Perspectives and policy of Italian Working Group on the standardisation and the use of metadata within cultural Institutions[1].

by

Cristina Magliano
Central Institute for the Union Catalogue of Italian Libraries
and for Bibliographic Information
Rome - Italy

I want to describe the results work on the metadata activities of Italian Metadata Group instituted by ICCU.

Many institutions in Italy are working to improve public access, teaching and researching about heritage information. Archivists, Librarians and Museum professionals are deeply involved in creating digital resources to access and understanding to their collections. Unfortunately in the digitalisation process the application of metadata is lacking and misunderstood.
The principles and practice in the digitalization of primary textual and image resources are important skills but at the same time the issue in creating metadata for the access to the reproduced collections, is an added value.

In particular some experiences has been just started between ICCU and the Central Institute for the catalog and documentation regard to the photograph material. In that project we have used the UNIMARC format to produce a minimum data set for interoperability and cross searching.
Consequently ICCU to continue the cooperative work and for the national co-ordination for the use of metadata in the description of digital objects, in the different projects of cultural heritage established a working group which is made up of the most important private and public Institutions of Libraries, Museums and Archives
The purposes of the group are:

- To hold an enquiry and to create an observatory on metadata use in Italian digital projects
- To start initiatives with the view to the interoperability of the different Institutions
- To co-ordinate the uniform application of the "metadata set" providing with libraries the production of the Guidelines and Standard specifications.

The integrated access to heterogeneous databases (bibliographic, museum, digital) is one of the key areas of the Fifth Framework programme for Research and Technological Development Cultural heritage institution that should work in order to find out tools and standards not only for sharing their information resources, but also for facilitating integrated access to the end users. Integrated access therefore implies an high level conceptual mapping enabling users to identify and to obtain all document pertaining to certain subject, irrespective of their support, ownership, geographical location etc.
The Central Institute for the Union Catalogue of Italian Libraries and for Bibliographic Information (ICCU) has been started some activities in the use of metadata within the Digital Libraries.

[1] Take part to the working Group: Simona Gatta, Gloria Cirocchi (Biblioteca della Camera dei Deputati) – Antonella Mulè (Ufficio Centrale Beni Archivistici) – Gianfranco Crupi (Biblioteca A. Monteverdi, Univ. La Sapienza) – Paul Gabriele Weston (Biblioteca Apostolica Vaticana/ Università degli studi di Pavia) – Claudio Leombroni (Provincia di Ravenna - Rete Bibliotecaria di Romagna) – Giovanni Bergamin (Biblioteca Nazionale Centrale di Firenze) – Marco Lattanzi, Matilde Amaturo (Istituto Centrale per il Catalogo e la Documentazione) – Francesco Baldi (Discoteca di Stato) – Antonio Scolari (Centro Servizio bibliotecario della fac. di Ingegneria Univ.- Di Genova) - Maria Bruna Baldacci (Istituto di elaborazione dell'Informazione. Area della Ricerca - CNR di Pisa) – Maurizio Messina (Biblioteca Nazionale Marciana, Venezia) –, Cristina Magliano, Patrizia Martini, Claudia Parmeggiani, Donatella Roveri (ICCU)

The implementation of the Dublin Core in the SBN-ON LINE System, is finalized on the integrated research among data bases in the different sectors of the national heritage (Bibliographic, Digital and of Museums), within the European project ONE-2 and according to the international Z39.50 specification "Bath Profile", where has been quoted the SBN on line project as a new on-line services for the Italian Union Catalogue users to search and retrieve in a integrated and simultaneous way on databases belonging to different application domains, like libraries, museums, digital collection using Dublin Core access points.
The same obiectives of interoperability among different sectors of the national heritage belongs to the european project DELOS Noe and to TEL, the European Library; the latter will investigate how to make a mixture of traditional and electronic farmats available in a coherent manner to both local and remote users.

In this regard, the ICCU initiative hope to get an agreement among Italian digital library projects.

In fact the first activities of the Group have been:

1. to produce a questionnaire in regard to the implementation of metadata in digitalization project in the different cultural assets. This will allow to produce registries of project profiles as a tool for harmonizing metadata usage in Italy. (www.iccu.sbn.it/quesmeta.htm)
2. comparing the data element, profile , format and standard of bibliographic description among different organizations of culturale heritage
3. evaluation of administrative information.

Survey

According to European initiatives (see Dublin Core Metadata) it has been produced a questionnaire (http://www.iccu.sbn.it/quesmeta.htm) and collected data about Italian projects (http.//www.iccu.sbn.it/metaprog.htm)-
The first results present a high percentage for access use and preservation in regards to administrative data. The mostly used type of metadata result Dublin Core followed by MARC and Library of Congress.

Descriptive Metadata

The work has started from the mapping of the core elements of Dublin Core.
First It has been analysed the mapping produced by ICCU for the One2 project and the crosswalks just published in Internet:
www.getty.edu/gri/standard//intrometadata/3_crosswalks/index.htm
http://www.loc.gov/marc/marc2dc.html
http://www2.sub.uni-goettingen.de/metaform/crosswalks.html
http://www.schemas-forum.org/registry/registry.html

In the perspective to produce a flexible schema within Italian metadata community and to analyze additional element according with the IFLA study "Functional requirement for bibliographic record", this encoding schema should provide information able to distinguish a resource from similar resources and support and carefully maintain name authority files and controlled vocabulary lists.
The analisys will define a minimum set for the identification of a digital object relating to the services to offer to the users.
Practical guidelines for expressing the schema types for web object are being developed in contexts such as the Dublin Core Metadata Initiative and the EU SCHEMAS Project, but these guidelines will surely evolve with implementation experience.
The working group has begun to map profiles, standard and format used in particular application or projects. The mapping has been done among bibliographic, archival, artistic data and official publications. The definition of a particular element has been compared with a definition in the other schema. For example in an application profile the "Title" element of the Dublin Core in different namespaces or profiles has a different semantic value.

For example, a profile for an application for describing archival unit might be redefined the Dublin Core element Title -- officially, "A name given to the resource" -- with a more restricted definition: "A name given to the unit of the collection". In the Archive's domain also a new issue evolves: the organisation of a collection.
Meantime in the museum domain the "Title " element might be redefined as Subject or Historical name of the painting or sculpture.
The type element is a different semantic meaning in the EAD standard where prescribes the hierarchical level of archival unit.
The group has identified some topics to be investigated.and defined:

- unique elements for identification
- semantic interoperability among metadata standard in different domains
- availability of controlled vocabularies encoding scheme, normalised list, authority file, etc., for productors and users to facilitate the access.

Administrative Metadata
The working Group has analized and focused the metadata set for administrative information and availability data, reproduction restrictions, etc. connected with the services and strategy of preservation as well.
The European Commission funded NEDLIB (Networked European Deposit Library) project has recently published its specification of Metadata for long term preservation. The specification defines core metadata elements for the preservation management of digital documents. As with the Cedars project, the specification generally follows the arrangement of the taxonomy of information objects defined in the Reference Model for an Open Archival Information System (OAIS), a draft ISO standard published by the Consultative Committee on Space Data Systems.
The group have considered Italian survey of Digital library the necessary step and framework for beginning.

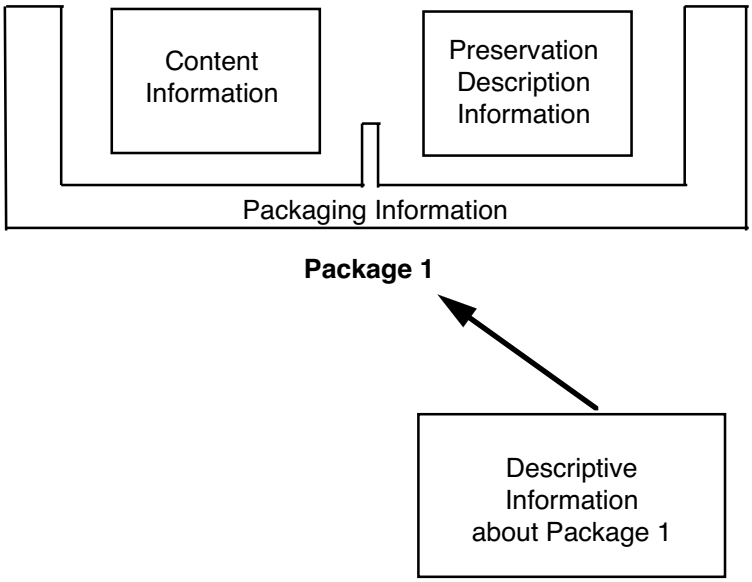
The first model analysed has been the OAIS (Open Archival Information System) (http://lcweb.loc.gov/standards/metadat.html).
The OAIS Reference Model is designed as a conceptual framework in which to discuss and compare archives. It has just used by libraries , archives and museum in many projects(NEDLIB ; CADARS and PANDORA).
The information being maintained has been deemed to need **Long Term Preservation**, even if the OAIS itself is not permanent. **Long Term** is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely. In this reference model there is a particular focus on digital information, both as the primary forms of information held and as supporting information for both digitally and physically archived materials. Therefore, the model accommodates information that is inherently non-digital (e.g., a physical sample).
It provides a framework for the understanding and increased awareness of archival concepts needed for Long Term digital information preservation and access;
The reference model addresses a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination.
It covers long-term preservation of information that is also not digital form (e.g., physical media, physical samples);

INFORMATION PACKAGE DEFINITION

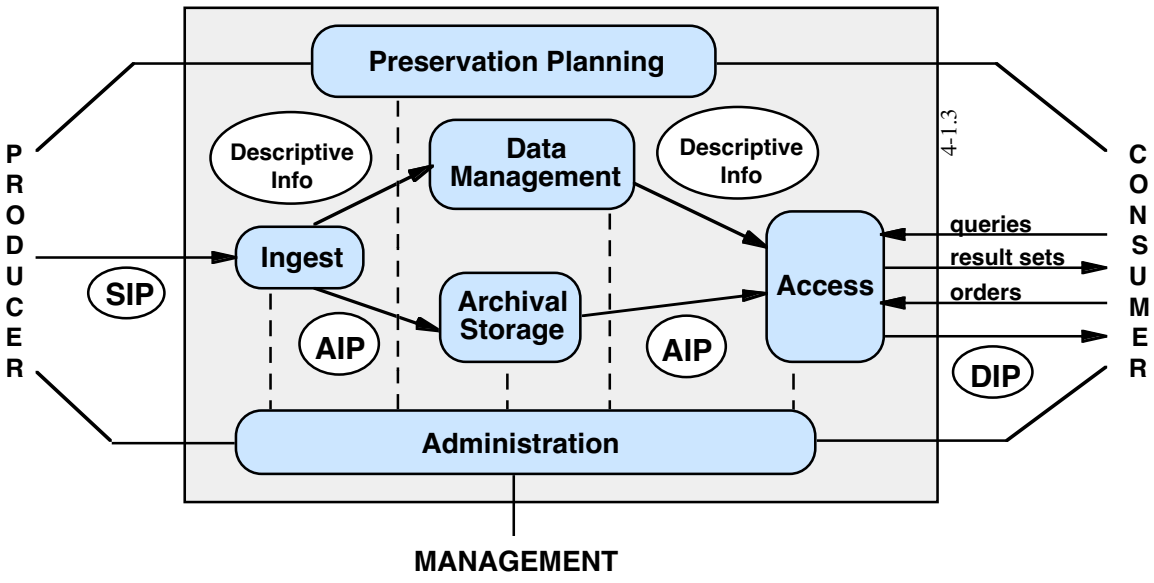
The definition of an Information Object is applicable to all the information types

An **Information Package** is a conceptual container of two types of information called **Content Information** and **Preservation Description Information (PDI)**. The Content Information and PDI are viewed as being encapsulated and identifiable by the **Packaging Information**. The resulting package is viewed as being discoverable by virtue of the **Descriptive Information**.

These Information Package relationships are shown schematically in figure



FUNCTIONAL MODEL



Data Management: This entity provides the services and functions for populating, maintaining, and accessing Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. Data Management functions include administering the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading new descriptive information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets.

The second model that has beeen considered is the one of the Library of Congress, in which hierarchy of information is proposed to accommodate the diversity of digital objects and to propagate data with some efficiency. Metadata elements may be supplied at multiple and various levels

- **Set -** Set-level metadata applies to what is currently known as a digital collection
- **Aggregate** - An aggregate organizes digital objects by digital type
- **Primary Object -** Primary objects are usually the digital equivalents of physical library items,
- **Intermediate Object - all** all the page images of the book, the other to an encoded text file
- **Terminal Object -** The terminal object is the digital content file or files that form the object

According to these analyses the Group has found points of convergence and common methodologies choices:
- OAIS as suitable model
- data stored independent of all operating systems

In the meantime, the group is working to create an environment useful as a tool for harmonizing metadata practice within particular communities and domains and that are harvestable and processable by machines.
In the initiatives called "The European Library" (TEL) (http://www.eureropeanlibrary.org) one of the specific objectives is the development of common standards for metadata to support wide scale access to digital, off-line digital and non-digital materials and the building of European Digital Library of traditional and electronic formats available in a coherent manner to local and remote users.
The works of the Group might necessary have the priority goal of the application of a common model in the Multimedia Digital Libraries. The model could provide a scalable approach to interoperability among multiple metadata set. The interoperability could be achieved by mapping through this common logical model.

Nome	Titolo								
Definizione	Un nome attribuito alla risorsa (A name given to the resource)								
Settore	ARTI (ProgettoNnetwork dei Beni Culturali)					ARCHIVI	Biblioteche		PUBBLICAZIONI UFFICIALI (Atti parlamentari Camera dei Deputati)
Standard / Profili Metadati	CIMI / DC					EAD	UNIMARC	Marc 21	DC
Etichetta	TITLE					<UNITTITLE>	200 $a,$c,$e 500, 517, 530	245, 130, 210, 240, 242, 246, 730, 740	TITOLO
Obbligatorietà MD							O	O	NO
Ripetibilità MD							R	R	NR
Vocabolario MD							NC	NC	NC
Standard norme catalografiche	OA-D-N	S-MI	FKO	RA-N	A	ISAD	ISBD /RICA		
Elemento / Campo	SGTT OGTN	SGTT	SGTT	SGTT OGTN	OGTD	3.1.2 Denominazione o titolo	1.1 Titolo proprio	1.3 Titolo parallelo	1.4 Complemento del Titolo
Obbligatorietà	NO	NO		NO	NO	O	O	NO	NO
Ripetibilità	NR	NR		NR		NR	NR	R	R
Vocabolario	NC	NC		NC		NC	NC	NC	NC
Commento	In assenza di titolo si riporta la denominazione dell'oggetto.					Riportare il titolo originale oppure attribuire un titolo conciso in conformità alle regole della descrizione in più livelli e alle convenzioni nazionali			
Note									
Esempi									

Nome	Tipo								
Definizione	La natura o il genere del contenuto della risorsa (the nature or genre of the content of the resource)								
Settore	ARTI (Progetto Network dei Beni Culturali)					ARCHIVI	**BIBLIOTECHE**		**PUBBLICAZIONI UFFICIALI (Atti parlamentari Camera dei Deputati)**
Standard / Profili Metadati	**CIMI / DC**					**EAD**	UNIMARC	Marc 21	DC
Etichetta	TIPO					<ARCHDESC> A) <LEVEL => B) <TYPE => C) <LEGALSTATUS	Guida posizione 6, 7 200$b	Guida 06, 07 245$h	TIPO DI RISORSA
Obbligatorietà MD						a) O b) NO c) NO	Guida pos. 6,7 = O 135, 230, 200$b = NO	Guida pos. 06, 07 = O 245$h = NO	NO
Ripetibilità MD						a) NR b) NR c) NR	NR/R	NR/R	NR
Vocabolario MD						a) C b) C c) C	C	C	C
Standard norme catalografiche	OA-D-N	S-MI	FKO	RA-N	A	ISAD	ISBD /RICA		
Elemento / Campo	OGTT OGTD	OGTT OGTD	OGTT OGTD	OGTT OGTD	OGTT OGTQ	a) 3.1.4. Livello di descrizione b) 3.4.1. Condizioni che regolano l'accesso	1.2 Designazione generica del materiale 3 Area specifica del materiale		
Obbligatorietà	O	O		O NO	NO	a) O b) NO c) NO	1.2 NO 3 O		
Ripetibilità	NR	NR		NR	NR	a) NR b) NR c) NR	1.2 NR 3 R		
Vocabolario	C	C		C	C	a) C b) C c) C	C		
Commento	In assenza di titolo si riporta la denominazione dell'oggetto.					- ISAD prevede i seguenti livelli: fondo, sub-fondo, serie, fascicolo, unità documentaria. - EAD prevede di utilizzare l'attributo Type per identificare il tipo di strumento di corredo. Il Getty Institute propone: archival finding aid - EAD prevede di utilizzare l'attributo per indicare la condizione giuridica del materiale descritto: pubblico, privato, altro.	Lista controllata ISBD(ER), (NBM), App. C: designazioni del materiale Liste controllate DCT1 (con estensioni locali)		
Note							DCT1: Dublin Core Type vocabulary		
Esempi									

Traduzione di Cristina Magliano

"Traduzione provvisoria; la traduzione uscira' in veste definitiva negli atti cartacei che saranno pubblicati dalla Editrice Bibliografica di Milano.