

# Dig *Italia*

Numero 2 - 2006

Rivista del digitale nei beni culturali

ICCU-ROMA



ICCU

Istituto centrale per il catalogo unico  
delle biblioteche italiane e per le informazioni bibliografiche  
[www.iccu.sbn.it](http://www.iccu.sbn.it)

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista  
è ammessa con obbligo di citazione

# DigItalia

Rivista del digitale nei beni culturali  
Numero 2 - Dicembre 2006

*In copertina:*

L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.c. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicicabile all'Apollo di Leochares (IV sec. a.c.)

**Direttore Responsabile**

Marco Paoli

**Comitato di Redazione**

Anna Maria Maiorano Mandillo (coordinatore)

Giovanni Bergamin

Simonetta Buttò

Marta Cardillo

Eva Gilmore

Andrea Giuliano

Cristina Magliano

Alessandra Ruggiero

Mario Sebastiani

Vittoria Tola

Varo Augusto Vecchiarelli

**Segreteria**

Antonella Corradi

Lucia Basile

Federico Circi

Federica Amerio

**Grafica & Impaginazione**

R&R di Roberta Micchi

**Produzione e Stampa**

Futura Grafica 70 s.r.l.

Via Anicio Paolino, 21 - 00178 Roma

**Editore**

ICCU - Roma

*Istituto centrale per il catalogo unico  
delle biblioteche italiane*

*e per le informazioni bibliografiche*

Viale Castro Pretorio, 105

00185 Roma

T +39 06 49.89.484

F +39 06 49.59.302

[www.iccu.sbn.it](http://www.iccu.sbn.it)

In attesa di registrazione al Tribunale di Roma.



ICCU

**Comitato Scientifico**

Osvaldo Avallone

Armida Batori

Rossella Caffo

Michele Ciliberto

Gabriella Contardi

Flavia Cristiano

Gianfranco Crupi

Gisella De Caro

Maurizio Fallace

Gigliola Fioravanti

Antonia Ida Fontana

Paolo Galluzzi

Daniela Grana

Tullio Gregory

Maria Guercio

Mauro Guerrini

Hans Hofman

Salvatore Italia

Claudia Leoncini

Claudio Leonardi

Franco Lotti

Cristina Magliano

Anna Maria Maiorano Mandillo

Massimo Menna

Maurizio Messina

Massimo Pistacchi

Giuseppe Proietti

Amedeo Quondam

Antonia Pasqua Recchia

Aldo G. Ricci

Seamus Ross

Padre Stefano Russo

Marco Santoro

Maria Rita Sanzi Di Mino

Luciano Scala

Salvatore Settis

Giuliana Sgambati

Giovanni Solimine

Maria Carla Sotgiu

Laura Tallandini

Anna Maria Tammaro

Costantino Thanos

Paul Weston

# SOMMARIO

dicembre 2006

## EDITORIALE

**Il primo anno di *Digitalia*** 7  
di Marco Paoli

## SAGGI

**Introduction to library standards  
and the players in the field** 11  
di Alan Hopkinson

**Le qualità delle immagini  
nei progetti di digitalizzazione** 22  
di Franco Lotti

**Biblioteche digitali e musei virtuali** 38  
di Franco Niccolucci

**Prospettive di rinnovamento  
della Legge sul diritto d'autore** 52  
di Giuseppe Corasaniti

**Granularità: un percorso di analisi** 60  
di Maurizio Zani

## PROGETTI

**Verba *Manent*: Teoria e prassi  
della digitalizzazione dei documenti sonori  
e video della Discoteca di Stato-  
Museo dell'audiovisivo (prima parte)** 131  
di Francesco Aquilanti, Massimo Baldi  
e Massimo Pistacchi

**Progetto InterPARES2:  
il case study sulle *moving images*** 149  
di Isabella Orefice

**La Biennale di Venezia verso un futuro in digitale** 163  
di Giovanna Pasini

**La raccolta dei siti web:  
un test per il dominio "punto it"** 170  
di Giovanni Bergamin

---

## **DOCUMENTI**

**Convegno internazionale:  
*Conservare il digitale. Un confronto internazionale*** 177  
**Asolo, 29 settembre 2006**  
di Laura Ciancio

---

## Il primo anno di *Digitalia*

Il nostro progetto, così come dichiarato nell'editoriale del n.0, si proponeva di colmare un vuoto nel panorama dell'editoria specializzata italiana. Ci sembra ora di poter affermare che quel vuoto *Digitalia* lo ha effettivamente colmato. Il merito va agli autori e ai membri del comitato scientifico. Senza il loro impegno assiduo ed estremamente qualificato, la rivista non esisterebbe. Ma sono state molto importanti anche le numerose espressioni di consenso ed approvazione che ci sono pervenute da molte parti, sia dentro il Ministero che fuori.

*Digitalia* è la prima, tra le tante riviste che vengono prodotte oggi nel mondo dell'editoria specializzata sulle biblioteche, a porre al centro dell'indagine proprio le tematiche relative alle moderne tecnologie del digitale. Con una caratteristica in più: che il campo di ricerca si estende al di là dei confini 'bibliotecari', per tentare di abbracciare l'intero settore del 'digitale nei beni culturali'.

Questa, naturalmente, ha finito per essere una scelta obbligata dalla natura stessa delle materie trattate. Le tecniche, gli strumenti, gli standard, le metodologie e i problemi del digitale sono per così dire trasversali, e musei, archivi e biblioteche, considerati sotto questa ottica, hanno più elementi in comune che non elementi di differenziazione.

Il primo anno è appena trascorso.

Marco Paoli



## Saggi



# Introduction to library standards and the players in the field

**Alan Hopkinson\***

*Systems and Bibliographical Services, Middlesex University Learning Resources*

*What are standards? Have you ever wondered why so many standards begin with "Z"? The present article wishes to provide an answer to these and many other questions you always wanted to know about standards. Standards can be de jure, de facto and proprietary, though there is not always clear cut as to which category a standard belongs to. In the area of librarianship and information management, many standards are guidelines advising particular practices: the subject which they seek to standardise is not one which has to be uniform for the situation to work. Standards are often conceived by professional associations or national institutions needing to promote standard methodologies within their constituency. They are then proposed to a national standards body before being offered to ISO for international standardisation. In the library and information field, the interest lies in each country's own national standard, in ISO standards and in NISO standards, which though American national standards sometimes get used internationally. Standards in the field can be grouped by their applications: describing and identifying information resources, information exchange, managing collections, and delivering services. The article provides an overview of the main standards in use in these four areas of work, and of the bodies which developed or adopted them. As a conclusion, the author looks at standard development in emerging areas such as RFID technology and metadata. Standards are an important part of library and information work, and will become even more so in an increasingly connected electronic environment. Library and information professionals will need to understand which standards apply to their work, and will have to play a significant role in the development of standards concerned with information management and retrieval.*

## Introduction

**H**ave you ever wondered why so many standard numbers begin with "Z"? Z39.50, which is often referred to as "Search and Retrieve" is the best known but there are also (to mention just two examples) Z39.71 *Holdings statements for*

\* Alan Hopkinson, head of Systems and Bibliographical Services, Middlesex University Learning Resources, represents the British Standards Institution on an ISO Committee and is chair of IFLA's Permanent UNIMARC Committee. He wishes to acknowledge the assistance of David Haynes of Atkins Consultants and Pat Harris, formerly of NISO, in the preparation of this document.

*bibliographic items and Z39.48-1992 (R2002) Permanence of paper for publications and documents in libraries and archives.*

If you read on you will find the answer to this and many other questions you always wanted to know about standards.

What are standards? Amongst standards you will find that some are *de jure*, some *de facto* and others *proprietary*.

To deal with *de jure* first, which means “according to the law”, within almost every country there is one organization tasked with being the national focal point for standards. In some countries this has a legal status as being part of national law, in others the body acts by common consent. In the United States there is a body for each industry which has responsibility delegated from the American National Standards Institute (ANSI). In the field of library and information this is the National Information Standards Organization (NISO), which covers also archives, records management, museums and some areas of publishing. In the United Kingdom, the situation is clearer because the British Standards Institution (BSI) deals with all areas of standardization. Standards which are approved by BSI have to go through various procedures to ensure that they are really the consensus and technically sound and then they have a certain status, in some fields a legal status if enshrined in law – but it is the law that endows that – though in the case of library standards they are generally advisory or recommendations for good practice. In Germany the standards body is DIN (remember films used to be labelled 100 ASA/21 DIN, not to mention DIN plugs for cabling) and in France it is AFNOR. All these bodies are members of the International Organization for Standardization (ISO) and more of that relationship later.

Many practices are governed by standards which appear no different for the most part from a *de jure* standards but which have not been through the procedures required of a national body or ISO. They may have been through equally rigorous procedures and may be under the auspices of a national institution that leads the field or a professional society. MARC (Machine Readable Cataloguing) is such a standard. Most likely MARC is a national standard somewhere because many countries have their own national MARC standard, but the almost defunct UK MARC has never by itself been a British Standard, purely a document produced by the British Library for its own system needs and because the British Library produced the records for others to use in their catalogues it became a *de facto* standard. MARC 21 started out as LC MARC (Library of Congress MARC) and later became US MARC. Now it is MARC 21 but all along it has been the joint work of the Library of Congress and a committee known as MARBI (American Library

Association's Machine Readable Bibliographic Information Committee). Having said this, the record structure of MARC is a standard, and has been from the very early days, initially a NISO standard and then adopted as an international standard.

There is no legal distinction between a *de facto* standard and a proprietary standard. A good example of proprietary standards are the "rules" which govern the way that a range of the features of Microsoft Windows work. These features may require, for example, a certain design for a mouse; drop down menus work in particular way. Software and hardware designers need specifications for these features so that they can develop further tools for these systems. The specifications are available but they do not have the status of standards though they have the same force of standards. These features were not agreed by any international committee but they are nevertheless used in many systems around the globe. They would appear to have the status of standards but are purely the development of one company, Microsoft and no doubt were "standardised" by Microsoft developers rather than any formally established committee.

It is not always clear cut as to which category a standard belongs. In the area of librarianship and information management, many standards are guidelines advising particular practices; the subject which they seek to standardise is not one which has to be uniform for the situation to work. Take BS 1629 which indicates how bibliographical references should be made; it is a variant of other "standards" such as Harvard or MLA which are proprietary. Cataloguing codes such as AACR are related in their function; like Harvard rules AACR has never been made into a formal standard though it has the same effect. However, in the late 1970s, there was a move to make the ISBD (International Standard Bibliographical Descriptions) into a formal international standard. ISBD is a set of rules for the descriptive part of the cataloguing record (as opposed to the access points or headings). These are intended to be incorporated into national cataloguing codes. In certain countries, particularly in the eastern European countries at that time, in order that this incorporation could be mandated, their legislation required that the ISBDs be enacted as international standards. In the case of the ISBDs, this was never achieved. When an international standard is under development, it has to go through certain procedures to ensure that it meets approval. During these procedures in the case of the ISBDs, the committees wished to make certain changes to the document that were not acceptable to the Committee of the International Federation of Library Associations (IFLA) that had developed the standard in the first place. The IFLA group who had rather reluctantly proposed the standard withdrew it from the international standardisation process. IFLA of course had already gone through its own rigorous procedures to develop the standard.

So far we already have a flavour of the organizations involved in standards making in our field but not a complete picture. Standards are made by standards bodies at both the national and international level, they are often initiated by professional associations and by national institutions. How do standards come into being? Often they are conceived by professional associations or national institutions needing to promote standard methodologies within their constituency. They are then proposed to a national standards body before being offered to ISO for international standardisation. Sometimes a national standards body may feel the need to cooperate in the development of a certain standard with other countries and will propose it initially as an international standard. To get this approved it is necessary to propose it to the appropriate ISO committee as a New Work Item. This then has to be voted on, and if there is a majority work can begin. Having said that, there have to be at least three countries involved. Currently, NISO is being very productive in developing standards, and they often invite representatives from outside the US to participate in their working groups. Nevertheless this does not give them an official international status, but it eases the path to subsequent adoption as an international standard if a standard has been given international consideration. The DCMES (Dublin Core Metadata Element Set) is an example of a document with the status of a *de facto* standard. This was then accepted as a NISO standard before being accepted as an ISO Standard. It is now incidentally a British Standard. Instead of having its own numbering in the BS series it is given an identifier BS ISO 15836 to indicate that it is approved as a British Standard.

In the library and information field, we are interested only in our own national standards, ISO standards and NISO standards, which though American national standards sometimes get used internationally. Other countries' standards become usually ISO standards when they are of international interest or application. National standards bodies and ISO have committees which deal with standards making in the field which is referred to officially as Information and Documentation. This covers librarianship, archives, records management, museum documentation and certain aspects of the publishing industry. In British Standards, for example, the committee that oversees this is IDT/2 and this has various subcommittees as in figure 1.

IDT/2 Information and documentation (responsible for oversight of the area)

IDT/2/1 Bibliographic references and descriptions

IDT/2/2 Indexes, filing and thesauri

IDT/2/7 Computer applications in information and documentation

IDT/2/9 Document preservation

IDT/2/10 Technical manuals

IDT/2/11 Coding of country names and related entities

IDT/2/15 Library and publishing statistics  
IDT/2/17 Archives/records management  
IDT/2/18 Identifiers and metadata

Standards can be grouped by their applications:

- describing and identifying information resources;
- information exchange;
- managing collections;
- delivering services.

## **Describing and identifying information resources**

Perhaps the most fundamental area of standards work for libraries is when it comes to identifying and describing information – whether it is books in a catalogue or information resources on a Web site.

Many library and information standards are guidelines advising particular practices. Take British Standard BS 1629:1989 *Recommendations for references to published materials*, which indicates how bibliographical references should be made. There are proprietary standards such as Harvard rules to which it is related as well as to ISO 690:1987 *Bibliographic references – content, form and structure* and ISO 690-2:1997 *Bibliographic references Part 2 – Electronic Documents or parts thereof*. Cataloguing codes such as AACR2 are related in their function; like Harvard rules AACR2 has never been made into a formal ISO standard though it has the same effect. Guidance for authors, publishers and editors also comes from: BS5605:1990 *Recommendations for citing and referencing published material* and BS6371:1983 *Recommendations for citation of unpublished documents*.

Some standards are used to describe the subject content of information resources and are widely used beyond the library and information domain. For instance, ISO 639-1:2002 and ISO 639-2:1998 deal with two-letter and three-letter codes respectively for languages. ISO 3166-1:1998 *Codes for the representation of names of countries and their subdivisions – Country codes* are used as the two-letter codes in URLs to designate country of origin. Fortunately these codes were originally developed long before the Internet age or else no doubt a small country such as Tuvalu would never have been given TV, the selling of which earns it half its Gross National Product!

Many librarians are concerned with classifying and indexing information resources, or developing vocabularies and classifications for specialist collections.

Some of the general classification systems such as DDC (Dewey Decimal Classification), UDC and the Root Thesaurus are actual standards or, in the case of DDC, *de facto* standards. The British Standards Institution publishes the English-language versions of the complete and abridged UDC (Universal Decimal Classification), as well as the online version, so it has authority of a British Standard.

Standards that cover indexing include ISO 5963:1985 *Methods for examining documents, determining their subjects, and selecting index terms* and the related British Standard BS 6529:1984 *Recommendations for examining documents, determining their subjects and selecting indexing terms*. Arranging indexing and items in catalogues are covered by ISO 999:1996 *Guidelines for the content organization and presentation of indexes* and ISO 7154:1983 *Bibliographic filing principles*.

Thesaurus development is an area of considerable activity where the BSI is taking a leading role internationally. The long-established ISO 2788:1986 *Guide to Establishment and Development of Monolingual Thesauri* has been superseded in the UK by BS 8723-2:2005 *Structured Vocabularies for Information Retrieval: Guide: Thesauri* to incorporate some of the lessons that have been learnt from applying thesauri in machine-readable environments. The second standard is ISO 5964:1985 *Guidelines for the establishment and development of multilingual thesauri*. Also of note is ISO 1987-1:2000 *Terminology work – Vocabulary – Part 1: Theory and Application*.

The International Organization for Standardization has a number of Technical Committees and Subcommittees. One such, Technical Interoperability in Information and Documentation ISO TC 46/SC 4, also oversees ISO 15511:2003 *International Standards Identifier for Libraries and Related Organizations (ISIL)* and a number of data element directories which cover inter-library loan, acquisitions, information retrieval and circulation applications and data elements for the exchange of cataloguing and metadata.

Identifiers and metadata are covered by a subcommittee entitled identifiers and Description, ISO TC 46/SC 9. The bulk of the standards of SC 9 relate to international numbering. The best known of these is ISO 2108:2005 *International Standard Book Number (ISBN)* which has grown from 10 to 13 digits in the latest standard. ISO 3297:1998 *International Standard Serial Number (ISSN)* is currently under revision. Other important identifiers are: ISO 3901:2001 *International Standard Recording Code (ISRC)*, ISO 10444:1994 *International Standard Technical Report Number (ISRN)*, ISO 10957:1993 *International*

*Standard Music Number (ISMN)*, ISO15706:2002 *International Standard Audiovisual Number (ISAN)* and ISO 15707:2001 *International Standard Musical Work Code (ISMWC)*.

On the Internet, Dublin Core has been widely adopted as a standard for describing Web content. This metadata standard originally arose from an open, informal international collaboration and became a *de facto* standard. It was then adopted as a NISO standard and is now an international standard, ISO 15836:2003 *The Dublin Core metadata element set*.

## Information exchange

When exchanging information between systems, standards are essential. This is important for interoperability (where different systems are involved in a single transaction or series of transactions) and for migration of data to new systems. Many libraries now require their library management systems to be able to handle UNIMARC or MARC 21 records, for instance. This allows libraries to migrate their catalogues on to new systems, regardless of manufacturer and to purchase standard bibliographic records from vendors. MARC was originally designed for the mass production of library catalogue cards, in the days when data processing was done on mainframe computers and access was severely limited. All the MARC formats conform to ISO 2709:1996 *Format for information exchange*.

MARC is very specifically aimed at producing catalogues, but in the last decade or so the book trade sector has become interested in bibliographic records for a wider range of functions. A standard known as ONIX has been developed by The International Group for Electronic Commerce in the Book and Serials Sectors (EDItEUR). Again, as in the case of MARC, this is a *de facto* standard and used generally across the sectors though it has never been formalized as an international standard. The record structure used is XML. This is a record structure used widely across many sectors unlike ISO 2709 which is used only in library data processing. Earlier EDItEUR worked with EDI for the book industry which was based on ISO 9735: *Electronic data interchange for administration, commerce and transport (EDIFACT)* which was developed outside the book trade by commerce in general and was originally sponsored by the United Nations. In the UK, Book Industry Communication (BIC) is the lead body in these activities spanning the library world and the book trade.

One of the most often talked about of standards in the library field is the NISO standard Z39.50-2003 *Information retrieval: application service definition and protocol specification* (also adopted an international standard ISO 23950) which

allows the development of virtual union catalogues. Z39.50 develops standard protocols for searching which can be incorporated into a search engine (in an OPAC or library management system, for instance) to search other catalogues even though they may be mounted on different systems. NISO has also produced Z39.88-2004 *OpenURL Framework for Context-Sensitive Services*, which enables catalogues of books and journal articles or lists of bibliographical references to link to a preferred source for an electronic resource which is available at more than one location.

There are several international standards which cover the way in which characters are coded and transliterated. The Unicode standard is based on the character set of ISO/IEC 10646:2003 *Universal Multiple-Octet Character Set (UCS)* which allocates a computer code for the representation of the majority of the world's scripts. This is based on ASCII (an equally important ANSI standard) which allocates numeric values to each character in the Latin alphabet, so that @ gets a value of 64, A of 65, and a of 97. The library world was first in the field in requiring non-Latin scripts to be produced by computers and the ISO Technical Committee 46 which deals with libraries and information science is responsible for standards for transliteration of non-Latin scripts:

- ISO 9:1995 *Transliteration of Cyrillic characters into Latin characters*
- ISO 233:1984 *Transliteration of Arab characters into Latin characters*
- ISO 259:1984 *Transliteration of Hebrew characters into Latin characters*
- ISO 843:1997 *Transliteration of Greek characters into Latin characters*
- ISO 3602:1989 *Romanization of Japanese (Kana script)*
- ISO 7098:1991 *Romanization of Chinese*

Metadata is a key resource for information exchange and apart from Dublin Core, there are standards dealing with the exchange of metadata: ISO 8459-5:2002: *Bibliographic data element directory – Part 5: Data elements for the exchange of cataloguing and metadata*; and ISO/IEC 11179-1 to -6 *Metadata Registries (MDR)*.

## Managing collections

Collection management includes standards for conservation, management of processes and housing of collections.

Some aspects of this are covered by:

- ISO 11108:1996 *Archival paper. Requirements for permanence and durability*
- ISO 11799:2003 *Document storage requirements for archive and library materials*

Records management has been the focus of considerable activity in recent years and the current international standard based on the Australian national standards was published in 2001:

ISO 15489-1:2001 *Records management – Part 1: General*  
ISO/TR 15489-2:2001 *Records management – Part 2: Guidelines*

And more recently:

ISO 23081-1:2006 *Records management processes. Metadata for records. Part 1: Principles*

Many libraries have started using RFID tags to manage circulation and loans. Several standards are emerging for bibliographic data on RFID tags; and there is as yet no international consensus on the issue. Some manufacturers are also concerned about keeping a competitive edge with their proprietary systems and are applying for patent protection before a standard is agreed. NISO is currently investigating this area and the Danish Standards Institute has probably gone furthest in developing a standard for bibliographic data held on RFID tags. An ISO TC46 working group is developing a standard data model for encoding information on RFID tags used in libraries and it is likely that a new international standard will eventually emerge.

### **Delivering services**

Several countries have worked on standards for service delivery. For instance the UK Department for Culture, Media and Sport (DCMS) has published Public Library Standards (<http://www.dcms.gov.uk>). In order to monitor service delivery, performance measures have been developed and there are several ISO standards for performance measures specifically aimed at libraries:

ISO 11620:1998 *Library performance indicators*  
ISO 11620 Amd.1:2003 *Library performance indicators. Amendment 1: Additional performance indicators for libraries*  
PD ISO/TR 20983:2003 *Performance indicators for electronic library services*

Performance indicators for electronic library services are rapidly developing and are dealt with comprehensively in *E-Metrics for library and information professionals* (Facet Publishing, 2005).

Many sectors monitor library performance and produce library statistics – see ISO

2789:2003 *International library statistics*. UNESCO (<http://www.unesco.org>) produces a series of statistics, as do several bodies in the UK such as LISU (<http://www.lboro.ac.uk/departments/lis/lisu>), SCONUL (<http://www.sconul.ac.uk>) and CIPFA (<http://www.cipfa.org.uk>).

### Where are standards going?

Standards will continue to consolidate in emerging areas such as RFID technology as it becomes more widely adopted in libraries. There will be more development of standards concerned with information exchange as the Internet develops and we see new ways of operating on Web 2.0. Metadata in particular has developed in the last five years in response to new demands placed on electronic information systems and no doubt new metadata schemas will begin to emerge in new areas of application.

Standards are an important part of library and information work and in an increasingly connected electronic environment we believe that they will become even more so. Library and information professionals will need to understand which standards apply to their work. Even more importantly they will have a significant role to play in the development of new standards concerned with information management and retrieval.

*Cosa sono gli standard? Vi siete mai chiesti perchè così tanti standard cominciano con "Z"? Questo articolo intende fornire una risposta a questo e a molti altri interrogativi che avete sempre voluto conoscere. Gli standard possono essere de jure, de facto e proprietari, sebbene non sia sempre possibile determinare con precisione a quale di queste categorie uno standard appartiene. Nel campo della biblioteconomia e dell'information management, molti standard costituiscono delle linee guida che forniscono consigli su determinate attività: cercano di standardizzare oggetti che, all'atto pratico, non devono essere necessariamente uniformi. Spesso gli standard vengono elaborati da associazioni professionali o istituzioni nazionali che hanno la necessità di promuovere le metodologie standard nell'ambito dei loro associati. Successivamente vengono sottoposti a un organismo nazionale di standardizzazione prima di essere trasmessi all'ISO per la realizzazione dello standard internazionale. Nel campo delle biblioteche e dell'informazione, l'attenzione si concentra sugli standard nazionali di ciascun paese, sugli standard ISO e sugli standard NISO che, sebbene siano standard nazionali americani, vengono talvolta usati in campo internazionale.*

*Gli standard in questo campo possono essere raggruppati secondo le loro applicazioni: descrivere e identificare le risorse informative, lo scambio di informazioni, la gestione delle collezioni e la fornitura di servizi. L'articolo fornisce una panoramica degli standard principali utilizzati in queste quat-*

tro aree di attività insieme agli enti che li sviluppano o li adottano. Per concludere, l'autore getta uno sguardo allo sviluppo di standard in settori emergenti quali la tecnologia RFID e i metadati. Gli standard costituiscono una componente importante del lavoro nel campo delle biblioteche e dell'informazione e lo diventeranno ancora di più in un ambiente sempre più connesso elettronicamente. Gli operatori professionali delle biblioteche e del settore dell'informazione avranno bisogno di comprendere quali standard trovano applicazione nel loro lavoro e quindi giocheranno un ruolo significativo nello sviluppo di standard implicati nell'information management e nell'information retrieval.

*Qu'est ce que les standards? Vous êtes-vous jamais demandés pourquoi un nombre si élevé de standards commencent par «Z»? Le but de cet article est répondre à cette question et aux nombreuses questions que vous vous êtes posés au sujet des standards. Les standards peuvent être de jure, de facto et propriétaires, quoiqu'il ne soit pas toujours possible d'établir avec précision la catégorie à laquelle un standard appartient. Dans le domaine de la bibliothéconomie et de l'information management de nombreux standards constituent les directives qui fournissent des conseils sur des activités précises: ils tentent de standardiser des objets qui, dans la pratique, ne doivent pas forcément être uniformes. Les standards sont souvent élaborés par les associations professionnelles ou les institutions nationales qui doivent promouvoir des méthodologies standards pour leurs associés. Ils sont par la suite évalués par un organisme national de standardisation avant d'être transmis à l'ISO pour la réalisation du standard international. Dans le domaine des bibliothèques et de l'information, l'attention se focalise sur les standards internationaux de chaque pays, sur les standards ISO et sur les standards NISO qui, bien qu'il s'agisse de standards américains nationaux, sont parfois utilisés dans le domaine international.*

*Dans ce domaine les standards peuvent être regroupés selon leurs applications: décrire et identifier les ressources informatiques, l'échange d'informations, la gestion des collections et la fourniture de services. Cet article offre un panorama des principaux standards utilisés dans ces quatre aires d'activités ainsi que des organismes qui les développent ou qui les emploient. Pour conclure l'auteur jette un regard sur le développement des standards dans des secteurs émergents tels que la technologie RFID et les métadonnées. Les standards sont une composante importante dans le domaine des bibliothèques et de l'information et ils le deviendront encore plus dans un milieu de plus en plus connecté électroniquement. Les opérateurs professionnels des bibliothèques et du secteur de l'information auront besoin de repérer les standards qu'ils appliqueront dans leur travail et qui joueront donc un rôle important pour le développement des standards impliqués dans l'information management et dans l'information retrieval.*

# La qualità delle immagini nei progetti di digitalizzazione

**Franco Lotti**

Istituto di Fisica Applicata "Nello Carrara" di Firenze

*L'articolo esamina le principali cause che influenzano la qualità delle immagini numeriche generate in un progetto di digitalizzazione. Il processo di formazione e acquisizione dell'immagine dalla sua forma analogica a quella numerica è determinante per la qualità dei file destinati alla conservazione (master). Esso dipende in gran parte dalle limitazioni delle prestazioni della strumentazione utilizzata e dalle modalità di acquisizione.*

*Un secondo aspetto non meno importante riguarda la qualità delle immagini destinate alla fruizione. Recenti sviluppi delle metodologie di compressione delle immagini e di nuovi standard consentono oggi di definire nuovi paradigmi per una fruizione dei contenuti culturali digitali orientata a una molteplicità di servizi e a un'utenza sempre più differenziata.*

## Introduzione

**S**i prenderanno in considerazione due momenti del processo che va dalla formazione alla fruizione del prodotto digitale: quello dell'acquisizione, che conduce alla produzione del cosiddetto file master destinato alla conservazione, e quello dell'immagine elaborata per essere fruita dall'utente attraverso un accesso locale o remoto.

Il flusso dei dati in un progetto di digitalizzazione può essere schematizzato nei suoi elementi essenziali come nella Figura 1. Il blocco relativo all'acquisizione comprende sia il processo di formazione dell'immagine a partire da un "originale", sia la sua conversione in forma numerica e la relativa codifica in un formato di file. Esso riassume quindi tutti i processi (ottici, elettronici, numerici) che concorrono alla produzione di un insieme di dati costituenti l'immagine numerica (digitale). Per prima cosa vengono esaminate le varie cause che influenzano la qualità dell'immagine generata, ponendoci a valle di questo blocco (punto A); in un secondo momento si prendono in considerazione gli aspetti qualitativi a valle del blocco successivo in Figura 1 (punto B), che provvede a generare e formattare i dati in funzione degli obiettivi di presentazione, trasmissione, distribuzione e in generale di fruizione, e che è rivolto quindi all'utenza.

## La qualità dei master digitali

Il sistema di acquisizione comprende la sorgente di luce, l'ottica, il sensore e infine il software di acquisizione e di calibrazione. Ciascuna di queste componenti lascia una sua impronta sull'informazione che si vuole catturare e contribuisce a modificarne

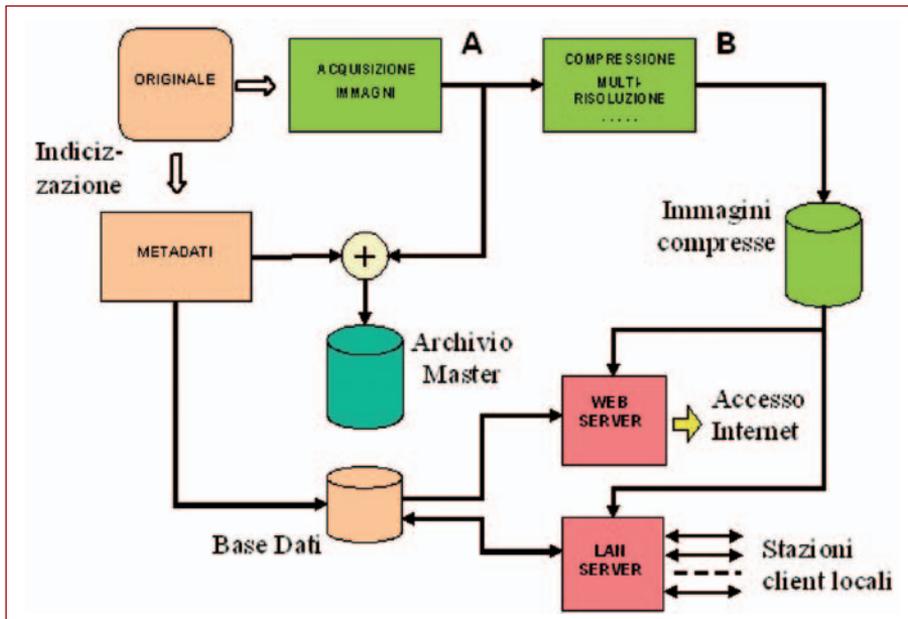


Figura 1. Esempio di schema del flusso dei dati in un progetto di digitalizzazione.

il processo di trasferimento dall'oggetto da riprodurre alla sua rappresentazione numerica finale.

Il livello di qualità ottenibile alla fine di tale processo è dunque il risultato dell'integrazione di tutte le limitazioni e distorsioni presenti nella catena di formazione e acquisizione.

Alcune limitazioni provengono dalle caratteristiche della strumentazione e sono migliorabili con l'impiego di tecnologia più sofisticata, altre invece sono intrinseche al processo stesso di digitalizzazione, come la rappresentazione di un oggetto "continuo" per mezzo di un numero finito di elementi (campionamento spaziale) o la successiva rappresentazione di questi elementi con un numero limitato di valori (quantizzazione). Queste ultime restrizioni impongono il rispetto di determinate regole per ridurre o minimizzare gli errori relativi.

Tutti questi fattori, comunque, dovrebbero essere oggetto di esame nella formulazione dettagliata del progetto di digitalizzazione, sebbene, spesso le specifiche relative alla qualità delle immagini da digitalizzare non vadano oltre la richiesta di due parametri, espressi in termini della cosiddetta *risoluzione* (termine a dire il vero inesatto, ma ormai entrato nell'uso; meglio sarebbe: *densità* o *frequenza di campionamento spaziale*), espressa in pixel per pollice (ppi) e di *profondità di colore* (bit per pixel, bpp). Con ciò si intenderebbe assolvere all'esigenza di garantire un certo livello di qualità delle immagini, commisurato agli obiettivi del progetto e alle disponibilità tecnologiche ed economiche.

Invero questi due parametri sono determinanti per generare correttamente il file numerico ottenuto dal processo di acquisizione, tuttavia non sono sufficienti a garantire un determinato livello di qualità dell'immagine. Vediamone più da vicino il significato.

### Risoluzione e campionamento

Il concetto di risoluzione è legato al massimo dettaglio percepibile o rappresentabile. Riferiamoci ad esempio a una configurazione di linee alternate bianche e nere come quelle riportate sui test di risoluzione (*test charts*); ogni coppia di linee (*line pair*) corrisponde a un ciclo bianco-nero, la cui dimensione determina la frequenza di tale alternanza e si esprime con il numero di cicli per millimetro o coppie di linee per mm (lp/mm). Linee più fitte corrispondono a maggiore dettaglio e a frequenze più elevate. La risoluzione è quindi da riferirsi alla massima frequenza riproducibile con un ammesso livello di distorsione.

La teoria dei segnali campionati ci indica la relazione teorica tra la risoluzione e la densità spaziale, o frequenza di campionamento, espressa in numero di campioni per unità di lunghezza sull'originale (ad es. ppi). Essa impone infatti che il campionamento avvenga a una frequenza superiore al doppio della massima frequenza (massimo dettaglio) presente sull'originale: la cosiddetta *frequenza di Nyquist*.

Con riferimento all'esempio della *test chart*, per campionare correttamente un *pattern* di 5 cicli/mm (o lp/mm) secondo il criterio di Nyquist occorre una frequenza di campionamento superiore a 10 campioni/mm, che equivale a 254 ppi. Attenzione però: mentre si può affermare che un campionamento più rado porterebbe all'impossibilità di ricostruire correttamente le oscillazioni di 5 o più cicli/mm, ciò non significa che con 254 ppi riusciremo sicuramente ad avere una perfetta ricostruzione del *pattern* originale. E ciò non solo perché intervengono altri fattori di approssimazione e altri fenomeni di degrado dell'informazione, ma anche per ragioni teoriche, poiché la frequenza di Nyquist si riferisce non tanto al massimo dettaglio che ci interessa riprodurre, quanto al massimo dettaglio (frequenza) che è presente sull'originale.

In altre parole, se nell'area campionata esistono rapide variazioni di intensità luminosa, cioè, nell'esempio precedente, frequenze spaziali superiori a quei 5 cicli/mm che ci interessano, l'energia di queste rientrerà nell'immagine campionata a 254 ppi sotto forma di disturbo "alieno" (da cui il termine *aliasing*), caratteristico del sottocampionamento. Tale disturbo è spesso evidente in forma di *moiré* quando si sottocampionano ad esempio oggetti a trama fitta come i tessuti, riproduzioni a rotocalco o con retinature molto fitte e quindi di frequenza superiore a quella di Nyquist.

Determinare la frequenza massima presente nelle immagini, per regolare in base a essa il campionamento da effettuare, non è generalmente cosa facile. Tanto meno se non è possibile analizzare ciascuna immagine prima di fissare i parametri dell'acquisizione. Nel caso della digitalizzazione di materiale fotografico ci si può riferire alle

caratteristiche della pellicola, considerando come frequenza massima quella potenzialmente rappresentabile con quel tipo di supporto<sup>1</sup>.

Più complesso è il caso di materiale documentario originale. Si può ricorrere in tali casi a una selezione di esempi significativi dell'insieme da digitalizzare, analizzando alcuni originali scelti a rappresentare classi omogenee per tipologia, con particolare riferimento alla presenza dei dettagli più sottili.

In teoria anche un solo tratto grafico a inchiostro nero su fondo chiaro, presentando alto contrasto e rapida variazione di livello, ha componenti di frequenza illimitata, perciò non sarebbe possibile evitare che una parte delle sue alte frequenze divenga energia di *aliasing*.

Un modo per risolvere in parte il problema è quello di eliminare, con un filtro passa-basso (*anti-aliasing*) prima del campionamento, le frequenze che superano quella di Nyquist e che comunque non saremmo in grado di rappresentare con il campionamento fissato. Spesso non è necessario ricorrere al filtro *anti-aliasing*, dato che altre parti del sistema (ottica, elettronica), avendo una risposta limitata nelle alte frequenze, funzionano esse stesse da filtro passa-basso. In entrambi i casi l'attenuazione delle alte frequenze si paga con una diminuzione della nitidezza (*sharpness*) dei contorni dell'immagine<sup>2</sup>.

Una volta determinato il campionamento in funzione della risoluzione spaziale desiderata dobbiamo verificare come questa sia effettivamente raggiungibile, tenendo conto di tutte le componenti del sistema, quali le caratteristiche dell'ottica, del sensore e dell'elettronica a esso associata, che dovranno essere tali da consentire il prelievo dei campioni con la prevista densità e precisione. Anche il sistema di illuminazione, specialmente quando non sia solidale con il sensore come negli scanner piani, può introdurre disomogeneità di illuminamento che vanno ottimizzate e di cui occorre tener conto in fase di post-elaborazione.

L'ottica introduce limitazioni soprattutto legate all'apertura dell'obiettivo (luminosità), alla profondità di campo (importante per la messa a fuoco di tutte le parti dell'originale), alle eventuali distorsioni geometriche e cromatiche. Il potere risolutivo delle ottiche associate a un buon sistema di acquisizione non rappresenta tuttavia un limite significativo alle prestazioni del sistema rispetto ad altre cause di errore, come quelle esibite dal sensore. Nella catena dei processi che intervengono sulla formazione dell'immagine il sensore è infatti per la maggior parte dei casi il principale responsabile delle più severe limitazioni alla qualità dell'immagine ottenibile.

Oltre al numero e alla dimensione degli elementi sensibili, che determinano, insieme alla geometria di ripresa e all'ottica, il campionamento spaziale, il sensore "soffre" di limiti intrinseci legati alla sua tecnologia, che ne fissa le caratteristiche di dinamica, efficienza, rumorosità.

Senza entrare nel dettaglio tecnico, tuttavia è utile distinguere tra i sensori che

<sup>1</sup> [http://www.iuav.it/CNBA/Giornate-d/2003-Le-Ot/Abstract--/Testo-degl/Buonora\\_relazione1.pdf\\_asc1.pdf](http://www.iuav.it/CNBA/Giornate-d/2003-Le-Ot/Abstract--/Testo-degl/Buonora_relazione1.pdf_asc1.pdf).

<sup>2</sup> Un esempio di filtraggio *anti-aliasing* si trova in: <http://www.photo.net/sigma/sd9>.

forniscono dati “non interpolati” (generalmente sensori a cortina lineari o tri-lineari) e quelli che forniscono dati “interpolati” (generalmente sensori a matrice rettangolare di elementi). Infatti, per la definizione dei criteri di campionamento, quello che conta è l’effettiva densità di informazione prelevata dall’originale (legata al numero di campioni per unità di superficie e quindi al numero di ppi).

Nelle acquisizioni a colori tale informazione riguarda le tre componenti primarie (ad es. rosso, verde e blu), perciò la definizione di un pixel a colori richiede la misura dei tre valori relativi. Questa operazione viene realizzata ponendo davanti al sensore opportuni filtri che separano le tre componenti cromatiche e operando in uno dei modi sotto descritti.

Utilizzando sensori lineari (come negli scanner piani o nei dorsi a scansione): con una cortina di sensori allineati e tre scansioni dell’originale (una per colore), oppure con scansione singola e una cortina tripla di sensori (tri-lineare). Utilizzando sensori a matrice: con tre esposizioni separate, oppure, come nella maggior parte delle attuali fotocamere digitali, con un solo scatto e interpolazione. Quest’ultimo metodo consiste nel disporre su ogni elemento sensibile della matrice un micro-filtro che lo rende sensibile esclusivamente a una sola delle tre componenti cromatiche. La matrice risulta così composta da un mosaico di elementi (*Bayer pattern*), alcuni dei quali sensibili al rosso (circa il 25%), altri al verde (circa il 50%) e i restanti al blu (circa il 25%). Dopo l’esposizione, un software provvede a ricostruire, a partire da quest’unico mosaico, tre matrici distinte, una per colore, riempiendo gli elementi mancanti di ciascuna di esse con valori interpolati sulla base dei pixel più vicini. Si passa così da una a tre matrici, interpolando in totale un numero di pixel pari al doppio di quelli acquisiti.

È evidente che il risultato, soprattutto cromatico, non è equivalente ad avere acquisito realmente tutti i pixel su ciascun colore, come avviene nei sensori “non interpolati”. Non tener conto di quanti sono i pixel realmente misurati e non interpolati porta a una sottostima della densità di campionamento e quindi a un’applicazione errata del criterio di Nyquist con possibile comparsa di artefatti.

Tra i sensori a matrice, interessante è infine il caso, a mia conoscenza l’unico a tutt’oggi, del sensore sviluppato da Foveon<sup>3</sup>, costituito da tre matrici sovrapposte di sensori, sensibili al rosso, verde e blu rispettivamente, che sfruttano la diversa trasparenza ai raggi luminosi di diversa lunghezza d’onda per operare con un solo scatto non interpolato.

### La quantizzazione

Il secondo livello di approssimazione che viene operato dal processo di conversione analogico/numerica dopo il campionamento è rappresentato dalla quantizzazione. Si intende con questo termine la discretizzazione (dopo quella spaziale avvenuta con il campionamento) della rappresentazione dei valori di tono corrispondenti alla

<sup>3</sup> <http://www.abc-fotografia.com/artic/foveon-x3.htm>.

luminosità di ciascun pixel. In altre parole del passaggio da valori che variano con continuità a un insieme finito di numeri rappresentabili con le tecnologie del calcolo numerico, che, come è noto, utilizzano l'aritmetica binaria. Si parlerà quindi di numero di bit per pixel (bpp), cioè di quante cifre binarie sono necessarie per rappresentare tutte le gradazioni di intensità luminosa del segnale acquisito (profondità di colore). Tipicamente: con 8 bit si possono rappresentare  $2^8 = 256$  diversi livelli di grigio o di una componente di colore; con 12 bit 4096 livelli, con 16 bit 65536 livelli. Per i pixel a colori si parlerà rispettivamente di  $3 \times 8 = 24$  bpp, 36 bpp e 48 bpp.

Anche nel processo di quantizzazione si perde dell'informazione, riducendo a un insieme di "gradini" la continuità delle sfumature. Questa perdita va sotto il nome di rumore di quantizzazione.

Così come il numero dei livelli, anche la precisione di rappresentazione dell'intensità luminosa, e quindi delle varie tonalità di colore, crescerà (in modo esponenziale) all'aumentare del numero di bit per pixel. Ad esempio con 8 bit si raggiunge una precisione nella rappresentazione della gamma dei grigi pari a circa lo 0,4% ( $1/256$ ) dell'escursione massima tra nero e bianco, mentre con 12 bit (4096 livelli) essa diventa circa lo 0,025%.

Il numero di bpp nativi esibiti da un sensore è generalmente commisurato alla effettiva capacità di quest'ultimo di distinguere due livelli vicini di luminosità o due tonalità vicine di colore. Viene determinato infatti dal costruttore in base alla effettiva dinamica e al rumore esibiti dal sensore. È quindi un parametro indirettamente indicativo, ma non affidabile, della sua qualità.

Generalmente nei sistemi di qualità medio-alta il numero di bit per pixel interni (utilizzati dal sensore), non è inferiore a  $12 \times 3$  e si arriva a 48 bpp ( $16 \times 3$ ), anche se in quest'ultimo caso non sempre tutti e 16 bit per colore sono realmente utilizzati dalla sua dinamica.

È il software associato al sensore (generalmente proprietario) che consente poi di ottenere in uscita un file finale rielaborato e corretto (secondo criteri di "ottimizzazione" o "gradevolezza" delle immagini, stabiliti dal costruttore) in formato TIFF non compresso a  $3 \times 8$  bpp, oppure compresso JPEG o altro. I sistemi di classe alta forniscono opzionalmente in uscita il formato nativo (RAW), non ancora elaborato, a 48 bpp. In questi casi, soprattutto nelle fotocamere digitali, si consiglia di estrarre il file in formato RAW e di effettuare fuori linea le elaborazioni necessarie per generare il formato definitivo per il master, utilizzando procedure non proprietarie o comunque note, che consentano sia di sfruttare tutte le variazioni tonali che offre il sensore quantizzato a 16 bit, sia di effettuare una migliore calibrazione del colore. Traccia di tali procedure dovrà restare nel file format stesso oppure nei metadati associati all'immagine.

Nella determinazione dei parametri di acquisizione, che dipendono dai dati di progetto, si cerca di tenersi sempre al di sopra degli stretti requisiti teorici, per compensare alcune delle inevitabili approssimazioni insite nel passaggio dalla teoria alla pratica.

È bene comunque sottolineare che l'intero sistema è affetto da numerose altre cause di errore, sulle quali i parametri suddetti non hanno influenza diretta. Tra queste vi sono le aperture finite degli obiettivi, i tempi finiti di esposizione, il rumore termico ed elettronico che è sempre presente nei sensori, l'efficienza di questi in funzione dell'intensità e della lunghezza d'onda della luce incidente, le fluttuazioni termiche degli illuminatori, le vibrazioni meccaniche, etc.

Da questo elenco sommario emerge un quadro abbastanza complesso delle cause di degrado del segnale "ideale" che influiscono sulla qualità dell'immagine numerica. Tener conto di tutti i fattori non è semplice e può non essere sempre fattibile. Va notato comunque che la qualità dell'immagine digitalizzata è solo in parte garantita da una corretta definizione dei parametri di "risoluzione" in termini di ppi e di "profondità di colore" (bpp).

### **MTF: una stima della nitidezza**

Un modo per caratterizzare il comportamento di un sensore è quello di descrivere come vengono riprodotte le variazioni di intensità luminosa (contrasto) via via che esse si fanno più fitte, descrivendo dettagli sempre più minuti. Tale metodo, utilizzato ampiamente nel mondo della fotografia per caratterizzare sia il materiale sensibile che l'ottica, conduce alla definizione di una funzione, detta "funzione di trasferimento della modulazione", MTF (*modulation transfer function*)<sup>4</sup>, che riassume in sé molte delle prestazioni del sistema. Come accennato sopra, essa descrive ciò che più ci interessa, e cioè come il sistema riproduce le variazioni di contrasto al crescere della loro frequenza e quindi del dettaglio.

L'andamento della MTF del sistema può essere valutato sottoponendo ad acquisizione apposite *test charts* standard, contenenti alternanze di chiari e scuri sempre più ravvicinati, (con profilo a barre o sinusoidale), come rappresentato nella Figura 2, che descrive un esempio virtuale di un sistema di buon livello: il valore dell'MTF raggiunge il 50% a 38 lp/mm e il 10% a 69 lp/mm. Una buona stima della MTF si può ottenere, con semplici calcoli o con l'ausilio di software proprietari<sup>5</sup>, anche valutando l'andamento di una linea che attraversa un brusco passaggio dal bianco al nero. Un esempio reale è riportato in Figura 3 per un dorso a scansione. La differenza di forma con la curva precedente è dovuta alla differente scala, che qui è lineare. Dalla curva riportata si vede che a 0,177 cicli/pixel (corrispondenti a 10 coppie di linee/mm nel caso specifico di campionamento a 1450 ppi) l'MTF scende al 50%, indicando che tale frequenza viene restituita dal sistema con la metà del contrasto che ha nell'originale.

L'MTF al 50% è riconducibile alla percezione della nitidezza di un'immagine da parte del sistema visivo umano, che dipende appunto dalla MTF e dal livello di rumore;

<sup>4</sup> <http://www.rlg.org/preserv/diginews/diginews21.html#technical>.

<sup>5</sup> <http://www.imatest.com/>.

mentre alla massima frequenza percepibile dall'occhio, l'MTF scende a valori al di sotto del 5% e variano abbastanza con la persona e le condizioni di illuminazione. La densità di campionamento spaziale, cioè il numero di ppi, anche se calcolata in modo da rispettare ampiamente il criterio di Nyquist, ci dà un'idea di quale sarà la frequenza massima rappresentabile senza *aliasing*, ma niente dice sulla fedeltà e nitidezza dell'immagine, legata al comportamento del sistema a tutte le altre frequenze. Viceversa, la conoscenza dell'MTF ci dà una descrizione più completa del comportamento globale del sistema, consentendo di stimarne quantitativamente le prestazioni e le caratteristiche essenziali e di effettuare *benchmark* per confrontare i vari apparati. Due sistemi diversi infatti possono esibire la stessa risoluzione massima, ma comportarsi assai diversamente alle frequenze più basse. L'MTF ci consente inoltre di stimare le aberrazioni cromatiche, se valutata separatamente per i tre canali.

### Le immagini destinate alla fruizione

Fino alla metà dell'ultimo decennio del secolo scorso la consultazione in linea di grandi banche di immagini generate dai progetti di digitalizzazione presentava problemi economici e gestionali a causa della disponibilità di supporti ancora limitati nella loro rapidità di accesso o nella capacità e affidabilità. Si pensi ad esempio ai jukebox di CD-ROM. Per questo motivo in molti casi si sceglieva di generare una versione compressa delle immagini destinate all'accesso, sia dai terminali delle reti locali, che da Internet. In alcuni casi anche i file master sono stati conservati in forma compressa, a causa dei problemi suddetti. Uno dei compiti del progetto di digitalizzazione era quindi quello di definire i parametri relativi alla compressione, stabilendo un giusto compromesso tra qualità della fruizione da un lato e dimensione dell'archivio di immagini e quindi di impegno della rete dall'altro. Tipicamente, si generavano fino a 3 formati della stessa immagine: la versione non compressa, cioè il master destinato alla conservazione, la versione compressa JPEG per la consultazione locale ad alta qualità e in alcuni casi anche un terzo formato assai più compresso, destinato alla consultazione libera via Internet.

La tecnologia dell'ultimo decennio ha rapidamente modificato lo scenario: nastri magnetici e DVD di consistente affidabilità e *disk arrays* a costi medi inferiori a 10 euro per gigabyte, hanno consentito l'impiego di nuove soluzioni per la memorizzazione dei master e dei dati in linea e indotto a una revisione dei precedenti criteri di progetto. Oggi per esempio nei file master si conservano solo immagini in formato non compresso o compresso senza perdita.

Di pari passo si è evoluto anche il software per la creazione di algoritmi di compressione di immagini, più efficienti e orientati a una fruizione flessibile, adatta alle esigenze di un'utenza differenziata.

Un terzo fattore che sta entrando in giuoco nel modellare lo sviluppo di nuove

modalità di fruizione è rappresentato dall'ingresso dell'*e-commerce* nella fornitura di servizi di tipo culturale.

L'integrazione armonica di queste opportunità, anche se stenta a decollare per una serie di motivazioni sia culturali che economiche, rappresenta un notevole potenziale per ridisegnare nuovi paradigmi di fruizione, in un quadro di maggiore efficienza e sostenibilità.

### Le tecniche multirisoluzione

Nello schema di Figura 1 il blocco relativo alla compressione fa riferimento alle tecniche multirisoluzione. Con questo termine vengono indicate genericamente varie metodologie di trattamento delle immagini, utilizzate sia nella compressione che nella visualizzazione, basate su una decomposizione dell'originale in una serie di "sotto-immagini" che contengono caratteristiche di informazione differenziate e ordinate gerarchicamente. Un esempio semplice può essere la cosiddetta "piramidizzazione", che consiste nel ricavare dall'immagine originale alcune versioni ridotte, formando una piramide di vari livelli a risoluzione progressivamente più bassa, che vengono utilizzate in alcuni algoritmi di visualizzazione rapida a risoluzione variabile (*zooming*)<sup>6</sup>.

Tali concetti vengono utilizzati ed estesi nel sistema di compressione JPEG2000, che si basa su una particolare decomposizione in sottobande dell'immagine originale, strutturata in multirisoluzione, cioè la trasformata *wavelet* discreta<sup>7</sup>.

### Compressione e fruizione

Ogni procedimento di compressione d'immagini genera, a partire da un file definito come originale, una sua versione compressa di minor "peso", cioè contenente un numero di byte inferiore a quello di partenza. Il "rapporto di compressione" (quoziente tra l'ingombro del file originale e quello compresso) è un parametro che misura l'efficacia di tale compressione. Questa operazione si dirà reversibile (*loss/less*) se dal file compresso si può risalire esattamente all'originale. Negli altri casi si parlerà di compressione con perdita di informazione (*lossy*).

Esistono molti modi per misurare tale perdita, collegati ai vari metodi di stima della qualità dell'immagine ricostruita. Escludendo i metodi soggettivi basati su gruppi di osservatori, i più diffusi misurano la differenza tra l'immagine originale e quella compressa (errore di ricostruzione) e ne valutano alcune funzioni statistiche come l'errore quadratico medio (*root mean square error*, RMSE), la deviazione standard o l'errore massimo.

<sup>6</sup> <http://www.xlimage.it>.

<sup>7</sup> Stephane Georges Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, «IEEE Transactions on Pattern Analysis and Machine Intelligence», 11 (1989), n. 7, p. 674.

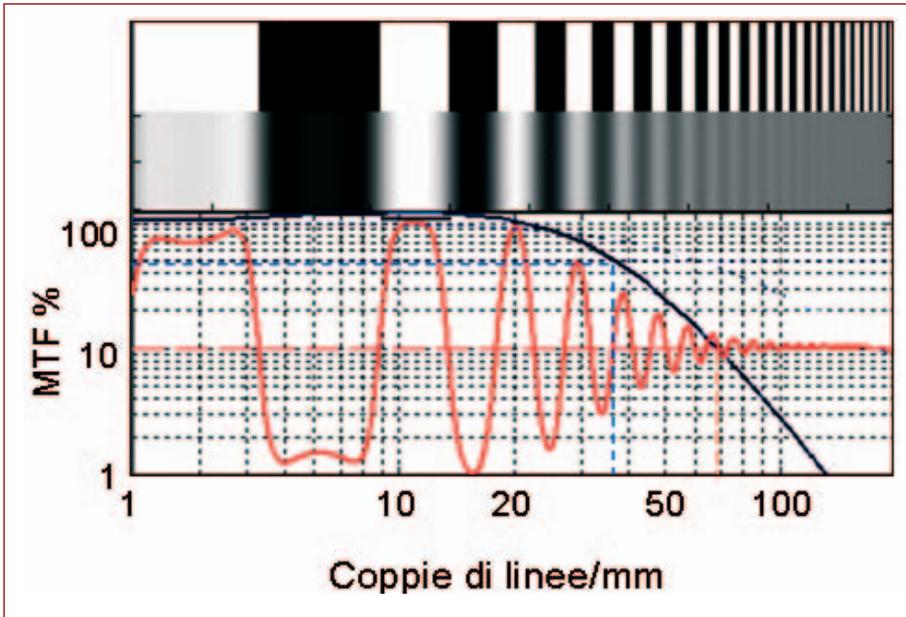


Figura 2. Valutazione della modulation transfer function (MTF). In alto: pattern originale a barre bianche e nere; al centro: immagine acquisita; sotto: in rosso il profilo del livello di grigio acquisito; in nero la funzione MTF.

Alcuni autori hanno proposto indici di qualità formati da combinazioni di parametri legati a considerazioni matematiche e psicovisive<sup>8</sup>. Negli esempi riportati nel seguito li useremo entrambi.

Il formato di compressione *lossy* più diffuso a tutt'oggi è certamente il JPEG. Esso consente di fissare un livello indicativo di qualità del file ricostruibile, da un minimo a un massimo secondo una scala arbitraria. Qualità più bassa produce un file compresso più piccolo, quindi un maggiore rapporto di compressione, il cui valore esatto non è peraltro prevedibile a priori, dipendendo dalle caratteristiche di ogni singola immagine. Il JPEG prevede varie modalità di compressione, invero poco utilizzate nella pratica, salvo quella di poter essere fruito progressivamente con risoluzione crescente, usata soprattutto nelle applicazioni Web<sup>9</sup>.

I limiti più gravi del JPEG risiedono nella comparsa di artefatti pesanti a valori di compressione medio-alta e nella scarsa flessibilità di fruizione: una volta compressa l'immagine con un certo livello di qualità, non è più possibile modificarla, se non con una nuova operazione di compressione.

<sup>8</sup> Zhou Wang – Alan C. Bovik, *A Universal Image Quality Index*, «IEEE Signal Processing Letters», 9 (2002), n. 3, p. 81-84.

<sup>9</sup> William B. Pennebaker – Joan L. Mitchell, *JPEG: Still Image Data Compression Standard*, New York: Van Nostrand Reinhold, 1992.

Da ricordare anche il JPEG-LS, che fornisce una versione *lossless*, comunque non integrabile nella versione *lossy* di JPEG.

Il JPEG2000, che ha ormai iniziato da tempo la sua, invero lenta, penetrazione nel mondo delle applicazioni, è da considerarsi molto di più che una nuova versione dell'ormai diffusissimo JPEG. Basato sui nuovi concetti di multirisoluzione e con un formato di file concepito per una fruizione remota e differenziata, esso consente al contempo di gestire la codifica dei file compressi e di regolamentarne la presentazione in una varietà di modalità che ne fanno uno strumento di grande flessibilità e potenzialità<sup>10</sup>.

Commentiamo solo alcune delle sue caratteristiche che riteniamo più importanti per le applicazioni di nostro interesse.

- 1 Apertura e interoperabilità, derivanti dalla sua aderenza allo standard ISO/IEC<sup>11</sup>.
- 2 Efficienza. Confrontato con il JPEG a parità di errori di ricostruzione, JPEG2000 raggiunge rapporti di compressione molto più elevati; oppure, viceversa, fissato il "peso" dell'immagine compressa, fornisce livelli di qualità molto più alti. Tale effetto si evidenzia soprattutto a elevati rapporti di compressione, come mostra il semplice confronto riportato nel seguito.
- 3 Gestione della profondità di colore: dal bianco e nero puro (bi-level: 1 bpp), fino ai 16 bit per ognuna delle componenti del pixel (ad es. 48 bpp per il normale colore RGB).
- 4 Gestione contemporanea della compressione *lossy* e *lossless* (v. successivo punto 7).
- 5 Possibilità di inglobare nel formato del file compresso varie tipologie di metadati, in speciali box in vari formati: XML, come ad es. quelli del *mag schema*, oppure IPR, o altri per la gestione dei diritti di sfruttamento e di proprietà intellettuale; tutti utilizzabili ai fini di un'efficiente consultazione della base dati.
- 6 Gestione di immagini multiple. Utile per effettuare la mosaicatura di grandi immagini, o la selezione di regioni di interesse su cui operare con differente qualità, come ad es. testo e grafica.
- 7 Progressività, disponibile nelle quattro modalità seguenti:
  - Progressività di risoluzione. Consente di trasmettere immagini a scala differente cioè a risoluzione crescente (*zooming*).
  - Progressività di qualità. L'immagine presentata a una determinata scala (dimensione) può aumentare progressivamente la sua qualità fino al livello desiderato, addirittura, fatto assai interessante, fino alla completa

<sup>10</sup> David S. Taubman – Michael W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, London: Kluwer Academic Publishers, 2002.

<sup>11</sup> JPEG2000 è pubblicato come standard ISO/IEC e come raccomandazione ITU-T (ex Consultative Committee of the International Telephone and Telegraph, CCITT).

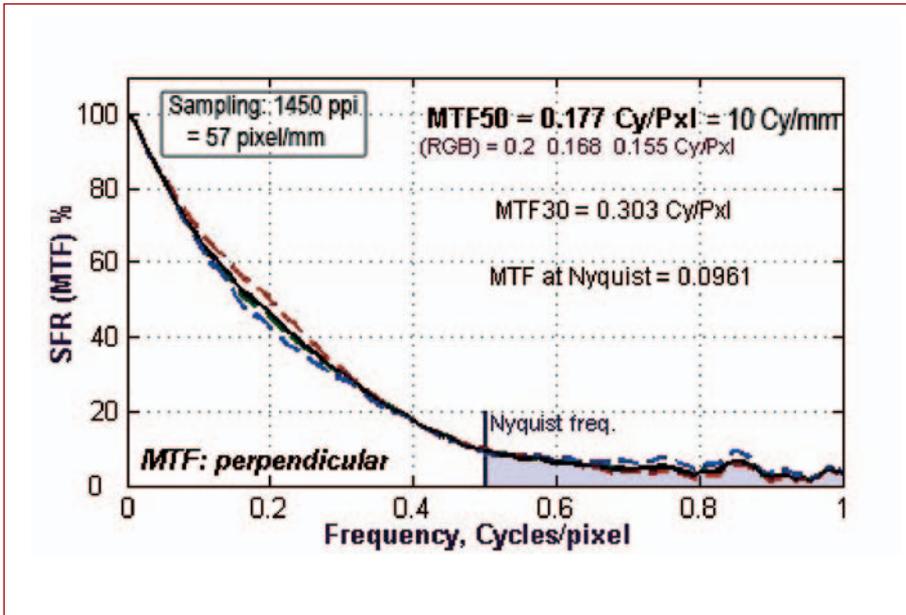


Figura 3. Esempio di Modulation transfer Function (MTF) di un dorso a scansione valutata acquisendo una test chart campionata a 1450 ppi.

- reversibilità (*lossless*).
- Sequenzialità. È possibile decodificare l'immagine compressa via via che viene ricevuta. Cosa non consentita da altri metodi di compressione, che possono presentare l'immagine solo dopo averla ricevuta interamente e poi decodificata. Ciò permette di ridurre i tempi di attesa e la memoria necessaria in ricezione sia dalla rete Internet sia anche da parte di una stampante.
- Infine la quarta modalità consente di gestire separatamente le varie componenti dell'immagine (ad es. per una foto in bianco e nero basta una componente, 3 per il colore RGB, 4 per CMYK; altre componenti possono riguardare informazioni in *overlay*, come note o grafica), fino a un numero notevolissimo di componenti per le immagini multi-spettrali.

### Comprimi una volta: decomprimi in molti modi

Con questa frase Taubman<sup>12</sup> sintetizza la caratteristica a mio avviso più promettente del JPEG2000: la possibilità di combinare le varie modalità di fruizione progressiva secondo un'architettura assai flessibile, adatta alle esigenze di un'utenza caratterizzata da profili diversi (utente generico, utente abilitato, pagante, etc).

<sup>12</sup>David S. Taubman, *JPEG2000* cit., p. 410.

Infatti, una volta generato il file compresso con certe caratteristiche di massima, si può progettare a piacere la sequenza di decompressione secondo un preciso ordine di risoluzione, qualità, scala cromatica, etc., così da poter stabilire diversi livelli di fruizione. Per esempio si può trasmettere prima un'ìcona in livelli di grigio, poi aggiungervi le componenti di colore con un livello medio di qualità, poi zoomare fino a 8x, poi aumentare la qualità di una sola regione di interesse prefissata, poi infine restituire tutta l'immagine ad alta qualità oppure nella versione *lossless*. In tutti i passi descritti si dovrà trasmettere solo l'informazione aggiuntiva necessaria per raggiungere il successivo livello desiderato, senza dover comprimere di nuovo l'immagine.

Ecco dunque che, avendo la possibilità di conoscere le caratteristiche dell'utente (persona con certi privilegi di accesso o dispositivo con un certo grado di risoluzione), si può conseguentemente differenziare il tipo di ricostruzione e arrestare al livello voluto il processo di trasmissione dell'informazione.

È evidente come queste caratteristiche di scalabilità sia in qualità che in risoluzione possano avere un ruolo decisivo nel progetto di digitalizzazione: una volta scelto il massimo livello di qualità della fruizione e le eventuali regioni di interesse da trattare in maniera differenziata (per es. le parti dipinte in un testo miniato), possiamo generare una sola versione di file compressi. Il progetto di fruizione potrà essere definito successivamente, nel rispetto di un ampio e flessibile paradigma di modalità distinte, e modificato in qualsiasi momento senza necessità di alterare la banca dati delle immagini già compresse.

In conseguenza di queste considerazioni, nello schema di flusso dei dati presentato in Figura 1, avendo supposto il ricorso a tecniche di multirisoluzione, abbiamo rappresentato con un unico blocco l'archivio delle immagini compresse destinate sia alla consultazione locale (LAN), sia al mondo Internet, abbandonando lo schema di due o più livelli di qualità distinti per la fruizione di alta qualità e via Web. Addirittura, scegliendo il *lossless* come massimo livello di qualità, si potrebbe anche eleggere questo archivio a livello di master, se alcuni "ragionevoli dubbi" sulla sicurezza non sconsigliassero al momento questa soluzione.

### Un breve confronto

Per dare corpo agli enunciati del paragrafo precedente, riportiamo alcuni risultati tipici ottenuti nel confronto tra compressione JPEG e JPEG2000. Una misura oggettiva, anche se molto sommaria, delle differenze tra i pixel dell'originale e quelli ricostruiti può essere data dall'errore quadratico medio, mediato a sua volta sulle tre bande cromatiche RGB (*average root mean square error, A-RMSE*).

La Figura 4 mostra l'andamento di tale errore, nel caso di una pergamena acquisita a 300 ppi, e compressa con i due algoritmi. Le prestazioni di JPEG2000 superano largamente quelle di JPEG, soprattutto a valori del rapporto di compressione

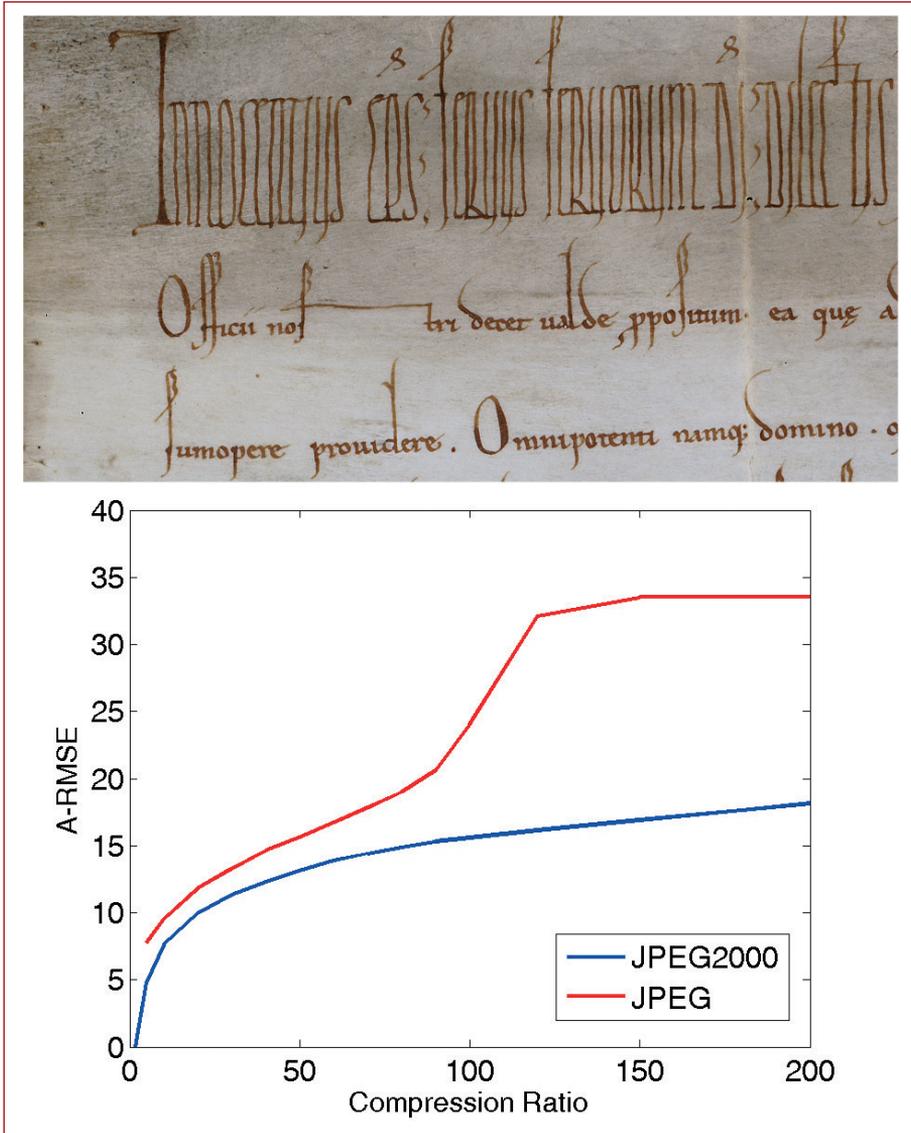


Figura 4 - Confronto tra JPEG e JPEG2000. Errore quadratico medio (A-RMSE) in funzione del rapporto di compressione, per la pergamena riportata in alto, acquisita a 300 ppi (media sulle tre componenti RGB).

superiori a 50. La curva del JPEG oltre CR = 110 diviene quasi costante, ma perde di significato per l'eccessiva distorsione, mentre JPEG2000 continua a dare risultati accettabili anche oltre CR = 200. Ciò è evidente nella Figura 5, in cui sono riportati anche i valori di un indice di qualità proposto da Bovik<sup>13</sup>, che varia tra 0 e 1,

<sup>13</sup> Zhou Wang – Alan C. Bovik, A Universal Image Quality Index, cit., p.81.

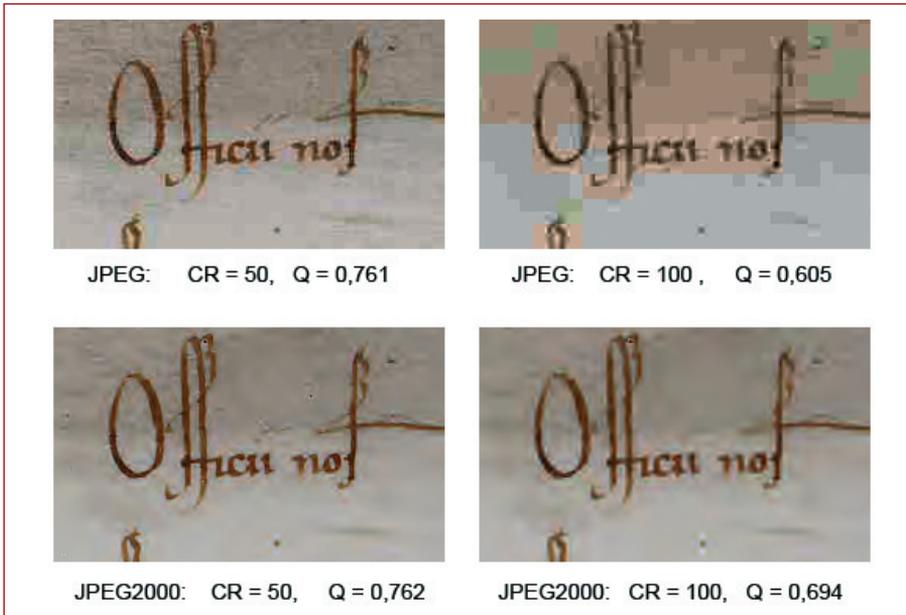


Figura 5 - Dettaglio dell'immagine di Fig. 4 ricostruita con i due algoritmi, per 2 valori medio-alti del rapporto di compressione ( $CR = 50$  e  $= 100$ ).  $Q$  è il valore dell'indice di qualità proposto in Zhou Wang -Alan C. Bovik, A universal image quality index, 2002 cit., p. 81.

attribuendo valore = 1 nel caso di identità tra immagine originale e decompressa. Il grafico di Figura 4 mostra anche che la curva di JPEG200 arriva fino a errore nullo. Infatti la ricostruzione *lossless* fornisce in questo caso un rapporto di compressione = 1,66, indicando un risparmio netto del 60% senza perdita di informazione.

## Conclusioni

Una breve descrizione dei processi tipici della formazione dell'immagine numerica e degli artefatti che essi possono introdurre, seguita da semplici considerazioni sulle altre cause fisiche di errore, evidenzia l'insufficienza dei soli due classici parametri di risoluzione e profondità di colore a garantire un preciso livello di qualità dell'immagine. Risulta perciò utile completare il criterio di scelta dei requisiti di ripresa nelle specifiche tecniche di un progetto di digitalizzazione. Potrebbe essere utile inserire tra i requisiti tecnici i valori minimi della funzione di trasferimento della modulazione (MTF) a frequenze spaziali ben definite, determinati in base al dettaglio e alla nitidezza che vogliamo riprodurre, nonché i valori massimi ammessi per l'*aliasing* e la distorsione cromatica. Ciò potrebbe evitare di fornire alle terze parti esecutrici del progetto l'alibi per trincerarsi dietro al numero di ppi richiesto e fornito, senza badare ad altri aspetti qualitativi del prodotto digitale generato.

Per quanto riguarda la qualità delle immagini destinate alla fruizione, abbiamo infine passato in rassegna alcune importanti caratteristiche del JPEG2000, concludendo che i metodi di compressione come questo, basati sui concetti della multirisoluzione e sulla trasformata *wavelet*, non solo offrono nell'immediato le prestazioni migliori, ma aprono un mondo di possibilità di gestione della presentazione e dell'accesso che meritano di essere meglio conosciute e sfruttate per dare nuovo valore aggiunto al progetto di fruizione.

*The main factors affecting digital image quality in a digitization project are outlined and discussed. The imaging and acquisition process has the main impact on the quality of the images to be stored as a master file for preservation. It mainly depends on the equipment features and acquisition methodology. A second not less important quality aspect concerns the images for presentation. Recent developments of image compression techniques and standards allow nowadays the design of new services and the access to digital cultural contents by ever more differentiated classes of users.*

*Les causes principales qui influencent la qualité des images numériques produites par un projet de numérisation sont ici examinées. Le procédé de formation et d'acquisition de l'image, de sa forme analogique à sa forme numérique, est fondamental pour la qualité des fichiers destinés à la conservation (master). Il dépend surtout des limites des prestations des instruments employés et des modes d'acquisition.*

*Un deuxième aspect tout aussi important concerne la qualité des images destinées à l'utilisation. Les développements récents de méthodes de compression d'images et de nouveaux standards permettent aujourd'hui de définir de nouveaux paradigmes afin d'orienter l'utilisation des contenus culturels numériques vers des services multiples et des utilisateurs de plus en plus diversifiés.*

# Biblioteche digitali e musei virtuali

**Franco Niccolucci**

Università di Firenze e Progetto EPOCH

*Le biblioteche digitali sono attualmente il fiore all'occhiello delle strategie europee in tema di digitalizzazione e applicazione delle tecnologie dell'informazione e della comunicazione ai beni culturali. Forse avendo ricevuto nuovo impeto da pressioni politiche, le iniziative sulle biblioteche digitali assorbono ora la maggior parte del precedente sostegno della Comunità Europea alla ricerca sulla comunicazione del patrimonio culturale, che fino al VI° Programma Quadro finanziava progetti tendenti a facilitare l'accesso dei cittadini al patrimonio contenuto nei musei e siti storici. L'articolo esamina lo stato dell'arte delle tecnologie utilizzabili per i musei virtuali e analizza se esse siano le stesse usate per gestire una biblioteca digitale. Risulta che c'è una vasta area comune, ma i musei virtuali richiedono un approccio particolare sostanzialmente differente da quello delle biblioteche digitali. La necessità di comunicare e di mediare i loro patrimoni, coinvolgendo interattivamente il pubblico, caratterizza i musei come diversi dalle biblioteche, e ciò determina differenze sostanziali nella tecnologia interessata. Inoltre, l'atteggiamento dei visitatori sta cambiando verso nuovi modelli, più simili ai consumi di massa, che richiedono maggiore attenzione e devono essere affrontati con strumenti nuovi e migliorati. D'altro canto, la tecnologia sembra essere il mezzo di comunicazione più adatto per dialogare con questi "nuovi" visitatori, che la usano quotidianamente. Tali considerazioni conducono a formulare scenari applicativi sia a breve che a medio-lungo termine. L'articolo conclude che sottostimare il rischio di un progressivo disinteresse per il patrimonio culturale europeo può decisamente impoverire questa componente fondamentale della nostra identità comune.*

## Introduzione

L'articolo di Anna Maria Tammaro sul numero 0 di questa rivista (Tammaro 2005) passa in rassegna alcune definizioni di biblioteca digitale, con l'intenzione non tanto di giungere alla "definizione" di biblioteca digitale, quanto piuttosto di evidenziarne le caratteristiche attraverso l'esame ragionato delle definizioni che ne sono state date. Si tratta, in effetti, di un argomento che travalica ormai i confini della comunità dei professionisti che si occupano di biblioteche, sia dal lato dei contenuti e dei servizi (i bibliotecari) sia da quello della tecnologia che ne consente la gestione e l'accesso (gli informatici).

Numerose iniziative di carattere politico hanno infatti posto in primo piano la tematica delle biblioteche digitali anche per i non addetti ai lavori. Fra queste, la più nota è quella per la Biblioteca Digitale Europea, originata dall'allarme lanciato dal presidente della Biblioteca Nazionale di Francia all'inizio del 2005 e fatto proprio

dal Presidente francese Jacques Chirac con un appello alla Commissione Europea, sostenuto anche dai governi di Italia, Germania, Spagna, Polonia e Ungheria.

L'appello ha portato a un'iniziativa congiunta – da tutti considerata la risposta europea a Google Print – sottoscritta da 19 biblioteche nazionali con il consenso di altre tre, mancando all'appello solo Irlanda, Regno Unito e Lettonia. La Biblioteca Digitale Europea è stata quindi oggetto di un documento specifico della Commissione Europea, nel quadro della strategia nota come "i2010", di cui è considerata uno dei punti di forza, con l'obiettivo di rendere il patrimonio culturale scritto e audiovisivo accessibile ai cittadini europei attraverso un'operazione che dovrebbe mettere *on line* due milioni di oggetti digitali (libri, film, fotografie, manoscritti) per la fine del 2008 e oltre sei milioni, almeno, per la fine del 2010. Dunque, le biblioteche, ancorché digitali, costituiscono uno degli assi portanti dell'azione dell'Unione Europea, com'è testimoniato anche dall'ingente quantità di risorse messe a disposizione all'interno di vari programmi di ricerca e sviluppo per ricerche nel campo delle biblioteche digitali, fra cui ad esempio e-Content+, Programma Quadro VI e VII.

Come segnalato nell'articolo citato di Tammaro, nel concetto di biblioteca digitale convergono le visioni e gli interessi della comunità di ricerca, che ha come obiettivo quello di predisporre strumenti di accesso efficienti e potenti e di arricchire e allargare i contenuti in termini di quantità e qualità, e della comunità di pratica, che si focalizza sui servizi e sulle problematiche di acquisizione, uso e sviluppo dei contenuti. Citando Borgman<sup>1</sup>, Tammaro fa inoltre presente che «[occorre] avere consapevolezza che le biblioteche digitali hanno un ruolo da svolgere nella società [...] il disegno, la strategia e la pratica della biblioteca digitale devono riflettere il contesto sociale di riferimento dell'utente insieme ai reali bisogni e ai comportamenti di ricerca dell'utente medesimo».

Dunque la biblioteca digitale è uno strumento aggiornato, "nuovo", come "nuove" sono le tecnologie che utilizza, teso a soddisfare i bisogni culturali dei cittadini; diventa promotore di cultura nel momento in cui non solo "riflette" tali bisogni, ma anche stimola nuovi interessi, in maniera attiva e non soltanto con la semplice disponibilità di accesso. Se ci si limita a «favorire l'accesso dei cittadini al patrimonio culturale europeo», come recitano i bandi europei relativi ai progetti nel campo delle applicazioni delle tecnologie informatiche ai beni culturali, non vi sono sostanziali differenze fra biblioteche digitali e musei virtuali. Questi ultimi comprendono, infatti, una varietà di contenuti: testi, ma soprattutto immagini, filmati e modelli tridimensionali. Dunque, ne conseguirebbe che, per favorire l'accesso alle collezioni museali digitali, sarebbe sufficiente una tecnologia simile, se non identica, a quella delle biblioteche digitali, che, nella concezione allargata di biblioteca introdotta dalla tecnologia digitale, prevedono anch'esse di ospitare diversi *media* e non più soltanto libri: immagini, filmati e forse anche modelli tridimensionali di oggetti.

<sup>1</sup> Per il riferimento completo si veda il citato articolo di Tammaro 2005, p. 14-33.

C'è però una differenza di fondo nell'uso previsto di questi contenuti. Facendo ancora riferimento all'articolo citato di Tamaro, la definizione ivi menzionata di biblioteca digitale fornita dalla Digital Libraries Foundation (DLF) individua come scopi della biblioteca digitale «selezionare, organizzare, dare l'accesso intellettuale, interpretare, distribuire, preservare l'integrità e assicurare la persistenza nel tempo delle collezioni digitali così che queste possano essere disponibili prontamente ed economicamente»<sup>2</sup>.

D'altro canto, sulla definizione di museo virtuale esiste un ampio dibattito che non è qui il caso di riportare. Converterà invece estendere quella di museo fornita dall'International Council of Museums (ICOM): «compie ricerche che riguardano le testimonianze materiali e immateriali dell'umanità e del suo ambiente; le acquisisce, le conserva, le comunica e, soprattutto, le espone a fini di studio, educazione e diletto»<sup>3</sup>. Il museo virtuale svolge le stesse funzioni su oggetti digitali che sono repliche virtuali di tali testimonianze; a volte, una parte delle collezioni virtuali nasce digitale, non ha cioè alcun corrispondente reale, mentre più spesso tali collezioni riproducono o interpretano oggetti materiali. Mettiamo a confronto le due definizioni:

1. Biblioteca digitale  
Funzioni: selezionare, organizzare, dare accesso, interpretare, distribuire, preservare, assicurare la persistenza.  
Scopo: disponibilità pronta ed economica dei contenuti.
2. Museo virtuale  
Funzioni: studiare, acquisire, conservare, comunicare, esporre.  
Scopo: educazione, studio e diletto.

Dunque, nel concetto di museo c'è qualcosa di più, la "comunicazione", con cui si si propone di educare, far studiare e "dilettare" l'utente. In modo schematico, si potrebbe scrivere l'equazione:

museo virtuale = biblioteca digitale + comunicazione

<sup>2</sup> Disponibile su <http://www.clir.org>. Questa e le altre referenze al web sono state verificate il 25 ottobre 2006.

<sup>3</sup> La definizione è tratta dal sito di ICOM Italia: <http://www.icom-italia.org>, e traduce quella ufficiale in inglese/francese «acquires, conserves, researches, communicates and exhibits, for purposes of study, education and enjoyment» fornita da ICOM su <http://icom.museum>. Vale la pena di ricordare che la definizione ufficiale legislativa italiana (Codice dei beni culturali e del paesaggio, art 101, comma 1) non include né la ricerca fra le attività, né il diletto fra gli scopi, condannando i curatori a mere funzioni burocratiche e di conservazione, e i visitatori alla noia; dimostrandosi, in questo, meno illuminata di quella adottata nel suo testamento dall'Elettrice Palatina, che nel 1743 definiva il lascito delle collezioni di famiglia, fra cui gli Uffizi, come finalizzato a «ornamento dello Stato, [...] utilità del pubblico e per attirare la curiosità dei Forestieri».

Rispetto all'efficientissimo, ricchissimo e onnicomprensivo deposito di conoscenza che sarà la Biblioteca Digitale Europea, dotata delle più avanzate funzioni di ricerca e accessibile a tutti e dovunque, un museo virtuale ha dunque in più l'obiettivo di "coinvolgere" l'utente in un'attività di comunicazione. Non che questa sia estranea alle biblioteche: basti pensare ad esempio all'esposizione del *Book of Kells* nella biblioteca del Trinity College a Dublino. In tali circostanze, però, la biblioteca svolge una funzione museale, perché espone e comunica – nella fattispecie anche attraverso un'esposizione permanente dedicata alla scrittura – sconfinando nell'altra categoria.

Le considerazioni precedenti possono apparire ovvie, e difatti lo sono finché si considerano biblioteche cartacee e musei "di mattoni e intonaco"<sup>4</sup>, per i quali enfatizzare la differenza può sembrare inutile pedanteria. La questione diventa invece più sottile quando i contenuti sono digitali, nella sostanza identici per le due tipologie, che si diversificano solo per le specifiche funzioni di ciascuna: mettere a disposizione, la biblioteca; comunicare, il museo. Confondere l'identità dei contenuti con l'identità delle funzioni, o comunque non dare importanza a questo aspetto aggiuntivo di comunicazione, rischia di produrre conseguenze sociali negative, come si vedrà in seguito.

Poiché una biblioteca digitale è anche la sua organizzazione e la tecnologia che ne permette il funzionamento, questa differenza di funzioni implica la necessità di considerare aspetti tecnologici e organizzativi aggiuntivi rispetto a quelli propri delle biblioteche digitali. Nei prossimi paragrafi si esamineranno succintamente le caratteristiche di queste tecnologie, che cosa esse danno in più ai musei virtuali rispetto alle biblioteche digitali e rispetto ai musei tradizionali, e quali precauzioni richiedono. Infine, si cercherà di tratteggiare uno scenario in cui il museo virtuale è appiattito sul concetto di biblioteca virtuale per aver preso in considerazione solo le affinità delle due tipologie e averne ignorato le peculiarità.

### Le tecnologie per i musei virtuali

Le tecnologie disponibili per la creazione di musei e collezioni virtuali e per "aumentare"<sup>5</sup> la comunicazione del patrimonio culturale sono sostanzialmente tecnologie di visualizzazione. Esse sono impiegate per offrire rappresentazioni vive globali, di particolari e dettagli o ricostruzioni dell'aspetto passato di oggetti – dal pezzo della collezione al monumento e al sito storico-archeologico – colmando lacune dovute all'inaccessibilità o alla lontananza oppure a una

<sup>4</sup> È questa (*bricks and mortar* nell'originale) la definizione un po' scherzosa data, nel dibattito sui musei virtuali, ai musei reali, in quanto contrapposti a quelli virtuali, fatti soltanto di "numeri".

<sup>5</sup> Questo termine è la meccanica traduzione dell'equivalente inglese; in questi casi si fa infatti riferimento alla AR (*Augmented Reality*), in cui si mescolano elementi visivi reali ed altri digitali, virtuali, tanto più realistici quanto indistinguibili dai precedenti.

parziale o totale distruzione degli stessi, spesso corredandole di spiegazioni. In molti casi, si tratta solo di “pannelli tecnologici”, perché richiedono un approccio passivo da parte del visitatore, la cui volontà si esplica soltanto nella scelta di dedicare la propria attenzione ad alcuni di essi piuttosto che ad altri. In altri casi, invece, il livello di interattività è più elevato, e la ricchezza dell’oggetto digitale e la sofisticazione della tecnologia utilizzata possono offrire esperienze più complete.

In tutti i casi è comunque necessario predisporre una replica digitale degli oggetti reali, adatta alle successive elaborazioni. L’esempio più semplice è la foto digitale o la scansione di oggetti bidimensionali da utilizzare nella presentazione del museo. Anche questi casi apparentemente più semplici presentano tuttavia sfide tecnologiche in ordine a due fattori. Il primo riguarda la qualità delle immagini: alta qualità significa anche grosse dimensioni degli archivi digitali corrispondenti e di conseguenza prestazioni informatiche più scadenti, con conseguente caduta di interesse del visitatore. La soluzione più semplice è cercare il miglior *trade-off* fra qualità e prestazioni nelle condizioni date; la più raffinata utilizza invece i LOD (*Level Of Detail*, livelli di dettaglio), cioè cambia la risoluzione delle immagini quando si fa uno zoom più ravvicinato e quindi utilizza una minore risoluzione per le viste d’insieme ed una maggiore per i dettagli. Questa tecnologia è tipica delle applicazioni alla pittura, e corrisponde al comportamento naturale di chi osserva un quadro da lontano per apprezzarne l’insieme e si avvicina per osservarne i dettagli.

L’altro aspetto importante riguarda la proprietà intellettuale delle opere rappresentate e trova applicazione soprattutto su Internet, dove tutti possono accedere alle immagini, piuttosto che su sistemi *stand-alone*. Gli accorgimenti generalmente adottati consistono nel fornire solo immagini a bassa risoluzione, inadatte a scopi commerciali, ed eventualmente a marcare le immagini di alta qualità con una filigrana digitale (*Digital Watermarking*) invisibile a occhio nudo ma evidenziabile con particolari algoritmi.

Nel caso di oggetti tridimensionali l’acquisizione della forma e dell’aspetto è, naturalmente, più complessa. Essa può avvenire mediante scansione tridimensionale, utilizzando uno scanner 3D, oppure attraverso la costruzione di un modello geometrico tridimensionale mediante un software di disegno CAD. In entrambi i casi la rappresentazione della superficie dell’oggetto è comunque approssimata, e si ripropone il problema del *trade-off* fra qualità della rappresentazione e le prestazioni di calcolo. Mentre questo problema ha minore importanza nel caso di presentazioni non interattive, come ad esempio la realizzazione di un filmato, in cui le posizioni degli oggetti digitali sono preordinate e i singoli fotogrammi possono essere realizzati in tempi anche molto lunghi, consentendo di ottenere i risultati di impressionante fotorealismo ai quali il cinema ci ha ormai abituato, l’efficienza computazionale è invece determinante nei sistemi interattivi, in cui è l’utente a deci-

dere le posizioni e i movimenti degli oggetti virtuali e sono richiesti tempi di risposta molto brevi, se non immediati, per mantenere l'impressione di "realtà". Sono dunque necessarie tecniche di eliminazione dei particolari inutili. Ad esempio, una scansione tridimensionale produce una "nube" di punti geometrici corrispondenti ai punti reali sulla superficie dell'oggetto, che verrà successivamente ricostruita come costituita da tanti triangoli, le cui dimensioni caratterizzano la qualità visiva del modello: più piccole, modello più dettagliato-più grandi, modello più grossolano. Tali punti e i relativi triangoli devono essere più fitti nelle zone più complesse, possono invece essere più radi nelle zone più semplici: a ciò provvedono algoritmi di decimazione che ne riducono il numero in modo intelligente.

Un sistema alternativo rispetto alla scansione 3D è costituito dalle tecniche basate sulla luce strutturata o sull'IBR (*Image Based Rendering*, rappresentazione basata sulle immagini), che sfruttano proprietà della luce o il confronto di più immagini prese da angolature leggermente diverse per ricostruire la forma dell'oggetto. In alcuni casi, il materiale di cui è fatto o rivestito l'oggetto può essere fonte di ulteriori problemi per le sue proprietà di riflettenza o assorbimento della luce. C'è dunque tutta una serie di tecnologie relative all'acquisizione automatica di oggetti digitali tridimensionali e alla loro ottimizzazione, che hanno come scopo anche quello di ridurre i costi<sup>6</sup>. Nonostante una loro riduzione progressiva, i costi delle apparecchiature si mantengono elevati, ed è comunque necessaria una buona competenza tecnica per svolgere tutte le fasi fino al modello 3D definitivo.

Le tecnologie sopra descritte sono importanti anche per le biblioteche digitali nonostante il fatto che i modelli 3D siano adesso e siano destinati a restare una componente molto ridotta del loro contenuto; esse fanno dunque parte del patrimonio tecnologico comune alle biblioteche digitali e ai musei virtuali. Questo comprende anche le metodologie e le tecnologie di documentazione. A questo proposito, si deve notare che un oggetto digitale culturale è tale solo perché è riferito in modo documentato, a un oggetto culturale reale.

Anche nell'era digitale, però, la documentazione risente in modo pesante delle sue origini cartacee. In passato, e spesso ancora tutt'oggi, la documentazione consisteva in un documento con una descrizione più o meno standardizzata (meno, nel caso del patrimonio culturale, i cui operatori non hanno la tradizione di standardizzazione dei bibliotecari), corredata da immagini, disegni o foto. Questa separazione fra archivi di testi, per cui esistono tecniche sofisticate di gestione, e altri ar-

<sup>6</sup> Va in questo senso, ad esempio il sistema realizzato dal progetto europeo EPOCH (European Research Network on Excellence in Processing Open Cultural Heritage, <http://www.epoch-net.eu>) che ha realizzato un servizio *on-line* di generazione di modelli 3D a partire da una serie foto digitali dell'oggetto. A parte i costi di gestione del servizio, per ora svolto all'interno del progetto e quindi gratuito, l'attrezzatura necessaria è solo una comune fotocamera digitale del prezzo di poche centinaia di euro e il collegamento a Internet. Si tratta quindi di un sistema che mette la creazione di modelli 3D alla portata di tutti.

chivi, per cui queste non esistono o esistono solo in parte, sopravvive anche nel caso della documentazione digitale, impedendo l'integrazione completa delle informazioni. Dunque le informazioni grafiche, bi o tridimensionali, sono gestite separatamente dalle descrizioni. Ciò apre tutta una serie di problemi: ad esempio, anche a una richiesta semplice come "Quante colonne ha sulla parte anteriore il Tempio della Concordia ad Agrigento?" è praticamente impossibile dare una risposta, perché la domanda mescola informazioni testuali (Tempio della Concordia ad Agrigento) con informazioni su componenti geometriche del modello. Per inciso, le colonne sono sei, come si può riscontrare da una fotografia. Questa operazione di riconoscimento è ancora impossibile, o almeno problematica, per un computer, però non più problematica che per una persona che non conosca il significato del termine "colonna". Si tratta in questo caso di utilizzare tecniche di riconoscimento automatico di immagini, per ora applicate solo sperimentalmente in alcuni progetti europei, come Agamemnon, che ha sperimentato il riconoscimento di monumenti o loro parti utilizzando immagini scattate sul posto dai visitatori e inviate per il riconoscimento a un server centrale; o in corso di sperimentazione in COINS per l'identificazione di monete detenute illegalmente e poste in vendita su Internet. Comunque, anche questi progetti non affrontano la semantica delle immagini né la loro annotazione.

Caratteristica comune a tutte le ricerche e i progetti in questo campo è lo sviluppo assai più pronunciato dell'altra estremità della catena della produzione di informazione culturale, la comunicazione al pubblico. In questo campo, vi sono stati numerosi progetti europei che hanno affrontato aspetti diversi della visualizzazione, in genere applicata a siti archeologici. In questo caso infatti è indispensabile fornire al visitatore una guida per la visita, che altrimenti si ridurrebbe nel migliore dei casi al fascino romantico delle rovine. Di particolare importanza sono perciò le ricostruzioni virtuali di monumenti e siti, di cui esiste un'ampia casistica e una buona bibliografia<sup>7</sup>, anche se quest'ultima non può quasi mai documentare adeguatamente le realizzazioni perché queste necessitano di un appropriato supporto informatico. Fra i progetti europei più recenti<sup>8</sup> si possono ricordare quelli realizzati nel V Programma quadro di ricerca e sviluppo della Comunità Europea come CHARISMATIC ( Cultural Heritage Attractions Featuring Real-time Interactive Scenes and Multi-functional Avatars as Theatrical Intelligent Agents), ARCHAEOGUIDE (Augmented Reality-based Cultural Heritage Onsite Guide) e 3D-MURALE (3D Measurement & Virtual Reconstruction of Ancient Lost Worlds of Europe), tutti

<sup>7</sup> Oltre al volume pionieristico edito da Barceló, Forte e Sanders nel 2000 (Barceló – Forte – Sanders 2000), per una visione d'insieme si possono ricordare gli atti, riportati in bibliografia, della Conferenza Internazionale VAST (International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage) che dal 2000 è dedicata a queste tematiche. VAST si è tenuta ogni anno a eccezione del 2002.

<sup>8</sup> Un elenco completo dei progetti europei in questo campo è disponibile sul già ricordato sito di EPOCH.

volti a sviluppare tecnologie ad alto grado di interattività e, nel caso di ARCHAEO-GUIDE, anche apparecchiature particolari da far indossare al visitatore per una immersività più spinta. Si tratta di progetti di carattere innovativo che in generale non hanno determinato un aumento significativo del numero di applicazioni simili, compito questo che si è assunto il progetto EPOCH, ma hanno piuttosto aperto la strada per un nuovo modo di intendere la comunicazione archeologica. Prova ne è il fatto che le ricostruzioni virtuali sono state recepite anche nella legislazione italiana, che nel D.L. 63/2005 individua all'art. 2-quater fra le forme di edizione e di divulgazione delle indagini archeologiche le «ricostruzioni virtuali volte alla comprensione funzionale dei complessi antichi». Non risulta tuttavia che questo strumento sia stato ancora applicato, come del resto l'intero complesso delle indicazioni della legge citata relative all'archeologia preventiva.

In campo espositivo, si può invece ricordare la mostra *Immaginare Roma antica* svoltasi nel 2005 al Museo dei Mercati Traianei, che ha presentato un'ampia rassegna delle realizzazioni in questo settore. Al di fuori della progettazione europea, si può citare il progetto tedesco TroiaVR (Jablonka – Kirchner – Serangeli 2003) che ha prodotto nel 2002 un modello ricostruttivo virtuale della famosa città dell'Asia Minore basandosi sul *corpus* delle indagini archeologiche e lo ha mostrato in una esposizione di grande successo di pubblico. L'importanza di TroiaVR (Troia Virtual Reality) sta anche nel fatto che è uno dei pochi che si è posto il problema dell'affidabilità delle ricostruzioni virtuali. Queste si basano sull'interpretazione degli esperti, che colmano le lacune delle testimonianze materiali effettivamente ritrovate basandosi su analogie con altri siti o monumenti simili, su affinità di carattere antropologico o, infine, su quella che gli anglosassoni definiscono *educated guess*, cioè una congettura non priva di fondamento. Di questo processo interpretativo non rimane però traccia nel risultato finale, che anzi per le caratteristiche del mezzo di comunicazione e per la qualità elevata della visualizzazione trasmette una sensazione di autorevolezza, quasi fosse un documentario. Per ovviare a questo inconveniente si è suggerito (Frischer *et al.* 2002) di adottare un metodo filologico basato su annotazioni, restando tuttavia ancora oggetto di sperimentazione le modalità con cui comunicare al pubblico il livello di certezza della ricostruzione o la possibilità di interpretazioni alternative. A questo proposito si è proposto di utilizzare tecniche di visualizzazione che ne diversifichino le parti per colore o intensità (Zuk – Carpendale – Glanzman 2005), tecniche di disegno non fotorealistico, ad esempio tipo acquerello (Roussou – Drettakis 2003), e altro ancora. Mancano però analisi sulla reazione dei visitatori e quindi indicazioni sull'efficacia di tali accorgimenti, che devono essere dosati in modo da evitare un eccesso di informazioni e la conseguente confusione.

Studiando queste tematiche, è emerso che l'intero processo ricostruttivo è carente per quanto attiene alla documentazione. Non vengono registrate, in altri termini, le scelte che il creatore del modello compie nella sua realizzazione né le loro moti-

vazioni. Di recente sono state definite delle linee guida (Bearman – Denard – Niccolucci 2006) chiamate *London Charter*, su cui si è aperto un dibattito fra gli studiosi<sup>9</sup>. Esse definiscono, sia pure ancora sommariamente, i principi da seguire e la documentazione da predisporre quando si intenda utilizzare un modello tridimensionale nell'ambito del patrimonio culturale; test su alcuni casi di studio sono attualmente in corso. Tali linee guida si ispirano ai principi enunciati nell'*Ename Charter*, un documento<sup>10</sup> che enuncia i principi cui attenersi nell'interpretazione e presentazione del patrimonio culturale.

Queste ultime considerazioni mostrano come soluzioni innovative di vecchi problemi come la comunicazione archeologica portino con sé nuove problematiche di cui inizialmente non si è consapevoli; è un segno di maturità disciplinare il fatto che si sia iniziata una riflessione metodologica sulle tecniche sino a ieri utilizzate con l'entusiasmo un po' ingenuo della novità.

A questo punto le strade delle biblioteche digitali e dei musei virtuali iniziano a divergere, e il fatto di dover considerare l'aspetto comunicativo inizia ad avere il suo peso anche sulle tecnologie, che devono prendere in considerazione aspetti complessi quali quelli sopra considerati. Quale però sia il pubblico a cui si comunica sembra essere il vero problema.

### Quale pubblico per i musei (virtuali e non)

In un recente convegno sul patrimonio culturale tenutosi a Ravello, "RavelloLab", Antinucci<sup>11</sup> ha fatto notare la situazione particolare dei musei italiani. Riferendosi ai musei e siti archeologici statali, le statistiche dei visitatori mostrano come il 10% dei musei raccolga metà delle visite, con il 20% delle istituzioni si arrivi al 90% delle visite, e un terzo dei musei abbia meno di 20 visitatori il giorno. Se fossero imprese economiche, quasi non ci si accorgerebbe della chiusura dell'80% dei musei. Il "mercato" è dunque dominato, in situazione di oligopolio, dai Top 20.

Al convegno, studiosi stranieri hanno confermato che questa tendenza è in atto nelle maggiori mete del turismo culturale. Sorge dunque il quesito sul modello di fruizione che questo tipo di visitatori adotta. L'ipotesi proposta è che si tratti di uno analogo ai beni di consumo, basato sul *branding* di oggetti di grande notorietà. Dunque la cultura si trasforma in una serie di nuovi miti, e genera nuovi riti un po' straccioni quali la visita alla Gioconda sulla scia del *Codice da Vinci*, la foto con il falso gladiatore al Colosseo, la gita in gondola a Venezia con il gondoliere che

<sup>9</sup> Le motivazioni, il testo completo del *London Charter*, la bibliografia correlata e le versioni aggiornate a seguito dei commenti degli esperti sono disponibili sul sito <http://www.londoncharter.org>.

<sup>10</sup> Il documento e le informazioni relative possono essere reperite sul sito <http://www.enamecharter.org>.

<sup>11</sup> Francesco Antinucci è un noto esperto di comunicazione in ambito museale. Ricercatore CNR, è autore di *Comunicare nel museo*, Bari: Laterza, 2005.

canta *O sole mio* (meglio se in versione inglese).

Per evitare di completare la trasformazione dei luoghi della cultura in parchi a tema, è dunque necessario esercitare un forte controllo sulla tecnologia, che nella sua pretesa neutralità tenderebbe a spingere in questa direzione con i suoi fantasmagorici effetti speciali. D'altro canto, la tecnologia, specie quella telematica, ha un punto di forza nell'essere il linguaggio di uso corrente specialmente fra le nuove generazioni. Il telefono cellulare con le sue funzionalità in continua espansione può dunque divenire il veicolo di un'informazione forse succinta nei contenuti ma somministrata nel modo giusto e al momento e nel luogo più appropriato. È per questo che si stanno studiando strumenti<sup>12</sup> utilizzabili ovunque e da parte di tutti. In effetti, la tecnologia è già (quasi) disponibile. Per la visualizzazione, si va dall'apparecchio sgradevolmente denominato "videofonino" agli occhiali con monitor incorporato collegati al cellulare, apparsi qualche anno fa in una pubblicità televisiva come oggetto fantascientifico ma oggi già distribuiti ai propri clienti da un fornitore francese di servizi di telefonia cellulare. Quello che manca sono i contenuti, o meglio lo stile adatto al mezzo. La tecnologia è più veloce della nostra capacità di usarla, forse, e gli stessi venditori propongono utilizzi non ottimali e a volte anche un po' ridicoli: anche qui, il *trash* va molto.

Nel museo virtuale, il lavoro del mediatore culturale è dunque fondamentale. Esso si esplica attraverso la narrazione di una storia in formato digitale (*digital storytelling*) che prende lo spunto dalle evidenze materiali, o meglio dalle loro repliche digitali, per raccontare la Storia. Questa attività richiede tecnologie che facilitino il lavoro di *storytelling* e competenze tecnico-comunicative in gran parte ancora da scoprire – certo abbandonando lo stile un po' libresco e pedante che ha caratterizzato certa informazione museale.

## Nuovi problemi, grandi sfide

Una frase pronunciata da un anonimo a ECDL2005 (9<sup>th</sup> European Conference on Research and Advanced Technology for Digital Libraries) e citata da Martin Doerr<sup>13</sup>

<sup>12</sup> Questi strumenti vengono chiamati in vario modo, da *mobile computing* a *ubiquitous computing* oppure *wearable computing* (elaborazione "indossabile") per sottolinearne la funzionalità ovunque senza bisogno di apparecchiature aggiuntive. Anche su questi temi il progetto EPOCH sta svolgendo una significativa funzione di integrazione e proposta, e sta sperimentando queste tecnologie in ambito museale nel corso dell'esposizione denominata *Interactive Salon* (<http://www.tii.se/v4m/is/index.htm>) che si tiene a Stoccolma fino alla fine di marzo 2007.

<sup>13</sup> Martin Doerr è un noto ricercatore del FORTH (Foundation for Research and Technology-Hellas), centro di ricerca greco situato a Creta. Uno dei maggiori esperti internazionali nel campo delle biblioteche digitali, è il padre di CIDOC-CRM (CIDOC Conceptual Reference Model), lo standard ISO (International Organization for Standardization) per la documentazione del patrimonio culturale. Parte delle idee esposte in questo paragrafo relativamente alle biblioteche digitali sono una riformulazione di sue intuizioni. Gran parte delle sue pubblicazioni sono accessibili sul sito <http://cidoc.ics.forth.gr>.

in una sua recente comunicazione al convegno VAST2006 (7<sup>th</sup> International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage) recita: «Non ci sono nuove sfide di ricerca nel campo delle biblioteche digitali. Ci sono solo quelle di trent'anni fa che non abbiamo ancora risolto».

Doerr continuava indicando i temi di ricerca che potrebbero cambiare il modo di concepire la biblioteca digitale, anzi di trasformarla in una "rete integrata di conoscenza" attraverso un nuovo approccio che passa per la creazione di modelli globali di conoscenza, l'integrazione, l'interoperabilità, l'interdisciplinarietà e la creazione di comunità che si facciano carico di tutto questo. A questa agenda di ricerca, per quanto riguarda i musei virtuali si aggiunge la consapevolezza che ci si rivolge a un pubblico variegato di visitatori da coinvolgere il più possibile nell'esperienza culturale.

Le grandi sfide della ricerca in questo campo possono essere quindi essere formulate come scenari d'uso: probabile, a breve termine; possibile, a medio; immaginato, nel lungo periodo, cioè entro dieci anni, poiché la velocità di sviluppo della tecnologia non permette di estendere oltre le previsioni.

Nel breve periodo (2-3 anni) ci si può proporre di rendere disponibili a costi ragionevoli, e con una diffusione assai più vasta, le applicazioni che qui sono state descritte come ancora sperimentali o di limitata diffusione. La presenza di strumenti multimediali di comunicazione deve diventare la regola nei musei, come pure, grazie alla banda larga, l'utilizzo su Internet di modelli più ricchi di quelli finora disponibili. Il *digital storytelling* diventerà un elemento sostanziale della comunicazione. Per questo, un elemento indispensabile è la formazione di personale con competenze appropriate. In questo campo, il sistema formativo italiano sconta anche qui gravi ritardi: le competenze interdisciplinari in campo culturale vengono relegate nell'effimero dei master universitari e la rigida gabbia dei raggruppamenti scientifico-disciplinari da un lato e delle classi di laurea dall'altro impedisce innovazione e sperimentazione; in termini di carriera, viene premiato chi è centrale alla disciplina, non chi opera al confine integrando conoscenze diverse; in termini di offerta didattica, si continua a pensare che basti mettere l'uno accanto all'altro nel curriculum un corso di informatica e uno di discipline umanistiche per avere interdisciplinarietà, lasciando allo studente il carico dell'integrazione dei due diversi ambiti. Per fortuna dell'Europa, non è così dappertutto: basta guardare il Progetto Methods Network<sup>14</sup> nelle università inglesi per rendersi conto di come potrebbero essere fatte le cose.

<sup>14</sup> Methods Network (<http://www.methodsnetwork.ac.uk/>) è un progetto finanziato dall'Arts and Humanities Research Committee inglese per istituire un forum nazionale per la disseminazione e lo scambio di esperienze sull'utilizzo delle Tecnologie dell'Informazione e della Comunicazione in campo umanistico. Si tratta di un progetto ombrello che coordina e finanzia sottoprogetti specifici gestiti da gruppi di università. Il confronto con i PRIN (Programmi di Ricerca di Interesse Nazionale) del nostro Ministero dell'Università è sconcertante.

Queste carenze formative si faranno notare in modo più sensibile nel medio periodo (4-6 anni) – a meno che i nostri studenti vadano a studiare all'estero, come molti stanno già facendo per i dottorati – in cui probabilmente si assisterà alla definitiva affermazione della comunicazione anche in ambito culturale con terminali mobili individuali multifunzione, risultato dell'evoluzione dei nostri telefoni cellulari. Come già accennato, la tecnologia in questo campo è già disponibile o in corso di realizzazione; il ritardo è dovuto semmai alla nostra limitata capacità di adattare la comunicazione al medium.

Più difficile definire gli scenari di lungo periodo. Su questo tema si stanno esercitando ben più autorevoli consessi, prima fra tutti la Comunità Europea, che li ha individuati come sfondo alla già ricordata strategia "i2010". Qui si avverte, però, la mancanza di una consapevolezza profonda della ricchezza e specificità del patrimonio culturale europeo. Non si tratta però di rivedere la strategia europea sulle *Digital Libraries*, ormai divenute una parola magica che apre tutte le porte incluse quelle dei finanziamenti, ma di non dimenticare che molti contenuti digitali sono stati, prima, oggetti reali, di cui l'Europa è ricchissima.

Senza una strategia globale, il patrimonio materiale verrà trattato in modo diseguale: accudito amorevolmente dove è meno presente, trasformato in mito di consumo dove abbonda. Tecnologie generiche, cioè buone per tutti gli usi, rischiano di accelerare questo processo che in inglese è denominato *commodification*, cioè trasformazione del bene culturale in risorsa meramente economica (*commodity*) e, in ultima analisi, in parco a tema.

Questo rischio è presente anche se ci si limiterà a valorizzare solo le tecnologie tipiche delle biblioteche digitali, perché non sono sufficientemente orientate al coinvolgimento attivo dell'utente e dunque accentuano la divisione fra gli utenti consapevoli del valore culturale, sempre più una minoranza, che le utilizzeranno per "educazione e studio" e quelli sempre più numerosi che le ignoreranno, scambiando il legittimo "diletto" culturale per intrattenimento di consumo.

*Digital Libraries (DL) are currently the flagship of European strategies on digitization and Information and Communication Technology applications to Cultural Heritage. Possibly receiving new impetus under political pressure, DL initiatives are now absorbing the largest part of the previous EU support to research on heritage communication, which until FP6 funded projects aimed at facilitating citizens' access to heritage content of museums and historic sites. The paper examines the state of the art of the technologies which may be used in virtual museums and analyzes whether they are the same used to manage Digital Libraries. It appears that there is a very wide common area, but virtual museums need a peculiar approach which is rather different from the DL one. The need of communicating and mediating their cultural assets, interactively involving the public, characterizes museums as different from libraries, and this determines substantial differences in the underlying technology. Moreover, the visitor's attitude is presently shifting to different models, more similar to mass*

*consumption. They require a greater attention and need to be faced with new, improved tools. On the other hand, technology seems to provide the most appropriate medium to dialogue with these “new” visitors, who use it in everyday life. These considerations lead to outline application scenarios, both short-term and medium to long-term ones. The paper concludes that underestimating the risks of a progressive unawareness, or lack of interest, of European cultural heritage may decidedly impoverish such substantial component of our common identity.*

*Les bibliothèques numériques sont actuellement la perle des stratégies européennes en ce qui concerne la numérisation et l'application des technologies de l'information et de la communication aux biens culturels. C'est peut-être grâce au nouvel élan dû à des pressions politiques, que les initiatives au sujet des bibliothèques numériques absorbent aujourd'hui la plus grande partie de l'aide de la Commission Européenne à la recherche sur la communication du patrimoine culturel qui jusqu'au VI<sup>ème</sup> Programme-Cadre finançait des projets visant à faciliter l'accès des citoyens au patrimoine contenu dans les musées et dans les sites historiques. Cet article étudie la situation de l'art des technologies utilisables dans les musées virtuels et analyse s'il s'agit des mêmes que l'on utilise pour gérer une bibliothèque numérique. Il en résulte qu'il existe un vaste domaine commun, mais les musées virtuels demandent une approche particulière, différente de celle des bibliothèques numériques. En effet, la nécessité des musées de communiquer et de transmettre leur patrimoine en établissant un rapport interactif avec le public les différencie des bibliothèques et entraîne une différence des technologies impliquées. De plus, l'attitude des visiteurs est en train de changer et se dirige vers de nouveaux modèles, plus proches de la consommation de masse et qui demandent plus d'attention et des instruments nouveaux et améliorés. La technologie semble d'ailleurs être le meilleur moyen de communication pour dialoguer avec ces “nouveaux” visiteurs qui l'utilisent quotidiennement. Ces considérations mènent à formuler des scénarios des applications aussi bien à bref qu'à moyen-long terme. Cet article conclut que sous-estimer le risque d'un désintérêt progressif à l'égard du patrimoine culturel européen peut appauvrir nettement cette composante fondamentale de notre identité commune.*

## RIFERIMENTI BIBLIOGRAFICI

- Barceló – Forte – Sanders 2000: Juan Antonio Barceló – Maurizio Forte – Donald Sanders (a cura di). *Virtual Reality in Archaeology*. Oxford: Archaeopress, 2000 (BAR International Series 843).
- Bearman – Denard – Niccolucci 2006: Richard Beacham – Hugh Denard – Franco Niccolucci. *An Introduction to the London Charter*. In: *The e-volution of Information Communication Technology in Cultural Heritage: Where Hi-Tech Touches the Past: Risks and Challenges for the 21<sup>st</sup> Century: Project papers from the joint event of VAST/CIPA/EG/EuroMed 2006*, a cura di Marinos Ioannides, David Arnold, Franco Niccolucci, Katerina Mania. Budapest: Archaeolingua, 2006, p. 263-270.
- Frisher *et al.* 2002: Bernard Frisher – Franco Niccolucci – Nick Ryan – Juan Antonio Barceló. *From CVR to CVRO: the Past, Present and Future of Cultural Virtual Reality*. In: *Virtual Archaeology: Proceedings of VAST 2000*, a cura di Franco Niccolucci. Oxford: Archaeopress, 2002, p. 7-18 (BAR International Series 1075).
- Jablonka – Kirchner – Serangeli 2003: Peter Jablonka – Steffen Kirchner – Jordi Serangeli. *TroiaVR: a Virtual Reality Model of Troy and the Troad*. In: *CAA2002: The Digital Heritage of Archaeology: Proceedings of the 30<sup>th</sup> CAA Conference, Heraklion, Crete, April 2002*, a cura di Martin Doerr, Apostolos Sarris. Athens: Archive of Monuments and Publications, Hellenic Ministry of Culture, 2003, p. 13-18.
- Roussou – Drettakis 2003: Maria Roussou – George Drettakis. *Photorealism and Non-Photorealism in Virtual Heritage Representation*. In: *VAST2003: The 4th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, a cura di David Arnold, Alan Chalmers, Franco Niccolucci. Aire-La-Ville: Eurographics Publications, 2003, p. 51-60.
- Tammaro 2005: Anna Maria Tammaro. *Che cos'è una biblioteca digitale*. «Digitalia», 1 (2005), n. 0, p. 14-33.
- Zuk – Carpendale – Glanzman 2005: Torre Zuk – Sheelagh Carpendale – William D. Glanzman. *Visualizing Temporal Uncertainty in 3D Virtual Reconstructions*. In: *VAST2005: The 6<sup>th</sup> International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, a cura di Mark Mudge, Nick Ryan, Roberto Scopigno. Aire-La-Ville: Eurographics Publications, 2005, p. 99-106.

## ATTI DELLE CONFERENZE VAST 2000-2006

- VAST2000: Franco Niccolucci (a cura di). *Virtual Archaeology: Proceedings of the VAST Euroconference*. Oxford: Archaeopress, 2002 (BAR International Series 1075).
- VAST2001: David Arnold – Alan Chalmers – Dieter Fellner (a cura di). *Virtual Reality, archaeology and cultural heritage: Proceedings of the 2001 conference on virtual reality, archaeology and cultural heritage*. New York: ACM Press, 2001.
- VAST2003: David Arnold – Alan Chalmers – Franco Niccolucci (a cura di). *VAST2003: The 4th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*. Aire-La-Ville: Eurographics Publications, 2003.
- VAST2004: Yorgos Chrisantou – Kevin Cain – Neil Silberman – Franco Niccolucci (a cura di). *VAST2004: The 5<sup>th</sup> International Symposium on Virtual Reality, Archaeology and Cultural Heritage*. Aire-La-Ville: Eurographics Publications, 2004.
- VAST2005: Mark Mudge – Nick Ryan – Roberto Scopigno (a cura di). *VAST2005: The 6<sup>th</sup> International Symposium on Virtual Reality, Archaeology and Cultural Heritage*. Aire-La-Ville: Eurographics Publications, 2005.
- VAST2006: Marinos Ioannides – David Arnold – Franco Niccolucci – Katerina Mania (a cura di). *VAST2006: Joint event of VAST/CIPA/EG WS G&CH/EuroMed*. Aire-La-Ville: Eurographics Publications, 2006.

# Prospettive di rinnovamento della Legge sul diritto d'autore

**Giuseppe Corasaniti**

Magistrato, Presidente del Comitato consultivo per il diritto d'autore

*Le norme sul diritto d'autore presentano questioni irrisolte nei rapporti economici e sociali, e appaiono soggette a una confusione duplice: da un lato si pretende una comunicazione senza garanzie per i creatori e gli organizzatori della produzione intellettuale, dall'altro si immaginano misure repressive spinte sino alla individuazione di ogni singolo uso non autorizzato e insensibili alle esigenze di natura sociale. La rivoluzione digitale impone tuttavia una completa revisione dell'attuale apparato normativo, tuttora basato sui precetti fondamentali posti dalla legge 633/1941, modificata finora con interventi di aggiornamento disorganici e scarsamente coordinati. Nell'adeguare la normativa vigente alla realtà digitale, occorrerà immaginare un'azione riformatrice più vasta che riconosca nuovi soggetti e oggetti dei diritti d'autore. I contenuti digitali inducono a moduli di consumo nuovo, moduli che possono paradossalmente assicurare una ancora più certa difesa dei diritti di autori ed editori: le tecnologie non assaltano la proprietà intellettuale, ma possono assicurarne la più ampia ed effettiva garanzia. Affinché ciò sia possibile, occorre avviare un'autoregolamentazione e un metodo di concertazione costante all'interno delle categorie interessate. Il ruolo della SIAE e del Ministero per i beni le attività culturali in questo quadro appare centrale. Bisogna superare la visione negativa che vede solo i rischi della realtà digitale, perché si pone in un'ottica limitata e anacronistica, così come è non meno limitata la visione che pretende di fare a meno di autori ed editori immaginando una rete di comunicazioni priva di regole. C'è necessità di nuove politiche pubbliche "aperte" e autenticamente europee che promuovano meccanismi di apertura sociale e di condivisione che, fatti slavi i diritti delle nuove figure d'autore, permettano di sfruttare fino in fondo le opportunità di diffusione globale dei contenuti offerte dalla rete.*

**S**ul diritto d'autore si prospetta una vera e propria questione irrisolta sia nei rapporti economici, sia in quelli sociali. Vi è una confusione duplice: da un lato, infatti, si pretenderebbe una comunicazione senza garanzie per i creatori e gli organizzatori della produzione intellettuale, dall'altro si immaginano misure repressive, spinte sino alla individuazione di ogni singolo uso non autorizzato, in particolare nella "copia privata" di documenti digitali. Si arriva in tal modo a estremizzare la legittima tutela degli interessi coinvolti, riconosciuta a livello internazionale, verso una forma costante di pressione indebita, insensibile a ogni esigenza sociale.

Ma, d'altra parte, se è lecito parlare di "proprietà" intellettuale, non si può non sottolineare come ogni proprietà sia destinata a una "funzione sociale" (art. 42 della Costituzione). Se tale funzione può concepirsi in relazione alla socialità della persona, tuttavia non può contenere, quale formula riassuntiva, ogni legittima aspirazione alla condivisione sociale dei contenuti, ma deve considerare anche il riconoscimento, altrettanto legittimo, dell'equa retribuzione per ogni forma di sfruttamento commerciale non espressamente autorizzato da autori ed editori.

La rivoluzione digitale impone una completa revisione dell'attuale apparato normativo, ancora basato sui precetti fondamentali posti dalla legge n. 633/1941. Molti interventi di aggiornamento sono stati fatti nel corso degli anni in modo disorganico e scoordinato, sulla spinta soprattutto dei cambiamenti richiesti dalle Direttive europee in materia di diritto d'autore. La legge, nel suo complesso, non ha avuto una sistematica revisione e attualizzazione delle sue disposizioni, ferme, quanto a descrizione di processi e di forme di sfruttamento delle opere intellettuali, persino nella stessa definizione dei soggetti e dell'oggetto dei diritti, a una immagine arcaica, persino ottocentesca della società sulla quale deve operare. È auspicabile, quindi, non solo un adeguamento per quanto riguarda la realtà digitale, ma anche un'azione riformatrice più vasta, che riconosca nuovi soggetti e nuovi oggetti dei diritti d'autore individuati nella realtà del nuovo millennio, certamente senza disperdere le garanzie minime, ma introducendo forme più efficaci di mediazione in ogni caso di conflitto, forme retributive più adeguate alla realtà tecnologica, sanzioni efficaci e non virtuali, rivolte soprattutto verso le reali violazioni dei diritti, che implicano un abusivo sfruttamento commerciale delle opere altrui, se non una vera e propria contraffazione.

Nessuna altra legge come la nostra legge sul diritto d'autore ha subito così numerosi e rilevanti interventi, ma questi l'hanno mutata in un ordito in alcuni punti inestricabile, che rende incerta la sua applicazione. Prosperano, in questa situazione, le iniziative di vera offesa ai diritti degli autori e degli editori, e, nell'incertezza normativa del testo, ormai incongruo e inadeguato nelle definizioni normative, vengono a essere colpite le categorie più deboli o più difficilmente organizzate, impotenti di fronte al fatto compiuto.

Come ogni norma civile, e soprattutto penale, anche la legge sul diritto d'autore deve confrontarsi, direi consapevolmente, con la società, deve poter essere applicata senza difficoltà, deve, soprattutto, chiarire nel precetto e nella coerente sanzione quali sono i beni giuridici fondamentali che intende presidiare, senza eccessi e paradossi.

In particolare occorrerebbe definire con maggiore chiarezza la nozione stessa di "pirateria", e verificare in concreto quelle che sono le forme di riproduzione "criminale", cioè di vera e propria contraffazione organizzata e sistematica dei contenuti, comunque diffusi nelle reti di comunicazione. D'altra parte devono essere evidenziati con chiarezza gli aspetti che invece riguardano forme di riproduzione a

uso personale, inevitabili in una società basata sulla interattività e sulla digitalizzazione dei contenuti stessi, sull'interscambio immediato e globale nelle reti di comunicazione, quando questo avvenga in modo non concorrenziale con i titolari dei diritti, che debbono essere sempre conosciuti e riconosciuti.

I nuovi contenuti digitali possono essere riconvertiti e adattati a moduli di consumo nuovo, ma questi stessi moduli di consumo possono paradossalmente assicurare una ancora più certa difesa dei diritti di autori ed editori.

Le tecnologie non assaltano la proprietà intellettuale, ma possono assicurarne la più ampia ed effettiva garanzia.

Così come sarà indispensabile riconoscere dignità ai diritti degli utenti della produzione intellettuale, consumatori di cultura, che devono poter interloquire rispetto a misure spesso approssimative e genericamente repressive (si veda il recente dibattito a proposito della cosiddetta Legge Urbani che, senza neppure un tentativo di analisi tecnica e funzionale improvvisamente e solo nel nostro Paese rendeva illegittimo il *file sharing*, basandosi sul presupposto che ogni forma di condivisione di contenuti deve considerarsi per se stessa lesiva dei diritti). Quest'ultima vicenda normativa ha dimostrato quanto poco convincente sia stata la scelta, tanto che le critiche forti hanno spinto lo stesso Parlamento, nel momento dell'approvazione della legge, a impegnarsi per la sua successiva modifica. È rimasto nella legge Urbani un eccesso di sanzioni amministrative, considerate dai più una panacea, e spesso destinate a essere solo virtuali perché sproporzionate in modo irragionevole: sono persino più gravi di quelle previste dal Codice della Strada.

Solo l'etica può riempire quel diaframma che c'è tra la norma e la coscienza sociale, tra la consapevolezza diffusa di una esigenza di tutela e il bisogno di conoscere che può creare, talvolta, nella comunità una domanda deviata e deviante, che non riconosce o ignora i legittimi titolari dei diritti sulle opere, autori e coautori.

È preferibile, a mio parere, utilizzare il termine autori piuttosto che quello di "titolari dei diritti connessi" perché ogni opera è sempre più oggetto di un lavoro e di un coordinamento operativo (ancora molto la legge deve esplorare in questa direzione), e tocca in fondo la sfera culturale che ha molte sfaccettature e molti gradi di rilevanza.

Ritengo che nella revisione della legge debba essere tenuto presente il primo – e fondamentale – problema di avviare una autoregolamentazione e una dialettica regolamentare costante all'interno delle categorie e nei rapporti interni tra le categorie. In altre parole ritengo necessario un dialogo proficuo e sistematico che può aiutare a segnalare e prevenire abusi e rischi di diffusione abusiva e incontrollata delle opere dell'ingegno.

Occorre pertanto rafforzare l'attenzione verso gli aspetti economici, giuridici, sociali, seguire l'evoluzione tecnologica in atto e le dinamiche a questa connesse, in particolare allo sviluppo di tecnologie "aperte" che danno la possibilità agli autori di rafforzare la tutela della identità creativa e culturale nelle scelte dei metodi comu-

nicativi, nelle strategie di interattività, nelle capacità di dialogo con i potenziali consumatori e fruitori delle opere intellettuali.

Il ruolo della SIAE e del Ministero per i beni e le attività culturali in questo quadro appare centrale e fondamentale.

È importante suggerire forme più attuali di tutela dei diritti, in grado di assicurare la composizione e la soddisfazione di esigenze apparentemente contrastanti (quelle degli autori/editori e dei fruitori, ma in realtà perfettamente compatibili, tenendo in equilibrio posizioni diverse. Ma le parti interessate dovranno tentare le une di riconoscere le ragioni delle altre, dovranno proporre forme di mediazione soddisfacenti e adeguate a implementare anche soluzioni tecniche e giuridiche alle nuove questioni che ogni giorno l'innovazione digitale presenta.

Verranno così ad affermarsi nuovi modelli e forme di diffusione dei contenuti e insieme nuovi rischi, ma anche nuove opportunità per raggiungere un ambiente di fruitori sempre più vasto e globale.

Una visione negativa che veda solo i rischi della realtà digitale è una visione in sé limitata, oltre che anacronistica, così come forse non è meno limitata una visione che pretenda di fare a meno di autori ed editori immaginando una rete di comunicazioni priva di regole e di garanzie. In primo luogo vanno perciò al più presto potenziati, a ogni livello, i meccanismi di autotutela, le forme di intervento "associativo" delle categorie interessate (consumatori, produttori di contenuti, produttori di hardware di "lettura", gestori di reti o servizi).

Occorre prendere atto della realtà tecnologica senza alcun preconcetto e pregiudizio verso le tecnologie interattive (internet e telefonia della terza generazione).

Sono proprio le tecnologie in rete che favoriscono un rapporto diretto autore-pubblico. Sono proprio le tecnologie digitali a poter garantire un nuovo e moderno sistema di distribuzione dei contenuti, e a poter assicurare ad autori e imprese produttrici nuove forme di diffusione editoriale, che però non possono ricalcare i modelli gerarchici e di controllo dell'editoria tradizionale.

Una aversione e una demonizzazione generica delle tecnologie non è neppure compatibile con le politiche comunitarie chiaramente espresse nelle più recenti Direttive europee.

Il punto è che i rapporti economici nella produzione culturale, e quindi nella gestione dei diritti degli autori, basate su posizioni di forza precostituite, non possono essere l'unico elemento condizionante e che occorre rimarcare il ruolo e l'identità del mercato italiano della produzione dei contenuti culturali.

La questione della tutela dei consumatori e del rapporto diretto consumatori-autori non è secondaria: è il vero punto di forza per una autentica alleanza per la legalità, quella che non passa dalle roboanti proclamazioni di guerra (l'antipirateria vista in modo generico, il richiamo improprio alla illegalità e al mondo del crimine organizzato utilizzato solo come pretesto per politiche e pratiche significativamente definite come "operazioni" via web da autentico Stato di polizia, magari lascian-

do spesso intoccate le grandi organizzazioni criminali che sfruttano gli extracomunitari) ma tiene conto dei costi dei servizi, delle dinamiche di mercato digitale, delle capacità del mercato di mantenere una sua irrinunciabile democraticità di fondo, che poi coincide con la tutela primaria della produzione culturale italiana e della sua identità, come produttore di contenuti che rispecchiano la nostra stessa storia e immagine sociale in un contesto economico europeo.

L'abuso deve essere chiaramente definito come tale, e almeno nei casi di effettivo rischio per le potenzialità di sopravvivenza di un soggetto deve poter essere individuato tempestivamente in una "spontanea" responsabilizzazione degli operatori. Il che deve coinvolgere tanto i produttori dei contenuti quanto i responsabili dei servizi di comunicazione. E in questo contesto tutte le forme di autoregolamentazione sono destinate a giocare un ruolo fondamentale.

La migliore forma di contrasto alla pirateria audiovisiva e informatica consiste non nella costante ricerca di sanzioni più pesanti per i consumatori di falso, ma nella progressiva riconversione di una domanda "deviata" attraverso misure di "attenzione" al prodotto originale, che può essere liberamente usato, esposto, collezionato perché esistono offerte percepite come socialmente ed economicamente convenienti e perciò accettate.

Solo il prodotto intellettuale originale, infatti, è in grado di dare soddisfazione stabile a un bisogno di cultura e non all'apparenza di una cultura "usata", ma non correttamente percepita come tale, e quindi vanamente posseduta, come oggetto utile o inutile a seconda delle circostanze e dei casi, destinata, come tutti gli oggetti, a essere abbandonata e dimenticata.

La cultura non è oggetto di proprietà, è oggetto di trasmissione della conoscenza umana, e richiede perciò attenzione alle esigenze umane, tanto quelle dei creatori e dei promotori delle opere intellettuali che quelle dei fruitori. Nessuna categoria può immaginarsi autoreferente e le istituzioni pubbliche, dal canto loro, non possono apparire inerti, disinformate, o peggio dominate da interessi diversi da quelli della collettività dei cittadini.

La chiave di volta di nuove politiche pubbliche "aperte" e autenticamente europee potrebbe essere l'ampliamento dei diritti connessi e delle forme anticipate e forfettarie di equo compenso per gli autori e il potenziamento del ruolo e della funzione pubblica di garanzia della SIAE, sia pure in una prospettiva di forte modernizzazione e di partecipazione, l'ampliamento delle possibili licenze multimediali, la pratica intelligente del "fair use" e dei "common contents", la capacità di predisporre, quindi, meccanismi di apertura sociale e di condivisione, che partano dal riconoscimento di nuove figure di autore (nel cinema e nella televisione in primo luogo, ma anche nella editoria tradizionale e multimediale) e la possibilità di sfruttare fino in fondo le opportunità che offrono la rete e le nuove tecnologie della comunicazione interattiva per diffondere in modo globale i contenuti. Senza produttore non ci sono garanzie, salvo quella apparente del prezzo molto ridotto, ma

ogni scelta di consumo o di produzione che si basa solo sul fattore economico e che si lascia alle spalle ogni altra valutazione è destinata a essere perdente. Su questi aspetti molte imprese farebbero bene a riflettere. Il problema vero è quello delle risorse e dei diritti nella società dell'informazione.

Forse è arrivato il momento di costruire, con la partecipazione di tutti gli operatori e dei soggetti pubblici interessati, un differente impianto normativo, moderno, meno ridondante, capace di adattarsi alla tecnologia e alle nuove forme di comunicazione multimediale, nel segno della difesa dei diritti degli autori e degli operatori nei nuovi mercati globali delle reti e dei servizi.

Serve una tutela giuridica, certa, ma anche forte ed efficace, perciò bisogna muoversi nel segno della semplificazione e della chiarezza normativa, soprattutto nelle definizioni di soggetti, forme creative tutelate e specifiche metodologie di tutela.

Denis Diderot nella sua *Memoria sulla libertà di stampa* del 1763 osservava:

«qual è il bene che può appartenere ad un uomo se non gli appartiene un'opera di intelletto, il frutto unico della sua educazione, dei suoi studi, delle sue veglie, del suo tempo, delle sue ricerche, delle sue osservazioni, se non gli appartengono le più belle ore, i più bei momenti della sua vita, i suoi pensieri personali, i sentimenti del suo cuore, la parte più preziosa di sé stesso, quella che non muore, quella che lo immortala? Può farsi il confronto fra l'uomo, la sostanza stessa dell'uomo, la sua anima ed il campo, il prato, l'albero o la vigna che all'origine la natura offriva in misura eguale a tutti e che il singolo ha fatto suoi soltanto con la cultura, il primo mezzo legittimo di possesso. Chi dunque ha più diritto dell'autore di disporre della propria cosa mediante il dono o la vendita? [...] Vi si griderà negli orecchi: Gli interessi dei singoli non sono niente, in concorrenza con l'interesse del tutto. Come è facile asserire una massima generale che nessuno osa contestare, ma come è difficile e raro avere tutte le conoscenze particolari necessarie per prevenire una falsa applicazione!»

*The current legislation on the rights of authors gives rise to a number of unsolved economical and social issues, and is the object of double confusion: on the one hand, the creators and producers of intellectual work are supposed to deliver their contents in the absence of adequate guarantees, on the other new repressive measures are being conceived which do not take social issues into consideration and are pushed to the point of identifying every single unauthorised usage of a resource. The digital revolution makes it nevertheless necessary to thoroughly revise the legislation in force, which is still founded on the principles laid out by the Law n. 633 of 1941 and has been the object of disharmonious and poorly coordinated interventions in the course of time. In order to adapt the law to the digital reality, it will be necessary to implement a wider reform capable of acknowledging the existence of new categories of rights holders and new forms of intellectual production. Digital contents are creating new modes of consumption which can paradoxically ensure stronger protection of the rights of authors and publishers: technologies are not attacking intellectual property, and can on the contrary offer wider and more effective guarantees. For this to happen, the actors and categories involved must engage in a process of self-regulation and constant dialogue. The role of SIAE, the Italian society for the rights of authors and publishers, and of Ministry for Cultural Heritage and Activities appears to be central. It also necessary to put aside a negative vision focused only on the risks of digital reality, which is a short-sighted and anachronistic way of looking at the problem, as much as it is short-sighted to think that we could ignore the rights of authors and publishers and build up networks where no rules exist. New, 'open' and truly European policies must be developed, so as to promote new mechanisms for the exchanging and sharing of contents for social purposes, and make the most out of the possibilities of global distribution through networks while protecting the rights of the new authors.*

*Les normes sur le droit d'auteur présentent des questions irrésolues au sujet des rapports économiques et sociaux et semblent sujettes à une double confusion: on prétend d'une part une communication dépourvue de protection pour les créateurs et les organisateurs de la production intellectuelle et de l'autre, on imagine des mesures de répression visant à repérer la moindre utilisation non autorisée et insensibles aux exigences sociales. La révolution numérique impose toutefois une révision totale de l'apparat normatif actuel, qui se base encore sur les règles fondamentales établies par la loi 633/1941 qui n'a été modifiée jusqu'aujourd'hui qu'avec des interventions incohérentes et peu coordonnées. Pour adapter les règles en vigueur à la réalité du numérique il faut concevoir une action réformatrice plus vaste, en mesure de reconnaître des nouveaux sujets et des nouveaux objets des droits d'auteurs. Les contenus numériques conduisent à des nouveaux modèles de consommation qui paradoxalement peuvent assurer une défense des droits d'auteur et des éditeurs plus ferme: les technologies n'attaquent pas la propriété intellectuelle, elles peuvent au contraire en assurer une protection plus ample et plus efficace. Pour que cela soit possible, il faut entamer une autorégulation*

*et un dialogue constant au sein des catégories intéressées. Dans ce contexte, le SIAE et le Ministère des Biens et des Activités Culturels jouent un rôle central. Il faut dépasser la vision négative qui ne perçoit que les risques de la réalité du numérique car elle ne se place que dans une perspective limitée et anachronique de même que la vision qui prétend éliminer les auteurs et les éditeurs en imaginant un réseau de communication dépourvu de règles. Il faut de nouvelles politiques publiques «ouvertes» et véritablement européennes qui puissent promouvoir des mécanismes d'ouverture sociale et de partage qui, mis à part les droits des nouveaux types d'auteurs, puissent permettre d'exploiter complètement les opportunités de la diffusion globale offertes par le réseau.*

# Granularità: un percorso di analisi

**Maurizio Zani**

Biblioteca Centrale della Facoltà di Ingegneria Alma Mater Studiorum dell'Università di Bologna

*Negli ultimi anni il termine granularità si è diffuso in maniera crescente nell'ambito delle discipline del libro e del documento, con usi spesso difformi. L'intento principale della ricerca è stato quello di "mappare" i significati assunti dal vocabolo. Granulare può essere una risorsa costituita da blocchi di testo (lessie) uniti da link ipertestuali; granulare può essere un metadato relativo a una risorsa; granulare è intesa la descrizione di o l'accesso a un articolo di una rivista piuttosto che al fascicolo al cui interno è stampato. Esigenze commerciali e sviluppo delle tecnologie si contendono anche queste definizioni. Il censimento effettuato evidenzia la diffusione del termine ma anche la profondità storica dei concetti e dei significati associati. La ricerca dell'accesso alle entità più minute di una risorsa (ricerca della parola, della frase, della citazione, dell'articolo, ecc.), favorito dallo sviluppo delle tecnologie digitali, non può d'altra parte ignorare la necessità che all'utente finale siano comunque sempre fornite anche le necessarie relazioni, in particolare quelle gerarchiche, che esistono tra le diverse parti componenti delle risorse e che possono consentirgli una più agevole navigazione.*

**N**egli ultimi anni il termine *granularità* si è diffuso in maniera crescente nell'ambito delle discipline del libro e del documento. I significati attribuitigli sono stati spesso difformi. Granulare può essere una risorsa costituita da blocchi di testo (*lessie*) uniti da link ipertestuali; granulare può essere un metadato relativo a una risorsa; granulare può essere la descrizione di un articolo di una rivista piuttosto che quella del fascicolo al cui interno è stampato. L'intento principale della ricerca è quello di "mappare" i significati assunti dal termine, per renderne più consapevole l'uso e contemporaneamente cercare di capire quali siano state le ragioni che hanno convinto tanti studiosi a utilizzarlo<sup>1</sup>.

<sup>1</sup> Modello di ricerca sul senso e il significato di un termine nel tempo rimane, per la sua completezza analitica e per l'ampiezza interdisciplinare, il classico studio, solo recentemente pubblicato, di Robert Merton e Elinor Barber sulla *seredipity* (Merton e Barber 2002).

La ricerca che qui si presenta nasce dalla tesi presentata per il conseguimento del Diploma di bibliotecario presso la Scuola Speciale per Archivistici e Bibliotecari dell'Università di Roma - La Sapienza. La tesi ha avuto come relatrice la professoressa Paola Castellucci, che ringrazio per i molteplici suggerimenti e il costante incoraggiamento. Ringrazio inoltre il professore Giovanni Solimine per i suoi preziosi consigli. La responsabilità di ogni errore è ovviamente da attribuirsi a chi scrive.

Alcuni testi che utilizzano il termine in questione possono darci una prima idea delle problematiche in discussione. Nel recente saggio sulla definizione e catalogazione delle risorse elettroniche, Stefano Gambari e Mauro Guerrini indicano nella *granularità* una caratteristica specifica di quelle risorse, definendola come

«la caratteristica della risorsa digitale di articolarsi in parti componenti (per esempio, sito Web, sezione di più pagine, singola pagina, sezioni della singola pagina, immagini, *link*); sono le parti atomiche o *lessie* che costituiscono i microtesti, i moduli di un ipertesto: il lettore dispone di varie possibilità di configurazione e di assemblamento delle unità atomiche che compongono il testo elettronico» (Gambari e Guerrini, 2002, p. 49-50).

I vantaggi per i lettori sono potenzialmente rilevanti: verificare facilmente se in un testo appare o meno un termine o una frase; accedere solo a parti specifiche di testo, senza dover procedere a una lettura sequenziale; riutilizzare facilmente il testo recuperato per costruire altri percorsi o altri documenti. Il termine *lessia* era stato introdotto da Roland Barthes per denotare le “unità di lettura” ritagliate, quasi “scomposte” in senso cinematografico, all’interno del testo, ed è poi entrato stabilmente nella letteratura sull’ipertesto<sup>2</sup>.

Leggermente diversa è la definizione del termine riconosciuta da Maria Guercio:

«L’innovazione tecnologica consente soluzioni progressivamente più raffinate e differenziate per gli utenti e, quindi, sempre meno standardizzabili e riducibili a formati unitari per chi è impegnato nella conservazione. La cosiddetta ‘granularità’ dell’informazione – un termine ricorrente nella letteratura tecnica per indicare la progressiva riduzione di oggetti un tempo unitari a un insieme di elementi informativi sempre più piccoli e tendenzialmente indipendenti – presenta rischi e difficoltà crescenti per chi ha il compito di rendere disponibile l’informazione all’interno di contesti storicamente definiti» (Guercio 2002, p. 165).

Cercando di identificare un nucleo comune di significato in queste due citazioni, si può avanzare l’ipotesi che granulare è un documento digitale articolato in più parti componenti (anche di diversa natura: testo, immagini, suoni, ecc.), che consente al lettore piste autonome di lettura e di ricerca. Gambari e Guerrini sottolineano le possibilità di ricerca e lettura per il lettore, Guercio evidenzia i rischi per una lettura decontestualizzata e per le difficoltà di gestione di quella che può diventare un’enorme mole di documenti.

Una riflessione interessante è offerta dalla recente traduzione in lingua italiana di un intervento tenuto da John D. Byrum –Jr. alla 71– a IFLA Conference di Oslo.

<sup>2</sup> Per i riferimenti specifici ai testi di Roland Barthes, e in particolare alla sua lettura di *Sarrasine* di Balzac, rinvio al testo di Landow (Landow 1997).

Il testo inglese presenta in due passaggi il termine in questione, tradotto in italiano in maniera non letterale.

#### Testo in inglese (Byrum 2005)

Increasingly, users do not depend on the OPAC to reach the resources in which they are interested. [...] Around the year 2000, however, new library systems for responding to *more granular information requests* [corsivo mio] began to appear

The Library of Congress recently commissioned a study by Marcia Bates for the purpose of exploring ways to enrich metadata records by focusing on providing additional subject and other access mechanisms (e.g. front-end user thesauri) *and increasing granularity of access and display* [corsivo mio]

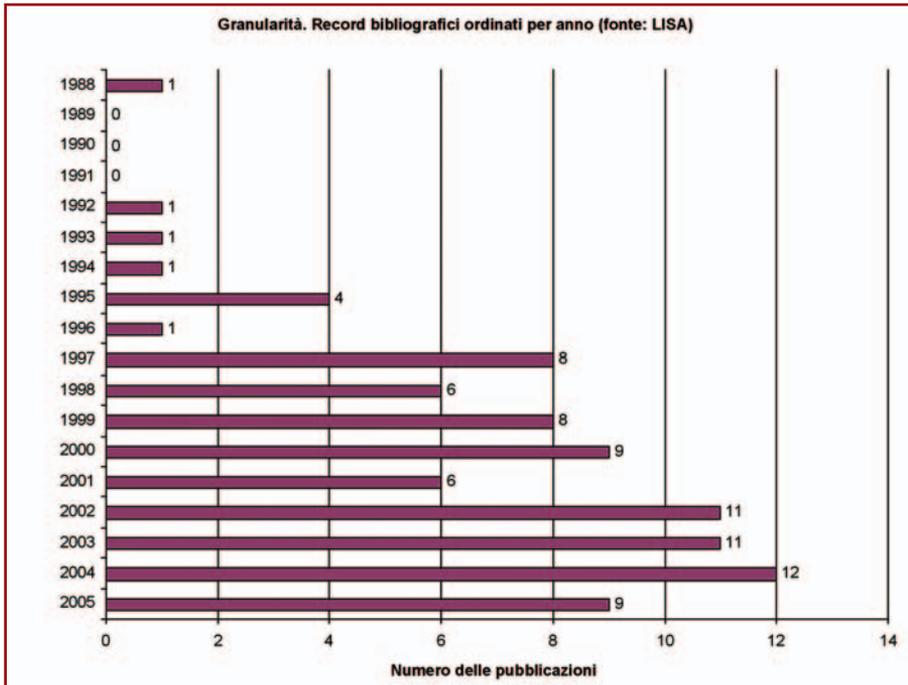
#### Traduzione italiana

Sempre più di frequente, gli utenti non si affidano agli OPAC per raggiungere le risorse di loro interesse. [...] Intorno al 2000, comunque, sono cominciati a emergere nuovi sistemi di biblioteca pensati per rispondere a *esigenze informative più sottili* [corsivo mio]

Di recente la Library of Congress ha incaricato Marcia Bates di effettuare uno studio avente l'obiettivo di esplorare le modalità di arricchimento dei record di metadati fornendo ulteriori meccanismi di accesso quali quelli per soggetto e altri (ad esempio, i tesauri di consultazione immediata per l'utente) *e maggiore sottigliezza di accesso e visualizzazione* [corsivo mio]

*Granular* e *granularity* sono stati tradotti con *sottile* e *sottigliezza*. Evidentemente il traduttore ha ritenuto ancora non ben definito e condiviso il significato dei due termini e ha scelto un termine più diffuso. In questo modo però si nasconde l'ampia diffusione del termine nella letteratura angloamericana, e soprattutto il secondo brano non appare del tutto chiaro, e non del tutto legato al passaggio successivo in cui la ricerca di Marcia Bates propone un'integrazione della ricerca nell'opac tramite l'uso di una struttura tesauroale. Come vedremo in seguito, l'uso del termine *granulare* da parte di Byrum, invece, nasce dalle molteplici riflessioni che negli ultimi dieci-quindici anni sono state elaborate dalla comunità internazionale dei bibliotecari e dei documentalisti sulle caratteristiche delle risorse informative e degli strumenti necessari per l'identificazione, la descrizione e il recupero dei documenti elettronici. In questi dibattiti il termine *granularità* è stato utilizzato con una frequenza via via maggiore<sup>3</sup>.

<sup>3</sup> Il grafico illustra l'aumento numerico dei record bibliografici presenti sul database *Library and information science abstracts (LISA)* in cui appaiono i termini *granular* o *granularity* nel titolo o nell'abstract. L'analisi delle parole chiave controllate assegnate a quegli stessi record conferma che i due termini appaiono in ambiti disciplinari non sempre omogenei e con significati diversi (il dato relativo al 2005 è incompleto).



*Granularity* nel repertorio bibliografico *Library and information science abstracts*, al 10.2.2006. La *query* effettuata in quell'occasione è stata: granular\*

Non vorrei anticipare riflessioni che sarà bene far maturare lungo il percorso di questa ricerca. Una breve indagine su dizionari, enciclopedie, basi dati e motori di ricerca dovrebbe chiarire i significati attribuiti al termine nel linguaggio comune e offrire riferimenti più precisi per proseguire la ricerca.

## I viaggi della granularità, dal concreto all'astratto

L'analisi effettuata dal *Grande dizionario della lingua italiana* di Salvatore Battaglia è ampia (Battaglia 1961-2002). Granulare è definito come ciò che è

«ridotto in granuli; composto di granuli; costituito da elementi a forma di granuli (una sostanza). Per estensione: ruvido, scabro (una superficie); granuloso».

Sono attestati riferimenti nell'ambito della geologia («struttura granulare o granitica o granitoides»), della biologia («che presenta granulazioni»), dell'elettronica<sup>4</sup>.

<sup>4</sup> «Effetto granulare: fluttuazioni piccolissime e casuali dell'intensità di corrente anodica di un tubo termoelettronico a circuito chiuso, dovute al fatto che l'arrivo degli elettroni sull'anodo del tubo è discontinuo e irregolare (e si manifesta come un caratteristico rumore di fondo)».

Il termine *granularità* rimanda direttamente invece alla *granulazione*, nell'accezione fotografica. *Granulazione* sta per «ridurre una sostanza o una materia in granuli», e consiste in un

«processo con cui vengono ridotte in piccoli granuli sostanze polverulente (come concimi, medicinali, ecc.) o fuse (come metalli, vetro) allo scopo di renderne più facili il maneggio, la distribuzione, l'utilizzazione, il trasporto, ecc.)».

La granulazione avviene o si produce nell'oreficeria antica, nell'astronomia (granulazione solare), nella fotografia. Nel linguaggio tecnico dei fotografi, si parla di granulazione a proposito dell'

«aspetto eterogeneo della struttura dell'immagine fotografica, dovuta alla sovrapposizione e al modo di dispersione dei granuli d'argento nello strato dell'emulsione».

Sono attestati comunque diversi usi in senso concreto: «insieme di granuli, corpuscoli o tubercoli» in biologia e in medicina.

Il processo della granulazione è rintracciabile anche in testi letterari. L'esperienza visiva dei colori consente quasi di coglierne l'effetto materiale e figurato. Ardengo Soffici, per esempio:

«Chiunque non è sensibile al sapore pittorico di un frego di carbone fatto da un ragazzo sul muro, al vellutato della sua granulazione nera sull'intonaco, sarà per sempre chiuso al godimento della bellezza plastica».

Emilio Cecchi, invece:

«La sua frase accompagna la linea del contorno e risponde alle sue varie tensioni, fermenta nella granulazione d'un impasto coloristico».

A parte questi ultimi, i significati attestati sono comunque tutti fortemente concreti. A sua volta, il termine "granulo" sta per «piccolo grano, granello». In particolare, il granulo è «ciascuno degli elementi che costituiscono alcuni materiali compatti (come il granito) o sciolti (come la sabbia)». Significati specifici sono attestati in astronomia (granuli del sole), in citologia e batteriologia, in botanica, in istologia, in chimica, in farmacia, e in fotografia.

«Fotogr. Ciascuno dei microcristalli d'argento che si formano nell'emulsione fotografica in seguito all'esposizione alla luce e all'azione del bagno di sviluppo e che formano l'immagine fotografica; il loro insieme, grana».

Il fenomeno chimico descritto si trasforma poi in immagine e in effetto visivo. Il significato registrato nelle discipline fisiche, invece, è forse uno dei più interessanti, e lo possiamo considerare un'altra traccia importante per il prosieguo:

«Fis. *Granulo di energia*: nome comune del *quanto* di energia».

La composizione, la tessitura, l'aspetto stesso dei corpi e degli oggetti costituiti da granuli o granelli più o meno piccoli è espressa tramite il sostantivo *grana*. In un'emulsione esposta alla luce e sviluppata, il termine *grana* esprime il numero e la consistenza dei granuli microscopici di sali d'argento, considerati in proporzione alla superficie di una pellicola. La grana di una pietra o di un metallo sono evidenziati soprattutto al momento della frattura, e la finezza o grossolanità di questi materiali dipendono dalla grandezza relativa dei grani in cui la massa si suddivide e dalla maggiore o minore vicinanza, per cui la massa assume un aspetto più o meno compatto.

Il termine *grana* è utilizzato anche per rendere l'aspetto più o meno ruvido di una superficie, di una pelle, di un tessuto o di una superficie metallica trattata da un artigiano, da un artista, da un orafo. La stessa voce umana ha una grana. È Roland Barthes, con una geniale sinestesia, a parlare di *grana della voce*, caratteristica unica di ogni voce individuale, e che ne rende il riconoscimento e l'ascolto un'esperienza unica e irripetibile<sup>5</sup>.

Per riconoscere e apprezzare l'individuale grana o granularità di un oggetto è dunque necessario attivare i propri sensi. Dalle caratteristiche più concrete, che si possono toccare o vedere, si passa a caratteristiche più astratte, che si colgono con l'udito o la vista (anche attraverso lo strumento fotografico), fino addirittura a utilizzare lo strumento della ragione per comprendere il quanto di energia.

Negli ambiti scientifici e tecnici, il termine *granularità* ha acquisito ulteriori significati. Il Dizionario tecnico scientifico McGraw-Hill Zanichelli riporta alla voce *granularity* tre significati, legati a tre ambiti disciplinari diversi:

«GRAF [ARTI GRAFICHE]. The distribution of grains in a portion of photographic material that has been uniformly exposed and processed.

ING SIS [INGEGNERI DEI SISTEMI]. The degree of which a system can be broken down into separate components, making it customizable and flexible.

<sup>5</sup> Riporto al proposito alcune indicazioni di un recente libro sulla filosofia dell'espressione vocale di Adriana Cavarero: «Secondo Roland Barthes [...] il proprio della voce starebbe invece in quel che egli chiama la sua grana. Piuttosto che la sfera del soffio e del respiro, essa riguarda 'la materialità del corpo che sgorga dalla gola, là dove si forgia il metallo fonico'. L'attenzione va dunque alla cavità orale, luogo erotico per eccellenza. La grana della voce concerne soprattutto il modo in cui, mediante la voluttà dell'emissione sonora, la voce lavora nella lingua. [...] Compito della voce è dunque quello di fare da tramite o, meglio, da snodo, fra corpo e parola» (Cavarero 2003; cito dalla versione online [http://www.feltrinelli.it/SchedaLibroEstratto?id\\_estratto=1011](http://www.feltrinelli.it/SchedaLibroEstratto?id_estratto=1011)).

PETROGR [PETROGRAFIA]. The feature of rock texture relating to the size of the constituent grains or crystals» (McGraw-Hill Zanichelli 2004).

Il significato del termine nell'ambito delle arti grafiche e della fotografia o nell'ambito della petrografia è quello attestato anche nei dizionari della lingua italiana, ma emerge ora anche la disciplina dell'ingegneria dei sistemi. La granularità indica il grado in cui un sistema può essere articolato in componenti separate, così da renderlo flessibile e adeguabile alle esigenze degli utenti. Queste esigenze possono essere soddisfatte tramite l'utilizzo degli strumenti informatici. È proprio l'ambito informatico a fare la parte del leone nei significati più diffusi del termine, e la conferma la si può ottenere anche utilizzando l'archivio di Google come una sorta di corpus, da cui estrarre le definizioni più in uso. La serie dei significati si allunga di giorno in giorno, e la grande maggioranza fa riferimento all'ambito informatico, in particolare alla progettazione dei data base, alla programmazione parallela, al calcolo parallelo, all'elaborazione del testo e all'organizzazione delle memorie. Il *Dictionary of Computing* tenta di sintetizzare il significato del termine:

«A measure of the size of the segments into which memory is divided for purposes of either memory protection or virtual-memory management»<sup>6</sup>.

Anche *Whatis.com*, ottimo strumento per muoversi tra le definizioni relative al mondo dell'information technology e di Internet, tenta di tirare le fila, con una definizione ampia del termine in grado di abbracciare tutti i diversi settori disciplinari.

«Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or activity. It may help to think of it as: which type of 'granule' are we looking at? This term is used in astronomy, photography, physics, linguistics, and fairly often in information technology. It can refer to the level of a hierarchy of objects or actions, to the fineness of detail in a photograph, or to the amount of information that is supplied in describing a person's age. Its meaning is not always immediately clear to those unfamiliar with the context in which it's being used»<sup>7</sup>.

La definizione intende "coprire" diversi significati. In tutte le discipline, la granularità indica il livello di dettaglio, la precisione con la quale si può cogliere un oggetto o un'attività. Non sono dati riferimenti precisi (misure, numeri, ecc.), visto che si può usare in tantissimi contesti e si può adattare alle esigenze di ciascun utente.

<sup>6</sup> *Granularity*, in: *A Dictionary of Computing*, Oxford: Oxford University Press, 1996, consultato su *Oxford Reference Online*, <http://www.oxfordreference.com>.

<sup>7</sup> Cfr. <http://whatis.techtarget.com/definition/0,289893,sid9gci212209,00.html>.

Rischiosa appare comunque la scelta di enumerare, quasi confondendoli, i diversi significati di dimensione relativa, livello di dettaglio, scala, profondità, quasi siano termini intercambiabili<sup>8</sup>.

Vale la pena chiedersi finalmente come si sia legato il termine di *grana* o di *granularità* al concetto e al termine *informazione*. Concepire come granulari l'informazione, un supporto documentario o uno strumento di computazione o ricerca appare come un'innovazione molto recente. Una ricerca sull'*Oxford english dictionary*, per esempio, non attesta storicamente significati assimilabili. L'ampia diffusione del termine nel settore informatico segnala che probabilmente è stato questo il canale privilegiato attraverso il quale il termine è entrato nello specifico delle scienze del libro e del documento. Credo però che il percorso non sia stato del tutto lineare, e probabilmente il percorso concettuale non sempre diretto. Le indicazioni raccolte suggeriscono che la diffusione del termine possa essere avvenuta attraverso percorsi molteplici e non sempre lineari. Le direzioni di ricerca sulle quali ritengo opportuno portare l'attenzione sono sostanzialmente tre.

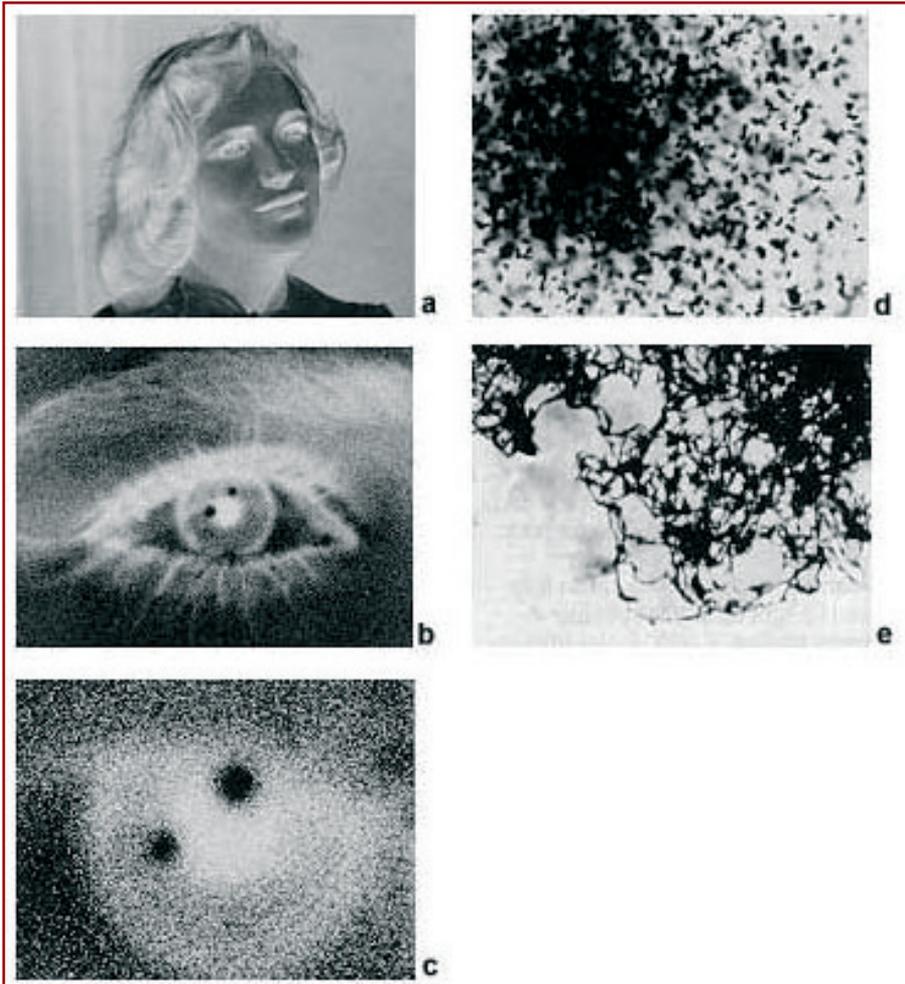
La prima riguarda il significato fotografico del termine granularità. Tra tutti i significati del termine, questo appare forse il più diffuso e comune. L'effetto granulare di un'immagine è basato sulle caratteristiche fisiche della pellicola fotografica, ma produce ben precise conseguenze visive. Si tratta di un effetto ben conosciuto, che si verifica quando, "zoomando" progressivamente, l'occhio percepisce la grana della fotografia, fino al punto in cui tale visione diventa sufficientemente fastidiosa per l'occhio e per le nostre capacità cognitive, l'immagine diventa granulosa e si finisce per perdere il contesto in cui si colloca il particolare che si intendeva cogliere. La sequenza di immagini di seguito riportate – dal titolo *Graininess and granularity area* – mostra come la granularità di una immagine fotografica si possa slabbare e sfrangere fino a perdere di significato per chi guarda.

Da un sito specificamente dedicato alla fotografia possiamo cogliere i significati tecnici dei termini in questione.

«GRANA. Osservando un'immagine negativa al microscopio o un forte ingrandimento si possono notare i piccoli ammassi di argento metallico che formano l'immagine dopo lo sviluppo. Questi ammassi prendono il nome di grana. [...]

GRANULARITÀ. Quantificazione oggettiva del concetto di grana. Indica la mancanza di uniformità della Densità di un'emulsione fotografica ovvero lo spostamento dei valori di Densità rispetto ad uno standard. Questo dato viene espresso dai fabbricanti in termini di granularità RMS diffusa. Più alto è il valore maggiore è la grossezza della grana. Un valore 8 è tipico per una pellicola 100 Iso.

<sup>8</sup> Non stupirà che la diffusione del termine abbia finito per trasformare la *granularità* addirittura in una *buzzword*, un termine di moda, usato per impressionare gli sprovveduti: (<http://www.buzzwhack.com/buzzcomp/indgk.htm>).



A 2.5X enlargement of a negative shows no apparent graininess. (b) At 20X, some graininess shows. (c) When a segment of the negative is inspected at 60X, the individual silver grains start to become distinguishable. (d) With 400X magnification, the discrete grains are easily seen. Note that surface grains are in focus while grains deeper in the emulsion are out of focus. The apparent 'clumping' of silver grains is actually caused by overlap of grains at different depths when viewed in two-dimensional projection. (e) The makeup of individual grains takes different forms. This filamentary silver, enlarged by an electron microscope, appears as a single opaque grain at low magnification. <http://www.kodak.com/US/en/motion/students/handbook/sensitometric6.jhtml>.

GRANULOSITÀ. Impresione soggettiva della grana che appare in un'immagine fotografica. Essa dipende dalla distanza di osservazione, dalle condizioni di visione e dal visus dell'osservatore»<sup>9</sup>.

<sup>9</sup> [http://www.reflex.it/enciclopedia/enciclopedia\\_g-i.htm](http://www.reflex.it/enciclopedia/enciclopedia_g-i.htm).

Se è stata esperienza comune per tanti anni osservare i piccoli puntini che caratterizzavano le fotografie riprodotti sui giornali quotidiani (effetto dovuto a un particolare e più veloce per l'epoca sistema di trasmissione), oggi osserviamo i video dei computer e parliamo della loro risoluzione esprimendoci in pixel<sup>10</sup>, come esprimiamo con il numero dei pixel anche la risoluzione di un'immagine. Un procedimento comune è quello di ingrandire anche le immagini digitali cogliendo proprio le unità minime che le compongono.

L'effetto granulare si può associare anche all'uso dello zoom, ovvero dell'obiettivo fotografico a distanza focale variabile che consente di ampliare o ridurre l'inquadratura. Lo zoom dovrebbe mantenere sempre a fuoco l'immagine, ma il movimento di esaltazione dei particolari e l'ansia di ampliare l'inquadratura può portare a perdere il riferimento al contesto, che solo riesce a farci comprendere appieno il significato di quell'immagine.

La diffusione attuale delle immagini fisse e in movimento e dei sistemi di riproduzione digitale — video, televisione, cellulari, ecc. — ha senz'altro giocato fortemente a favore della diffusione del termine granularità. Come si esprimeva un recentissimo articolo apparso su un giornale non scientifico, *The Christian science monitor*, "We're all going dotty" (Walker 2006)<sup>11</sup>.

La seconda traccia da seguire rimanda al mondo delle scienze fisiche. Il "granulo di energia" citato da Battaglia come «nome comune del quanto di energia» riporta alla ricerca effettuata dai fisici nel settore dell'infinitamente piccolo, in particolare sulla composizione e le caratteristiche del mondo subatomico, e le lunghe, secolari diatribe sulle caratteristiche ondulatorie o corpuscolari della luce. Il tentativo di identificare la realtà fisica più piccola ha portato a teorie sempre più sviluppate e a ipotesi radicali, come quelle della fisica quantistica, per la quale l'uso stesso degli strumenti di misurazione rende impossibile determinare con precisione meccanistica velocità e posizione delle particelle subatomiche. Ogni realtà fisica subatomica diventa sempre più piccola e astratta, e le caratteristiche fondamentali dei "granuli" più piccoli non sono la massa o l'energia, bensì le informazioni trasportate. Su questa base, l'informazione viene concepita da alcuni come linguaggio universale della nuova scienza fisica<sup>12</sup>.

Alla metà del secolo scorso, negli anni successivi all'affermazione della fisica quantistica, un fisico prestigioso, Donald Mackay, ipotizzava che, dato che ogni

<sup>10</sup> "Pixel. Un pixel (abbreviazione inglese di picture element) è uno dei molti minuscoli puntini che compongono la rappresentazione di un'immagine nella memoria di un computer. Solitamente i punti sono così piccoli e così numerosi che se vengono stampati su carta o visti su un monitor appaiono fusi in una unica immagine. Colore e intensità di ciascun punto sono usati dal computer per rappresentare una piccola area dell'immagine" (<http://it.wikipedia.org/wiki/Pixel>).

<sup>11</sup> L'autore associava i punti (dot) ai pixel e grani di sabbia, approfondendo inoltre il significato di *granularity*.

<sup>12</sup> Sull'argomento, cfr. il recente libro divulgativo del fisico Hans Christian Von Baeyer, dal titolo significativo: *Informazione. Il nuovo linguaggio della scienza* (Baeyer 2004).

realtà fisica presuppone comunque uno scambio di informazioni, si potesse identificare una «unit of structural information [...] for this unite there already exists the suitable name of a *logon*»<sup>13</sup>:

«Mackay suggests that the apparently ‘hard’ science of physics is epiphenomenal to the real science of information: it was because information was inherently quantal that physical reality was discovered to have the same property (Mackay, 1950)» (Rayward 1996).

La strada è diversa da quella battuta da Shannon e Weaver, che misuravano la quantità e la qualità dell’informazione scambiata in una transazione comunicativa, e per questo coniarono il bit<sup>14</sup>. Nella prospettiva di Mackay, e di diversi studiosi nei decenni successivi, la ricerca di una unità minima di informazione era sostenuta dall’ipotesi che fosse proprio l’informazione lo strumento di base per l’interpretazione della realtà fisica, in tutti i suoi aspetti.

La terza strada da seguire riguarda la scienza dell’informazione in senso lato, o per meglio dire l’informatica e la scienza dei computer. In queste discipline l’uso è oggi attestato massicciamente<sup>15</sup>, e la consultazione dei materiali disponibili in Rete lo può confermare. La consultazione di INSPEC, il repertorio bibliografico prodotto dall’Institution of Electrical Engineers di Londra<sup>16</sup>, permette di tentare di cogliere il momento in cui il termine entra nel linguaggio informatico. Se dalla fine del XIX secolo il significato del termine è fortemente ancorato al contesto fotografico, a partire dal 1969-1970 articoli si occupano della granularità dei materiali utilizzati nella registrazione di informazioni in formato digitale, e poi nel 1976 il sostantivo *granularity* appare usato in un senso nuovo, più precisamente nell’ambito della ricerca sulla gestione ottimale dei data base<sup>17</sup>. L’uso astratto, e vorrei dire già propriamente metaforico, del termine granularità (o granuli) si conferma negli anni successivi e si estende alla ricerca sui sistemi di gestione dei data base, prima di

<sup>13</sup> Traggo l’indicazione della citazione da un intervento di W. Boyd Rayward sulla storia e la storiografia dell’*information science* (Rayward 1996, p. 66).

<sup>14</sup> «Unità di misura della quantità di informazione. Non esistono sottomultipli del bit: la minima quantità d’informazione è pari a 1 bit ed equivale alla scelta tra due valori (sì/no, vero/falso, acceso/spento), quando questi hanno la stessa probabilità di essere scelti» ([http://it.wikipedia.org/wiki/Bit\\_\(informazione\)](http://it.wikipedia.org/wiki/Bit_(informazione))).

<sup>15</sup> L’interrogazione della base dati bibliografica multidisciplinare *Current Contents* conferma per il periodo 2001-2005 come proprio la *computer science* sia nettamente il settore disciplinare in cui il termine è più diffuso, superando le discipline delle scienze dei materiali e registrando un’elevata diffusione anche nei settori affini dell’information technology e dell’intelligenza artificiale e della robotica.

<sup>16</sup> INSPEC indicizza articoli di riviste e atti di convegno nei settori dell’ingegneria informatica ed elettronica, della fisica applicata e dell’ottica.

<sup>17</sup> Gray-JN – Lorie-RA – Putzolu-GR – Traiger-IL, *Granularity of locks and degrees of consistency in a shared data base*, in: Nijsen-GM, *Proceedings of the IFIP Working Conference on Modelling in Data Base Management Systems*, North-Holland: Amsterdam, 1976, p. 365-94.

espandersi ulteriormente in altri settori delle scienze dell'informazione.

I tre percorsi che ho brevemente indicato costituiscono il quadro di riferimento più ampio al cui interno si potrebbe essere svolta questa possibile vicenda di "contrabbando" di significato tra diverse regioni disciplinari confinanti. All'interno di questa cornice, vale la pena indicare almeno altre due aree fortemente interdisciplinari che utilizzano in maniera estremamente interessante il termine *granularità* e che ritengo possono avere concorso alla diffusione del termine nel campo delle discipline del libro e del documento.

Il termine *granularità* è infatti attestato nei manuali e nelle ricerche che si occupano di sistemi di informazione geografici (GIS). Il *GIS Dictionary* della ESRI (un'importante azienda produttrice di software), alla voce *Granularity* riporta due significati:

- «1. The coarseness or resolution of data. Granularity describes the clarity and detail of data during its capture and visualization.
2. The objective measure of the random groupings of silver halide grains into denser and less dense areas in a photographic image»<sup>18</sup>.

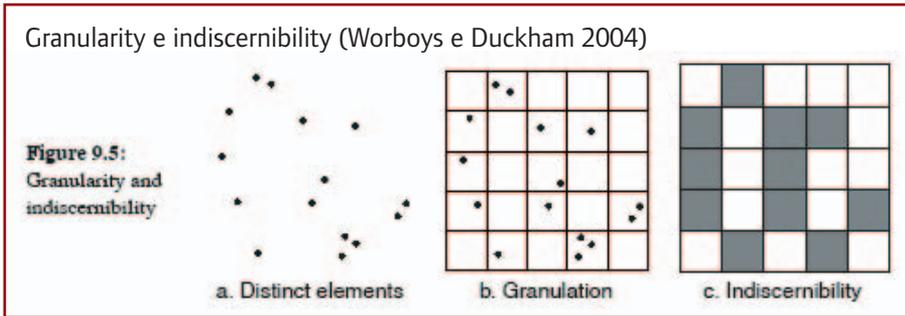
Come si vede il termine ha nel primo caso un'accezione abbastanza generale relativa al settore dell'elaborazione dei dati nei sistemi informativi territoriali, e poi riprende la definizione tecnica di origine fotografica. Una più ampia e interessante definizione la offre un accreditato e aggiornato testo di riferimento dedicato ai GIS, i cui esempi, che utilizzano anche il linguaggio visuale, danno un senso più chiaro al termine.

«Granularity is closely related, but not identical to imprecision. Granularity refers to the existence of clumps or grains in observations or representations, in the sense that individual elements in the grain cannot be distinguished or discerned apart from each other.

For example, imagine a remotely sensed image of a region of the Earth. In this image, any two locations on the ground that fall within the same pixel cannot be distinguished in characteristics by the observation. So, granularity in an observation makes things *indiscernible* that in principle are distinguishable from each other» (Worboys 2004).

Il problema è il trattamento di un numero rilevante di dati, che debbono essere resi intellegibili in fase di osservazione o rappresentazione, selezionando un particolare livello di dettaglio. La scelta di un determinato dettaglio o grado di precisione comporta sempre una perdita di informazioni, per cui il termine *granularità* si associa anche all'imprecisione e all'indiscernibilità dei dati rilevati, registrati ed elaborati.

<sup>18</sup> <http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.Search&search=true&searchTerm=granularity>.



La letteratura sui GIS potrebbe assumere un notevole interesse per questa ricerca, come snodo interdisciplinare dove s'incontrano le scienze del territorio, dell'informazione e le scienze cognitive. In pratica, l'osservazione di dati relativi alla realtà geografica (fenomeni, processi ed entità fisiche, ma anche economiche e sociali) avviene tramite più passaggi, quello fotografico (dove cioè sono implicati i processi chimici della definizione dell'immagine), quello della rappresentazione tramite riproduzioni cartografiche (che usano da secoli lo strumento della rappresentazione del territorio su scale diverse) e quello dell'implementazione di data base numerici.

La rappresentazione e la ricerca di informazioni in questo settore richiede dunque competenze e abilità cognitive diverse, che debbono collaborare strettamente. Non credo possa sorprendere che questo sia un ambito disciplinare di grande interesse per un possibile sviluppo della ricerca sulla granularità. Proprio seguendo il filo di queste problematiche, è comprensibile che il termine granularità appaia in un interessante riflessione offerta da un recente intervento di alcuni studiosi di GIS (Frederico Fonseca, Max Egenhofer, Clodoveu Davis e Gilberto Camara), che hanno sottolineato sì la sempre maggiore disponibilità di dati relativi al nostro pianeta, ma anche come queste sistemi informativi siano sempre più distribuiti ed eterogenei. Ciò costituisce una vera sfida alla ricerca, il cui problema diventa quello di integrare le diverse informazioni, raccolte ed elaborate a differenti livelli di dettaglio o di granularità. Se facciamo riferimento all'immagine del testo di Worboys e Duckham, quando debbono dialogare data base diversi la griglia utilizzata per raccogliere e manipolare gli originali dati granulari deve essere simile, pena la non comunicabilità.

«Frequently, the information exists, but integration is very difficult to achieve in a meaningful way because the available information was collected by different agents and also with diverse purposes. The effective integration of multiple resources and domains is known as interoperation. [...] In the past, exchanging geographic information was as simple as sending paper maps or raw data tapes through the mail. Today, computers throughout the world are con-

nected and the use of GIS has become widespread. The scope of interoperability has changed from static data exchange using flat files to global systems, interconnected using sophisticated protocols to exchange information on-line. In the future, computers are expected to be able to share not only information but also knowledge» (Fonseca e al. 2002).

Il problema di integrare o interoperare archivi diversi è dunque anche un problema di «different levels of detail», in pratica un problema di granularità, che i computer, da soli, non riescono a risolvere.

Per completare queste avventurose incursioni in campi disciplinari così lontani dalle abituali letture di un bibliotecario, vale la pena sottolineare come secondo alcuni uno strumento in grado di coordinarsi velocemente e in maniera flessibile tra più sistemi informativi, spesso a livelli diversi di granularità, esista già: si tratta del cervello umano. Se è consentito pescare nella letteratura dell'intelligenza artificiale – disciplina che mantiene da decenni la leadership tra le scienze della cognizione e dell'informazione – vale la pena leggere almeno la parte introduttiva di un breve ma fondamentale articolo pubblicato da Jerry Hobbs una ventina di anni fa. Il titolo appare decisamente pertinente: *Granularity*.

«We look at the world under various grain sizes and abstract from it only those things that serve our present interests. Thus, when we are planning a trip, it is sufficient to think of a road as a one-dimensional curve. When we are crossing a road, we must think of it as a surface, and when we are digging up the pavement, it becomes a volume for us. [...] Our ability to conceptualize the world at different granularities and to switch among these granularities is fundamental to our intelligence and flexibility. It enables us to map the complexities of the world around us into simple theories that are computationally tractable to reason in» (Hobbs, 1985, p. 432).

Questo approccio suggerisce a Hobbs che quello della granularità è un modo di concepire l'intelligenza tipico degli esseri umani, e che le macchine dovranno per forza avere.

«It is that our knowledge consists of a global theory together with a large number of relatively simple, idealized, grain-dependent, local theories, interrelated by articulation axioms. In a complex situation, we abstract the crucial features from the environment, determining a granularity, and select the corresponding local theory» (Hobbs 1985, p. 435).

Come dotare le macchine di questa capacità di *switchare* tra diversi livelli di dettaglio dei sistemi informativi e delle realtà fisiche è problema che ha interessato non

solo gli studiosi dell'intelligenza artificiale, ma anche gli studiosi delle scienze dell'informazione e i ricercatori di molteplici settori disciplinari, che in questi anni si sono avvicinati al tema delle ontologie. Per quanto riguarda le discipline del libro e del documento, sicuramente i tesauri disciplinari e i sistemi classificatori possono funzionare in questo senso, sebbene non in forma automatica. La riflessione degli studiosi su questo tipo di problemi si è recentemente risvegliata soprattutto da quando si è imposto il tema delle ontologie. Avrò modo di sfiorare di nuovo questo argomento.

Tutti i percorsi intellettuali qui troppo velocemente indicati hanno visto un progressivo abbandono dei significati più concreti del termine, con il passaggio a un uso sempre più astratto e quasi metaforico. In questa progressiva mutazione dalla definizione delle caratteristiche concrete e fisiche di un oggetto granulare all'analisi di una realtà impalpabile e incorporea come quella dei dati, delle memorie, dell'informazione in generale, l'uso metaforico ha forse approfittato dell'abitudine all'uso del termine da parte dei ricercatori di formazione tecnica o scientifica. Il termine granularità al posto del più tradizionale "dettaglio" potrebbe avere dato agli utilizzatori una sensazione di maggiore scientificità. Nello stesso tempo, avere perso il riferimento concreto – le diverse forme e composizioni di minerali e metalli – può avere contribuito a espanderne l'uso e i significati fino quasi a perdere molto della precisione e affidabilità scientifica, e la ricerca informatica potrebbe avere finito per utilizzarlo in senso metaforico. A ogni livello, infatti, l'informatica ha visto un'elevatissima diffusione di ogni tipo di metafora legata all'interazione uomo-macchina. Non c'è praticamente alcun termine di quotidiano uso informatico che non sia fortemente segnato da metafore (windows, mouse, banche dati, rete, ragni, agenti intelligenti, motori di ricerca, data mining, ecc.)<sup>19</sup>. L'ambito informatico non può non utilizzare metafore in cui è fortemente implicato il senso della vista. Alla diffusione della metafora della granularità può dunque avere giovato la possibilità di "pescare" via via da esperienze concrete come quella della scienza dei materiali, dell'ottica e della fotografia, per imporsi nel campo informatico – settore disciplinare oggi al centro di numerosissime intersezioni disciplinari – e da qui sbarcare nell'area delle scienze del libro e del documento<sup>20</sup>.

<sup>19</sup> Gerald Johnson sostiene che l'impiego della metafora nel mondo della *computer science* ha una sua specificità che spiega la sua particolare diffusione, e cioè che oltre a descrivere cose innovative, serve a dare consistenza a realtà talmente astratte da risultare altrimenti incomprensibili e di difficile utilità pratica (Johnson 1994). Ronald Day ricorda come gli stessi Shannon e Weaver usassero termini metaforici in diversi testi collocati alle origini della teoria della comunicazione (Day 2000).

<sup>20</sup> L'uso delle metafore – e in particolare di quelle che hanno utilizzato esplicitamente anche il linguaggio delle immagini, attivando gli stili cognitivi visivi di tanti ricercatori – ha sempre avuto anche in queste discipline un'importanza rilevante. In particolare, Werner Bies offre diverse suggestioni e ipotesi di ricerca sull'uso delle metafore nel campo della *knowledge organization*, per esempio quella dell'albero del sapere (Bies 1996).

L'affermazione della granularità come metafora e ponte di transizione tra settori disciplinari originariamente distinti credo possa rientrare in un processo tipico dello sviluppo del sapere. È anzi ampiamente accettato che in ambito scientifico il momento più propriamente creativo è difficilmente formalizzabile, al di là di quanto non appaia nel momento della formulazione teorica e della pubblicazione formale a stampa. «Formulazioni di tipo analogico o metaforico, attraverso invenzioni sperimentali o utilizzazioni eretiche di sistemi precedenti», collaborano in maniera complessa alla costruzione di sperimentazioni e teorie scientifiche.

«La formazione dei concetti [...] può avere anche a che fare con i linguaggi del senso comune. Il modo in cui si usa questo senso comune, così come le terminologie di discipline diverse da quella che è al lavoro, sono dati dalla pratica metaforica; attraverso questa pratica si formano concetti 'pescandoli' dai linguaggi di discipline diverse da quella in oggetto» (Gagliasso 2002, p. 5).

Secondo la teoria di Black<sup>21</sup>, le metafore sono distinguibili come sostitutive o esegetiche (pedagogiche) e metafore interattive o costitutive. Le prime costituiscono un "fragile ponte temporaneo" offerto all'intuizione in mancanza di spiegazioni causative o ricostruttive lineari. Anche le metafore sostitutive possono comunque svolgere una funzione costitutiva, aprendo possibilità di ricerca nuove, attraverso slittamenti progressivi del significato e un continuo scivolamento di significato da una disciplina all'altra. Viene da chiedersi se questa, come altre metafore scientifiche, consentirà di «allargare l'ambito osservativo e ritagliare diversamente, riorganizzandoli, i campi tematici» del nostro specifico disciplinare (Gagliasso, p. 6). Certamente credo che potrebbe essere stimolante rileggere alla luce di questa nuova metafora riflessioni teoriche e soluzioni pratiche prodotte nel corso dei secoli da bibliotecari e documentalisti e, più recentemente, dagli stessi informatici. Se proviamo a pensare al lavoro concreto di chi si occupa di testi e documenti – siano essi bibliotecari e documentalisti o informatici che si cimentano con le problematiche relative alle nuove realtà testuali e documentarie create in ambito elettronico e digitale – è evidente che alcune delle domande di base che si pongono ai cultori delle diverse discipline sono in qualche modo largamente simili, anche per quanto riguarda la granularità e il livello "atomico" dell'informazione e della documentazione. Qual è il livello di dettaglio a cui può giungere il trattamento dell'informazione, o meglio del testo o del documento? Quale è la misura ottimale del "segmento di memoria" (scritta) in cui si può dividere un testo o un documento al fine della sua migliore indicizzazione, circolazione o conservazione? Quali realtà "discrete" o "granulari" si possono riconoscere e individuare in un documento o in

<sup>21</sup> Riprendo la teoria della metafora di Max Black dal testo di Carmela Morabito sulla metafora nelle scienze cognitive (Morabito 2002).

un testo: una parola, una frase, un concetto, un'immagine?

La metafora della granularità credo abbia riportato alla ribalta temi e problematiche dalle profonde radici storiche. Ciò non per trovare a tutti i costi dei precursori, piuttosto per tentare di rileggere una parte della lunga storia delle discipline bibliografiche e documentarie, e del possibile intrecciarsi con la nascita delle scienze informatiche.

Storicamente, il documento cartaceo nella sua unità fisica ha costituito un supporto indispensabile per la registrazione, la conservazione e il recupero dei dati. L'unità documentaria è una realtà che il bibliotecario e documentalista può studiare, gestire, indicizzare, ma non può modificare. Costituisce anzi una sorta di vincolo al recupero dei dati registrati. La definizione che di unità documentaria danno Anderson e Perez-Carballo è in questo senso molto precisa.

«Documentary unit. The document, document segment, or collection of documents to which entries in an index refer and on which they are based. Examples of verbal documentary units include sentences, paragraphs, pages, complete articles, books, complete serial runs, collections of archival materials, microform sets, and entire library collections» (Anderson e Perez-Carballo 2005).

Individuare un documento all'interno di una biblioteca o di una collezione costituisce una delle tradizionali sfide dei bibliotecari, ma un ulteriore problema è stato per secoli proprio quello di "liberare" i dati contenuti nell'unità documentaria. La produzione di indici relativi a questi due problemi si è evoluta nella storia secondo metodi e con strumenti spesso molto diversi. Negli ultimi decenni, proprio il superamento dell'unità documentaria ha costituito una vera e propria sfida per la ricerca nel settore dell'*information retrieval*.

Non esiterei a rintracciare le radici del tema della granularità anche nella storia dell'ideazione e dello sviluppo dei sistemi di indicizzazione dei testi scritti, a partire dalla creazione di note, glosse, indici generali, tematici e alfabetici per l'accesso a punti o soggetti particolari dei libri (Rouse 1989). Tra gli strumenti paratestuali in grado di consentire una produttiva lettura non lineare del testo, Fayet Scribe annovera il riassunto e l'abstract, dai registri diplomatici delle corti papali ai riassunti degli articoli scientifici, dal «Journal des Savants» alle riviste del nostro secolo. Il periodo finale del Medio Evo, in questa prospettiva, può essere considerato addirittura il periodo del "triomphe de l'indexation".

«Jamais, sauf peut-être à notre époque, l'intérêt ne s'est autant porté vers le contenu sémantique et l'usage qui en est fait. Du point de vue du repérage, la révolution actuelle du texte numérique serait à rapprocher de la 'révolution gothique', plus qu'à celle de l'imprimé de la Renaissance. [...]

De fait, l'apparition de l'imprimerie fait porter l'accent plus sur 'l'objet livre' lui-même que sur son contenu. À partir de la Renaissance jusqu'au milieu du XXe siècle, ce sont les voies d'accès par la référence (bibliographie, catalogue) et leur système de classifications qui seront les plus exploitées» (Fayet Scribe 1997).

L'indicizzazione completa del documento e la liberazione dei concetti e dei dati dai loro confini documentari è al centro anche della riflessione di Paul Otlet, fondatore della scienza della documentazione. L'archivio bibliografico ideato e implementato per anni da Otlet era basato sul *monographic principle*, secondo il quale occorre registrare analiticamente ogni documento o parte di documento dedicata a un determinato argomento. Il lavoro dell'indicizzatore doveva essere mirato non solo al livello dell'intera opera, di un suo capitolo o di una sua sezione, ma anche di ogni dato o "fatto" che fosse possibile estrarre dal testo (Rayward 1994, p. 169). Lo studioso che in fase di ricerca avesse voluto recuperare tutte le unità informative dedicate a un singolo argomento avrebbe potuto utilizzare lo schema costituito dalla classificazione decimale universale. La sicurezza di poter individuare i "fatti" e "ricostruire" la conoscenza, evitando le disfunzioni, i vincoli e le ripetizioni presenti in letteratura, costituisce ai nostri occhi la testimonianza di una precisa epoca storica, quella in cui il sapere scientifico poteva sperare di giovare di un sistema di accrescimento continuo e senza scosse, in cui ogni nuovo libro, ogni nuovo articolo aggiungeva ulteriori informazioni e conoscenze, senza stravolgere il quadro preesistente. La visione di Otlet mantiene comunque a tutt'oggi il suo fascino, soprattutto per le discipline scientifiche, in cui l'apprendimento è fortemente strutturato e l'ideale scientifico è quello di un sapere che cresce costantemente sulle esperienze precedenti.

Gli obiettivi di Otlet non erano molto diversi da quelli che si sono posti negli ultimi decenni i primi sistemi di *information retrieval*. Un'ipotesi di lavoro che appassionava informatici e documentalisti anche degli anni '60 era quella della possibilità di creazione di sistemi di "fact retrieval", distinti dai sistemi di "document retrieval" e di "data retrieval" (intesi come dati numerici). I sistemi di "fact retrieval" avrebbero dovuto essere in grado di rispondere a un utente in maniera diretta, piuttosto che recuperare un testo che può contenere la risposta o, più comunemente, il riferimento a tale testo. Secondo Frederick Wilfrid Lancaster e Amy Warner, quella del *fact retrieval* è stato «the most ambitious project in information retrieval» (Lancaster e Warner 2001, p. 43), ed è ancora all'ordine del giorno. Attualmente i sistemi di *information retrieval* consentono di ipotizzare non solo un efficiente ed efficace recupero delle unità documentarie, ma anche un altrettanto avanzato recupero delle informazioni salvate in "segmenti" più piccoli dei documenti, l'identificazione dei concetti sottesi e la produzione in forma automatica di nuova conoscenza. Particolarmente sviluppati gli approcci nati dall'elaborazione del linguaggio naturale (NLP, *natural language*

*processing*), e quello che viene chiamato *text extraction*<sup>22</sup>.

Le scienze dell'informazione hanno cercato dunque di accelerare in ogni modo la circolazione dei documenti scientifici e di superare le barriere "fisiche" che, anche secondo Otlet, frenavano la stessa ricerca scientifica. L'evoluzione delle tecniche ha visto una decisa accelerazione con l'affermazione del documento elettronico e la creazione di un nuovo strumento di recupero dell'informazione, l'ipertesto. In questo modo si sono aperte nuove strade alla realizzazione dell'antico sogno del documento granulare. Prima di parlare dell'ipertesto, è obbligatorio affrontare alcune altre riflessioni di ordine storico.

### L'unità minima dell'informazione

Consultando il repertorio LISA, si può ragionevolmente riportare l'entrata della *granularità* nell'ambito biblioteconomico e documentalistico al 1992 quando il termine appare nel titolo dell'articolo di Frans C. Heeman, *Granularity in structured documents* (Heeman 1992). Un'attenta (e paziente) analisi dei molti risultati di una query su Google consente però di risalire a un interessante *abstract* di un intervento che già nel 1993 si poneva l'obiettivo di una possibile storia della granularità dell'informazione. Fu Stephen E. Robertson, uno dei più qualificati ricercatori nel settore dell'*information retrieval*<sup>23</sup>, a presentarlo alla sessione di Storia dell'Information Science del Convegno annuale dell'American Society of Information Science.

«One particular theme, that of granularity of information, is followed through developments in information technologies (taken broadly), over the last four thousand years. Some classes of systems in the context of modern information technology are analyzed for their approach to information granularity. Different kinds of systems (e.g. wordprocessors, relational databases, text retrieval systems, knowledge-based systems) assume and use different levels of granularity» (Robertson 1993)<sup>24</sup>.

<sup>22</sup> Riprendo la descrizione della *text extraction* dalla monografia di Lancaster e Warner (Lancaster e Warner 2001, p. 44-46.) Per l'applicazione degli strumenti dell'intelligenza artificiale al settore dell'*information retrieval*, rimando per una breve introduzione a manuali di Heting Chu (Chu 2003, in particolare alle p. 229-236). L'evoluzione dell'analisi automatica del testo ha comunque dovuto fare i conti con la complessità del linguaggio naturale, che rende complessa e problematica la "liberazione" di dati e fatti dai vincoli dei supporti. Qualunque aspirazione in questo senso deve forzatamente tenere conto dei diversi livelli ai quali è necessario elaborare i testi, e cioè il livello fonetico, quello morfologico, lessicale, sintattico, semantico e pragmatico (Chowdhury 2004, p. 373).

<sup>23</sup> Stephen Robertson, tra gli altri, ha contribuito a sviluppare il modello probabilistico nel settore dell'*information retrieval*; cfr. al proposito *il manuale di information retrieval* di Heting Chu (Chu 2003, p. 106).

<sup>24</sup> Contattato personalmente per posta elettronica, il prof. Robertson mi ha confermato che non ha mai pubblicato il suo intervento e mi ha inviato gli appunti scritti con i quali aveva preparato l'intervento, indicandomi in particolare che l'oggetto della sua polemica fosse proprio il libro di Keith Devlin. Lo ringrazio per la sua squisita cortesia.

In questo intervento Robertson respingeva le teorie avanzate da Keith Devlin, che nel 1991 aveva ipotizzato l'esistenza logica di un'unità informazionale minima, battezzata *infon* (Devlin 1991). A parere di Robertson tale idea era fuori della realtà e della storia. Il tentativo di Devlin di fondare in senso logico-matematico una nuova scienza dell'informazione non poteva che scontrarsi con l'approccio di un informatico e di uno studioso dell'*information retrieval*, orientato a impostare ogni ricerca sulla concreta – e storicamente determinata – interazione uomo-macchina. Robertson sottolineava che l'informazione è «an human construct», un artefatto umano, nel quale non si possono individuare realtà atomiche reali come quelle che contraddistinguono il mondo della fisica. Tutte le unità che hanno consentito che l'informazione potesse essere elaborata (dai caratteri alfabetici al bit o al pixel), sono creazioni dell'uomo. Ogni tipo di sistema informatico in considerazione ha assunto come unità minime rappresentazioni diverse della realtà, o *different levels of granularity*. Le unità informative che si possono individuare nei documenti e che si possono manipolare grazie ai diversi software a disposizione non sono considerabili realtà fisiche o logiche preesistenti rispetto all'intervento e all'azione dell'uomo. Il quadro concettuale presentato da Robertson è chiarito dalle due tabelle che ho provato a ricostruire sulla base degli appunti gentilmente forniti dallo stesso autore.

Il tentativo di Devlin di identificare e teorizzare l'esistenza di una unità informazionale minima si può forse leggere come il prosieguo di un filone di ricerca quasi cinquantennale, iniziato dal grande successo della teoria di Shannon e Weaver<sup>25</sup>. L'ipotesi di un *infon* non mi appare molto distante da quella del *logon* di Donald Mackay, almeno nel tentativo di identificare un'unità minima dell'informazione come punto di raccordo tra i molteplici settori disciplinari che negli ultimi decenni hanno utilizzato il termine e il concetto di informazione (la biologia, la fisica, l'economia, ecc.). Mackay aveva parlato di «quantal aspects of information», altri recentemente hanno parlato di «quantum theory of information»<sup>26</sup>, con riferimento a una realtà ultima dell'informazione strettamente legata al concetto fisico di quanto di energia.

In realtà, diversi tentativi in questo senso non hanno avuto successo. Il bit appare l'unica unità di misura realmente utilizzata, prima dagli ingegneri e poi da noi tutti, come base di tutto il mondo digitale. Il successo della teoria di Shannon e Weaver era basato anche sulla deliberata rinuncia ad affrontare gli aspetti seman-

<sup>25</sup> Mi riferisco ovviamente al lavoro alla base della teoria della comunicazione (Shannon e Weaver 1949). Sulle origini e il successo di quella teoria, anche nello specifico delle scienze della documentazione, cfr. almeno il recente lavoro di Ronald Kline (Kline 2004).

<sup>26</sup> Per esempio H. Lire, come indicato nella già citata rassegna di Rafael Capurro e Birger Hjørland (Capurro e Hjørland 2003). Si può inoltre seguire sull'argomento la completa ricostruzione storica di Gernot Versig (Versig 1997).

ELEMENTI ATOMICI E TECNOLOGIE DELL'INFORMAZIONE E DELLA COMUNICAZIONE (Robertson, comunicazione personale)	
Tecnologie dell'informazione	Elementi informativi minimi
<i>Comunicazione (tecnica)</i>	
Linguaggio	Messaggio
Scrittura	Parola
Alphabet	Carattere
Telegraph	Bit
<i>Numero</i>	
Scrittura	Numero
Sistema arabo	Digit
Punched card	Bit
<i>Image</i>	
Fotografia	Immagine fissa
Film	Frame
Televisione	Line
Schermo del computer	Pixel <sup>27</sup> /bit
<i>Comunicazione (organizzativa)</i>	
Posta	Messaggio
Biblioteche	Libro
Editoria	Libro
Editoria	Rivista

ti (e pragmatici, verrebbe da aggiungere) della comunicazione, e sull'esclusivo trattamento degli aspetti tecnici. La ricerca relativa agli aspetti semantici e pragmatici ha imboccato altre strade, e almeno per quanto riguarda le scienze dell'informazione non ha più cercato "unità minime"<sup>28</sup>.

La riflessione presentata nel 1993 da un informatico come Robertson costituisce in qualche modo un punto fermo e sancisce la rottura tra la ormai acquisita specificità delle scienze dell'informazione (nella loro accezione più lata) e le sirene di una "metascienza" dell'informazione, che avrebbe come base le scienze fisiche<sup>29</sup>.

<sup>27</sup> «Pixel. Un pixel (abbreviazione inglese di picture element) è uno dei molti minuscoli puntini che compongono la rappresentazione di un'immagine nella memoria di un computer. Solitamente i punti sono così piccoli e così numerosi che se vengono stampati su carta o visti su un monitor appaiono fusi in una unica immagine. Colore e intensità di ciascun punto sono usati dal computer per rappresentare una piccola area dell'immagine» (<http://it.wikipedia.org/wiki/Pixel>).

<sup>28</sup> Forse l'ultimo tentativo in questo senso è stato quello di Anthony Debons, che ha proposto la definizione di realtà come *informs* e *knowgs* sulla base della possibilità di ciascun paragrafo di poter dare o informazioni (rispondere alle classiche domande: *when, where, who, what*) o conoscenza (*why* e *how*) (Debons 1992). Una proposta che non sembra avere avuto seguito.

ELEMENTI ATOMICI E STRUMENTI SOFTWARE (tabella costruita su dati forniti da Robertson, comunicazione personale, i testi in corsivo sono dello scrivente)	
<b>Elementi minimi dell'informazione trattati da strumenti software o strumenti di information retrieval</b>	
Word processing	L'unità minima è il carattere, a metà tra la parola e il pixel Altre unità minime possono comunque essere la parola, la frase e il paragrafo
Fogli elettronici di calcolo	Il numero e la formula di calcolo
Sistemi di gestione dei database	La tabella, il record, il campo, il sottocampo. Si assume che l'analista abbia definito i campi a un tale livello che il sistema non ha bisogno di accedere a un livello più basso <i>Rientrano in questo ambito i database catalografici, qualunque formato di meta dati essi usino.</i>
Text retrieval	Le parole, le frasi, le sentenze, i passaggi
Natural language processing	La struttura della "frase" e la parola e la frase
<i>Iper testo</i>	<i>Il chunk, che esprime un concetto, e può essere di ampiezza variabile; il link o nodo tra più concetti</i>
<i>Linguaggi di marcatura</i>	<i>Il paragrafo, la frase, la formula, il nome, un elemento, un attributo, un concetto</i>
<i>Video</i>	<i>Il frame, lo shot, la scena, il clip, il programma, l'episodio<sup>30</sup></i>

<sup>29</sup> Questo senza voler criticare chi nelle scienze fisiche prosegue la ricerca di una realtà informazionale minima, in particolare a partire dalla meccanica quantistica, e vede nell'informazione proprio una prospettiva in grado di unificare le diverse visioni della realtà. Cfr. al proposito il recente, divulgativo volume di Hans Christian Von Baeyer, *Information. The new language of science* (Baeyer 2004), che presenta l'ipotesi dell'esistenza del qbit.

<sup>30</sup> «A 'piece' of video is variously referred to as a frame, a shot, a scene, a clip, a program, or an episode. To provide anything more than linear navigation through video information, we need to structure video in some way; as has been discussed, this structuring needs to be done automatically. Video is made up of frames grouped together into *shots*, which are defined as the contiguous set of frames taken by a single uninterrupted camera over time. During a shot, the camera may move by zooming in or out, panning to the left or right, tracking, booming up or down, tilting, or by a combination of any of these motions. Shots are often grouped into logical or semantic units called *scenes*, which will have some interpretation in terms of the overall story being related in the program or episode that constitutes the complete video file. A clip of video is any unit; it may be as large as a set of multiple shots or scenes, or as small as a shot fragment» (Smeaton 2004).

In questo senso, l'intervento di Robertson conferma da un altro punto di vista quanto da tempo sostenuto da Alfredo Serrai e lucidamente confermato in un intervento pubblicato nel 2001:

«è necessario ripetere, con forza, come sia indispensabile in merito liberarsi di due pregiudizi assai comuni, e tuttavia entrambi esiziali: il primo consiste nell'erroneo valore che generalmente si attribuisce alla informazione, che viene considerata alla stregua di elemento o particella di conoscenza, nucleo di sapere, unità di giudizio discriminativo o almeno di perizia orientativa; il secondo nella caratterizzazione dell'era attuale e della nostra società quali era e società della informazione.

Niente di più fallace: l'informazione è unicamente la segnalazione di una differenza, e tale differenza risulta incisiva se, eventualmente, suscita e mette in azione altri effetti. Quel che conta sono le modifiche e gli esiti che l'informazione induce quando viene opportunamente ricevuta da un sistema predisposto.

Occorre sbarazzarsi dell'idea che l'informazione posseda una qualità intrinseca, che possa cioè venir considerata e valutata al di fuori del sistema in cui dovrà o potrà produrre i propri effetti informativi» (Serrai 2001, p. 43).

Non si può negare il fascino di alcune riflessioni interdisciplinari sul concetto di informazione che Serrai giustamente sottopone a critica. Un approccio al tema della granularità dell'informazione penso non possa fare a meno della consapevolezza dei temi affrontati in altre discipline e sulla base di approcci pure non da tutti condivisi. D'altra parte trovo condivisibile e in linea con l'atteggiamento di Serrai il richiamo di Rafael Capurro e Birger Hjørland che, nel ricostruire le origini greche e le radici latine del termine e nel segnalarne i diversi significati assunti negli ultimi decenni, hanno sottolineato la necessità di non focalizzarsi eccessivamente sul concetto di informazione e piuttosto di concentrarsi sui segni, i testi e, aggiungerei, i documenti (Capurro e Hjørland 2003).

L'intervento di Robertson è del 1993, agli albori del fenomeno Internet. Le realtà documentarie e gli strumenti software a disposizione non avevano ancora raggiunto il numero e la raffinatezza che conosciamo oggi e le risorse di Rete non avevano ancora assunto le dimensioni attuali. Gli sviluppi registrati nel settore dell'information retrieval, la creazione dell'ipertesto e del World Wide Web e soprattutto il trattamento sempre più analitico delle diverse entità informative possono a mio parere essere in gran parte inseriti all'interno delle griglie predisposte da Robertson. Nel frattempo, la metafora della granularità dell'informazione ha continuato ad agire e probabilmente ad aprire nuovi orizzonti a coloro che si occupano specificamente di testi, di documenti e di surrogati dei documenti.

## Granularità dell'ipertesto e architettura dell'informazione

Seguendo le indicazioni di George Landow, l'ipertesto è

«qualsiasi forma di testualità – parole, immagini, suoni – che si presenti in blocchi o lessie o unità di lettura collegati da link. Si tratta, essenzialmente, di una forma di testo che permette al lettore di abbracciare o di percorrere una grande quantità di informazione in modi scelti dal lettore stesso, e, nel contempo, in modi previsti dall'autore» (Landow 1998).

Questa nuova realtà testuale, realizzata e soprattutto sviluppata grazie all'uso delle nuove tecnologie elettroniche, consente al lettore una libertà molto più grande di quella concessa dal tradizionale documento cartaceo. Attraverso la rete dei legami tra i nodi concettuali dei testi e dei documenti, l'autore dell'ipertesto traccia molteplici percorsi di lettura. Il lettore può approfittare di questi percorsi e talvolta, seguendo le proprie intuizioni e associazioni mentali, crearne di nuovi, in maniera più attiva di quanto non succedesse nella tradizionale lettura lineare e sequenziale.

Nella visione di Landow, l'ipertesto nasce nella seconda metà del secolo scorso dal convergere delle correnti della teoria critica post-strutturalista e dal contemporaneo sviluppo tecnologico. Nel complesso percorso che va da Roland Barthes (in particolare, la sua analisi critica di *Sarrasine* di Balzac e la creazione del termine stesso *lessia*) a Michel Foucault, da Derrida e Deleuze e Guattari, Landow inserisce l'ingegnere Vannevar Bush e il suo futuristico *Memex* e Theodor Nelson, che ha propriamente coniato il termine *ipertesto*.

La frammentazione del testo in lessie e i collegamenti (link) tra queste secondo percorsi di lettura e interpretazione sempre nuovi sono gli elementi centrali di un ipertesto.

«Si potrebbe dire che l'ipertesto è una forma di testo composta da blocchi di 'scrittura' e immagini collegati da link, che permette una lettura multilineare: non una lettura non lineare o non sequenziale, ma una lettura multisequenziale» (Landow 1998).

Landow sostiene che i link avvengono tra «blocchi o lessie o unità di lettura», senza dare ulteriori specifiche sulle dimensioni fisiche di queste unità. Creare ipertesti significa disegnare forme di collegamento unidirezionali o bidirezionali tra lessia e lessia, tra stringa e stringa, tra stringa e lessia. È l'affermazione pieno del desiderio di "mobilizzazione" del testo e delle sue parti componenti, non più prestabilite in forme rigide dall'autore o dall'editore. Nel linguaggio di Landow, il testo appare "frammentato", "disperso" o "atomizzato" (Landow 1997, p. 96-97). In altri interventi relativi alla teoria o alla pratica dell'ipertesto – interventi di autori di formazione più

propriamente scientifica – si usano anche altri termini. Jeff Conklin, per esempio, ancora nel 1987, prima dell'affermazione di massa del Web, affermava che

«Most users of hypertext favor using nodes which express a single concept or idea, and are thus much smaller than traditional files. When nodes are used in this fashion, hypertext introduces an intermediate level of machine support between characters and files, a level which has the vaguely semantic aspect of being oriented to the expression of ideas. But this sizing is completely at the discretion of the hypertext writer, and the process of determining how to *modularize a document* [corsivo mio] into nodes is an art, because its impact on the reader is not well understood» (Conklin 1987, p. 35).

*Modularize a document*, si esprime Conklin, che prosegue con *modularization of ideas*. Credo si possa essere certi che siamo vicini vicini al concetto espresso con *granularity*, almeno in relazione alla realtà dei documenti ipertestuali e ipermediali.

«Hypertext invites the writer to modularize ideas into units such that (a) an individual idea can be referenced elsewhere (i.e. this is a scope of reference concern) and (b) alternative successors of a unit can be offered to the reader (e.g. more detail, an example, bibliographic references, or the logical successor). But the writer must reckon with the fact that a hypertext node, unlike a textual paragraph, is a strict unit which does not blend seamlessly with its neighbors» (Conklin 1987, p. 35).

Non si tratta di problemi sostanzialmente nuovi. Da secoli il testo è *modularizzato* in paragrafi, così da consentire sia una chiara esposizione di ciascun idea e di ciascun concetto, sia al fine di consentire una più agevole lettura. Il nodo ipertestuale costituisce però una realtà nuova, per la lettura, la ricerca e il recupero.

Geoffrey Nunberg è uno dei primi nel 1993 a utilizzare il termine *modularity* per descrivere alcune caratteristiche tipiche del nuovo documento elettronico, in cui viene messa in crisi la corrispondenza tra unità fisica e contenuto (libro, capitolo, articolo, ecc.).

«And finally, it is the individual user who determines the modularity of the document – that is, the particular chunk of text that it contains [corsivo mio].

All of this affects the way documents are used and reproduced. Let me begin with modularities. A mass-produced book is both bound and bounded in a way that's replicated for all its instances: each copy contains the same text in the same order. But the computational representations of texts can be divided and reassembled in an indefinitely large number of documents, with the final form left to the decision of the individual user» (Nunberg 1993).

La crisi dell'unità cartacea "libro" e la distinzione tra struttura fisica del volume e struttura logica del testo aprono al lettore innumerevoli possibilità di creazione e rielaborazione, a cominciare dallo specifico ritaglio di queste unità minime del testo per passare al numero dei link che collegano queste unità atomiche. Le dimensioni della lessia sono difficili da identificare in termini assoluti, forse è possibile valutare solo quale può essere la dimensione ottimale in grado di facilitare il compito del lettore. Per esempio, attraverso un'analisi del comportamento di alcuni lettori di ipertesti, Erping Zhu ha cercato di valutare se la leggibilità di un documento è influenzata dal numero di link ipertestuali e dalla loro granularità.

«In hypermedia literature, granularity is defined as the grain size or size of a node. A high or large granularity means a larger node or larger unit of information. A larger node indicates that a node is larger than a few sentences or one screen. A low or small granularity indicates that a node has few sentences and its length is less than one screen, normally explaining a single concepts or idea. This concepts was used in defining the size of a node in this study. Larger nodes in this study consisted of 29-61 lines, roghly 360 to 740 words. Smaller nodes ranged from 5-26 lines, 60 to 312 words» (Zhu 1999, p. 342).

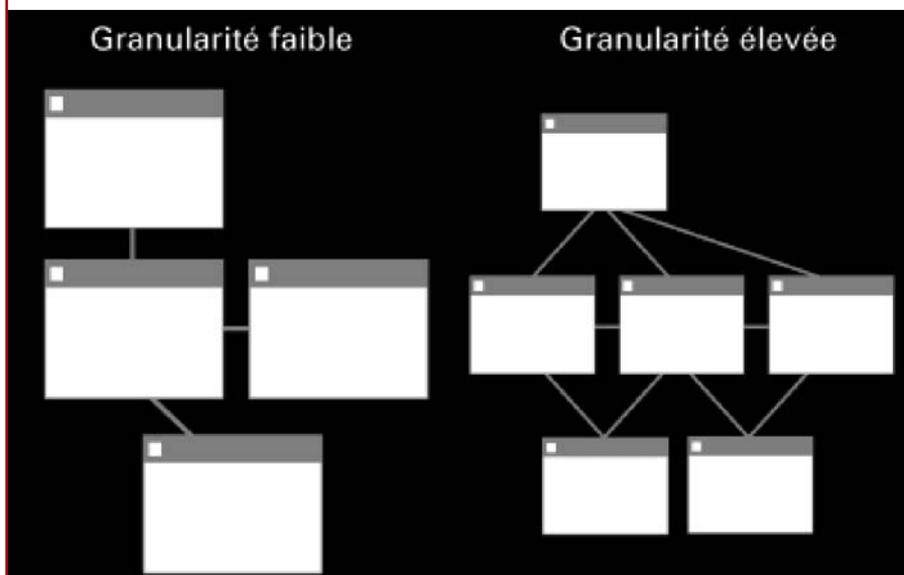
Gli studi di Zhu hanno dato un risultato sostanzialmente aperto, ma impongono la necessità di focalizzare la nostra attenzione sulle capacità e potenzialità cognitive dei lettori. Il problema della *granularità* diventa quello di rafforzare queste potenzialità e non di disorientare chi legge con troppi link. La frammentazione del testo e la moltiplicazione dei percorsi, quindi, non deve eccedere, oppure deve essere sostenuta da molteplici aiuti alla lettura e alla ricerca. In questo senso si esprime un giornalista italiano ben documentato sulla realtà del Web, Franco Carlini, che parla del pericolo di un «sovraccarico cognitivo» e di uno «smarrimento granulare» per un utente alle prese con troppi link, portati a unità o frammenti di testo troppo piccoli (Carlini 1999, p. 56-59).

Anche il *Groupe de recherche en arts médiatiques* dell'University de Quebec de Montreal segue un approccio simile, nella definizione che tenta del concetto di granularità. Nel *Dictionnaire des arts médiatiques*, stilato dal nel 1996, granularità è il

«Degré d'interactivité d'un programme, représenté par la quantité de points décisionnels qu'il comporte. Un grain est en ce sens l'unité fondamentale d'interaction dans un programme interactif»<sup>31</sup>.

<sup>31</sup> <http://www.comm.uqam.ca/~GRAM/C/term/tcm/tcmt67.html>.

Granularité faible e granularité élevée (Groupe de recherche en arts médiatiques)



La definizione del *grain* come *unità fondamentale d'interazione* ben evidenzia il ruolo cardine assunto dal lettore. È sulla base delle capacità cognitive di quest'ultimo che deve essere pensata l'organizzazione del sistema interattivo. Cercando di definire le caratteristiche che dovrebbe assumere l'unità minima di un ipertesto, Robert Horn ha stilato una vera e propria 'tabella dimensionale', che va dalla frase, al brano di testo arbitrariamente "ritagliato", al video a cui si affaccia il lettore<sup>32</sup>.

I problemi connessi alla granularità del link o del nodo di informazione si sono ovviamente ampliati da quando l'ipertesto è diventato strumento di uso quotidiano attraverso Internet. Le dimensioni delle unità informative minime (la pagina, il paragrafo, la videata del computer) sono studiate e valutate nei rapporti con il loro significato semantico, alla luce delle capacità cognitive del lettore e al fine di migliorare l'usabilità del sito web. Se le informazioni sono presentate in maniera corretta, semplice, facile da leggere e memorizzare, l'ipertesto arriverà meglio all'attenzione del lettore. È questa la strada percorsa per esempio da chi si occupa di quella che viene chiamata l'usabilità delle pagine web, cioè del «grado in cui un prodotto può essere usato da particolari utenti per raggiungere certi obiettivi con efficacia, efficienza e soddisfazione in uno specifico contesto d'uso» (UNI EN ISO 9241-11, p. 2).

L'esigenza di ben progettare un sito web così da renderlo maneggevole e funzionale è alla base della nascita di quella che viene talvolta considerata una giovane disciplina, la cosiddetta "architettura dell'informazione".

<sup>32</sup> Horn 1989, cit. <http://www.comm.uqam.ca/~GRAM/A/comm/032.html>.

Gli autori di un testo basilare, *Information architecture for the World Wide Web*, Louis Rosenfeld e Peter Morville, definiscono l'architettura dell'informazione come:

«La combinazione di organizzazione, categorizzazione e schemi di navigazione all'interno di un sistema di informazioni.

Il design strutturale di uno spazio informativo atto a facilitare l'esecuzione dei compiti e l'accesso intuitivo ai contenuti.

L'arte e la scienza della strutturazione e classificazione dei siti web e delle intranet per assistere gli utenti a trovare e gestire le informazioni» (Rosenfeld e Morville 2002).

Sono questi "architetti" a usare ripetutamente il concetto di granularità e, in alcuni casi, a tentare la definizione di alcuni parametri. Nel definirne le basi disciplinari, sempre Rosenfeld e Morville indicano come centrali la strutturazione, l'organizzazione e la catalogazione dell'informazione.

«La strutturazione implica la determinazione degli appropriati livelli di granularità degli 'atomi' di informazione del vostro sito e il decidere come correlarli. L'organizzazione riguarda il raggruppamento di quei componenti in categorie significative e distinte. Classificazione significa definire le categorie e la serie di collegamenti ipertestuali che condurranno ad esse. [...] La *granularità* si riferisce alla dimensione relativa o grezza dei singoli elementi informativi di base. I vari livelli di granularità possono includere il numero di un giornale, articoli, paragrafi e frasi» (Rosenfeld e Morville 2002, p. 5).

Non ci sono indicazioni più specifiche o più precise. La granularità rimane come un concetto – o una metafora – estremamente "elastico", che consente di evocare facilmente il problema della dimensione di questi "atomi", e soprattutto della loro gestione, della loro classificazione, dei link da creare, della strutturazione del sito. Evidentemente, più crescono il loro numero e la loro specificità, maggiori saranno l'impegno del sistema dal punto di vista informatico e lo sforzo di raccolta e classificazione attorno al loro nucleo concettuale.

### **Forme granulari del documento elettronico**

La creazione dell'ipertesto ha segnato un momento fondamentale di una più complessiva trasformazione del documento stesso. La sua nuova natura digitale, la possibilità di combinare realtà multimediali diverse, la distribuzione e l'accesso via rete hanno finito per mettere in discussione la natura originaria del documento.

Ancora oggi, il documento cartaceo rimane come modello di riferimento ma, una volta aperto il vaso di Pandora della tradizionale unità documentaria, sia il creatore

che il lettore possono pensare di poter scegliere o creare autonomamente il documento più consono ai propri scopi. Per quanto riguarda l'*information retrieval*, ne sono riuscite esaltate le possibilità dell'indicizzazione automatica e di reperimento di documenti e frammenti di documenti. Inoltre, lo sviluppo del World Wide Web ha portato alla creazione di uno "spazio" informativo assolutamente inimmaginabile solo pochi anni fa. Molti utenti pensano forse oggi alla Rete attraverso il filtro dei motori di ricerca, e sono abituati a raccogliere frammenti di siti, pagine, immagini, la cui realtà ha ormai ben poco a che fare con il tradizionale modello cartaceo. Molto probabilmente la realtà documentaria appare sempre più articolata e sfuggente, come pure sempre più indefinibile e tutto sommato ininfluente o inutile appare loro la distinzione tra documento e surrogato. Ciò che emerge è uno spazio di memoria informativa di durata indefinibile, in cui ogni documento, ogni testo, ogni immagine, ogni frammento, può essere oggetto di continue ridefinizioni, può mutare di significato, riaggregato in percorsi e contenitori sempre diversi. I motori di ricerca consentono – o obbligano – la ricerca di dati, parole, frasi, pagine, paragrafi, *chunk* o pezzi vari, nel formato più semplice e "nudo", senza alcuna mediazione indicale; le lunghe liste di risultati di ciascuna *query* possono essere ricollocate successivamente all'interno di un nuovo contesto, in un puzzle da riassemblare e ricomporre vorticosamente.

Non si fatica dunque a comprendere perché la metafora della granularità abbia continuato a operare e il termine a riaffiorare costantemente e sempre più ampiamente. Liberata da ogni vincolo fisico, che le teneva raccolte in un libro o in un altro documento cartaceo, queste ideali realtà informative atomiche che percorrono lo spazio informazionale – un po' onde, un po' granuli – possono essere riaggregate in modi sempre nuovi da parte dei navigatori della rete<sup>33</sup>. In questo tipo di realtà, la granularità appare sì una caratteristica fondamentale, ma risulta in qualche modo connotata in modo "orizzontale", quasi piatto. I documenti sono articolati in tante parti, ciascuna di esse facilmente recuperabile per le sue dimensioni, il suo formato, i suoi metadati, il testo o le caratteristiche multimediali. I legami ipertestuali possono legare questi documenti e le loro parti, ma non è obbligatorio, questi ultimi possono anche muoversi liberamente in rete, senza rapporti definiti tra di loro, incrementando lo "smarrimento granulare". Questa dimensione "orizzontale" non è forse sufficiente a connotare come granulare un documento, un surrogato o un'altra realtà informativa. Per approfondire questi aspetti, occorre provare forse a riprendere dalle basi il concetto stesso di documento elettronico.

Descrivere e concettualizzare le modificazioni subite dalle realtà documentarie è uno sforzo che dura oramai da almeno un paio di decenni.

<sup>33</sup> Ammesso che non siano proprio le realtà minime unitarie, i concetti, i *memi*, a costituire l'unica realtà e gli utenti della rete i loro strumenti di navigazione. Ma non vorrei seguire qui ulteriormente l'approccio à *la Matrix* derivato dal *gene egoista* di Dawkins e proposto, per esempio, da Susan Blackmore in *La macchina dei memi. Perché i geni non bastano* (Blackmore 2002).

La rigidità del documento cartaceo, forse più mito che realtà, ha lasciato il campo a realizzazioni molto più flessibili e articolate che consentono modi nuovi di elaborazione, scrittura, stampa, conservazione, pubblicazione, ricerca. La separazione concettuale tra aspetto grafico dei caratteri e della pagina (font, carattere, organizzazione della pagina, ecc.) e il suo contenuto è diventata per taluni divergenza tra il testo in sé e il testo come letto<sup>34</sup>, ma ha anche consentito di “portare” il testo su diverse piattaforme software, e di aprire nuove possibilità di interrogazione e di analisi a più livelli.

Il documento elettronico appare oggi suddiviso e nello stesso tempo integrato in parti componenti, ciascuna connotata da caratteristiche diverse (testo, suoni, immagini, video, ecc.). A fronte della perdita dei vincoli fisici, e quindi della tradizionale identità o fissità del documento, questa nuova organizzazione composita e i link ipertestuali lo hanno reso aperto all’accesso e all’indicizzazione.

L’unità documentaria, la stessa pagina web, può venire a comporsi di unità più piccole e dinamiche, elementi multimediali spesso linkati da altre realtà e composti in entità informative nuove. Centrale qui non è più il documento, ma quella che viene chiamata *granularity of browsing*, la capacità degli utenti di interagire non solo con la pagina web nella sua interezza, ma anche con unità più piccole, quasi delle ulteriori finestre, al cui interno possono essere richiamate risorse provenienti da altre fonti e dotate di caratteristiche diverse. Il loro accostamento finisce per produrre quella che qualcuno ha chiamato *recombinant information* o *metadocuments*, realtà nuove che sollecitano le capacità cognitive degli utenti. L’esperienza in questione è stata effettuata presso il Department of Computer Science dell’Università del Texas, i cui ricercatori hanno usato espressamente in questo contesto il termine di granularità.

«Granularity refers to the size of the fundamental units that browsers present. combinFormation shifts the granularity of browsing down from documents to smaller units. It makes *information elements* essential. In the current implementation, supported media elements consist of images and chunks of text. [...] users regularly need to deal with these finer grains. combinFormation translates hypermedia documents into *containers* of media elements and references to other containers. Together, the media elements, their containers, and hyperlinks comprise the set of hypermedia *information elements*, on which the program conducts its operations. Documents are automatically downloaded every few seconds; information elements stream into the visual space once per

<sup>34</sup> Cfr. in questo senso le riflessioni di Aarseth e il conio delle espressioni “textons” e “scriptons” per sottolineare come il testo elettronico si distingua in realtà differenziate: «A text is any object with the primary function to relay verbal information. Two observations follow from this verbal definition: (1) a text cannot operate independently of a material medium, and this influences its behaviour, and (2) a text is not equal to the information it transmits» (Aarseth 1997, p. 62).

second. Temporally, the granularity of browsing is extended from a single static view that settles and waits for the user to click a hyperlink, to a dynamic one that evolves continuously» (Kerne et al. 2003a)<sup>35</sup>.

Sempre più questi nuovi documenti sono prodotti dai browser o da veri e propri agenti informatici che navigano in rete alla ricerca dell'informazione, riportando agli esseri umani i risultati di questa ricerca, in forme sempre nuove.

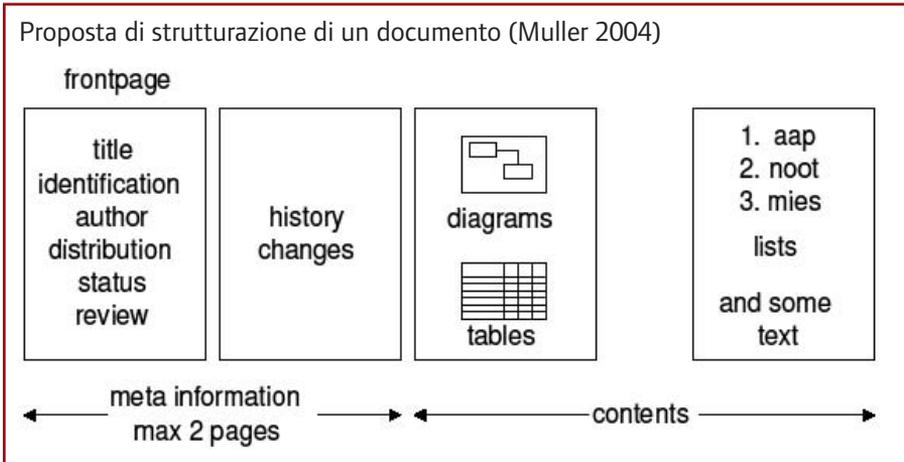


A fronte delle sperimentazioni più ardite, vale comunque la pena di sottolineare come sia riconosciuto da tutti coloro che hanno esperienza di documenti – digitali e cartacei – che le capacità di lettura dell'uomo sono esaltate da una ben articolata organizzazione del documento, preceduta da una ben studiata guida alla lettura. Libri e articoli cartacei si sono storicamente dotati di indici di diverso tipo e sommari, oppure sono stati ripartiti secondo precise strutturazioni. Le tecnologie digitali possono facilitare questo tipo di organizzazione consolidata, che può aiutare la lettura anche quando effettuata sullo schermo di un computer<sup>36</sup>.

<sup>35</sup> Cfr. anche Kerne, Khandelwal e Sundaram 2003.

<sup>36</sup> Sull'organizzazione di testi e documenti in grado di facilitare le capacità cognitive dei lettori, cito per esempio il manuale di Information design di Enrica Cavina e Rossella La Picciarella ispirato dalle metodologie di Graziella Tonfoni (Cavina e La Picciarella 2002). Tra i tanti studi sull'argomento, segnalo quello di Heather Silyn-Roberts, che attesta la maggiore leggibilità dei testi tecnici presentati in forma strutturata, con un abstract e una ripartizione in sezioni, ciascuna seguita da una preve presentazione (Silyn-Roberts 2002).

L'organizzazione strutturata del documento cartaceo viene anch'essa associata al termine di granularità. Il titolo di un documento facilmente recuperabile tramite un'interrogazione di Google è *Granularity of documentation*. L'autore, Gerrit Muller, enumera quali dovrebbero essere le caratteristiche principali e l'organizzazione di un buon documento, e presenta graficamente un modello di documento "granulare", organizzato in maniera per l'appunto da facilitarne la lettura e la gestione<sup>37</sup>.



La strutturazione "granulare" è ancora più essenziale quando l'obiettivo è quello di gestire un intero data base di documenti, di provenienza diversa, ma tutti orientati a un unico scopo. In questi casi, un accordo preventivo e una standardizzazione del documento dovrebbe facilitare la soluzione del problema. Qualcosa del genere è stato tentato da qualche anno a questa parte dall'*International conference on harmonization of technical requirements for registration of pharmaceuticals for human use*, che ha promosso la redazione di apposite linee guide per la registrazione di farmaceutici.

«The 'Granularity Document' of the M4 Organisation defines a series of standards for the physical construction of CTD (*Common technical document*) submissions. [...] The appendix defines the granularity of the documents, their pagination, the use of section numbers, the use of tab dividers (for the paper CTD) and what to include in the Tables of Content for each module (for the paper CTD). The granularity aspects describe what constitutes a document, i.e., divided by tabs and numbered separately in the paper submission and, in

<sup>37</sup> Gerrit Muller, *Granularity of documentation, draft*, 3 luglio 2006, <http://www.gaudisite.nl/DocumentationGranularitySlides.pdf>.

an eCTD, provided as separate files. It also defines at what level in the table of contents documents/files should be placed and where multiple documents/files may be provided for a particular section»<sup>38</sup>.

Nel documento è indicata con precisione la necessità di suddivisione gerarchica delle specifiche sezioni di questo *Common technical document*. La sua granularità è costantemente definita come l'organizzazione in parti strutturate gerarchicamente.

Organizzato e strutturato è anche l'articolo di interesse scientifico che appare sulle principali riviste internazionali. Essenziali per renderne più rapida la lettura e la circolazione tra gli studiosi di tutto il mondo sono il ridotto numero di pagine e la struttura formale. Quella che viene chiamata *topical structure* o "struttura canonica" è il prodotto di una lunga elaborazione storica.

«Il secolo XIX portò il cambiamento che condusse al principio organizzativo, oggi universalmente accettato, noto come *struttura canonica*:

- 1) Intestazione (titolo, lista e credenziali degli autori)
- 2) Riassunto
- 3) Introduzione
- 4) Corpo (metodi, teoria, esperimento, risultati)
- 5) Discussione
- 6) Conclusione
- 7) Ringraziamenti
- 8) Riferimenti bibliografici (bibliografia)» (Matriccioni 2003, p. 8)<sup>39</sup>.

Al di là delle critiche epistemologiche relative alla scarsa corrispondenza che correla questa organizzazione formale al reale procedimento di ricerca<sup>40</sup>, la struttura canonica ha il pregio di consentire ai lettori una lettura standardizzata e veloce, mirata ai "fatti", ai dati e alle caratteristiche delle esperienze di laboratorio. Veloce risulta anche la valutazione relativa all'affidabilità dei risultati e a un'eventuale possibile riproducibilità dell'esperimento stesso (Harmon 1989).

Recentemente alcuni lavori di un gruppo di ricerca olandese, tra i quali sono da citare quelli di Joost Kircs e Frédérique Harmsze, hanno avanzato la proposta di uti-

<sup>38</sup> *International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, ICH harmonised tripartite guideline, Organisation of the Common technical document for the registration of pharmaceuticals for human use, M4*, ultima correzione del 13 gennaio 2004, <http://www.ich.org/cache/compo/1325-272-1.html>.

<sup>39</sup> Sulla struttura canonica dell'articolo scientifico, cfr. soprattutto gli interventi di Joseph E. Harmon (Harmon 1989 e 1992).

<sup>40</sup> Mi limito a rimandare agli studi di Robert K. Merton, e in particolare tra le tante alle sue riflessioni sullo «standard scientific article e le obliterated scientific serendipities» (Merton e Barber 2002, p. 407-430).

lizzare in maniera più efficiente le possibilità ipertestuali, favorendo una maggiore modularità o granularità degli articoli scientifici. La proposta è basata sul riconoscimento che ogni articolo non nasce dal nulla, ma si inserisce all'interno di una tradizione scientifica, riutilizzando e richiamandosi a teorie, modelli, esperienze, dati e strumentazioni preesistenti e già attestati in letteratura. Le ipotesi e i dati che vengono pubblicati in ciascun articolo vanno a integrarsi in un panorama e in un contesto scientifico che in questo modo finisce per consolidarsi sempre di più, almeno fino a quando nuovi dati non falsificano le ipotesi precedenti e si assiste a una rottura del paradigma dominante. La ricerca e la lettura degli interventi più nuovi e originali sarebbero sostanzialmente migliorate qualora si desse la possibilità di recuperare facilmente solo ed esclusivamente quelle parti degli articoli specificamente dedicate ai metodi seguiti, o agli strumenti utilizzati, o alle ipotesi avanzate. In questo modo sarebbero evitate le ripetizioni delle ipotesi di partenza, ed evidenziati solo quei dati e quelle ipotesi che per la loro novità o originalità giustificano la "pubblicazione" in rete al cospetto del "collegio invisibile" dei ricercatori di tutto il mondo. Le potenzialità dell'ipertesto potrebbero in questo modo essere preziose, consentendo di articolare il testo dei documenti scientifici e consentendo una navigazione che faciliti il lettore a diversi livelli (Kircz 1998 e Harmsze 1999).

Sia che si vada verso un documento digitale in grado di sfruttare appieno le potenzialità ipertestuali e ipermediali, sia che ci si richiami alle tradizionali virtù del documento cartaceo strutturato in sezioni, paragrafi, ecc., è ormai ampiamente accettata la visione del documento elettronico come di una realtà in cui l'aspetto grafico e il contenuto sono separati, il testo è strutturato da linguaggi di marcatura e organizzato in parti componenti, correlate tra di loro all'interno di una vera e propria gerarchia. È stato l'influente paper di Steven DeRose, David Durand, Elli Mylonas e Allen Rehear che nel 1990 ha segnato l'abbandono dei precedenti modelli statici e lineari, rappresentati in tabella.

Modelli alternativi del documento secondo DeRose et al. 1990
Testo come bitmap: aspetto grafico
Testo come stream di caratteri: ogni carattere è codificato, l'unico livello di marcatura è costituito dallo spazio bianco e dal <i>return</i>
Testo come carattere e istruzioni per la stampa
Testo come layout di pagina
Testo come corrente di oggetti, in senso lineare
Testo come "ordered hierarchy of content objects" o "OHCO"

Non si può dire che questa concezione del documento abbia risolto tutti i problemi e sia stata accettata senza riserva, e nello stesso *paper* non erano ignorati gli aspetti problematici. Quello che preme sottolineare è che il modello OHCO ha

portato a conseguenze rilevanti per quanto riguarda i temi della granularità. Il documento digitale è dinamico e flessibile, accogliendo altre realtà non testuali (video, immagini, ecc.), e soprattutto non solo statiche, sebbene di non sempre facile gestione in fase per esempio di conservazione e recupero. I linguaggi di marcatura si sono affermati come base indispensabile per la portabilità dei testi su altre piattaforme, per marcare le unità logiche del testo (il paragrafo, per esempio) e per annotarle dal punto di vista semantico. Ogni paragrafo, ogni parola, ogni segno grafico può oggi essere annotato, a livelli di dettaglio estremamente analitici, così da rendere più potente la lettura e la ricerca, e individuare nel testo ulteriori livelli “logici” e concettuali. Queste realtà (testuali e non testuali) che compongono il documento possono essere indagate e indicizzate attraverso l’uso di sistemi di gestione della conoscenza (tesauri e ontologie) che possono sanare la tradizionale distinzione tra indicizzazione dei documenti (tradizionalmente effettuata grazie a classificazione e soggetti) e indicizzazione all’interno dei documenti (come si esprimeva negli indici analitici dei libri)<sup>41</sup>.

L’evoluzione dei linguaggi di marcatura assume un ruolo fondamentale nell’ottica dell’affermazione del documento granulare. A partire dalla creazione nel 1974 di SGML – linguaggio nato per consentire la lettura nel tempo dei numerosi testi manualistici prodotti dalle grandi aziende commerciali dell’area della ricerca tecnologica e militare<sup>42</sup> – il cosiddetto *markup* descrittivo assunse un notevole interesse anche per lo sviluppo del documento di interesse scientifico (Coombs, Renear e DeRose 1987). La complessità di SGML obbligò i creatori del Web alla creazione di un linguaggio di marcatura più semplice, HTML, dall’inaspettato, straordinario successo al momento della diffusione delle pagine web. Proprio la diffusione del Web ha fatto toccare con mano da una parte i possibili vantaggi dei documenti marcati in maniera logica, dall’altra il bassissimo valore semantico di quello specifico linguaggio. Presa coscienza dei limiti, la comunità internazionale ha creato e accettato un linguaggio più maturo, derivato da SGML, ma non così pesante e complesso, designato come XML.

XML è oggi utilizzato in applicazioni estremamente diverse. In un manuale di informatica molto diffuso<sup>43</sup>, si afferma la natura “ibrida” di XML, a cavallo tra il modello di rappresentazione dei dati e lo standard per la definizione di documenti. Questa doppia natura ha comunque contribuito ad assicurargli il ruolo di anello di congiunzione tra il Web (formato da documenti non strutturati) e le basi di dati (contenenti informazioni strutturate). Come HTML, XML è utilizzato per la rappre-

<sup>41</sup> A questo proposito, cfr. come XML sia oggi presentato anche agli indicatori di professione da un recente articolo di Bill Kasdorf come lo strumento per sanare la storica distinzione tra *open-system indexing* (quella appunto operata su monografie e articoli di rivista attraverso vocabolari controllati), *closed-system indexing* (l’indicizzazione analitica tipica delle monografie e presente al termine del testo) e *word lists* (la ricerca automatica della parola nel testo) (Kasdorf 2004).

<sup>42</sup> Su SGML cfr. almeno il lavoro introduttivo di Joan M. Smith (Smith 1992).

<sup>43</sup> Riprendo qui e nel paragrafo successivo le indicazioni del manuale a più mani sulle basi dati, il cui primo curatore è Paolo Atzeni (Atzeni 2003, p. 222-224).

sentazione di documenti, sia per il web che tradizionali, e per la costruzione di motori di ricerca. L'uso di questo linguaggio consente una migliore e più facile elaborazione di dati presenti in quei contenitori "semi-strutturati" che sarebbero le pagine in formato XML. Una delle caratteristiche essenziali e innovative di XML è poi quella di garantire, molto più di HTML, una maggiore attenzione all'aspetto semantico dell'informazione e, tra le altre cose, una sempre maggiore granularità. Secondo Terrence Brooks,

«The revolutionary aspect of XML is the *modularization* [corsivo mio] of information. Information presents itself as a self-describing unit that can does not inhibit processing, storing or display. Topical subject qualifiers (e.g., attributes) are placed at the appropriate level of granularity. [...] Meaning resides in the semantic structure of information. Meaning can also reside in the meaning-bearing terms and phrases placed at the appropriate level of granularity that serve to qualify a specific element of information» (Brooks 2001).

In questa prospettiva, grazie ai linguaggi di marcatura, l'informazione è sminuzzabile fino a dimensioni realmente minime. A ogni livello, dal più elevato (l'intero documento) al minore (il paragrafo, la frase, la parola, un qualunque "segno") è possibile riconoscere e assegnare un valore semantico, anche senza la necessità di una intermediazione umana, a un dettaglio al quale non era mai stato possibile arrivare. Un esempio offerto dal testo *XML companion* di Neil Bradley<sup>44</sup> consente di illustrare come si possa marcare un documento XML secondo diversi livelli di granularità. Si può dire che un documento possiede una granularità fine (*fine*) se è possibile isolare e identificare ogni possibile unità di informazione. Per esempio, il nome dell'autore viene suddiviso nel massimo numero di sottounità.

```
<name>
  <salutation>Dr.</salutation>
  <initial>J</initial>
  <last-name>Smith<last-name>
</name>
```

Quando invece un documento ha un numero di elementi inferiore o insufficiente, si dice che possiede una granularità grossolana (*coarse*):

```
<name>Dr. J. Smith</name>
```

<sup>44</sup> Riporto la definizione *granularity*: «The degree to which an element is divided into child elements. A complex hierarchy denotes a 'fine' granularity. Simple structures with few levels indicate a 'coarse' granularity» (Bradley 2001, <http://proquest.safaribooksonline.com/0201770598/d1e91826>).

Più fine o maggiore è la granularità, più largo e pesante il documento, e più costoso quando è marcato manualmente. D'altra parte, il documento più ricco è anche quello potenzialmente suscettibile di un'analisi e di un'elaborazione più approfondita, con un recupero (un'"estrazione") più avanzato dei dati presenti<sup>45</sup>.

Un testo marcato con XML consente di individuare facilmente le già indicate unità minime del pensiero (paragrafi, concetti, citazioni, formule chimiche, ecc.) e rende direttamente disponibili i "dati" di prima mano della ricerca. XML è oggi uno strumento indispensabile non solo per la gestione di dati in database strutturati, ma anche per quanto riguarda i documenti scientifici, giuridici e letterari. La codifica e marcatura del testo è effettuata anche in realtà complesse come quella biblica, che pure ha richiesto soluzioni articolate. Grazie a schemi di marcatura come il TEI (Text Encoding Initiative, basato originariamente sulla sintassi SGML e poi allineato alla sintassi XML) è possibile rappresentare la struttura astratta di varie tipologie testuali (testo in prosa, testo poetico, testo drammaturgico, fonte manoscritta, etc.), e le caratteristiche testuali rilevanti per diverse aree di ricerca (filologia, analisi linguistica, tematica, narratologica, etc.) (DeRose 1999)<sup>46</sup>.

In ambito scientifico, una marcatura molto attenta e granulare del testo può raggiungere non solo il paragrafo o la citazione, ma anche i dati presenti nei grafici e nelle tabelle, le formule chimiche e quelle matematiche. Negli ultimi anni sono nati "dialetti" XML pensati specificamente per marcare i documenti scientifici, così da consentire un più facile recupero con interrogazione mirate di dati e concetti<sup>47</sup>. Il settore scientifico forse più avanzato è quello della ricerca chimica, in particolare grazie al lavoro di Peter Murray Rust, che ha proposto uno specifico linguaggio (CML). Secondo Peter Murray-Rust e Henry S. Rzepa, grazie all'utilizzazione dei linguaggi di marcatura si potrebbe assistere alla nascita di un nuovo tipo di documento scientifico, il "datumento", realtà strutturalmente e funzionalmente progettata per "salvare" e pubblicare i dati scientifici, prodotti in maniera massiccia dai laboratori di tutto il mondo.

<sup>45</sup> Usa l'esempio della distinzione del nome e cognome di un autore in campi o sottocampi diversi anche il dizionario online ODLIS (*Online Dictionary for Library and Information Science*, a cura di Joan M. Reitz): «Granularity. The level of descriptive detail in a record created to represent a document or information resource for the purpose of retrieval, for example, whether the record structure in a bibliographic database allows the author's name to be parsed into given name and surname» ([http://lu.com/odlis/odlis\\_g.cfm](http://lu.com/odlis/odlis_g.cfm)).

<sup>46</sup> Non mi soffermo in questa sede sui problemi aperti. Terye Hillesund, per esempio, paventa una vera e propria "single source ideology", e cioè la presunzione che contenuto e presentazione possano essere trattati in maniera completamente indipendente, e che ogni testo possa essere offerto su qualunque supporto e in qualunque occasione (Hillesund 2005). Un'altra critica è quella avanzata, per esempio, da Paul Caton che sottolinea come ogni testo realizzi un originale atto comunicativo, che i linguaggi di marcatura tendono a ignorare, offrendo una visione statica del testo e del documento (Caton 2001).

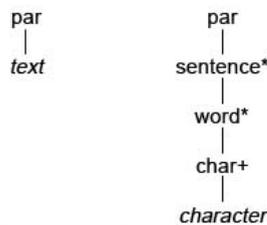
<sup>47</sup> Un elenco dei "dialetti" XML di marcatura dei testi scientifici, ciascuno creato secondo le esigenze della comunità scientifica di riferimento, è offerto da Zhongyu Lu 2005.

«Datument: a combination of data and document created using formal markup to allow the processing of individual data components» (Murray Rust 2003).

La marcatura dei testi è comunque operazione estremamente onerosa qualora effettuata da operatori umani<sup>48</sup>. I vantaggi che si possono prevedere da una marcatura dei testi, e pure da una loro organizzazione granulare e funzionale (come ipotizzata da Kircz e Harmsze) necessitano di un investimento massiccio, dal ritorno non garantito, e comunque da spalmare su tempi lunghi se non lunghissimi. Questo è il problema a cui si cerca di ovviare con l'elaborazione automatica del linguaggio naturale (*natural language processing*) e che ha avuto sviluppi di rilievo a livello scientifico e commerciale.

Per tornare al tema della granularità, credo non sia inutile rileggere il primo articolo nel cui titolo appare il termine *granularity*, e cioè il saggio di Frans C. Heeman pubblicato nel 1992, *Granularity in structured documents*. Heeman metteva a confronto tre diversi linguaggi di marcatura dell'epoca (Grif, ODA e SGML) e due diversi approcci alla strutturazione di un documento, l'approccio top-down e l'approccio bottom-up. Tutti e tre i linguaggi ed entrambi gli approcci non riuscivano secondo l'autore a eliminare tutte le *grey area* che venivano a crearsi nel momento in cui la marcatura e la strutturazione logica dovevano confrontarsi con il contenuto stesso del documento. Ogni paragrafo di testo, ogni formula matematica poteva essere marcata in maniera differenziata, secondo visioni diverse ed egualmente ammissibili. In altre parole, il documento – sia dal punto di vista logico, che dal punto di vista del contenuto – offriva per la complessità di alcuni suoi contenuti (tabelle e formule matematiche, per esempio) e per la complessità dei rapporti tra struttura e contenuto, soluzioni molteplici di organizzazione e marcatura.

Rapporti tra struttura e contenuto del documento: il problema della granularità (Heeman 1992)



*Figure 5. The granularity problem has its effects on content types. At the left, the logical structure stops at the level of a paragraph, and a content type 'text' takes over. At the right, the logical structure continues down to the level of a single character, which is defined as a content type*

<sup>48</sup> L'alternativa è costituita dall'estrazione automatizzata di nomi, formule, ecc., oppure la marcatura in forma automatica. Cfr. su questi temi anche l'intervento di Jon Bosak sui costi della marcatura dei testi scientifici (Bosak 2001): «The principle here is quite simple: the more information we add to a document, the better use we can make of it. But the information has to come from us; it is not going to come from the computer».

Secondo Heeman non può esistere un'unica partizione logica di un documento e il rapporto tra struttura logica e contenuto può essere risolto ragionevolmente secondo modalità diverse, tutte ugualmente compatibili.

«There is no clear distinction between structure and contents. The granularity, the level at which the structure part stops and the content part takes over, differs among systems. On some occasions, the primary structure is (mis)used to express the logical structure of a content type, on other occasions, logical structure is enclosed within contents» (Heeman 1992, p. 153).

Heeman riscontrava inoltre un difficile trattamento di tavole, formule chimiche e matematiche, suoni, ecc, auspicando soluzioni uniformi e integrate per gestire strutture e contenuti così eterogenei rispetto al semplice testo. Le riflessioni di Heeman appaiono interessanti ancora oggi, sia per la messa in guardia verso un eccessivo entusiasmo nell'uso dei linguaggi di marcatura, sia per l'uso del termine *granularità*, del resto ben poco definito. L'ipotesi che oggi XML lascia intravedere – navigare tra i diversi livelli di dettaglio, quasi zoomando in maniera fotografica – non può che essere considerata un ideale punto di arrivo, considerato il difficile, mai del tutto risolto rapporto tra struttura logica e contenuto del documento stesso.

La manipolazione in formato elettronico del testo e la sua strutturazione logica hanno comunque aperto un'approfondita riflessione sulla possibilità di organizzare in forma strutturata e gerarchica ogni tipo di documento e sugli effetti che tale strutturazione poteva avere ai fini della leggibilità e della ricerca nel testo stesso. L'uso di XML è alla base dell'ambizioso progetto ideato da Tim Berners-Lee per la creazione di un Web semantico in grado di superare i limiti del Web attuale e consentire un recupero dell'informazione maggiormente pertinente agli interessi degli esseri umani attraverso l'uso degli strumenti automatizzati. Riprenderò in seguito questo argomento.

La gestione e il recupero di informazioni da documenti non più costituiti da una sequenza di testo, ma organizzati in una struttura gerarchica in cui sono inserite parti componenti, ciascuna con una propria autonomia, apre inedite possibilità di lettura e di recupero. L'elaborazione automatica di suoni e immagini (fisse e in movimento) è oggi settore di ricerca sempre più sviluppato e che già offre possibilità di ricerca tramite strumenti di *content-based retrieval*, strumenti cioè che recuperano attributi e caratteristiche direttamente dal contenuto dell'immagine o del suono, piuttosto che da parole chiave derivanti da indicizzazione umana (Raieli e Innocenti 2004). L'obiettivo più ambizioso è quello di integrare non solo indicizzazione umana e indicizzazione automatica, ma anche le diverse abilità cognitive degli utenti e le potenzialità delle macchine automatiche. Attualmente, i documenti ipertestuali e ipermediali favoriscono soprattutto l'attività umana del *browsing*, mentre i testi facilitano l'attività del *searching*. Queste due attività cognitive

debbono collaborare in maniera più armonica di quanto non facciano adesso. Utilizzando l'esempio fornito da Yves Chiamarella,

«This integrated view of a hypermedia system allows users to issue queries combining content and structure as parts of the needed information. For example, one can issue a query aimed to retrieve all 'medical reports' that are about 'lung cancer' and which are 'illustrated by images of a given type' (X-ray etc.). In this query 'lung cancer' is the topic assigned to retrieve medical reports (i.e. textual information) while 'X-ray images' refers to a particular type of images. The expression 'illustrated by' refers to links that relate both kinds of data in the hyperbase. As a consequence the answers to such a query will be structured objects extracted from the hyperbase, the structure and the content of which having to match the requirements stated by the query» (Chiamarella 2000).

Modello di iperdocumento (Chiamarella 2000)

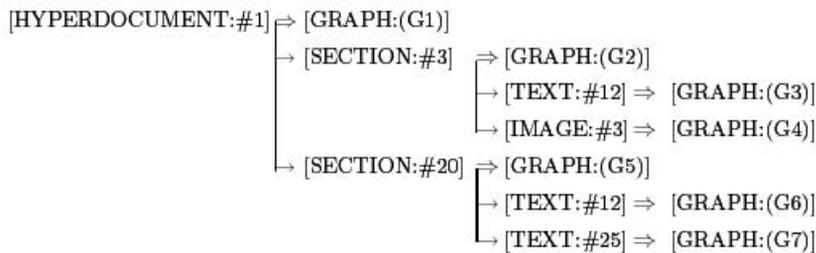


Figure 4. An Example of Hyperdocument.

Il riconoscimento e l'apprezzamento della segmentazione del documento strutturato in piccole unità di base (*granularity choices*, secondo Yves Chiamarella), inserite all'interno di una struttura gerarchica, è ovviamente essenziale per un corretto funzionamento di questo ancora ipotetico strumento di ricerca.

Leggere, sfogliare e cercare un testo sfruttandone appieno le caratteristiche ipertestuali, multimediali di documento strutturato è dunque problema ancora non risolto pienamente, sia per quanto riguarda le capacità degli esseri umani che per quelle delle macchine.

Un altro modo di rafforzare la cognizione umana attraverso lo sfruttamento delle molteplici possibilità di manipolazione offerte dal documento elettronico gerarchizzato, marcato e ipertestuale è offerto da diverse sperimentazioni verso una visualizzazione flessibile. Vale la pena citare come esempio un prototipo che può lavorare a diversi livelli di granularità, nella cui presentazione è esplicitamente usata la metafora dello zoom.

Questo strumento:

«allows readers to view an electronic text in multiple simultaneous views, providing insight at several different levels of granularity, including a reading view. This prospect display is combined with a number of tools for manipulating the text, for example by highlighting sections of interest for a particular task. [...] This project extends research in hypertext and kinetic text theory to provide readers with a text document display that combines simultaneous prospect – an overview of the entire text – and detail views, with related tools» (Ruecker, Homich e Sinclair 2005).

Ma se la visualizzazione di un documento tutto sommato statico può essere favorita da uno strumento di questo genere, la natura composita del documento ipertestuale e ipermediale suggerisce anche la nascita di realtà documentarie nuove e dinamiche. Un documento può essere o apparire granulare non solo perché marcato con uno specifico linguaggio, o perché organizzato in modo tale da consentire una più agevole lettura, ma anche perché può consentire un'agevole riutilizzazione dei suoi contenuti e delle sue parti componenti per creare nuovi documenti. La segmentazione è in questi casi cercata per consentire non tanto e non solo una trasmissione di contenuti, bensì un vero e proprio processo di apprendimento. Nel settore dell'e-learning, in particolare, i termini della famiglia *granularità* sono utilizzati in maniera massiccia.

«Granularité: Niveau de découpage d'un contenu pédagogique en une série d'items élémentaires, appelés grains, que l'on peut re-combiner dans le déroulement de parcours pour répondre aux besoins individuels de formation. [...] Le granule ou le grain est un objet pédagogique. Les objets pédagogiques microscopiques, unités élémentaires d'apprentissage sont associés pour constituer les parcours individuels de formation»<sup>49</sup>.

La partizione granulare di un corso di e-learning è finalizzata all'obiettivo di consentire una formazione del tutto autonoma del discente, che possa ritagliarsi dei moduli formativi su misura. Lester Gilbert, dell'Università di Southampton, richiama comunque una certa dose di cautela. I granuli o i chunk di formazione non devono essere connotati semplicemente da una identica dimensione fisica, bensì debbono correttamente essere definiti all'interno di una vera e propria *learning transaction*, che coinvolge docenti e discenti in un processo comunicativo in più fasi (Gilbert parla delle seguenti fasi: *tell, show, ask, response, feedback*), che non possono essere così nettamente segmentate (Gilbert, Sim, Wang 2005).

<sup>49</sup> <http://www.educnet.education.fr/dossier/eformation/modularite3.htm>.

Per offrire indicazioni omogenee al fine della creazione di quelli che vengono chiamati *learning object*, è stato elaborato dall'associazione professionale degli ingegneri elettronici americani lo standard IEEE 1484.12.1-2002, *IEEE Standard for learning object metadata*. La definizione di *learning object* rimane abbastanza elastica («any entity, digital or non-digital, that may be used for learning, education, or training»), ma è stata definita con maggiore precisione quale possano i diversi livelli di aggregazione e granularità, anzi di granularità funzionale. Li presento nella tabella successiva, riportando a fianco di ciascun livello l'esempio tratto dallo stesso standard.

Granularità funzionale (IEEE 1484.12.1-2002)	
Aggregation Level	
The functional granularity of this learning object	
1: the smallest level of aggregation, e.g. raw media data or fragments	Una fotografia digitale di Monna Lisa
2: a collection of level 1 learning objects, e.g., a lesson	Una lezione con la fotografia digitale di Monna Lisa
3: a collection of level 2 learning objects, e.g., a course.	Un corso o una collezione di lezioni su Monna Lisa
4: the largest level of granularity, e.g., a set of courses that lead to a Certificate	Un set di corsi comprensivi dell'intera storia, della descrizione, dell'interpretazione su Monna Lisa

Al di là della possibile ulteriore e pragmatica definizione della durata e della composizione di ciascun livello<sup>50</sup>, mi interessa sottolineare come ancora una volta il termine “granularità” venga utilizzato per definire una realtà strutturata e gerarchizzata. Il discente o il docente possono così meglio muoversi tra i diversi livelli di dettaglio, senza perdere i riferimenti al contesto. Si tratta di una visione “verticale” del concetto di granularità, che integra a mio parere in maniera importante il senso “orizzontale” che emerge dalla letteratura sull'ipertesto.

### Descrivere e identificare i documenti e le loro parti

L'uso del termine granularità costituisce come detto un interessante segnale delle novità emerse da quando la comunità internazionale dei bibliotecari ha cominciato

<sup>50</sup> In alcuni casi è stato tentato di definire con ancora maggiore precisione la consistenza di queste unità. Una proposta pragmatica in merito è quella reperibile in rete di Pete Richardson, che parla di 1 ora per il livello 2, 30 ore per il livello 3, un anno di corso per il livello 4. Pete Richardson, *Learning Objects & Granularity - Does Size Matter? (A Pragmatic Approach to LO design)*.

a gestire, descrivere e rendere accessibili le nuove risorse elettroniche. Ho già avuto occasione di riportare come Stefano Gambari e Mauro Guerrini abbiano definito la granularità, ma anche altri bibliotecari hanno utilizzato il termine proprio discutendo i problemi relativi all'identificazione, descrizione e recupero dei documenti elettronici.

Il progetto olandese che nel 1997-1998 prese il nome di DONOR (Directory of Netherlands Online Resources) è stato una delle iniziative che, a livello nazionale, sono state intraprese nel decennio scorso per censire e catalogare le risorse di qualità presenti in Rete e offrire agli utenti percorsi disciplinari guidati (Van der Werf-Davelaar 1999). In un rapporto stilato in quell'occasione, il concetto di granularità è oggetto di un'analisi completa.

«The concept of granularity refers to the fact that in the Web environment, where information entities are hyperlinked, complex hierarchical and/or associative information structures exist. A user may want to identify a network service, a document on that service, a chapter in a document, a collection of images, an individual image, a video or an individual scene or even one frame in that video».

Nell'ambiente di Rete, i confini fisici del documento tradizionale si dileguano e i link, i legami tra i file, o gli oggetti, definiscono realtà documentarie nuove, correlate logicamente ma di difficile gestione da parte dell'utente. Questo ha la possibilità di cercare e trovare le realtà più minute (l'immagine, il video, la scena o il singolo frame di un video) ma corre il rischio di perdere ogni cognizione relativa ai legami – talora molteplici – che legano le diverse entità informative e che danno vita non più ai singoli documenti fisici ma a veri e propri "documenti logici", di lettura e interpretazione non sempre immediata. È quindi necessario identificare e censire non solo le minute realtà e entità, ma soprattutto dare conto dei rapporti che li legano.

«Several 'formal entities' (e.g. files) may be part of the same logical document, e.g. an 'entity of information'. It is possible for a 'formal entity' to be part of more than one 'logical document'. Granularity has to be taken into account when:

you want to assign unique identifiers to information entities

you want to create metadata for information entities

you want to index information and reflect relations between information entities in the presentation of search results.

Retrieval of Web information can be improved when it is made clear of which formal entities a logical document consists. The provider of the information should indicate the relation of an entity with other entities. Indexing services

should be able to link to the logical document when the keyword is found in part of this document in a search» (Peereboom 1998).

Il rapporto elaborato in occasione del progetto DONOR si concentrava sul terzo aspetto, quello della definizione delle relazioni tra le diverse entità documentarie. In via sperimentale, si suggeriva di compilare il campo *Relation* dei metadati Dublin Core dopo avere individuato con precisione le relazioni tra i diversi documenti<sup>51</sup>. Si proponeva conseguentemente di sperimentare quelle soluzioni in grado di consentire all'utente, nel momento in cui riceveva una risposta alle proprie query, di apprezzare proprio i rapporti eventualmente esistenti tra i file diversi reperiti in fase di ricerca.

Le problematiche discusse in occasione del progetto DONOR sono state successivamente riprese e discusse in diverse sedi. Tra queste, una delle più autorevoli è stata quella della *Bicentennial Conference on bibliographic Control in the New Millennium*, tenutasi alla Library of Congress nel novembre 2000. Tra i diversi interventi, quello di Glenn Patton – Manager of Cataloging Products Department presso OCLC – affrontava il tema della granularità in fase di catalogazione. In pratica, all'interno di siti articolati e complessi come spesso si trovano in rete, quale doveva essere il livello di dettaglio a cui doveva scendere un catalogatore? Il titolo dell'intervento di Patton risultava lievemente provocatorio: *Granularity. Are we seeing a return to the 19th-century mixed catalog?* Il riferimento era alla tradizione del XIX secolo quando – prima che fossero disponibili i servizi di indicizzazione degli articoli delle riviste – si procedeva a una indicizzazione degli articoli delle riviste, dei saggi nelle collezioni o nelle collane, dei capitoli nelle opere di grandi dimensioni. Lo spoglio dei periodici e dei seriali costituiva infatti, e costituisce ancora oggi, un paragone preciso a cui richiamarsi quando si discute su quale debba essere oggi l'entità documentaria minima da descrivere.

Nel tentare una prima, concreta risposta alle sollecitazioni degli intervenuti, Patton ricorreva a un'analisi del lavoro dei catalogatori che partecipavano al progetto CORC – progetto lanciato da OCLC per la catalogazione partecipata delle risorse elettroniche remote. In quell'occasione non erano state date indicazioni cogenti ed era stata lasciata una certa libertà ai catalogatori

«At what level of granularity are CORC participants creating records? [...] Do you describe only at the 'site' level or at a level below that – a subsite that forms some kind of logical unit – or at the individual item level, be that a paper or article, an image, or some other kind of resource?» (Patton 2000).

<sup>51</sup> Cfr. la dettagliata formalizzazione dei rapporti tra documenti nel campo *Relation* dei metadati Dublin Core in Hillmann 2005.

Se il 60% delle risorse descritte appariva simile a un sito nella sua interezza, il 33% rappresentava qualcosa che era una parte di un insieme più grande e il rimanente 7% ricadeva in una sorta di area grigia di difficile definizione. A fronte della sfida quotidiana nel decidere cosa e fino a che punto catalogare, Patton auspicava la redazione di una guida – un codice si direbbe – alle opzioni relative alla scelta della granularità dei documenti da catalogare. Soluzione sicuramente auspicabile, soprattutto per chi deve gestire un archivio di record bibliografici di enormi dimensioni, frutto della catalogazione partecipata di migliaia di catalogatori, ma che non appariva – e non appare neanche oggi – di immediata realizzazione<sup>52</sup>.

Un taglio più teorico ha assunto la discussione tenutasi al *First IFLA Meeting of Experts on an International Cataloguing Code* del 2003 a Francoforte. L'intervento di Gunilla Jonsson affrontava proprio questi temi:

- «1. What is a bibliographic unit?
2. Is there always a one to one relationship between a bibliographic record and a bibliographic unit?
3. Do we need descriptions at a deeper level of granularity for the digital environment?» (Jonsson 2003).

Riprendendo concetti e terminologia proposte dal rapporto FRBR<sup>53</sup>, il problema di maggior rilievo appare quello della moltiplicazione degli item relativi a ciascuna manifestazione di ciascuna opera, fenomeno a cui si assiste nell'ambiente digitale e che non è affatto semplice tenere sotto controllo. Creare un record bibliografico per ciascun item o parte di item potrebbe essere un'operazione realizzabile in forme automatizzate, magari utilizzando la strutturazione dei documenti elettronici secondo i linguaggi di marcatura. La soluzione non risolve il problema di creare e mantenere un corretto sistema di correlazioni tra i record degli archivi di autorità, quelli bibliografici e quelli relativi ai singoli item.

<sup>52</sup> Cito dal sito della *Digital library for Earth system education* (<http://www.dlese.org/Metadata/collections/granularity.htm>) alcune sintetiche indicazioni per i catalogatori: «Granularity of metadata records.

Generally create a single metadata record for resources as a whole  
 Create additional metadata records for a whole resources if individual parts of a resource differ substantially in technical requirements, descriptions and educational information  
 Create relationships between resources using the *fields of relation* and occasionally, *learning resources type* and *description* [...]

Apply the relationship concept above to strike a balance between supporting resource discovery with reasonable user effort and to avoid potential user frustration with retrieving too many redundant records.

Recognize that community developed collections may have varying levels of granularity».

<sup>53</sup> Su FRBR, cfr. almeno come riferimento di partenza il testo introduttivo curato da Carlo Ghilli e Mauro Guerrini (Ghilli e Guerrini 2001).

«The bibliographic unit in the digital context is not different from the bibliographic unit in the analog environment. It comes in different constellations and it has different physical properties, and we must build a new infrastructure to be able to obtain items from the digital stacks, the archive. There may be possibilities to produce bibliographic records for a deeper level of granularity than what we achieve for publications on paper, due to economical constraints. It will, however, require that text be structured according to specified standards to allow automatic extraction, or generation, of metadata» (Jonsson 2003).

Questo tipo di riflessione rapporta in maniera diretta la discussione sul concetto di granularità ai temi sui quali negli ultimi anni la comunità internazionale dei bibliotecari si è a lungo confrontata. Mi riferisco non solo al problema della gestione e delle caratteristiche delle risorse elettroniche, ma in senso più ampio a quello della configurazione di tutto l'universo delle relazioni bibliografiche. Il tema è stato affrontato in seguito all'ideazione e alla pubblicazione del modello concettuale conosciuto come FRBR, ma aveva già trovato importanti anticipazioni nelle riflessioni di Alfredo Serrai sull'ontologia catalografica (Serrai 2001a) e nel lavoro con cui Barbara Tillett aveva compiutamente identificato le relazioni bibliografiche (Tillett 1991). Tra queste ultime, si riconnette a mio parere direttamente alla discussione sulla granularità la relazione tutto-parte che sussiste tra

«una parte componente di un oggetto bibliografico o un'opera e la sua totalità, come tra un racconto breve e l'antologia nel quale è contenuto. Le componenti possono essere parti di alcune particolari manifestazioni di un'opera, cioè parti di un oggetto bibliografico, o possono essere parti di un'opera astratta»<sup>54</sup>.

Il punto di arrivo potrebbe essere quello dell'identificazione e della descrizione di tutte le parti componenti che è possibile isolare in un documento, registrando con opportuni legami relazionali i rapporti che esistono tra di loro e tra i documenti e le collezioni dei documenti. Questo consentirebbe all'utente un accesso più facile e "granulare" tramite OPAC rinnovati, come ipotizzato da John Byrum nell'intervento citato nell'introduzione<sup>55</sup>. Credo che secondo questo approccio, il tema della granularità – fino a questo punto della ricerca affrontato in termini "orizzontali" e "verticali" – verrebbe ad assumere una dimensione più compiutamente relazionale. La disponibilità di risorse e strumenti elettronici dischiude la possibilità di accessi più granulari non solo rispetto ai tradizionali documenti cartacei o ai documenti testuali elettronici. Banco di prova e oggetto di riflessioni di particolare interesse è

<sup>54</sup> Riprendo la traduzione del passo di Barbara Tillett dal recente lavoro di Carlo Bianchini sulle funzioni, gli oggetti e i modelli della catalogazione per autore e titolo (Bianchini 2005).

<sup>55</sup> Sebbene l'intervento di Byrum faccia riferimento anche all'uso di strutture tesauroali; su questo, cfr. la definizione di Roy Prytherch citata in seguito.

oggi costituito dai documenti digitali multimediali, soprattutto gli audiovisivi. In un recente intervento James Turner e Abby Goodrun sottolineano opportunamente come da tempo i video siano oggetto di una indicizzazione che si spinge al livello della singola scena, se non addirittura al livello del singolo frame. Ciò avviene negli archivi televisivi, per esigenze commerciali e talvolta anche legali (si pensi al caso eclatante delle sequenze relative all'omicidio del presidente degli Stati Uniti John F. Kennedy, analizzate per anni frame per frame). L'esigenza di offrire agli utenti una sempre maggiore granularità in fase di identificazione, descrizione e recupero delle parti componenti dei documenti visivi appare oggi sempre più pressante. Questo richiede uno sforzo preliminare di analisi delle immagini in movimento prodotte nelle diverse fasi di registrazione, preparazione, utilizzazione e riutilizzazione dei documenti visivi, e la preventiva creazione di una sorta di tassonomia dei diversi tipi di materiali che possono trovarsi in una collezione di video (Turner 2002).

La disponibilità dei documenti digitali in ambiente di Rete e di strumenti automatici per il loro reperimento ha suggerito nuove soluzioni anche per quanto riguarda i formati dei record bibliografici. I sostenitori dei nuovi formati di metadati si sono ritrovati ad affrontare diverse problematiche – magari già affrontate e risolte in ambito MARC – relative alla creazione e alla descrizione delle risorse digitali. Soprattutto quando si procede alla progettazione di nuove banche dati o biblioteche digitali, è necessario riflettere approfonditamente sulle specifiche caratteristiche dei metadati associati alla risorsa, sull'architettura dei record, dei campi e dei sottocampi. Interpreto in questo senso i ripetuti interventi di Roy Tennant sul termine "granularità". In un articolo del 2002 sulla sua rubrica periodica che appare *Library journal*, Tennant ha indicato la granularità come caratteristica fondamentale di qualunque tipologia di metadati, anche di quelli creati in fase di marcatura dei testi con il linguaggio TEI (*Text encoding initiative*). Tennant sottolineava:

«Our libraries are increasingly dependent on metadata [...] Granularity refers to how finely you chop your metadata. For example, in the standard for encoding the full text of books using the Text Encoding Initiative (TEI) schema, a book author may be recorded as: <docAuthor> William Shakespeare(</docAuthor>. That's all well and good, if you never need to know which string of text comprises the author's last name and which the first. If you do and most library catalogs should have this capability, you're not going to get very far with information extracted from a book encoded using the TEI tag set».

Chi ha pratica del formato Marc sa bene dell'estremo dettaglio, e quindi dell'estrema "granularità" raggiunta dai record bibliografici creati secondo quello standard. Il richiamo di Tennant – semplice, ma efficace – richiamava comunque gli utilizzatori di questi nuovi standard a progettare con attenzione i nuovi archivi digitali e i

loro metadati. Un campo autore, insomma, deve prevedere un dettaglio adeguato e la possibilità di creare sottocampi che consentano, per esempio, di ordinare i surrogati e i documenti secondo il cognome dell'autore. La soluzione è già stata da tempo trovata nei cataloghi di biblioteca, ma in ambiente di rete i diversi formati di metadati sono creati e implementati in realtà diverse. Se coloro che gestiscono gli archivi digitali non descrivono risorse in maniera dettagliata e per l'appunto granulare, i risultati finali saranno di cattiva qualità. Ne risulterebbero danneggiati gli utenti, sarebbe impossibile un dialogo gestito in forme automatiche tra le diverse biblioteche digitali e sarebbe eventualmente necessario procedere a difficili correzioni di basi dati di notevoli dimensioni.

«Granularity is good. It makes it possible to distinguish one bit of metadata from another and can lead to all kinds of additional user services. [...] Remember: select (or create) and use metadata containers that are granular enough for any purpose to which you can imagine putting them. If you do this, not only can you serve your own purposes, but you can also share your metadata with anyone you wish. If this is not practical, then you must decide which needs will remain unfulfilled. Highly granular metadata doesn't come cheap. There is a trade-off between all possible uses that you may wish to support and the staff time required to capture the metadata required to do so. In some cases, the benefit will not warrant the cost; in others, it will be worth it» (Tennant 2002).

Uno schema di metadati deve quindi essere sufficientemente granulare per rispondere alle esigenze ipotizzabili nel prossimo futuro. I costi necessari per raggiungere questo livello di dettaglio devono d'altra parte essere valutati con attenzione, per evitare uno sforzo troppo oneroso in fase di catalogazione – o di predisposizione della banca dati – in vista di benefici solo ipotetici da parte degli utenti. Verrebbe da commentare che, per esempio, i record in formato MARC sono connotati da una granularità – almeno nell'organizzazione dei campi e dei sottocampi – che certamente non è propria del formato Dublin Core. D'altra parte, anche il formato MARC non è in grado di trattare in maniera soddisfacente né le relazioni tra documenti in ambiente di rete né le relazioni tra le diverse entità dell'universo bibliografico. Solo da alcuni a questa parte, come già detto a proposito di FRBR, partendo dall'insoddisfacente trattamento riservato a queste relazioni dai tradizionali formati delle descrizioni bibliografiche, sono state ipotizzate delle notizie bibliografiche strutturate in modo da tenere conto delle relazioni tra diverse entità documentarie (l'opera, l'espressione, la manifestazione e l'item).

Una visione che mirava a identificare le specifiche caratteristiche "granulari" dei metadati e a integrarle in un quadro complessivo, è stata offerta qualche anno fa in un intervento di Carl Lagoze, che presentava la complessità dei processi di ricer-

ca e scoperta delle risorse informative tra le diverse prospettive e dimensioni della “granularità”.

«Another way to view the process is from the perspective of granularity. Over the term of the discovery process, information seekers move along an information granularity spectrum that impacts their location and examination criteria. Generally, early in the process of information seekers have relatively coarse granularity criteria. They may compose and submit relatively vague queries that, by their nature, lead to large results. In many cases, the examination of the result set may lead to successive refinement of the discovery criteria to finer granularity, as the seekers seeks to decrease the size of the result set. Again, however, the process is not linear and over the long-term the information seeker may move back and forth over the granularity spectrum» (Lagoze 1997).

L'attività di ricerca dell'informazione poteva essere favorita secondo Lagoze dai diversi tipi di metadati allora disponibili, che si componevano secondo lui all'interno del *Warwick Framework*, un meccanismo per associare molteplici set di metadati.

«For example, using the Warwick Framework, we can associate content objects with general descriptive forms such as the Dublin Core and MARC, and domain-specific descriptive forms such as that encoded in the Content Standard for Digital Geospatial Metadata (CSDGM). Each descriptive form can provide data appropriate for a relatively specific niche in the resource discovery process. The Dublin Core is appropriate for coarse-granularity, domain-independent resource discovery. The MARC record is more appropriate for the finer granularity stages of resource discovery. Finally, the CSDGM package enables fine-granularity, domain-specific resource discovery» (Lagoze 1997).

Secondo Lagoze, dunque, la maggiore o minore granularità dei processi di ricerca dell'utente, può essere relativamente assecondata da ciascuna tipologia di metadati. Il rilievo dell'intervento sta nell'impostazione complessiva, attenta sia alle caratteristiche intrinseche dei metadati che alla necessità di venire incontro e accogliere i diversi comportamenti di ricerca degli utenti.

Al di là delle relazioni tra documenti e della gestione dei surrogati descrittivi, un'altra tematica che ha assunto grande importanza nell'età di Internet è quello dell'identificazione delle entità informative, anche e soprattutto ai fini di una corretta attribuzione dei diritti economici. In un catalogo di biblioteca, la descrizione bibliografica è strumento sufficiente per un'identificazione del documento cartaceo. Fino a circa un decennio fa, l'identificazione univoca del libro, per esempio, era garantita a fini commerciali dal numero ISBD, che si integrava nel record bibliografico stilato secondo gli standard bibliografici internazionali. Questa soluzio-

ne non risolve ovviamente i problemi dell'identificazione delle copie, delle ristampe, delle edizioni. Problemi ampliati a dismisura con la moltiplicazione dei supporti e l'avvento delle realtà digitali in ambiente di rete. Identificare e localizzare un documento, gestirne le diverse versioni, consentire l'identificazione delle opere dell'ingegno a fini di tutela dei diritti intellettuali ed economici, autorizzare l'accesso a utenti con caratteristiche e diritti differenziati; si tratta di sfide che richiedono un impegno sempre maggiore.

Crescente appare poi l'esigenza di identificare le parti componenti delle diverse entità informative, come gli articoli dei periodici e soprattutto le diversificate realtà multimediali (testi, immagini, suoni) che compongono i nuovi documenti elettronici diffusi in rete. I collegamenti telematici lasciano intravedere scenari nuovi per la diffusione delle opere dell'intelletto. È stato più volte scritto e affermato che in rete i documenti non possono che essere granulari, così da consentire transazioni intellettuali e commerciali estremamente fini. In pratica, ogni singolo ricercatore potrebbe accedere a un solo articolo senza acquistare l'intera annata o anche l'issue della rivista scientifica. Questo commercio selettivo o granulare potrebbe tutelare la paternità intellettuale e i relativi diritti economici addirittura di una singola frase o immagine, controllandone efficacemente la riproduzione o distribuzione nei più diversi contesti.

Riflessioni attente in questo senso sono state effettuate in occasione della creazione di un nuovo sistema di identificazione dei documenti digitali, il DOI<sup>56</sup>. In diversi articoli e interventi, e nello stesso *DOI Handbook*, Norman Paskin, presidente dell'International DOI Foundation, ha avuto modo di tornare più volte sul tema della granularità, spiegandone il significato in questo contesto e avanzando il concetto di "granularità funzionale".

«Granularity denotes the degree of resolution or granule size of the information items identified. For example, an identifier could be applied to a complete article; for some purposes it is useful to identify individual figures or tables within that article, a finer level of granularity. Similar issues arise in the context of: identification of components and collections; a series or group; an item and parts of an item; a high-resolution, uncompressed image and a thumbnail image; a database; a record in the database; and an image of an item described by a record» (Paskin 1999).

Da una parte, come detto, un documento può essere costituito da parti diverse con caratteristiche differenziate (in questo caso, lascio i termini nella loro vaghez-

<sup>56</sup> Per una breve presentazione del DOI, all'interno di un più generale inquadramento dei sistemi di identificazione, rinvio agli interventi di Juha Hakala (Hakala 2002) e Giuseppe Vitiello (Vitiello 2004); il *Doi Handbook* è disponibile su <http://www.doi.org/hb.html>.

za), sulle quali potrebbero essere esercitati diritti diversi nel tempo e nello spazio. Identificare preliminarmente ogni possibile elemento “atomico” è però irrealizzabile. Qualunque parte di un documento potrebbe essere oggetto di interesse. La soluzione attualmente ipotizzata è quella di pensare il DOI come un sistema di identificazione aperto, in grado di accogliere gli identificativi standard attualmente esistenti (ISBD, ISSN, SICL, ecc.), consentendo eventualmente integrazioni e specificazioni, sulla base delle esigenze che via via si manifesteranno concretamente (Paskin 1999)<sup>57</sup>. In questo senso, il principio della “granularità funzionale” non assume caratteristiche rigidamente normative, ma rimane aperto: «it should be possible to identify an entity whenever it needs to be distinguished». È l’ambiente digitale a consentire, secondo Paskin, questa elasticità.

«In the digital environment, things can be relatively easily managed at extreme levels of granularity as minute as a single bit. Each of these processes will apply identifiers of different types at different levels of (functional) granularity in different ‘dimensions’, these may need to be reconciled to one another at a point of higher granularity» (Paskin 1999).

Le entità di cui parla Paskin non sono solo quelle delle parti componenti di un documento, ma le collezioni di documenti, le diverse opere presenti in un unico documento e tutti quei documenti che hanno strette relazioni tra loro. Ciascun documento derivato o correlato e ciascuna sua parte potrebbero all’occorrenza essere dotati con un DOI per la sua identificazione in rete, magari utilizzando la partizione creata con i linguaggi di marcatura (Paskin 2003).

La locuzione “granularità funzionale” è stata utilizzata anche da altri studiosi proprio a proposito del tema dei rapporti tra opere, documenti e collezioni e dei modi con i quali questi rapporti possono essere resi evidenti agli utenti. Di granularità funzionale parla Patrick Le Boeuf, alla ricerca di una (non facile) integrazione tra le esigenze che hanno portato alla creazione del DOI e quelle sottese al modello concettuale FRBR. L’obiettivo è quello della redazione di uno standard internazionale relativo a un possibile identificativo di un’opera, o meglio, con un termine FRBR e con ogni possibile cautela, dell’espressione di un’opera. Lo standard ISTC (International Standard Text Code) potrebbe identificare le parti componenti di un volume, le diverse opere all’interno di un volume in commercio, le diverse traduzioni di un’opera. Questo sforzo potrebbe avere interessanti ricadute per i bibliotecari (sia in fase di catalogazione sia al momento del recupero negli OPAC delle espressioni e delle manifestazioni di un’opera) e ovviamente per gli editori, a fini commerciali e a fini di tutela dei diritti economici di un’opera.

<sup>57</sup> Ciò non implica secondo Paskin nessuna corrispondenza predeterminata tra questi granuli atomici e brani di testo marcati per esempio in formato SGML o XML, a meno che non si reputi che questi stessi frammenti possano essere commercializzati separatamente.

Granularità funzionale nell'interpretazione di Le Boeuf (Le Boeuf 2005)

## "Functional Granularity"



(print, online CD-ROM, etc.)

ISSN 1234-567X



Publisher A



ISSN 5674-7727

(online)



ISSN 7654-3216

(CD-ROM)



ISSN 3265-442X

(print)



Publisher B

Diverse proposte di applicazione della "granularità funzionale" per l'assegnazione del numero standard ISSN

La proposta di Le Boeuf è basata proprio sul concetto di *functional granularity*:

«Functional granularity means:

If I need it, I can do it

If I don't need it, I don't have to do it

As a consequence, ISTC applies to various things, meets various needs» (Le Boeuf 2005).

La flessibilità di questo concetto ha il suo fascino. Secondo questo principio, sarebbe dunque possibile creare un identificativo (DOI o ISTC) qualora fosse ritenuto necessario. Non sorprende però che le soluzioni proposte a fini commerciali non siano ancora del tutto conciliabili con quelle avanzate dai bibliotecari. La gestione dei numeri ISTC è tutt'altro che risolta e la prospettiva di creare un DOI per qualunque entità si ritenga opportuno non risolve – secondo Kristin Antelman – il problema di mantenere traccia delle relazioni che legano le diverse entità che vanno a comporre un documento (Antelman 2005). Problemi simili sono stati presi in considerazione in fase di revisione dello standard ISO 3297, relativo all'identificazione dei seriali. La moltiplicazione dei supporti ha infatti creato notevoli problemi

nella gestione dei legami che rapportano tra di loro le riviste pubblicate su supporti diversi e tra ciascun seriale e gli articoli pubblicati al suo interno<sup>58</sup>.

La locuzione *functional granularity* viene ampiamente utilizzata dunque da diversi autori e in diversi contesti, con significati parzialmente diversi. L'individuazione dello specifico livello di dettaglio sembra comunque sempre accompagnarsi a una (non facile) discussione dei legami, tra le diverse entità bibliografiche in gioco.

Un'altra importante traccia di lavoro è offerta al proposito da alcuni interventi e alcune esperienze che si sono registrate in particolare in Scozia. George MacGregor, ricercatore presso il Centre for digital library research, dell'Università di Strathclyde, Glasgow, ha fissato nel 2003 alcune di quelle riflessioni. MacGregor evidenziava innanzitutto alcuni presupposti – se non pregiudizi – dell'attività dei bibliotecari e degli studiosi di *information science*, e cioè quella che lui valuta un'eccessiva enfasi sulle attività (la catalogazione, per esempio) relate al livello dell'item, al contrario di quelle portate sul livello delle collezioni.

A fronte delle gigantesche dimensioni assunte dalle biblioteche digitali negli ultimi anni, è inevitabile secondo MacGregor una diminuzione dell'importanza della descrizione dei singoli documenti e la crescita di valore per l'utente delle informazioni relative al livello della collezione (CLD, collection-level description). L'attenzione si sposta dunque alla definizione di collezione.

«We can consider a CLD to be a structured, open, standardised and machine-readable form of metadata providing a high-level description of an aggregation of individual items. Such descriptions disclose information about their existence characteristics and availability, and employ the use of implicit item-level metadata and, more particularly, contextualise that aggregation of item-level descriptions» (MacGregor 2003).

Nell'ambiente di rete, è probabile che sia più urgente dare e ricevere informazioni sulle collezioni digitali di documenti, piuttosto che sulla singola entità. In realtà, la dimensione e il livello della collezione sono volutamente lasciate imprecisate, come indefinite rimangono sono le parti componenti.

«An institution that agrees upon the particular aggregations that form its collections will invariably discover that these collections are related on a variety of levels. Thus, relationships could be applied to collections of varying sizes and granularity so that, for instance, a 'collection' may contain numerous 'sub-collections', and vice versa ('super-collections')» (MacGregor 2003).

<sup>58</sup> Cfr. ISO/TC46/SC9/WG 5 N 16, *Revision of the ISSN Standard (ISO 3297), Report from WG5 Adhoc group 1 as follow up to the May 2004 meeting in Amsterdam*, 14 ottobre 2004, <http://www.collectionscanada.ca/iso/tc46sc9/wg5/wg5n16.pdf>; cfr. l'immagine relativa.

Solo l'approccio della granularità funzionale (a "functional granularity approach", nella definizione che MacGregor recupera da Heaney) potrebbe consentire alle istituzioni di adottare in maniera flessibile un diverso livello di dettaglio e gestire così al meglio le parti componenti. La determinazione e descrizione delle collezioni e delle relative sottocollezioni sarebbero lasciate alle conoscenze dei bibliotecari e dei documentalisti e al loro rapporto con gli utenti. In particolare, il concetto di collezione elaborato nell'occasione dell'esperienza scozzese di SCONE richiama direttamente esperienze simili negli ambiti delle collezioni museali e archivistiche, proponendo riflessioni e soluzioni interessanti per ciascuno di questi ambiti, per tanti versi ancora oggi distanti<sup>59</sup>. Il valore di questa esperienza in termini di flessibilità rispetto alle esigenze "locali" degli utenti e alle conoscenze "locali" dei documentalisti trova probabilmente un limite nella difficoltà di rendere comunque disponibili i metadati di queste collezioni e sotto-collezioni locali agli utenti "globali". Se la definizione della collezione avviene a livello "locale", è evidente che la possibilità di riutilizzare i metadati relativi in ambiente di rete globale potrebbe risultarne limitata.

L'attenzione scozzese al livello della collezione è parallela all'attenzione americana al livello dell'item. La crescita delle biblioteche digitali ha trovato un momento di ulteriore sviluppo con il successo del protocollo *Open Archives Initiative*. I metadati relativi ai documenti di interesse scientifico stoccati nei *repository* che rispettano questo protocollo sono raccolti in fase di *harvesting* e le banche dati così create sono messe a disposizione degli utenti. Il limite riscontrato, per esempio, da Muriel Foulonneau, Timothy Cole, Thomas Habing e Sarah Shreeves è che una messe di metadati a livello di documento (il singolo item) ha il difetto di cancellare i legami con la collezione e il contesto specifico in cui gli stessi documenti sono stati creati. L'esempio scelto è quello di una raccolta di documenti su Theodore Roosevelt. Il soggetto assegnato a uno specifico documento di questa collezione digitale può essere, per esempio, «On a horse», che in quello specifico contesto chiunque non può che leggere «Theodore Roosevelt on a horse». Chiaramente, quando i metadati relativi a questo specifico documento finiscono in una base dati di dimensioni mondiali, si può perdere il contesto originale e le relazioni che lo legavano agli altri documenti della collezione.

<sup>59</sup> «SCONE is based on An Analytical model of collections and their catalogues, a study carried out by Michael Heaney on behalf of the UK Office for Library and Information Networking with support from OCLC. Heaney says: The preliminary work done by UKOLN with respect to collection description identified 'collection' as encompassing the following types of entity: Internet catalogues (e.g. Yahoo); Subject gateways (e.g. SOSIG, OMNI, ADAM, EEVL, etc.); Library, museum and archival catalogues; Web indexes (e.g. Alta Vista); Collections of text, images, sounds, datasets, software, other material or combinations of these (this includes databases, CD-ROMs and collections of Web resources); Collections of events (e.g. the Follett Lecture Series); Library and museum collections; Archives; Other collections of physical items; Digital archives» (<http://scone.strath.ac.uk/service/Guide/gdefinition.cfm>). Cfr. inoltre Powell 2000.

L'utente finale, in fase di interrogazione e lettura dei risultati della ricerca, non ha elementi sufficientemente chiari per comprendere il significato e il valore di quel documento. Per evitare questo effetto – che è poi il tipico effetto prodotto da qualunque motore di ricerca – occorre valorizzare in qualche modo le notizie relative al livello della collezione, che pure esistono, o dovrebbero esistere, in ogni record relativo a ciascun item. In senso inverso a quanto proposto da George MacGregor, occorrerà recuperare queste informazioni, proponendo soluzioni particolari e innovative a livello di interfaccia utente.

«Collection-level description can be a useful adjunct when searching item-level metadata aggregations. Collection-level description can be used as a way to preserve or restore context otherwise lost when item-level metadata are harvested from disparate and heterogeneous repositories and can also provide an additional level of descriptive granularity that may be better suited for some user queries» (Foulonneau 2005).

Questa rapida presentazione di quelle riflessioni biblioteconomiche che hanno utilizzato il termine "granularità" ha consentito di evidenziare alcuni temi rilevanti. Esigenze di identificazione a fini commerciali (DOI, per esempio) e a fini intellettuali (FRBR e ISTC, per esempio) non sembrano ancora comporsi in un quadro preciso, bensì si intrecciano e si scontrano, non riuscendo talvolta a pervenire neanche a linguaggi comuni, mentre altre volte gli stessi termini (*functional granularity*, per esempio) sono usati per sostenere tesi divergenti. Vale comunque la pena di sottolineare come la discussione sulla granularità abbia finito per investire alcuni concetti fondamentali delle scienze del libro: quello del documento, della sua unità e dei suoi confini; quello della collezione, della sua articolazione e del rapporto con il singolo item; e quello del surrogato bibliografico. Il problema più rilevante appare oggi non tanto quello di descrivere, bensì quello di identificare le entità bibliografiche e di rendere chiare e trasparenti – "tracciare", come dicono in gergo gli informatici – le relazioni che esistono tra queste entità e le loro diverse parti componenti. Questo consentirebbe agli utenti non solo un accesso immediato alla singola parte o alla singola entità, ma soprattutto la possibilità di non perdere quei legami che soli possono rendere conto del contesto effettivo in cui si pone un'entità bibliografica o informativa.

Un tentativo di sintesi delle discussioni in corso sulla granularità è stato compiuto nel 2000 da Ray Prytherch in vista della compilazione della nona edizione del prestigioso glossario di biblioteconomia curato originariamente da Leonard Montague Harrod<sup>60</sup>. Credo che quella definizione costituisca a tutt'oggi il più riuscito tentati-

<sup>60</sup> Ringrazio il prof. Prytherch per la cortesia con la quale, contattato per posta elettronica, mi ha confermato come l'inclusione del termine sia avvenuta nella 9ª edizione (e mantenuta nella 10ª) soprattutto sulla base delle riflessioni svolte all'interno dell'UK Office for Library Networking.

vo di arrivare a una sintesi delle riflessioni sul tema nell'ambito delle discipline della documentazione:

«Granularity. The level of detail expressed in a search term and in an information resource that influences the relevance of retrieved data to the end user» (Harrod 2000).

In fase di indicizzazione il dettaglio con cui è descritto un documento deve potere consentire all'utente un ampio margine di manovra, e la possibilità di avere risultati molto rilevanti per le proprie richieste. Per evitare che il livello di dettaglio sia troppo ampio o troppo ristretto, e ampliare o restringere i risultati della ricerca sulla base della necessità, occorre che lo stesso utente possa di muoversi da una parte all'altra dello spettro della granularità, dal dettaglio più minuto a una visione più ampia e complessiva.

«The end user can move along an information granularity spectrum, beginning with coarse granularity criteria (broad terms) that lead to large results, subsequently narrowing these to a finer granularity that reduces the size of the result set» (Harrod 2000).

Questa definizione sintetizza correttamente gran parte dei significati del termine che abbiamo rintracciato nei diversi contesti. Il focus di questa definizione è correttamente identificato nelle esigenze di ricerca dell'utente, che vanno agevolate dalla strutturazione degli archivi, dei database, dei documenti o dei loro surrogati. Essenziale è dunque il dettaglio con cui viene descritto un documento, ma altrettanto necessaria appare la rete dei legami tra le diverse parte componenti e le diverse versioni dei documenti. Strumenti in grado di consentire una navigazione di questo genere sono i sistemi classificatori e i tesauri, utilizzati da diversi decenni per l'organizzazione dei documenti e dei surrogati bibliografici, sia in biblioteca che nei repertori bibliografici. La navigazione gerarchica consentita dai tesauri e dai sistemi di classificazione gerarchico-enumerativi come la Classificazione Decimale Dewey possono essere concepiti come strumenti granulari, secondo la definizione di Pritchard. Tradizionalmente la granularità dei tesauri e sistemi classificatori consentiva al lettore una navigazione tra i libri e/o tra gli articoli scientifici. La liberazione dal supporto cartaceo e la creazione di un documento elettronico organizzato in maniera articolata e dal testo finemente marcato e semanticamente annotato hanno aperto la possibilità di riconciliare l'indicizzazione esterna e quella interna ai documenti<sup>61</sup>.

<sup>61</sup> A questo proposito, cfr. come XML sia oggi presentato anche agli indicizzatori di professione da un recente articolo di Bill Kasdorf come lo strumento per sanare la storica distinzione tra *open-system indexing* (quella appunto operata su monografie e articoli di rivista attraverso vocabolari controllati), *closed-system indexing* (l'indicizzazione analitica tipica delle monografie e presente al termine del testo) e *word lists* (la ricerca automatica della parola nel testo) (Kasdorf 2004).

Navigare tra concetti e relazioni è la prospettiva più ambiziosa che informa attualmente le sperimentazioni portate avanti in tutto il mondo dai ricercatori che si dedicano alla creazione del Web semantico. La prospettiva indicata da Tim Berners-Lee è quella di un Web evoluto in grado di consentire agli strumenti automatizzati di leggere e “comprendere” – dove “semantico” sta per “leggibile dalla macchina” – le informazioni, anzi le metainformazioni disponibili in Rete. Alla base di questo sistema, la marcatura XML dei documenti e la creazione da parte delle diverse comunità di utilizzatori della rete di nuove mappe concettuali, conosciute come ontologie (Berners-Lee e Hendler 2001)<sup>62</sup>, in grado di guidare gli automi alla ricerca dell’informazione.

Il Web semantico costituisce il principale sforzo teorico e sperimentale a livello internazionale che si propone di arrivare a un più efficiente ed efficace recupero dell’informazione in rete. Si tratta di un’esperienza che potrebbe integrare i risultati e le riflessioni migliori di settori diversi come quelli dell’informatica, delle scienze dell’informazione e dell’intelligenza artificiale, portando a evoluzione strumenti già conosciuti come quelli dei tesauri e dei sistemi di classificazione, e procedure informatiche come l’elaborazione del linguaggio naturale e l’estrazione automatica dell’informazione dai testi. Essenziale in questa prospettiva risulta la creazione da parte delle diverse comunità di scienziati e ricercatori di ontologie in grado di consentire la navigazione tra i documenti. Un’ontologia è un’esplicita specificazione della concettualizzazione di un dominio, di una realtà, ed è effettuata da una comunità di ricerca, sulla base della letteratura esistente e della consultazione degli esperti di quel settore, generalmente con un obiettivo specifico.

«Domain ontologies are constructed by capturing a set of concepts and their links according to a given context. A context can be viewed as various criteria such as the abstraction paradigm, the granularity scale, interest of user communities, and the perception of ontology developer» (Vickery 1997).

Questi strumenti andrebbero a innestarsi sulla preesistente esperienza dei tesauri disciplinari, considerati non più in grado di rendere in modo esplicito tutte le possibili relazioni tra concetti, che non possono essere ristrette esclusivamente alle relazioni gerarchiche.

Il livello di dettaglio raggiunto da una ontologia – la sua granularità, come per l’appunto viene chiamata nella letteratura specializzata – è una caratteristica essenziale per valutare le sue potenzialità in fase di trattamento e recupero dell’informazione. La scelta di decidere quali e quanti livelli di granularità assegnare a un’ontologia può costituire un problema e necessitare di una ridefinizione qualora

<sup>62</sup> Per una presentazione più completa del Web semantico, rimando al chiaro e affidabile manuale di Grigoris Antoniou e Frank Van Harmelen (Antoniou e Van Harmelen 2004).

si decida di modificare gli obiettivi per i quali si era creata l'ontologia oppure si desideri "comunicare" tra ontologie diverse<sup>63</sup>.

Le ontologie potrebbero essere collegate ai testi opportunamente marcati, ma soprattutto a quelle unità documentarie fino a oggi non raggiunte dall'indicizzazione umana e collocate al livello inferiore rispetto a quello dell'articolo. Il recupero di paragrafi, dati, fatti, e nomi sarebbe poi consentito agli agenti intelligenti (strumenti software costruiti per ricercare l'informazione sulla base di compiti e indicazioni assegnati loro dai ricercatori umani) in grado di effettuare la navigazione concettuale a livello di ontologie.

Il problema della granularità delle ontologie non si esaurisce in una semplice scelta di "livello di dettaglio". Se ne rende conto uno studioso come Brian Vickery, riportando le opinioni di alcuni studiosi dell'intelligenza artificiale su cosa possa essere la granularità di un'ontologia.

«There has been discussion of what has been called the 'granularity' or 'grain size' of an ontology – to what degree of specificity should the concept hierarchy be continued. Some authors, e.g. Poli, concentrate on the top-level types of concept occurring in the domain – the ontological *categories*. For him, 'an ontology is not a catalogue of the world, a taxonomy, a terminology of a list of objects, things or whatever else: it is the general framework or structure within which catalogues, taxonomies, terminologies may be given suitable organisation'. Others, e.g. Guarino, see the need to include all the specific concepts occurring in the domain, and urge that an ontology should be of high granularity» (Vickery 1997).

In una ontologia, le relazioni tra concetti potranno essere di tipo diverso, per esempio:

- *tassonomico* (IS\_A)
- *meronimico* (PART\_OF)<sup>64</sup>
- *telico* (PURPOSE\_OF)

<sup>63</sup> Il tema della granularità è stata oggetto di lunghe ricerche nel campo dell'intelligenza artificiale, a partire dall'articolo di Hobbs del 1985. Un libro di testo di riferimento, compilato da Elaine Rich e Kevin Knight affronta il tema della scelta della granularità di rappresentazione.

«Indipendentemente dal particolare formalismo di rappresentazione scelto, è necessario chiedersi a che livello di dettaglio si debba rappresentare il mondo reale, ovvero quali debbano essere le primitive da utilizzare: ne dobbiamo prevedere poche e di basso livello, o molte in grado di ricoprire un ampio spettro di livelli» (Rich e Knight 1992, p. 126).

<sup>64</sup> Per definire il significato di meronimico, riprendo un brano di una recente introduzione al concetto di ontologia in ambito filosofico, curata da Achille C. Varzi: «Il capitolo più studiato dell'ontologia formale è senz'altro la mereologia (dalla parola greca *meros*, parte), cioè la teoria delle relazioni parte-tutto: le relazioni che sussistono tra un'entità *x* e le sue parti, e tra le diverse parti di una data entità *x*, indipendentemente dalla natura di *x*. [...] il tema principale su cui negli ultimi decenni si è concentrata l'attenzione riguarda la definizione dei principi fondanti della teoria, intesa appunto come una teoria formale applicabile a *qualsiasi* dominio di entità» (Varzi 2005).

In questo quadro, la granularità andrebbe a inserirsi tra le relazioni parte/tutto, ma ciò non appare sufficiente a descrivere tutte le realtà naturali che si intende descrivere. La riflessione teorico-pratica sull'argomento ha assunto negli ultimi anni un certo rilievo, per l'impatto che l'organizzazione di ontologie finisce per avere in ogni settore di ricerca. Ciascun nel proprio specifico disciplinare, i ricercatori dovranno definire o ridefinire formalmente tutti i concetti, i fenomeni e le relazioni abitualmente utilizzati, proprio per renderli meglio leggibili dagli strumenti automatici. La ricerca sulle relazioni che sostanziano le ontologie disciplinari ha assunto una sua autonomia, che tenta da una parte di abbracciare le problematiche simili a livello generale e interdisciplinare, dall'altra è chiamata a confrontarsi con ciascuno specifico disciplinare. Tra gli studiosi che si occupano di questo settore, da anni Thomas Bittner lavora specificamente sul concetto di granularità, e lo ha affrontato in diversi ambiti, da quello biomedico a quello geografico. In un recentissimo articolo Alan Rector, Jeremy Rogers e Thomas Bittner, tentano di illuminare proprio il concetto di granularità nelle ontologie, districandosi dai diversi significati che sono stati dati nel tempo a questo termine. Gli esempi richiamati sono i seguenti:

Significati del termine <i>granularity</i> riscontrati in letteratura (Rector, Rogers e Bittner 2005)	
Specialisation	Category vs. kind—the usual notion of “is-kind-of,” e.g., that “mammal” is a generalisation including, amongst other things, dogs and elephants. Sometimes also labelled ‘abstraction’
Degree of detail	The amount of information represented about each entity, regardless of its level of specialisation. Crudely in an ontology represented in OWL [linguaggio formale con cui viene rappresentata un'ontologia], the number of axioms and restrictions concerning each entity
Density	The number of semantically ‘similar’ concepts in a particular conceptual region. How “bushy” the subsumption graph is. High local density in an ontology usually co-occurs with high levels of specialisation and degree of detail, but in two different ontologies of the same overall depth, in a particular section one may find the same two categories separated by different numbers of intervening categories or possessing very different numbers of sibling categories
Connectivity	The number of entities connected directly and indirectly to a given entity either through generalisation/specialisation or by other properties

Volendo sintetizzare al massimo il senso della loro ricerca – che appare ancora in fieri – i tre autori sostengono che due sono sostanzialmente le nozioni confuse sotto il termine di granularità: la scala dimensionale, che si riferisce sostanzialmente alle dimensioni fisiche, e la collettività, che si collega all'aggregazione di entità individuali in entità collettive che hanno proprietà e caratteristiche diverse da quelle delle entità poste al livello inferiore. Situazioni di questo tipo sono costituite dalle molecole nell'acqua, dalle cellule rispetto ai tessuti, dalle persone componenti folle, dalle stelle rispetto alle galassie. Affrontando su queste basi le ulteriori riflessioni, gli autori giungono a individuare le caratteristiche specifiche di questi due modi di affrontare il concetto di granularità, riservando maggiore attenzione proprio al concetto di "collectivity", che si rivela il più complesso e interessante, proprio al fine di "ricucire" le differenze di scala che, soprattutto nell'ambito biologico, distinguono a vari livelli le molecole, le cellule, gli organi, gli organismi, fino agli ambiente ecologici.

La strada da percorrere su questa tematica è ancora lunga, ma valeva la pena portare l'attenzione su di un approccio e un metodo che appare in prospettiva estremamente fecondo.

### **Un quadro riassuntivo**

Il sondaggio effettuato in questo articolo ha consentito di individuare una serie di usi e significati non sempre convergenti ma ciascuno di grande interesse. Le incursioni effettuate in altre aree disciplinari hanno ampliato le prospettive della ricerca e credo abbiano aperto a ulteriori approfondimenti di indagini in settori non ancora sfiorati. Per offrire una sorta di cartografia orientativa in grado di riepilogare i principali risultati dell'indagine, ritengo opportuno sintetizzare in un unico strumento di riferimento tutte quelle esperienze in cui il termine "granularità" ha assunto un particolare significato o è stato utilizzato in maniera degna di nota.

Dal quadro sommariamente presentato nelle pagine seguenti, credo emerga evidente come nella metafora della granularità appaiano leggibili approcci diversi, via via più elaborati e raffinati: uno più "orizzontale" (il documento e le sue plurime parti componenti in formato digitale), uno più "verticale-gerarchico" (documento organizzato in maniera da consentire la lettura e il recupero in maniera efficiente), uno più compiutamente "relazionale-contestuale", che si pone l'obiettivo di dare conto al lettore delle relazioni che legano i documenti, le loro parti componenti, i concetti stessi. Quest'ultimo è l'approccio sicuramente più complesso, che richiede un impegno costante da parte di chi è chiamato a costruire le ontologie concettuali e di chi sarà deputato a curare i legami relazionali tra le diverse espressioni e manifestazioni documentali.

## QUADRO RIASSUNTIVO DELLA GRANULARITÀ NELLE

DOCUMENTO A TESTO PIENO	
Caratteristiche granulari	Strumenti e soluzioni riscontrate in letteratura
Granularità del testo sequenziale	<p>Stream di caratteri, ogni carattere è codificato, l'unico livello di marcatura è lo spazio bianco e il <i>return</i></p> <p>Ricerca della lettera, della parola, delle frasi o dei passaggi</p> <p>Identificazione di importanti concetti in paper scientifici altamente strutturati (Paice &amp; Jones 1993) per la sommarizzazione automatica</p>
Granularità ipertestuale	<p>Ripartizione del testo in lessie (o chunk) e creazione dei link ipertestuali</p> <p>Lettura e ricerca tramite il browsing</p>
Granularità nella marcatura del testo	<p>Strutturazione del documento secondo formati logici: marcatura del testo a livelli sempre più dettagliati (specificazioni e qualificazione a livello di paragrafo o di elemento oppure con la creazione di dettagliati DTD)</p> <p>Dialetti scientifici XML based: SVG, MathML, UnitsML, STXML, CML (Lu 2005)</p> <p>Tensione tra strutturazione logica e contenuto del documento (Heeman 2002)</p> <p>Proposte di marcature sempre più semantiche (utilizzazione della teoria della struttura retorica del testo)</p> <p>Concetto di datumento: Murray-Rust 2003</p>
Composizione granulare del documento	<p>Partizione del documento in paragrafi</p> <p>Paragrafi, sezioni, citazioni, note, (struttura topica o irmdc)</p> <p>Strutturazione gerarchica del documento e segmentazione delle sue parti</p> <p>Strutturazione del testo e dell'abstract</p> <p>Kircs e Harmsze e modularità / granularità del documento scientifico</p> <p>Presenza di realtà documentarie diverse (suoni, testi, animazioni)</p>
Searching granulare	<p>Necessità di integrare searching e browsing per il documento strutturato (Chiarabella)</p> <p>Integrazione dell'indicizzazione con tesauri, sistemi classificatori e ontologie</p> <p>Navigazione per concetti (tesauri, ontologie, topic maps, ecc.) tra documenti e parti componenti del documento</p>

## SCIENZE DEL LIBRO E DELLA DOCUMENTAZIONE

	Estrazione dell'informazione e associazione dei testi marcati a ontologie XML e RDF, ontologie > Web semantico
Visualizzazione granulare	Multilevel document visualization
Architettura granulare dei documenti e delle risorse	Granularità a livelli degli strumenti di e-learning  Documento organizzato per concetti (topics): DITA, Darwin information typing architecture  Architettura dell'informazione (Rosenfeld e Morville)  Analisi grafica e interpretazione semantica delle diverse parti del documento (per es., digitalizzazione di documenti cartacei)

## SURROGATO DEL DOCUMENTO

Caratteristiche granulari	Strumenti e soluzioni riscontrate in letteratura
Granularità della descrizione	<p>Scelta dell'unità documentaria da descrivere</p> <ul style="list-style-type: none"> <li>• Monografie: capitoli, sezioni, paragrafi</li> <li>• Riviste: articoli, sezioni, paragrafi, ecc.</li> <li>• Video: sezioni, scene, shot, ecc.</li> <li>• RER e REL: siti, pagine, parti componenti</li> <li>• Opposizione tra collezione e item</li> </ul> <p>Problema: rapporto 1:1 tra doc. e metadato</p> <p>Granularità del metadato: campi, sottocampi, relazioni, ecc. per una ricerca nei cataloghi e nelle bibliografie</p> <p>Necessità di rendere i metadati il più possibile dettagliati; granularità del formato TEI (Tennant 2002)</p> <p>Diversa capacità dei metadati di favorire la ricerca granulare degli utenti (Lagoze 1997)</p>
Granularità dell'identificazione	<p>Utilizzazione di numeri standard relativi a diverse entità bibliografiche: opere, documenti, ecc.: DOI, ISTC, ISSN</p> <p>Concetto di granularità funzionale (Paskin)</p>
Gestione delle relazioni tra documenti	<p>Necessità di evidenziare i rapporti tra i diversi documenti (fa parte di, altra ed. di, ecc.), (Donor, Tillett, Jonsson)</p> <p>Descrizione della collezione a molteplici livelli</p> <p>Concetto di granularità funzionale (Le Boeuf, ISTC)</p> <p>Equilibrio tra descrizione dell'item e collezione, grazie anche a strumenti di visualizzazione (Foulonneau 2005)</p>

Il quadro complessivo delle caratteristiche “granulari” delle risorse informative – siano esse testi, documenti o surrogati bibliografici – potrebbe rischiare di apparire eccessivamente sbilanciato in senso “datocentrico”. Testi e documenti non possono essere ripartiti, strutturati o marcati in forma “oggettiva” e una volta per tutte, come non è possibile stabilire un unico o ottimale modo di definire una volta per tutte i metadati relativi a una risorsa o a una collezione di risorse. Se è lecito concludere questo censimento con l’indicazione di una prospettiva ulteriore di indagine, indicherei l’obiettivo di rileggere in maniera più esplicita le problematiche della granularità attraverso il focus costituito dagli orientamenti, dalle esigenze, dai comportamenti di chi si trova a cercare informazione.

*In the recent years the term ‘granularity’ has increasingly come into usage within the librarian and archival field, though it does not seem to be employed in a uniform way. The research here presented was mainly focused on “mapping” the various meanings of the term. ‘Granular’ can be used when referring to a resource structured as blocks of text (lexia) kept together by hyperlinks; to a resource’s metadata; to the description of or access to an article contained within a periodical or within one of the periodical’s files. The research highlighted the large usage of the term, but also the historical profoundness of the associated concepts and meanings. Nevertheless, in a context in which digital technologies have made it possible to search for the smallest entities of a resource (you can search for a word, a line, a citation, for the article itself, etc.), it becomes necessary to always deliver information on the relations –especially the hierarchical ones– that exist between a resource’s various components, so as to facilitate end users’ navigation.*

*Au cours des dernières années le terme granularité s’est de plus en plus diffusé dans le domaine des disciplines du livre et du document avec des utilisations souvent différentes. Le but principal de cette recherche a été celui de «cartographie» les significations que ce terme a assumé. Granulaire peut indiquer une ressource constituée par des blocs de textes (lexies) unis par des liens hypertextuels ; granulaire peut indiquer une métadonnée relative à une ressource ; granulaire renvoie à la description de ou à l’accès à un article d’un périodique plutôt qu’au fascicule dans lequel il est imprimé. Les exigences commerciales et le développement des technologies se disputent aussi ces définitions. Le recensement effectué souligne non seulement combien ce terme est diffusé mais aussi la profondeur historique des concepts et des significations qui lui sont associés. D’autre part, la recherche de l’accès aux plus petites entités d’une ressource (recherche du mot, de la phrase, de la citation, de l’article, etc.), favorisée par le développement des technologies numériques, ne peut ignorer la nécessité de fournir à l’utilisateur les relations nécessaires, surtout les relations hiérarchiques, qui existent entre les différentes parties composantes des ressources et qui peuvent lui permettre une navigation plus aisée.*

## RIFERIMENTI BIBLIOGRAFICI

- Aarseth 1997: Espern J. Aarseth. *Cybertext: perspectives on ergodic literature*. Baltimora: John Hopkins University Press, 1997.
- Anderson e Perez-Carballo 2005: James D. Anderson – Jose Perez-Carballo. *Information retrieval design: principles and options for information description, organization, display, and access in information retrieval databases, digital libraries, and indexes*. St. Petersburg (FL): Ometeca Institute. 2005. <http://www.scils.rutgers.edu/~carballo/ird2005.html>.
- Antelman 2005: Kristin Antelman. *Identifying the serial work as a bibliographic entity*. «LRTS», 48 (2005), n. 4, p. 238-255.
- Antoniou e Van Harmelen 2004: Grigoris Antoniou – Frank Van Harmelen. *A semantic web primer*. Cambridge, London: Mit Press, 2004.
- Atzeni 2003: Paolo Atzeni et al. *Basi dati. Architetture e linee di evoluzione*. Milano: McGraw-Hill, 2003.
- Baeyer 2004: Hans Christian Von Baeyer. *Information: the new language of science*. Cambridge (Mass.): Harvard University Press, 2004. Trad. it.: *Informazione. Il nuovo linguaggio della scienza*. Bari: Dedalo, 2005.
- Battaglia 1961-2002: Salvatore Battaglia. *Grande dizionario della lingua italiana*. Torino: UTET, 1961-2002.
- Bellardo Hahn e Buckland, 1998: Trudi Bellardo Hahn – Michael Buckland (a cura di). *Historical studies in information science*. Medford (NJ): Information Today for the American Society for Information Science, 1998.
- Berners-Lee e Hendler 2001: Tim Berners-Lee – James Hendler. *Publishing on the semantic web. The coming Internet revolution will profoundly affect scientific information*. «Nature», 410, 2001, p. 1023-1024. <http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>.
- Bianchini 2005: Carlo Bianchini. *Riflessioni sull'universo bibliografico: funzioni, oggetti e modelli della catalogazione per autore e titolo*. Milano: Bonnard, 2005.
- Bies 1996: Werner Bies. *Thinking with the help of images: on the metaphors of knowledge organization*. «Knowledge organization», 23 (1996), n. 1, p. 3-8.
- Blackmore 2002: Susan Blackmore. *La macchina dei memi: perchè i geni non bastano*. Torino: Instar libri, 2002.
- Bowden, Bellardo Hahn e Williams 1999: *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, a cura di Mary Ellen Bowden, Trudi Bellardo Hahn, Robert V. Williams. Medford (NJ): Information Today for American Society for Information Science and Chemical Heritage Foundation, 1999.
- Bradley 2001: Neil Bradley. *Xml companion*. Boston: Addison-Wesley, 2002.
- Brooks 2001: Terrence A. Brooks. *Where is meaning when form is gone?: knowledge representation on the Web*. «Information Research», 6 (2001), n. 2. <http://InformationR.net/ir/6-2/paper93.html>.
- Byrum 2005: John D. Byrum Jr. *Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it*. In: *Libraries: a voyage of discovery. atti del World Library and Information Congress, 71th IFLA General Conference and Council, Oslo, 14-18 agosto 2005*. <http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf>. Trad. it.: *Raccomandazioni per miglioramenti urgenti dell'OPAC*. «Biblioteche oggi», 2005, p. 5-14.

- Capurro e Hjørland 2003: Rafael Capurro – Birger Hjørland. *The concept of information*. «Annual review of information science & technology», 37 (2003), p. 343-411.
- Carlini 1999: Franco Carlini. *Lo stile del Web: parole e immagini nella comunicazione di rete*. Torino: Einaudi, 1999.
- Caton 2001: Paul Caton. *Markup's current imbalance*. «Markup Languages: Theory & Practice», 3 (2001), n. 1, p. 1-13.
- Cavarero 2003: Adriana Cavarero. *A più voci. Filosofia dell'espressione vocale*. Milano: Feltrinelli, 2003.
- Cavina e La Piccirella 2002: Enrica Cavina – Rossella La Piccirella. *Information design: progettare la comunicazione su basi cognitive*. Torino: Utet, 2002.
- Chiarabella 2000: Yves Chiarabella. *Information retrieval and structured documents*. In: Maristella Agosti – Fabio Crestani – Gabriella Pasi (a cura di). *Lectures on Information Retrieval: Third European Summer School, Essir 2000 Varenna, Italy, September 11-15, 2000: Revised Lectures (Lecture Notes in Computer-Science)*. Berlin [ecc.]: Springer, 2001. p. 286-309.
- Chowdhury 2004: Gobinda G. Chowdhury. *Introduction to modern information retrieval*. London: Facet, 2004.
- Chu 2003: Heting Chu. *Information representation and retrieval in the digital age*. Medford (NJ): Information today for American Society for Information Science and Technology, 2003.
- Conklin 1987: Jeff Conklin. *Hypertext: an introduction and survey*. «Computer», 20 (1987), n. 9, p. 17-41.
- Coombs, Renear e DeRose 1987: James H. Coombs – Allen H. Renear – Steven J. DeRose. *Markup systems and the future of scholarly text processing*. «Communications of the ACM archive», 30 (1987), n. 11, p. 933-947.
- Day 2000: Ronald E. Day. *The "conduit metaphor" and the nature and politics of information studies*. «Journal of the American Society for Information Science», 51 (2000), n. 9, p. 805-811.
- Debons 1992: Anthony Debons. *The measurement of knowledge*. In: *Proceedings of the Fifth Annual Meeting of the American Society for Information Science, Pittsburgh, 26-29 ottobre 1992*, a cura di Debra Shaw. Medford (NJ): Learned Information for American Society for Information Science, 1992, p.212-15.
- De Mauro 1999-2000: *Grande dizionario italiano dell'uso*, ideato e diretto da Tullio De Mauro. Torino: UTET, 1999.
- DeRose et al. 1990: Steven J. DeRose – David G. Durand – Elli Mylonas – Allen H. Rehear. *What is text, really?*. «Journal of Computing in Higher Education», 2 (1990), n. 1, p. 3-26.
- DeRose 1999: Steven J. DeRose. *XML and the TEI*. «Computers and the Humanities», 33 (1999), p. 11-30.
- Devlin 1991: Keith J. Devlin. *Logic and Information*. Cambridge: Cambridge University Press, 1991.
- Fayet-Scribe 1997 : Sylvie Fayet-Scribe. *Chronologie des supports, des dispositifs spatiaux, des outils de repérage de l'information*. «Solaris», 1997, n. 4. [http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d04/4fayet\\_2com.html](http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d04/4fayet_2com.html).
- Fonseca et al. 2002: Frederico Fonseca – Max Egenhofer – Clodoveu Davis – Gilberto Câmara. *Semantic granularity in ontology-driven geographic information systems*. «Annals of Mathematics and Artificial Intelligence», 36 (2002), p. 121-151.
- Foulonneau et al. 2005: Muriel Foulonneau – Timothy W. Cole – Thomas G. Habing – Sarah L. Shreeves. *Using collection descriptions to enhance an aggregation of harvested item-level metadata*. In: *Proceedings of the 5th ACM/IEEE Joint Conference on Digital Libraries, 7-11 giugno 2005. Denver, Colorado*. New York: ACM Press, 2005, p. 32-41.

- Gagliasso 2002: Elena Gagliasso. *Usi epistemologici della metafora e metafore cognitive*. In: Morabito 2002, p. 1-17.
- Gambari e Guerrini 2002: Stefano Gambari – Mauro Guerrini. *Definire e catalogare le risorse elettroniche: un'introduzione a ISBD(ER), AACR2 e metadati*. Milano: Bibliografica, 2002.
- Ghilli e Guerrini 2001: Carlo Ghilli – Mauro Guerrini. *Introduzione a FRBR*. Milano: Bibliografica, 2001.
- Gilbert, Sim, Wang 2005: Lester Gilbert – Yee-Wai Sim – Chu Wang. *Modelling the learning transaction*. In: *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*. Piscataway: IEEE, 2005, p. 615-616.
- Guercio 2001: Maria Guercio. *Rischi e promesse dell'innovazione tecnologica*. «Bollettino AIB», 41 (2001), n. 2, p. 157-173.
- Hakala 2002: Juha Hakala. *Principi di identificazione: prospettive europee*. In: Mauro Guerrini (a cura di). *Le risorse elettroniche: definizione, selezione e catalogazione: atti del convegno internazionale, Roma, 26-28 novembre 2001*. Milano: Bibliografica, 2002, p. 77-91.
- Harrod 2000: *Harrod's librarians' glossary and reference book*, a cura di Ray Prytherch. Aldershot: Gower, 2000.
- Heeman 1992: Frans C. Heeman. *Granularity in structured documents*. «Electronic publishing», 5 (1992), n. 3, p. 143-155.
- Hillesund 2005: Terye Hillesund. *Digital text cycles: From medieval manuscripts to modern markup*. «Journal of Digital Information», 6 (2005), n.1.  
<http://jodi.tamu.edu/Articles/v06/i01/Hillesund/>.
- Hillmann 2005: Diane Hillmann. *Using Dublin Core: The Elements*. 2005.  
<http://dublincore.org/documents/2005/11/07/usageguide/elements.shtml>.
- Hobbs 1985: Jerry R. Hobbs. *Granularity*. In: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, 1985, p. 432-435.
- Horn 1989: Robert E. Horn. *Mapping Hypertext*. Lexington (Mass.): The Lexington Institute, 1989.
- Johnson 1994: Gerald J. Johnson. *Of metaphor and the difficulty of computer discourse*. «Communications of the ACM», 37 (1994), n. 12, p. 97-102.
- Jonsson 2003: Gunilla Jonsson. *The bibliographic unit in digital context: how to define it?*, presentato al *First IFLA Meeting of Experts on an International Cataloguing Code*.  
[http://www.ddb.de/news/ifla\\_conf\\_index.htm](http://www.ddb.de/news/ifla_conf_index.htm).
- Kasdorf 2004: Bill Kasdorf. *Indexers and XML: an overview of the opportunities*. «The indexer», 24 (2004), n. 2, p. 75-78.
- Kerne et al. 2003: Andruid Kerne – Vikram Sundaram – Jin Wang – Madhur Khandelwal – J. Michael Mistrot. *Human + agent: creating recombinant information*. In: *International Multimedia Conference archive: proceedings of the eleventh ACM international conference on Multimedia, Berkeley (CA), 2003*. New York: ACM, 2003, p. 454-455.
- Kerne, Khandelwal e Sundaram 2003: Andruid Kerne – Madhur Khandelwal – Vikram Sundaram. *Publishing evolving metadocuments on the web: conference on Hypertext and Hypermedia archive: proceedings of the fourteenth ACM conference on Hypertext and Hypermedia, Nottingham, UK, 2003, Session: Link aggregation*. New York: ACM Press, 2003, p. 104-105.
- Kline 2004: Ronald R. Kline. *What is information theory a theory of?: boundary work among information theorists and information scientists in the United States and Britain during the Cold War*. In: *The history and heritage of scientific and technological information systems: proceedings of the 2002 conference*. Medford (NJ): Information Today for American Society for Information Science and Technology and Chemical Heritage Foundation, 2004.

- <http://www.chemheritage.org/events/asist2002/01-kline.pdf>.
- Lagoze 1997: Carl Lagoze. *From static to dynamic surrogates: resource discovery in the digital age*. «D-Lib magazine», giugno 1997. <http://www.dlib.org/dlib/june97/06lagoze.html>.
- Lancaster e Warner 2001: Frederick Wilfrid Lancaster – Amy Warner. *Intelligent technologies in library an information service applications*. Medford (NJ): Information today for American Society for information science and technology, 2001.
- Landow 1997: George P. Landow. *Hypertext 2.0: the convergence of contemporary critical theory and technology*. Baltimore (MD): Johns Hopkins University Press, 1997. Trad. it.: *L'ipertesto: tecnologie digitali e critica letteraria*. Milano: Bruno Mondadori, 1998.
- Le Bouef 2005: Patrick Le Bouef. *Identifying 'textual works': ISTC: controversy and potential*. In: *FRBR in 21st Century Catalogues: An Invitational Workshop, 2-4 maggio 2005, Dublin (Ohio)*. <http://www.oclc.org/research/events/frbr-workshop/presentations/leboeuf/ISTC.ppt>.
- Lu 2005: Zhongyu Lu. *A survey of Xml applications on science and technology*. «International Journal of Software Engineering and Knowledge Engineering», 15 (2005), n. 1, p. 1-33.
- Macgregor 2003: George Macgregor. *Collection-level descriptions: metadata of the future?*. «Library Review», 52 (2003), n. 6, p. 247-250.
- MacKay 1950: Donald M. Mackay. *Quantal aspects of scientific information*. «Philosophical Magazine», 41 (1950), p. 289-311.
- McGraw-Hill Zanichelli 2004: *Il McGraw-Hill Zanichelli. Dizionario enciclopedico scientifico e tecnico inglese italiano, italiano inglese*. Bologna: Zanichelli, 2004.
- Merton 2002: Robert K. Merton – Elinor G. Barber. *The travels and adventures of Serendipity: a study in historical semantics and the sociology of science*. Trad. it.: *Viaggi e avventure della Serendipity: saggio di semantica sociologica e sociologia della scienza*, Bologna: Il Mulino, 2002.
- Morabito 2002: Carmela Morabito. *La metafora nelle scienze cognitive*. Milano: McGraw-Hill, 2002.
- Murray-Rust e Rzepa 2003: Peter Murray-Rust – Henry S. Rzepa. *The next big thing: from hypermedia to datuments*. «Journal of Digital Information», 5 (2003), n. 1. <http://jodi.tamu.edu/Articles/v05/i01/Murray-Rust/>.
- Nunberg 1993: Geoffrey Nunberg. *The places of books in the age of electronic reproduction*. «Representations», 24 (1993), estate. Rip. in: Howard Bloch – Carla Hesse (a cura di). *Future libraries*. Berkeley: University of California Press, 1994. <http://www-csli.stanford.edu/~nunberg/places3.html>.
- Oxford English Dictionary: *The Oxford English dictionary*, a cura di John A. Simpson, Edmund S. C. Weiner. Oxford: Clarendon Press, 1989. <http://www.oed.com>.
- Paskin 1999: Norman Paskin. *Toward unique identifiers*. «Proceedings of the IEEE», 87 (1999), n. 7, p. 1208-1227.
- Paskin 2003: Norman Paskin. *On making and identifying a "copy"*. «D-Lib Magazine», 9 (2003), n. 1. <http://www.dlib.org/dlib/january03/paskin/01paskin.html>.
- Patton 2000: Glenn Patton. *Crossing a digital divide: AACR2 and unaddressed problems of networked resources*. In: *Library of Congress Bicentennial Conference on Bibliographic Control in the New Millennium, 15-17 novembre 2000*. [http://www.loc.gov/catdir/bibcontrol/patton\\_paper.html](http://www.loc.gov/catdir/bibcontrol/patton_paper.html).
- Peereboom 1998: Marianne Peereboom. *DONOR WP1 Granularity and the use of DC.Relation: WP1.1. Discussion paper, version 1.0*. 1998. <http://www.kb.nl/coop/donor/project-en-index.html?coop/donor/rapporten/granularity.html>.
- Powell, Heaney e Dempsey 2000: Andy Powell – Michael Heaney – Lorcan Dempsey. *RSLP collection description*. «D-Lib Magazine», 6 (2000), n. 9.

- <http://www.dlib.org/dlib/september00/powell/09powell.html>.
- Raieli e Innocenti 2004: Roberto Raieli – Perla Innocenti (a cura di). *Multimedia information retrieval: metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione*. Roma: AIDA, 2004.
- Rayward 1996: W. Boyd Rayward. *The history and historiography of information science: some reflections*. «Information Processing & Management», 32 (1996), n. 1, p. 3-17; Rip. in: Bellardo Hahn e Buckland, 1998, p. 7-21.
- Rayward 1998: W. Boyd Rayward. *Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext*. In Bellardo Hahn e Buckland, 1998, p. 65-80.
- Rector, Rogers e Bittner 2005: Alan Rector – Jeremy Rogers – Thomas Bittner. *Granularity scale and collectivity: when size does and does not matter*. «Journal of biomedical informatics», 39 (2006), n. 3, p. 333-349.
- Rich e Knight 1991: Elaine Rich – Kevin Knight. *Artificial intelligence*. 2. Ed., New York: McGraw-Hill, 1991. Trad. it.: *Intelligenza artificiale*, Milano: McGraw-Hill, 1992.
- Robertson 1993: Stephen E. Robertson. *Information granularity: a theme in the history of information science and technology*. Intervento (non pubblicato) all'Annual Meeting dell'American Society for Information Science, Session on the History of Information Science; il riassunto è in: <http://www.sims.berkeley.edu/~buckland/asiconf.html>.
- Rosenfeld e Morville 2002: Louis Rosenfeld – Peter Morville. *Information architecture for the World Wide Web*. Cambridge: O'Reilly, 1998. Trad. it.: *Architettura dell'informazione per il World Wide Web*. Milano: Hops, c2002.
- Rouse e Rouse 1989: Richard H. Rouse – Mary A. Rouse. *La naissance des index*. In: Roger Chartier – Henri-Jean Martin (sous la direction de), *Histoire de l'édition française: le livre conquérant, du moyen âge au milieu du XVIIIème siècle*. Paris: Fayard, 1989, p. 95-108.
- Ruecher, Homich, Sinclair 2005: Stan Ruecker – Eric Homich – Stefan Sinclair. *Multi-level document visualization*. «Visible language» 39 (2005), n. 1, p. 33-41.
- Serrai 2001a: Alfredo Serrai. *L'ontologia catalografica*. In: Alfredo Serrai. *Il cemento della bibliografia*. Milano: Bonnard, 2001, p. 13-32.
- Serrai 2001b: Alfredo Serrai. *Linguaggio informazione cultura*. In: Alfredo Serrai, *Il cemento della bibliografia*. Milano: Bonnard, 2001, p. 43-52.
- Shannon e Weaver 1949: Claude E. Shannon – Warren Weaver. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949. Trad. it.: *La teoria matematica delle comunicazioni*. Milano: ETAS Compass, 1971.
- Smeaton 2004: Alan F. Smeaton. *Indexing, browsing, and searching of digital video*. «Annual review of information science and technology», 38 (2004), p. 371-407.
- Smith 1992: Joan M. Smith. *SGML and related standards: document description and processing languages*. Upper Saddle River (NJ): Horwood, 1992. Trad. it.: *SGML e altri standard: linguaggi di descrizione ed elaborazione dei documenti*. Bologna: Clueb, 1997.
- Tennant 2002: Roy Tennant. *The Importance of being granular*. «Library Journal», 127 (2002), n. 9, p. 32-34.
- Tillett 1991: Barbara B. Tillett. *A taxonomy of bibliographic relationships*. «Library resources & technical services», 35 (1991), n. 2, p. 150-158.
- Turner e Goodrum 2002: James M. Turner – Abby A. Goodrum. *Modeling videos as works*. «Cataloging & classification quarterly», 33 (2002), ? , p. 27-38. Rip. in: Richard P. Smiraglia (a cura di), *Works as entities for information retrieval*. New York: Haworth, 2002, p. 27-38.

- UNI EN ISO 9241-11: *Requisiti ergonomici per il lavoro di ufficio con videotermini: guida sull'usabilità*. Norma UNI EN ISO 9241-11, maggio 2002.
- Van der Werf-Davelaar 1999: Titia van der Werf-Davelaar. *Identification and resolution systems for networked access to information*. «Online & CD-Rom Review», 23 (1999), n. 6, p. 325-331.
- Vickery 1997: Brian C. Vickery. *Ontologies*. «Journal of Information Science», 23 (1997), n. 4, p. 277-286.
- Vitiello 2004: Giuseppe Vitiello. *L'identificazione degli identificatori*. «Biblioteche oggi», 22 (2004), n. 2, p. 67-80.
- Walker 2006: Ruth Walker, *We're all going dotty*, «The Christian science monitor», 15 marzo 2006, p. 18.
- Worboys e Duckham 2004: Michael F. Worboys – Matt Duckham. *GIS: A Computing Perspective*. Boca Raton: CRC Press, 2004.

## Progetti



# Verba *Manent*

## Teoria e prassi della digitalizzazione dei documenti sonori e video della Discoteca di Stato-Museo dell'audiovisivo (prima parte)

Massimo Pistacchi – Francesco Aquilanti – Francesco Baldi  
*Discoteca di Stato-Museo dell'audiovisivo*

### Introduzione: la Discoteca di Stato e Museo dell'audiovisivo\*

Nell'arco dell'ultimo decennio il ruolo e le funzioni della Discoteca di Stato sono andate ampliandosi esponenzialmente e hanno conquistato un ruolo centrale nell'ambito delle attività istituzionali finalizzate alla tutela, valorizzazione e promozione del patrimonio audiovisivo nazionale.

È sufficiente evidenziare il progressivo sviluppo normativo dell'istituto a partire dalla sua fondazione per capire la portata di tale evoluzione.

Vittorio Emanuele III con il Regio decreto del 10 agosto 1928<sup>1</sup> sanciva la nascita della Discoteca di Stato: «ritenuta la necessità assoluta e urgente di disciplinare e sviluppare mediante l'istituzione di una Discoteca di Stato la raccolta e la diffusione di dischi fonografici riproducenti la voce dei cittadini italiani benemeriti della Patria».

La legge 130 del 18 gennaio 1934 recepì le idee di sviluppo della Discoteca di Stato di Gavino Gabriel, relative alla necessità di allargare gli scopi e le funzioni dell'istituto a tutto quanto nel campo dei suoni interessi la cultura scientifica, artistica e letteraria con particolare riferimento ai canti, ai dialetti di tutte le regioni e le colonie d'Italia e agli studi di glottologia e di storia. Tale legge, prevedendo la collaborazione con gli istituti affini degli altri stati per contribuire al progresso della cultura a mezzo della fonografia, cercava di sfruttare appieno le potenzialità rappresentate dalla riproduzione sonora, superando l'utilizzazione prettamente commerciale e celebrativa prevista dalla legge istitutiva.

Accanto all'attività di raccolta di documentazione sonora, nel corso degli anni trenta, la Discoteca di Stato iniziò a creare una collezione degli strumenti per la riproduzione del suono.

\* A cura di Massimo Pistacchi, Direttore.

<sup>1</sup> Cfr. <http://www.dds.it/livello.php?cat=1&sot=1&liv=1>.

Con il Regio decreto legge del 1 aprile 1935, la Discoteca di Stato passò alle dipendenze del Sottosegretariato di Stato per la stampa e la propaganda; divenne poi parte del Ministero per la cultura popolare, con la legge 467 del 2 febbraio 1939 *Riordinamento della Discoteca di Stato e istituzione di una speciale censura sui nuovi testi originali da incidere sui dischi*.

Questa legge, nella quale vengono ribaditi gli scopi previsti dal decreto istitutivo e dalla legge del 1934, rappresentò un approccio più meditato alla gestione di un materiale, quale quello sonoro, in continua crescita di importanza e diffusione. Permise, con la possibilità di acquisire la produzione discografica, l'arricchimento della collezione.

Venne inoltre riformulata la disposizione per il deposito delle copie delle pubblicazioni discografiche oggi sostituita dalla legge 106 del 15 aprile 2004 *Norme relative al deposito legale dei documenti di interesse culturale destinati all'uso pubblico*, il cui regolamento è stato oggetto del dpr 252 del 3 maggio 2006.

Le vicende belliche provocarono un'interruzione nelle attività dell'istituto e una parziale perdita dei materiali e dei documenti durante il trasferimento al nord d'Italia. Trascorso questo periodo la Discoteca fu riorganizzata – con dl 274 dell'8 aprile 1948 – alle dipendenze della Presidenza del Consiglio dei Ministri – dove erano confluite le competenze del Ministero della cultura soppresso – divisione dei Servizi informazioni e proprietà letteraria.

Dal dopoguerra la Discoteca di Stato si stabilì a Palazzo Mattei di Giove, a Roma, dove tuttora risiede.

Nel 1975 venne a far parte del Ministero per i beni culturali e ambientali, alle dipendenze dell'Ufficio centrale per i beni librari e gli istituti culturali.

A seguito delle recenti modificazioni della struttura organizzativa del Ministero per i beni e le attività culturali, l'istituto attualmente fa parte del Dipartimento per i beni archivistici e librari – Direzione generale per i beni librari e gli istituti culturali. Con la legge 237 del 12 luglio 1999<sup>2</sup> nell'ambito della Discoteca di Stato è stato istituito il Museo dell'audiovisivo «con il compito di raccogliere, conservare e assicurare la fruizione pubblica dei materiali sonori, audiovisivi, multimediali, realizzati con metodi tradizionali o con tecnologie avanzate».

Le leggi 237/1999 e 106/2004 costituiscono gli ultimi, rilevanti atti di un lungo iter legislativo che ha finalmente confermato la centralità della Discoteca di Stato-Museo dell'audiovisivo nell'ambito del patrimonio culturale italiano.

Attualmente, l'istituto vive un momento di forte evoluzione gestionale e organizzativa. In riferimento alla programmazione delle iniziative di digitalizzazione del patrimonio audiovisivo della Discoteca di Stato e Museo dell'audiovisivo, approvate dal Comitato dei Ministri per la società dell'informazione, sono state definite una serie

<sup>2</sup> Cfr. <http://www.dds.it/livello.php?cat=1&sot=1&liv=3>.

di linee operative per conseguire lo sviluppo della teca digitale, per garantire i servizi di conservazione e tutela nonché per erogare un più ampio servizio informativo. Tali iniziative di digitalizzazione comportano la risoluzione di peculiari problemi comuni agli altri archivi audiovisivi in ordine alla movimentazione, alla gestione tecnica, alla catalogazione, al controllo e alla tutela dei supporti sonori con particolare riguardo a quelli storici (cilindri, fili magnetici, lacche, 78 giri non standardizzati, ecc.), e alla nastroteca, per i quali, tra l'altro, è necessario prevedere specifiche apparecchiature di trascrizione (piatti a doppia lettura, macchine per la trascrizione ottica dei supporti storici, ecc.).

Presupposto essenziale per il conseguimento degli obiettivi prefissati è l'entrata a regime dell'intera filiera della digitalizzazione audiovisiva.

È evidente che il sistema hardware della Discoteca di Stato è stato potenziato, sia per consentire l'operatività dell'organico addetto al processo di riversamento (postazioni di lavoro per la catalogazione, stazioni di riversamento), sia per rendere il sistema nel suo complesso idoneo alle mutate attese di capacità e di potenza di calcolo.

Si è accennato che il Museo dell'audiovisivo (MAV)<sup>3</sup> è stato istituito, ai sensi del terzo comma dell'articolo 1 della legge 237 del 12 luglio 1999, nell'ambito della Discoteca di Stato.

La realizzazione del progetto porterà alla costituzione di un polo di attrazione e di riferimento per la documentazione audiovisiva, nonché bibliografica, di settore e multimediale off e on line, capace di abbracciare tutto il complesso della comunicazione contemporanea. Obiettivo primario è di assicurare, nel tempo, il valore della testimonianza storica dei beni culturali nella loro interezza – nella forma scritta, su supporto multimediale, sonoro e audiovisivo – con il contestuale recupero architettonico e ripristino della destinazione originaria di un edificio (individuato nel Palazzo della civiltà italiana) che, negli anni, ha subito un degrado strutturale ed estetico.

L'iniziativa si allinea con realtà già presenti in autorevoli contesti internazionali e il progetto si contraddistingue, oltre che per il luogo simbolico individuato, di grande visibilità, per la coerenza interiore e la rigorosa impostazione scientifica direttamente dimostrata dalle funzioni strettamente intrecciate fra loro in una unica missione: creare un marchio da comunicare in tutte le forme possibili che tenga costantemente conto dell'obiettivo di realizzare un luogo che amplii significativamente l'offerta culturale e turistico-impresoriale italiana in direzione della modernità e mediante un appropriato *merchandising*.

In data 10 Dicembre 2005 è stato firmato l'*Accordo di modifica e integrazione della Convenzione in data 10 maggio 2001, tra Ministero per i Beni e le Attività Culturali rappresentato dall' On. Ministro Prof. Rocco Buttiglione e l'Eur S.p.a.*

<sup>3</sup> <http://www.dds.it/sottcategory.php?cat=1&sot=2>.

*rappresentata dal Prof. Mauro Miccio*, accordo che regola l'uso di parte del Palazzo della civiltà italiana per destinarlo a Museo dell'audiovisivo.

In particolare, l'individuazione del Palazzo della civiltà italiana quale centro di attività culturali e promozionali e in particolare come sede del Museo dell'audiovisivo, con annessi audio/mediateca (in collaborazione con importanti enti quali le Teche Rai e l'Istituto Luce), depositi, archivi, uffici, auditorium, ecc., segna un momento essenziale nella storia della Discoteca di Stato, indirizzandola decisamente come grande centro polifunzionale di informazione e di accesso ad alto livello tecnologico e di grande impatto pubblico, nonché come ente di eccellenza nella tutela, valorizzazione e promozione del patrimonio sonoro e audiovisivo italiano.

Il Museo dell'audiovisivo consentirà di fatto di ampliare l'attività culturale della Discoteca di Stato in termini gestionali e di servizi, in ottemperanza alle premesse dell'articolo 3 della legge 237 del 1999 che recita espressamente: «È istituito, nell'ambito della Discoteca di Stato, il Museo dell'audiovisivo con il compito di raccogliere, conservare e assicurare la fruizione pubblica dei materiali sonori, audiovisivi, multimediali, realizzati con metodi tradizionali o con tecnologie avanzate». Al riguardo, per il Museo dell'Audiovisivo è previsto il perseguimento dei seguenti obiettivi:

- 1 Creazione di una audio/mediateca, al fine di predisporre un significativo interfaccia informativo sui beni audiovisivi, destinata ad ampliare in modo rilevante l'offerta di servizi per la consultazione e l'accesso alle banche dati nazionali e internazionali nonché ai documenti sonori e audiovisivi della Discoteca di Stato nella loro complessità. L'audio/mediateca risulta pertanto di importanza essenziale e connotativa e si attesta su circa 80/100 postazioni, con una disponibilità giornaliera, a fasce orarie di circa un'ora, con una previsione di servizio a favore di circa 500 utenti al giorno. A tale area si aggiunge la presenza di un centro di informazione/*references* di livello nazionale sull'audiovisivo.
- 2 Trasferimento e sistemazione definitiva del patrimonio sonoro della Discoteca di Stato (circa 300.000 supporti) nei locali adibiti a magazzini del Palazzo della civiltà italiana. La disponibilità dei nuovi depositi permetterà di risolvere i problemi relativi alla carenza degli spazi necessari alla conservazione delle collezioni storiche nonché di affrontare in modo positivo l'esplosivo incremento delle acquisizioni di beni conseguente alla legge 106/2004 sul deposito legale<sup>4</sup> che vede la Discoteca di Stato destinataria dei documenti sonori e audiovisivi prodotti e distribuiti in Italia. L'ingente patrimonio della Discoteca di Stato è infatti destinato a incrementarsi annualmente di circa 25.000/30.000 unità, secondo una stima effettuata in base all'analisi della produzione nazionale di audiovisivi nel biennio 2002/3.

<sup>4</sup> Cfr. <http://www.dds.it/sottcategory.php?cat=2&sot=7>.

- 3 Potenziamento degli uffici attualmente operanti presso la Discoteca di Stato nelle attività di catalogazione, di digitalizzazione, di registrazione nonché afferenti ai laboratori di restauro dei beni audiovisivi. In particolare, il restauro dei beni audiovisivi nonché la loro digitalizzazione, assolutamente necessari per le finalità dell'istituto, risultano attività in Italia frammentate e non coordinate, a differenza di quanto in atto in altri paesi europei.
- 4 Avvio di una programmazione di promozione culturale a livello nazionale e internazionale con l'utilizzo dei nuovi spazi (auditorium, sale espositive temporanee, spazi esterni ecc.). È in corso di elaborazione il progetto preliminare relativo all'allestimento funzionale del Museo dell'audiovisivo in grado di tener conto delle citate specifiche necessità che debbono trovare spazi, caratterizzati da una decisa vocazione tecnologica.

Il progetto del Museo dell'audiovisivo dovrà assicurare non solo l'aspetto strettamente conservativo del patrimonio della Discoteca, ma anche garantire la sua funzione di centro informativo ad alto contenuto tecnologico, capace di prevedere programmi differenziati sulle diverse fasce d'utenza (esperti, studiosi, scuole, cultori di musica e dell'audiovisivo), attraverso una consultazione rapida ed efficace e spazi organizzati secondo i più aggiornati criteri di vivibilità e di amichevolezza.

### **La digitalizzazione dei documenti della Discoteca di Stato e Museo dell'audiovisivo\***

L'obiettivo primario con cui la Discoteca di Stato e Museo dell'audiovisivo fu istituita nel 1928 era di «raccolgere e conservare per le generazioni future la viva voce dei cittadini italiani che in tutti i campi abbiano illustrato la Patria e se ne siano resi benemeriti». Successivamente, oltre a raccogliere materiale documentario a carattere storico, la Discoteca di Stato e Museo dell'audiovisivo ha rivolto il proprio interesse verso documenti di natura socio-culturale e si è fatta promotrice di iniziative quali concerti, manifestazioni culturali, rassegne, ecc., e della edizione di collane discografiche.

Nel tempo, quindi, si è venuto a costituire un ingente patrimonio di documenti, sia editi che inediti, il cui contenuto intellettuale è tra i più vari: si distinguono comunque alcune grandi classi quali le tradizioni popolari italiane (raccolte etnomusicologiche e di tradizioni orali), le voci storiche, la musica di vario genere, il teatro, la danza, e altro.

Altro obiettivo centrale è stato quello di definire e mettere in atto criteri e strutture per la corretta tutela e conservazione del materiale disco-nastrografico e video. Tutte queste attività sono in ogni caso finalizzate al principale obiettivo che la

\* A cura di Francesco Aquilanti, Promozione culturale.

Discoteca di Stato e Museo dell'Audiovisivo attualmente si pone: offrire al pubblico, esperto e non, un valido servizio di diffusione didattica e ricerca culturale.

Ma esiste una peculiarità propria di quest'istituto in termini di rapporto tra gestione e conservazione dei beni: la totale coincidenza, con la digitalizzazione, tra sistemi di gestione e di conservazione.

La digitalizzazione dei documenti della Discoteca di Stato e Museo dell'Audiovisivo è infatti legata a uno dei più significativi sviluppi tecnologici di questo periodo: la convergenza di strumenti che consentono di produrre e allo stesso tempo di accedere a una varietà enorme di informazioni in rete. L'utilizzazione futura dell'informazione viene così integrata con la garanzia della conservazione a dispetto della fragilità fisica dei supporti e dei condizionamenti dovuti ai programmi e alle attrezzature per la loro riproduzione: tecnologie in costante evoluzione e quindi soggette a rapida obsolescenza.

Il riversamento in digitale risulta pertanto operazione strategica per la tutela e valorizzazione di un patrimonio come quello della Discoteca di Stato e Museo dell'audiovisivo per la peculiare dualità di soddisfacimento delle esigenze informative e di servizio da una parte e definitiva conservazione dei contenuti dall'altra.

In realtà "reti informatiche, distribuzione e fruizione remota, ecc." fanno parte del mondo dell'informazione, un'informazione che la tecnologia ha reso radicalmente più approfondita, permettendoci addirittura di usufruire, spesso pienamente, dei contenuti.

Tali "contenuti", anche se copie fedeli, in nessun modo però potranno essere confusi, in altri ambiti come ad esempio quello bibliotecario, con l'oggetto originale, se non appunto nel loro "contenuto di informazione". Per fare un esempio potrà esserci per un bibliotecario qualche difficoltà di "identità" sui cosiddetti prodotti dell'editoria elettronica, ma un incunabolo che può essere letto su uno schermo o riprodotto da pc su un foglio di carta formato A4, mai nessun bibliotecario potrà confonderlo con l'incunabolo originario. Nessuno penserà di confondere la dignità "materica" dell'originale con la sua copia digitale...

Per i "beni audiovisivi" la cosa però cambia radicalmente. Una semplice considerazione: tutto ciò che è conservato come patrimonio presso la Discoteca di Stato-Museo dell'audiovisivo può essere utilizzato, "fruito", soltanto attraverso un macchinario: meccanico, elettrico, elettronico, digitale, ecc. Nulla, assolutamente nulla di quanto raccolto e conservato in questo istituto è di fatto utilizzabile se non attraverso una tecnologia che con la lettura di un sistema tecnico peculiare ne consenta l'utilizzo. I cilindri in cera inventati da Edison, i dischi, i nastri magnetici audio e video, fino agli attualissimi file-wave o i DVD, non sono che oggetti muti e ciechi fino a quando non vengono decodificati da una macchina. L'inezienza di questo patrimonio, pur fisicamente esistente, è inerte, muta e cieca senza l'ausilio di una tecnologia. E questo produce una sensazione di quasi "immaterialità" di questi beni.

Questa è una identità assolutamente propria e specifica rispetto al mondo bibliotecario e ai principi di tutela dei beni culturali in genere. E la differenza diventa ancor più evidente quando si entra nel contesto della digitalizzazione. Infatti quando si parla di digitalizzazione anche per il mondo bibliotecario questa rappresenta comunque un processo “immateriale”. La “digitalizzazione” infatti è o pura “informazione” (cataloghi, data base, ecc.) o “copia” asservita all’informazione o all’uso alternativo dell’originale. Non così per i beni audiovisivi: nell’epoca della “opera d’arte nella riproducibilità tecnica” la copia digitale, realizzata secondo elevati parametri di campionatura – nel caso del suono di molto maggiori addirittura di quanto percettibili da orecchio umano – e soprattutto realizzata “flat”, lineare, cioè senza alcuna correzione tecnica, è teoricamente e sostanzialmente un clono assoluto, indistinguibile, dell’originale.

L’ascolto di un 78 giri da un file digitale è pertanto identico all’ascolto effettuabile direttamente dal riproduttore originale. Il dato digitale pertanto, benché apparentemente immateriale, finisce per coincidere esattamente con l’oggetto, con il suo contenuto e con l’informazione.

Un “originale” infinitamente riproducibile.

Entriamo nel dettaglio del sistema di digitalizzazione e nelle sue procedure tecniche peculiari fino a illustrare la “filiera” di lavorazione dei supporti disco-nastografici.

### **L’architettura del sistema di digitalizzazione e gestione dei documenti della Discoteca di Stato e Museo dell’audiovisivo\***

L’architettura del Sistema di digitalizzazione e gestione dei documenti sonori e audiovisivi della Discoteca di Stato e Museo dell’audiovisivo è costituita da quattro componenti principali:

- Catalogo gestionale: fornisce informazioni catalografiche e meta-dati tecnici relativi ai documenti archiviati. Permette la documentazione di quanto trascritto.
- Sistema Teca digitale: costituisce l’archivio dei dati (documenti sonori e allegati) in formato numerico. Supporta inoltre la segmentazione dei documenti sonori.
- Stazioni di trascrizione: permettono il popolamento dell’archivio numerico, a partire dai supporti originali, sia analogici che digitali.
- OPAC-DDS/MAV: l’OnLine Public Access Catalogue. Permette la consultazione e ascolto dei documenti via Web.

\* A cura di Massimo Baldi, Servizi informatici.

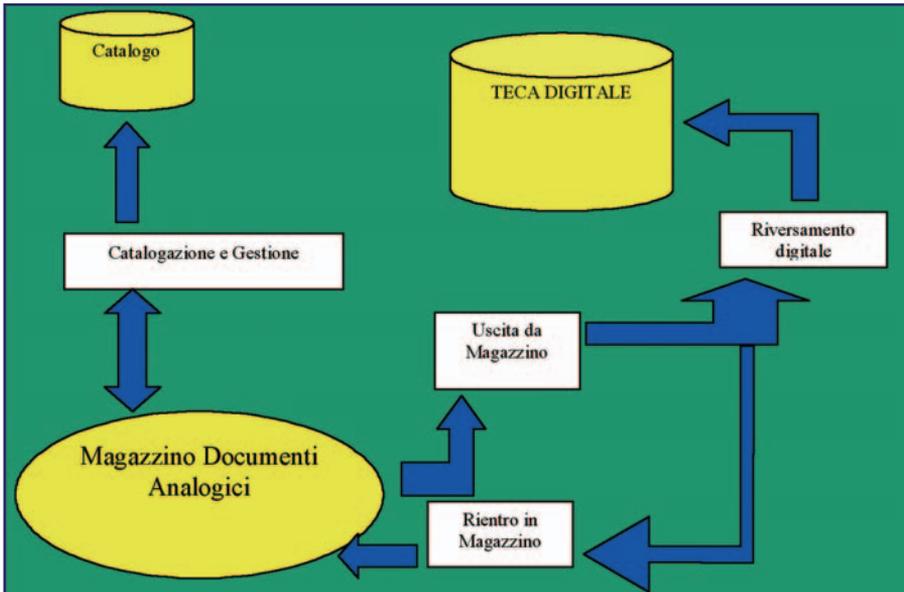


Figura 1. Architettura del Sistema di digitalizzazione e gestione dei documenti sonori e audiovisivi della Discoteca di Stato e Museo dell'audiovisivo.

### SISTEMA CATALOGO GESTIONALE

Il catalogo, governato dal software WinDJ di proprietà della DdS-MAV, è costituito da un database relazionale contenente le schede relative ai documenti sonori e video di proprietà della Discoteca di Stato-Museo dell'audiovisivo.

WinDJ consente la gestione (ricerca, inserimento, cancellazione, variazione) di schede elettroniche, la gestione inventariale dei documenti, il governo di flussi di dati in ingresso e in uscita a partire da formati standard (Unimarc e MAG).

Il catalogo dei documenti audiovisivi guida inoltre l'intero processo della trascrizione:

- contiene le descrizioni catalografiche dei documenti sonori e audiovisivi;
- fornisce informazioni sui supporti da trascrivere;
- memorizza i dati sulle trascrizioni digitali effettuate;
- fornisce i metadati descrittivi da collegare agli oggetti digitali;
- fornisce i record Unimarc con le informazioni catalografiche da trasferire sull'OPAC ADM e da collegare agli oggetti digitali;
- fornisce informazioni sulla consistenza e lo stato di conservazione dei documenti.

### SISTEMA TECA DIGITALE

La Teca digitale, Teca nel seguito di questo documento, permette la gestione di grosse moli di dati multimediali rendendoli disponibili alla fruizione.

Teca offre servizi di archiviazione e movimentazione di dati multimediali sia su base LAN che WAN. L'architettura di Teca, a oggetti distribuiti, permette di riguardare la singola istanza di TECA come un rappresentante di una rete di collaboratori che offrono servizi omogenei per natura e per logica. Ogni istanza di Teca può rispondere ed effettuare servizi in collaborazione con altre istanze, provvedendo all'archiviazione locale o centralizzata di dati multimediali, allo spostamento degli stessi su diversi sottosistemi e in genere a effettuare ricerche e rintracciare i dati multimediali su una rete configurabile di singole istanze di Teca.

Nel seguito, è illustrata la struttura attuale del sistema Teca le cui componenti principali sono:

- Sistema Gestione archivio.
- Area di cache.
- Robotica.
- Sottosistema MagTeca.

### **Sistema Gestione archivio**

Tutti i dati multimediali in transito da e per la robotica e tutti i file accessori necessari a Teca sono gestiti da questo sottosistema. Quest'ultimo controlla non solo il flusso dei dati dentro il sottosistema di archiviazione (comanda cioè i registratori della robotica), ma gestisce tutte le linee dati attraverso cui i dati memorizzati sono distribuiti agli utenti e/o da questi ritornano per essere memorizzati. Il calcolatore che deve svolgere questa funzione possiede le seguenti caratteristiche:

- *Elevata capacità di ingresso/uscita (I/O).* È fondamentale che il sistema sia in grado di gestire gli elevati flussi dati, che i registratori e le linee trasmissione saranno in grado di trasferire. La capacità di I/O può crescere in modo tale da garantire la crescita del sistema (aggiunta di nuovi archivi robotizzati, aggiunta di nuove linee di trasmissione dati) salvaguardando l'investimento effettuato. Deve essere possibile aumentare in modo semplice sia il numero di periferiche collegate su di un singolo canale di I/O, sia il numero stesso di canali.
- *Elevata quantità di memoria centrale.* La manipolazione di grossi archivi implica che debba essere possibile manipolare queste grandi quantità di dati in modo semplice e con un elevato parallelismo, se non si vogliono avvilire le prestazioni dell'intero sistema.
- *Elevata capacità di calcolo.* La gestione di un grande archivio richiede anche una elevata capacità di calcolo. La capacità di calcolo può crescere al pari della memoria e della capacità di I/O, con l'aumentare delle esigenze.

### **Area di cache**

L'area di cache costituisce la memoria di transito da e per l'archivio su nastro ma-

gnetico (LTO2). Nell'area di cache risiedono i brani di più recente acquisizione e/o quelli richiesti con maggiore frequenza. Il rapporto tra la dimensione dell'area di cache e l'archivio su nastro, dipende dalla tipologia delle richieste in lettura e dal carico di materiale da archiviare giornalmente. Per le richieste di brani in uscita, meno queste sono concentrate sullo stesso brano più grande deve essere questo rapporto, poiché lo scopo dell'area di cache è di consentire una cache di dati di primo livello dei dati conservati su supporto magnetico all'interno dell'armadio robotizzato. Nell'architettura del sistema sono presenti due aree di cache:

- *In ingresso allo storage permanente.* Qui sono temporaneamente depositati i file prodotti dai sistemi di trascrizione, in attesa di essere caricati su nastro in robotica.  
Si tratta di un'area molto critica, poiché deve garantire che tali file non vadano persi o danneggiati. Il dimensionamento della cache è stato progettato in funzione dell'attività di trascrizione e della capacità di smaltimento della robotica.
- *In uscita dalla robotica, verso gli utilizzatori.* In quest'area sono presenti i brani estratti dall'archivio, perché richiesti dagli utenti. Tali brani sono conservati nella cache finché non occorre recuperare spazio, il criterio di pulizia privilegia i brani più richiesti: se la cache è abbastanza ampia, i brani più "popolari" saranno tipicamente disponibili immediatamente, senza doverne attendere la lettura dalla robotica. Questa seconda cache non è critica, giacché duplica dati in ogni caso disponibili nel sistema, benché sia importante per conseguire un più elevato livello di prestazioni. Il suo dimensionamento è funzionale al traffico atteso e allo "sparpagliamento" delle richieste.

### Robotica

L'Armadio robotizzato costituisce l'archivio su nastro magnetico dei dati multimediali in grado di supportare il traffico in ingresso (nuovi brani da archiviare) e in uscita (richieste dagli utenti) e crescere in conseguenza della crescita continua dei brani archiviati.

Elemento sostanziale è la scalabilità del sistema. Il sistema di *storage* selezionato (ADIC Scalar 1000) consiste di un *control module* (ospitante la robotica vera e propria e un numero limitato di cassette LTO2) più una serie di *expansion modules* (dedicati a ospitare LTO2 ulteriori). La scalabilità consiste nell'affiancare i moduli d'espansione in funzione della dimensione raggiunta dall'archivio digitale.

### Sottosistema MagTeca (fruizione interna e in Internet dei dati della TECA)

Teca Oracle per la connessione al Network turistico culturale (NTC) conserva i dati multimediali (immagini e audio) in bassa e media qualità per il servizio di fruizione interno e su rete Internet. È basata su un modello di archiviazione OAIS (Open Archival Information System) e prevede la consultazione automatica attraverso il

protocollo OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting). Al proprio interno sono memorizzati i metadati in formato standard MAG 2.01 con una sezione Bib contenente gli elementi descrittivi estratti automaticamente dal Catalogo gestionale.

Teca Oracle, compatibile con lo schema MAG 2.01, è basata su modello di archiviazione OAIS:

- memorizza i dati digitali in media e bassa qualità per la consultazione;
- fornisce il servizio di ascolto su rete Intranet e Internet;
- garantisce le funzionalità di Harvesting (PMH) per la rete NTC.

### STAZIONI DI TRASCRIZIONE

Le Stazioni di trascrizione audio sono anch'esse componenti del sistema Teca digitale. Sono differenziate per tipologia di supporto. Per i supporti di tipo analogico sono composte dai dispositivi audio e da convertitori analogico digitali. Le stazioni di acquisizione per i supporti digitali prevedono l'acquisizione direttamente in formato digitale.

Le versioni su pc sono state messe a punto nell'ambito del progetto europeo Presto (Preservation Technology for European Archives)<sup>5</sup>, progetto relativo alla preservazione d'archivi analogici audio, video e film e guidato da BBC, RAI e Institut National de l'Audiovisuel con ACS e ITC/IRST come partner tecnici per la parte audio.

La trascrizione ad alta qualità di tutti i supporti genera file secondo lo standard europeo Broadcast Wave Format (BWF). Il formato di campionamento è generalmente 48 kHz 24 bit stereo salvo casi particolari da definirsi per tutti i supporti analogici (secondo le necessità per materiale che potrà poi essere sottoposto a restauro sonoro si utilizzerà un formato di campionamento di 96 kHz 24 bit). Il materiale digitale nativo (CD e DAT) viene trascritto mantenendo il formato originale. Vengono inoltre generate due copie compresse per il *browsing*: una ad alta qualità a uso degli utenti presso l'istituto e una adatta alla consultazione su Internet.

Il *bit rate* e l'algoritmo di codifica vengono definiti al momento dell'entrata in servizio del sistema di trascrizione, in modo da poter utilizzare la tecnologia più promettente.

La copia mp3 ad alta qualità viene effettuata a 256 kbit/s con codifica MPEG 1 Layer 3, mentre quella a bassa qualità viene effettuata a 32 kbit/s con codifica MPEG 1 Layer 3.

- Il materiale analogico deve essere trascritto nel formato di campionamento 48 kHz, 24 bit o 96 kHz 24 bit, stereo o mono a seconda delle necessità.

<sup>5</sup> <http://presto.joanneum.ac.at/index.asp>.

- Il materiale già su supporti digitali deve essere importato in Elettra mantenendo il formato di campionamento nativo.
- Oltre alla copia a qualità piena dovranno essere generate altre due copie del materiale trascritto, dette copie di *browsing*, utilizzando un algoritmo di compressione standard al fine di ridurre la banda di trasmissione e memorizzazione locale; le due copie di *browsing* sono destinate alla fruizione del materiale rispettivamente all'interno dei locali dell'istituto e via Internet.

La trascrizione in digitale dei supporti e degli allegati è effettuata adottando le indicazioni dell'International Association of Sound and Audiovisual Archives (IASA)<sup>6</sup>, in conformità allo standard BWF prodotto dalla European Broadcasting Union (EBU)<sup>7</sup> in collaborazione con la Audio Engineering Society (AES)<sup>8</sup>, accogliendo le indicazioni della risoluzione dell'Unesco sulla conservazione del digitale e le indicazioni pervenute dalle attività del progetto Presto.

### L'OPAC DDS-MAV

Rappresenta l'interfaccia verso l'utente finale per la fruizione dei servizi offerti dall'archivio digitale e offre diverse modalità di consultazione (accesso a catalogo sintetico, ricerca specialistica).

- Contiene le informazioni descrittive sui documenti audiovisivi.
- Contiene i collegamenti ai documenti multimediali.
- Permette la ricerca dei contenuti audio e delle descrizioni catalografiche.
- Consente la fruizione dei documenti multimediali attraverso rete Intranet e Internet.

L'alimentazione del sistema OPAC si realizza utilizzando gli standard definiti a livello internazionale:

- *Unimarc*: per il riversamento dei dati di catalogo si utilizza il formato di scambio Unimarc nella versione definita dal Gruppo Unimarc italiano per la musica e rilasciata dall'ICCU.
- *Mag 2.01*: per i metadati è stato adottato lo standard MAG emanato dal gruppo italiano di studio sui metadati coordinato dall'ICCU (vers. 2.01) e comprendente anche le sezioni audio e video.

<sup>6</sup> <http://www.iasa-web.org>.

<sup>7</sup> [http://www.ebu.ch/pmc\\_bwf.html](http://www.ebu.ch/pmc_bwf.html).

<sup>8</sup> <http://www.aes.org>.

### LA "FILIERA" DELLA DIGITALIZZAZIONE

Vediamo da vicino, in breve, il ciclo completo di lavorazione in atto sul materiale presente in Discoteca di Stato e Museo dell'audiovisivo per conseguire la generazione della Teca digitale e garantire i servizi di conservazione/tutela e l'erogazione dei servizi di ascolto.

I documenti descritti nel catalogo sono prelevati dal magazzino e resi disponibili per le operazioni di riversamento. Di ciascun documento viene verificato lo stato di conservazione ed eventualmente pianificata una operazione di pulizia/restauro. I documenti vengono inviati alle postazioni di trascrizione dove tutti i supporti analogici vengono riversati in formato digitale e caricati nella Teca digitale. Contestualmente alla digitalizzazione vengono scansionati i documenti cartacei allegati. I documenti riversati vengono ricollocati in magazzino e da questo momento saranno usati esclusivamente in caso di particolari esigenze che prevedano l'utilizzo del supporto originale.

Le tipologie di supporti attualmente in possesso, che si stanno riversando in formato digitale per essere rese disponibili per la consultazione, sono:

- dischi 78 gg (musica classica, leggera, jazz, voci storiche);
- dischi 45 gg (musica classica, leggera, jazz, tradizioni popolari, teatro, voci storiche);
- dischi 33 gg (musica classica, leggera, jazz, tradizioni popolari, teatro, voci storiche);
- nastri 1/4" (musica classica, tradizioni popolari, teatro, voci storiche);
- compact disc (musica classica, leggera, jazz, tradizioni popolari, teatro, voci storiche).

Il processo di trascrizione consiste nel riversamento di tutti i supporti analogici in formato digitale sotto forma di file che vengono caricati nella Teca. Contestualmente alla digitalizzazione, vengono anche generate le versioni compresse (file mp3 in media e bassa qualità) del materiale e vengono acquisite le immagini relative alle etichette e ai documenti cartacei allegati.

#### **Il flusso di lavorazione: un esempio concreto - i dischi a 78 giri**

Il sistema è costituito da una catena di riversamento composta da uno o più apparati analogici in parallelo collegati a un convertitore analogico/digitale, a sua volta collegato tramite un'interfaccia digitale a un personal computer, e da un secondo pc collegato a uno scanner per l'acquisizione delle immagini relative alle etichette e a eventuali materiali allegati.

L'operatore viene guidato da un software dedicato in tutte le fasi di lavoro, mentre il sistema gestisce automaticamente e in maniera trasparente l'indirizzamento dei file prodotti al sistema di archiviazione centralizzato.

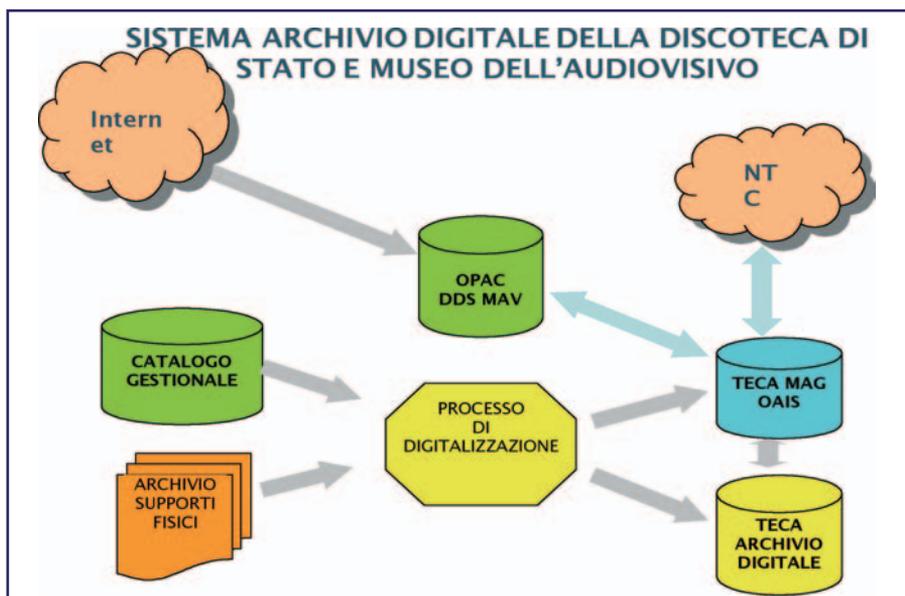


Figura 2. Sistema archivio digitale della Discoteca di Stato e Museo dell'audiovisivo.

Per ogni tipo di supporto è stata individuata una particolare procedura al fine di ottimizzare i tempi di lavoro pur garantendo la sicurezza e la qualità del risultato. Al termine delle attività di riversamento giornaliera, il materiale viene riconsegnato al personale della Discoteca di Stato e Museo dell'audiovisivo incaricato per il controllo dei supporti originali e la ricollocazione.

### Attività preliminari

- Individuazione dei supporti da trascrivere: sulla base dei dati contenuti nel catalogo gestionale si individuano i lotti di documenti da trascrivere. In questa fase si effettua anche un controllo sulla coerenza dei dati di inventario inseriti in catalogo e la consistenza dei documenti e si verifica lo stato di conservazione dei documenti.
- Presenza nel catalogo informatizzato: per una gestione razionale del processo è indispensabile che tutti i documenti da digitalizzare siano descritti nel catalogo gestionale. Infatti è importante che il catalogo raccolga tutti i dati sui documenti che sono in fase di digitalizzazione, sui documenti che sono già stati riversati in formato digitale e infine sia in grado di fornire informazioni sui documenti ancora da digitalizzare. Si producono le liste cartacee dei documenti da digitalizzare, comprensivi di moduli di richiesta per i magazzini con codici a barre di inventario e collocazione.
- Materiale cartaceo da allegare ai documenti da trascrivere:
  - moduli di richiesta per il materiale da prelevare al magazzino con i codici

a barre contenenti il numero del documento e il numero di inventario da allegare ai supporti da inviare alle stazioni di trascrizione.

– liste di prelievo (*picking lists*) sulle quali riportare annotazioni sullo stato di conservazione e sull'integrità dei dati (dischi mancanti o rotti, errori nelle collocazioni). Le liste di prelievo vengono compilate a partire dai dati presenti nel catalogo in base ai criteri sopra descritti. Nelle liste viene segnalato l'esito di ogni fase della digitalizzazione (prelievo, *scheduling*, trascrizione), fino alla ricollocazione nei magazzini.

### Attività al magazzino

- Prelievo del materiale dai magazzini: per garantire il livello di produttività richiesto agli operatori della trascrizione è necessario prelevare e ricollocare giornalmente oltre 100 supporti (dischi a 78 giri). La movimentazione di una tale quantità di supporti richiede un adeguato numero di collaboratori. Tale attività viene effettuata dal personale ausiliario coadiuvato dal personale interno.
- Controllo dello stato di conservazione: sulla base delle liste di prelievo si controlla lo stato di conservazione e si individuano le copie uniche (escludendo le doppie copie) da trascrivere. Nell'eventualità che un documento non sia idoneo al riversamento deve essere immediatamente escluso dal flusso di lavorazione. Se disponibile si procede al riversamento della copia.
- Il materiale viene prelevato dagli scaffali, lasciando al suo posto i moduli di richiesta.
- Etichettatura dei supporti da trascrivere applicando le etichette con codici a barre.
- Assegnazione e trasporto alla stazione di trascrizione: il personale addetto alla movimentazione giornalmente provvede alla consegna dei supporti da digitalizzare agli operatori e alla raccolta dei supporti già digitalizzati, incluse le liste di prelievo che contengono le informazioni sui supporti riversati e l'indicazione dei supporti che hanno presentato problemi e sono stati esclusi dalla digitalizzazione.
- Lavaggio dei dischi: i supporti a 78 giri vengono lavati, prima del riversamento, con appositi macchinari.

### Stazione di *scheduling*

La fase di *scheduling* prevede l'inserimento dei dati relativi ai task che verranno inviati alle diverse stazioni di trascrizione. In questa fase si inseriscono i dati relativi a codice di documento, numeri di inventario identificativi del supporto, numero dei supporti che compongono il documento, eventuali metadati descrittivi provenienti dal catalogo, dati sulla digitalizzazione delle immagini allegate (copertine, libretti, etichette dei supporti, ecc.). Il codice documento rappresenta l'identificativo univoco della trascrizione (denominato «codice di Album»).



Queste immagini sono tratte dalla collezione di buste commerciali per dischi a 78 giri della Discoteca di Stato-Museo dell'Audiovisivo e si riferiscono agli anni 1910-1950



- Il materiale prelevato viene trasportato alla stazione di *scheduling*.
- Il codice documento e i numeri di inventario si inseriscono utilizzando etichette con codici a barre e pistole a lettura ottica.
- I documenti vengono trasferiti alle diverse stazioni di trascrizione.

### **Trascrizione**

Vengono eseguiti i diversi task previsti; il materiale che presenta problemi viene escluso dalla catena di trascrizione e trattato a parte.

### **Digitalizzazione e scansione degli allegati**

L'operatore di trascrizione riceve i supporti assegnati e l'elenco degli stessi e procede alla digitalizzazione dell'audio e alla scansione degli allegati secondo le istruzioni ricevute e riconsegna i documenti trattati al personale addetto alla movimentazione, allegando la documentazione prescritta.

### **Processamento *batch* e archiviazione nella Teca**

Gli oggetti digitali generati durante il processo di trascrizione vengono automaticamente processati e trasferiti nella Teca digitale per l'archiviazione. Durante tale fase vengono prodotte le copie per la consultazione in media e bassa qualità destinate all'utenza locale e al «servizio Web» via Internet.

### **Aggiornamento del catalogo con le informazioni sulle trascrizioni effettuate**

I dati delle trascrizioni effettuate vengono riportati nel catalogo in linea della Discoteca di Stato; si riportano anche tutte le segnalazioni derivanti dalle varie fasi dell'intero processo di digitalizzazione.

### **Ricollocazione**

Il materiale viene ricollocato ai magazzini.

*Nel prossimo numero di DIGITALIA la seconda parte di Verba Manent. Un percorso tra: i "fasti" e i "nefasti" delle moltitudini di supporti audiovisivi, attraverso i temi della conservazione e restauro dei vari standard tecnologici della riproduzione del suono dalla fine dell'ottocento a oggi; i temi legati alla conservazione dei dati digitali; le molteplici attività promozionali, on line e off line, consentite dalla disponibilità di fonti sonore e video digitalizzate. (A cura di Francesco Aquilanti e Luciano D'Aleo)*

# Progetto InterPARES2: il *case study* sulle *moving images*

**Isabella Orefice**

*Presidente dell'Associazione Nazionale Archivistica Italiana (ANAI)*

## Introduzione

Con l'accelerazione che la rivoluzione digitale ha imposto all'incremento della produzione di fonti elettroniche negli ultimi anni, il problema della conservazione dei documenti generati in digitale è divenuta oggetto da qualche tempo di specifici studi, ricerche e discussioni. La crescente rapidità dell'obsolescenza dei sistemi di hardware e di software ha cominciato a destare preoccupazioni sempre più serie, nel momento in cui ci si è resi conto che una parte della memoria documentaria della nostra società, più recente, creata e conservata in digitale, come è noto, è già compromessa: il primo caso notevole si è infatti avuto quando a metà degli anni Novanta si è scoperto che non si potevano più "leggere" per questo motivo i nastri dei tabulati della NASA del primo viaggio sulla Luna. Non esistevano infatti più le macchine in grado di farli "girare" e non esisteva più il software nel quale erano stati codificati e col quale potevano essere letti. Lo stesso problema della permanenza della traccia magnetica dei documenti nei supporti fisici fissi e mobili è emerso solo di recente, quando questi supporti di relativamente recente origine hanno cominciato ad avere un'"età" cronologica e soprattutto tecnologica – in termini di obsolescenza della loro produzione e utilizzazione – tale da porlo concretamente.

Ma il problema della conservazione dei documenti elettronici non è solo una questione tecnica di conservazione o rinnovamento degli hardware e dei software originari, anche se già questa pone problemi ancora non risolti. Per secoli, la conservazione dei documenti si è basata su quella del supporto cartaceo su cui erano scritti, ed era l'unico, in quanto la lettura diretta a vista non poneva alcun problema di obsolescenza del sistema di decodificazione. Il fatto fondamentale per cui il documento digitale non è più basato su un'entità fisica corrispondente stabile e direttamente leggibile, ma è piuttosto il risultato di un complesso processo di codificazione e decodificazione, alla base del quale stanno altri processi informatici ed entità fisiche (memorie, circuiti, tastiere, schermi, ecc.) distribuite, nessuna delle quali corrisponde da sola in modo fisso ed esclusivo al vero e proprio "documento", fa sì che il problema della conservazione del documento digitale diventi il problema di assicurare la conservazione del risultato di un processo, ben più complesso di quello di conservare un'entità fisica circoscritta.

Ma il documento tradizionale, dal punto di vista archivistico, non è un semplice testo o complesso astratto di informazioni linguistiche, bensì anche un oggetto materiale, l'integrità dei caratteri fisici del quale ne garantisce il requisito indispensabile dell'autenticità. Perciò, dal punto di vista archivistico – e non semplicemente informatico – il problema della conservazione del documento digitale non è solo quello di sostituire a quello obsoleto un nuovo diverso supporto e processo di ricodificazione e decodificazione tale da produrre di regola un testo uguale a quello originario, cioè dal punto di vista archivistico una semplice copia, ma soprattutto quello di produrne una copia "autentica" o, in altri termini più generali, di garantire anche la conservazione della sua autenticità. Si pone quindi per gli archivisti il problema di individuare, in ambito digitale, i requisiti che possono assicurare al documento la conservazione della sua autenticità in modo equivalente a quello finora garantito dal documento tradizionale mediante la sua stabilità fisica come oggetto materiale.

L'impostazione organica di questo problema e l'indicazione delle linee per una sua soluzione è stato il grande obiettivo del progetto internazionale InterPARES (International Research on Permanent Authentic Records in Electronic Systems) sulla conservazione permanente della documentazione elettronica, avviato nel 1999. Il progetto<sup>1</sup> è stato promosso e coordinato da Luciana Duranti, docente di archivistica di origine e formazione italiana della British Columbia University di Vancouver, che ha raccolto un ampio gruppo di ricerca internazionale, che ha affrontato in modo articolato le tematiche multidisciplinari implicate dal suo obiettivo generale. Il progetto ha coinvolto paesi come il Canada, gli Stati Uniti, l'Australia, il Belgio, la Francia, l'Italia, l'Irlanda, la Svezia, il Regno Unito, la Cina, Singapore, il Giappone e si è articolato in due fasi.

InterPARES1 è stato dedicato all'individuazione della teoria e delle metodologie di conservazione dei documenti amministrativi, creati e mantenuti in database e su sistemi di *document management* non più utilizzati dal loro produttore. I primi risultati di questa fase 1 si sono concretizzati nella formulazione di teorie e metodi relativi all'autenticità, alle pratiche di selezione dei documenti, allo sviluppo di politiche e strategie, di strumenti per l'analisi e la valutazione di indirizzi e strategie per la conservazione a lungo termine dei documenti elettronici, grazie all'adozione di appositi standard di formattazione e trattamento. È infatti il rigoroso rispetto di questi ultimi, sui quali non mi soffermerò in dettaglio in questa introduzione al nostro *case study*, che potrà, secondo le conclusioni di InterPARES1, garantire la conservazione dell'autenticità del documento digitale; in altri termini, essa può essere garantita non tanto dalle caratteristiche del contenuto del documento, quanto da quelle delle procedure di formazione e riproduzione del documento stesso, e quindi la conservazione dell'autenticità non è solo questione di mera formattazio-

<sup>1</sup> I documenti prodotti dal progetto InterPARES si possono consultare sul sito <http://www.interpares.org>.

ne del documento, ma anche di una rigorosa definizione e controllo del contesto operativo del suo successivo trattamento.

La seconda fase del progetto, InterPARES2, iniziata nel 2002 e ora in conclusione, ha avuto come obiettivo quello di sviluppare e articolare requisiti e metodologie per garantire la produzione e la conservazione di documenti autentici, correnti, dalla loro nascita fino alla conservazione, senza forma fissa, prodotti nell'ambito di attività artistiche, scientifiche e di *e-government*, spesso caratterizzate dall'uso di sistemi dinamici, esperienziali e interattivi. Garantire l'autenticità di questo tipo di documenti digitali dinamici e interattivi – e ci riferiamo per esempio ai film contenuti in un sito web, a un sistema multimediale virtuale, alle banche dati *on-line* interattive – dipende dalla capacità di ricreare le caratteristiche dell'ambiente in cui questi documenti sono stati "sperimentati".

InterPARES2 è articolato in tre aree di ricerca, o *focus*, in base al tipo di attività che produce i documenti: il *Focus 1* studia i documenti prodotti da attività artistiche; il *Focus 2* analizza i documenti prodotti da attività scientifiche; il *Focus 3* esamina i documenti prodotti dall'interazione amministrativa *online* con i cittadini (*e-government*). Ciascun *focus* è diviso in tre parti, o *domains*, dello stesso tipo, ciascuna dedicata a un gruppo specifico di questioni da esaminare: il *Domain 1* esamina la natura dei documenti di una delle attività che rientra in quel *focus* e il loro processo di produzione, ovvero il ciclo di vita dei documenti; il *Domain 2* studia i concetti di affidabilità, accuratezza, e autenticità nel contesto delle discipline che appartengono a quel *focus*; il *Domain 3* monitorizza metodi esistenti o proposti di selezione e conservazione su casi specifici, sviluppando e suggerendo nuove metodologie, quando necessario. Per ognuna di queste articolazioni si è formato un gruppo di ricerca internazionale, per un totale di nove, coordinato da un *International Team* generale di progetto.

Le problematiche comuni al progetto generale sono a loro volta esaminate da ciascun gruppo nell'ambito di quattro *cross-domains*:

- 1) Terminologia, per il controllo dell'uso dei termini e relative definizioni per garantire coerenza di linguaggio tra tutti i gruppi di ricerca.
- 2) Strategie, per analizzare politiche e linee-guida per la produzione, tenuta e conservazione dei documenti e svilupparne delle nuove.
- 3) Descrizione, per studiare schemi descrittivi esistenti in ciascuna disciplina e svilupparne dei nuovi, al fine di renderli trasferibili e applicabili nel contesto di tutte le discipline coinvolte.
- 4) *Modeling*, per elaborare funzioni, informazione e risorse necessarie per la produzione, tenuta e conservazione di documenti accurati, affidabili e autentici e supportare la rappresentazione e l'analisi dei risultati degli studi di casi. Cerca, in pratica, di tradurre e rappresentare gli obiettivi della ricerca in modelli formali, con lo scopo di verificarne la validità e la comunicazione interna ed esterna, nonché la disseminazione dei risultati del progetto.

I ricercatori di InterPARES2 sono organizzati in unità di ricerca a seconda del loro interesse e della loro specializzazione. L'attività di ricerca consiste nello studio di casi, in cui si esaminano la creazione, il mantenimento e la conservazione dei record risultanti dalle attività comprese in ciascun *focus*, e include anche attività come l'effettuazione di inchieste sul campo e la compilazione di bibliografie.

Per quello che riguarda il *focus* artistico, i ricercatori sono esperti nei campi della musica, film, danza, *performing art*, fotografia, teatro, con l'aggiunta di archivisti e di esperti di scienza informatica. Gli oggetti dei *case studies* includono i record delle *performing arts* multimediali, musica elettroacustica e immagini digitali.

I principi metodologici del progetto sono:

- 1) **Interdisciplinarietà:** gli obiettivi del progetto si realizzano grazie al contributo di diverse discipline. Ad esempio, per sviluppare metodi di creazione e di affidabilità dei *records* nel corso della produzione digitale della musica, si fa riferimento e si consultano teorici della musica, compositori, ingegneri informatici, scienziati e storici della musica.
- 2) **Trasferibilità:** un altro obiettivo del progetto è lo sviluppo di un sistema affidabile di tenuta e trasferimento dei *records*, al fine di garantirne l'autenticità nel tempo. Questo implica che il lavoro svolto durante il progetto nelle varie aree disciplinari, deve essere costantemente tradotto in termini archivistici e legato a concetti archivistici, che sono il fondamento sul quale i sistemi intendono proteggere i record. D'altro canto, per poter comprendere la complessità della ricerca, i sistemi archivistici hanno necessità di essere resi comprensibili per i creatori dei *records*, organizzazioni e istituzioni e ricercatori della disciplina.
- 3) **Ricerca aperta (*Open inquiry*):** i ricercatori di InterPARES2 in ogni gruppo di lavoro identificano le prospettive di ricerca, i metodi che ritengono più appropriati alla loro ricerca. La ragione di questa apertura è che InterPARES2 è concepito come un lavoro di conoscenza per gradi, stratificata. Infatti alcuni membri del gruppo di ricerca si servono della conoscenza sviluppata nel corso del progetto InterPARES1, alcuni di quella maturata in differenti aree, e tutti possono seguire le linee di ricerca così intraprese per svilupparle a fini archivistici proponendo lo studio di nuove entità e nuovi problemi.
- 4) **Disegno basato su una molteplicità di metodi (*Multimethod design*):** ogni attività di ricerca è svolta usando la metodologia e gli strumenti che il *team* di ricerca considera il più appropriato. I principali metodi impiegati sono: inchiesta, *case study*, *modeling*, analisi diplomatico-archivistica e analisi testuale. L'inchiesta è basata su *research questions* organizzate secondo i *domains* e i *cross domain* e appositamente predefinite per la singola ricerca. Nel preparare i loro strumenti per i *case studies* e nel predisporre le inchieste, i ricercatori sono guidati da 23 "domande", sviluppate dall'*International Team*.

Nell'ambito di tale complessa articolazione delle attività di ricerca del Progetto InterPARES2, i *case studies* comportano l'analisi di oggetti digitali prodotti da sistemi specifici nel corso dell'attività di un determinato tipo di produttore d'archivio, all'interno della competenza di uno dei tre *focus* sopra descritti. I *case studies* sono condotti da gruppi di ricerca i cui membri appartengono ai tre *domains* entro ciascun *focus*. Vorrei illustrare ora in particolare un *case study* del *focus* 1 condotto in Italia, a cui ho partecipato, relativo alle attività artistiche e, in particolare, a quello sulle *moving images*. Non anticipiamo in questa sede i risultati conclusivi di InterPARES2 comprendenti le indicazioni relative ai requisiti e standard da applicare per la conservazione a lungo termine di documenti digitali esaminati che sono state elaborate successivamente al rilevamento del *case study*, di cui qui ci occupiamo esclusivamente, verranno presentati il 12 e il 13 dicembre 2006 a Milano durante il seminario internazionale di InterPARES2 che si svolgerà nell'ambito della I edizione di Archiexpò.

### **Il *case study* sulle *moving images*, ovvero sul film digitale**

Il *case study* sul processo produttivo di un film digitale è stato motivato da una serie di criticità, quali la mancanza in generale di standard di produzione e tenuta dei documenti filmici; la difficoltà di stabilire l'originale e la copia autentica di un film, l'uso di speciali tecnologie come parte dell'espressione artistica, la mancanza di copie multiple dello stesso documento; il problema del *copyright* (diritto di creare copie identiche, di creare lavori derivati, di presentare o rappresentare il lavoro, problema di stabilire a chi appartiene e per quale versione, con le relative conseguenze per la conservazione, la consultazione, l'accesso e il riuso del documento).

La metodologia di lavoro si è concretizzata in una serie di attività o fasi, quali: seguire un film attraverso le fasi di pre-produzione, produzione e post-produzione; acquisire informazioni sull'origine del progetto e sul materiale prodotto durante il lancio e lo sviluppo della fase di pre-produzione, concentrarsi sull'analisi del film e mettere in relazione il materiale documentario che lo accompagna con i risultati di InterPARES1. Tra gli obiettivi del *case study* si segnalano in particolare: quello di costruire un modello che descriva e spieghi i processi connessi alla creazione del film digitale in merito alle tre aree esaminate (commerciale, pubblica e indipendente multimediale); seguire un film attraverso le fasi sopra menzionate, identificare i documenti che risultano da ciascuna attività, identificare le entità digitali generate in ogni fase del processo produttivo, rappresentare in un modello la struttura di ciascuna entità digitale creando campi marcati (*tagged*), scoprire quali entità digitali hanno strutture standardizzate o usate comunemente, rendere esplicita la funzione di ciascuna entità digitale nell'ambito di ogni attività e processo, mostrare le relazioni tra le varie entità digitali e tra queste e l'intera produzione, mettere in relazione i metadati identificati in questo studio e

l'MPEG-7<sup>2</sup>. Tale *case study* è stato finalizzato ai processi di produzione delle *moving images*, a identificare la documentazione creata ad ogni *step* del processo e a riflettere su metadati utili per la gestione della documentazione prodotta.

Il gruppo di ricerca, che poi si è allargato, inizialmente era formato da Marta Braun (Ryerson University), Randal Luckow (DreamWorks), Michael Murphy (Ryerson University), Isabella Orefice (ANAI), Andrew Rodger (National Archives of Canada) e da James M. Turner (Université de Montréal), direttore.

Questo lavoro è stato svolto analizzando, in tre paesi, quattro soggetti produttori corrispondenti a quattro tipi di produzione di *moving image*: uno studio commerciale di Hollywood (USA), un'istituzione pubblica nazionale di produzione filmica (Canada), un produttore *broadcaster* di una televisione pubblica (USA) e una casa di produzione multimediale indipendente (Italia).

Come metodologia di lavoro si è scelto di intervistare il personale delle varie strutture, di descrivere dettagliatamente entità, attività e processi, di descrivere le infrastrutture tecnologiche utilizzate e di avviare l'attività di *modeling* di processi, entità, controlli e meccanismi, individuando i criteri seguiti dai soggetti esaminati nell'assegnazione dell'autenticità e affidabilità dei documenti digitali.

La descrizione e l'analisi del processo, nell'ambito di ciascuna struttura, sono stati portati avanti grazie a una rilevazione basata sul tracciato UID (identificatore unico, quale produzione, quale dipartimento, quale sequenza, quale scena), che comprende i campi relativi a:

- creatore
- motivo di creazione, uso
- struttura dell'informazione (analisi diplomatica)
- supporto (se digitale, software e sistema operativo)
- versione: come si controlla
- parole chiave
- data di creazione o modifica
- scheda di conservazione e scarto
- *backups*
- contesto documentario e procedurale
- note

I dati raccolti nell'ambito del *case study* su tre strutture di produzione audiovisiva digitale inizialmente individuate nel contesto di InterPARES2 mostrano come que-

<sup>2</sup> Con lo standard MPEG-7. L'MPEG (Moving Picture Export Group) propone di associare a ogni oggetto audio-video una descrizione, che possa essere estratta e indicizzata da un motore di ricerca. Evidentemente, MPEG-7 è basato su MPEG-4, che fornisce gli strumenti per la codifica delle informazioni multimediali tramite AVO. Oltre a una descrizione del contenuto, è utile avere altre informazioni sui dati multimediali, come: la forma, le condizioni di accesso, la classificazione, i link ad altro materiale, il contesto.

sto lavoro possa contribuire a incrementare l'accesso alle *moving images* e a fornire una base teorica più solida per una migliore organizzazione delle collezioni di film.

### Studio cinematografico commerciale (USA)

Il nostro partner per questa parte del *case study*, una società di produzione di Hollywood, ha richiesto l'anonimato. I dati sono stati raccolti dai ricercatori attraverso le interviste dell'archivista della società, utilizzando le 23 domande che tutti i ricercatori di InterPARES debbono usare. Oggetto dell'analisi è stata la sequenza e le procedure degli stadi della creazione di un film animato mediante la computer grafica, che saranno di seguito esposte. Sono stati studiati gli aspetti del processo, dalla creazione delle immagini, alla realizzazione del prodotto finale, escludendo aspetti amministrativi come diritti, contratti e accordi. Sebbene InterPARES2 si occupi unicamente di file digitali, sono stati considerati nel nostro *case study* anche elementi analogici, dato che questi fanno parte del processo di produzione e non possono essere isolati dagli aspetti digitali.

Per poter comprendere come questo lavoro si è svolto bisogna considerare il processo iniziale della pianificazione di un film. Le idee vengono visualizzate in forma di *artwork* manuale su carta e questo lavoro viene montato in uno *storyboard*. Molti singoli disegni portano a un'ulteriore sequenza, poi mostrata a coloro che hanno l'autorità di decidere se l'idea è abbastanza interessante per farne un film. È facile immaginare come l'uso dei disegni analogici sia vantaggioso. Gli artisti preferiscono fare piccoli disegni su carta perché la matita e la carta sono più rapide e facili da usare. Questo metodo di consultazione è molto più efficiente, rispetto alla manipolazione di molte finestre su un computer che dia la stessa informazione. Sistemando i fogli su una lavagna, si possono poi mettere in sequenza per essere mostrati più facilmente in una presentazione per coloro che decideranno la realizzazione o meno del film. Se il progetto è approvato, le immagini sono scansionate e diventano le entità digitali iniziali utilizzate maggiormente per la realizzazione del film.

Nello studiare questo processo, si scopre che, in effetti, viene creato un numero relativamente piccolo di entità digitali. Questi oggetti digitali, durante la lavorazione, sono costantemente alterati con un ulteriore lavoro di elaborazione e integrazione prima di passare nella catena della produzione.

Questo processo sviluppa un nuovo *artwork* digitale, che permette di modificare la consistenza e il contenuto degli elementi, per esempio prendendo le immagini base di un carattere che è stato approvato per l'uso nel film, e aggiungendo un costume a esso. Una volta che il nuovo lavoro prodotto è soddisfacente, l'*artwork* è composto e pianificato a ciascuno stadio, al fine di ridurre il formato del file. Questo comporta l'integrazione mediante un ulteriore lavoro artistico. Quando il file è così ridotto, a un singolo *layer*, richiede meno

spazio per l'immagazzinamento rispetto a un file con strati multipli. In questo modo il numero dei file è ridotto, anche se ogni file è costantemente modificato, a mano a mano che il lavoro procede. Tutta l'attività è focalizzata nel produrre il lavoro artistico nel formato TIFF, che sarà l'unico delle immagini del film finale.

Come parte del nostro lavoro, abbiamo chiesto come vengono identificate le entità digitali. Abbiamo compreso che all'inizio della produzione a ognuno che lavora sul film è richiesto di apprendere le convenzioni delle denominazioni dei file. Queste includono informazioni sulla sequenza e la scena del film, l'identificazione di oggetti digitali particolari inclusi, la versione particolare dell'oggetto digitale, l'estensione del file che indica il software necessario per leggere il file. Tutte queste informazioni sono racchiuse nel nome del file, e quando questi file passano da una persona all'altra per la lavorazione, il numero della versione cambia.

Tutto l'immagazzinamento e il recupero delle entità digitali è basato su queste convenzioni nominative e i numeri della sequenza e delle scene includono nel file l'assemblaggio delle immagini con il proprio ordine per costituire il film. Il metodo di denominazione convenzionale dei file è l'unico a dare una stabilità alle procedure che portano alla realizzazione del film, dal momento che le procedure di produzione non sono altrimenti documentate. Queste ultime possono cambiare radicalmente da un film all'altro, così si ritiene che documentarle non sia utile, mentre le informazioni circa le procedure seguite sono comunicate direttamente da un *team* di produzione a un altro nell'ambito delle discussioni comuni che si svolgono nei primi stadi della produzione.

Noi ci domandiamo anche come siano assicurate l'affidabilità e l'autenticità dei file digitali. In questo caso è il numero della versione dei file a conferire ad essi l'autenticità e l'assegnazione corretta del numero è assicurata mediante l'uso delle denominazioni convenzionali adottate per la produzione. Così, se il documento risponde al software nell'aprire ed eseguire i comandi dati per il successivo stadio del lavoro, è riconosciuto come affidabile.

### **Casa di produzione multimediale indipendente Altair4 Multimedia (Italia)**

Questa parte del *case study* è stata realizzata in collaborazione con l'ANAI che ha organizzato un gruppo di lavoro specifico di esperti nell'area delle *moving image*. Un membro del nostro *team* ha avuto la possibilità di contattare un interessante gruppo di professionisti che lavorano in una piccola società multimediale, Altair4 Multimedia, che ha sede a Roma e adotta metodologie progettuali dell'*industrial design* per la realizzazione di siti web e supporti informativi per musei e istituzioni. Il gruppo creativo di Altair4 proviene da diverse esperienze nell'ambito delle arti applicate della computer grafica e della regia televisiva, e ha trovato nello stru-

mento multimediale un modo nuovo e più organico di comunicare, caratterizzato dalla fusione di diverse discipline e metodologie.

L'amore per l'arte e la storia dei membri del gruppo ha influenzato la scelta degli argomenti trattati negli ultimi anni, che sempre più spesso, senza disdegnare parentesi nel mondo dell'animazione e della televisione, contengono tematiche inerenti la divulgazione e valorizzazione del patrimonio artistico culturale. Proprio in questo ambito è da segnalare l'ampia produzione di ricostruzioni archeologiche tridimensionali realizzate dall'*atelier* per musei, produzioni televisive e Cd-Rom interattivi.

Il progetto di Altair4, che abbiamo studiato in questo contesto, e del quale esporremo l'analisi procedurale, riguarda un progetto di ricostruzione virtuale, mediante la tecnologia multimediale, intitolato a un sito archeologico di Pompei, distrutta nel 79 d.C. dall'eruzione del Vesuvio, la casa di Giulio Polibio. Il progetto, commissionato dall'università di Tokyo che ne ha fatto oggetto di una mostra itinerante con intenti didattici, è stato presentato nell'ambito dell'incontro dell'*International Team* di InterPARES2, tenutosi a Roma nel dicembre 2002.

In campo archeologico il linguaggio della realtà virtuale, basandosi sulla percezione visiva e spaziale, si pone come mezzo per una trasmissione di dati complessi a una vasta platea. Interessante è capire quali caratteristiche debba avere chi materialmente realizza il restauro virtuale e in che termini questi si relaziona con i ricercatori. Una soluzione possibile potrebbe essere quella di addestrare gli archeologi a operare in questo particolare settore, rendendo così il flusso dei dati diretto; tuttavia sembra difficile che si possa giungere a risultati qualitativi propri di chi ha fatto di questo lavoro (cioè la ricostruzione virtuale) una professione specifica. Nonostante la straordinaria novità apportata dagli strumenti tecnologici informatici, chi si pone oggi l'obiettivo di una ricostruzione archeologica deve però essere cosciente che sta perpetuando una tradizione radicata nel tempo, che dalle tavole di Palladio passando per gli *Envois des pensionnaires* dell'Accademia di Francia del XVIII secolo, mantiene oggi sempre lo stesso spirito, anche se si sono sostituiti gli acquarelli con dei pixel.

Nel nostro studio della produzione di questa società abbiamo infatti rilevato un dato per noi fondamentale, ovvero l'interdisciplinarietà del processo di produzione. Nella ricostruzione virtuale della casa di Giulio Polibio, Altair4 ha infatti lavorato con archeologi, botanici, matematici e vulcanologi per creare un'immagine la più autentica e accurata possibile, un filmato con scopi divulgativi di supporto alla mostra *Pompei e i suoi abitanti*, commissionato dall'Università di Tokyo. Scopo del filmato è portare gli spettatori indietro nel tempo, mostrando la casa di Polibio così come era nel 79 d.C. e la successione scientificamente corretta degli eventi catastrofici che portarono alla sua distruzione.

Abbiamo investigato mediante le domande predisposte dal *team* internazionale le fasi sviluppate nella ricostruzione archeologica virtuale della casa di Giulio Polibio

e le attività che hanno generato le entità digitali principali (tutti i componenti della villa, come i modelli della casa, gli arredi, gli oggetti e i tessuti), e si è venuta delineando la metodologia di lavoro usata dalla società per la produzione digitale, che costituisce sicuramente la parte preponderante della ricostruzione virtuale.

Nel modellare le entità digitali, Altair4 ha iniziato dall'esterno procedendo all'interno. Ciò era necessario anche in ordine al seguente corso degli scavi archeologici. Una camera digitale veniva usata per riprendere tutte le parti esistenti dell'edificio, comprese le pareti affrescate, e le foto erano scansionate e lavorate. Queste entità sono state associate mediante un apposito kit di elaborazione grafica con un set di coordinate spaziali e di proprietà spaziali relazionali, includendo caratteristiche fisiche e modelli di riflessione e assorbimento della luce. Con l'aiuto di un'équipe di vulcanologi è stata ricostruita la sequenza delle fasi eruttive del Vesuvio e i suoi effetti sulle strutture architettoniche dell'abitazione. Per rendere maggiormente chiaro il rapporto della ricostruzione con ciò che rimane oggi della casa, sono state realizzate delle sequenze animate in dissolvenza tra il modello virtuale e particolari elaborazioni tridimensionali di fotografie realizzate al momento dello scavo. Proiettando le fotografie su un modello tridimensionale congruente, è stato possibile, restituendogli la terza dimensione, realizzare anche dei veri e propri movimenti di macchina.

Come nel caso precedente, era stato stabilito un sistema di denominazione convenzionale dei file prodotti mediante la computer grafica e a tutti gli operatori coinvolti nella manipolazione dei file digitali veniva richiesto di aderire a tali convenzioni. Queste includono il nome del progetto, il nome dell'oggetto del file e il numero della versione. Mentre il lavoro progredisce, i file sono organizzati per l'immagazzinamento in *directory* e *subdirectory*. Il *workflow* stesso determina quindi il modo in cui le entità digitali sono organizzate nel passare da un gruppo di lavoro al successivo per la manipolazione e il trattamento.

Il prodotto finale è interamente digitale, ma alcuni artefatti analogici vengono creati in forma di immagini su carta, poi scansionate e inserite durante il processo produttivo. Anche in questo caso le procedure impiegate non vengono altrimenti documentate; ciascun processo di produzione di *moving image* o multimediali è talmente unico che molti produttori non ritengono sia utile lavorare attraverso procedure documentate, perché altri progetti successivi potrebbero avere dei requisiti differenti e richiedere procedure differenti.

L'autenticità delle entità digitali è conferita anche in questo caso dal numero di versione incluso nel nome del file. Se l'entità digitale può essere aperta dal software per la nuova fase del lavoro è considerata autentica e affidabile. A livello di sistema, un *backup* giornaliero costituisce la garanzia dell'affidabilità. Le entità digitali sono usate in molti modi durante la produzione e tutte le attività sono focalizzate nel creare il prodotto finale. Una volta che il lavoro è completato, l'uso delle entità prodotte dipende da uno specifico contratto con il cliente; la

loro proprietà legale è infatti anche un problema complesso, dato che anche le attività di promozione e di marketing costituiscono un altro uso delle medesime entità digitali.

Mentre i requisiti di qualità, accuratezza e autenticità sono presi in considerazione da Altair4 nello svolgimento del progetto, la conservazione a lungo termine degli oggetti digitali non è stata presa in considerazione. Trattandosi di una società di non grandi dimensioni, mentre viene curato l'aspetto del marketing e della promozione, non c'è una figura che si occupa dell'archiviazione del digitale con una preparazione specifica. Questo ruolo viene svolto a grandi linee da una segretaria che ha diversi compiti.

I file sono conservati su Cd-Rom e su dvd e in aggiunta c'è un *backup* globale, ma non ci sono particolari strategie di conservazione a lungo termine. La compagnia salva circa il 90% delle entità digitali create. La selezione viene fatta in base all'importanza delle entità, dato che non c'è un sistema apposito per archiviare a lungo termine le entità digitali, non vengono utilizzati particolari schemi o standard per metadati descrittivi, né c'è una specifica catalogazione o indicizzazione separata dei prodotti, affidandosi per il recupero agli ordinari meccanismi disponibili nel software.

### **Rete televisiva WGBH di Boston (USA)**

Il terzo studio considerato è quello della rete produttrice ed emittente televisiva pubblica WGBH di Boston (USA), che ha coinvolto specialisti provenienti da diversi settori disciplinari e da diversi paesi, presentato a Los Angeles nel febbraio 2004. La rete è in corso di trasformazione dall'analogico al digitale e così al momento si trova a operare in ambiente misto analogico-digitale, nel senso che il sistema digitale è già operativo ma l'archivio di film, nastri video e audio, risalente agli anni Cinquanta, è mantenuto in forma analogica.

Oggetto dello studio è stata la produzione di un programma televisivo documentario e in particolare le riprese originali e le rispettive note di accompagnamento. Le riprese originali sono effettuate in analogico all'interno degli *studios* e in esterni. Benché alcune delle riprese originali comprendano il suono, gli elementi sonori non sono inclusi nello studio in quanto vengono registrati separatamente. Il materiale in studio è usualmente registrato su videocassette con diversi formati di nastro o anche su pellicola. I materiali sono accompagnati da elenchi separati delle riprese che riportano informazioni come il soggetto del materiale visivo, l'ambiente, il contesto e il tempo di registrazione.

Le entità digitali successivamente prodotte sono individuate mediante un unico identificatore che collega il record del catalogo informatizzato delle entità digitali alle note relative alla ripresa originale. I record del catalogo sono strutturati gerarchicamente per serie, stagione e programma. A differenza dei casi precedenti,

queste informazioni non sono quindi registrate nel nome stesso del file, ma in un database separato. A differenza delle reti commerciali, questa rete si preoccupa inoltre della gestione e conservazione a lungo termine dell'informazione e investe un maggior impegno in questo campo, sviluppando standard interni per organizzare questi materiali.

Il sistema digitale può ricevere le riprese e le note originali in forma digitale e il materiale risultante può essere usato per scopi documentari ed editoriali. Gli utenti del sistema devono preventivamente registrarsi, in modo da garantire la sicurezza del materiale e da consentire al personale dell'archivio di controllare l'uso dei materiali stessi. Una volta terminata la produzione, i file digitali definitivi sono inviati all'archivio per il controllo di qualità e per il trasferimento finale al deposito digitale istituzionale.

In questo caso, anche i procedimenti di organizzazione dell'informazione sono documentati. Il servizio d'archivio fornisce al *team* di produzione addestramento e manuali di lavoro per documentare le procedure di completamento delle note di ripresa, e sono anche previste misure per assicurare la qualità, affidabilità e autenticità dei materiali digitali. Questi sono confrontati con i dati delle note per controllare l'accuratezza dell'informazione e per aggiungere informazioni utili a reperire la collocazione dei materiali. L'autenticità è garantita dal fatto che di regola le note possono essere lette, ma non modificate dagli utenti.

Il sistema digitale consente ad altri team di produzione di individuare le riprese e le scene e di utilizzarle in nuove opere e consente di costruire nuovi *storyboard* con immagini estratte dal sistema stesso. I metadati usati per gestire le immagini digitali possono essere modificati solo dal personale d'archivio, in modo da garantire un buon sistema di controllo. Un apposito sistema digitale monitorizza i cambiamenti nei file e può determinare se sono stati eventualmente apportati cambiamenti ad un file. È quindi da sottolineare che, a differenza delle reti commerciali, qui la creazione e la gestione dei metadati di controllo sono attività incorporate nel sistema, che riflette la diversa missione del produttore.

Nel sistema analogico di questa rete l'informazione prodotta, una volta terminato il ciclo, è inviata all'archivio. Il nuovo sistema digitale consente di gestire contestualmente sia le attività di produzione che quelle di archiviazione. I responsabili del sistema stanno elaborando procedure per archiviare correttamente e a lungo termine i materiali correnti, creando collegamenti a livello di singole scene fra i materiali analogici originali e quelli archiviati in forma digitale per un loro più facile reperimento, anche mediante un *thesaurus* apposito.

La conservazione dei materiali è un'attività a cui quest'istituzione dedica molta attenzione. La conservazione del materiale analogico si attua effettuando copie del materiale obsoleto o deteriorato su nuovi supporti digitali. Gli standard descrittivi dei metadati sono da tempo in uso in questa istituzione e corrispondono alle norme per le collezioni di immagini ferme e in movimento

adottate in tutto il mondo, ma alle volte vengono usati standard provenienti dal settore bibliotecario, che ha standard ben consolidati. Il nuovo sistema, cioè quello digitale, continuerà a impiegare personale con formazione professionale archivistica e bibliotecaria e includerà i metadati Dublin Core e la base di riferimento PB Core.

### Conclusioni

Come abbiamo rilevato nei precedenti resoconti, la cura dell'organizzazione e della conservazione dei materiali varia molto da un soggetto all'altro. Tuttavia la circostanza che emerge più chiaramente è che i produttori commerciali dedicano poche cure a organizzare il materiale in modo sistematico al di là di quanto è strettamente richiesto dalle esigenze di produzione. Una volta terminato il processo di produzione, vi è poi fra questi soggetti scarsa preoccupazione per la conservazione a lungo termine dei materiali, al di fuori dell'esigenza di poter provare legalmente il diritto d'autore su di essi. D'altra parte considerazioni a lungo termine di conservazione dei materiali come beni culturali esulano dai loro compiti, e da questo punto di vista la cosa sembra comprensibile: essi lasciano di fatto alle pubbliche istituzioni, che eventualmente acquisiscono i materiali da essi prodotti, il compito di conservarli a più lungo termine e pertanto non investono risorse particolari in questa attività. Inoltre, c'è da tener presente che i produttori non hanno molte ragioni di aspettarsi un adeguato tornaconto da tale conservazione, dato che la tecnologia di riproduzione di immagini digitali in movimento è in continua evoluzione e per il riutilizzo dei vecchi materiali conservati, richiederebbe la loro conversione ad hardware e software differenti da quelli usati per archivarli, sensibilmente costosa.

Tuttavia, si annunciano anche alcune tendenze a una maggiore preoccupazione circa l'archiviazione dei materiali nell'ambiente commerciale, dato che sempre più spesso emerge che questi possono anche essere riutilizzati per nuovi prodotti che non facevano parte della rispettiva precedente strategia di marketing. I motivi del profitto che ispirano in definitiva tutta l'attività dell'ambiente commerciale possono insomma determinare in misura anche significativa l'intrapresa di ulteriori attività di conservazione.

Le pubbliche istituzioni hanno, da parte loro, il dovere di investire risorse nello sviluppo di buone pratiche di conservazione dei *record* e di una gestione ottimale dei materiali di cui sono responsabili. Esse non dovrebbero essere astrette ai criteri del profitto e alcune hanno anzi l'esplicito compito della conservazione a lungo termine dei beni culturali e come parte essenziale di tali compiti si richiede che adottino standard e metodi adeguati a tal fine. Dato che questi ultimi hanno cominciato a essere sviluppati solo di recente nell'ambito delle immagini in movimento, attualmente e per un certo tempo prevarranno ancora sistemi

*ad hoc* diversi, anche se si riscontra una crescente pressione, imposta dal relativamente recente ambiente di Rete che ci avvolge tutti, ad adottare standard e metodi comuni.

In conclusione, il lavoro di InterPARES2 potrà gettare qualche luce sulle presenti e future pratiche di organizzazione dell'informazione digitale dinamica di diversi generi, incluse le immagini in movimento sulle quali ci siamo concentrati negli studi presentati, che naturalmente verranno contestualizzati con gli altri effettuati dal progetto InterPARES nelle sue conclusioni generali. I nostri risultati potranno così contribuire a delineare la situazione della produzione di immagini in movimento nel contesto più generale della problematica della produzione di documenti autentici e affidabili e della loro conservazione, e hanno fornito indicazioni per la loro conservazione a lungo termine, basate anche sui risultati di InterPARES1, che saranno oggetto di diffusione e più ampia discussione fra gli operatori del settore e potranno intanto già essere utili a chi vorrà cominciare a sperimentarle.

# La Biennale di Venezia verso un futuro in digitale

**Giovanna Pasini**

*Archivio storico delle arti contemporanee (ASAC) della Fondazione la Biennale di Venezia*

La Biennale di Venezia è da oltre un secolo una delle istituzioni culturali più prestigiose al mondo. Fin dalla sua origine (1895) è all'avanguardia nella promozione delle nuove tendenze artistiche, organizzando manifestazioni internazionali nelle arti contemporanee, secondo un modello pluridisciplinare che ne caratterizza l'unicità. Si colloca ai vertici mondiali sia per la Mostra internazionale d'arte cinematografica (63 edizioni), sia per l'Esposizione internazionalele d'arte (51 edizioni) e la Mostra internazionale di architettura (10 edizioni), e prosegue la grande tradizione dei Festival della musica (50 edizioni) e del teatro (38 edizioni), affiancati ora dal Festival della danza (4 edizioni).

Per mezzo dell'Archivio Storico delle Arti Contemporanee (ASAC), la Biennale tutela, conserva e valorizza il suo patrimonio documentale raccolto dal 1895 a oggi. L'ASAC è una struttura pluridisciplinare e multimediale e si articola in:

- *Fondo storico*: 3.000.000 di documenti.
- *Fototeca*: 600.000 positivi; 40.000 diapositive; 37.000 negativi; 28.000 lastre.
- *Cineteca*: 1.080 film.
- *Mediateca*: 8.282 video; 3.604 audionastri; 5.000 dischi sonori; 233 CD-Rom.
- *Collezione manifesti*: 3.100 tra manifesti e locandine.
- *Raccolta documentaria*: 1.500.000 documenti.
- *Biblioteca*: 127.700 tra libri e cataloghi.
- *Collezione periodici*: 3.000 titoli.
- *Collezione partiture e spartiti*: 4.000 tra partiture e spartiti.
- *Fondo artistico*: 2.515 opere (194 dipinti; 55 sculture; 547 bozzetti di scena; 916 stampe; 85 disegni; 531 fotografie d'artista; 39 progetti; 67 multipli; 14 plastici; 67 oggetti vari tra cui mosaici, piatti, vasi, calchi).
- *Archivio microfilm*: circa 100 microfilm; circa 6.000 microfiche.
- *Archivio progetti*: circa 200 contenitori di progetti e disegni.
- *Fondo e magazzino editoriale*: 135.000 tra monografie, cataloghi, depliant, brochure, volantini.

A partire dal 2000, durante la presidenza di Paolo Baratta, la Biennale mise a punto un progetto di recupero del suo archivio, che prevedeva anche la digitalizzazio-

ne di alcuni documenti scelti tra quelli più significativi e particolarmente richiesti dal pubblico e quelli ridotti in un precario stato di conservazione:

- 30.321 scatti (lastre, positivi, negativi, diapositive di cui 2100 riproduzione dei manifesti);
- 2.500 circa fotografie, relative al Fondo Artistico;
- 870 film riversati;
- 100 ore di materiale audio riversato su supporto digitale.

A questa descrizione del “fondo digitale” si devono aggiungere anche le riprese fotografiche digitali delle manifestazioni di tutti i settori di attività della Biennale realizzate dal 2003. Purtroppo attualmente risulta impossibile una loro precisa quantificazione, che si aggira intorno ai 25.000 scatti.

Nel 2005 Giorgio Busetto, direttore dell'ASAC dal 1 settembre 2004, decise di fare analizzare dalla Fratelli Alinari Edizioni artistiche e dal Dipartimento di ingegneria dell'informazione (DEI) dell'Università di Padova il patrimonio digitale dell'archivio. Tale indagine si rese necessaria per avere un quadro più chiaro sulla qualità del materiale a disposizione, anche in vista di una sua possibile commercializzazione, ma soprattutto per sapere come procedere per eventuali nuovi progetti di digitalizzazione. Il report di analisi, realizzato da Leonardo Lemmi della Fratelli Alinari il 10 maggio 2005, permise di evidenziare tale situazione:

«A seguito dell'incontro avvenuto in data 10 Maggio 2005 nella sede di Venezia si evince che l'archivio digitale della Biennale di Venezia è costituito da:

- n. 10.294 LASTRE (fondo L) acquisite nel formato 2000x3000 pixel a una risoluzione di 72 dpi; il fotoritocco risulta accettabile.
- n. 16.411 DIAPOSITIVE (fondo D) acquisite nel formato 800 pixel (lato lungo) a una risoluzione di 72 dpi; il fotoritocco per queste dimensioni non è determinante.
- n. 2.529 NEGATIVI (fondo N) acquisite nel formato 800 pixel (lato lungo) a una risoluzione di 72 dpi; il fotoritocco per queste dimensioni non è determinante.
- n. 788 STAMPE/POSITIVI (fondo F) acquisite nel formato 1400 pixel (lato lungo) a una risoluzione di 72 dpi; il fotoritocco risulta accettabile.
- n. 299 VARIE (la cui fonte analogica non è deducibile) acquisite nel formato 800 pixel (lato lungo) a una risoluzione di 72 dpi; il fotoritocco per queste dimensioni non è determinante.

Per un totale di n. 30.321 immagini digitali salvate nel formato PNG (lo standard è peraltro TIFF o JPEG) pari a uno spazio disco di 45,5 GB».

Tale disamina evidenziò che nessuna di queste immagini rispondeva agli standard ministeriali e solo le digitalizzazioni delle lastre avevano una dimensione appena sufficiente per consentirne anche una gestione commerciale; le restanti immagini potevano essere utilizzate solo per applicazioni video, ovvero Internet, presentazioni, *slide-show*, documenti elettronici. Ciò rispondeva in effetti agli obiettivi del progetto avviato nel 2000, che intendeva digitalizzare il patrimonio solo per renderlo fruibile in rete.

Il patrimonio conservato, come descritto all'inizio, è vasto ed eterogeneo, quindi risulta difficile programmare una sistematica, ma onerosa, digitalizzazione di tutti i fondi, soprattutto quando i finanziamenti sono sempre meno adeguati. Ciò nonostante tra novembre 2005 e febbraio 2006, seguendo il criterio del particolare e critico stato di conservazione dei documenti, si è scelto di passare in digitale i microfilm, prodotti da personale interno ad ASAC tra il dicembre 1990 e il febbraio 1994, dei 113 volumi della serie copialettere. Ogni volume consta di circa 500 documenti e pertanto i microfilm da digitalizzare comprendevano circa 56.500 fotogrammi. Tale intervento venne affidato alla Ditta Gallo Pomi di Padova, poiché assicurava:

- una digitalizzazione dei microfilm secondo gli standard elaborati dal Ministero per i Beni e le Attività Culturali nell'ambito del progetto della Biblioteca Digitale Italiana;
- una lavorazione effettuata manualmente da personale qualificato e non attraverso attrezzature d'acquisizione automatica, al fine di ottenere per ciascuna immagine la migliore resa;
- dimensioni di acquisizione dei documenti (regolabili attraverso zoom ottico), tenute all'interno dei limiti che consentono il massimo ingrandimento dell'immagine e la migliore leggibilità.

Pianificare un progetto di digitalizzazione ha senso solo se a questo si accompagna anche un piano di lavoro che preveda la gestione e l'accesso ai materiali digitalizzati. Infatti la creazione di un archivio digitale è normalmente di sprone per l'avvio di procedure di acquisizione in formato digitale del materiale contenuto nei fondi. Le digitalizzazioni fatte nel 2000 rientravano in un più vasto e articolato progetto, chiamato Sistema multimediale (SMM), per l'ideazione e la creazione di un sistema informativo unitario per la gestione degli eventi e delle attività culturali correnti, la loro documentazione, la loro archiviazione permanente e la loro fruizione tramite reti Intranet e Internet. Sempre tra il 2000 e il 2001 si stava realizzando anche un altro data base, la Mostra delle Mostre (in collaborazione con l'Ogilvy Group di Milano), con l'intento di rendere disponibile su Web tutte le informazioni possibili relative alle diverse edizioni della Biennale provenienti dai set-

tori come se si trattasse di un unico grande catalogo o di un'unica grande mostra. La prima fase del progetto prese in considerazione i dati delle Esposizioni internazionali d'arte sin dalla prima edizione del 1895. Il data base di inserimento era costituito da una scheda esposizione, una scheda artista e una scheda opera. Vennero inseriti 15.000 nominativi artisti partecipanti alla sezione Arti visive dalla prima Esposizione internazionale del 1895. Purtroppo, entrambi i progetti si arenarono nel 2002 con la fine dei finanziamenti.

Un'esigenza, più volte ribadita dallo stesso direttore Giorgio Busetto, è quella di non voler avviare nuove attività senza prima verificare la possibilità di recuperare quei progetti che già erano stati avviati per la soluzione del problema. Tra il 2004 e il 2005 si tentò quindi di riprendere entrambi questi lavori. Ciò si è dimostrato possibile per il database Mostra delle Mostre, che è stato recuperato e aggiornato (ad oggi sono stati inseriti dati relativi a dieci edizioni dell'Esposizione internazionale d'arte), mentre il progetto del Sistema multimediale non ha potuto essere ripreso. La proposta della ditta che aveva seguito i lavori sin dalla loro genesi di non utilizzare software *open source*, oltre alla necessità di stanziare, solo ed esclusivamente per questo progetto, risorse finanziarie non disponibili, hanno reso questa via impraticabile.

Nel frattempo si era stipulata una convenzione con il Dipartimento di ingegneria di Padova, grazie alla quale all'inizio del 2005 cominciò una fase di attenta e scrupolosa disamina del patrimonio digitale dell'archivio, disamina che naturalmente riguardò anche i diversi e frammentati data base. Tale indagine evidenziò la ricchezza del patrimonio di dati dell'archivio ma anche la fragilità dei suoi strumenti che ne impedivano la divulgazione. Oltre al Sistema Multimediale e al Mostra delle Mostre in ASAC si utilizzavano alcuni data base per la gestione del lavoro corrente delle seguenti raccolte:

- Catalogo dei Periodici correnti, cessati e rari.
- Catalogo della Mediateca (video, pellicole, audio, CD, CD-Rom).
- Catalogo del Fondo artistico (opere d'arte).
- Catalogo del Fondo editoriale (cataloghi e pubblicazioni della Biennale).
- Fototeca (le circa 30 mila immagini digitalizzate a campione dai fondi fotografici dell'ASAC).

Tale situazione però si presentava estremamente frammentata: esisteva una diversa base di dati per quasi ciascun fondo e ciascuna delle manifestazioni. Si trattava di fogli di lavoro o semplici tabelle inserite in programmi di videoscrittura. In altri casi, anche se veniva utilizzato un sistema di base di dati, le informazioni erano rappresentate da una singola tabella. Ciò era determinato dalla necessità di un rapido sviluppo del sistema per l'archiviazione di informazioni e dalla necessità di

privilegiare la semplicità di interazione da parte del personale dell'archivio. Ma un simile approccio presentava forti limiti:

- rischio di inconsistenza dei dati: le stesse informazioni dovevano essere riportate su sistemi diversi;
- impossibilità di effettuare delle interrogazioni complesse sul contenuto dei fondi e delle manifestazioni;
- impossibilità di rendere disponibili i dati on line. Non tutti i software utilizzati consentivano un accesso sicuro tramite interfaccia web.

Venne allora richiesto uno studio di fattibilità alla *spin-off* del Laboratorio di tecnologia e telecomunicazioni multimediali del Dipartimento di ingegneria dell'informazione dell'Università di Padova, la 3Deverywhere, che propose la creazione di una base di dati unificata realizzata con un software di tipo *open source* che garantisse maggior sicurezza e stabilità rispetto a software proprietari, secondo un programma di attuazione che prevedeva tre scadenze temporali nell'arco di tre anni per il completamento del progetto.

La realizzazione di un tale lavoro, impegnativo dal punto di vista delle risorse finanziarie e umane, era però assolutamente in linea con il programma di attività dell'ASAC, che dal 2004 era stato definito sulla base dell'esigenza prioritaria di avviare attività collegate alla valorizzazione, la tutela e la conservazione delle raccolte. Inoltre la realizzazione di un archivio digitale del materiale in possesso dell'ASAC si presentava particolarmente interessante per la tipologia e la varietà dei fondi e delle informazioni riguardanti le diverse manifestazioni della Biennale. Nel settembre del 2005 la prima fase del progetto ASAC-DATI ebbe quindi inizio.

Essenziali si sono dimostrate le numerose interviste effettuate a un ristretto numero di curatori di fondi, in seguito alle quali durante i primi incontri si sono gettate le basi di un linguaggio comune con il quale descrivere il campo di applicazione ed evidenziare le componenti fondamentali che accomunano i diversi fondi, come anche per spiegare le similarità, differenze e connessioni esistenti tra le diverse tipologie di informazioni gestite dall'archivio.

Durante gli incontri con il personale è inevitabilmente emerso, come uno dei primi problemi da affrontare, il diverso significato attribuito ad alcuni termini del linguaggio comune, vocaboli che invece nella terminologia adottata comunemente nel mondo delle basi di dati assumono una particolare accezione. Sulla base di quanto emerso dai colloqui vennero evidenziati alcuni requisiti di cui era necessario tenere conto nella progettazione della base di dati:

- Rappresentare il contenuto dei fondi in modo conforme alle schede ministeriali. Ciò costituisce un vincolo importante, che impone alcune scelte sugli attributi delle entità che contribuiscono alla definizione delle informazioni dei fondi.

- Evidenziare la peculiare eterogeneità dei fondi dell'archivio.  
Le interrelazioni possibili tra i dati contenuti nei fondi sono innumerevoli, e sono dovute principalmente al fatto che i medesimi artisti hanno avuto negli anni funzioni diverse, come ad es.: regista di un film premiato a un'edizione della Mostra del cinema, componente della giuria di un'edizione successiva, soggetto del materiale fotografico che correda uno dei suoi film, soggetto di fotografie scattate per documentare un evento della Biennale, e possibilmente anche coinvolto in altri tipi di eventi, come la Biennale musica o la Biennale teatro.
- Promuovere maggiormente l'immagine e le attività dell'ASAC. E la promozione migliore si ottiene proponendosi come fornitore di servizi nel Web, ovvero nel caso dell'ASAC mettendo a disposizione le informazioni sul contenuto dei fondi e degli eventi. In particolare che la memoria storica degli eventi promossi dalla Biennale rappresenta forse il patrimonio di maggior interesse per gli studiosi.

È apparso chiaro dal principio che il sito della Biennale di Venezia poteva soddisfare le esigenze degli utenti in cerca di informazioni dell'ultima ora o relative agli eventi temporanei organizzati di anno in anno, ma non poteva rispondere alle diverse esigenze dell'archivio. Era quindi necessario passare dall'aver una sezione del sito Biennale dedicata ad ASAC, al costruire un meccanismo di rinvio a quest'ultimo dotandosi di una serie di altri strumenti per la consultazione dei suoi fondi.

La realizzazione di ASAC-DATI ha anche costituito l'occasione per avviare una nuova campagna di digitalizzazione dei materiali della fototeca: tra aprile e ottobre del 2006 sono state digitalizzate 6.000 foto, e questa volta lo si è fatto seguendo gli standard ministeriali. Sono state quindi acquistate attrezzature adeguate a garantire tali standard — non solo nuovi scanner, ma anche una nuova macchina digitale per documentare le manifestazioni, anch'essa in regola con le norme ministeriali.

Inoltre, il nuovo deposito dell'ASAC, collocato presso il Parco scientifico e tecnologico Vega di Marghera e destinato a divenire operativo entro la fine di gennaio 2007, ospiterà i laboratori attrezzati per la digitalizzazione e la catalogazione automatizzata dei fondi.

La necessità di suddividere l'intero progetto in diverse fasi, vista la quantità, la varietà e l'eterogeneità dei materiali dell'archivio, ha determinato l'identificazione di fondi e manifestazioni che per primi sarebbero stati trattati. Ecco allora che la realizzazione di una base di dati per la gestione di Cineteca, Fototeca e di notizie relative alla Manifestazione del cinema, ha rappresentato la prima fase di questo importante e complesso progetto. Grazie anche al lavoro di nove catalogatori, tra i mesi di aprile e settembre del 2006, sono stati inseriti nel nuovo data base ASAC-DATI:

- 13.000 schede minimali di film (titolo del film, regia, anno, nazione) presentati alla Mostra del cinema dal 1932 al 2006. Inoltre, schedatura completa delle edizioni del 2004, 2005 e 2006.
- 30.000 schede di foto relative alla Mostra di cui 15.500 disponibili anche in formato digitale (6.000 digitalizzazioni da stampe e 9.500 native digitali dal 2003 al 2006).
- Dati minimali relativi a tutte le manifestazioni organizzate dalla Biennale nei suoi 111 anni di attività culturale (nome manifestazione, data, presidente, curatore, riproduzione digitale di tutti i manifesti generali dal 1895 al 2006).
- Catalogo completo della cineteca, con schede corredate di trailer degli 870 film digitalizzati.

ASAC-DATI è stato testato alla 63a Mostra internazionale d'arte cinematografica: i mille scatti giornalieri effettuati in tempo reale fatti dai fotografi Biennale venivano selezionati, catalogati e resi disponibili alla consultazione attraverso alcune postazioni. Nonostante l'interesse in quei giorni fosse giustamente rivolto alla manifestazione, molti giornalisti, critici e fotografi hanno avuto modo di interrogare la base di dati, dimostrandosi piacevolmente sorpresi da tale lavoro.

Entro dicembre 2006, ASAC-DATI sarà on line. Il suo varo è necessario anche a fronte della convenzione stipulata nel dicembre 2005 con il Ministero per i beni e le attività culturali al fine di partecipare al portale della cultura: da fine settembre sino a dicembre gli ingegneri di Padova lavoreranno con gli ingegneri di Pisa per rendere interoperabili i due sistemi.

Giunti a questo punto del progetto, occorrerà decidere circa il futuro del data base Mostra delle Mostre: si tratterà di stabilire se continuare ancora per qualche tempo su binari distinti e collegati eventualmente solo da un semplice link, o cercare piuttosto di sfruttare al massimo il lavoro fatto in passato per l'analisi e la realizzazione dello schema di questa base di dati, facendola confluire in ASAC-DATI.

ASAC-DATI è un progetto complesso, meticoloso e per certi aspetti interminabile, ma reso necessario proprio dalla chiusura dell'ASAC. ASAC-DATI garantisce infatti la valorizzazione dell'archivio, le cui collezioni diverranno accessibili da qualsiasi luogo e in qualsiasi momento, ed è inoltre un valido strumento per la gestione e, di conseguenza, la conservazione delle sue preziose raccolte. Un lavoro che, come è facile immaginare, necessita di un grande sforzo economico e che si scontra con la sempre più difficile reperibilità delle risorse. Cionostante, a un anno di distanza dall'avvio del progetto, sono stati individuati altri fondi dell'archivio (manifesti, mediateca, spolio periodici) e si è dato il via all'analisi e quindi alla realizzazione della seconda parte del progetto.

# La raccolta dei siti web: un test per il dominio “punto it”

**Giovanni Bergamin**

*Biblioteca Nazionale Centrale di Firenze*

L'uscita della nuova legge sul deposito legale (106/2004) è stata subito seguita da una vivace discussione soprattutto per quanto riguarda la lettera r) dell'art. 4: l'estensione del deposito legale ai «documenti diffusi tramite rete informatica». Le differenti posizioni che si sono confrontate possono essere riassunte con tre aggettivi: impossibile, inutile, civile.

Il primo aggettivo – impossibile – si riferisce alla posizione che sostiene l'impossibilità tecnica di effettuare un deposito di oggetti che per loro natura non sono legati a un determinato supporto, come ad esempio un sito web. Secondo questa posizione la consegna alla biblioteca depositaria su CD o su altro supporto di copie di siti web risulterebbe essere del tutto impraticabile: alti costi di gestione per avere risultati del tutto discutibili.

Il secondo aggettivo – inutile – si riferisce invece ai contenuti del web. Secondo i sostenitori di questa posizione il web è il mondo dell'effimero e quindi assolutamente non paragonabile alle pubblicazioni tradizionali. Anche se tecnicamente esistono strumenti per il deposito legale del web, l'operazione avrebbe uno scarso risultato dal punto di vista culturale.

L'ultima posizione si riferisce invece al fatto che l'estensione del deposito legale al mondo del digitale è da considerarsi una grande conquista di civiltà: il prendere atto che oggi la conoscenza si trasmette anche attraverso i bit. In questo senso vanno le raccomandazioni dell'UNESCO del 2003 e il recente adeguamento della legislazione in molti paesi<sup>1</sup>.

Tenendo presente che nella categoria “documenti diffusi tramite rete informatica” oggi rientrano tipologie molto differenti di risorse digitali (dal periodico elettronico con contenuti scientifici “certificati” al sito web casalingo) questo intervento intende presentare i risultati di una sperimentazione finanziata nel contesto della Biblioteca Digitale Italiana a partire da una discussione del tutto provvisoria e aperta sui presupposti e sui fondamenti delle posizioni sopra richiamate.

La prima posizione molto diffusa deriva da una scarsa conoscenza della natura della risorsa digitale. Nel mondo del digitale, l'informazione non è più legata a un determinato supporto. Si parla di “rappresentazione informatica” ovvero di «una sequen-

<sup>1</sup> *Charter on the Preservation of the Digital Heritage/UNESCO*, agg. 2006  
[http://portal.unesco.org/ci/en/ev.php-URL\\_ID=13366&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=13366&URL_DO=DO_TOPIC&URL_SECTION=201.html).

za di bit, che, elaborata da un sistema informatico, può essere resa visibile su uno schermo, stampata sulla carta o inviata a distanza». Si tratta di «un cambiamento radicale nella concezione e nell'uso del documento, così come lo conosciamo da migliaia di anni, nella sua natura di *res signata*, cioè di una cosa che riporta dei segni, delle informazioni». Più in generale la "risorsa digitale" non dipende quindi necessariamente da un supporto, ma assume «una funzione autonoma rispetto alla sua (eventuale) fissazione su un supporto materiale»<sup>2</sup>. Per "risorsa digitale" si intende qui fare riferimento a un insieme di bit dove «informazioni di tipo diverso possono essere tutte ridotte allo stesso codice di base, alle lunghe *catene di 0 e di 1* dell'informazione digitalizzata»<sup>3</sup>: la posta elettronica, le pagine web, la musica su CD, i film su DVD sono esempi di risorse digitali che negli ultimi anni hanno profondamente modificato la nostra vita e le nostre abitudini. Non senza qualche difficoltà, il concetto di documento come "rappresentazione" – non legata a un particolare tipo di supporto – è entrato anche nella legislazione: ad esempio nei meccanismi che regolano la firma elettronica il supporto fisico non è mai determinante.

Del resto la Legge 106/2004 non entra nel merito delle tecnologie utili al deposito. L'art. 5, comma 5 lettera g) rinvia al regolamento per la definizione di «speciali criteri e modalità di deposito» per i documenti h) [manifesti] q) [su supporto informatico] r) [diffusi tramite rete informatica]. Ed è evidente che con questa precisazione il termine deposito non va inteso in senso letterale (fisicamente portare un oggetto in un determinato luogo) ma nel contesto del "deposito legale" come "istituto" dove le procedure operative devono tener conto della natura dell'oggetto.

La seconda posizione si basa sulla considerazione che il web è dominato dalla immediatezza e dalla quantità. La mancanza di mediazione che invece caratterizza l'editoria tradizionale (scelte editoriali, *peer review*, ecc.) mette a serio rischio la qualità di quanto viene messo in rete: si tratta di un *mare magnum* dove occorrerebbe procedere con una accurata selezione e non con un deposito indiscriminato.

La selezione tuttavia presenta innegabili aspetti problematici. Nel contesto del "deposito legale" le biblioteche nazionali hanno sempre cercato di assicurare la più ampia copertura possibile riducendo al minimo i rischi e i costi collegati alla selezione. In ogni caso la scelta dovrebbe essere fatta con criteri pubblici e oggettivi (non dipendenti dalle convinzioni e dai pregiudizi di chi sceglie). La formalizzazione dei criteri di selezione fa subito emergere contraddizioni di non facile soluzione. Se ad esempio un criterio di selezione è il genere, potremmo escludere dal deposito legale tutti i weblog: ma con que-

<sup>2</sup> Manlio Cammarata – Enrico Maccarone, *Introduzione alla firma digitale: 9, La natura del documento informatico*, 2000, <http://www.interlex.it/docdigit/intro/intro9.htm>.

<sup>3</sup> Fabio Ciotti – Gino Roncaglia, *Il mondo digitale. Introduzione ai nuovi media*, Roma [ecc.]: Laterza, 2000, p. 348.

sta decisione perderemo la possibilità di archiviare diari in rete di autori citati anche nella letteratura scientifica “certificata”<sup>4</sup>.

A livello internazionale il Consorzio fra biblioteche nazionali e Internet Archive (International Internet Preservation Consortium – IIPC)<sup>5</sup> è partito nel 2002 proprio dalla terza posizione (l’estensione del deposito legale al digitale come conquista di civiltà). La convinzione del Consorzio è che lo strumento dell’*harvesting* (della raccolta dei siti web) sia una tecnologia oggi in grado di far diventare il deposito legale dei siti web una attività sostenibile con risultati misurabili. Come è noto l’*harvesting* viene usato dai motori di ricerca (es. *Google*) per indicizzare il web; ma viene usato anche da oltre 10 anni per “l’archiviazione del web” da parte di Internet Archive<sup>6</sup>. Esistono ormai da tempo attività di *harvesting* portate avanti da biblioteche nazionali.

L’*harvesting* ha ovviamente anche “controindicazioni”: con l’*harvesting* si ottengono “fotografie a intervalli di tempo” di un determinato spazio web (con la conseguente perdita degli intervalli); tutto il cosiddetto “web profondo” rimane inaccessibile all’agente software che si occupa dell’*harvesting* (*crawler*). In altre parole l’*harvesting* non è una tecnologia in grado di risolvere tutti i problemi di deposito legale del digitale in rete, ma una tecnologia in grado di offrire un’ampia base di risultati.

Occorre ricordare che i *crawler* di nuova generazione permettono di impostare delle regole di priorità nella raccolta per evitare ad esempio che un sito di una agenzia di viaggi venga raccolto con la stessa frequenza di un sito di una università. In altre parole anche per quanto riguarda l’*harvesting* è possibile impostare regole di selezione che riguardano essenzialmente:

- la definizione – sulla base di determinati criteri – della lista di indirizzi di partenza (URL) chiamati anche “semi” del *crawler*. Nel caso del deposito legale italiano i semi di partenza potrebbero essere dati da tutti i domini “punto it”;
- la definizione di regole con le quali i semi di partenza si accresceranno con nuovi semi durante una sessione di *harvesting* (ad esempio gli indirizzi di altri siti trovati nella pagina raccolta, ma non quelli che non sono “punto it”);

<sup>4</sup> Il progetto Pandora della Biblioteca nazionale australiana si basa proprio sulla selezione operata dal bibliotecario con i criteri definiti in <http://pandora.nla.gov.au/selectionguidelines.html>.

Attualmente la stessa Biblioteca sta parallelamente sperimentando anche l’*harvesting* dei siti web.

<sup>5</sup> <http://netpreserve.org>.

<sup>6</sup> <http://www.archive.org>.

<sup>7</sup> Per convenzione si parla di *deep web* con riferimento a siti non raggiungibili dai tradizionali motori di ricerca (e quindi non raggiungibili nemmeno da un *crawler*). Tra questi si indicano di solito: siti non accessibili liberamente (per esempio a pagamento e protetti da password); siti che presentano le pagine HTML come risultato di una interrogazione da parte di un utente (per esempio la ricerca in un catalogo dove l’utente inserisce il titolo, l’autore, ecc., di un libro). Nel primo caso l’*harvesting* potrà funzionare solo se il sito “aprirà le porte” al *crawler* (per esempio fornendo la password all’istituzione depositaria). Nel secondo caso occorrerà una forte collaborazione tra il produttore dell’informazione e la biblioteca depositaria. Non è ovviamente pensabile che l’istituzione depositaria installi e mantenga tutti i database e tutte le applicazioni che generano le pagine HTML. Ci sono sperimentazioni a questo proposito (Francia e Australia) di invio alla biblioteca depositaria di record esportati in formato XML da database che “alimentano” il *deep web*.

- la definizione della frequenza di raccolta (quante volte in un determinato periodo di tempo si desidera che quella pagina venga raccolta).

In generale si può dire che la definizione degli insiemi di partenza (i semi) per l'harvesting sono il risultato di interrogazioni su archivi esistenti: ad esempio è possibile estrarre da Internet Archive i siti "punto it". Inoltre la possibilità di impostare la frequenza della raccolta (ogni quanto tempo si ritorna sullo stesso sito) è una caratteristica che differenzia i *crawler* di nuova generazione da quelli storici. Oggi non si parla più di "istantanee periodiche" dello spazio web ma di "raccolta continua" dove il *crawler* è in grado dinamicamente di gestire i ritorni a partire dalle "priorità" che vengono date in input. I criteri di priorità sono comunque liste di siti (modificabili dinamicamente nel corso della raccolta) basati su criteri "oggettivi" quali:

- le pubblicazioni di fonte pubblica (Stato, Regioni, Enti locali, ecc.);
- la produzione scientifica delle università e dei centri di ricerca;
- criteri usati dai motori di ricerca come la frequenza di aggiornamento del sito, i link in entrata (numero di siti che citano un determinato sito) ecc.<sup>8</sup>

La Biblioteca Nazionale Centrale di Firenze nel corso del progetto Crawler – finanziato dalla Biblioteca Digitale Italiana – ha dato vita a una prima sperimentazione su larga scala di raccolta dello "spazio web nazionale". In collaborazione con Internet Archive nel corso di quattro settimane (tra maggio e giugno) sono stati raccolti 7.22 terabyte di dati in formato WARC<sup>9</sup> (oltre i 2 milioni i server contattati per circa 240 milioni di documenti raccolti). Il software usato per il *crawler* è Heritrix<sup>10</sup> (open source prodotto da IIPC). Punto di partenza sono stati 648.255 siti "punto it": la lista dei "semi" è stata prodotta da Internet Archive sulla base delle sue raccolte precedenti. Naturalmente i siti "punto it" non esauriscono lo "spazio web italiano" (come è noto molti siti italiani sono registrati come "punto com", "punto net", ecc.), ma questa scelta è stata un compromesso inevitabile. A parte la problematicità della definizione di "spazio web italiano", occorre ricordare che questa definizione va poi tradotta in istruzioni che permettono di creare automaticamente la lista di partenza. Ad esempio se si definisce come appartenente allo "spazio web italiano" un sito che contiene uno o più documenti in lingua italiana, per realizzare la lista dei "semi" di partenza si potrebbe analizzare – tramite un software di riconoscimento della lingua – tutti i documenti presenti su Internet Archive (a oggi intorno ai mille terabyte).

<sup>8</sup> Julien Masanès. *Towards continuous web archiving*, «D-Lib Magazine», 8 (2002), n. 12., <http://www.dlib.org/dlib/december02/masanés/12masanes.html>.

<sup>9</sup> Si tratta di un formato che rappresenta l'evoluzione a cura di IIPC del formato ARC di Internet Archive e che sta per essere standardizzato in ambito ISO: <http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>.

<sup>10</sup> Informazioni sul crawler Heritrix: <http://crawler.archive.org/>.

Naturalmente il progetto Crawler deve essere visto come un punto di partenza. Con questo progetto siamo in grado di fare una prima stima dei tempi e dei costi per l'*harvesting* italiano: un importante e concreto contributo per la sperimentazione che il nuovo Regolamento sul deposito legale prevede per quanto riguarda i "documenti diffusi via rete informatica"<sup>11</sup>.

<sup>11</sup> DPR n. 252 del 3 maggio 2006, art. 37.

## Documenti



# Convegno internazionale: *Conservare il digitale.* *Un confronto internazionale*

**Asolo, 29 settembre 2006**

**Laura Ciancio**

ICCU

**C**onservare il digitale: un confronto internazionale. Questo il tema del convegno che si è svolto ad Asolo il 29 settembre 2006, nella cornice dello storico Castello, residenza della regina Caterina Cornaro, luogo suggestivo per la felice ispirazione del lavoro di Pietro Bembo, *Gli Asolani*, e per il teatro recentemente ricostruito secondo l'impianto originario, che porta il nome della più moderna e popolare asolana, Eleonora Duse. La struttura interna del teatro antico, smontato per far posto a una sala cinematografica negli anni Trenta, fu venduto in antiquariato e ricostruito in Florida, a Sarasota nel Ringling Art Museum.

Il convegno è stato organizzato dalla ULSS 8 del Veneto in collaborazione con il Politecnico di Milano per presentare una importante realizzazione, l'Archivio unico della ULSS e la gestione documentale integralmente realizzata in digitale, e occasione di incontro per esperti internazionali e operatori archivistici della pubblica amministrazione, per mettere a confronto le diverse esperienze. È stato preceduto dai saluti di Gino Redigolo, direttore generale dell'ULSS 8, e di Giustiniana Migliardi O'Riordan, sovrintendente archivistico per il Veneto, che ha voluto nel suo intervento di benvenuto presentare il punto delle attività e degli interventi della sovrintendenza regionale. L'introduzione al convegno "Scenari di frontiera del digitale", accompagnata da una ricca sequenza iconografica, è stata tenuta da

Kim Henry Veltman, direttore scientifico dell'Istituto McLuhan Maastricht e Network Europeo dei Centri di eccellenza del patrimonio culturale digitale. Ne risulta una visione globale dei temi delle culture occidentale e orientale, il ripetersi di simboli e significati tra le diverse culture per sottolineare come Internet, in pochi anni, abbia velocemente elaborato linguaggi nuovi e che se si intende creare una memoria culturale del pianeta bisogna preoccuparsi di riuscire a conservare un documento due-tremila anni.

La sessione antimeridiana è stata presieduta da Paola Carucci, dell'Archivio storico della Presidenza della Repubblica, che ha brevemente messo in luce alcuni temi del digitale dal punto di vista dell'archivista, evidenziando la distanza tra normativa di riferimento e attualità del digitale, dalla creazione stessa del documento alla sua conservazione; le contraddizioni tra le norme del Garante della privacy e la necessità di mantenere banche dati stabili; la questione del deposito archivistico oggi previsto dopo i quarant'anni, inimmaginabile per i dati digitali.

Mariella Guercio, docente di archivistica e di gestione informatica degli archivi dell'Università di Urbino, è componente dei gruppi di lavoro che hanno prodotto la recente normativa sul trattamento informatico dei documenti della pubblica amministrazione; è coordinatrice del *team* italiano della ricerca internazionale InterPARES (International Research on Permanent Authentic

Records in Electronic Systems) sulla conservazione permanente dei documenti elettronici autentici. Espone le tematiche del digitale per il documento archivistico e i nodi organizzativi: da una parte la necessità di definizioni certe per procedure, regole, responsabilità specifiche e l'occorrenza di certezze giuridiche nella produzione del documento archivistico digitale; dall'altra conservare e rendere leggibili a distanza di tempo i documenti e gestirli correttamente, controllarli e conservarli con qualità, per ritrovarne il ruolo originario. La certezza della qualità è affidata alla responsabilità dell'amministrazione ma anche alle realizzazioni dei produttori di software, per questo è necessaria una campagna di sensibilizzazione, ragionando in termini di formazione, produzione di manuali di documentazione e realizzando infrastrutture per la conservazione, luoghi fisici e responsabilità mirate a questo fine. Di fatto, ancora oggi, gli archivi di Stato non conservano un solo byte ma, a breve termine, sarà necessario predisporre reti di strutture intermedie, quali ad esempio, poli archivistici regionali.

Luciana Duranti, docente della School of Library, Archival and Information Studies dell'Università della British Columbia di Vancouver, InterPARES Project Director, propone un quadro della ricerca internazionale. I progetti oggi propongono una identificazione condivisa dei problemi attraverso una visione multi/inter/trans-disciplinare, generale e globale. A questo consenso all'identificazione dei problemi è ugualmente chiaro che le risposte sono molteplici e legate ai diversi contesti culturali, ecc. Perciò i problemi vengono suddivisi e le soluzioni adattabili alle diverse circostanze. Le idee condivise dai gruppi di ricerca sostengono che il processo della conservazione nasce già alla produzione del documento stesso ed è un processo continuo; non è possibile conservare un documento digitale, ma è possibile conservare la capacità di riprodurlo e l'unico modo di conservare questo documento è fare una copia autentica della sua ultima manifestazione attiva; la conservazione

digitale deve consentire nel tempo l'accesso al documento autentico. Luciana Duranti individua aree problematiche in parte non ancora risolte, come il mantenimento in condizioni stabili di documenti dinamici e interattivi e la conservazione a lungo termine, ovvero decenni e secoli; espone ancora i progetti in corso e la ricerca da sviluppare: modelli economici (costi e organizzazione), modelli etici (scelte di metodo e conseguenze per singoli e società), modelli di valutazione (misurare il successo delle soluzioni proposte e attivate). Alla fine del suo intervento ha annunciato che è stato approvato il finanziamento a InterPARES3.

Micaela Procaccia della Direzione generale per gli archivi del Ministero per i beni e le attività culturali interviene con una relazione dal titolo *Il quadro normativo italiano*, un panorama della legislazione italiana e dei principali provvedimenti adottati in materia di gestione informatica dei documenti, ponendo l'accento sul ritardo dell'intervento del legislatore in una materia che diventa sempre più difficile da gestire e contraddittoria, come sottolineato anche da Paola Carucci.

Conclusi gli interventi di carattere generale si sono susseguiti interventi relativi a esperienze specifiche: April Miller della Banca mondiale *Documenti in pericolo: imparare dagli eventi avversi*, Chen Bitan della Cyber-Ark software di Tel Aviv *Gestione e protezione dei contenuti medici digitali*.

Irmgard Schuler della Biblioteca Apostolica Vaticana ha esposto le *Problematiche e prospettive della digitalizzazione dei manoscritti. L'esperienza della Biblioteca Vaticana*. Ha illustrato un progetto, realizzato dall'Università di Rio de Janeiro e dall'IBM, di digitalizzazione di 150 manoscritti per un totale di 55.000 immagini realizzate con scanner in formato TIFF e archiviate su nastro, che con gli anni ha mostrato una perdita di dati a causa della corruzione del 5% dei nastri. Il progetto sui codici ebraici prevede la digitalizzazione con dorso digitale Leaf da 33 megapixel e l'archi-

viazione su hard disk in tre formati: DNG, TIFF, JPEG. Un altro progetto attende all'acquisizione in digitale con scanner planetario di 400 codici (i più preziosi e delicati) e infine il Progetto Cicero (from Cooperation to Integration in our Common Europe) si occuperà dei palinsesti. Qui verranno usati scanner e le immagini saranno duplicate, una a luce naturale, l'altra a luce UV, inoltre è stato predisposto un software di analisi delle scritture che utilizza contestualmente le due immagini elaborando i contenuti testuali. Per la conservazione dei dati digitali è in progetto la creazione di un servizio di *storage* da 50 terabyte. Seamus Ross, Visiting Fellow all'Oxford Internet Institute, professore di Humanities Informatics and Digital Curation, direttore di HATII (University of Glasgow), e direttore associato del Digital Curation Centre, è intervenuto sul ruolo delle reti di collaborazione e cooperazione, intese a produrre consapevolezza, orientamento e formazione, indirizzate alla ricerca. Ross presenta gli obiettivi e le finalità di ERPANET (Electronic Resource Preservation and Access Network), DELOS DPC (Digital Preservation Cluster), DCC (Digital Curation Centre), DPE (Digital Preservation Europe), descrive le reti come ambienti collaborativi, capaci di facilitare gli sviluppi della ricerca altrimenti non realizzabili, di creare comunità di soggetti operanti, di promuovere la crescita della consapevolezza e di una informazione distribuita, infine servono all'eliminazione della ridondanza degli sforzi.

Nel pomeriggio sono proseguiti gli interventi con l'esperienza di Nitya Karmakar, della School of Management della University of Western Sydney, con *Progress in digital preservation: an Australian experience*; Alfredo Ronchi, del Politecnico di Milano e di EC Medici Framework Secretariat, sulle *Fragilità del mondo digitale: rapporto sul progetto 1999-2005 Conservazione a lungo termine del digitale*; Guenther Kolar, di Ars Electronica di Linz, sull'attualità dei prodotti digitali; Mathias Jehn, della Deutsche Nationalbibliothek di

Frankfurt am Main, ha parlato delle *Attività e processi nella salvaguardia dei contenuti digitali nel lungo termine*, presentando i tre progetti della Deutsche Nationalbibliothek: Nestor (Network of Expertise in Long-Term Storage of Digital Resources), Kopal (Kooperativer Aufbau einer Langzeitarchivs digitaler Informationen) e Persistent Identifier.

Stefano Pigliapoco, docente dell'Università di Macerata, interviene con *Depositi digitali: requisiti e prospettive*. Ha parlato di complessità di procedure e della necessità di costituire, come soluzione più realistica, Centri di Conservazione Digitale (CeCoDi) o depositi digitali, cioè strutture dedicate alla conservazione delle memorie digitali di più soggetti produttori, secondo il modello OAIS. Ne descrive la struttura, l'organizzazione, la regolamentazione dei rapporti tra soggetti produttori e deposito, ecc.

Nel pomeriggio poi sono stati presentati progetti in corso: Rossella Bonora per la Provincia di Bologna, ha presentato il Progetto DocArea; Mario Pò, Annamaria Brosolo, *Il deposito digitale clinico della ULSS di Asolo*; Loreno Zandri del Comune di Pesaro-Urbino ha parlato del *Registro delle imprese della Camera di Commercio di Pesaro-Urbino*; Elettra Cappadozzi del CNIPA ha parlato degli *Archivi di Stato civile*; Giangiulio Radivo, delegato della Diocesi di Roma presso Radio Vaticana, è intervenuto con *L'archiviazione digitale tra storia e conservazione: il Tabularium Urbis*.

Per un approfondimento dei temi trattati si segnala il sito <http://www.ulssasolo.ven.it/> dove, alla voce convegni, sono presenti gli interventi dei relatori in .pdf e .mp3 con le registrazioni.



ICCU

### **Pubblicazioni recenti**

Catalogo delle biblioteche d'Italia: Lombardia (2004), 4 vol..  
Euro 150,00 ISBN 88-7107

Futuro delle memorie digitali e patrimonio culturale (2004)  
Euro 50,00 ISBN 88-7107-7

Memorie digitali: rischi ed emergenze (2005)  
Euro 15,00 ISBN 88-7107-111-5

Linee guida per registrazioni d'autorità e di rinvio (2005)  
Euro 20,00 ISBN 88-7107-105-0

Linee guida per la digitalizzazione del materiale fotografico (2005)  
Euro 15,00 ISBN 88-7107-112-3

EAD. Descrizione archivistica codificata: dizionario dei marcatori (2005)  
Euro 20,00 ISBN 88-7107-115-8

MAG: metadati amministrativi e gestionali. Manuale d'uso (2006)  
Euro 40,00 ISBN 88-7107-117-4

Linee guida per la digitalizzazione del materiale cartografico (2006)  
Euro 15,00 ISBN 88-7107-116-6

Inter omnes: contributo allo studio delle marche dei tipografi e degli editori italiani del XVI secolo (2006)  
Euro 100,00 ISBN 88-7107-114-X

Le edizioni italiane del XVI secolo: censimento nazionale – Vol. 5: D (2006)  
Euro 100,00 ISBN 88-7107-113-1

Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti (2006)  
Euro 25,00 ISBN 88-7107-118-2

Per l'acquisto delle edizioni ICCU rivolgersi a:

*ICCU Ufficio vendite*  
Viale Castro Pretorio 105  
00185 Roma  
tel: 06 4989468  
fax: 06 4959302  
e-mail: [venditapubl@iccu.sbn.it](mailto:venditapubl@iccu.sbn.it)

*Editrice Bibliografica*  
Via Bergonzoli 1/5  
20127 Milano  
Tel: 02 28315996  
Fax: 02 28315906  
e-mail: [bibliografica@bibliografica.it](mailto:bibliografica@bibliografica.it)

Sito web di *Digitalia*:  
<http://digitalia.sbn.it/>