelines and standards for all aspects of digital textbook publishing, including peer review, metadata and editorial procedures, legal issues as well as a business model will be developed, resulting in a best practice guide to digital textbook publishing. The software and workflows will not be developed from scratch, but build upon existing open-source software, using the example of a scientific textbook of hand surgery as its first use case. This paper introduces the project and gives a prospectus of the proposed framework.

**Keywords.** open access, publishing, open educational resources, textbooks, medicine, hand surgery

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=111\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=111_elpub2012)

urn:nbn:se:elpub-111\_elpub2012

---

<sup>1</sup> Corresponding author: Anita Eppelina, German National Library of Medicine (ZB MED), Cologne, Germany; E-mail: [eppelin@zbmed.de](mailto:eppelin@zbmed.de).

# OAPEN-UK: an Open Access Business Model for Scholarly Monographs in the Humanities and Social Sciences

Caren MILLOY<sup>a,1</sup> and Graham STONE<sup>b</sup> and Ellen COLLINS<sup>c</sup>

<sup>a</sup>*Head of Projects, JISC Collections*

<sup>b</sup>*Information Resources Manager, University of Huddersfield*

<sup>c</sup>*Research Officer, Research Information Network*

**Abstract.** This paper presents the initial findings of OAPEN-UK, a UK research project gathering evidence on the social and technological impacts of an open access business model for scholarly monographs in the humanities and social sciences.

**Keywords.** open access, scholarly monograph, humanities and social sciences, culture, technology, business models, research study

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=105\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=105_elpub2012)

urn:nbn:se:elpub-105\_elpub2012

---

<sup>1</sup> Corresponding Author: Head of Projects, JISC Collections, Brettenham House, 5 Lancaster Place, London, UK; E-mail: [c.milloy@jisc-collections.ac.uk](mailto:c.milloy@jisc-collections.ac.uk)



# Accessibility and Self Archiving of Conference Articles: A Study on a Selection of Swedish Institutional Repositories

Peter LINDE<sup>a</sup>, Jörgen ERIKSSON<sup>b</sup>, Lars KULLMAN<sup>c 1</sup>, Margareta FATHLI<sup>d</sup>,  
Klemens KARLSSON<sup>d</sup>, Marjatta SIKSTRÖM<sup>e</sup>, Ylva SKÖLD<sup>f</sup>, Ingela TÅNG<sup>e</sup>

<sup>a</sup>*Blekinge Institute of Technology*

<sup>b</sup>*Lund University*

<sup>c</sup>*Chalmers University of Technology*

<sup>d</sup>*KTH Royal Institute of Technology*

<sup>e</sup>*Stockholm University*

<sup>f</sup>*Luleå University of Technology*

**Abstract.** The main purpose of this project has been to examine the accessibility of refereed conference articles and the OA- and publishing policies of conferences in order to in this way elucidate different aspects concerning self-archiving in Swedish institutional repositories. For this purpose, the project participants have examined a number of conferences and references to conference articles via their institutional repositories during a specific time period and described these from the perspective of a common scheme. The study has showed that the local institutional repositories fill an important role to make conference publications visible. We have found that ca. 50 % of the conference papers can be published as postprints in our institutional repositories. We have noticed that ca. 15% of the studied conference articles are not available at all. It is, therefore, of great importance to use local institutional repositories as a publishing channel, not only for primary published material such as dissertations and reports, but also as a source for finding these conference articles “without a home”. Between 20-25 % of the examined articles were found in some type of OA archive; ca. half of these were found in one of the project participants’ own institutional repositories. This indicates that the publishing database of respective higher education institution is an important factor for open accessibility. Ca. 10% of the conferences in the study had an explicit OA policy or expressed such a policy by openly making conference articles accessible on their conference sites. A big problem when it comes to self-archiving of conference articles is the lack of information about OA policy. The landscape of conference publishing is complex and the self-archiving of documents from conferences is very time-consuming. Above all, we would wish a policy resource for conferences similar to the SHERPA/RoMEO. At present, however, there is no other alternative than scrutinizing the conferences’ copyright information to the authors and from this attempt to draw conclusions about possible self-archiving.

---

<sup>1</sup> Corresponding Author: Lars Kullman, Chalmers University of Technology, The Library, Sweden;  
E-mail: lars.kullman@chalmers.se

To facilitate the future handling and classification of conference articles in Swedish institutional repositories a number of recommendations are suggested.

**Keywords.** Open Access, conference articles, institutional repositories, self-archiving

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=102\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=102_elpub2012)

urn:nbn:se:elpub-102\_elpub2012

# Text vs Visual Metaphor in Mobile Interfaces for Novice User Interaction

Chrysoula GATSOU<sup>a,1</sup>, Anastasios POLITIS<sup>b</sup> and Dimitrios ZEYGOLIS<sup>a</sup>

<sup>a</sup> *School of Applied Arts Hellenic Open University, Greece*

<sup>b</sup> *Graphic Arts Technology Faculty of Fine Arts and Design, TEI of Athens, Greece*

**Abstract.** This paper explores the effectiveness, efficiency and ease of use of two alternative interface interactions in a mobile tablet application. More specifically, the study employs a navigation task which novice users were required to complete by means of two types of main menu button, one utilizing text and the other utilizing a visual metaphor. Furthermore, little evidence exists on the effect of the use of metaphor on users of various ages and levels of computer experience. Our results show that young users prefer buttons that display a visual metaphor in the form of an icon, in contrast to older participants, who preferred the button that employed text. They also show that the users' performance is dependent upon age.

**Keywords.** Mobile tablet, visual metaphor, interface, interaction design, novice user.

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=103\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=103_elpub2012)

urn:nbn:se:elpub-103\_elpub2012

---

<sup>1</sup> Corresponding Author: Chrysoula Gatsou, Botasi 25, 18537, Piraeus, Greece. E-mail: [cgatsou@teiath.gr](mailto:cgatsou@teiath.gr)

# Digital Reading: The Transformation of Reading Practices

Gustavo CARDOSO<sup>a,1</sup>, Carla GANITO<sup>b</sup> and Cátia FERREIRA<sup>b</sup>

<sup>a</sup>*CIES, ISCTE-IUL*

<sup>b</sup>*CIES, ISCTE-IUL and UCP*

**Abstract.** The paper aims to present the preliminary results of a two-year project having as scope the future of the book and libraries stemming from the current reading practices in Portugal. In the scope of the project, the presentation of the findings will be focused on the mobile consumption practices in Portugal. This research is based on a mixed methodology: a quantitative survey – Network Society in Portugal – articulated with a qualitative analysis of the discourses of the representatives of what Thompson calls the *publishing chain* (librarians, publishing houses, authors, and content and soft/hardware providers). To understand the impact of mobile devices on reading practices is crucial for libraries and publishing houses. Mobile devices offer augmented mobility – a mobility that is connected, networked and collaborative. Although the hype is currently around eBooks, we are still faced with a market where the vast majority still reads books on paper. The sales of devices have exploded but eBooks are lagging behind. What do people use their tablets, iPads, and eReaders for? What are they reading and where? How do they articulate their readings with other media and cultural consumptions? Those are the central questions that we are aiming at answering.

**Keywords.** digital reading, mobility, publishing chain, Portugal. libraries

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=112\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=112_elpub2012)

urn:nbn:se:elpub-112\_elpub2012

---

<sup>1</sup> Corresponding Author: CIES, ISCTE-IUL – Avenida das Forças Armadas, 1649-026 Lisboa – Portugal; E-mail: [leituradigital@googlegroups.com](mailto:leituradigital@googlegroups.com).

# Are Turkish Universities Ready for E-learning: A Case of Hacettepe University Faculty of Letters

İrem SOYDAL,<sup>a</sup> Gülten ALIR<sup>b</sup> and Yurdagül ÜNAL<sup>a,1</sup>

<sup>a</sup>*Department of Information Management, Hacettepe University*

<sup>b</sup>*Department of Information Management, Yıldırım Beyazıt University Ankara, Turkey*

**Abstract.** Universities play an important role in lifelong learning. E-learning is one of the key elements in today's networked and knowledge oriented world. It is crucial to know whether the universities are ready to transform their courses to e-learning systems. In this paper the e-learning readiness of the academic staff of Hacettepe University Faculty of Letters (HUFL) is investigated. A 37-item questionnaire along with some demographic questions is used for obtaining the data. Results show that title might be a significant factor for e-learning readiness and in general, HUFL academic staff are not ready for the e-learning environment. The findings of this study will help to conduct a larger study throughout Turkey in order to determine a model for course of action for transition to an e-learning system.

**Keywords.** e-learning readiness, higher education, academic staff readiness

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=110\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=110_elpub2012)

urn:nbn:se:elpub-110\_elpub2012

---

<sup>1</sup> Corresponding Author: Yurdagül Ünal, Department of Information Management, Hacettepe University, 06800 Beytepe, Ankara, Turkey; Email: yurdagul@hacettepe.edu.tr.

# The Biographical Portal of the Netherlands

Experiences in collecting and curating data from the web and elsewhere

Jelle GERBRANDY<sup>a,1</sup> and Els KLOEK<sup>b</sup>

<sup>a</sup>*Gerbrandy SRL, Torino, Italy*

<sup>b</sup>*Huygens Institute for Netherlands History, Den Haag, Netherlands*

**Abstract.** The Biographical Portal aims to collect, organize, and make available all biographical information that is available about Dutch people. In this paper, we describe the design process and what choices we made, the tools we used and may be useful for a larger audience, and “lessons learned” - things that we feel we have done well, and things we would have done differently if we had known from the start what we know now.

**Keywords.** Biography, Standards, Cultural Heritage, Data Aggregation

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=104\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=104_elpub2012)

urn:nbn:se:elpub-104\_elpub2012

---

<sup>1</sup> Corresponding Author: Jelle Gerbrandy, Gerbrandy SRL, Torino, Italy. E-mail: [jelle@gerbrandy.com](mailto:jelle@gerbrandy.com)

# The Future of Digital Magazine Publishing

Dora SANTOS SILVA<sup>a,1</sup>

<sup>a</sup>*Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa*

**Abstract.** The publishing industry went through more structural changes in the past ten years than in the whole second half of XX century. In magazine publishing in particular, since the emergence of the first sites to complement the print edition in the 1990s to the release of iPad in 2010, these media have enjoyed numerous opportunities provided by new technologies, which, in turn, demanded the reinvention of editorial and business models. This article seeks to bring together some of the opportunities that digital magazines should embrace (or try) to be sustainable in the future and get the readers' attention and loyalty.

**Keywords.** Digital journalism, digital magazine publishing, magazine journalism, business models, digital magazine publishing software.

Full text available at:

[http://elpub.scix.net/cgi-bin/works/Show?\\_id=109\\_elpub2012](http://elpub.scix.net/cgi-bin/works/Show?_id=109_elpub2012)

urn:nbn:se:elpub-109\_elpub2012

---

<sup>1</sup> Corresponding author: Dora Santos Silva, Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, Lisboa, Portugal; E-mail: [dora.santos@netcabo.pt](mailto:dora.santos@netcabo.pt)

This page intentionally left blank



## Subject Index

Abc	82	history	120
academic publication	12	historical archives	52
academic staff readiness	127	humanities	119
Africa	12	humanities and social sciences	122
authorship	43	hybrid publishing	94
biography	128	hypertext	52
business models	122, 129	information and communication	
co-authorship	43	technologies	52
collaboration	43	information and technology	3
collaborative work	82	information retrieval	34
commons	3	information science	3
conference articles	124	institutional repositories	24, 124
copyright	3, 24	interaction design	125
creative commons	3	interactive design	104
cultural heritage	128	interface	125
culture	118, 122	knowledge networks	72
data aggregation	128	knowledge representation	72
designers as authors	43	libraries	126
developing countries	12	library and information science	34
digital book applications	43	linked data	72
digital journalism	129	magazine journalism	129
digital library	62	medicine	121
digital magazine publishing	129	metadata	62
software	129	mobile tablet	125
digital preservation	104	mobility	126
digital reading	126	multi touch	104
document retrieval	62	music publishing	82
e-book	94	music transcription	82
economic models	119	novice user	125
editorial design	94	OA	117
educational goals	34	ontology	62
e-journals	118	open access	12, 117–122, 124
e-learning readiness	127	open access movement	24
electronic publication	120	open archives	12
electronic scientific		open educational resources	121
communication	24	open publishing	117
enhanced ebook	43	open repository	12
e-science	72	Portugal	120, 126
freemium	119	post-custodial	52
hand surgery	121	printed book	94
health sciences	34	public university	24
higher education	127	publishing	118, 121
		publishing chain	126

research study	122	terminological knowledge bases	72
scholarly communication	120	textbooks	121
scholarly monograph	122	University Press	117, 118
search engine	62	virtual exhibitions	104
self-archiving	124	visual metaphor	125
semantic publishing	72	Web 2.0	3
semantic web	72	Web publishing	52
social sciences	119	wiki	82
standards	128	X3D	104
technology	118, 122		

## Author Index

Aitken, J.	43	Kloek, E.	128
Alir, G.	127	Kullman, L.	123
Almeida, J.J.	82	Lavesson, N.	v
Baptista, A.	v	Linde, P.	v, 123
Bastos, F.	24	Marcondes, C.H.	72
Borges, M.M.	94, 120	Martin, C.	43
Böttcher, R.	121	Mass Araya, E.R.	3
Brito, M.A.	v	McGreal, R.	117
Cardoso, G.	126	McNamara, T.	117
Carvalho, N.R.	82	Milloy, C.	122
Chalabi, L.	12	Möstl, R.	104
Chen, N.-S.	117	Mounier, P.	119
Collins, E.	122	Nisheva-Pavlova, M.	62
Dahmane, M.	12	Oddone, N.	24
Derler, C.	104	Oliveira, J.N.	82
Eggeling, E.	104	Pavlov, P.	62
Eppelin, A.	121	Politis, A.	125
Eriksson, J.	123	Santos Silva, D.	129
Fathli, M.	123	Serrano, A.	34
Fellner, D.W.	104	Sikström, M.	123
Ferreira, C.	126	Silva, A.C.	94
Ganito, C.	126	Sköld, Y.	123
Gatsou, C.	125	Soydal, Í.	127
Gerbrandy, J.	128	Stone, G.	118, 122
Guardado, M.C.	120	Tång, I.	123
Hecher, M.	104	Ünal, Y.	127
Jorente, M.J.V.	52	Vidotti, S.	3, 24
Karlsson, K.	123	Zevgolis, D.	125

This page intentionally left blank