



InformaticaUmanistica

Lezione 7
Introduzione ai Metadati
IFLA/FRBR

Pasquale Savino

ISTI - CNR



UNIVERSITÀ DI PISA

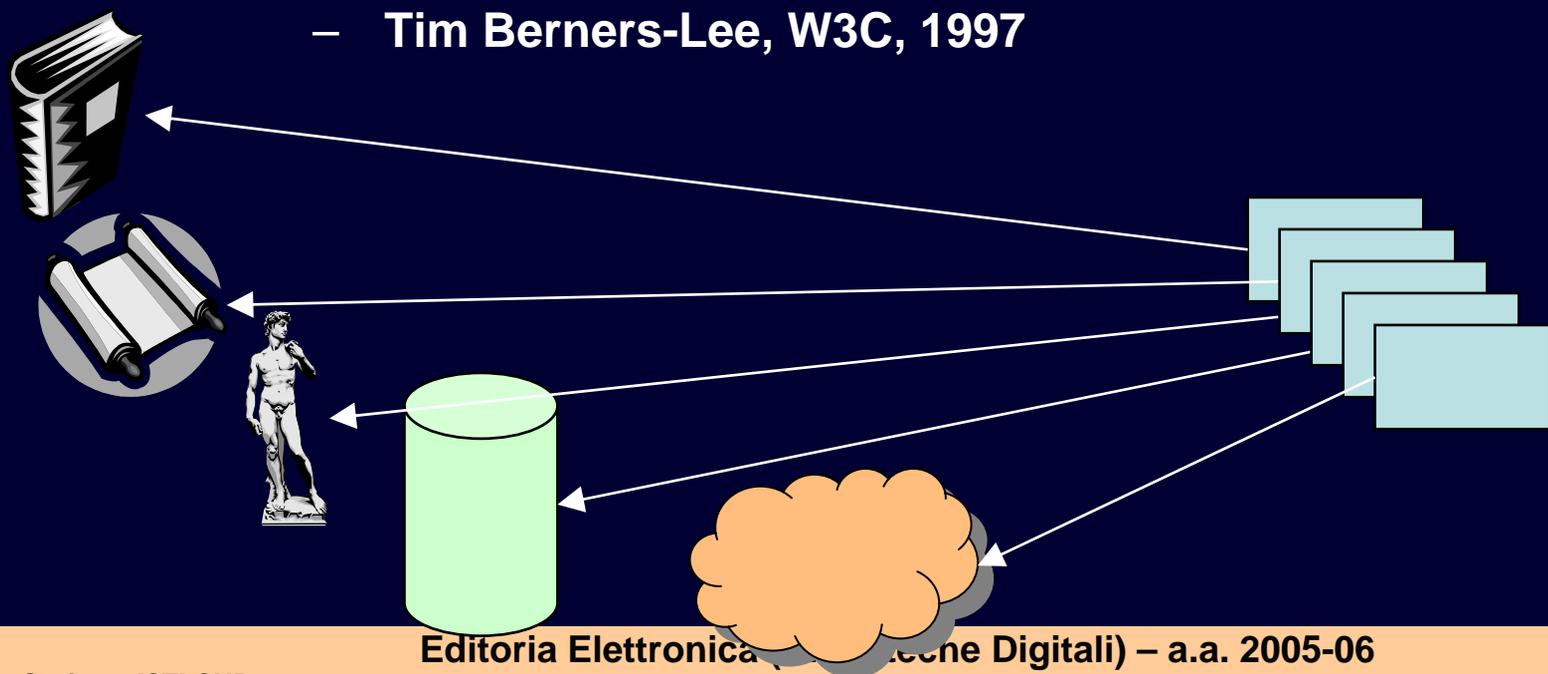
Sommario

- ◆ **Definizione**
- ◆ **I vari tipi di metadati**
- ◆ **Modelli di metadati**
 - IFLA/FRBR
 - MARC21
 - Dublin Core
 - MPEG-7
 - ECHO

Metadati - Definizione

Cosa sono i metadati?

- Alcune definizioni semplici...
- 'Structured data about data'.
 - **Dublin Core Metadata Initiative FAQ, 2003**
- Machine-understandable information about Web resources or other things.
 - **Tim Berners-Lee, W3C, 1997**



Una visione “funzionale” dei metadati

- Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics.
A user might be a program or a person.
 - **Lorcan Dempsey & Rachel Heery, 1998**
- Structured data about resources that can be used to help support a wide range of operations
 - **Michael Day, 2001**

A quali risorse possono essere associati dei metadati?

- I metadati possono essere associati a qualunque cosa – digitale, fisica, risorsa “astratta”

- HTML documents
- digital images
- databases
- books
- museum objects
- archival records
- metadata records
- Web sites
- collections
- services
- physical places
- people
- institutions
- abstract “works”
- concepts
- events

Che informazione contengono i metadati?

- I record di metadati includono
 - Record bibliografici appartenenti a cataloghi o ottenuti da servizi di indicizzazione e creazione di abstract
 - Descrizione del materiale archiviato
 - Record relativi ad oggetti contenuti nella documentazione del museo
 - Descrizione di oggetti digitali (documenti, immagini, video software, ...)
 - Descrizione di collezioni di oggetti digitali
 - Descrizione di servizi in rete
 - Descrizione di record di metadati

Quali operazioni possono essere effettuate sui metadati?

- Operazioni da parte degli utenti o da parte di tool software
- I metadati possono essere usati per supportare diverse funzioni, quali ad esempio
 - **resource disclosure & discovery**
 - **resource management, (incluso preservation)**
 - **intellectual property rights management**
 - **Autenticazione ed autorizzazione**
 - **Personalizzazione e localizzazione dei servizi**
- **Le diverse funzioni richiedono diversi tipi e classi di metadati**
 - **Non è possibile trovare una soluzione unica per tutti gli usi**
 - **È necessario specificare i requisiti funzionali per definire le caratteristiche dei metadati da utilizzare**

Elementi ed insiemi di elementi dei Metadati

- I Metadati descrivono gli attributi e le proprietà di una risorsa
- Ogni attributo o proprietà è descritto da un **metadata element**
 - Può essere identificato e definito in modo preciso
 - Ogni metadata element può essere rappresentato in diverse forme
- Un **metadata element set** è
 - Insieme coerente di elementi
 - Creato per uno scopo specifico, come una singola unità
- Schema
 - Rappresentazione strutturata di un element set

Autore = J Smith
Data = 2001-11-05
Titolo = Report

Resource1

I tipi di metadati

Tipi di metadati

- ◆ **Metadati descrittivi o per “resource discovery”**
 - Utilizzati per descrivere le risorse della Biblioteca Digitale
 - Utilizzati per facilitare la ricerca di documenti nella BD
 - Cataloghi, indici, ecc.
- ◆ **Metadati strutturali**
 - Descrivono la struttura degli oggetti della BD e le relazioni che esistono tra oggetti (o loro parti) ed altri oggetti
- ◆ **Metadati amministrativi**
 - Utilizzati per la gestione delle collezioni di documenti e per il controllo dell'accesso

Metadati per resource discovery

- Gli utenti desiderano
 1. Individuare le risorse in accordo a determinati criteri
 - Ad es. *Trova tutti i documenti scritti da ... che parlano di ...*
 2. (oppure) identificare una specifica risorsa
 - Essere in grado di distinguere tra diverse risorse simili tra loro
 3. Selezionare
 - **Scegliere una risorsa che sia appropriata ai bisogni**
 4. Localizzare la risorsa
 5. *Ottenere/accedere alla risorsa*
 6. *Utilizzare la risorsa*
 - ***open, read, display, run, play, copy, unpackage/repackage***
 - ***Interpretare il contenuto***
- I metadati per resource discovery sono utilizzati principalmente per le operazioni 1-4

Metadati per resource discovery

Continuum di complessità/funzionalità

Indici full-text, feature per immagini	Generato con tool software in modo automatico. Secondo alcune definizioni potrebbero non essere considerati metadati	discovery per contenuto, localizzazione
Forme semanticamente semplici (e.g. Dublin Core)	Permettono la descrizione di un gran numero di risorse In alcuni casi possono essere generati in modo automatico, in altri sono prodotti manualmente	discovery, identificazione, selezione, localizzazione
Forme complesse più ricche (e.g. MARC, ECHO, etc)	Relativo a tipi di risorse specifiche Spesso associato con domini/comunità specifici La creazione può richiedere un significativo intervento umano	discovery, identificazione, selezione, localizzazione, accesso, uso

Associazione di una risorsa ai metadati [1/3]

Creator = J Smith

Date = 2001-11-05

Title = Report

Le metadati possono essere contenuti all'interno della risorsa

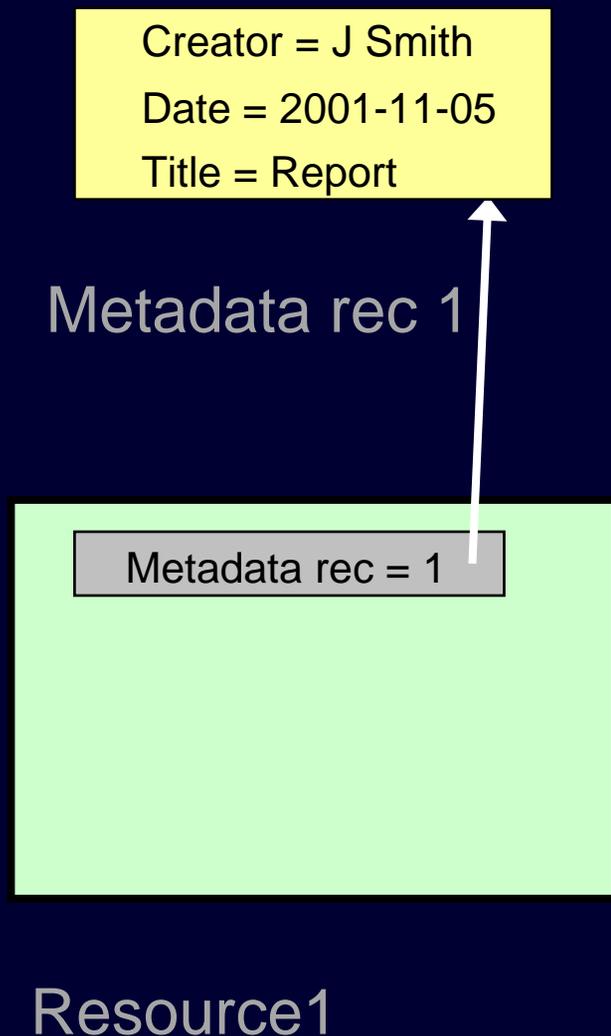
e.g. elementi `meta` in doc HTML;

Proprietà del documento nei word processor

- Le risorse possono permettere che i metadati siano contenuti al loro interno (embedding)?
- Il creatore dei metadati può avere i diritti di **scrittura** sulla risorsa?
- Può un servizio **estrarre** i metadati che sono contenuti nella risorsa?
- Possono esistere metadati per **aggregati di risorse**? Come vengono gestiti?

Resource1

Associazione di una risorsa ai metadati [2/3]



*Il record dei metadati è un oggetto separato
L'identificatore del record e' contenuto
nella risorsa*

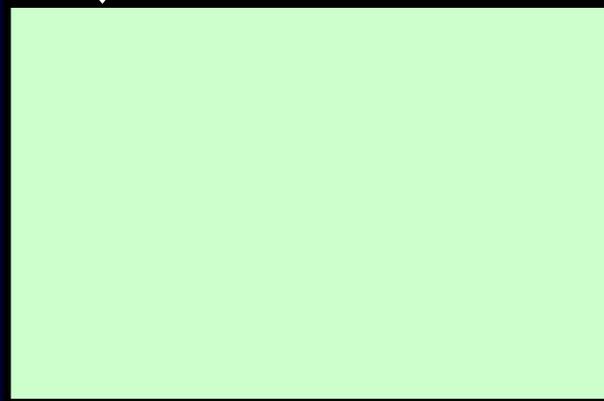
e.g. elementi `link` elements in doc HTML
I record di metadati possono essere remoti
rispetto alla risorsa

- Le risorse devono poter contenere i link
- Il creatore dei metadati può avere i diritti di **scrittura** sulla risorsa?
- Devono esistere dei meccanismi per seguire i link
- Cosa succede se una risorsa viene cancellata?
- Possono esistere metadati per **aggregati di risorse**? Come vengono gestiti?

Associazione di una risorsa ai metadati [3/3]

Doc = 1
Creator = J Smith
Date = 2001-11-05
Title = Report

Metadata rec 1



Resource 1

*Il record dei metadati è un oggetto separato
L'identificatore della risorsa e' contenuto nel
record*

Il record dei metadati può essere remoto rispetto alla risorsa

Non richiede che nessuna informazione sui metadati sia contenuta nella risorsa Non richiede che il creatore dei metadati possa modificare la risorsa

Il record dei metadati può essere creato indipendentemente dalla risorsa – possibilità di record multipli

I servizi utilizzano i metadati indipendentemente dalla risorsa

Il record dei metadati può continuare ad esistere anche se la risorsa viene cancellata

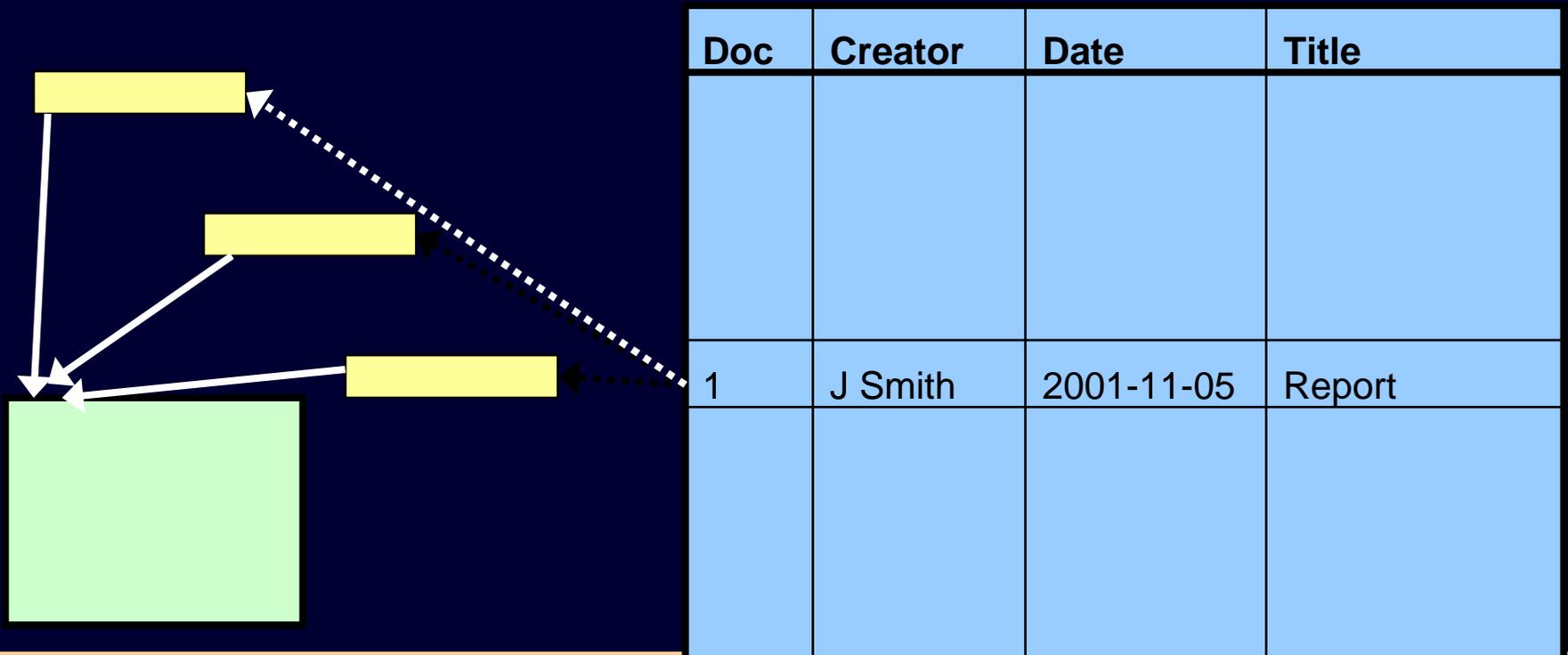
Il record dei metadati può descrivere qualunque oggetto che abbia un identificatore

Gestione dei metadati

Il record dei metadati viene usato separatamente dalla risorsa che viene descritta

I metadati sono essi stessi una risorsa

Il contenuto dei metadati può essere memorizzato nella forme più appropriate per il servizio: in un database, in forms, ecc.



IFLA – FRBR

Cos'è IFLA – FRBR

- ◆ **IFLA è la International Federation of Library Association** (<http://www.ifla.org/>)
- ◆ **A partire dal 1990 IFLA si è posta l'obiettivo di definire i "Functional Requirements for Bibliographic Records (FRBR)"**
 - The purpose of the study is to delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs. The study is to cover the full range of functions for the bibliographic record in its widest sense--i. e., a record that encompasses not only descriptive elements, but access points (name, title, subject, etc.), other "organizing" elements (classification, etc.), and annotations.
- ◆ **Questa attività si è conclusa nel 1997 con l'approvazione del report finale** (<http://www.ifla.org/VII/s13/frbr/frbr.htm>)

Concetti generali

◆ **Obiettivi di FRBR**

- Correlare i dati contenuti nei record bibliografici ai requisiti degli utenti
- Definire un livello minimo di funzionalità per i record bibliografici creati dalle diverse agenzie nazionali

◆ **Vari tipi di dati vengono considerati**

- Testo, musica, audio-video, grafica, cartografia

◆ **... vari tipi di media**

- Carta, pellicola, nastro magnetico, supporti di memorizzazione ottica, ecc.

◆ **... vari tipi di formati**

- Libri, dischi, cassette, ecc.

Utenti ed applicazioni

◆ Vari tipologie di utenti

- Lettori, studenti, ricercatori, personale di una biblioteca, publishers, distributori, gestori dei diritti intellettuali, ecc.

◆ Varietà di applicazioni

- Sviluppo, acquisizione e catalogazione di una collezione di documenti
- Produzione di bibliografie
- Conservazione
- Information retrieval

◆ Esempi di utilizzo

- Determinare se un certo tipo di risorsa esiste
- Verificare se uno specifico documento è disponibile
- Tenere traccia degli spostamenti di un determinato documento
- Controllare gli accessi ad un documento
-

Metodologia

- ◆ **Un modello di metadati definisce in modo chiaro ed univoco i metadati utilizzati da una biblioteca digitale. Esso è associato ad uno schema di una base di dati (dove sono memorizzati i metadati veri e propri).**
- ◆ **FRBR utilizza il modello Entità – Relazione utilizzata per la rappresentazione del modello concettuale delle basi di dati.**

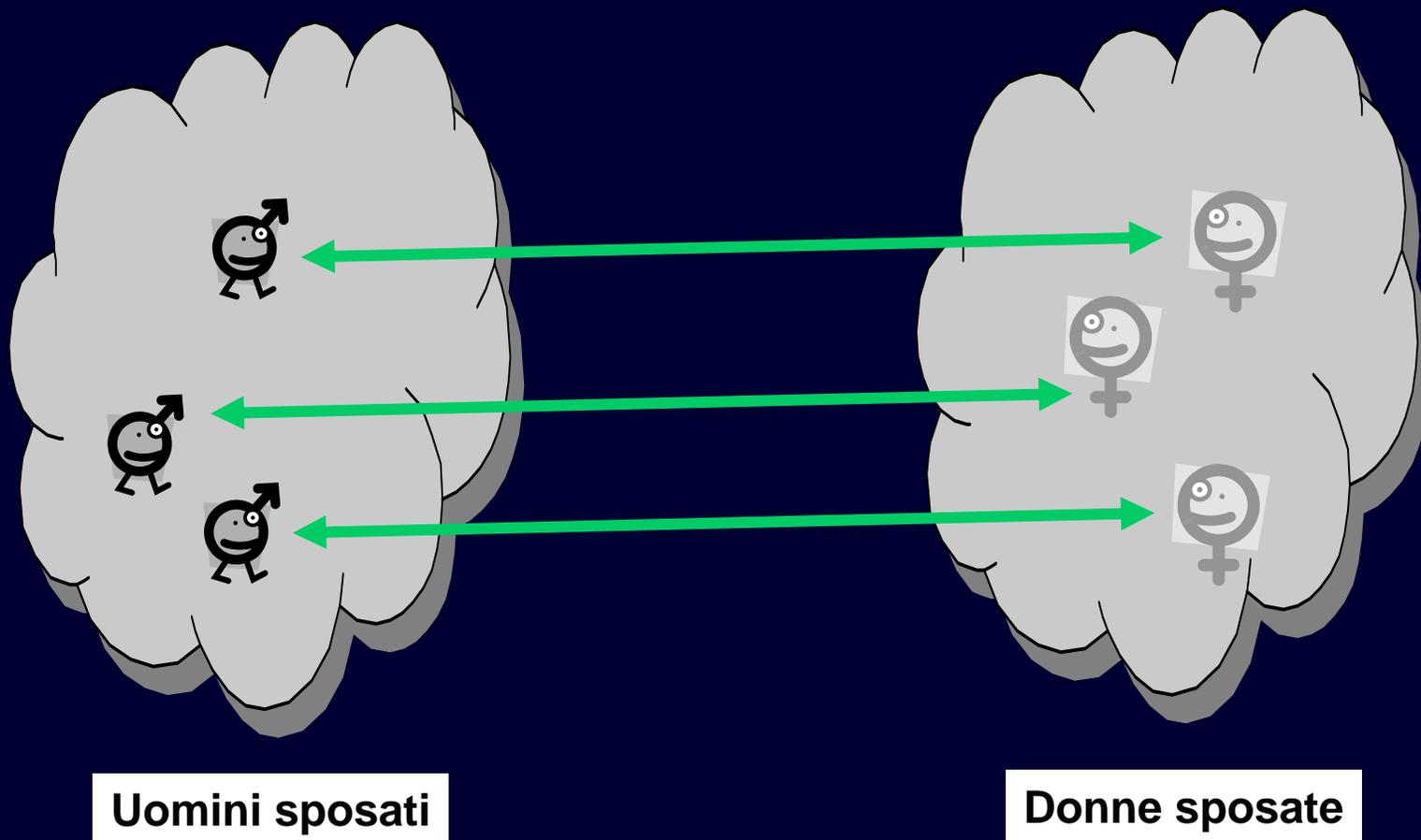
Le entità

- ◆ Per Entità si intende in generale un qualsiasi elemento della realtà che può essere chiaramente individuato.
- ◆ Ogni entità possiede un certo numero di attributi, che nel nostro caso sono i campi dei metadati.
- ◆ Le entità sono raggruppate in classi e quindi l'entità è in pratica un'istanza di una classe. Per non appesantire la trattazione useremo quasi sempre il termine entità, anche quando si parlerà di classi. Il termine classe sarà adottato quando ci sarà qualche possibilità di confusione.

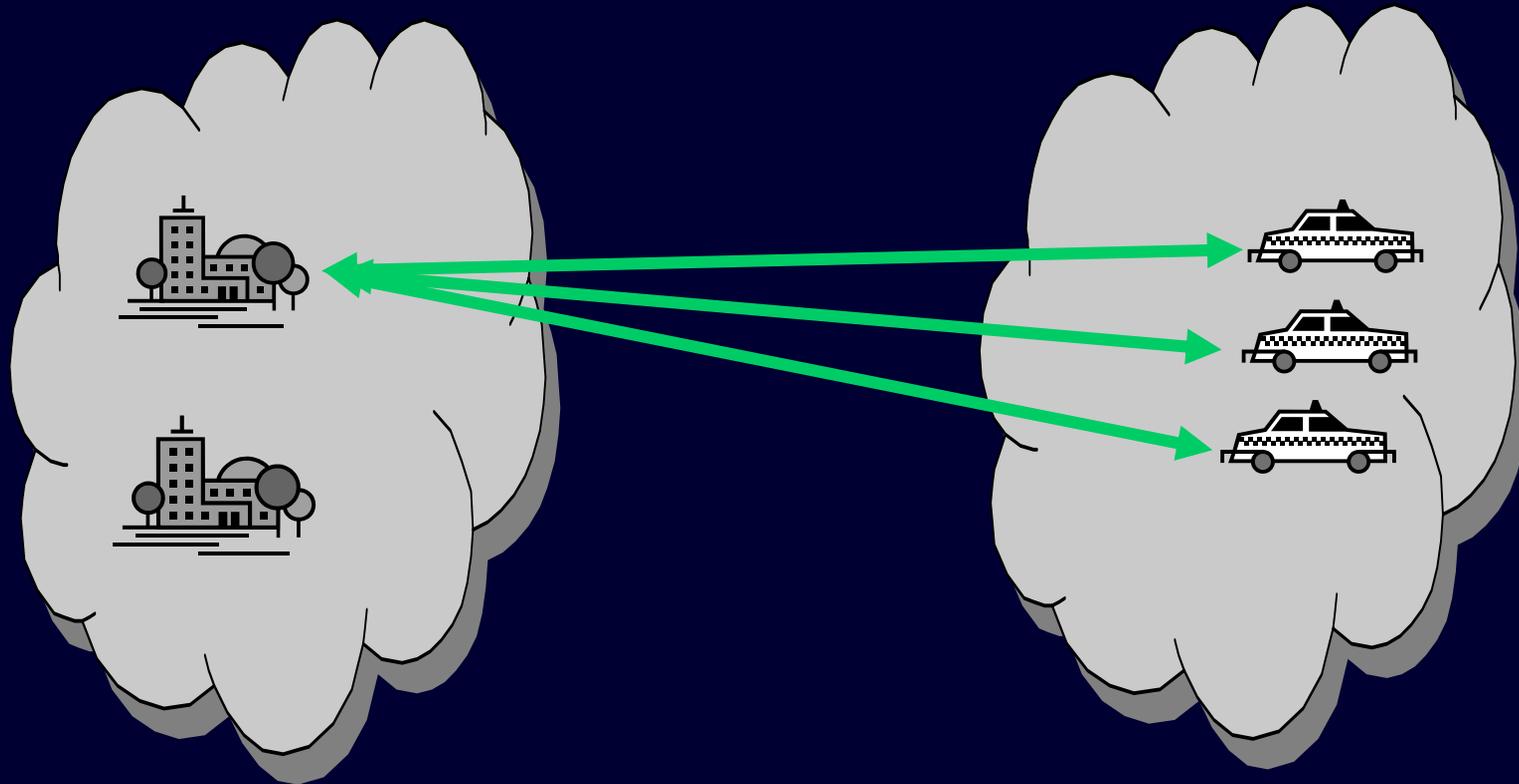
Le relazioni

- ◆ **Le relazioni permettono di mettere di collegare tra loro le entità.**
- ◆ **Date due entità arbitrarie di due classi A e B esistono tre tipi di relazioni che possono legarle:**
 - *uno–a–uno*, fa corrispondere 1 entità di A ad 1 entità di B. Non può esistere alcun entità, né di A né di B, che compaia due volte, abbinata con entità diverse.
 - *uno–a–molti*, in tal caso ad un'entità di A possono corrispondere più entità di B. Non è, invece possibile che la stessa entità di B corrisponda a entità distinte di A.
 - *molti–a–molti*, in tal caso a entità distinte di A possono corrispondere entità distinte di B e viceversa.

Esempio di relazione uno-a-uno



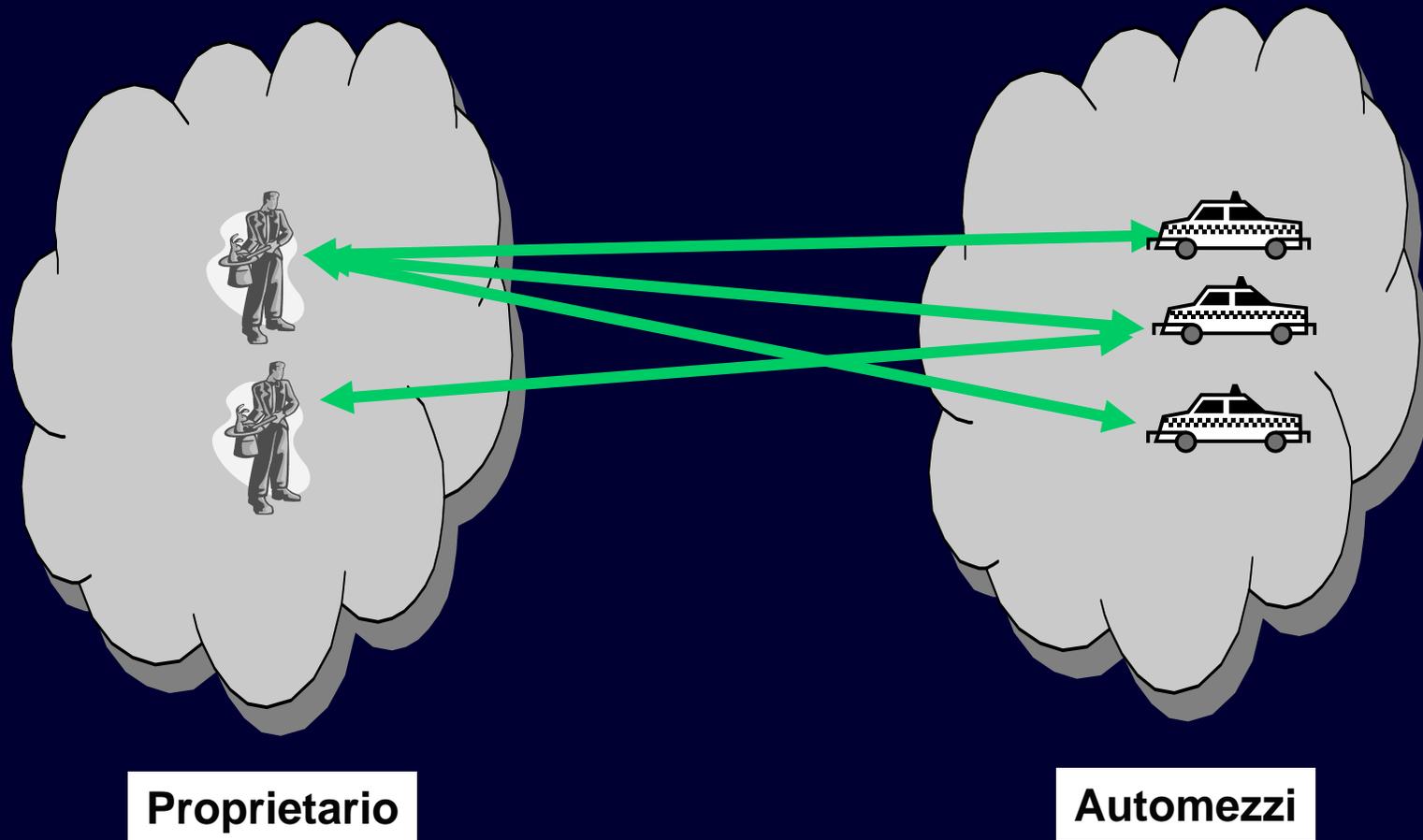
Esempio di relazione uno-a-molti



Compagnie di Assicurazione

Automezzi

Esempio di relazione multi-a-molti



Le Entità in FRBR

◆ Gruppo 1

- Rappresentano i diversi aspetti di interesse per un utente in un'opera intellettuale o artistica
- Work, Expression, Manifestation, Item

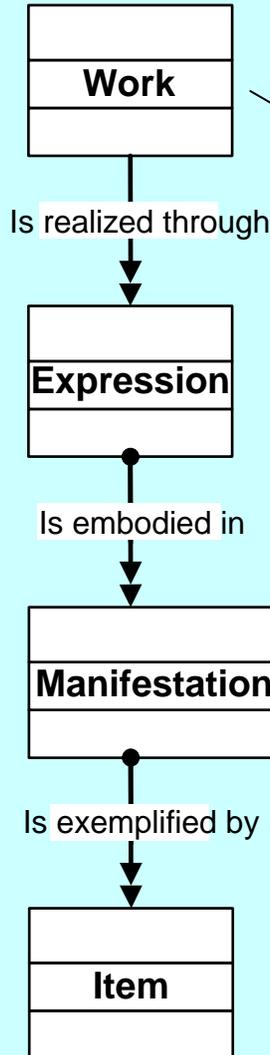
◆ Gruppo 2

- Entità responsabili per il contenuto intellettuale o artistico dell'opera, per la sua produzione e disseminazione
- Person, Corporate Body

◆ Gruppo 3

- Entità utilizzate come soggetti dei “work”
- Concept, Object, Event, Place

Gruppo 1

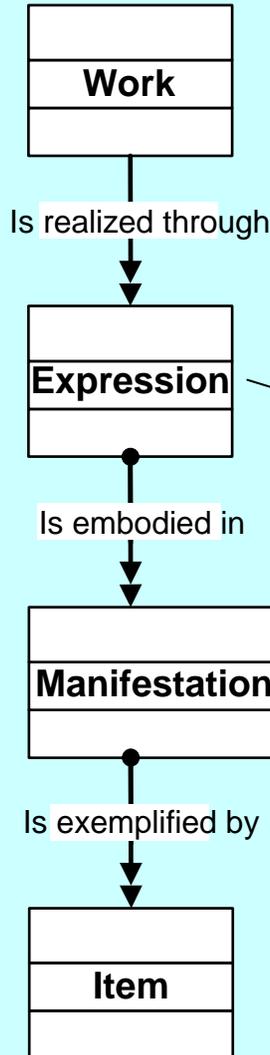


Una opera intellettuale o artistica.

Si tratta di una entità astratta che non rappresenta nessun oggetto fisico specifico. Può essere individuata attraverso le espressioni.

Ad es. L'Iliade di Omero è un WORK. In questo caso non ci riferiamo ad una specifica edizione dell'Iliade, ma alla creazione intellettuale.

Gruppo 1

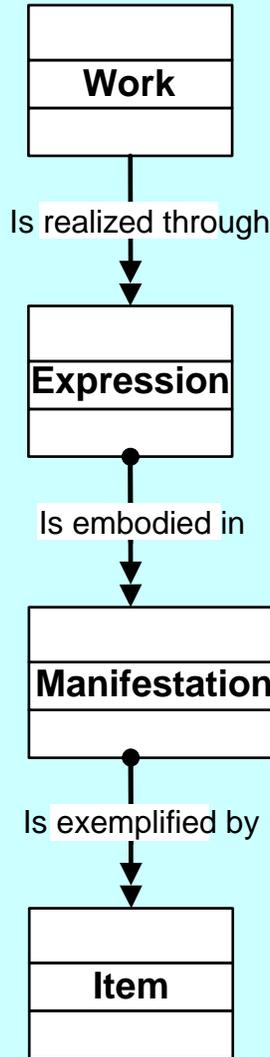


L'entità di Expression è la realizzazione del WORK sotto forma di testo, musica, suono, immagini, ecc. o una qualunque loro combinazione.

L'Expression non specifica aspetti che riguardano la forma fisica della realizzazione.

Ad esempio, possiamo avere una o più espressioni del WORK Iliade di Omero: una espressione narrata, una scritta, una rappresentata come opera teatrale, ecc.

Gruppo 1

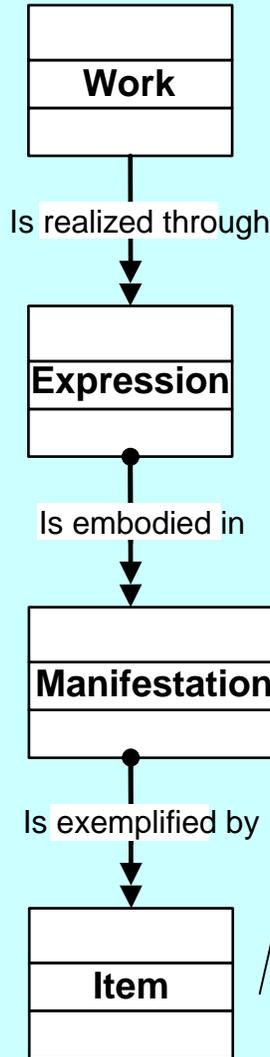


Definisce la rappresentazione fisica di una specifica espressione.

Può includere un ampio numero di materiali quali manoscritti, libri, registrazioni sonore, film, CD-ROM, ecc.

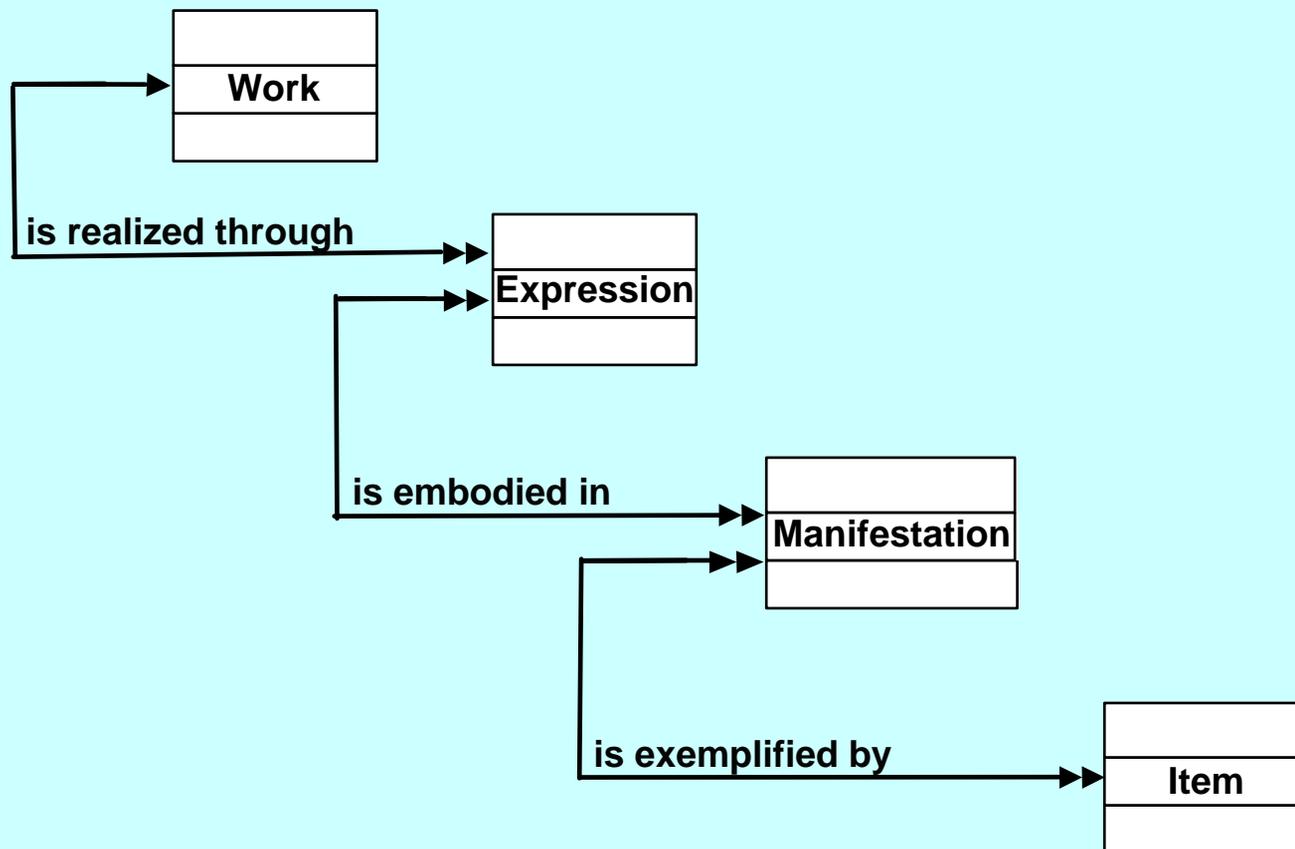
Ad esempio, ogni espressione dell'Iliade di Omero avrà una o più manifestazioni: l'espressione scritta può avere diverse edizioni, oppure l'espressione video può avere una manifestazione su CD-ROM ed una su nastro.

Gruppo 1

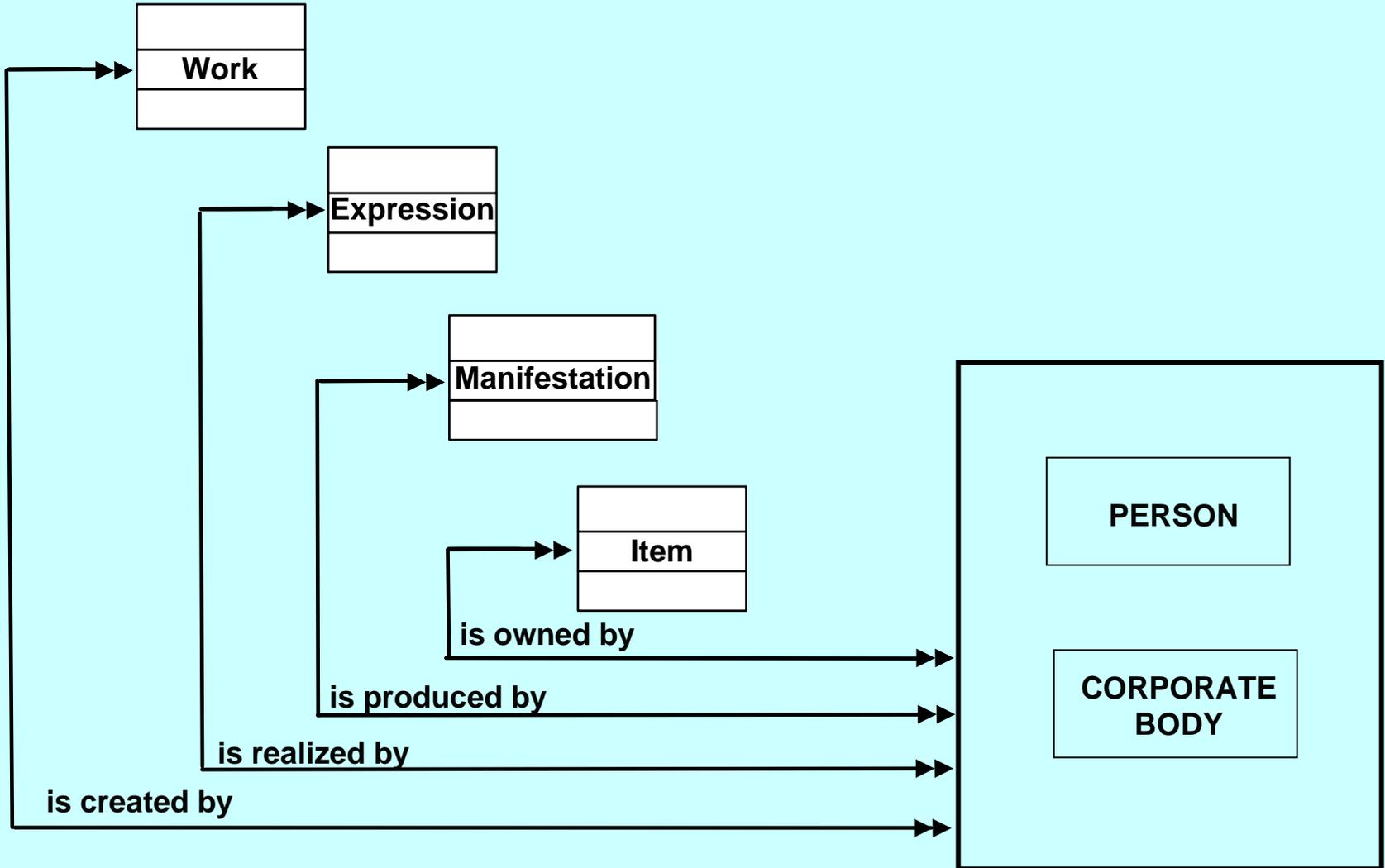


L'entità ITEM corrisponde ad un singolo esemplare di una manifestazione. Si tratta di una entità concreta che descrive le singole istanze di una manifestazione. Ad esempio, per una certa edizione della versione a stampa dell'Iliade di Omero, avremo un determinato numero di copie che possono essere individuate singolarmente.

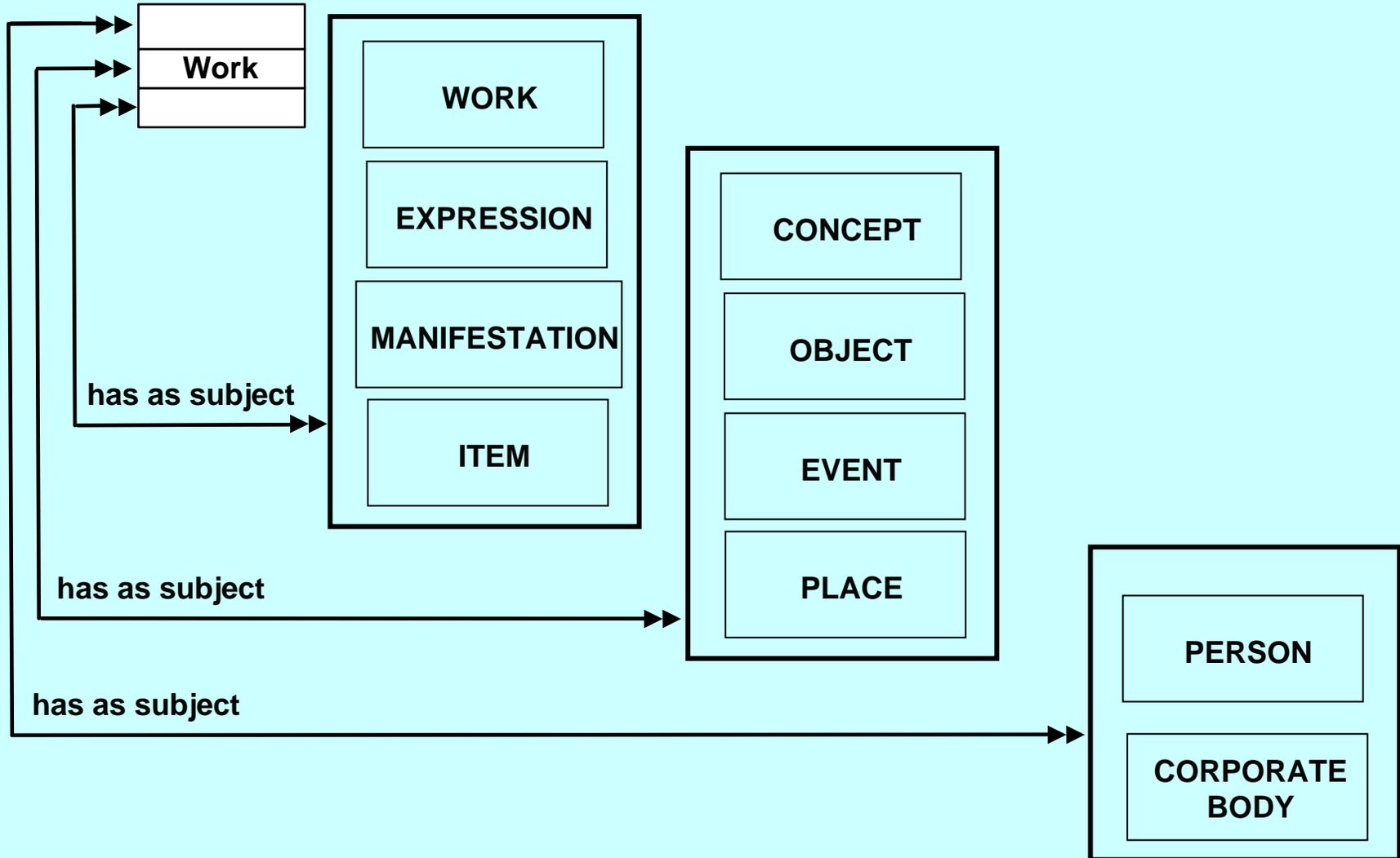
Relazioni tra entità del Gruppo 1



Gruppo 2



Gruppo 3



WORK [1/3]

◆ Una creazione artistica o intellettuale distinta

◆ Entità astratta

- Non esiste nessun oggetto materiale associato a un WORK
- Individuato attraverso le sue ESPRESSIONI
- Difficile definire precisamente le separazioni tra WORK
 - Ad es. revisioni diverse dello stesso testo, testi con parti aggiunte, traduzioni, arrangiamenti musicali diversi, ecc.
 - w_1 – J.S. Bach: L'arte della fuga
 - > e_1 : lo spartito originale per organo
 - > e_2 : un arrangiamento per orchestra da camera
 - w_2 – A. Manzoni: I promessi sposi
 - > e_1 : testo ed illustrazioni della prima edizione
 - > e_2 : testo ed illustrazioni della seconda edizione
 - w_3 – S. Kubrik: 2001 Odissea nello spazio
 - > e_1 : versione originale inglese
 - > e_2 : versione italiana

WORK [2/3]

- **Difficile definire precisamente le separazioni tra WORK**
 - **Viceversa, se esistono differenze significative per quanto riguarda il valore artistico o intellettuale, si possono avere WORK diversi**
 - **Ad es. adattamenti da una forma artistica ad un'altra vengono considerati nuovi WORK**
 - **Analogamente le riscritture o adattamenti per un tipo di pubblico diverso.**
 - **w_1 – J.S. Bach: L'arte della fuga**
 - **w_2 – Edizione semplificata dell'arte della fuga di J.S. Bach**

 - **w_1 – S. Kubrik: 2001 Odissea nello spazio**
 - **w_2 – Arthur C. Clark: 2001 Odissea nello spazio**

WORK [3/3]

◆ WORK come entità permette

- Connettere tutte le diverse entità del WORK
 - Ad es. possiamo definire la relazione tra tutte le espressioni del WORK **Divina Commedia** e del WORK **Commento alla Divina Commedia**
- Stabile una relazione implicita tra le diverse espressioni di uno stesso WORK
 - Ad es. tutte le traduzioni

EXPRESSION

- ◆ **La realizzazione del WORK sotto forma di testo, musica, immagini, video, ecc.**
- ◆ **Non include gli aspetti relativi alla forma fisica delle realizzazione (ad es. font di caratteri, formato video, ecc.)**

MANIFESTATION [1/2]

- ◆ **Corrisponde alla rappresentazione fisica di una espressione**
- ◆ **Possono essere utilizzati diversi tipi di materiali, quali manoscritti, libri, registrazioni sonore, film, CD-ROM, ecc.**
- ◆ **Possono esistere copie singole o diverse copie. L'insieme delle copie costituisce la manifestazione.**
- ◆ **La distinzione tra diverse manifestazioni della stessa espressione è data dalle caratteristiche fisiche dell'oggetto**
 - Ad es. variazioni nel font, nella forma delle pagine
 - Variazioni nel supporto fisico
 - Variazioni nel contenitore

MANIFESTATION [2/2]

◆ Esempi

- w_1 – J.S. Bach: L'arte della fuga
 - e_1 : lo spartito per organo
 - > m_1 : edizione originale
 - > m_2 : edizione pubblicata da Ricordi nel 1972
 - e_2 : un arrangiamento per orchestra da camera
 - > m_1 : registrazione su dischi 33 giri del 1965
 - > m_2 : registrazione su CD-ROM del 1991
- w_2 – A. Manzoni: I promessi sposi
 - e_1 : testo ed illustrazioni della prima edizione
 - e_2 : testo ed illustrazioni dell'ultima edizione
 - > m_1 : edizione in broccura
 - > m_2 : edizione economica

ITEM

- ◆ **Singolo esemplare di manifestazione**
- ◆ **Possono esservi differenze tra i diversi item associati alla stessa manifestazione, se queste differenze non sono state opera di chi ha prodotto gli item (ad es. per danneggiamento)**
- ◆ **Esempi**
 - **w₂ – A. Manzoni: I promessi sposi**
 - **e₁: testo ed illustrazioni della prima edizione**
 - > **m₁: edizione del 1823**
 - **i₁: copia autografa dell'autore**
 - **i₂: copia acquisita nel 1923**

Attributi

- ◆ **Ad ogni entità è associato un insieme di attributi, utilizzati per le interrogazioni e per la descrizione dell'entità**
- ◆ **Due categorie di attributi**
 - Attributi inerenti, ad es. caratteristiche fisiche, informazioni ottenibili dal documento (titolo, autore, ecc.)
 - Attributi esterni, quali ad es. categoria di appartenenza, informazione contestuale, ecc.

Relazioni

- ◆ **Permettono di descrivere i legami tra le diverse entità**
- ◆ **Possono essere utilizzate per “navigare” tra le entità**
- ◆ **Le relazioni più semplici sono quelle tra entità di tipo diverso, quali ad es. tra WORK ed EXPRESSION o tra EXPRESSION e MANIFESTATION**
- ◆ **Esistono relazioni tra entità dello stesso tipo, ad es. tra WORK e WORK**
 - Ad es. supplement, summarization sono relazioni tra WORK diversi

Riferimenti

- ◆ **Functional Requirements for Bibliographic Records – Final Report, 1998**
(<http://www.ifla.org/VII/s13/frbr/frbr.pdf>)
- ◆ **What is FRBR?**
(<http://www.loc.gov/cds/downloads/FRBR.PDF>)
- ◆ **Carlo Ghilli, Mauro Guerrini, Introduzione a FRBR, Editrice Bibliografica**