

Catalogazione e rappresentazione della conoscenza: dalla scheda cartacea al web semantico

Seminario LIMES, 29 marzo 2019

Oreste Signore

Pisa

oreste.signore@gmail.com



C'era una volta... (1)

(Tutta questa parte storica remota è un contributo di Irene Buonazia)

❖ **Adolfo Venturi (con Cavalcaselle, Aleardi e Prati)**

- ◆ **1888: mancano linee guida unitarie e persone con competenze specifiche**
- ◆ **Catalogazione secondo criteri emanati nel 1875 dal ministro Bonghi**
- ◆ **Esperienze di Cavalcaselle e Morelli per Marche e Umbria (1861-61)**
 - ✓ *Sostanzialmente elenchi di beni, per conoscere cosa si possedeva a seguito dell'unità d'Italia*

C'era una volta... (2)

❖ **1888: circolare ministeriale contenente le «Norme per la catalogazione delle schede di catalogo degli oggetti d'arte»**

◆ **Triplice copia**

- ✓ *Consegnatario*
- ✓ *«chi possa o debba assumersi, all'occorrenza, la rinnovazione della consegna»*
- ✓ *Ministero*

C'era una volta... (3)

❖ **Informazioni:**

- ◆ **Provincia, comune, collocazione precisa (chiesa, oratorio, monumento)**

- ◆ **Indicazioni generali**

 - ✓ *Soggetto, descrizione, materia e dimensioni, nome dell'autore*

- ◆ **Ubicazione storica e corrente**

 - ✓ *(essenzialmente per poter ricollocare l'opera)*

- ◆ **Stato di conservazione e restauri subiti**

- ◆ **Condizioni giuridiche**

- ◆ **Studio dell'oggetto**

 - ✓ *Indicazioni storico critiche e bibliografia*

C'era una volta... (4)

- ❖ **Venturi prevede un continuo aggiornamento delle schede**
 - ◆ **Catalogo a schede mobili**
 - ◆ **Impostazione scientifica e collettiva**
- ❖ **Convegno dei Soprintendenti del 1938**
 - ◆ **Roberto Longhi** propone il catalogo delle cose d'arte e delle pubblicazioni connesse
 - ✓ *Evitando descrizioni vaghe e articolate*
 - ✓ *Corredando la descrizione di documentazione fotografica*
 - ◆ **Giulio Carlo Argan** propone la creazione dell'Istituto Centrale del Restauro

Nasce l'ICCD

❖ 1964-67

- ◆ **commissione di studio** sui problemi della catalogazione costituita dal CNR d'intesa col Ministero della P.I. (Argan)

❖ 1969

- ◆ nasce **l'Ufficio Centrale del Catalogo** (Ferrari)

❖ 1975

- ◆ nasce il **Ministero dei Beni Culturali**
- ◆ l' **Ufficio Centrale del Catalogo** diventa l' **Istituto Centrale per il Catalogo e la Documentazione** (Ferrari)

Conoscere per conservare

- ❖ È il concetto chiave che sta alla base del Catalogo
- ❖ La **conoscenza**:
 - ◆ è il fattore fondamentale
 - ◆ è un patrimonio dello studioso
 - ◆ non viene comunicata esplicitamente agli altri
- ❖ Lo studioso è in grado di **contestualizzare** le informazioni

La scheda cartacea

- ❖ **Compilata da esseri umani, è destinata ad un lettore umano**
- ❖ **La presenza di conoscenza tacita da parte dell'esperto non costituisce un problema**
- ❖ **Strutturazione dell'informazione molto ridotta:**
 - ◆ **suddivisione in sezioni contenenti informazioni semanticamente pertinenti**
- ❖ **Informazioni talvolta ambigue o incoerenti**
- ❖ **Principio informatore: “*una scheda per ogni oggetto*”**

Quanti tipi di scheda? (1)

❖ Beni mobili

- ◆ RA (reperto archeologico)
- ◆ OA (opera d'arte)
- ◆ N (numismatica)
- ◆ E (etnografia)
- ◆ MI (matrice d'incisione)
- ◆ D (disegni)
- ◆ S (stampe)

Quanti tipi di scheda? (2)

❖ Beni immobili

- ◆ A (architettura)
- ◆ PG (parchi e giardini)
- ◆ MA (monumenti archeologici)
- ◆ CA (complessi archeologici)
- ◆ SAS (saggio stratigrafico)

Quanti tipi di scheda? (3)

❖ Beni urbanistico-territoriali

- ◆ SU (settore urbano)
- ◆ CS (centro storico)
- ◆ TP (toponimo)
- ◆ T (territorio comunale)

Quanti tipi di scheda? (4)

❖ Beni demo-antropologici

- ◆ FKO (oggetti)
- ◆ FKN (narrativa)
- ◆ FKM (musica)
- ◆ FKC (cerimonie)

Perché tanti tipi di schede?

- ❖ **Descrizione di oggetti intrinsecamente diversi**
 - ◆ **Proprietà diverse**
 - ◆ **Proprietà non applicabili allo specifico oggetto**
- ❖ **Retaggio di distinzioni disciplinari**

Le prime esperienze informatiche

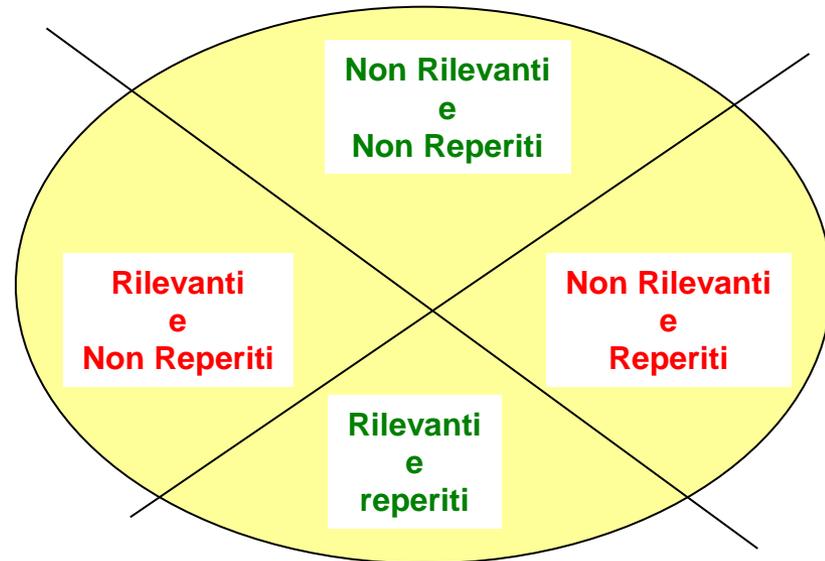
- ❖ A metà degli anni '70
- ❖ **Trasposizione** della scheda cartacea
- ❖ **Utilizzo di sistemi di Information Retrieval**
 - ◆ Molto entusiasmo
 - ◆ Qualche risultato deludente
 - ✓ *Dove sono le anfore del I a.C.?*
 - ✓ *Come mai non riesco a selezionare gli oggetti che hanno delle parti in bronzo, ma non sono solo di bronzo?*

Ricordiamo sempre...

❖ Recall e Precision

$$R = \frac{\|Doc_{rilevanti} \cap Doc_{reperiti}\|}{\|Doc_{rilevanti}\|}$$

$$P = \frac{\|Doc_{rilevanti} \cap Doc_{reperiti}\|}{\|Doc_{reperiti}\|}$$



❖ Parametri da valutare

sempre contemporaneamente

❖ **Indicizzazione esaustiva + linguaggio specifico =
alto RICHIAMO e alta PRECISIONE**

Verso quale ambiente orientarsi?

❖ Information Retrieval Systems

- ◆ Gestiscono bene le informazioni descrittive in testo libero
- ◆ Determinano ridondanze e gestiscono male le relazioni

❖ Database Management Systems

- ◆ Approccio che richiede una forte formalizzazione
- ◆ Alcune esigenze verrebbero costrette

❖ Artificial Intelligence

- ◆ Forse troppo avveniristico all'epoca, soprattutto per un contesto così vario e ricco?

Un approccio unificante

- ❖ Identificazione di **proprietà comuni** per descrivere oggetti di pertinenza di ambiti disciplinari diversi
 - ◆ RA e OA hanno molti punti di contatto
- ❖ Perché non utilizzare le stesse proprietà per descrivere aspetti sostanzialmente omogenei?
- ❖ Sarebbe possibile superare le differenze disciplinari?

La scelta

- ❖ **Nasce il modello di oggetto**
- ❖ **Maggiore formalizzazione**
- ❖ **Approccio:**
 - ◆ **non legato ad uno specifico software**
 - ◆ **unificato**
 - ✓ ***S. Papaldo, M. Ruggeri, R. Gagliardi, D. R. Matteucci, G. Romano, O. Signore, Strutturazione dei dati delle schede di catalogo. Beni mobili archeologici e storico-artistici, Roma-Pisa 1985 (prima edizione), 1988 (seconda edizione)***

L'oggetto di catalogazione

- ❖ Oggetto per il quale siano stati rilevati dati descrittivi e/o storico artistici
- ❖ Ogni oggetto deve essere individuato sia nella sua **interezza** che in base alle caratteristiche di **particolari rilevanti**
- ❖ Ogni voce deve descrivere uno specifico aspetto (attributo) dell'oggetto
- ❖ Quindi non più:
 - ◆ «*Monumento sepolcrale con statue*», o «*altare con ancona*» o «*calice e patena*»
- ❖ Ma la voce oggetto deve contenere il nome di **un solo oggetto**

Il modello di oggetto

- ❖ **Categorie di oggetti:**
 - ◆ oggetto **semplice**
 - ◆ oggetto **complesso**
 - ◆ **aggregazione** di oggetti
- ❖ **Il modello**
 - ◆ non vuole interferire nella classificazione scientifica inerente le varie discipline
 - ◆ è puramente **strumentale** per la gestione automatica
 - ◆ vuole permettere un certo grado di **soggettività**
- ❖ **L'appartenenza di un oggetto ad una categoria non è implicita nel tipo di oggetto, ma dipende esclusivamente dalla quantità e dalla qualità delle informazioni che si possono riportare**

Oggetto semplice

- ❖ I dati anagrafici, fisico-descrittivi e storici sono validi per l'oggetto nella sua **totalità**, che non presenta componenti meritevoli di menzioni specifiche
 - ◆ *Fibula, vaso, statua, dipinto*
- ❖ **Insieme di oggetti** per cui non è possibile o opportuno schedare separatamente i singoli pezzi
 - ◆ Oggetti riprodotti con tecnica o criterio seriale
 - ◆ Paramento liturgico (di cui non si ritiene opportuno schedare i singoli componenti)

Oggetto complesso

- ❖ **Oggetto singolo logicamente e/o fisicamente composito**
 - ◆ **Crocifisso di cui è opportuno schedare separatamente la croce e il Cristo**
 - ◆ **Sarcofago con coperchio scolpito**
- ❖ **Serie di più oggetti diversi formanti un oggetto complesso (anche astratto) **identificato da un nome****
 - ◆ *Ciclo di affreschi*
 - ◆ *Paramento liturgico (composto da pianeta e stola)*
- ❖ **Gli oggetti componenti possono a loro volta essere degli oggetti complessi**

Aggregazione di oggetti

- ❖ Insieme di due o più oggetti di catalogazione correlati tra di loro da un **criterio concettuale**
- ❖ Formano un complesso unitario per il quale **non esiste un nome** che li identifichi nel loro insieme
 - ◆ *Calice e patena*
 - ◆ *Coppa e coperchio*
 - ◆ *Tazza e piattino*

Le relazioni tra le schede

- ❖ **Non più una scheda per ogni oggetto:**
 - ◆ una **serie di schede** collegate tra di loro
 - ◆ per la **struttura** riferimenti verticali (RVE)
 - ◆ per le **associazioni** riferimenti orizzontali (ROZ)
- ❖ **Le componenti degli oggetti complessi **ereditano** alcuni attributi dall'oggetto padre**
 - ◆ Alcune proprietà del componente sono **ereditate automaticamente** dalla scheda descrittiva dell'oggetto nel suo complesso
 - ◆ Alcune proprietà del componente sono da considerare **mancanti** se non esplicitamente specificate

Esempio di riferimento verticale

Oggetto	Livello	Rif. ogg. Principale
-----	-----	-----
Altare		10
Paliotto	1	10
Tabernacolo	2	10
Sportello tabernacolo	2.1	10

Esempio di riferimento orizzontale

Oggetto	Codice Catalogo	Rif. orizzontale
-----	-----	-----
Altare	10	10
Dipinto	90	10

La strutturazione: le grandi categorie

- ❖ **Codici di catalogazione**
- ❖ **Struttura dell'oggetto**
- ❖ **Localizzazioni**
- ❖ **Oggetto**
- ❖ **Cronologia**
- ❖ **Definizione culturale**
- ❖ **Dati tecnici**
- ❖ **Dati analitici**
- ❖ **Dati amministrativi**
- ❖ **Documentazione**

Strutturazione: campi, sottocampi...

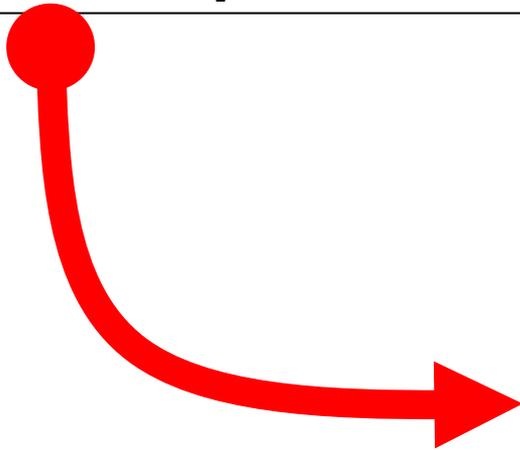
- ❖ Per ogni scheda una serie di:
 - ◆ **campi** (codice di tre lettere)
 - ✓ *Semplici o strutturati*
 - ◆ **sottocampi** (codice di quattro lettere)
 - ◆ spesso **ripetibili** singolarmente e nel loro complesso
 - ◆ spesso con **vocabolario controllato**

... e paragrafi

- ❖ Successivamente vengono aggiunti i **paragrafi** (codice di due lettere) anch' essi talvolta **ripetibili**
- ❖ L' informazione viene **frammentata**:
 - ◆ meno ambiguità ed errori
 - ◆ possibilità di ricombinare i singoli elementi
- ❖ **Perché codici di due, tre o quattro lettere?**

Oggetto e soggetto

OGGETTO: Dipinto raff. L'Annunciazione



OGT:
OGTD: Dipinto
SGT:
SGTI: L'Annunciazione

Un esempio di scheda

CD:

TSK: RA

[...]

OG:

OGT:

OGTD: parete affrescata

LC:

PVC:

PVCP: NA

PVCC: ERCOLANO

LDC:

LDCT: casa

LDCN: V 15 (DEL BICENTENARIO)

LDCS: 13 (tablino); parete N

[...]

DA:

DES:

DESS: Zona mediana rossa con tre pannelli riquadrati da bordi di tappeto: centrale (giallo in orig.) con quadro (Pasifae e Dedalo), laterali con medaglioni (Bacco a s.; baccante a d.) separati da fasce nere con grottesche, cornice a ovoli. Fregio nero: tre pannelli con scene di amorini in caccia con scudo e sfondo naturalistico a s., con cervo e cane in lotta al centro, con amorino su cavallo alato a d., separati da riquadri con maschere tragiche. Zona superiore rossa con architetture (quasi illeggibile).

[...]

I principi informatori

- ❖ **Compromesso tra:**
 - ◆ le esigenze **specialistiche** del mondo della ricerca
 - ◆ la definizione di una normativa **onnicomprensiva**
- ❖ **Modo di rappresentare la **conoscenza****
 - ◆ non sempre erano disponibili thesauri e dizionari
 - ◆ in alcuni casi rappresentata mediante la strutturazione
- ❖ **Pensato come un modello **interoperabile** per la raccolta delle informazioni**

- ❖ **Superamento dell' approccio dipendente dalla soluzione tecnologica adottata**
 - ◆ libertà di realizzare il proprio data entry
 - ◆ libertà di usare qualunque sistema per storage/retrieval
- ❖ **Ha colmato il divario esistente con altri paesi**
- ❖ **Salvaguardia degli investimenti in catalogazione:**
 - ◆ recupero dei dati dei “**Giacimenti culturali**”
 - ◆ il formato di scambio nasce in meno di un' ora!
- ❖ **Durevole nel tempo:**
 - ◆ le successive normative in XML sono sostanzialmente una trascrizione dello schema originale e delle sue evoluzioni

Le ombre

- ❖ **La percezione di**
 - ◆ «calata dall'alto»
 - ◆ eccessiva frammentazione
 - ◆ schema rigido
- ❖ **La mancanza, inizialmente, di strumenti per la compilazione**
- ❖ **I vincoli derivanti dai vocabolari controllati, authority files, thesauri, di aggiornamento lento e complesso**

I limiti dell' approccio

- ❖ **Visione centrata sulla “*scheda dell' oggetto*”**
 - ◆ informazioni ridondanti (es. autore, dati di scavo, etc.)
 - ◆ manca una visione “ricca” di associazioni
- ❖ **L'illusione dello «*schema concettuale universale*»**
- ❖ **Il modello:**
 - ◆ parte da una *rappresentazione concettuale* (mutuata dalla progettazione di basi di dati)
 - ◆ si attua in una “*linearizzazione*” dello schema
 - ◆ pone eccessiva *enfasi sull' oggetto*
 - ◆ non rappresenta tutta la varietà di *associazioni interdisciplinari* esistenti
- ❖ **La *conoscenza* dell' esperto rimane inespressa**

I vincoli

- ❖ **Tradizioni disciplinari**
- ❖ **Resistenza degli esperti a certe formalizzazioni**
 - ◆ **Datazioni**
 - ◆ **Rifiuto di un approccio basato sulla modellazione concettuale**
 - ◆ **Poca fiducia nella possibilità di restituzione della «scheda» a partire da informazione frammentata**
- ❖ **Flusso delle informazioni e modalità organizzative**
- ❖ **Sindrome della «**scheda completa e perfetta**»**

Formidabili quegli anni...

- ❖ **Tra il 1985 e il 1990 viene condotta una sperimentazione approfondita**
 - ◆ **realizzazione di thesauri**
 - ✓ *Suppellettile ecclesiastica*
 - ✓ *Banca dati storico-geografica*
 - ◆ **sperimentazioni dell' approccio **ipertestuale/ipermediale****
 - ◆ **recupero dei dati dei “Giacimenti culturali”**
- ❖ **Nel 1990 cessa la collaborazione tra CNUCE e ICCD**
 - ◆ **Il gruppo di progetto si scioglie e il progetto non viene completato**
 - ✓ *il formato di acquisizione e scambio viene inteso come formato di restituzione*
 - ✓ *l' attività sui thesauri perde impulso*

Beh, nessuno è perfetto!



La "Web Revolution"

- ❖ **Superamento dei meccanismi tradizionali di accesso all' informazione**
- ❖ **Architettura completamente decentralizzata**
 - ◆ **interoperabilità tecnica**
 - ◆ **interoperabilità semantica**
- ❖ **Emerge l' esigenza di rappresentare, esportare e condividere conoscenza**

Uno schema comune di metadati

❖ **Dublin Core Metadata Element Set**

✓ (<http://dublincore.org/>)

❖ **Pro**

- ◆ **Ampia accettazione**
- ◆ **Semplice e generale**

❖ **Contro**

- ◆ **Talvolta troppo generico (quindi i qualifiers)**
- ◆ **Item-centrico (specifica le proprietà dell'oggetto)**
- ◆ **Non è adatto per il ragionamento automatico**



Lo standard Dublin Core

- ❖ **Largamente utilizzato nei progetti internazionali per interoperabilità**
- ❖ **Viene spesso utilizzato nella versione “qualified”**
- ❖ **È basato su una visione che pone al centro l'oggetto**
- ❖ **La struttura della scheda può essere mappata agevolmente sullo standard DC**
(salvaguardia degli investimenti in catalogazione)

Ragionando su Dublin Core...

❖ Esempi (con una sintassi semplificata):

- ◆ **dc:title=Pietà**
dc:creator=Michelangelo
dc:date=1499
dc:subject=Madonna
dc:subject=Cristo

- ◆ **dc:title=Madonna del cardellino**
dc:creator=Raffaello
dc:date=1505
dc:subject=Madonna
dc:subject=Bambino

Dublin Core è sufficiente?

- ❖ Adotta una strutturazione meno “fine” rispetto a quella della scheda di catalogazione
- ❖ La **conoscenza** dell’ esperto rimane ancora inespressa
- ❖ Non permette di fare «*ragionamenti*» e dedurre nuova conoscenza: **non è un’ontologia**
- ❖ Non è possibile:
 - ◆ verificare la **coerenza** tra **dc:creator** e **dc:date**
 - ◆ trovare oggetti **coevi**, dello stesso artista o di suoi **contemporanei**
 - ◆ avere informazioni sul **contesto** storico o politico (**multidisciplinarietà**)
 - ◆ **Ritratti immaginari** (ragionamento temporale)

Un approccio ontologico

❖ CIDOC CRM

◆ Standard internazionale

✓ (<http://www.cidoc-crm.org/>)

◆ Alto livello di astrazione

◆ Estensibile

◆ Idoneo per supportare il **ragionamento** (spaziale, temporale)



Perché un approccio ontologico?

❖ Core metadata Vs core ontology

- ◆ Entrambi intendono supportare l'integrazione dell'informazione
- ◆ Differiscono per quanto concerne la comprensibilità da parte degli esseri umani
 - ✓ *Metadati previsti per l'elaborazione da parte degli **esseri umani***
 - ✓ *Core ontology modello formale per gli **strumenti automatici***
- ◆ **Vocabolari basati su ontologie:**
 - ✓ *I termini hanno una semantica ben definita*
 - ✓ *(quindi...) è possibile operare dei ragionamenti*
- ◆ **Limite dell'approccio basato sui metadati**
 - ✓ *Assunzione implicita di relazioni 1:N o N:N*

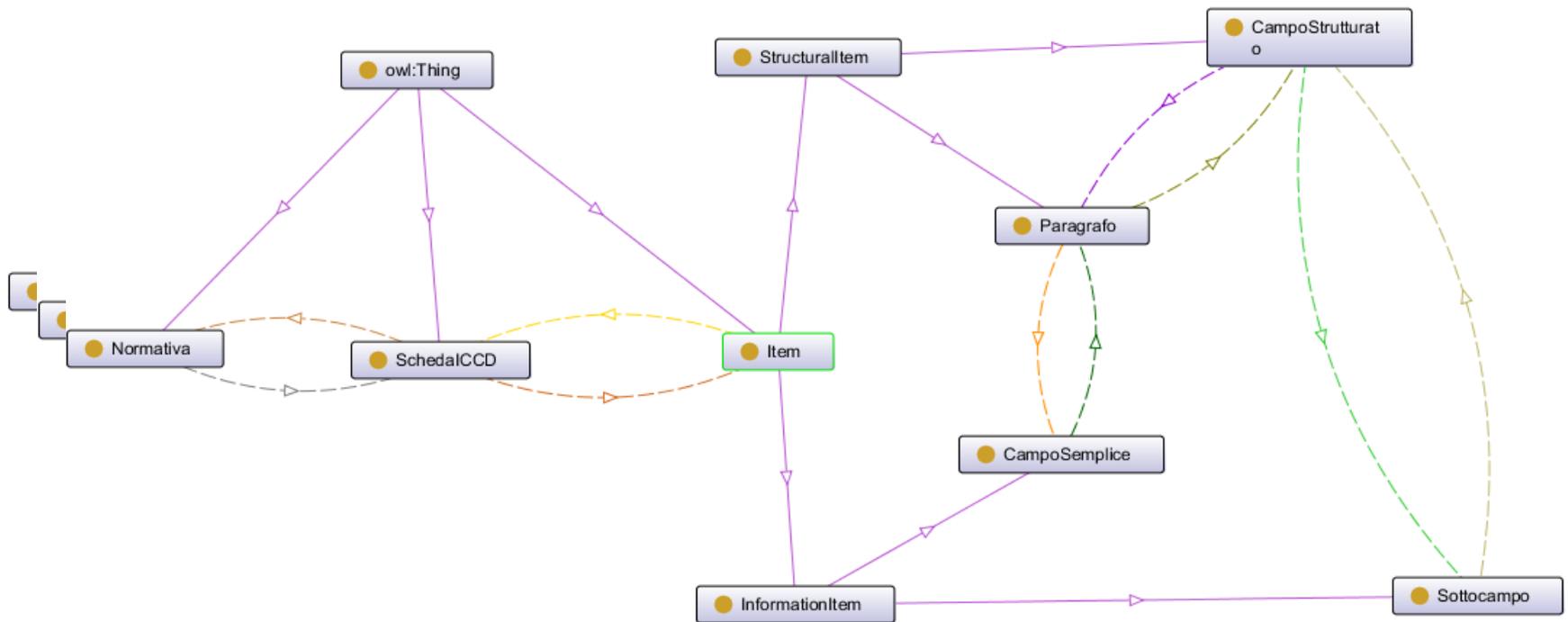
Un esempio: l'approccio catalografico classico

- ❖ Una “scheda” per ogni oggetto
- ❖ Schemi differenti nelle diverse culture
 - ◆ Strutture e metadati imposti da istituzioni
- ❖ Generalmente uno standard nazionale (ICCD in Italia)
 - ◆ Studiati per esperti
- ❖ Le differenze **sintattiche** prevalgono sulle equivalenze **semantiche**
- ❖ **Scarse o nulle** possibilità di integrazione e interoperabilità

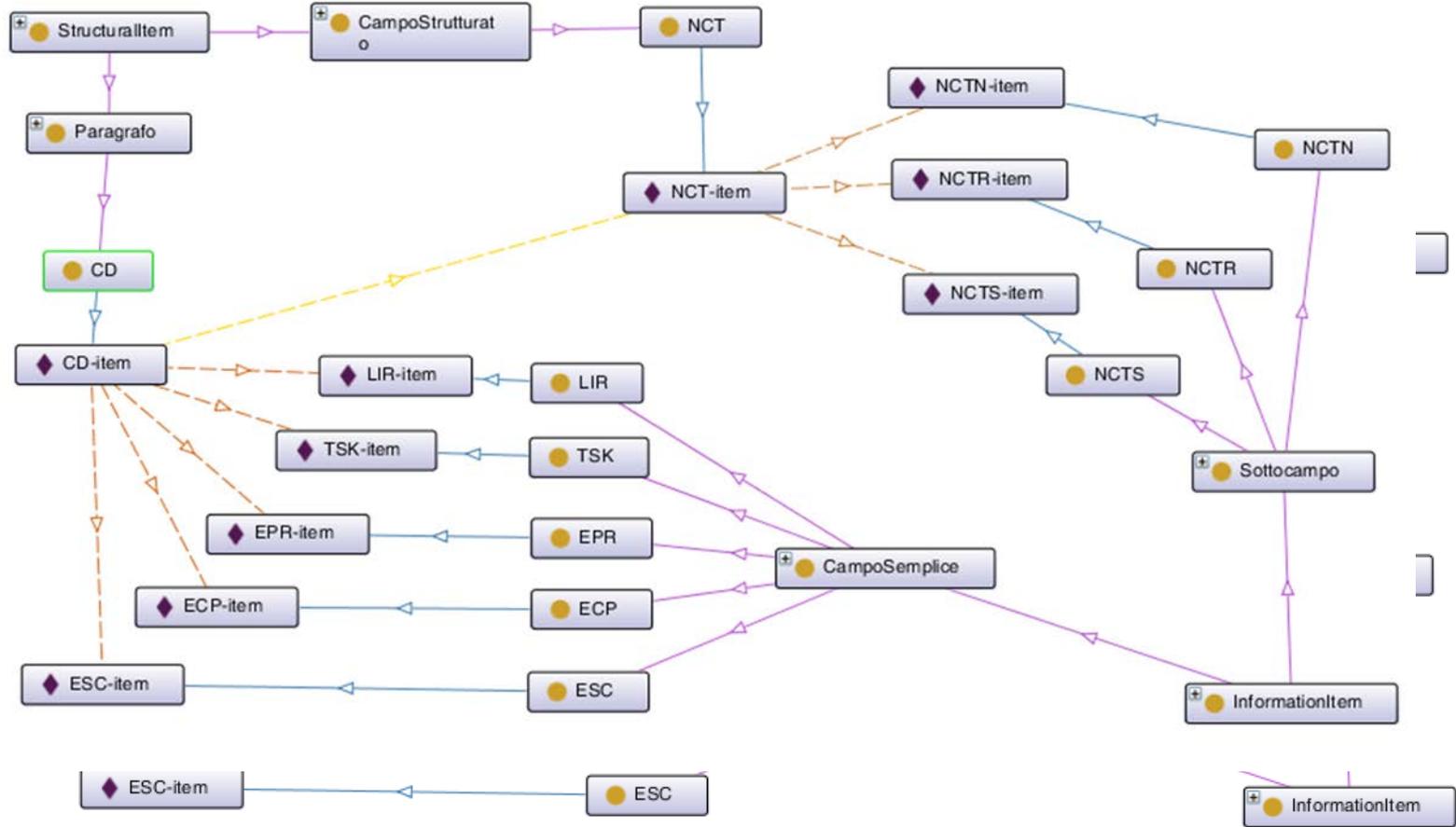
Il Web Semantico

- ❖ È un' infrastruttura basata sui **metadati** per poter operare ragionamenti sul Web
- ❖ **Estende** il Web attuale
- ❖ **La conoscenza**
 - ◆ può essere **rappresentata** in maniera **esportabile**
 - ◆ è **comprensibile dalle macchine**, che possono operare ragionamenti
 - ◆ è **condivisa**
- ❖ **Le ontologie** contengono la **rappresentazione esplicita** della conoscenza e le **regole** per **dedurre** di nuova
- ❖ È possibile **dedurre** nuova conoscenza

La struttura astratta della scheda



Dalla scheda all'ontologia



Un nuovo scenario?

- ❖ **Autorevolezza**, non **autorità**
- ❖ L'utente (o anche la **macchina**) potrà accedere all'informazione pertinente senza incontrare barriere
- ❖ Il sapere esistente nei vari luoghi potrà essere combinato in un **approccio interdisciplinare**
- ❖ La conoscenza degli studiosi sarà **patrimonio di tutti**

... o nihil sub sole novi?

- ❖ Il sapere degli studiosi dovrà essere **esplicitato e formalizzato** come ontologia
- ❖ Le **categorie concettuali** restano le stesse
- ❖ La conoscenza formalizzata in thesauri e authority file dovrà essere rappresentata in maniera esportabile (**ontologie**)
- ❖ Sarà necessario passare da una visione incentrata sull' oggetto a una visione in cui la **conoscenza, distribuita sul web**, dovrà essere integrata
- ❖ Il nuovo ambiente tecnologico **valorizzerà** la conoscenza esistente ed il lavoro svolto

Conclusioni

- ❖ La catalogazione è basata sulla **conoscenza**
- ❖ Nella catalogazione cartacea la conoscenza degli studiosi era un loro patrimonio e non veniva comunicata (**conoscenza tacita**)
- ❖ L' avvento dell' informatica ha reso disponibili le **informazioni**, ma non la **conoscenza**
- ❖ Le tecnologie del Semantic Web consentono un approccio **interdisciplinare** per **condividere la conoscenza**

Ringraziamenti

- ❖ Agli organizzatori della giornata, che mi hanno permesso di rivivere un periodo importante
- ❖ A **Irene Buonazia** per le informazioni «storiche»
- ❖ Ai funzionari ICCD con i quali ho condiviso tante esperienze (ed in particolare **Serenita Papaldo** e **Maria Ruggeri**)
- ❖ Un commosso ricordo al collega **Roberto Gagliardi**
- ❖ Un commosso ricordo a **Oreste Ferrari**, che con tanta lungimiranza individuò gli obiettivi del Catalogo.

Grazie per l' attenzione

Domande?

