



Elementi di modellazione e programmazione di contenuti digitali

**Corso di Laurea Magistrale in Metodologie
informatiche per le discipline umanistiche**

Ing. Michela Paolucci

Department of Systems and Informatics

University of Florence

Via S. Marta 3, 50139, Firenze, Italy

tel: +39-055-4796523, fax: +39-055-4796363

Lab: DISIT, Sistemi Distribuiti e Tecnologie Internet

<http://www.disit.dsi.unifi.it/>

paolucci@dsi.unifi.it

<http://www.dsi.unifi.it/~nesi>, <http://www.dsi.unifi.it/~paolucci>,

<http://www.axmedis.org> <http://mobmed.axmedis.org/>





Elementi di modellazione e programmazione di contenuti digitali

Corso di Laurea Magistrale in Metodologie informatiche per le discipline umanistiche

- **Metadati e standard per la descrizione dei contenuti digitali**
 - DC, MPEG-7, MARC, MODS, RDF
- **Metadati usati nel campo dei beni culturali**
 - OAI-ORE, SBN, MAG, FRBR





- | La Fondazione Rinascimento Digitale (<http://www.rinascimento-digitale.it/>), il MiBAC (Direzione Generale per gli archivi e Direzione Generale per i beni librari, gli istituti culturali ed il diritto d'autore) e la Library of Congress organizzano la conferenza:

CULTURAL HERITAGE on line

15-16 December 2009, Florence - Italy
Teatro della Pergola, Via della Pergola 30

- | La Conferenza intende analizzare :
 - ♣ stato dell'arte e le future linee di sviluppo della produzione di contenuti culturali nel web
 - ♣ diffusione della conoscenza del patrimonio culturale e del web interattivo
 - ♣ diffusione delle biblioteche digitali
 - ♣ cooperazione fra musei, archivi e biblioteche
 - ♣ conservazione digitale
 - ♣ l'utilizzazione delle tecnologie dell'informazione da parte delle diverse comunità di utenti



La conferenza avrà inizio il 15 dicembre con interventi di esperti internazionali per illustrare le esigenze e le aspettative delle diverse tipologie di utenti per delineare soluzioni in grado di dare risposte concrete e di coinvolgere attivamente gli utilizzatori nella creazione e nella fruizione delle risorse culturali in formato digitale e nell'utilizzo delle tecnologie informatiche e della comunicazione (ICT).

Inoltre la Commissione Europea presenterà le opportunità di finanziamento nel settore delle digital libraries e della digital preservation nell'ambito della call 6 del FP7 aperta il 24 novembre 2009.

La giornata del 16 dicembre inizierà con la presentazione dei più importanti scenari nazionali e internazionali; seguiranno due sessioni parallele con interventi relativi alle seguenti tematiche: digital library applications, interactive Web e sustainable policies for digital culture preservation.



| Programma dettagliato: Cultural Heritage On-line

- ♣ http://bibc.emailsp.it/frontend/track.aspx?idUser=1636&idnl=17&url=http://bibc.emailsp.it/documenti/8/17/programma_cultural_heritage_online.pdf

| Per maggiori informazioni sulla conferenza e i due tutorial:

- ♣ www.rinascimento-digitale.it/conference2009
- ♣ <http://bibc.emailsp.it/frontend/track.aspx?idUser=1636&idnl=17&url=http://www.rinascimento-digitale.it/conference2009>

| Per iscriversi:

- ♣ www.rinascimento-digitale.it/conference2009-registration
<http://bibc.emailsp.it/frontend/track.aspx?idUser=1636&idnl=17&url=http://www.rinascimento-digitale.it/conference2009-registration>

| **Deadline: Le iscrizioni chiuderanno ufficialmente l'11 dicembre 2009**



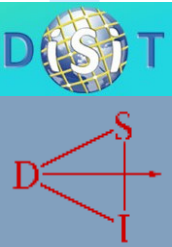
Metadati e standard per la descrizione dei contenuti digitali - Introduzione -





I metadati (1)

- | I metadati sono informazioni che descrivono un insieme di dati
 - | Sono quindi utili per la classificazione delle risorse
 - | Hanno un ruolo fondamentale nella ricerca di materiali, nello scambio di informazioni, nella riusabilità e tracciabilità degli oggetti digitali, ...
-
- | Esempi:
 - ♣ tag delle pagine web in formato html. Si tratta di un insieme di tag con il compito di definire e descrivere ulteriormente il significato ed il contenuto della pagina web
 - ♣ Campi informativi usati per catalogare libri e riviste in una biblioteca
 - | I metadati sono informazioni stabili, che si mantengono nel tempo e sono gli stessi su sistemi informativi diversi
 - | Una corretta associazione di metadati alle risorse permette di renderle disponibili ad un pubblico più vasto, di migliorare la ricerca, etc...
 - | Ad esempio è possibile associare ai formati dei documenti opportune strutture di metadati:
 - ♣ Video: durata, soggetto, produttore, ...
 - ♣ Immagini: risoluzione, formato, soggetto, ...
 - ♣ ...





I metadati (2)

- | I metadati vanno uniformati il più possibile in modo da permettere risultati ottimali ai motori di ricerca
- | "The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself." (Weibel and Lagoze, 1997)
[<http://dublincore.org/documents/usageguide>]
- | "L'associazione di metadati descrittivi standard con oggetti in rete ha il potenziale per migliorare in modo sostanziale le possibilità di reperire le risorse attraverso la ricerca basata su campi (ad esempio autore, titolo), permettendo l'indicizzazione di oggetti non testuali, e consentendo l'accesso a un contenuto surrogato che è distinto dall'accesso al contenuto della risorsa stessa."





Tipi di metadati (1)

- | Esistono vari tipi di metadati, si possono principalmente suddividere in tre categorie:
 - ♣ Descrittivi
 - ♣ Gestionali
 - ♣ Strutturali
- | **Metadati Descrittivi:** sono usati per descrivere le risorse gestite/condivise/pubblicate/etc., in modo da facilitare la ricerca di libri, documenti, riviste, etc.
- | Sono applicati sia a documenti digitali che cartacei
- | Alcuni esempi:
 - ♣ Dublin Core: <http://dublincore.org/>
 - ♣ MARC: <http://www.loc.gov/marc/>
 - ♣ OAI-ORE: <http://www.openarchives.org/ore/>





Tipi di metadati (2)

| **Metadati gestionali:** Sono utilizzati per la gestione delle risorse:

- ♣ Controllo dell'accesso alle risorse
- ♣ Archiviazione, manutenzione e conservazione degli oggetti digitali
- ♣ Certificazione di autenticità e integrità della risorsa
- ♣ Identificazione univoca della risorsa

| Alcuni esempi di metadati gestionali :

- ♣ Comitato Metadati Amministrativi Gestionali, MAG (<http://www.iccu.sbn.it/genera.jsp?id=99&l=en>)
- ♣ Preservation Metadata for digital Collections della National Library of Australia: <http://www.nla.gov.au/preserve/pmeta.html>
- ♣ Metadata for digital preservation CEDARS: <http://www.webarchive.org.uk/ukwa/target/99695>





Tipi di metadati (3)

- | **Metadati Strutturali:** Servono per descrivere la struttura delle risorse e le relazioni che esistono tra di esse (o tra le loro parti)
- | Collegano le varie componenti delle risorse per un'adeguata e completa fruizione, che spesso avviene attraverso la mappatura di schemi di metadata diversi
- | Questi metadata inoltre forniscono dati di identificazione e localizzazione del documento, come il codice identificativo, l'indirizzo del file sul server, l'archivio digitale di appartenenza e il suo indirizzo Internet
- | Alcuni esempi di metadati strutturati:
 - ♣ SGML: <http://xml.coverpages.org/sgml.html>
 - ♣ XML: <http://www.w3.org/XML/>
 - ♣ RDF: <http://www.w3.org/RDF/>





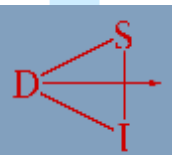
MPEG-7





MPEG-7 (1)

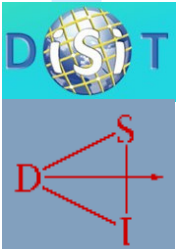
- | È uno standard ISO/IEC sviluppato da MPEG (Moving Picture Experts Group, <http://www.chiariglione.org/mpeg/>), organizzazione che lavora con l'International Organization for Standardization (ISO) e sviluppa le norme per la compressione video e audio digitale:
 - ♣ MPEG-1 : The standard on which such products as Video CD and MP3 are based
 - ♣ MPEG-2: The standard on which such products as Digital Television set top boxes and DVD are based
 - ♣ MPEG-4: The standard for multimedia for the fixed and mobile web
 - ♣ MPEG-7: The standard for description and search of audio and visual content
 - ♣ MPEG-21: The Multimedia Framework
- | Si tratta di una famiglia di standard utilizzati per la codifica delle informazioni audiovisive (ad esempio, filmati, video, musica) in formato digitale compresso
- | MPEG-7, chiamato anche "Multimedia Content Description Interface", è uno standard per la descrizione dei contenuti **multimediali**
- | MPEG-7 non è rivolto ad una particolare applicazione, al contrario gli elementi che descrive supportano una vasta gamma di applicazioni





MPEG-7 (2)

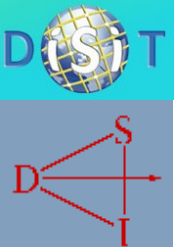
- | Il valore delle informazioni spesso dipende da quanto semplicemente una risorsa può essere trovata, estratta, accessibile, etc.
- | MPEG-7:
 - ♣ classifica i contenuti multimediali
 - ♣ fornisce una serie di strumenti standardizzati per descrivere i contenuti multimediali: si tratta di strumenti audiovisivi di supporto alla descrizione delle risorse digitali (a partire dalla descrizione del significato dei metadati e della loro struttura fino a mettere a disposizione veri e propri programmi di descrizione)
- | MPEG-7 è stato sviluppato da:
 - ♣ Esperti che rappresentano le emittenti televisive
 - ♣ Produttori di componenti elettronici
 - ♣ Creatori e responsabili di contenuti
 - ♣ Editori
 - ♣ Gestori dei diritti di proprietà intellettuale
 - ♣ Fornitori di servizi di telecomunicazione
 - ♣ Mondo accademico





MPEG-7: campi di applicazione

- | **Broadcast media selection** (e.g., radio channel, TV channel).
- | **Cultural services** (history museums, art galleries, etc.).
- | **Digital libraries** (e.g., image catalogue, musical dictionary, film, video and radio archives).
- | **E-Commerce** (e.g., personalised advertising, on-line catalogues, directories of e-shops).
- | **Education** (e.g., repositories of multimedia courses, multimedia search for support material).
- | **Home Entertainment** (e.g., systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke).
- | **Investigation services** (e.g., human characteristics recognition, forensics).
- | **Journalism** (e.g. searching speeches of a certain politician using his name, his voice or his face).
- | **Multimedia directory services** (e.g. yellow pages, Tourist information, Geographical information systems).
- | **Multimedia editing** (e.g., personalised electronic news service, media authoring).
- | **Remote sensing** (e.g., cartography, ecology, natural resources management).
- | **Shopping** (e.g., searching for clothes that you like).
- | **Social** (e.g. dating services).
- | **Surveillance** (e.g., traffic control, surface transportation, non-destructive testing in hostile environments).





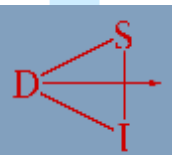
MPEG-7 Objectives (1)

- | Le descrizioni MPEG-7 sono utilizzati nel caso sia necessario descrivere contenuti audiovisivi: foto, grafica, modelli 3D, audio, discorso, video, risorse multimediali
- | Tali descrizioni:
 - ❖ Non dipendono dalle modalità di codifica o memorizzazione della risorsa
 - ❖ Sono strutturate in modo tale da permettere una descrizione granulare delle risorse
 - ❖ Possono sfruttare i vantaggi forniti da risorse MPEG-4 (es: sarà possibile collegare le descrizioni delle risorse all'interno della scena, come gli oggetti audio e video)
 - ❖ Le modalità di descrizione dipendono dal contesto (le funzioni descrittive sono ottimizzate in base alla applicazione usata)

| Esempio:

| Dato un Video (si ipotizzano due livelli di astrazione):

- ❖ Il livello di astrazione:
 - ➔ Video: descrizione di forma, dimensione, consistenza, colore, movimento (traiettoria) e posizione (es: in che punto della scena può essere trovato un oggetto);
 - ➔ Audio: chiave, umore, ritmo, cambiamenti di ritmo, la posizione nello spazio sonoro
- ❖ Il livello di astrazione: potrebbe fornire informazioni semantiche: "questa è una scena in cui sulla sinistra si trova un cane marrone che abbaia, mentre sulla destra c'è una palla blu che cade. Come sottofondo musicale c'è il rumore delle macchine che viaggiano"





MPEG-7 Objectives (2)

- | Il livelli di astrazione dipendono dalle modalità con cui le informazioni possono essere estratte dalla risorsa
 - ♣ Metodi automatizzati (bassi livelli)
 - ♣ Interazione umana (alti livelli)

- | MPEG-7 definisce solo il **formato** da usare per descrivere queste features e non le metodologie usate per estrarle
 - ♣ Per lo scambio, ricerca, ecc. è necessario conoscere solo il formato delle descrizioni, non come sono state ottenute
 - ♣ Questo permette di avere uno standard che si “adatta” alle evoluzioni tecnologiche (nuovi algoritmi di estrazione di feature possono essere utilizzati senza modificare lo standard)

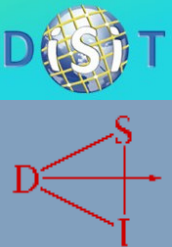




MPEG-7: parts

The MPEG-7 Standard consists of the following parts:

- ♣ **MPEG-7 Systems** – the tools needed to prepare MPEG-7 descriptions for efficient transport and storage and the terminal architecture
- ♣ **MPEG-7 Description Definition Language** - the language for defining the syntax of the MPEG-7 Description Tools and for defining new Description Schemes
- ♣ **MPEG-7 Visual** – the Description Tools dealing with (only) Visual descriptions
- ♣ **MPEG-7 Audio** – the Description Tools dealing with (only) Audio descriptions
- ♣ **MPEG-7 Multimedia Description Schemes** - the Description Tools dealing with generic features and multimedia descriptions
- ♣ **MPEG-7 Reference Software** - a software implementation of relevant parts of the MPEG-7 Standard with normative status
- ♣ **MPEG-7 Conformance Testing** - guidelines and procedures for testing conformance of MPEG-7 implementations
- ♣ **MPEG-7 Extraction and use of descriptions** – informative material (in the form of a Technical Report) about the extraction and use of some of the Description Tools
- ♣ **MPEG-7 Profiles and levels** - provides guidelines and standard profiles
- ♣ **MPEG-7 Schema Definition** - specifies the schema using the Description Definition Language





MPEG-7: main elements (1)

I principali elementi che costituiscono lo standard sono:

- ♣ **Description Tools:**

- Descriptors (D), definiscono la sintassi e la semantica di ogni funzione

- Description Schemes (DS), specificano la struttura e la semantica dei rapporti tra gli oggetti che compongono la risorsa. Possono essere sia descrittori che programmi di descrizione

- ♣ **Description Definition Language (DDL):**

- Serve per definire la sintassi degli strumenti MPEG-7, per permettere l'estensione e/o modifica dei Description Schemes

- ♣ **System tools:**

- Serve per supportare i meccanismi di trasporto e di codifica per l'archiviazione e la trasmissione efficiente delle informazioni, la gestione della sincronizzazione delle descrizioni con contenuti, la gestione e protezione dei diritti di proprietà intellettuale, etc.





MPEG-7: Descriptors

- | Le descrizioni di MPEG-7 possono includere informazioni:
- | Orientate all'archiviazione
 - ♣ creazione e produzione del contenuto (es. titolo, regista/autore, attori)
 - ♣ uso del contenuto (es. copyright, broadcast scheduling, cronologia di utilizzo della risorsa, etc)
 - ♣ memorizzazione e presentazione (es. formati, codifica)
- | Orientate alla percezione
 - ♣ struttura spazio-temporale (es. cambi di scena, motion tracking, regioni)
 - ♣ caratteristiche a basso livello (es. colore, texture, timbro, melodia)
 - ♣ interpretazione semantica della realtà rappresentata (es. soggetti raffigurati, eventi e loro relazioni)
- | Orientate alla gestione
 - ♣ organizzazione di oggetti in collezioni
 - ♣ supporti per il browsing (es. sommari, indici)
 - ♣ Interazione dell'utente con la risorsa (es. profilo e preferenze dell'utente, storia di utilizzo della risorsa, etc.)



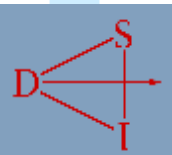


MPEG-7: Description Schemes

- | I Description Schemes (DS) sono:
 - ♣ strutture di metadati necessari per la descrizione e le annotazioni dei contenuti audiovisivi (AV)
 - ♣ definiti mediante il DDL MPEG-7, basato su xml-schema
 - ♣ una 'combinazione' dei descrittori (D) o di altri DS
 - ♣ progettati principalmente per descrivere i livelli più alti di astrazione di una risorsa AV (regioni, segmenti, oggetti, eventi; aspetti semantici, metadati immutabili relative alla creazione e produzione, utilizzo, etc.)

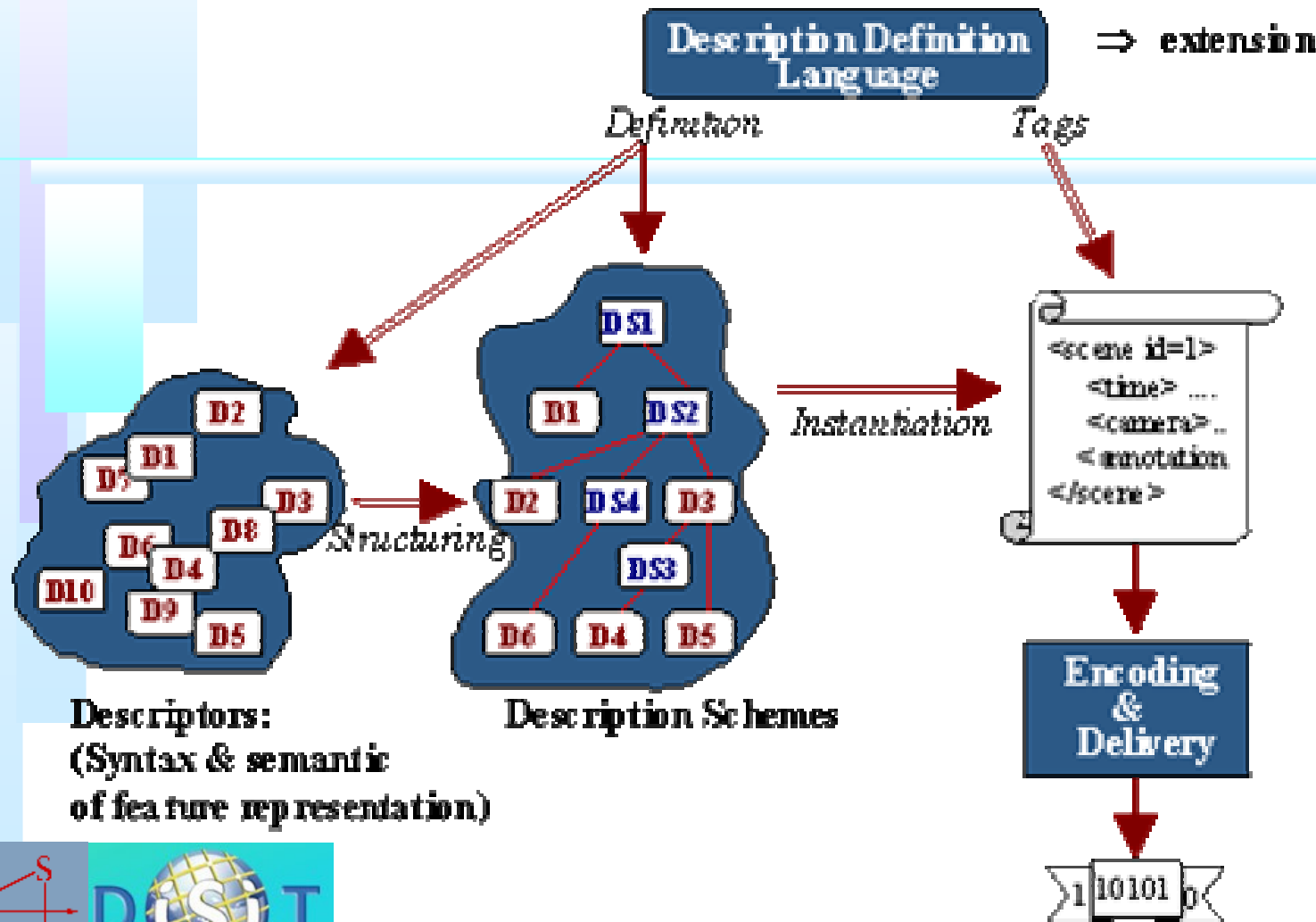
- | I DS descrivono la risorsa come combinazione di vari aspetti: audio, video, dati legati alla visualizzazione degli oggetti nello spazio, dati testuali, etc.

- | Richiedono solitamente l'intervento dell'uomo





MPEG-7: Description Definition Language (1)

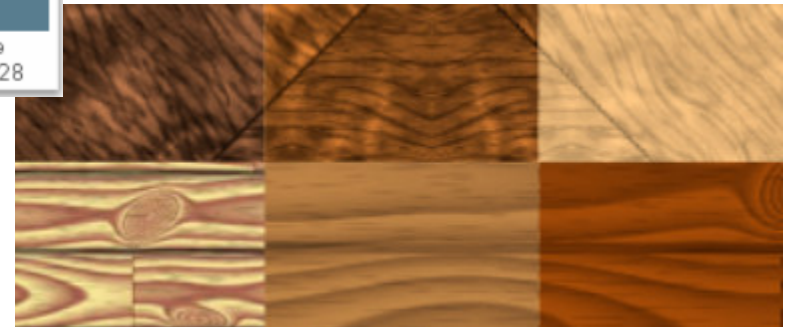




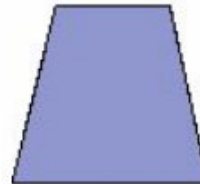
MPEG-7 Visual

- | MPEG-7 Visual – the Description Tools dealing with (only) Visual descriptions
- | **Tool che** permettono di descrivere le seguenti feature visuali:

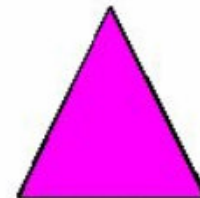
- ♣ **Colour**
- ♣ **Texture**
- ♣ **Shape**
- ♣ **Motion**
- ♣ **Localisation**
- ♣ **Faces**
- ♣ **etc.**



cylinder



trapezoid



triangle





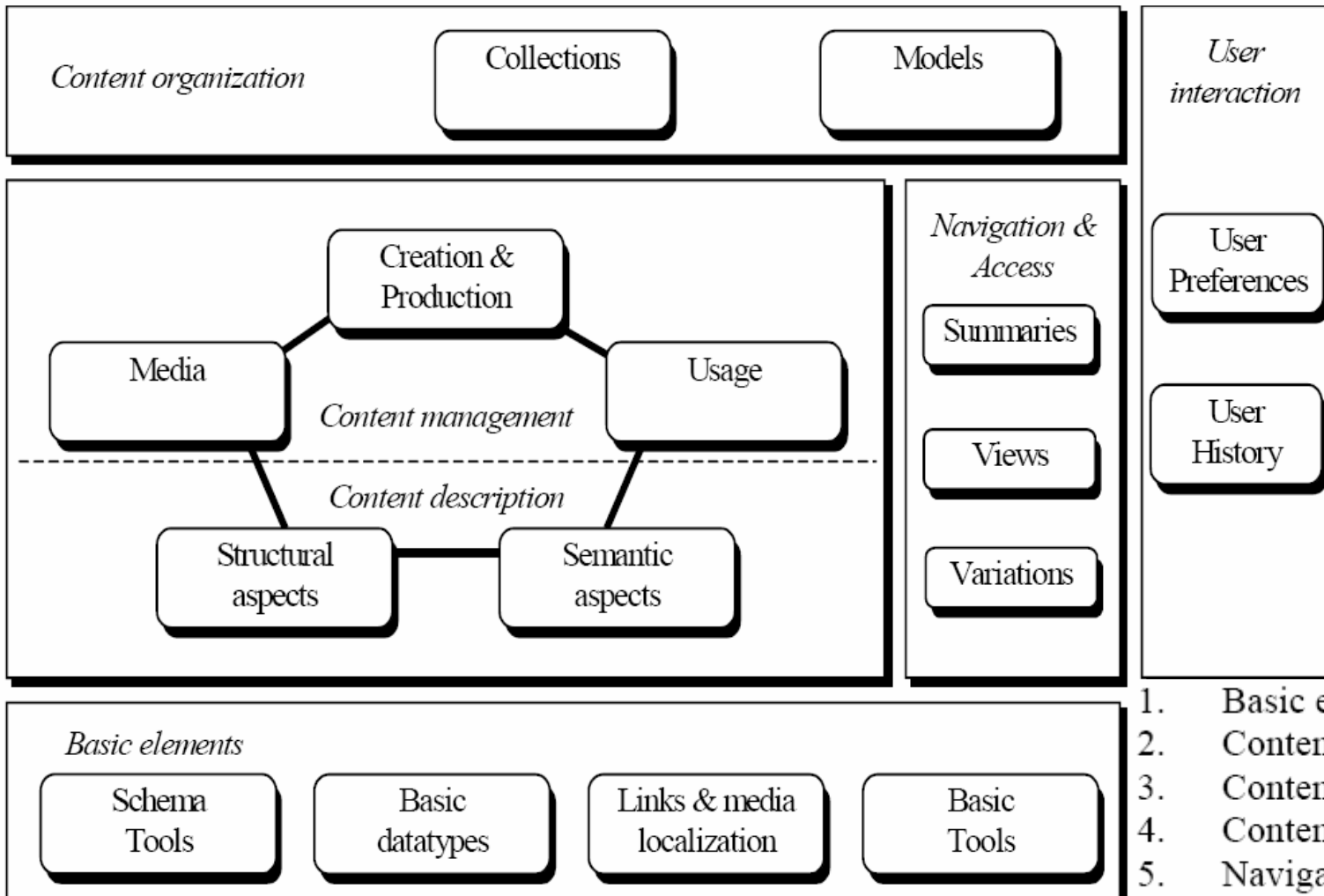
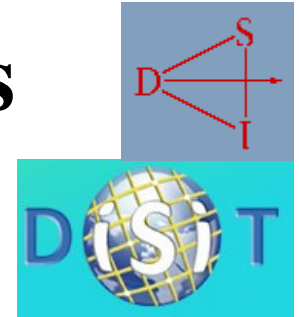
MPEG-7 Audio

- | MPEG-7 Audio – the Description Tools dealing with (only) Audio descriptions
- | Prevede tool che, affiancati ai Description Schemes, sono usati per la descrizione dei contenuti audio delle risorse
- | Prevede:
 - ♣ una serie di descrittori di basso livello per le funzioni audio basilari e comuni alle varie applicazioni (spettro, funzionalità temporale di un segnale, etc spectral, parametric, and temporal features of a signal),)
 - ♣ strumenti per le descrizione di alto livello (specifiche in base alla applicazine usata): riconoscitori musicali, indexing Description Tools, instrumental timbre Description Tools, spoken content Description Tools, an audio signature Description Scheme, and melodic Description Tools to facilitate query-by-humming, etc.



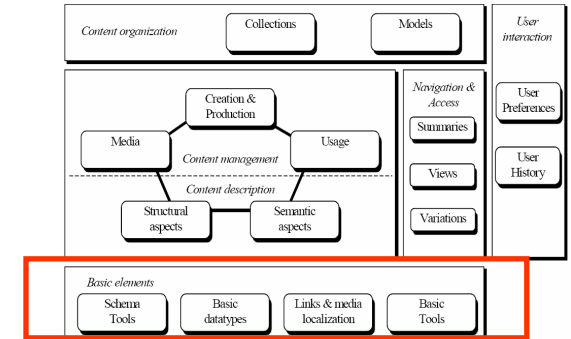


The MPEG-7 Description Tools organized into areas



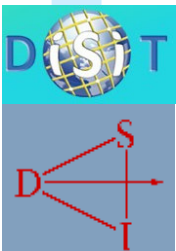
1. Basic elements
2. Content Management
3. Content Description
4. Content Organization
5. Navigation and Access
6. User Interaction

The MPEG-7 Description Tools organized into areas (2)

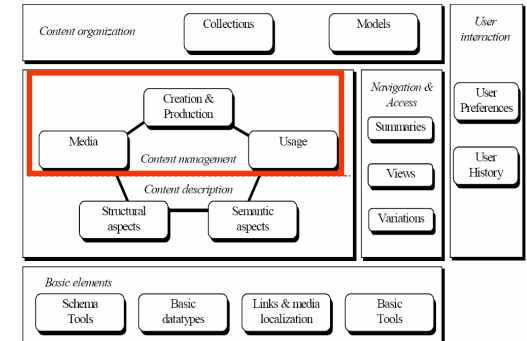


Basic Elements: Components necessary for the development of complex and compound description schemes:

- ♣ **Schema Tools** that assist in the formation, packaging, and annotation of MPEG-7 descriptions
- ♣ The **Basic data types** provide a set of extended data types and mathematical structures such as vectors and matrices, which are needed by the DSs for describing AudioVisual (AV) content
- ♣ **Links and media localization.** Constructs for linking media files and localizing pieces of content
- ♣ **Basic Tools** for describing time, place, persons, individuals, groups, organizations, and other textual annotation. In addition, constructs for classification schemes and controlled terms



The MPEG-7 Description Tools organized into areas (3)



Content Management: Include tools describing the life cycle of the content, from content to consumption; creation and production, media coding (storage and file formats), and content usage. "content" is an entity that has a specific structure to represent the reality. The content described by MPEG-7 descriptions can be available in different modalities, formats, Coding Schemes, and there can be several instances:

- ♣ **Creation and Production** Description Tools: Describing the creation process of an entity (title, textual annotation, creators, creation location, dates) and classification (categories such as genre, subject, purpose, language etc) of the AV content and of other related materials.
- ♣ **Media** Description Tools: Describes the storage features of the media such as the format, compression and coding of the AV content
- ♣ **Content** Usage Description Tools: Describing information related to the usage and usage process of the AV content such as usage rights, usage records, and financial information.



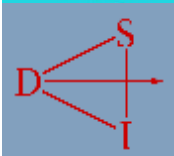
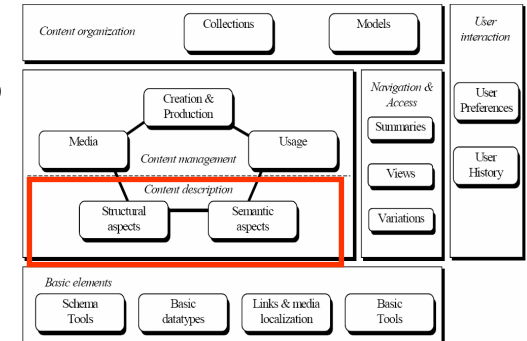
The MPEG-7 Description Tools organized into areas (4)



Content Description

MPEG-7 provides Description Schemes for describing the structure and semantics of the AV content. The structural tools describe the structure of the AV content in terms of video segments, frames, still and moving regions and audio segments. The semantic tools describe the objects, events, and notions from the real world that are captured by the AV content.

- ♣ **Structural Aspects:** description of the content structural aspect. Structural information regarding physical, spatial, temporal or spatial-temporal segments of the content. Each segment may be described by information regarding the low level features of the content (colours, textures, sound timbres, melody description)
- ♣ **Semantic Aspects:** description of the content conceptual aspects: The conceptual aspects describe the AV content from the viewpoint of real-world semantics. Conceptual information of the reality captured by the content (objects and events, interactions among objects)

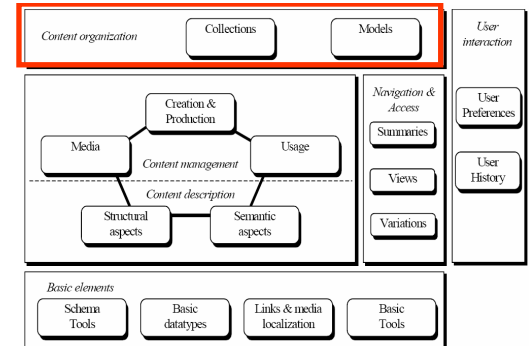


The MPEG-7 Description Tools organized into areas (5)



Content Organization

- | MPEG-7 provides DSs for organizing and modelling collections of AV content (segments, events, and/or objects), and describing their common properties
- | The collections can be further described using different models and statistics in order to characterize the attributes of the collection members

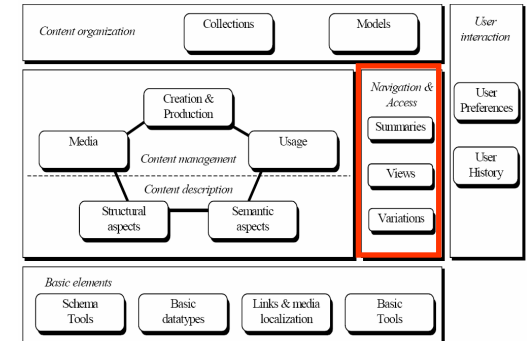


The MPEG-7 Description Tools organized into areas (6)



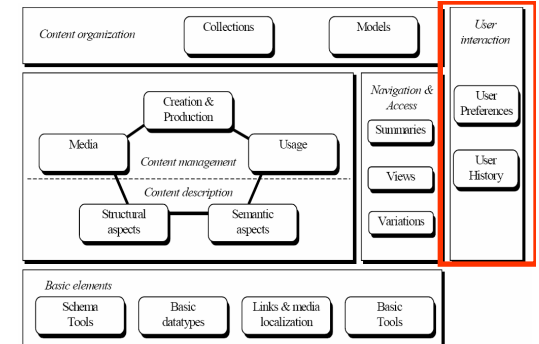
Navigation and Access

- | MPEG-7 facilitates navigation and access of AV content by describing summaries, views and partitions, and variations. Information about how to browse the content in an efficient way.





The MPEG-7 Description Tools organized into areas (6)



| User Interaction

- | The UserInteraction DSs describe preferences of users pertaining to the consumption of the AV content, as well as usage history. The MPEG-7 AV content descriptions can be matched to the preference descriptions in order to select and personalize AV content for more efficient and effective access, presentation and consumption.





Link MPEG-7

- | <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- | <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm#E9E1>
- | <http://www.dsi.unive.it/~sim/slide/sim-mpeg.pdf>





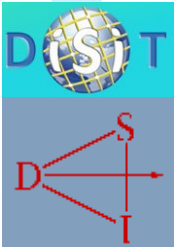
MARC





MARC (1)

- | **MARC** is an acronym for **MA**chine **R**eadable **C**ataloguing
- | The MARC format was primarily developed as a convenient way of storing and exchanging **bibliographic** records
- | It has been extended to include non-bibliographic forms of library material, such as maps and music scores
- | There is also a MARC format for the storage and exchange of **authority** records
- | MARC refers to a suite of related standards used for bibliographic control within the library science and 'digital libraries' communities
- | The MARC standards are based upon ISO 2709:1996, Format for Information Exchange (INEX)





MARC History

- | The original MARC format was developed by the Library of Congress in the mid-1960s
- | A pilot project, known as MARC I, was set up to investigate the feasibility of producing catalogue data in machine-readable form
- | Similar work was being done in the United Kingdom by the Council of the British National Bibliography Ltd, which had set up the BNB MARC Project to examine whether the production of the printed **British National Bibliography** (BNB) could be simplified by using machine-readable bibliographic records
- | These two initiatives came together in 1968 in the MARC II Project which was an Anglo-American effort to develop a standard communications format





MARC Formats

| MARC formats:

- ♣ handles resources of more complex structures, also for this reason are widely adopted for cataloguing of bibliographic information
- ♣ are not restricted to any particular type of information, so MARC-based catalogue will be able to handle collections of different medium

| MARC refers to a suite of related standards:

- ♣ MARC 21
- ♣ UKMARC, UNIMARC, CNMARC, USMARC,... (country-specific MARC formats)
- ♣ etc.

| used for bibliographic control within the library science and 'digital libraries' communities

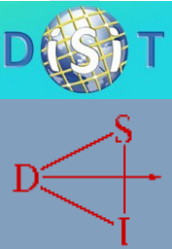
| The generic MARC 21 format, rather than country-specific MARC formats such as USMARC or CNMARC are used to ensure that the catalogue developed is not dependent on or limited by standards of any particular country





MARC Standards: MARC 21 (1)

- | MARC 21 is a result of the combination of the United States and Canadian MARC formats (USMARC and CAN/MARC)
- | The MARC 21 formats are maintained by the Library of Congress in consultation with various user communities
- | The British Library adopted MARC 21 as its cataloguing format in June 2004 as part of the implementation of an integrated library system
- | The MARC 21 formats are used for “representation and communication of bibliographic and related information in machine-readable form”
- | MARC 21 allows the use of two character sets:
 - ♣ MARC-8
 - ♣ Unicode, encoded as UTF-8
- | MARC-8 is based on ISO 2022 and allows the use of Hebrew, Cyrillic, Arabic, Greek, and East Asian scripts. MARC 21 in UTF-8 format allows all the languages supported by Unicode





MARC Standards: MARC 21 (2)

- | MARC 21 was designed to redefine the original MARC record format for the 21st century and to make it more accessible to the international community
- | There are five types of MARC 21 formats devised for different types of data:
 - ♣ Bibliographic
 - ♣ Authority
 - ♣ Holdings
 - ♣ Classification
 - ♣ Community Information
- | MARC 21 Format for Bibliographic Data is used for all types of bibliographic information, ranging from textual material, audio and visual material to computer files and mixed materials



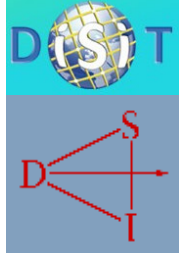


Country-specific MARC formats :

UNIMARC

- | UNIMARC - The primary purpose of UNIMARC is to facilitate the international exchange of data in machine-readable form between national bibliographic agencies
- | UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats
- | The interests of users of UNIMARC records are represented by the Permanent UNIMARC Committee(PUC) on behalf of IFLA-CDNL Alliance for Digital Strategies (ICADS), which has ultimate responsibility for UNIMARC
- | The British Library is a member of the Permanent UNIMARC Committee





MARC Record (1)

- | A MARC record is composed of three elements (RFC2220):
 - ♣ record structure
 - ♣ content designation
 - ♣ data content of the record
- | The **content designation** - the codes and conventions established explicitly to identify and further characterize the data elements within a record and to support the manipulation of that data - is defined by each of the MARC formats
- | The **content** of the data elements that comprise a MARC record is usually defined by standards outside the formats. Examples are:
 - ♣ the International Standard Bibliographic Description (ISBD)
 - ♣ Anglo-American Cataloguing Rules
 - ♣ Library of Congress Subject Headings (LCSH), or other cataloging rules,
 - ♣ subject thesauri, and classification schedules used by the organization that creates a record.



MARC record types

- | MARC prevede due tipi principali di record necessari per la catalogazione delle risorse:
 - ♣ **Bibliographic records**, che contengono informazioni su un libro, il numero di serie (ISBN), la registrazione di un suono o di un video, etc.
 - ♣ **Authority records**, che descrivono gli standard con cui devono essere memorizzati i nomi, i titoli, e in generale tutti i metadati bibliografici. Inoltre definiscono i riferimenti incrociati e danno la possibilità di aggiungere alcune note collegate alla risorsa in esame





MARC Bibliographic record

- | A MARC record is a MACHine-Readable Cataloging record
- | **Cataloging record**, means a bibliographic record, or the information traditionally shown on a catalog card
- | The record includes (not necessarily in this order):
 - ♣ Description of the item
 - ♣ Main entry and Added entries
 - ♣ Subject Headings
 - ♣ Classification or call number(MARC records often contain much additional information)





MARC Bibliographic record: Description

| Librarians follow the rules in Anglo-American Cataloguing Rules, 2nd ed., 2002 revision to compose the bibliographic description of a library item (AACR2)

| Description includes:

- ♣ Title
- ♣ Statement of responsibility
- ♣ Edition
- ♣ Material specific details
- ♣ Publication information
- ♣ Physical description
- ♣ Series
- ♣ Notes
- ♣ Standard numbers





MARC Bibliographic record: Main entry and Added entries

- | AACR2, Anglo-American Cataloguing Rules, also contains rules for determining "access points" to the record (usually referred to as:
 - ♣ "main entry"
 - ♣ "other added entries"and the form these access points should take
- | Access points are the retrieval points in the library catalog
- | In other words, the rules in AACR2 are used to answer questions such as:
 - ♣ For this book, should there be entries in the catalog for more than one author or more than one title?
 - ♣ Should the title of the series be noted?
 - ♣ How should the author's name be written?
 - ♣ Is this a "title main entry" item (no author)?





MARC Bibliographic record: Subject Headings

- | The librarian uses:
 - ♣ Sears List of Subject Headings (Sears)
 - ♣ Library of Congress Subject Headings (LCSH)
 - ♣ some other list of standard subject headingsto select the subjects under which the item will be listed
- | Use of an approved list is important for consistency, to ensure that all items on a particular subject are found under the same heading and therefore in the same place in the catalog





MARC Bibliographic record: Classification or call number

- | The librarian uses:
 - ♣ Dewey Decimal
 - ♣ Library of Congress classification scheduleto select the call number for an item
- | The purpose of the call number is to place items on the same subject together on the same shelf in the library
- | Most items are sub-arranged alphabetically by author
- | The second part of the call number usually represents:
 - ♣ the **author's name**, facilitating this subarrangement





MARC Bibliographic record: why it is necessary (1)

- | Why can't a computer just read a catalog card?
- | The information from a catalog card cannot simply be typed into a computer to produce an automated catalog
- | The computer needs a means of interpreting the information found on a cataloging record
- | The MARC record contains a guide to its data, or little "signposts," before each piece of bibliographic information
- | The place provided for each of these pieces of bibliographic information (author, title, call number, etc.) is called a "**field**"
- | The records in simpler computer files sometimes have a fixed number of fields, and each field contains a fixed number of characters
- | Therefore each MARC record contains a little "table of contents" to the record, according to a predefined standard.
- | Data "**signposts**:" The computer must have assistance if it is to read and interpret the bibliographic record





Bibliographic Record with textual "signposts"



| "SIGNPOSTS" | DATA |
|--|---|
| Main entry, personal name with a single surname: The name: | Arnosky, Jim. |
| Title and Statement of responsibility area, pick up title for a title added entry, file under "Ra..." Title proper: Statement of responsibility: | Raccoons and ripe corn / Jim Arnosky. |
| Edition area: Edition statement: | 1st ed. |
| Publication, distribution, etc., area: Place of publication: Name of publisher: Date of publication: | New York : Lothrop, Lee & Shepard Books, c1987. |

| | |
|--|---|
| Physical description area: Pagination: Illustrative matter: Size: | 25 p. : col. ill. ; 26 cm. |
| Note area: Summary: | Hungry raccoons feast at night in a field of ripe corn. |
| Subject added entries, from Library of Congress subject heading list for children: Topical subject: | Raccoons. |
| Local call number: | 599.74 ARN |
| Local barcode number: | 8009 |
| Local price: | \$15.00 |



Marc Bibliographic Record

Same record with MARC tags

| "SIGNPOSTS" | DATA |
|-------------|--|
| 100 1# \$a | Arnosky, Jim. |
| 245 10 \$a | Raccoons and ripe corn / |
| | \$c Jim Arnosky. |
| 250 ## \$a | 1st ed. |
| 260 ## \$a | New York : |
| | \$b Lothrop, Lee & Shepard Books, |
| | \$c c1987. |
| 300 ## \$a | 25 p. : |
| | \$b col. ill. ; |
| | \$c 26 cm. |
| 520 ## \$a | Hungry raccoons feast at night in a field of ripe corn. |
| 650 #1 \$a | Raccoons. |
| 900 ## \$a | 599.74 ARN |
| 901 ## \$a | 8009 |
| 903 ## \$a | \$15.00 |





MARC Bibliographic Terms and Their Definitions

- | How to read, understand, and use a MARC record it is fundamental
- | It is related to:
 - ♣ How librarians use their library automation system (adding, editing, or examining records)
 - ♣ Areas commonly used in cataloging for schools and small public libraries (i.e. books and audiovisual materials)
 - ♣ All forms of materials in which the resources are maintained (sound recordings, computer software, maps, and other non-book items, etc)
- | The box chart in the previous section showed a MARC record labelled with "signposts"
- | The proper names of these "signposts" are *field*, *tag*, *indicator*, *subfield*, *subfield code*, and *content designator*





MARC Bibliographic Fields are marked by Tags

- | A **field** is the place provided for each piece of bibliographic information such as: author, title, call number, etc.
- | Each bibliographic record is divided logically into fields
- | These fields are subdivided into one or more "subfields."
- | As previously noted, the textual names of the fields are too lengthy to be reproduced within each MARC record
- | Each field is associated with a 3-digit number called a "tag."(100, 245, etc.)
- | A **tag** identifies the field (the kind of data) that follows





MARC Bibliographic Tags used most frequently



010 tag marks the **Library of Congress Control Number** (LCCN)

020 tag marks the **International Standard Book Number** (ISBN)

100 tag marks a **personal name main entry** (author)

245 tag marks the **title information** (which includes the title, other title information, and the statement of responsibility)

250 tag marks the **edition**

260 tag marks the **publication information**

300 tag marks the **physical description** (often referred to as the "collation" when describing books)

490 tag marks the **series statement**

520 tag marks the **annotation or summary note**

650 tag marks a **topical subject heading**

700 tag marks a **personal name added entry** (joint author, editor, or illustrator)

Here is an example of a field. The number 100 is the tag, defining it as a personal name main entry (author) field.

```
100 1# $a Pirsig, Robert M.
```

NOTES:

- In the MARC record, 10% of the tags are used over and over, and the other 90% are seen rarely
- It is not unusual to hear librarians speaking in "MARCese.": they memorize the numbers for the fields common to the materials they catalog



MARC Bibliographic Indicators

- | Some fields are further defined by INDICATORS
- | Two character positions follow each tag (with the exception of Fields 001 through 009). One or both of these character positions may be used for indicators. All the following cases are used:
 - ♣ only the first or second position
 - ♣ both positions
 - ♣ neither
- | When an indicator position is not used, that indicator is referred to as "undefined" and the position is left blank. It is the convention to represent a blank, or undefined, indicator position by the character "#" (e.g. 100 1# , 24510,300 ##)
- | Each indicator value is a number from 0 to 9
- | Even though two indicators together may look like a 2-digit number, they really are two single-digit numbers

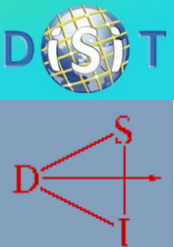




MARC Indicators: example

```
245 14 $a The emperor's new clothes /  
$c adapted from Hans Christian Andersen  
and illustrated by Janet Stevens.
```

- | The first 3 digits (245) defines this as a title field
- | The next 2 digits are indicator values:
 - ♣ **first indicator** value (1):
 - ♣ **second indicator** value (4):
- | A **first indicator** value of 1 in the title field In the card catalog environment, this means that a title card should be printed for this item and an entry for "Title" added to the tracings. A first indicator value of 0 would mean that a title main entry is involved; the card would be printed with the traditional hanging indention, and no additional tracing for the title would be required (since it is the main entry).
- | **Nonfiling characters:** One of the more interesting indicators is the second indicator for the title field. It displays the number of characters at the beginning of the field (including spaces) to be disregarded by the computer in the sorting and filing process. For the title *The emperor's new clothes*, the second indicator is set to "4" so that the first four characters (the "T," the "h," the "e," and the space) will be skipped and the title will be filed under "emperor's."





MARC Subfields

- | Most **Fields** (except Fields 001 through 009) contain several related pieces of data called **Subfields**, which are marked by Subfields codes and Delimiters
- | For example, the field for a book's physical description (tag 300) includes:

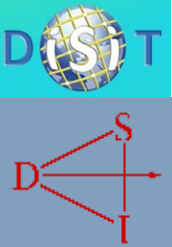
- ♣ a subfield (\$a) for the extent (number of pages)
- ♣ a subfield (\$b) for other physical details (illustration information)
- ♣ a subfield (\$c) for dimensions (centimeters)

- | 300 ## \$a 675 p. : \$b ill. ; \$c 24 cm.

- | **Subfield codes** are one lowercase letter (occasionally a number, eg. a,b,c) preceded by a delimiter (a character used to separate subfields, e.g. \$). Each subfield code indicates what type of data follows it. (For each field the MARC documentation lists and describes the valid subfield codes.)

- | **Delimiter:** Different software programs use different characters to represent the delimiter on the screen or on printouts. Examples are a double dagger (†), an "at sign" (@), a dollar sign (\$), an underline (_), or the graphic symbol " ".

- | In the example above, the subfield codes are \$a for the extent, \$b for other physical details, and \$c for dimensions.



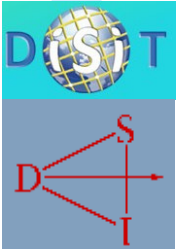


MARC General Rules: Tags divided by hundreds

- | There are some general rules that help define what all the numbers used as field tags mean (the notation XX is often used to refer to a group of related tags; e.g. 1XX refers to all the tags in the 100s: 100, 110, 130, and so on.)

| | |
|-----|--|
| 0XX | Control information, numbers, codes |
| 1XX | Main entry |
| 2XX | Titles, edition, imprint (in general, the title, statement of responsibility, edition, and publication information) |
| 3XX | Physical description, etc. |
| 4XX | Series statements (as shown in the book) |
| 5XX | Notes |
| 6XX | Subject added entries |
| 7XX | Added entries other than subject or series |
| 8XX | Series added entries (other authoritative forms) |

- | The 9XXs have been left for locally-defined uses, such as local barcode numbers. Local libraries, vendors, or systems can define and use them for attaching other types of information to records
- | The list of the most common tags shows how each fits into one of these divisions: 100 is an author main entry, 520 is a summary note, and so on.





MARC General Rules: Access Points (1)

| Access points (a main entry, subject added entries, and other added entries) are an important part of the bibliographic record. These are the headings for which separate cards were created for the traditional card catalog, and which a librarian can search in an online catalog

| Most of the access points are:

- 1XX fields (main entries)
- 4XX fields (series statements)
- 6XX fields (subject headings)
- 7XX fields (added entries other than subject or series)
- 8XX fields (series added entries)





MARC General Rules: Access Points (2)

- | For **names**, the best authority is:
 - ♣ the **Library of Congress Name Authority**
(<http://authorities.loc.gov>)
- | The form of the name used (personal name, corporate name, conference or meeting name, series title, or uniform title) can be checked against this authority
- | For **topics** or **geographic names**, the common subject authority lists are
 - ♣ the **Library of Congress Subject Headings**
 - ♣ the **Sears** List of Subject Headings.
- | The form of a subject heading should match one on the list or follow the rules for construction





MARC General Rules: Parallel content (1)

- | The fields requiring authority control are also the fields that use parallel tag construction. In general, in the 1XX, 4XX, 6XX, 7XX and 8XX fields, a personal name will have the last two digits 00.

- | Examples:

- ♣ for a main entry (1XX) that is a personal name (X00), the correct tag is 100
- ♣ for a subject heading (6XX) that is a personal name, the tag is 600, and so on

This parallel content can be summarized as follows:

| | |
|-----|----------------------|
| X00 | Personal names |
| X10 | Corporate names |
| X11 | Meeting names |
| X30 | Uniform titles |
| X40 | Bibliographic titles |
| X50 | Topical terms |
| X51 | Geographic names |

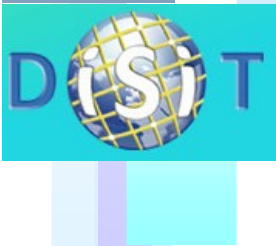
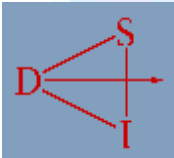




MARC General Rules: Parallel content (2)

I By combining this chart with the rule "Tags divided by hundreds", it becomes evident that:

- ♣ if the subject of a book (6XX) is a person (Lincoln, Abraham), the tag will be 600;
- ♣ if the subject of the book is a corporation (Apple Computer, Inc.), the tag will be 610;
- ♣ if the subject of the book is a topic (Railroads), the tag will be 650;
- ♣ if the subject of a book is a place (United States), the tag will be 651
- ♣ An added entry (7XX) for a joint author (a personal name) will have tag 700



Parallel content

X00 Personal names
X10 Corporate names
X50 Topical terms
X51 Geographic names

Tags divided by hundreds

6XX fields (subject headings)
7XX fields (added entries other than subject or series)



Unique information appears at the beginning of the MARC record (1)

- | Preceding the main bibliographic record parts the MARC record contains some less familiar information (automated cataloging systems usually provide default data or prompts to help a cataloger input this information):

- ♣ Leader
- ♣ Directory
- ♣ The 008 field

- | For example, there is a code in this field to indicate whether a book is large-print, a code to identify the country of publication, a code to identify juvenile materials, a code to indicate the language of the text, and so on





Unique information appears at the beginning of the MARC record: Leader

- | There are 24 positions in the Leader, numbered from 00 to 23
- | Each position has an assigned meaning, but much of the information in the leader is for computer use
- | MARC record creation and editing programs usually provide a window or prompts to assist the cataloger in filling in any leader data elements that require input.

Example:

05 Record status
a = increase in encoding level
c = corrected or revised
d = deleted
n = new
p = increase in encoding from prepublication (previous CIP)

- | More details required for interpreting the leader are defined here: <http://www.loc.gov/marc/umb/um07to10.html#part9>





Unique information appears at the beginning of the MARC record: Directory

- | Immediately following the Leader, is a block of data called a **Directory**: it tells what tags are in the record and where they are placed (by a count of the characters to the position where each field begins)
- | The directory is constructed (by computer) from the bibliographic record, based on the cataloging information, and, if any of the cataloging information is altered, can be reconstructed in the same way.
- | Here you can find additional information related to directory:
 - | <http://www.loc.gov/marc/umb/um11to12.html#part11>





Unique information appears at the beginning of the MARC record: the 008 field

- | The 008 field is referred to as Fixed-Length Data Elements, or Fixed Field Codes
- | Its 40 characters contain important information, but in an abbreviated form
- | Although it is not yet used to its fullest in online catalog systems, this field can be used to identify and retrieve records matching specific criteria
- | Examples, there are codes:
 - ♣ to indicate whether a book is large-print,
 - ♣ to identify the country of publication
 - ♣ to identify juvenile materials
 - ♣ to indicate the language of the text
 - ♣ Etc.
- | Here you can find a more detailed description:
 - | <http://www.loc.gov/marc/umb/um07to10.html#part10>





MARC Authority Fields are marked by Tags

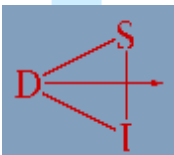
- | Each authority record (as described for the bibliographic records) is divided logically into **fields**
- | There is a field for the heading, fields for reference tracings, and so on
- | These fields are subdivided into one or more **subfields**
- | As previously noted, the textual names of the fields are too lengthy to be reproduced within each MARC record: they are represented by 3-digit tags

| | |
|---------|--|
| 100 tag | marks a personal name heading (author) |
| 400 tag | marks a personal name see from reference tracing |
| 500 tag | marks a personal name see also from reference tracing |
| 670 tag | marks a source data found note |



```
100 1# $a Woolf, Virginia, $d 1882-1941
```

- | More details here:
- | <http://www.loc.gov/marc/uma/pt8-11.html#pt8>





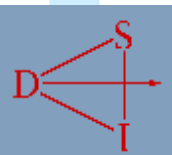
MARC Authority Indicators

- | As in the bibliographic records, two character positions follow each tag (except of Fields 001 through 009): one or both of these character positions may be used for **indicators**
- | When an indicator position is not used, that indicator is referred to as "undefined" and the position is left blank (character #)
- | Each indicator value is a number from 0 to 9
- | Even though two indicators together may look like a 2-digit number, they really are two single-digit numbers

Example:

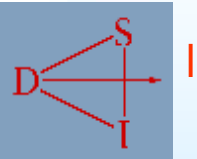
```
100 3# $a Guelf, House of
```

- | The **first indicator** value (3) in the personal name heading field indicates that the name is a family name, rather than an individual person's name
- | The **second indicator** value in field 100 is undefined (#)





MARC Authority Subfields



Most fields contain several related pieces of data, each type of data within the field is called a subfield, and each subfield is preceded by a delimiter-subfield code combination (Fields 001 through 009 have no subfields)

Example: the field for a personal name heading includes a subfield for:

- ♣ Personal name (\$a)
- ♣ Numeration (\$b)
- ♣ Titles (\$t)
- ♣ Dates (\$d)

associated with the name, among several others

```
100 0# $a Gustaf  
      $b V,  
      $c King of Sweden,  
      $d 1858-1950
```



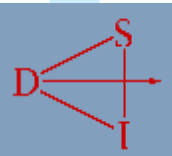
MARC Authority record

- | An authority record includes three basic components:
 - ♣ Headings
 - ♣ Cross references
 - ♣ Notes. (MARC authority records often also contain additional information.)

- | **Heading:** The standardized "authoritative" form of a name, subject, or title that is used for access points on bibliographic records. The purpose of using standardized names and subjects on bibliographic records is to help related records be retrieved together

- | **Tracings and references:** References that direct a user from a variant form of a name or subject to the authoritative form (this is called a see reference) or from one authoritative form to another authoritative form because they are related to one another (this is called a see also reference)

- | **Notes:** Notes that contain general information about standardized headings or more specialized information, such as citations for a consulted source in which information is either found or not found about a heading.





MARC Authority record: Headings (1)

- | An authority record is made for each authorized form of a heading
 - | For names, one of the most widely used authority files is the Library of Congress Name Authority File (or LCNAF, <http://authorities.loc.gov/>)
 - | For topics or geographic names, the most used subject authority file is the LCSH (Library of Congress Subject Headings)
 - | There are many other subject heading lists:
 - ♣ Sears List of Subject Headings
 - ♣ Art and Architecture Thesaurus
- The form of a subject heading should match one on these lists or follow their rules for construction





MARC Authority record: Headings (2)

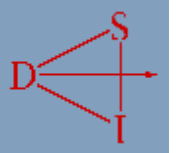
- | All headings are one of the following types: names, name/title combinations, uniform titles, or subjects
- | Examples of name headings include:
 - ♣ 1) A name heading (\$a) may be a personal, corporate, meeting, or jurisdiction (including geographic) name.

```
100 1#  $a Woolf, Virginia, $d 1882-1941  
          (personal name heading)  
110 2#  $a Association for Childhood Education International  
          (corporate name heading)  
111 2#  $a La Crosse Health and Sports Science Symposium  
          (meeting name heading)  
151 ##  $a Mexico  
          (geographic name heading)
```





MARC Authority record: Headings (3)



Examples of name headings include:

- 2) A name/title heading consists of both a name and a title of a work. The name portion (\$a) contains a personal, corporate, meeting, or jurisdiction name. The title portion (\$t) contains the title by which an item or a series is to be identified for cataloging purposes. It is especially used for classic works that have been published under different titles over time. An example of a name/title heading is:

```
100 1# $a Dostoyevsky, Fyodor,  
$d 1821-1881.  
$t Crime and punishment
```

- 3) A uniform title heading consists of a title not associated with a particular author. A familiar example is:

```
130 #0 $a Bible. $l Latin. $s Vulgate
```



MARC Authority record: Headings (4)

I Examples of name headings include:

- 4) A subject heading may be a single term, phrase, or a group of terms. The tag for the subject is based on the first term or phrase in the heading -- topic, genre/form, geographic, chronological, personal name, corporate name, meeting name, or title. The other terms in the heading are in specific subfields that are called subdivisions.

```
100 1# $a Shakespeare, William, $d 1564-1616  
      $x Criticism and interpretation $x History  
      $y 18th century  
150 ## $a Books and reading $z Argentina
```





MARC Authority record: Tracings and references

- | MARC authority records contain two types of cross references that either lead a user:
 - ♣ from an unauthorized heading to an authorized heading (these are called "see references" and use 4XX tags)
 - ♣ from one authorized heading to another (these are called "see also references" and use 5XX tags)

Authority record information:

100 1# \$a Twain, Mark, \$d 1835-1910

[Standardized heading]

400 1# \$a Conte, Louis de, \$d 1835-1910

[See from tracing]

500 1# \$a Clemens, Samuel Langhorne, \$d 1835-1910

[See also from tracing]





MARC Authority record: Tracings and references examples

- | Tracing cross references rather than making separate authority records for references enables efficient and accurate display of authority information by catalogs and the users see the information in the form that is appropriate for them
- | Cross reference displays may then be generated by systems which combine the contents of a tracing reference field and the 1XX heading field of a record

Display generated from the above record:

Twain, Mark, 1835-1910

see also: Clements, Samuel Langhorne, 1835-1910

Clemens, Samuel Langhorne, 1835-1910

see also: Twain, Mark, 1835-1910

Conte, Louis de, 1835-1910

see: Twain, Mark, 1835-1910





MARC Authority record: Notes

- | The MARC Authority record may contain notes
These may be intended for display in the public catalog, or only for viewing by librarians using the records in their work

Authority record information:

670 ## \$a Phone call to National Register of Historic Places
[Note for cataloging use]

680 ## \$i Surgery performed on an outpatient basis. May be
hospital-based or performed in an office or
surgicenter.
[Note for public catalog use]





MARC Authority record: Structural components - Leader

- | Like MARC bibliographic records, MARC authority records consist of three main components:
 - ♣ Leader
 - ♣ Directory
 - ♣ Variable Fields
- | The **Leader** provides information required for the processing of a record (is a 24 character positions long)
- | Much of the information in the Leader is for computer use in reading and processing the record and is computer generated
- | More information about the Leader:
 - | <http://www.loc.gov/marc/uma/pt8-11.html#pt10>





MARC Authority record: Structural components – Directory

- | The **Directory** is a series of entries that contain the tag, length, and starting location of each variable field within a record (each Directory entry is 12 character)
- | The Directory is always generated by the computer
- | More information about the Directory:
 - | <http://www.loc.gov/marc/uma/pt12.html#pt12>

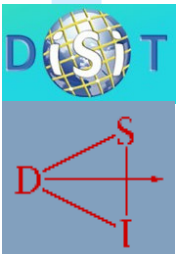




MARC Authority record: Structural components - Variable fields (1)

- | **Variable fields** are identified by a three-character numeric tag, each field ends with a field terminator character
- | There are two types of variable fields:
 - ♣ Variable *control* fields (the 00X fields)
 - ♣ Variable data fields
- | The variable data fields are grouped into blocks according to the first character of the tag:

| | |
|-----|---|
| 0XX | Standard numbers, classification numbers, codes |
| 1XX | Headings (authoritative and reference) |
| 2XX | Complex see references |
| 3XX | Complex see also references |
| 4XX | See from tracings |
| 5XX | See also from tracings |
| 6XX | Notes |
| 7XX | Linking entries |
| 8XX | Alternative graphics |
| 9XX | Reserved for local implementation |



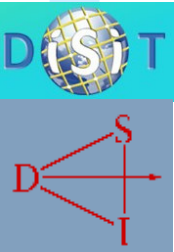


MARC Authority record: Structural components - Variable fields (2)

- | The following meanings, with some exceptions, are given to the final two characters of the tag of fields:

| | |
|-----|---------------------|
| X00 | Personal names |
| X10 | Corporate names |
| X11 | Meeting names |
| X30 | Uniform titles |
| X48 | Chronological terms |
| X50 | Topical terms |
| X51 | Geographic names |
| X55 | Genre/form terms |

- | More information about the Variable fields:
- | <http://www.loc.gov/marc/uma/pt8-11.html#pt8>

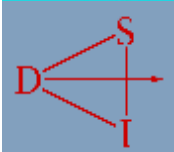




```
- <collection xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  - <record>
    <leader>00834nw 2200193n 4500</leader>
    <controlfield tag="001">CF 91000008</controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19960528091722.0</controlfield>
    <controlfield tag="008">910215acaaaaaa</controlfield>
    - <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a">CF 91000008</subfield>
    </datafield>
    - <datafield tag="040" ind1=" " ind2=" ">
      <subfield code="a">DLC</subfield>
      <subfield code="c">DLC</subfield>
    </datafield>
    - <datafield tag="084" ind1="0" ind2=" ">
      <subfield code="a">lcc</subfield>
    </datafield>
    - <datafield tag="153" ind1=" " ind2=" ">
      <subfield code="a">HF1</subfield>
      <subfield code="c">HF6182</subfield>
      <subfield code="j">Commerce</subfield>
    </datafield> [ ... ]
  </record>
  - <record>
    <leader>00877nw 2200181n 4500</leader>
    <controlfield tag="001">CF 91000424</controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19960528091739.0</controlfield>
    <controlfield tag="008">910605acaaaaaa</controlfield>
    - <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a">CF 91000424</subfield>
    </datafield> [ ... ]
  </record>
</collection>
```

MARC XML Instance

<http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml>





MARC Conversion formats

| Conversion from:

- ♣ MARC to SGML/XML (and the reverse)
- ♣ Dublin Core
- ♣ RDF
- ♣ FRBR
- ♣ OAI-ORE
- ♣ other metadata formats

| have been addressed





MARC Dublin Core Transformation

```
- <dc>
  <title>Arithmetic / </title>
  <creator> Sandburg, Carl, 1878-1967. </creator>
  <creator> Rand, Ted, ill. </creator>
  <type/>
  <publisher>San Diego :Harcourt Brace Jovanovich,</publisher>
  <date>c1993.</date>
  <language>eng</language>
- <description>
  A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
  </description>
  <description>One Mylar sheet included in pocket.</description>
  <subject>Arithmetic</subject>
  <subject>Children's poetry, American.</subject>
  <subject>Arithmetic</subject>
  <subject>American poetry.</subject>
  <subject>Visual perception.</subject>
</dc>
```





Link per marc marc21

- | <http://www.loc.gov/marc/uma/>
- | <http://www.loc.gov/marc/>
- | <http://www.bl.uk/bibliographic/exchange.html>
- | <http://www.loc.gov/marc/bibliographic/>
- | <http://www.loc.gov/marc/umb/um01to06.html>
- | <http://authorities.loc.gov/help/disphlp4.htm>
- | <http://www.loc.gov/marc/marcxml.html>





Elementi di modellazione e programmazione di contenuti digitali

**Corso di Laurea Magistrale in Metodologie
informatiche per le discipline umanistiche**

Ing. Michela Paolucci

Department of Systems and Informatics

University of Florence

Via S. Marta 3, 50139, Firenze, Italy

tel: +39-055-4796523, fax: +39-055-4796363

Lab: DISIT, Sistemi Distribuiti e Tecnologie Internet

<http://www.disit.dsi.unifi.it/>

paolucci@dsi.unifi.it

<http://www.dsi.unifi.it/~nesi>, <http://www.dsi.unifi.it/~paolucci>,

<http://www.axmedis.org> <http://mobmed.axmedis.org/>

