

Corso di dottorato in Scienze bibliografiche, archivistiche e
documentarie e per la conservazione e restauro
dei beni librari ed archivistici

Ciclo XXII

Frammenti Semantici

Riflessioni su descrizioni archivistiche e web semantico:
Il caso dell'archivio Giovanni Testori

Salvatore Vassallo

Oggetto del lavoro di ricerca

- Descrizione archivistiche e web semantico
- “Data Esci” con informazioni granulari e fortemente strutturate
- L'archivio Giovanni Testori come casus studi
- Applicazioni in diversi campi: nuovi strumenti di ricerca, interfacce e sistemi flessibili, interscambio di dati archivistici
- Riflessioni corollarie su rapporto con dati bibliografici e strumenti del web 2.0

L'oggetto del presente lavoro di ricerca è stato in primo luogo quello di valutare e verificare la possibilità di esprimere descrizioni archivistiche direttamente del web semantico (descrizioni archivistiche in senso lato, quindi sia guide, censimenti, inventari etc e, in particolare, tutte le descrizioni conformi agli standard descrittivi archivistici).

L'output web, l'uscita a video, il cosiddetto “data esci” in questo caso non sarebbe una mera derivata, ma presenterebbe i dati mantenendo un'adeguata granularità dell'informazione, rendendoli in un certo senso autoesplicativi (soprattutto per la macchina) e facilmente ricercabili, collegabili etc.

Una simile riflessione è nata a partire da un caso concreto, rappresentato dall'archivio Giovanni Testori, ma è stata poi generalizzata e estesa con applicazioni pratiche che vanno dalla creazione automatica di interfacce software flessibili, alla creazione di strumenti di corredo e di ricerca di nuova generazione fino a fornire soluzioni per l'interscambio dei dati archivistici in particolar modo fra i grossi sistemi informativi aggregatori di dati (e fra questi e il neo nato Sistema Archivistico Nazionale).

La discussione intorno a una nuova generazione di strumenti di ricerca archivistici ha portato, come corollario, ad accenni sui possibili rapporti con dati bibliografici (particolarmente rilevanti nel caso di archivi letterari) e con gli strumenti del web 2.0

L'archivio Giovanni Testori

- Archivio o fondo? Un problema non banale
- Consistenza dell'archivio:
 - 107 quaderni
 - 71 cartelle contenenti 5291 fogli dattiloscritti
 - 4 quadernoni con parziale inventario dipinti
 - 2 scatoloni e 2 buste contenenti 306 articoli a stampa di Testori e 605 su Testori
 - 273 volumi di opere di Testori
- Problemi e scelte di descrizione

I materiali documentari costituenti l'Archivio Giovanni Testori e la biblioteca storica dell'autore sono stati acquisiti dalla Regione Lombardia in due tranches nel 2002 e nel 2003 e affidati in deposito alla Fondazione Mondadori. Si tratta di materiale composito e in effetti la distinzione fra archivio e fondo non è banale. Si tratta per molti aspetti di ciò che in letteratura, specificatamente dal prof. Romiti, è definito un archivio improprio, in quanto non è sempre possibile determinare la presenza di un vincolo che leghi le carte o quantomeno il legame fra le carte è meno viscoso, un vincolo attenuato.

Materiali come gli articoli su e di Testori, raccolti anche postumi da Alain Toubas si avvicinano maggiormente alla collezione che all'archivio.

Come descrivere un materiale tanto fluido?

La scelta operata in fase di riordino da parte di Antonella Bilotto e Paola Gallerani è stata quella di considerare come unità archivistica l'opera, filo di arianna per guidare il lettore nella consultazione delle carte. La realtà fisica del fondo è stata relegata a un inventario sintetico, che riproduce invece il susseguirsi di materiali nei quaderni, nelle cartelle etc.

Questa scelta è sicuramente molto forte e per certi versi criticabili, a tal punto che Lombardia Beni Culturali Archivi (il vecchio PLAIN) ha scelto di recepire (sotto la supervisione di Saverio Almini) esclusivamente l'inventario sintetico che riproduce la struttura fisica dell'archivio e solo dei materiali realmente prodotti dall'autore.

La domanda è se sia possibile compendiare le esigenze diverse in strumenti ibridi. E se sì come.

Obiettivi ricerca

- Possibilità di descrivere archivi direttamente nel web semantico
- Soluzioni per affrontare difficoltà di descrizione di documentazione fluida
- Verificare se simili problemi possono essere risolti attraverso gli strumenti del web semantico
- Estendere simili soluzioni ad altri contesti

Gli obiettivi del lavoro di ricerca sono dunque riassumibili in una serie di domande:

- è possibile descrivere archivi direttamente nel web semantico?
- come affrontare la necessità di descrivere situazioni fluide, archivi impropri etc senza compiere scelte archivisticamente criticabili?
- gli strumenti del web semantico possono essere adottati per fornire una soluzione a questo problema?
- se un simile approccio permette di ottenere descrizioni archivistiche maggiormente flessibili, è auspicabile estendere tali considerazioni dalla valle (le descrizioni archivistiche) alla fonte (il software che le produce)?

Gli strumenti usati: le Topic Maps

- Standard ISO 13250
- I soggetti e la loro rappresentazione: i topics
- Parliamo la stessa lingua: i PSI
- Le caratteristiche di un topic: nomi, occorrenze e associazioni
- L'ambito di validità di una determinata asserzione
- Un linguaggio vincolato per determinare schemi di validazione
- La necessità di una notazione grafica chiara per esperti di dominio

La tecnologia usata per questa sperimentazione è stata quella delle Topic Maps, una tecnologia del web semantico affine al forse più noto RDF. Si tratta di uno standard ISO 13250, che è in realtà una famiglia di standard volti a definire vari aspetti della tecnologia: dal data model, a un linguaggio di serializzazione XML sino a una notazione grafica per creare esempi. A questa famiglia di standard si aggiungono lo standard ISO 18048 (un linguaggio di interrogazione specifico per questa tecnologia) e lo standard ISO 19756 (un linguaggio vincolato per esprimere schemi di validazione).

L'elemento chiave di questa tecnologia è il Topic che è la rappresentazione nel mondo delle macchine degli altrimenti ineffabili soggetti. A ogni soggetto (qualunque elemento del discorso) corrisponde un solo topic per il principio della collocation objective (non ambiguità di un soggetto all'interno di un sistema).

Un topic può essere identificato da un identificativo di soggetto, per garantire la possibilità di parlarsi fra diversi sistemi esistono identificativi di soggetto pubblicati che, se usati da entrambe le parti, assicurano che i topic di entrambi i sistemi rappresentano lo stesso soggetto. Parte del lavoro di ricerca, come vedremo, è costituito dalla definizione di PSI per permettere la condivisione di descrizioni archivistiche.

Un topic può avere diverse caratteristiche: il tipo di topic (tesi, persona etc), il nome (che a sua volta può essere caratterizzato tipologicamente) può essere associato con altri topic (ogni relazione rappresentata da un'associazione può essere caratterizzata tipologicamente) e infine può avere un'occorrenza che è, semplificando, un campo descrittivo interno anche questo con specifico topic (ad esempio citazione, abstract, descrizione etc). Ognuna di questa caratteristica, nel gergo asserzione, è valida in uno specifico ambito che può essere illimitato o limitato (ad esempio cronologicamente, linguisticamente, geograficamente etc.)

Descrizione archivistiche e web semantico

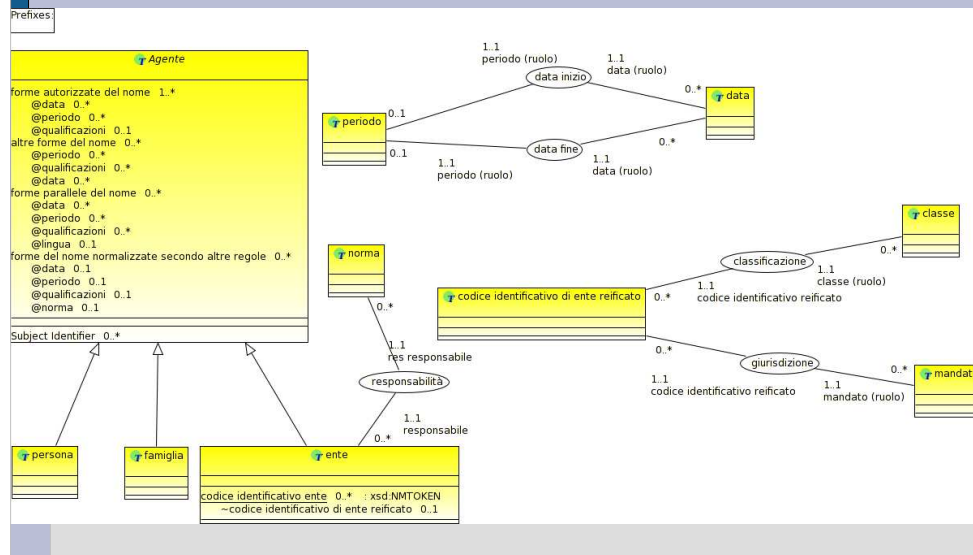
- Tradurre gli standard archivistici nel modello delle Topic Maps
- Fornire indicazione per ogni singolo elemento previsto dagli standard
- Mantenere granularità dell'informazione prevista da standard e dai software esistenti
- Fornire schemi di validazione TMCL
- Usare una notazione grafica per modellare la traduzione degli standard archivistici in Topic Maps

L'assunto di base che ha guidato il nocciolo centrale del lavoro di ricerca è che se riusciamo a esprimere ogni elemento previsto dagli standard di descrizione archivistica attraverso gli elementi della tecnologia scelta, non perdendo la ricchezza di informazioni in questo passaggio. Allora abbiamo dimostrato implicitamente che sia possibile esprimere descrizioni archivistiche nel web semantico. Per mostrare i passaggi logici effettuati si sono utilizzati alcuni strumenti quali un linguaggio di modellazione grafica, attualmente in discussione come standard iso 13250-7. Per gli esempi invece nella comunità è in uso un'altra notazione grafica, gtmalpha, specifica per il mondo delle Topic Maps. Per colmare la distanza fra queste due notazioni grafiche si è sviluppato nel corso della tesi un nuovo linguaggio di modellazione (che ho chiamato gtmalpha+) che potrà poi essere proposto nella discussione internazionale che porterà alla definizione del relativo standard ISO.

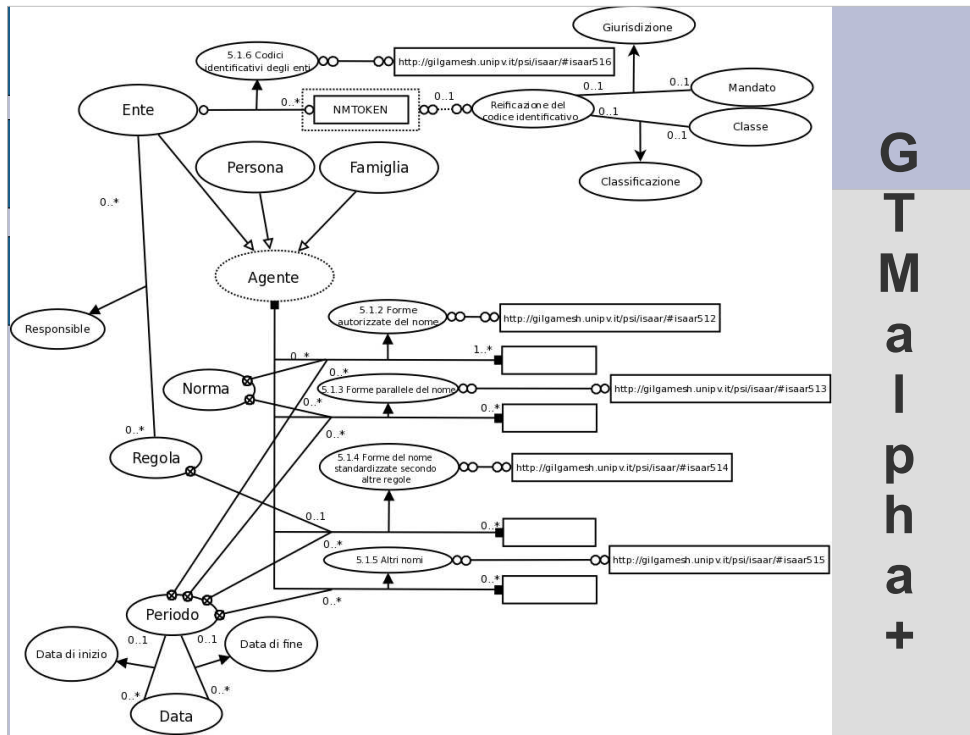
Per codificare le regole suggerite per tradurre una descrizione archivistica in Topic Maps si è utilizzato il TMCL, un linguaggio vincolato che permette di creare schemi di validazione (una sorta di ciò che sono DTD e XML Schema per XML... pur considerando la differenza tra validazione sintattica e semantica).

Con il TMCL noi possiamo ad esempio dire che un topic di tipo agente o, meglio, un topic il cui tipo è sottoclasse di agente deve per forza avere almeno un topic name di tipo “forma autorizzata del nome”, può avere uno o più topic name di tipo forma parallela del nome e questi possono essere qualificati con uno scope linguistico etc

Esempio di mappatura (I) Onotoa



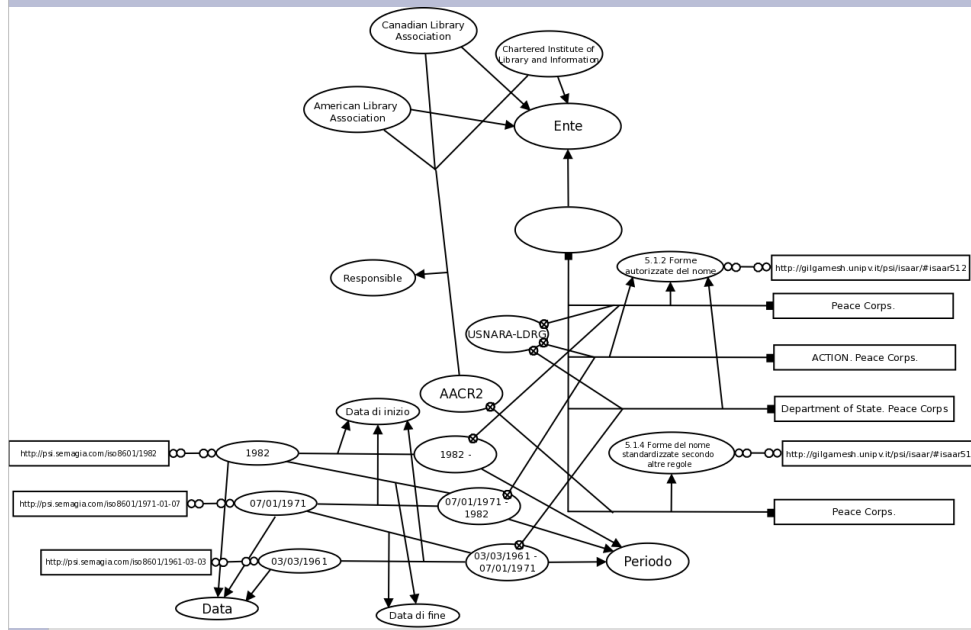
Esempio della prima area di ISAAR (International Standard for Archival Authority Record), utilizzando il software Onotoa che utilizza la sintassi attualmente in discussione come standard ISO 13250-7



**G
T
M
a
l
p
h
a
+**

Stesso esempio nella notazione grafica GTMalphaplus, notazione definita all'interno del lavoro di ricerca

Esempio - GTMalpha



Ci spostiamo dal piano dell'ontologia a quello delle istanze. Ecco un esempio concreto basato sulla notazione GTMalpha e, come si può vedere, è sicuramente maggiormente collegato alla notazione gtmalpha+ da me creata rispetto alle scelte grafiche simil UML di onotoa

TMCL (Estratto)

```
isaar-tm:agente isa tmcl:topic-type;
- 'Agente';
is-abstract();
has-name(isaar-tm:forma-autorizzata, 1, *).
isaar-tm:forma-autorizzata isa tmdm:topic-name-type;
- 'Forma autorizzata del nome
<http://gilgamesh.unipv.it/psi/isaar/#isaar512>'.
isaar-tm:persona isa tmcl:topic-type;
- 'Persona';
ako isaar-tm:agente.
isaar-tm:famiglia isa tmcl:topic-type;
- 'Famiglia';
ako isaar-tm:agente.
isaar-tm:ente isa tmcl:topic-type;
- 'Ente';
ako isaar-tm:agente;
has-occurrence(isaar-tm:codice-identificativo-ente, 0, *).
isaar-tm:codice-identificativo-ente isa tmcl:occurrence-type;
- 'Codice identificativo dell\'ente'
<http://gilgamesh.unipv.it/psi/isaar/#isaar512>
has-datatype(xsd:NMTOKEN);
is -unique().
```

Ecco un estratto dello schema di validazione TMCL.

In questo caso si definisce il tipo di topic Agente, che è un abstract, cioè non avrà istanze direttamente collegate, ma solo attraverso i suoi sottotipi persona, ente, famiglia.

Le istanze di questi sottotipi dovranno avere almeno un topic name con topic name type “Forma autorizzata del nome” identificato da uno specifico PSI.

Le istanze di tipo ente potranno avere uno o più codici identificativi dell'ente, di tipo NameToken e che dovranno essere univoci

Risultati - Studi

- Regole per creare descrizioni archivistiche conformi allo standard basate su Topic Maps
- Valutazione ed esempi di applicazioni pratiche a inventari e descrizioni della flessibilità garantita da un simile approccio
- Studio su come creare interfacce flessibili e sistemi informativi modulari sulla base di questa tecnologia
- Utilizzo delle indicazioni e mappature prodotte per ipotizzare un interscambio dei dati basato su XTM (XML Topic Maps) con particolare attenzione al Portale Archivistico Nazionale del Sistema Archivistico Nazionale

I risultati ottenuti nel corso del lavoro di ricerca possono essere distinti fra studi e risultati pratici.

Dal punto di vista dello studio si sono definite le regole per permettere di descrivere un archivio direttamente nel web semantico.

Se abbiamo visto e determinato la possibilità tecnica, non abbiamo ancora sottolineato i vantaggi di un simile approccio.

La lista dovrebbe essere necessariamente lunga ma basti ricordare che ogni asserzione nel mondo delle Topic Maps è valida in un determinato ambito.

Ciò ci permette facilmente di creare descrizioni multilingua (ambito o scope linguistico), elaborare ricostruzioni sulla carta (come nel caso dell'archivio Testori) definendo un ambito di validità "es Inventario Gallerani" delle associazioni gerarchiche, proporre descrizioni valide per uno specifico pubblico (es ragazzi, ambito di target) etc.

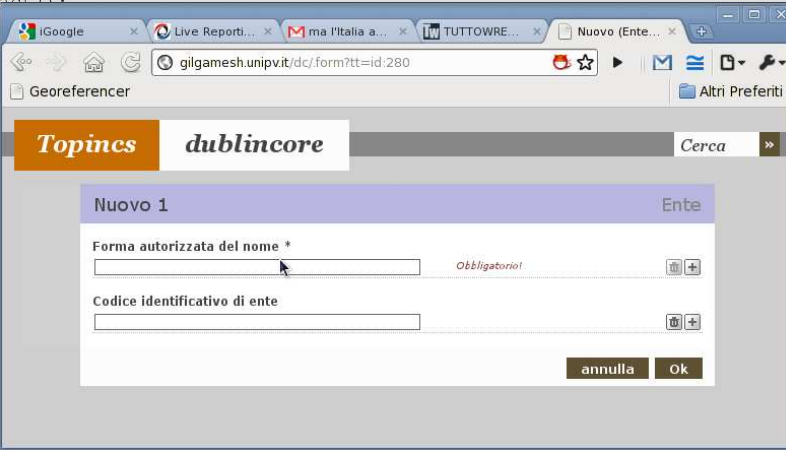
In seconda istanza si è analizzato come trasferire questa flessibilità dal risultato (le descrizioni) alla fonte (il software per realizzarle).

Si dimostra come in un software che implementi il linguaggio vincolato TMCL sia possibile creare facilmente interfacce di inserimento e modificarle senza toccare una stringa di codice, ma solo cambiando le regole definite nel linguaggio vincolato.

Infine ho sperimentato l'utilizzo della soluzione proposta nell'interscambio di dati fra sistemi archivistici con una proposta di soluzione per il Sistema Archivistico Nazionale

Esempio interfacce flessibili con Topincs

```
isaar-tm:agente isa tmcl:topic-type;
- 'Agente';
is-abstract();
has-nam
isaar-tm:fo
- 'Form
<http:/
isaar-tm:pe
- 'Pers
ako isa
isaar-tm:fa
- 'Fami
ako isa
isaar-tm:er
- 'Ente
ako isa
has-occ
*).
isaar-tm:cc
type;
- 'Codi
<http:/
has-dat
is -unique().
```



Ad esempio l'estratto di codice TMCL evidenziato in precedenza, dato in pasto a software capaci di leggere simili schemi (allo stato attuale solo Topincs di Robert Cerny) produce automaticamente un'interfaccia di inserimento.

Volessi cambiare la cardinalità e obbligatorietà degli elementi o se volessi inserire nuovi campi, basterebbe modificare lo schema tmcl senza toccare il codice del programma.

Risultati - Pratici

- Schemi TMCL per validare una descrizione archivistica secondo le regole proposte
- Produzione, mantenimento e diffusione di PSI
- Definizione di una notazione grafica (GTAlpha+) per la modellazione in Topic Maps
- TM4BAD una serie di fogli di stili per convertire da EAD, EAC in Topic Maps
- Un prototipo di strumenti di ricerca di nuova generazione per accedere alla documentazione dell'archivio Testori

Tra i risultati pratici di questo lavoro di ricerca è necessario sottolineare ancora una volta gli schemi tmcl per validare descrizioni archivistiche di cui si è visto fin qui un estratto. Contestualmente alla definizione della mappatura tra le Topic Maps e gli standard archivistici sono stati definiti e diffusi dei PSI affinché le stesse soluzioni potessero essere utilizzate altrove e iniziare a parlare, per ciò che riguarda le descrizioni archivistiche e il web semantico, la stessa lingua. Come si è detto e visto, per favorire la comprensione degli schemi è stato sviluppato una nuova proposta di notazione grafica per descrivere la modellazione in topic maps. Proposta che potrà concorrere nella discussione per la definizione dello standard iso. Sulla base delle mappature proposte sono stati predisposti un set di fogli di stile xsl-t per trasformare da standard di struttura dati archivistici come EAD e EAC-CPF a XTM (xml topic maps). A compendio e a dimostrazione delle potenzialità è stato approntato un prototipo di sistema informativo, strumento di ricerca per il fondo testori.

Esempio Testori (I)

Titolo attribuito (1) <ul style="list-style-type: none">Tre Lai	Hierarchy (Struttura archivistica (inventario Bilotto - Gallerani)) <ul style="list-style-type: none">Fondo Testori<ul style="list-style-type: none">Quaderni<ul style="list-style-type: none">Tre Lai<ul style="list-style-type: none">...
Titolo originale (1) <ul style="list-style-type: none">1 Cleopatras1^o Cleopatras2 Erodias2^o Erodias3^o Mater strangosciásMater strangosciásTre LaiTre lai, 1 Cleopatras, 2 Erodias, 3 Jesu matras o Mater strangosciásTre lai: Branciatrilogia seconda 3 Tre compianti, 1 Jesu [...?], 2 Cleopatras, 3 King liro [cancellato e sostituito con] erodias	Internal Occurrences (1) <ul style="list-style-type: none">3.1.1 Segnatura/e o codice/Identificativo/1<ul style="list-style-type: none">UA46
Associations (6) <ul style="list-style-type: none">3.1.3 Date(s)<ul style="list-style-type: none">1991-12 - 1992-02-18 (more)Struttura archivistica (Inventario Bilotto - Gallerani)<ul style="list-style-type: none">Quaderno 82 Foglio di guardia recto e verso - Quarta di copertina rectoQuaderno 84 Coperta recto e verso - Foglio di guardia recto e verso - Pagine 1-137, 139, 141, 143 - Foglio di guardia recto e verso - Quarta di copertina recto e versoQuaderno 85 Coperta recto e verso - Foglio di guardia recto e verso - Pagine 1-84 - Foglio di guardia recto e verso - Quarta di copertina recto e versoQuaderno 86 Coperta recto e verso - Foglio di guardia recto e verso - Pagine 1-157, 160 - Foglio di guardia recto e verso - Quarta di copertina recto e versoStruttura archivistica (Inventario Bilotto - Gallerani)<ul style="list-style-type: none">Quaderni	<p>Rimanda allo standard ISAD(g)</p> <p>Collegamento a note sulla datazione espresse in Topic Maps come una reificazione</p> <p>Questa unità logica si estende per più quaderni</p>

Qui l'ontologia sottostante il prototipo. Visualizzata attraverso il software opensource ontopia

Titoli originali (type="182") nel caso dei Tre Iain sono stati individuati nove ricorrenze nei diversi quaderni. Qualora si volesse sarebbe possibile indicare le informazioni su dove si sia riscontrato un determinato titolo e possibile utilizzare una replicazione del topic name

```

<topic id="t218">
  <instanceOf>
    <topicRef href="#t180"/>
  </instanceOf>
  <name>
    <type>
      <topicRef href="#t181"/>
    </type>
    <value>{CDATA[Tre Iain]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[Tre Iain]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[Tre Iain; Branciatrilogia seconda; 3 Tre compunti, 1 Ossu[...], 2 Cleopatra, 3 King Iiro; [cancellato e sostituito con] Erodias]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[Tre Iain, 1 Cleopatra, 2 Erodias, 3 Jesu matris o Mater strangocias]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[1 Cleopatra]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[2 Erodias]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[Mater strangocias]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[1* Cleopatra]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[2* Erodias]}</value>
  </name>
  <name>
    <type>
      <topicRef href="#t182"/>
    </type>
    <value>{CDATA[3* Mater strangocias]}</value>
  </name>
  <occurrence>
    <type>
      <topicRef href="#t183"/>
    </type>
    <resourceData>{CDATA[UM4]}</resourceData>
  </occurrence>
</topic>
          
```

(XML)

E
s
e
m
p
l
i
o
t
e
s
t
o
r
i

Un estratto del sorgente XTM che è sotto l'esempio mostrato

Esempio Testori (III)

Naviga per

Opera
Articolo
Quaderno
Persona
Data

Filtra per

Lingua

- Inglese
- Italiano
- Francese

Target

- Tutti
- Ragazzi

Relazioni

- FRBR
- Gerarchia strutturale
- Inventario Vassallo
- Inventario Vassallo (2009-07-28)
- Inventario Bilotto - Gallerani

Cerca per

La documentazione relativa ai *Tre Lai* comprende l'arco cronologico tra il dicembre 1991 e il 18 febbraio 1992 .
All'interno della documentazione l'opera è denominata dall'autore stesso anche:

- Tre lai; Branciatrilogia seconda 3 Tre compianti, 1 Jesu [..?], 2 Cleopatràs, 3 King liro [cancellato e sostituito con] Erodiàs"
- "Tre lai, 1 Cleopatràs, 2 Erodiàs, 3 Jesumatràs o Mater strangosciàs"
more..

Documentation about *Tre Lai* covers the period between december 1991 and 18 february 1992 .

Inside the documentation, the author also called the work:

- Tre lai; Branciatrilogia seconda 3 Tre compianti, 1 Jesu [..?], 2 Cleopatràs, 3 King liro [deleted and substituted with] Erodiàs"
- "Tre lai, 1 Cleopatràs, 2 Erodiàs, 3 Jesumatràs o Mater strangosciàs"
more..

Inventario Bilotto - Gallerani

- Fondo Testori
 - Quaderni
 - **Tre Lai**

FRBR

- Tre Lai
 - Tre Lai (Manoscritto - testo - italiano)
 - **Tre Lai**
 - Tre Lai (Stampa - testo - italiano)
more..

16

Ovviamente all'utente il lavoro sotto la superficie interessa ben poco, così come risulta difficoltosa l'interfaccia di default di ontopia.

Spesso i siti basati su topic maps sono accusati (giustamente) di essere rigidi e troppo legati alle caratteristiche del topic...

In realtà come si può vedere i risultati possono essere ben più armoniosi dell'interfaccia di default dell'omnigator di ontopia

Corollari alla ricerca

- FRBR e Archivi
 - CIDOC-CRM
 - Utilizzo in archivi storici
 - Prospettiva per gli archivi correnti
- FRAD e ISAAR
 - Mappatura fra gli standard
 - il name project
- Archivi 2.0

Nel corso del lavoro di ricerca si sono toccati anche altri aspetti corollari del nocciolo dello studio.

A ognuno di questi è dedicato uno specifico capitolo, mi preme solo riassumerli brevemente.

Si sono valutate le interazioni fra frbr e il mondo degli archivi. Ciò è di particolare rilevanza per gli archivi letterari dove potrebbe essere possibile legare la documentazione alle opere edite strutturate nella gerarchia work-expression-manifestation-item... questa gerarchia, nel nostro modello, sarebbe rappresentata ancora una volta da associazioni gerarchiche valide in uno specifico ambito. In realtà possibili fruttuosi sviluppi potrebbero esserci nell'applicazione del modello frbr al mondo degli archivi correnti, basti pensare a tutta la documentazione bilingue e in più formati prodotti dagli archivi dell'unione europea.

Parallelamente si sono analizzati i rapporti fra frad e isaar, standard con scopi “simili” seppur di mondi diversi. Una simile riflessione è stata condotta nel 2008, a ricerca iniziata, anche dal name project.

Nel caso del mio lavoro di ricerca il lavoro è stato diametralmente opposto e complementare a quello svolto dal name project: una mappatura tra i due standard, per verificare quali elementi di frad possano essere aggiunti agli elementi isaar in particolare nel caso di descrizioni di soggetti produttori che siano autori.

Infine è stata svolta una ricognizione su un filone di ricerca assolutamente attuale e in crescita che va sotto l'etichetta di Archive 2.0. Aldilà del marchio 2.0 inflazionato si è cercato di sottolineare la necessità di flessibilità e di dati strutturati a cui sembra ben rispondere il lavoro di ricerca.