FROM PAPER TO BYTES: DIGITAL KNOWLEDGE SHARING
A Multi Level Approach to Digitalization of Dictionary of the Italian Resurgency by Michele Rosi for a Collaborative Discipline of the Historical Studies

Letizia Bollini

Psycology Department, University of Milano-Bicocca, P.za dell'Ateneo Nuovo 1, 20126 Milano, Italy

Keywords: Knowledge management, Knowledge sharing, Collaboration, e-Government.

Abstract: Archives have a lot of documental collections relevant for studies; however items within are difficult to access or consult. Web will be an efficient tool to disseminate the content of documents rare or undisclosed for preservation purpose.
The research "Dictionary of the Italian Resurgency" by Rosi has allowed to public, historians to access four Volumes fundamental to studies on Italian Modern History.
Project has been developed in collaboration with *Civiche Raccolte Storiche di Milano* (Milano Historical Downtown Collections) following a proposal in the framework of the Network of Historical Museums.
The project was a transformation in digital form of books; then a web system was implemented to publish, to search text, using booth a textual format (indexed by Google and WW searchable) and the original printing (downloadable in jpg format). A search system using key criteria allows to crosscorrelate dates, overcoming the labour intensive traditional examination. A distributed system for data insertion allows many students collaborate, revising and upgrading the contents.
Next step will be the creation of links between the contents of the first tree volumes: *Peoples* and the content of the fourth: *Events* where peoples quoted in the dictionary were involved as main actors.

1 INTRODUCTION

The historical collections, the libraries, the museums have a wealth of history and valuable heirlooms of past times; some documents are even fundamental in order to collect background information for the historical research.

A relevant part of the historical heritage however is not available for consultation as a result of different situations:

- the precarious conditions of preservation or the perishability of the historical documents;

- the insufficient number of the document copies for loan;

- a rather inaccessible reading rooms (i.e. inadequate opening hours, placement a long way off, and so on)

- deficiency of structures, staff shortage, inadequacy of the spaces.

The core problems are always the limitations for the physical protection of the original documents and the inadequate access service to the collections (Camurri, 2008).

On the opposite the new information technologies of the web are producing accessibility to remote information, are enabling the search of documents outside the actual physical domain, and allows at the end to share the knowledge.

The pilot project "Dizionario Nazionale del Risorgimento Italiano" is a research pathway that, starting from a real case is aimed to produce a possible general model to make accessible and shared the historical culture and the documental heritage by means of the web tools.

Starting from a real case history the project has been developed following the guidelines of the *Minerva working Group 5* for the quality of the cultural web sites (Filippi, 2004).

2 RESEARCH FRAMEWORK

The project has been started inside the framework of a larger research pathway established in 2005 based on the Network of the Museums of Milan Down Town: an historical network of museums and of

public art galleries located in the historic city center of Milan.

The seven museums (Museum of the Risorgimento, House of Alessandro Manzoni, Theatrical Museum alla Scala, Contemporary Museum of History, Museum of Milan, Museum Bagatti-Valsecchi) exploited the network to create a territorial, real and virtual system in order to get visibility, to have efficient tools of communication and visibili*y* and to maximize the investments - always inadequate for the public Italian institutions - shared between the several museums to create common and shared web platform.

As a consequence of the success of this first pilot project sponsored and founded by the Milano town council, the Municipal Historical Collections on initiative of the director Dott. Roberto Guerri decided to invest in a series of projects, not more on the territorial network, but on specific Documental Archives. Many projects where started between 2007 and 2008; among the projects one was aimed to render in digital form and to allow the on line consultation of the Rosi Dictionary.

The *National Dictionary of the Italian Resurgency* is a fundamental work for the Italians national historical studies; the work is composed by 4 volumes subdivided in two parts.

The first three volumes are dedicated to the personalities of the Risorgimento history, the fourth one refers to facts, that is to say the historical events, the battles and so on that have as protagonists the figures cited in the preceding three volumes.

Editor of this work was Michele Rosi; he was involved in the compilation starting from the year 1930 until its death; many authors were involved that continued and finished the work. The dictionary is regarded as a fundamental work for the historical research for the abundance of the mentioned personalities, for the completeness of the information and for the extension of the in-depth studies.

Few copies of the work are still preserved by Italian Historical Archives; amidst them the Foundation of the Italian Risorgimento Museum, member of the Municipal Historical Collections of Milano, that collects the library, the periodicals library and the Documental Collections.

3 TECHNOLOGY AS A NEW FORM OF KNOWLEDGE CONSERVATION

A first tentative of search was tied to a Ph.D. thesis in modern and contemporary history (in the late '90s): the volumes have been already scanned in graphical format (every single page was saved in format jpg in high resolution) rather than through OCR; the material was indeed barely useable.

The research project arose from this fundamental problem: how to make somehow operable this collection of information already digitalized, but in a not correct format and how to translate them in a database.
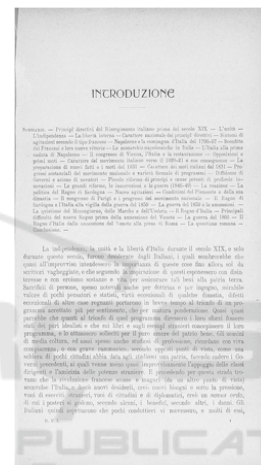

Figure 1: First version: scanned paper.

In this preliminary phase a first software was implemented: exploiting the peculiarity of the more recent versions of Acrobat Reader™ and the possibility to select blocks of text inside the jpg image, the translation of the contents from the graphic format to an editable text was efficiently performed.

The preparation and the importation of the data was finished by the human supervision of the curators of the project working in close connection with (coordinating) a group of university students involved in the search; texts were submitted to a first revision and a final correction of the drafts.

3.1 Technology as a Form of Shared Information

The preparatory phase has come to results, the project was continued in order to fully exploit the

opportunities of the network and of the digital
technologies:

- data were laid on line for the dissemination and
the sharing of the contents (accessibility);

- data were done searchable via web
(availability);

- actual criteria of the historical domain were set
up to search inside the same system
(searcheability);

- correlation of the contents (ipertextuality)
inside the different volumes;

- a system of administration and management of
the contents has to be implemented based on a
widespread editing (community);

- an user interface very focussed on the contents
search and presentation was developed. The
simplicity of use and a look&feel graphic design
were the prevalent guidelines.

3.2 Knowledge Accessibility

The first aim pursued by the project was to have on
line the contents of the Dictionary. The network for
its global pervasiveness - as stated by Rifkind -
(2003) is the most relevant medium to share the
knowledge. To develop a site and a publishing
system of the four volumes content is a way to
spread out this historical wealth from the narrow
space of a library, limited in a physical place, and in
consultation hours and with formal and inefficient
procedures of transcription and coping. The project
has given the Dictionary a new universal access and
diffusion, applying virtuous models that at the
present are developed to support the Cultural
Institutions pertaining to the public Administration

Not only: the web site achieves a double effect:
offers the scholars, historians and researchers that
are the most frequent users an easy and efficient way
to investigate the contents; in addition web
disseminates the historical knowledge of the Italian
Resurgency to the common people. They are
allowed to recognize the historical roots of the
country, to search for ancestors and events of the
Resurgency heritage and to get acquainted with the
process of building our Nation.

The accessibility, how underlines Visciola
(2002) is not based only on the possibility to find out
some data in network, but is based mainly on the
organization and the elaboration of published data.
Opportunity will be offered to non specialist but
interested users to benefit and enjoy the access to the
data. The archive open again on a demonstrative
virtual space thanks to the diffusion technology
becomes again collective heritage and open memory.

3.3 Availability and Searcheability

If the on line publishing makes the data available,
only a new arrangement according to complex
search criteria makes data searchable.

In the printed dictionary the returns, marked by
headwords and from notes that cross connect across
the different volumes, are still a static tool of a
consultation that is still sequential or rather punctual.

Figure 2: Internal research system.

In the framework of a digital version instead the
database is seacheable applying different criteria of
selection; criteria can be used singularly or cross
correlated to obtain even more specific and precise
results (for example date of birth, place of birth,
keywords, like the name of a battle or of an event,
name of the curator of the biographical note etc.).

As an example, the name of a battle can be
searched, in order to verify the name of the peoples
involved in the event , and at the same time the dead
date, in order to exclude peoples that died before the
battle or before an event.

3.4 Hypertext as Knowledge Extension

Furthermore the system allows a sophisticated
search covering all the contents and crossing the
borders of all the volumes of the dictionary. The
actual hypertextuality of the dynamic system of
search of the DB allows to create and discover
unthinkable correlations or that can be inferred
with extreme difficulty working on the paper version
support.

The extent of the contents, the number of the
personalities, the presence of hononymies, the
necessity of to have significant results for the search
domain have been the foundations of the
implementation of the present system and will be
taken into account in the design of an upgrade of the

software and upgrade of the contents for the next release.

Using the correlations produced from the search, the distributed editorial staff that is in charge of revision of the contents will introduce using the ad hoc software hypertextuals connections between the headwords of the Dizionario dedicated to the personalities and the events of 4° volume. The operation is aimed to support student and historians with a complete set of cross correlations between personalities and events, giving an easy and efficient tool to exploit the Dictionary Historical thesaurus.

Figure 3: Research out put.

3.5 Technology and Historicity of the Sources

Beyond to be a tool of research and a fundamental historical source of documentation the Dictionary is by oneself an heirloom belonging to a specific period with a peculiar graphic and editorial quality (typographic quality).

In the digital rendering the original appearance has been carefully preserved.

The graphic interface, based on a extended use of XHTML and CSS in order to create structural and semantic hierarchy of the code and of the contents, offers a quality almost materic of the page and plays with caution on the typographical appearance and on pagination style.

Next to the text paged up in code the opportunity is given to download in JPEG format the scanning of the original page of the text, restoring the value not only to the contents, but also to the historical apparence of their edition.

Figure 4: Internal fomat

4 CULTURAL DISSEMINATION AND DISTRIBUTED EDITING

The finished project has produced a relevant two sided benefit.

An historical knowledge treasure was quite inaccessible; a virtual image of the book can be accessed now by network in a complete and efficient way (Colombo, 1986).

In addition the project has generate a positive innovation spin-off effecting on the operative practices of the Municipality Historical Collections

The translation in digital form of the Dictionary and the web availability has produced the opportunity to work on the cataloguing and on the reorganization of the archives and on the classification of the collected material. These benefits will be positively extended also on the consultation and reconnaissance of the similar collections present and available inside books and newspaper collections of the Municipal Historical Collections.

The rediscovery of the heritage has fostered the organization of events pursuing the communication and of activities in order to focus the public attention on the project and on the project contents, reviving the interest and the studies on the project. The activities have been carried out by means of the Museum periodical channels of communication (newspapers).

5 CONCLUSIONS

The research project is based on a specific case history; however the project can be generalized - easily extended - and represents a reproducible experience for different disciplines.

The stratified/multi step approach to the problem of the knowledge sharing allows actually to realize in many phases the translation in digital form and the consequent sharing of the cultural heritage avoiding the risk that the historical documents will remain unknown and on the worst case will get forgiven and lost.

A first level is the easy and general requirement of web availability of the cultural information that allows to disseminate specialized information to the public at large, students of the specific discipline or crossing the border to different disciplines and reaching for interested peoples.

This is an important level, fundamental in the cultural dissemination, but in some way generic and rather oversimplified in respect to the study of sources and of documents.

The second level - that of the implementation of search tools specific of the Data Basis contents and mainly focussed on the research and cultural domain the historical documents pertains to.

The content publishing on the Web is then more an efficient operation tool then a simple low level itemization.

The third refers to the creation of a scientific community of interest that starts from the distributed editing and produces links and exchanges extending to a wider field and longer lasting than the original project.

A last level produces the maximization of the resources and optimization of the investments.

The actions of exploiting innovation technologies and of designing user experience will be carriers of innovation and winning strategy to overcome the chronic lack of grants for the public projects fostering the sharing and the exploitation of the treasures of a cultural heritage.

REFERENCES

Camurri, D., 2008. Il progetto "Una città per gli archivi". In *Storia e futuro. Rivista di Storia e Storiografia,* n.17, Ibc, Bologna

Colombo, F., 1986. *Gli archivi imperfetti. Memoria sociale e cultura elettronica*, Vita e Pensiero, Milano.

Rifkin, J., 2000. *L'era dell'accesso. La rivoluzione della new economy,* Oscar Mondadori. Milano

Filippi, F., 2004. *Manuale per la qualità dei siti Web pubblici culturali,* Ministero per i beni e le attività culturali. Roma

Visciola, M., 2002. *Progetto CABI. Accessibilità e usabilitàper la lettura dei testi digitali*

[http://marciana.venezia.sbn.it/catalogazione.php?sst=47]