



Workshop
“Progettazione di Ontologie e
integrazione di dati geografici aperti”

**Sistema Archivistico Nazionale
(SAN)**

Costantino Landino
(Responsabile Informatico dell’Istituto Centrale per gli Archivi)
27 Febbraio 2014

istruzioni
per
il
RI
USO



Roma
27-02-2014

2^a

Conferenza
Open
Geo
Data
Italia

Abstract

Nel corso dell'intervento verrà illustrato il percorso per la creazione dei Linked Open Data del Sistema Archivistico Nazionale attraverso le varie fasi della progettazione dai dati, del loro controllo e normalizzazione fino all'utilizzo come repository da cui attingere informazioni.

Verrà illustrato il processo di creazione dell'ontologia, dallo schema concettuale dei dati alla definizione in linguaggio OWL dell'ontologia stessa ed alla definizione in linguaggio SKOS dei vocabolari e dei tesauri utilizzati.

Successivamente verranno presentati attraverso l'uso del tool OpenRefine sia i passaggi di normalizzazione degli elementi geografici dell'ontologia sia i passaggi di normalizzazione per il collegamento ad ontologie esterne.

Saranno presentati degli esempi di record SAN descritti in formato RDF ed i vocabolari definiti in SKOS.

In questi passaggi verranno illustrate le problematiche di allineamento e di gestione dei dati per fornire open data di buona qualità in ambito beni culturali.

Agenda

- 1** Sistema Archivistico Nazionale
- 2** Progetto Linked Open Data
- 3** Struttura dei dati CAT-SAN
- 4** Disegno di Ontologie
- 5** Ontologia SAN
- 6** SKOS e Thesaurus SAN
- 7** Openrefine
- 8** Conclusioni

Agenda

1 Sistema Archivistico Nazionale

2 Progetto Linked Open Data

3 Struttura dei dati CAT-SAN

4 Disegno di Ontologie

5 Ontologia SAN

6 SKOS e Thesaurus SAN

7 Openrefine

8 Conclusioni

Il Sistema Archivistico Nazionale: principali caratteristiche e funzioni

- costituisce un **punto di accesso primario** al patrimonio archivistico italiano, rendendo disponibili archivi di natura eterogenea, finora consultabili separatamente.
- intende far conoscere a un pubblico non di soli specialisti **quali risorse archivistiche esistano** a livello nazionale, **dove siano dislocate** e **come vi si acceda**;
- consente l'**interconnessione con altri portali / archivi** e il **recupero di risorse** dal web tramite web crawler, l'**harvesting dei sistemi aderenti al SAN**, l'**importazione ed esportazione di dati** relativi alle risorse archivistiche e digitali, non solo rispetto ai sistemi dell'Amministrazione archivistica, ma anche a quelli curati da Enti statali e non e da organizzazioni pubbliche e private, come Regioni, Province, Comuni, Università, Enti ecclesiastici, Istituzioni culturali, Fondazioni;
- coopera con il **Portale europeo APex**;
- é **interoperabile con CulturalItalia** - il Portale della Cultura Italiana di MIBAC che rende accessibili risorse provenienti da archivi, musei, biblioteche, soggetti pubblici e privati – e quindi con il **Portale Europeo Europeana**.

Il Sistema Archivistico Nazionale gestisce i seguenti contenuti:

- **catalogo delle risorse archivistiche**: descrizioni dei soggetti conservatori, dei soggetti produttori, dei complessi archivistici e degli strumenti di ricerca, trasmesse al SAN attraverso standard, tracciati di scambio e protocolli siglati con i responsabili dei sistemi aderenti.
- **digital library**: risorse digitali riferite a documenti di diversa natura e tipologia (immagini, audio, video), descritte e rintracciabili sulla base di un tracciato di metadati condiviso (profilo METS SAN). Ciascun documento riprodotto è contestualizzato per la sua appartenenza a un archivio (conservatore e complesso documentario) e a un progetto di digitalizzazione (scheda progetto di digitalizzazione). METS SAN consente anche la gestione di oggetti digitali non archivisticamente contestualizzati;
- **risorse bibliografiche** rilevanti per il dominio archivistico, con collegamento al Sistema Bibliotecario Nazionale;
- **percorsi, notizie, informazioni, testi e contenuti** arricchiti elaborati dalla redazione e da istituti universitari e di ricerca, che hanno aderito al SAN attraverso specifici protocolli di intesa;
- **portali tematici** a diverso titolo integrati nel SAN

Sistemi cooperanti

Regioni

Emilia-Romagna
Lombardia
Piemonte
Toscana
Umbria
e Provincia Autonoma di Trento

Sistemi Aderenti

Archivi di Stato (SIAS)
Soprintendenze Archivistiche (SIUSA)
SIASBO, SIASFI, SIASMI, SIASNA,
SIASVE, SIASPG
Guida Generale Archivi Stato italiani
IBC ARCHIVI
Lombardia Beni Culturali (LBC)
INSMLI
CASTORE Imago Tusciae

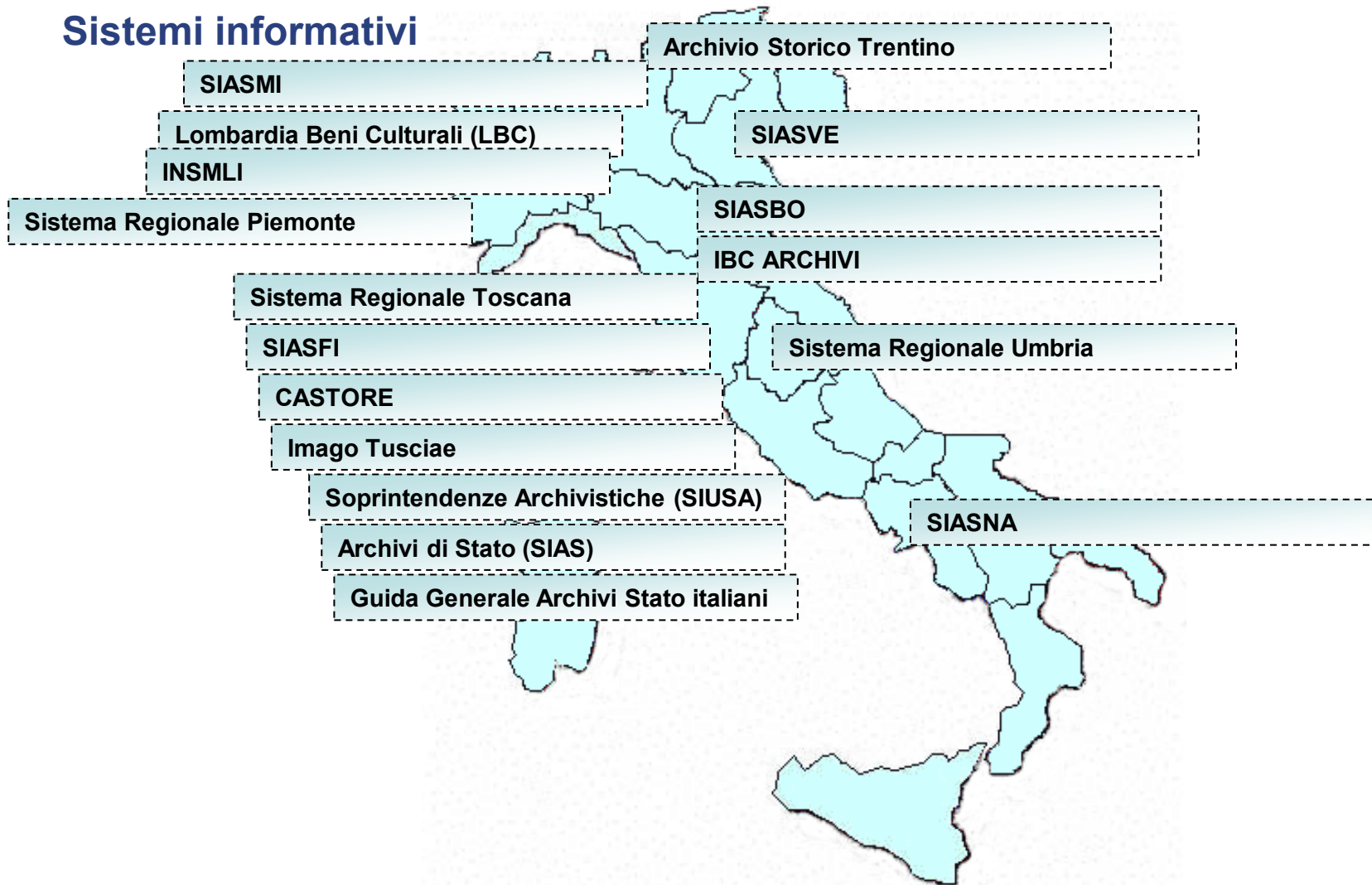


CEI-AR

Portali Tematici

Impresa	Territori
Antenati	Per non dimenticare
Musica	Architetti
Moda	“Verdi on line”
Carte da legare	Giovanni Pascoli
Archivio Storico Multimediale del Mediterraneo	
Portale della Storia degli Italiani	

Sistemi informativi



Dimensioni del sistema al 21 Febbraio 2014:

Tipologia di contenuti	Publici	Backoffice
Soggetti conservatori	8.750	9.230
Soggetti produttori	62.816	68.215
Complessi archivistici	85.471	94.012
Strumenti di corredo	17.606	17.606
Oggetti digitali (immagini, video, audio, documenti, ecc.) e metadati	362.341	364.207
Totale	536.980	553.270

Portali tematici

Il Sistema Archivistico Nazionale integra, alcuni portali tematici dedicati anche ad un pubblico di non specialisti.

- **Archivi di impresa:** fonti archivistiche e bibliografiche dal 1861 a oggi, provenienti da archivi di imprese. <http://www.impreses.an.beniculturali.it>
- **Rete degli archivi “per non dimenticare”:** fonti archivistiche e bibliografiche sul terrorismo, sulla violenza politica e sulla criminalità organizzata dal 1946 a oggi in Italia. <http://www.memoria.an.beniculturali.it>
- **Archivi della Moda del '900:** fonti archivistiche, bibliografiche, iconografiche, audiovisive, conservate presso diversi soggetti afferenti al mondo della moda. <http://moda.an.beniculturali.it>
- **Archivi di Musica:** fonti archivistiche, approfondimenti biografici, percorsi tematici, cronologia sulla musica del Novecento e sulla musica popolare. <http://www.musica.an.beniculturali.it>
- **Archivi di Architetti:** fonti archivistiche, approfondimenti biografici, progetti e immagini sugli archivi di architetti e ingegneri del Novecento. <http://www.architetti.an.beniculturali.it>
- **Antenati. Gli Archivi per la ricerca anagrafica:** fonti archivistiche per ricerche anagrafiche e genealogiche. <http://www.antenati.an.beniculturali.it>


Portali tematici

Il Sistema Archivistico Nazionale integra, alcuni portali tematici dedicati anche ad un pubblico di non specialisti.


- **Territori** : fonti archivistiche catastali e cartografiche provenienti da 5 archivi di stato . <http://www.territori.san.beniculturali.it>
- **Archivio Storico Multimediale del Mediterraneo**, che contiene documentazione sul Mediterraneo conservati negli archivi storici Italiani. <http://www.archividelmediterraneo.org/>
- **Portale "Verdi on line"**: luogo di elezione per l'aggregazione delle principali fonti verdiane. <http://www.verdi.san.beniculturali.it>
- **Carte da legare**: progetto realizzato per salvaguardare il patrimonio archivistico degli ex ospedali psichiatrici. <http://cartedalegare.filosofia.sns.it>
- **Giovanni Pascoli nello specchio delle sue carte**: l'archivio Pascoli ha creato un nuovo e affidabile patrimonio di informazioni sul poeta. <http://pascoli.archivi.beniculturali.it>
- **Portale della Storia degli Italiani**: un portale per la ricerca di documentazione tratte da una pluralità di fonti documentarie demografiche conservate negli archivi storici. <http://www.poster.beniculturali.it>

Portali Tematici


Ministero per i beni e le attività culturali • Direzione generale per gli archivi




ARCHIVI DELLA MODA DEL NOVECENTO




Il progetto



Partner




La moda ieri e oggi




I protagonisti




Percorsi



Strumenti di ricerca




Multimedia



News

© MINISTERO PER I BENI E LE ATTIVITA' CULTURALI - DIREZIONE GENERALE PER GLI ARCHIVI - SISTEMA ARCHIVISTICO NAZIONALE

Ministero per i beni e le attività culturali • Direzione generale per gli archivi

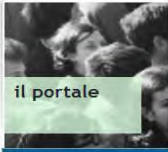


home | mappa sito | stampa | news


Rete degli archivi **per non dimenticare**

27 MARZO 1935 12 FEBBRAIO 1980
 23 DICEMBRE 1984 2 AGOSTO 1980
 12 DICEMBRE 1969 10 LUGLIO 1929 19 MARZO 20
 1 MAGGIO 1947 9 MAGGIO 1978

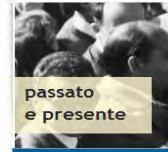
cerca nel portale




il portale




la rete




passato e presente




muro della memoria




per approfondire



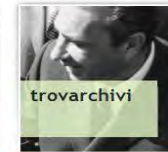
per non dimenticare




didattica



galleria multimediale



trovarchivi



biblioteca

crediti | note legali

Portali Tematici



Antenati
Gli Archivi per la Ricerca Anagrafica

Ministero dei beni e delle attività culturali e del turismo | Direzione Generale per gli Archivi

Home Link News Help Partecipa al progetto!

Il Portale | Il territorio e le fonti | Sfoggia i registri | Trova i nomi

Il Portale Antenati - promosso e curato dalla Direzione Generale per gli Archivi - nasce dall'esigenza di rendere disponibile l'enorme patrimonio documentario esistente negli Archivi di Stato per condurre ricerche anagrafiche e genealogiche, finalizzate alla ricostruzione della storia di famiglie e di persone, ma anche alla storia sociale in senso lato. Attraverso il Portale è possibile sfogliare a video milioni e milioni di immagini di registri di anagrafe e di stato civile, trovare nomi di persona presenti negli atti, ottenere informazioni sulle fonti.

Il Portale

Il territorio e le fonti

Sfoggia i registri

Trova i nomi

Scuola materna, attività ludiche (AS Mantova, Istituto pro lattanti e slottati E.Soncini)

Ultime news
22 agosto 2013 Caricate le immagini relative all'Archivio di Stato di Campobasso

Sono terminate le operazioni di caricamento dei registri dell'Archivio di Stato di Campobasso; sul sito sono ora disponibili le immagini relative ai registri provenienti dai fondi Napoleonico (150.695 immagini dal 1809 al 1815), della Restaurazione (1.405.245 immagini dal 1816 al 1860) ed Italiano (689.850 dal 1861 al 1913).

Crediti | Note legali



TERRITORI
il portale italiano dei catasti e della cartografia storica

Home | Il Portale | Accedi alle mappe | Partner | Protagonisti | In evidenza | Biblioteca

ACCEDI ALLE MAPPE

TROVARCHIVI | SOGGETTI PRODUTTORI | FONTI | TIMELINE | ACCESSO GEOGRAFICO | RICERCA

IL PORTALE | PARTNER | PROTAGONISTI | IN EVIDENZA | BIBLIOTECA

GALLERY

Fonti | Accesso Geografico | Timeline

Crediti | Note Legali | Hyperborea s.r.l.

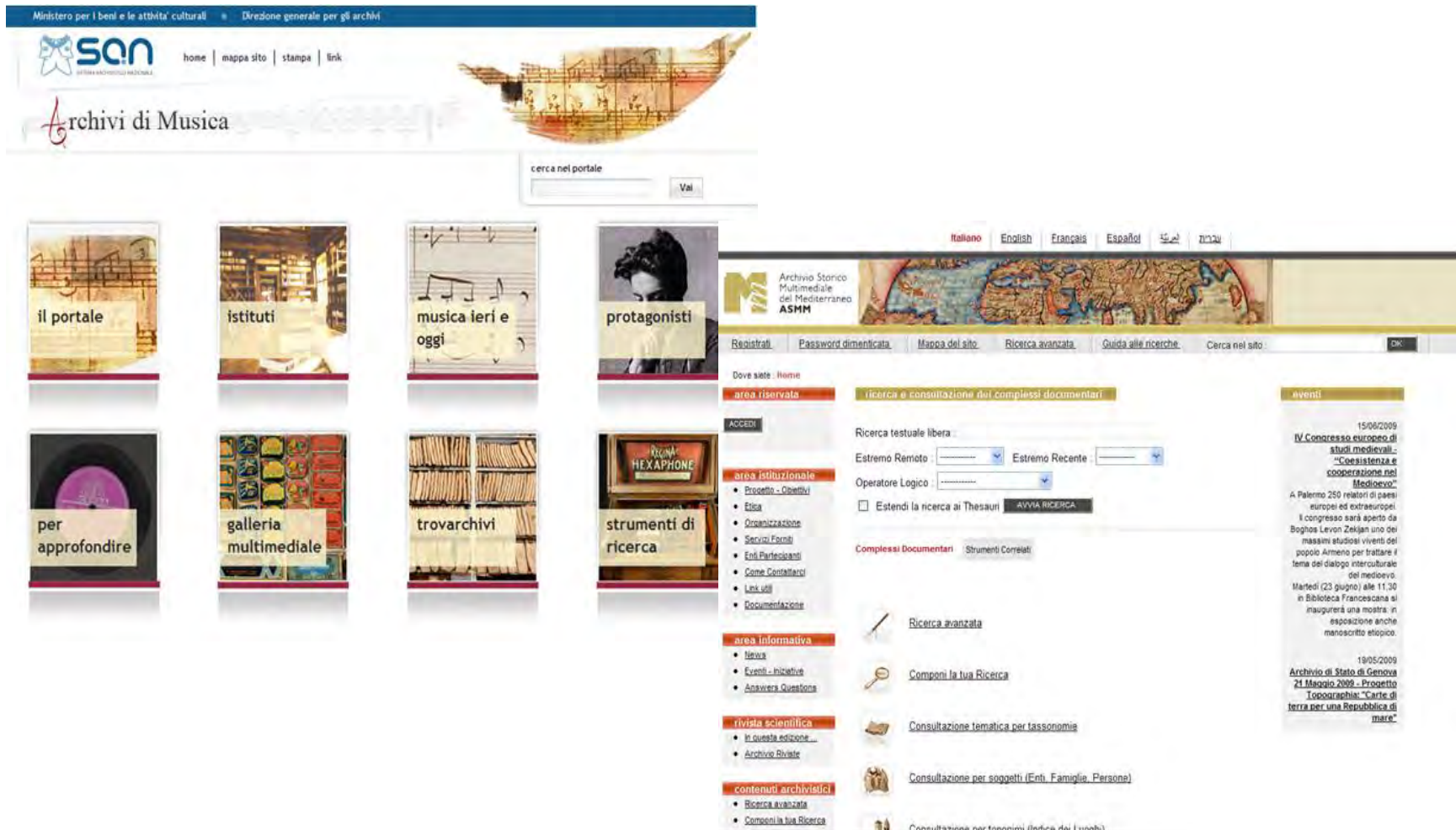
Questo sito richiede Internet Explorer 8+, Mozilla Firefox 4+, Google Chrome 9+, Safari

©2011-2012 Portale dei Territori

Portali Tematici

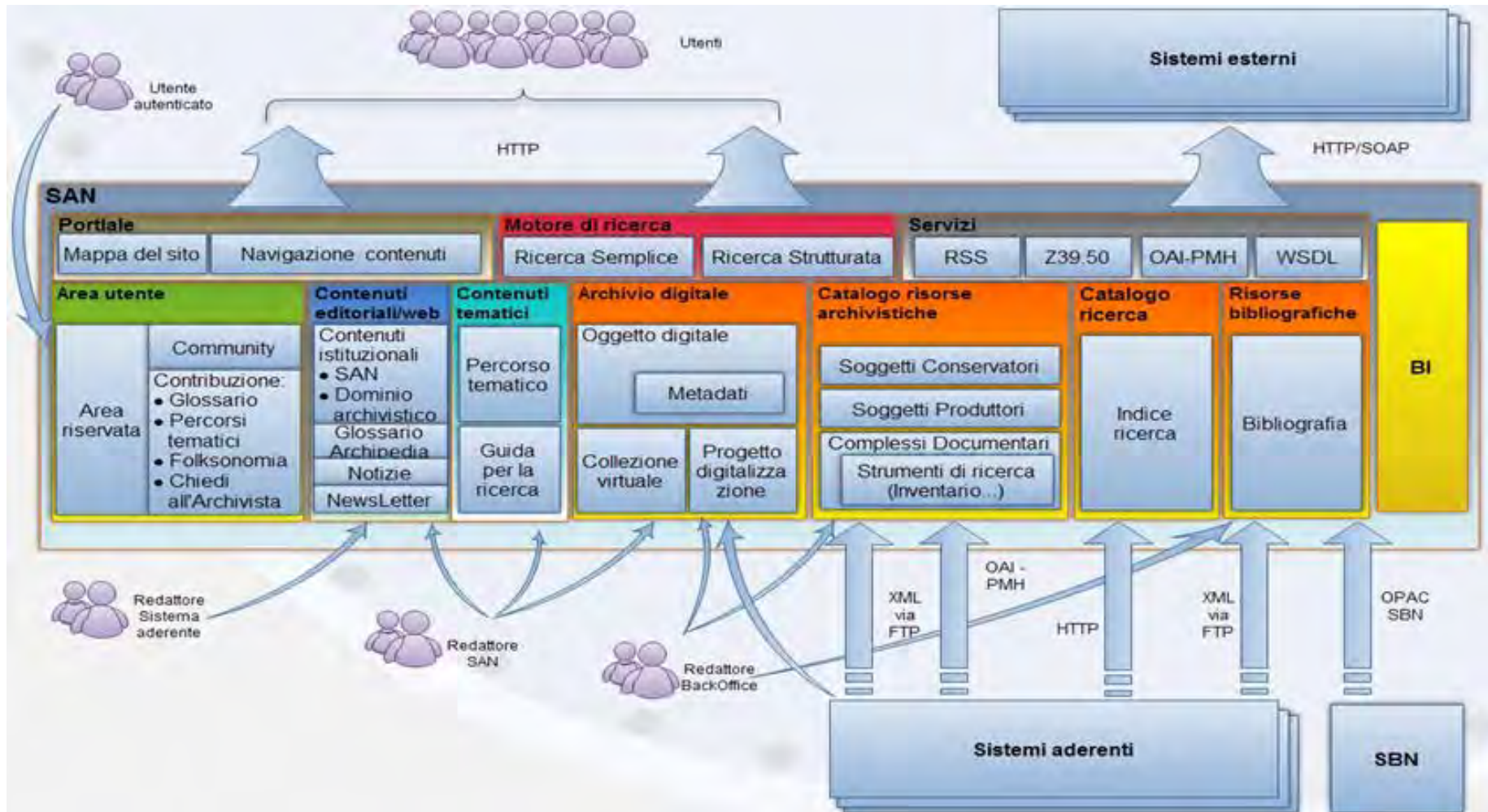


Portali Tematici

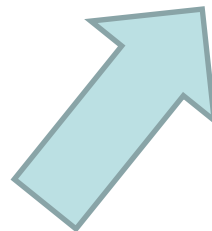


The screenshot displays the 'Archivi di Musica' portal. At the top, it features the SAN logo and navigation links: 'home', 'mappa sito', 'stampa', and 'link'. A search bar is positioned below the header with a 'Vai' button. The main content area is organized into a grid of eight thematic tiles: 'il portale', 'istituti', 'musica ieri e oggi', 'protagonisti', 'per approfondire', 'galleria multimediale', 'trovarchivi', and 'strumenti di ricerca'. To the right, a detailed search interface is visible, including a language selector (Italiano, English, Français, Español, العربية, עברית), a search bar, and a 'Vai' button. Below the search bar, there are sections for 'area riservata', 'area istituzionale' (with links like 'Progetto - Obiettivi', 'Etica', 'Organizzazione', 'Servizi/Fondi', 'Enti Partecipanti', 'Come Costituirsi', 'Link Utili', 'Documentazione'), 'area informativa' (with links like 'News', 'Eventi - iniziative', 'Answera Questioni'), 'rivista scientifica' (with links like 'In questa edizione...', 'Archivio Riviste'), and 'contenuti archivistici' (with links like 'Ricerca avanzata', 'Componenti la tua Ricerca'). The search interface also includes a 'Ricerca testuale libera' section with dropdown menus for 'Estremo Remoto' and 'Estremo Recente', an 'Operatore Logico' dropdown, and an 'Estendi la ricerca ai Thesauri' checkbox. A 'Ricerca avanzata' section is also present with a 'Vai' button. On the far right, there is an 'aventi' section with a date '15/06/2009' and a link to 'IV Congresso europeo di studi medievali: "Coesistenza e cooperazione nel Medioevo"'. Below this, there is a notice about a congress in Palermo on June 23rd at 11:30 in the Biblioteca Franciscana, featuring a manuscript exhibition.

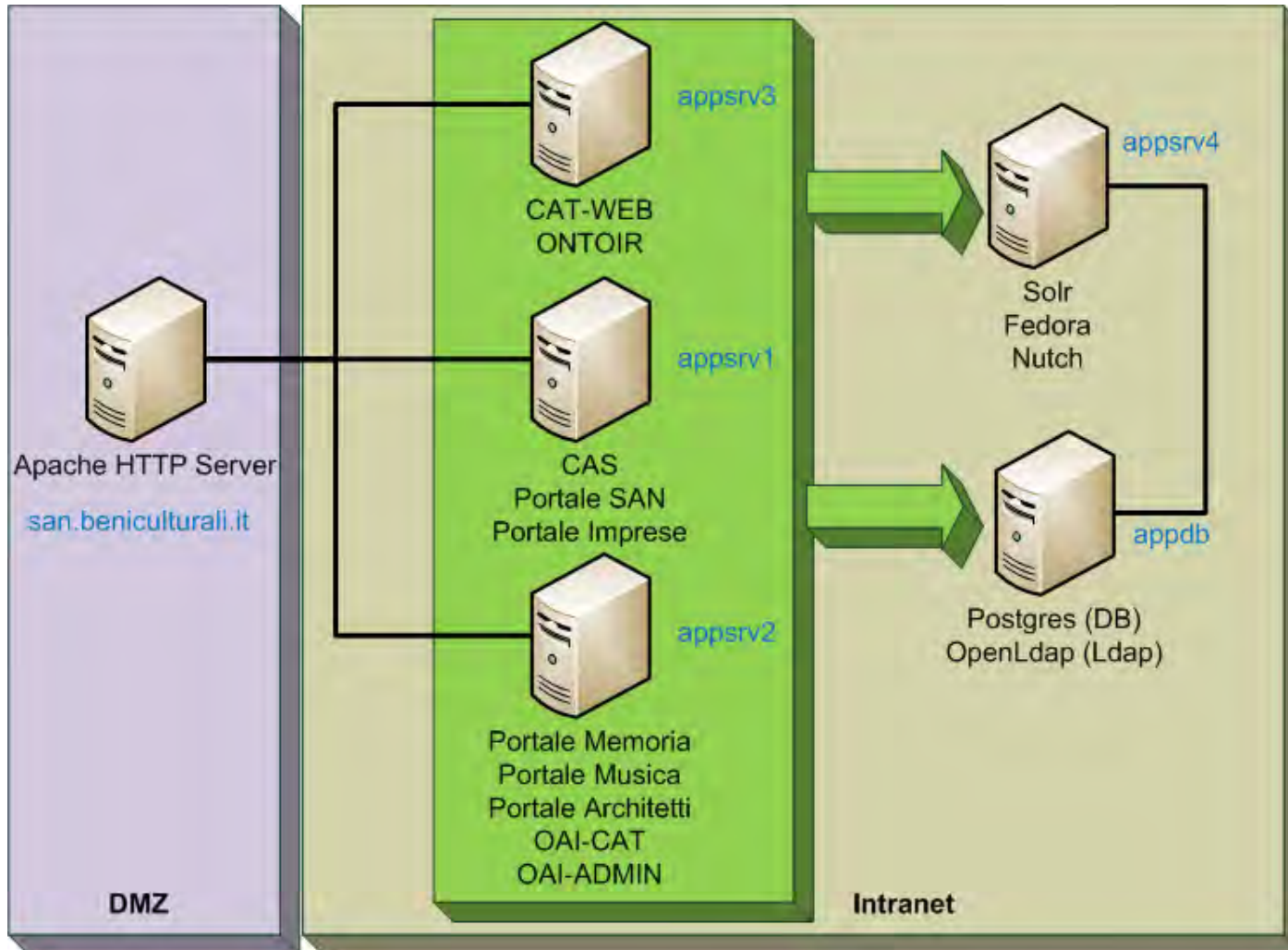
Il modello concettuale del SAN



Il modello concettuale del SAN



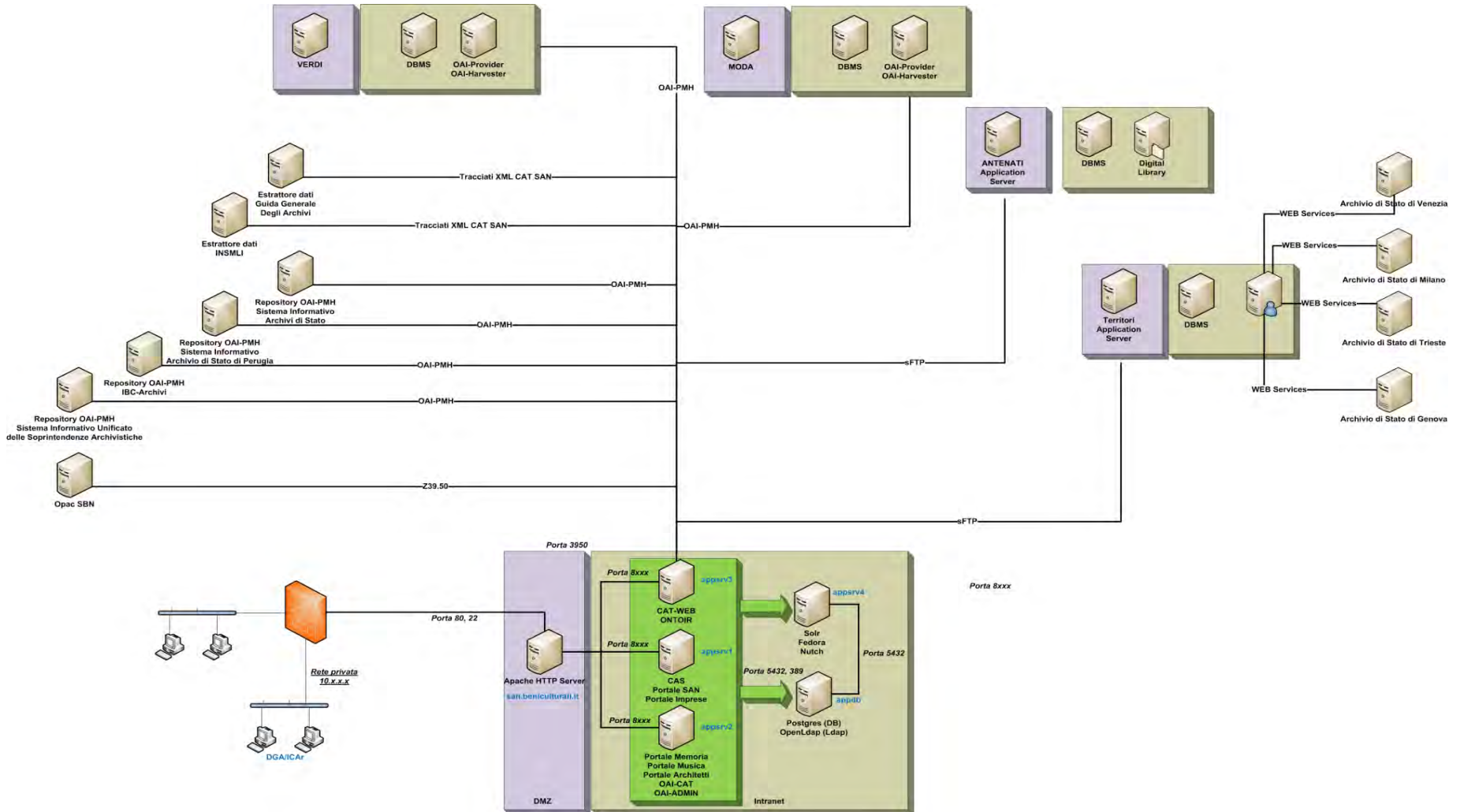
Architettura: Struttura delle applicazioni



Componenti software

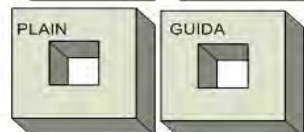
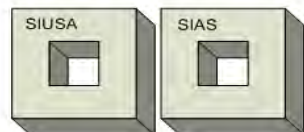
- **JBoss 4.2.3 GA.** Application Server open source
- **Solr 1.3.** Search engine basato su Apache Lucene
- **Fedora Commons 3.2.1.** Repository per oggetti digitali open source
- **Generic search 2.2.** Estensione di Fedora per l'integrazione con Solr
- **Nutch 1.1.** Motore di crawling open source
- **Liferay 5.2.3.** Portale di classe enterprise open source
- **Postgresql 8.3.1.** RDBMS open source

Architettura e sistemi esterni



Architettura: Schema dei flussi

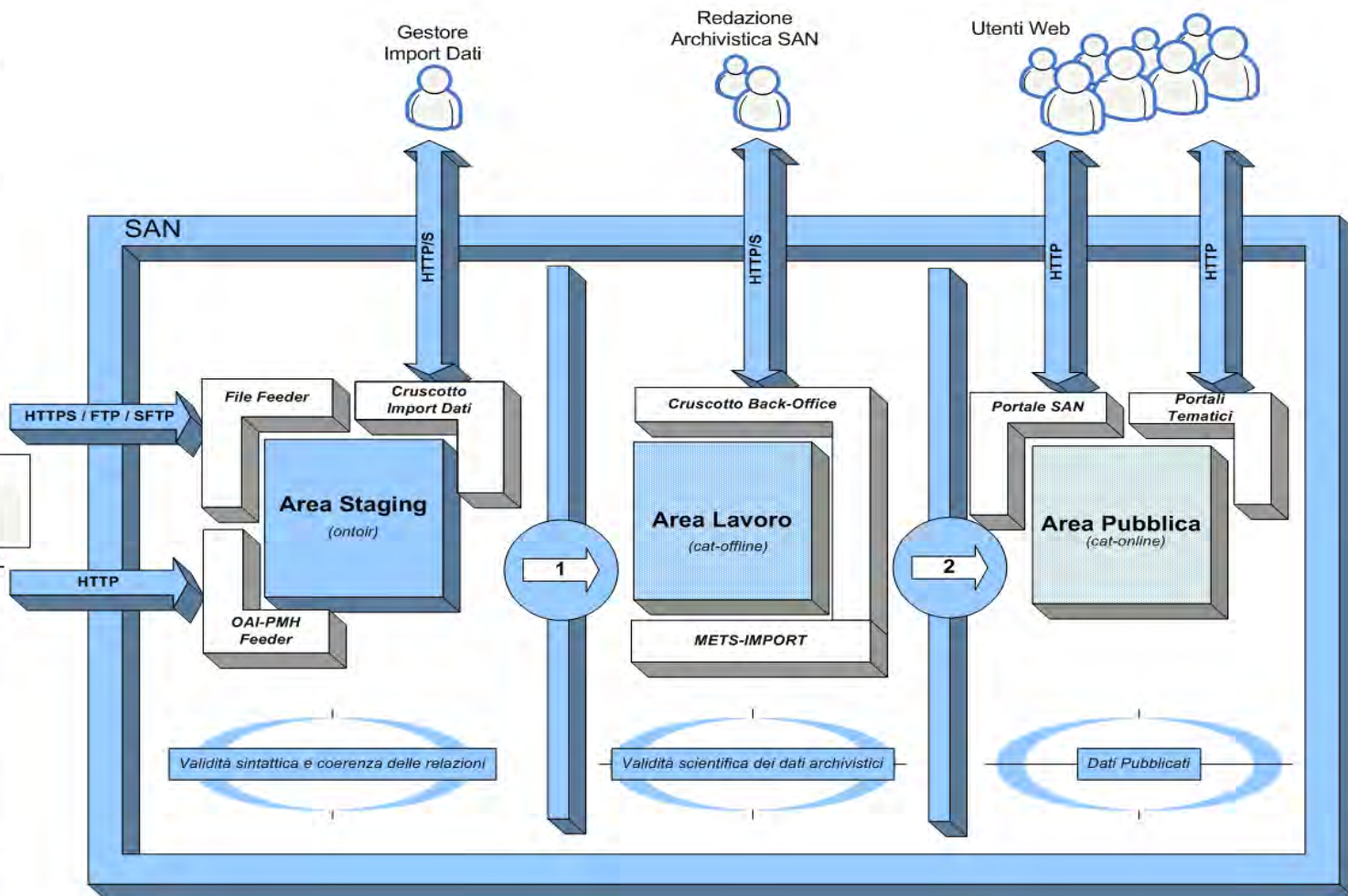
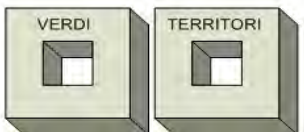
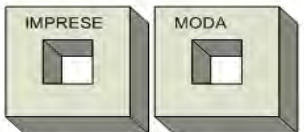
Altri Sistemi / Archivi



Standard Internazionali

XML
CAT

• XML



1

Sistema Archivistico Nazionale

2

Progetto Linked Open Data

3

Struttura dei dati CAT-SAN

4

Disegno di Ontologie

5

Ontologia SAN

6

SKOS e Thesaurus SAN

7

Openrefine

8

Conclusioni

SAN e Open Data : Obiettivi

L' **Istituto Centrale per gli Archivi** ha intrapreso un percorso per rendere accessibili i contenuti del SAN in **formato Open Data** secondo le modalità, i criteri e le tempistiche previste dal “**Decreto Crescita 2.0**” convertito nella legge 221/2012

Sul sito dell'ICAR, sono disponibili i primi set di Open Data sui Conservatori di Archivi all'indirizzo <http://www.icar.beniculturali.it/getFile.php?id=548>.

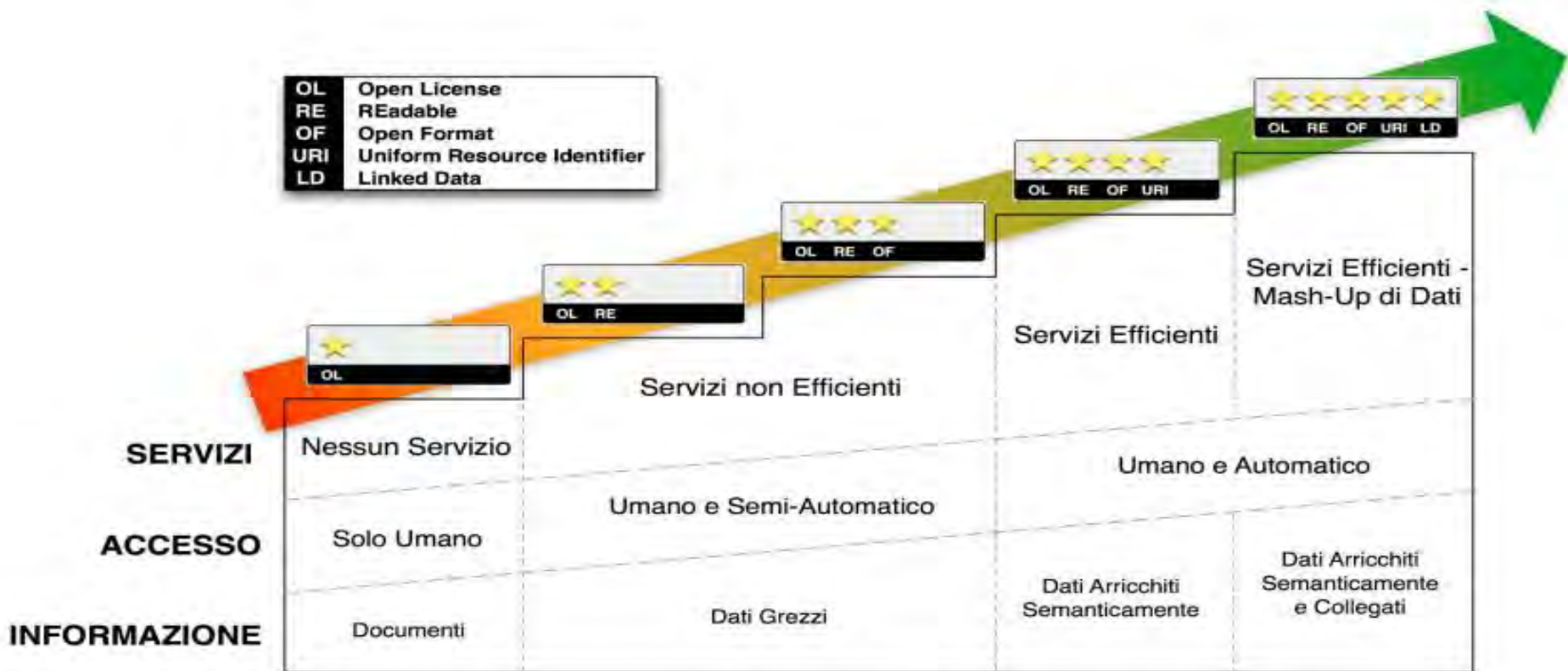
I dati esposti sono da intendersi come dati aperti e liberamente acquisibili da chiunque e riutilizzabili con licenza IODL 2.0 (<http://www.dati.gov.it/iodl/2.0/>).

Alla prima azione di pubblicazione di Open Data, prevista per tutta la Pubblica Amministrazione nel suo complesso, l'ICAR ha affiancato un progetto per lo sviluppo e la pubblicazione di **Linked Open Data di ambito archivistico**.

Trasversalmente a tutte le attività e gli obiettivi, si è adottata una strategia progettuale che prevede di sperimentare le criticità legate alla operazione, per arrivare a definire compiutamente un modello concettuale e una metodologia operativa in grado di integrare nel SAN “**OPEN**”, nella loro pienezza informativa, complessi archivistici e soggetti produttori forniti dai sistemi aderenti che intendano cooperare.

I Linked Open Data

I **Linked Open Data** (LOD) sono una particolare tipologia di dati strutturati semanticamente che vengono pubblicati sul web con licenze di consultazione e d'uso aperte o poco restrittive (es. Creative Commons).



Obiettivi

Obiettivi del progetto per lo sviluppo e la pubblicazione di **Linked Open Data di ambito archivistico** :

1. Definizione una **Ontologia di base** che rappresenti lo schema concettuale della struttura informativa espressa dai tracciati di rappresentazione dei dati SAN, previa analisi di similari esperienze nazionali ed internazionali.
2. Definizione una serie di **estensioni all'Ontologia di base** che permetta l'integrazione di entità esterne e relazioni tra oggetti attualmente non presenti all'interno del SAN, ma residenti nei sistemi di provenienza o in altre fonti individuate (ufficiali o rese disponibili come risorse LOD: geonames, dbpedia, etc etc);
3. Realizzazione di **liste di authority, repertori, strumenti** su scala nazionale da mettere a disposizione della comunità archivistica
4. Disponibilità di una serie di **strumenti di accesso ai dati LOD** prodotti (endpoint sparql, oai provider, set specifici di dati LOD, etc.);
5. Realizzazione di una sperimentazione di un **metodo di estensione** della struttura informativa del SAN, realizzata attraverso l'integrazione della ontologia di base con entità e relazioni tra oggetti non previsti in sede di elaborazione dei tracciati CAT.

Open Data e SAN : toponimi e storia

Il progetto prevede l'elaborazione di una lista di autorità di **toponimi** su scala nazionale, basata principalmente sulla **rielaborazione in formato LOD** di dati istituzionali di fonte ISTAT integrabili con fonti di particolare rilevanza storica, che sarà sperimentalmente raccordata coi dati SAN LOD di provenienza Guida Generale degli Archivi di Stato.

Il **Sistema Informativo Storico delle Amministrazioni Territoriali (SISTAT)** gestisce gli Enti territoriali definiti nell'ordinamento costituzionale italiano, organizzati gerarchicamente rispetto al territorio di competenza amministrativa (dall'ambito territoriale più grande al più piccolo: Regioni; Province; Comuni).

Tutti gli Enti territoriali descritti sono definiti genericamente nel sistema Unità Amministrative (UA). Delle UA vengono descritte:

- data di istituzione e/o variazione
- estremi del provvedimento amministrativo che le istituisce o ne sancisce una variazione, inclusa la soppressione
- codici identificativi Istat e nomenclature
- entità di livello gerarchicamente superiore o inferiore

Il SISTAT incamera dati a partire dall'Unità d'Italia (1861). Per le fonti precedenti si utilizzeranno dati desumibili da fonti archivistiche.

Open Data e SAN : toponimi e storia

Nell'ambito del modello concettuale dell'Atlante storico istituzionale del SAN l'espressione *Regione spazio temporale* viene usata per rappresentare **entità** di vario tipo (geografica, politico amministrative, religiose, storiche) a qualunque livello di posizione gerarchica (es. comune, provincia, regione, stato, etc.) assumibile in un contesto storico/temporale definito.

Il modello si presenta pertanto come naturalmente predisposto ad innalzare ovvero approfondire i criteri di individuazione e definizione delle singole entità, arrivando potenzialmente a contemplare nel concetto di "regione" tanto il micro toponimo (es. località o frazione) quanto estese aggregazioni di stati (es. Ducato di Savoia e Regno di Sicilia, poi Regno di Sardegna).

Ci si propone di mettere a punto un vero e proprio Atlante storico istituzionale in formato LOD, realizzando un modello concettuale basato sulla identificazione di una classe astratta denominata **Regione Spazio Temporale (RST)**.

"Regione" è un termine usato comunemente in differenti branche della geografia per riferirsi ad una porzione di territorio – «luogo» - caratterizzata da requisiti di unitarietà definiti sulla base di caratteristiche geografiche, storiche o antropiche.

Strumenti di accesso ai dati LOD



OAI-PMH



Un **OAI Provider** che distribuisca dati pubblici secondo gli schemi CAT – SAN.



www.san.beniculturali.it



Un'area di **download** dove poter scaricare i vari set di LOD resi disponibili dal SAN.



dati.san.beniculturali.it



Un **endpoint sparql** per la pubblicazione e l'interrogazione dell'ontologia e dei dataset SAN disponibili in formato LOD.



Digital Library SAN Repository SAN

1

Sistema Archivistico Nazionale

2

Progetto Linked Open Data

3

Struttura dei dati CAT-SAN

4

Disegno di Ontologie

5

Ontologia SAN

6

SKOS e Thesaurus SAN

7

Openrefine

8

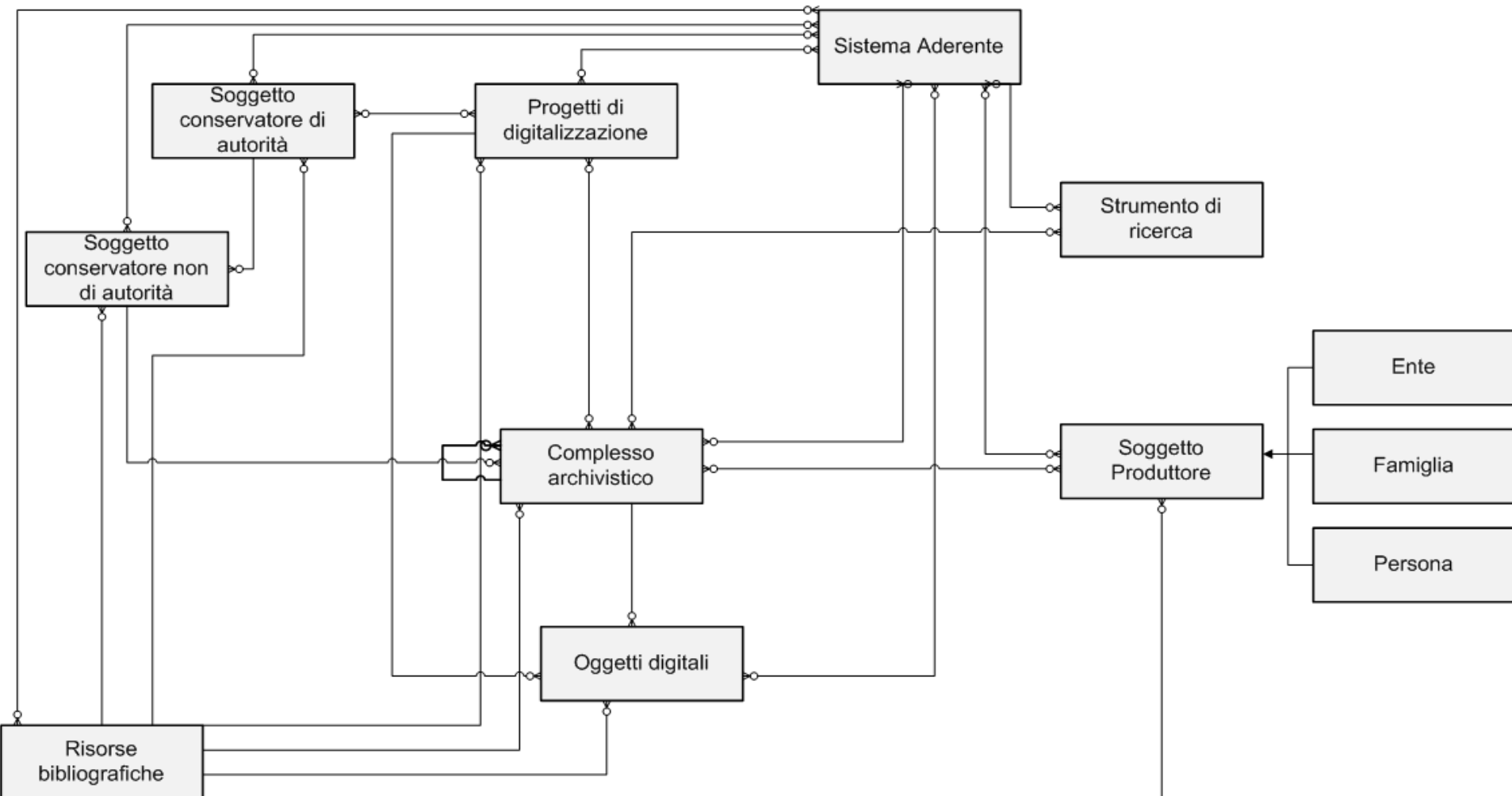
Conclusioni

Gestione catalogo: Entità

Le entità trattate nel SAN

- **Soggetti conservatori**, ovvero i soggetti che conservano i complessi archivistici
- **Complessi archivistici** di primo livello o di livello inferiore, che siano fondi autonomi, oppure che abbiano propri soggetti produttori, distinti da quelli dei complessi archivistici nel quale sono inclusi
- **Soggetti produttori**, ovvero i soggetti collegati ai complessi archivistici, in quanto loro produttori, importati dai sistemi aderenti
- **Strumenti di ricerca**, solo in funzione dei complessi archivistici cui si riferiscono e ai fini di una sommaria valutazione della pertinenza di questi ultimi rispetto ai loro interessi di ricerca
- **Progetti di digitalizzazione**, ovvero un insieme caratterizzato da metainformazioni che identifica gli oggetti digitali collegati ad una specifica attività di digitalizzazione
- **Oggetti digitali** ovvero oggetti quali immagini, video, audio, testo che caratterizzano i complessi archivistici per i quali è prevista l'acquisizione nel repository del sistema accompagnati da relativa metadattazione

Schema dei dati : Modello concettuale degli oggetti SAN



Tracciati CAT SAN: descrizione

- Le entità archivistiche sono importate tramite lo standard cat-import, il formato di scambio per le entità archivistiche, definito dalla Commissione tracciati CAT coordinata dal dott. Stefano Vitali.
- Il formato cat-import è un contenitore degli schemi delle quattro entità soggette ad importazione, espressi secondo i seguenti standard:
 - Per i **soggetti conservatori** è utilizzato lo standard **EAC – Encoded Archival Context**
 - Per i **complessi archivistici** è utilizzato lo standard : **EAD – Encoded Archival Description**
 - Per i **soggetti produttori** è utilizzato lo standard **EAC-CPF – EAC-Corporate Body, Persons and Families**
 - Per gli **strumenti di ricerca** è utilizzato lo standard **EAD – Encoded Archival Description**
- Per i progetti di digitalizzazione non è previsto un formato di scambio ma un tracciato utilizzato nell'ambito del SAN

Tracciati CAT-SAN: Complessi Archivistici

N.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
01	tipologia del complesso	Il sistema di provenienza riporterà l'indicazione della tipologia del complesso così come compare nel relativo campo	/ead/archdesc[@level="otherlevel"]/@otherlevel	obbligatorio
02	codice identificativo (id) del record descrittivo nel sistema di provenienza;	Codice identificativo del record esportato nel sistema di provenienza; nell'attributo @type andrà indicato l'acronimo del sistema di provenienza del record che servirà per stabilire il collegamento con i record SAN di descrizione del sistema di provenienza. il codice identificativo (id) del record nel sistema di provenienza deve essere un id. stabile ed univoco, così come lo deve essere l'URI/URL indicato nell'attributo @identifier	/ead/archdesc/did/unitid/@type "sistemadiprovenienza" @identifier	Obbligatorio Univoco insieme all'id del Sistema aderente
03	denominazione o titolo	Il sistema individuerà il campo (o i campi) che riporta la denominazione del complesso; se vengono esportate anche altre denominazioni, accoppierà a quella principale il relativo attributo	/ead/archdesc/did/unitid/@type="principale" se vengono esportate anche altre denominazioni	obbligatorio
04	altre denominazioni	Il sistema individuerà il campo (o i campi) che riporta le eventuali altre denominazioni del complesso e accoppierà a ciascuna di esse la relativa qualifica, se presente nel sistema di provenienza	/ead/archdesc/did/unitid/@type="\$qualifica"]	Ripetibile

Tracciati CAT-SAN: Complessi Archivistici

N.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
05	date testuali (per la visualizzazione); date codificate	Il sistema di provenienza riporterà la stringa alfanumerica adottata per la visualizzazione sul web, comprese le eventuali notazioni per date critiche, associando l'attributo @normal per la relativa codifica (sotto forma di sequenza di 8 cifre per data remota e recente); vanno incluse qui anche eventuali ulteriori estensioni alla data (ad es. "con documentazione dal"); per queste estensioni se si forniscono anche altre date normalizzate per estendere la datazione, si accoppia all'elemento l'attributo indicato	/ead/archdesc/did/unitdate/ attributi: @normal, per le date codificate @datechar "principale"	Obbligatorio Le date sono codificate in una sequenza di 8 cifre per data remota e recente nella forma AAAAMMGG/AAAAMMGG.
06	estensione delle datazione (date codificate)	Il sistema di provenienza indicherà eventuali ulteriori date codificate che estendano il range della data principale; l'elemento sarà vuoto e solo l'attributo sarà valorizzato. Per specificare la qualità della data (es: "con documentazione dal") sarà associato un attributo (@datechar) valorizzato come da vocabolario controllato, secondo una mappatura fra le qualifiche adottate nel sistema e il vocabolario controllato. Se nel sistema di provenienza le altre denominazioni non sono qualificate, la qualifica assumerà il valore "qualifica non indicata"	/ead/archdesc/did/unitdate/@normal @datechar	Ripetibile Le date sono codificate con una sequenza di 8 cifre ed n range di date con due coppie di 8 cifre nella forma AAAAMMGG/AAAAMMGG Voci del vocabolario controllato: (con documentazione dal, con documentazione fino al, data della documentazione compresa, con documentazione in copia dal, - con lacuna)

Tracciati CAT-SAN: Complessi Archivistici

N.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
07	consistenza dell'unità di descrizione	Il sistema individuerà i campi che riportano le indicazioni sulla consistenza e, se necessario, assocerà in un'unica stringa alfanumerica, quelli che indicano la consistenza numerica e quelli che specificano la tipologia documentaria; l'assenza di informazioni sulla consistenza nel sistema di provenienza sarà segnalata nel file d'esportazione/importazione con una formula che potrà essere adottata anche in sede di visualizzazione (es.: "non indicata")	/ead/archdesc/did/physdesc/extent	obbligatorio
08	descrizione sintetica	Se il sistema afferente non è dotato di un elemento descrittivo direttamente corrispondente a questo (tipo abstract della descrizione), esso individuerà le modalità che ritiene più opportune per valorizzarlo, predisponendo in sede di generazione del file d'esportazione le informazioni che dovranno confluirci, per un massimo di 1.500 caratteri; indicare il codice linguistico della lingua della descrizione nell'attributo corrispondente	/ead/archdesc/did/abstract@langcode	non obbligatorio
09	status del record nel sistema di provenienza	Il sistema stabilirà una mappatura fra la terminologia o le convenzioni adottate per indicare lo status del record esportato e il vocabolario controllato del CAT, che includerà una voce per segnalare l'assenza dell'informazione nel sistema di provenienza (es: "non indicato")	/ead/archdesc/processinfo	obbligatorio con vocabolario controllato

Tracciati CAT-SAN: Complessi Archivistici

N.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
10	relazione con scheda SAN di descrizione del Soggetto Conservatore	la Denominazione del soggetto conservatore nel sistema di provenienza e/o nel CAT se reso disponibile dal SAN prima della generazione dei file d'esportazione; Nell'attributo @id riportare l'id del record del soggetto conservatore nel sistema di provenienza oppure nel CAT. Precisare nell'attributo @label il sistema cui l'id si riferisce Essendo un gruppo ripetibile possono essere indicati la denominazione e/o l'id del soggetto conservatore sia nel sistema di provenienza che nel CAT	/ead/archdesc/did/repository/ @id @label	obbligatorio uno dei due codici ripetibile
11	link al complesso di livello superiore eventualmente presente nel CAT	Codice identificativo del record del complesso archivistico di primo livello;	/ead/archdesc/ relatedmaterial/archref	obbligatorio se il complesso descritto è già un primo livello si riporterà qui l'id dello stesso complesso.
12	relazione con scheda CAT di descrizione del Soggetto/i Produttore/i collegato/i nel sistema di provenienza.	la Codice identificativo dei soggetti produttori con i quali il record esportato del complesso archivistico è in relazione nei sistemi di provenienza	/ead/archdesc/did/origin/	non obbligatorio ripetibile
13	relazione con scheda CAT degli strumenti di ricerca collegati	la Codice identificativo (id) del record descrittivo dello strumento di ricerca nel sistema di provenienza	/ead/archdesc/otherfaindaid/extref/	non obbligatorio ripetibile

Tracciati CAT-SAN: Complessi Archivistici

N.	Tracciato CAT	EAD CAT import profile	Note
14	Codice identificativo (id) del record descrittivo nel CAT	Generato automaticamente in CAT. Dovrà essere compreso o nel formato che SAN userà per esporre all'esterno i propri dati	obbligatorio
15	relazione con la scheda SAN relativa al sistema/progetto di digitalizzazione, nel caso in cui il complesso selezionato disponga di immagini digitali	Il collegamento sarà stabilito dalla redazione SAN a partire dal progetto di digitalizzazione	non obbligatorio ripetibile

Tracciati CAT-SAN: Soggetto Produttore

n.	Tracciato CAT	Sistemi di provenienza	EAC CAT import profile	Note
01	Codice identificativo (id) del record descrittivo nel sistema di provenienza;	Codice identificativo del record esportato nel sistema di provenienza; nell'attributo @localtype andrà indicato l'acronimo del sistema di provenienza del record. il codice identificativo (id) del record nel sistema di provenienza deve essere un id. stabile ed univoco	/eac-cpf/control/otherRecordId @localtype "sistemadiprovenienza"	obbligatorio Univoco insieme all'id del sistema aderente
02	Collegamento alla scheda descrittiva del soggetto produttore nel sistema di provenienza	URI della scheda nel sistema di provenienza. L'URI deve essere stabile e in ogni caso devono comunque essere previste procedure di aggiornamento delle schede CAT nel caso di modifica dell'URI nel sistema di provenienza.	/eac-cpf/control/sources/source/@xlink:href	obbligatorio
03	Tipologia del soggetto produttore	Il sistema riporterà la tipologia del soggetto, eventualmente mappandola sul vocabolario controllato del campo	/eac-cpf/cpfDescription/identity/entityType/	obbligatorio a vocabolario controllato: corporate body (per tipologia soggetto conservatore ente), person (per tipologia soggetto conservatore persona), family (per tipologia soggetto conservatore famiglia)
04	Qualifica di profilo istituzionale	Da valorizzare unicamente se il record esportato è un profilo istituzionale e se la tipologia del soggetto è "corporate body"	/eac-cpf/cpfDescription/identity/@localType="profilo istituzionale"	se l'attributo è valorizzato, il record importato è collegato soltanto ad altre schede soggetto produttore e non a schede complesso archivistico

Tracciati CAT-SAN: Soggetto Produttore

n.	Tracciato CAT	Sistemi di provenienza	EAC CAT import profile	Note
05	Forme autorizzate del nome nel sistema di provenienza	Forme autorizzate del nome così come indicate nel sistema di provenienza;	/eac-cpf/cpfDescription/identity/nameEntry/part ogni forma autorizzata genera una nuova nameEntry	obbligatorio (ma cfr. elemento 6) e ripetibile
06	Forma parallela nel caso di bilinguismo	Forme parallele del nome così come indicate nel sistema di provenienza; Ogni forma parallela deve essere accoppiata alla forma autorizzata pertinente Nel caso di forme parallele si usa quindi questo elemento per entrambe le forme e non si usa perciò l'elemento 05. Nell'attributo @langcode saranno indicate le lingue in cui sono formulate le denominazioni	/eac-cpf/cpfDescription/identity/nameEntryParallel/nameEntry/part @langcode	obbligatorio (nel caso di bilinguismo)
07	Altre denominazioni	Eventuali altre denominazioni del soggetto produttore	/eac-cpf/cpfDescription/identity/nameEntry[@localType="altradenominazione"]/part	non obbligatorio ripetibile

Tracciati CAT-SAN: Soggetto Produttore

n.	Tracciato CAT	Sistemi di provenienza	EAC CAT import profile	Note
08	Date testuali Date codificate	Ogni sistema selezionerà le date da esportare accoppiando a ciascuna data testuale per la visualizzazione (costituita dalla stringa alfanumerica adottata per la visualizzazione nel sistema di provenienza, comprese le eventuali notazioni per date critiche) una corrispondente data codificata singola o una range di date codificate ed una qualifica che indichi la tipologia della data esportata (es: data di nascita, data di morte, date di esistenza, ecc.); se la qualifica non è presente o non può essere ricavata nel sistema di provenienza, il suo valore sarà sostituito nel file d'esportazione/importazione da una formula che ne segnali l'assenza e che potrà essere adottata anche in sede di visualizzazione (es.: "non indicata")	Ogni data selezionata per l'esportazione sarà composta da un elemento /eac- cpf/cpfDescription/description/ existDates/dateSet/ che conterrà, sia per date singole che per range di date il seguente insieme di elementi ripetibili: a) <date> contenete la data o il range di date testuali; b) l'attributo @localType per la qualifica della data (secondo le voci del vocabolario controllato); c) l'attributo @standarDate per le date codificate singole oppure per un range di date.	obbligatorio ripetibile le date sono codificate con una sequenza di 8 cifre; un range di date è codificato con due coppie di 8 cifre nella forma AAAAMMGG/ AAAAMMGG voci del vocabolario controllato: data di nascita, data di morte, date di esistenza, data di istituzione, data di soppressione
09	Luoghi	Ogni sistema selezionerà i luoghi, da esportare associando a ciascuno di essi la qualifica come da vocabolario controllato; se la qualifica non è presente o non può essere ricavata nel sistema di provenienza, il suo valore sarà sostituito da una formula che ne segnali l'assenza e che potrà essere adottata anche in sede di visualizzazione (es.: "non indicata")	Tutti i luoghi selezionati per l'esportazione saranno contenuti nell'elemento eac-cpf/cpfDescription/placeDates/ placeDate/place/ che conterrà sempre, per ciascuno dei luoghi, a) <place> per il nome del luogo; b) <descriptiveNote> per la qualifica da vocabolario controllato	Obbligatorio, eccetto che nel caso in cui sia stato valorizzato l'elemento 4 ripetibile voci del vocabolario controllato: luogo di nascita, luogo di morte, sede, ambito territoriale, Giurisdizione, Residenza

Tracciati CAT-SAN: Soggetto Produttore

n.	Tracciato CAT	Sistemi di provenienza	EAC CAT import profile	Note
10	Qualificazione	Ogni sistema selezionerà le qualificazioni del soggetto produttore, da esportare associando a ciascuna di esse una qualifica come da vocabolario controllato (es: tipologia dell'ente, professione, titoli nobiliari, ecc ; se la qualifica non è presente o non può essere ricavata nel sistema di provenienza, il suo valore sarà sostituito nel file d'esportazione/importazione da una formula che ne segnali l'assenza e che potrà essere adottata anche in sede di visualizzazione (es.: "non indicata")	Tutte le qualificazioni selezionate per l'esportazione saranno contenute nell'elemento eac-cpf/cpfDescription/description/descriptiveEntries/ che conterrà, per ciascuna delle qualificazioni, un <descriptiveEntry> che a sua volta conterrà sempre a) <term> per specificare la qualificazione b) <descriptiveNote> per la qualifica da vocabolario controllato	non obbligatorio ripetibile voci del vocabolario controllato: - professione - titolo nobile - natura giuridica - tipologia ente - sottotipologia ente
11	Descrizione sintetica	Se il sistema afferente non è dotato di un elemento descrittivo direttamente corrispondente a questo (tipo abstract della descrizione), esso individuerà le modalità che ritiene più opportune per valorizzarlo, predisponendo in sede di generazione del file d'esportazione le informazioni che dovranno confluirci, per un massimo di 1.500 caratteri.	/eac-cpf/cpfDescription/description/biogHist/abstract	non obbligatorio

Tracciati CAT-SAN: Soggetto Produttore

n.	Tracciato CAT	Sistemi di provenienza	EAC CAT import profile	Note
12	Status del record nel sistema di provenienza	Il sistema stabilirà una mappatura fra la terminologia o le convenzioni adottate per indicare lo status del record esportato e il vocabolario controllato del CAT, che includerà una voce per segnalare l'assenza dell'informazione nel sistema di provenienza (es: "non indicato")	/eac-cpf/control/maintananceStatus	obbligatorio con vocabolario controllato: scheda pubblicata, scheda in via di compilazione, pubblicazione locale/interna, scheda di prova, scheda in via di pubblicazione, da rivedere, rivista, cancellata
13	Relazione con le schede CAT dei Complessi archivistici collegati	Codice identificativo dei complessi archivistici con i quali il record esportato del soggetto produttore è in relazione nel sistema di provenienza	/eac-cpf/cpfDescription/relations/resourceRelation[@resourceRelationType="creatorOf"]/relationEntry	non obbligatorio ripetibile
14	Relazione con le schede CAT di altri Soggetti produttori collegati	Codice identificativo dei soggetti produttori con i quali il record esportato è in relazione nel sistema di provenienza; la relazione può essere qualificata, importando la relativa qualifica, anche temporale, presente del sistema di provenienza; andrà anche indicato se il soggetto produttore ente (o il profilo istituzionale) è connesso ad un profilo istituzionale.	Ogni relazione sarà contenuta nell'elemento /eac-cpf/cpfDescription/relations/cpfRelation, con l'elemento <relationEntry> con l'indicazione dell'Id del soggetto produttore connesso. Se la relazione è con un profilo istituzionale <cpfRelation> conterrà anche l'elemento <descriptiveNote> con la specifica "profilo istituzionale".	non obbligatorio ripetibile

Tracciati CAT-SAN: Strumenti di ricerca

n.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
01	Codice identificativo (id) del record descrittivo nel sistema di provenienza	Codice identificativo del record esportato nel sistema di provenienza; nell'attributo @identifier andrà indicato l'acronimo del sistema di provenienza; il codice identificativo (id) del record nel sistema di provenienza deve essere un id. stabile ed univoco	ead/eadheader/eadid @identifier @URL	Obbligatorio
02	Autore/i	Se nel sistema di provenienza gli autori dello strumento di ricerca sono contenuti in un unico campo, l'elemento sarà valorizzato con il contenuto del campo. Se nel sistema di provenienza il campo autore è ripetibile, nell'elemento saranno riportati in sequenza gli autori, separati da una virgola.	ead/eadheader/filedesc/ titlestmt/author	non obbligatorio nel caso di strumenti di ricerca editi il valore dell'elemento potrà essere inserito anche nel campo autore della relativa scheda bibliografia, se non già valorizzato
03	Titolo	Titolo dello strumento di ricerca così come indicato nel sistema di provenienza.	ead/eadheader/filedesc/ titlestmt/titleproper	Obbligatorio nel caso di strumenti di ricerca editi il valore dell'elemento potrà essere inserito anche nel campo titolo della relativa scheda bibliografia, se non già valorizzato.
04	Date	Data/e dello strumento di ricerca espressa con una sequenza alfanumerica non strutturata, così come indicata nel sistema di provenienza. Se il sistema di provenienza non riporta la data dello strumento di ricerca, l'elemento sarà valorizzato con apposita espressione: (es.: "non indicata")	ead/eadheader/filedesc/ publicationstmt/date	Obbligatorio nel caso di strumenti di ricerca editi il valore dell'elemento potrà essere inserito anche nel campo data della relativa scheda bibliografia, se non già valorizzato.

Tracciati CAT-SAN: Strumenti di ricerca

n.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
05	[Luogo di pubblicazione]	Luogo di edizione nel caso di strumenti di ricerca editi, se l'informazione è disponibile in forma strutturata nel sistema di provenienza	ead/eadheader/filedesc/publicationstm/address	Se l'informazione è disponibile in forma strutturata nel sistema di provenienza, il valore dell'elemento potrà essere inserito anche nel relativo campo della scheda bibliografia, se non già valorizzato.
06	[Casa editrice]	Casa editrice nel caso di strumenti di ricerca editi, se l'informazione è disponibile in forma strutturata nel sistema di provenienza	ead/eadheader/filedesc/publicationstm/publisher	Se l'informazione è disponibile in forma strutturata nel sistema di provenienza, il valore dell'elemento potrà essere inserito anche nel relativo campo della relativa scheda bibliografia, se non già valorizzato
07	Descrizione	A seconda della presenza e delle modalità di organizzazione dei loro dati, i sistemi di provenienza potranno far confluire in questo elemento, per un massimo di 1500 caratteri, informazioni di natura diversa: a) descrizione dello strumento di ricerca; e/o b) tipologia dello strumento; c) la parte del fondo cui lo strumento si riferisce; d) altre informazioni pertinenti. L'attributo @type deve essere valorizzato soltanto nel caso di strumenti di ricerca editi, con il termine "edito".	ead/eadheader/filedesc/notestmt/note@type	non obbligatorio

Tracciati CAT-SAN: Strumenti di ricerca

n.	Tracciato CAT	Sistemi di provenienza	EAD CAT import profile	Note
08	Relazione con la scheda CAT dei Complessi archivistici collegati	Codice identificativo (id) nel sistema di provenienza del record descrittivo del complesso archivistico esportato e cui lo strumento di ricerca è collegato in CAT	/ead/archdesc/did/unitid	
09	URI nel caso di strumenti di ricerca in formato digitale presenti in linea		ead/eadheader/filedesc/editionstmt/edition/extptr @href @title (per nota descrittiva del link che funge da titolo del link)	non obbligatorio ripetibile
10	Codice identificativo (id) del record descrittivo nel CAT	Generato automaticamente in CAT. Dovrà essere compreso o nel formato che SAN userà per esporre all'esterno i propri dati		10
11	Relazione con la scheda bibliografica allo strumento di ricerca edito se presente nel modulo repertorio bibliografico del SAN	Il link sarà stabilito dalla redazione SAN. La valorizzazione degli elementi 5 e 6 e/o l'esplicita segnalazione nel campo 7 costituirà una spia della necessità di stabilire il collegamento e/o di creare la relativa scheda bibliografia		11
10	Codice identificativo (id) del record descrittivo nel CAT	Generato automaticamente in CAT. Dovrà essere compreso o nel formato che SAN userà per esporre all'esterno i propri dati		10
11	Relazione con la scheda bibliografica allo strumento di ricerca edito se presente nel modulo repertorio bibliografico del SAN	Il link sarà stabilito dalla redazione SAN. La valorizzazione degli elementi 5 e 6 e/o l'esplicita segnalazione nel campo 7 costituirà una spia della necessità di stabilire il collegamento e/o di creare la relativa scheda bibliografia		11

Tracciati CAT-SAN: Soggetto conservatore

N.	Tracciato CAT	Note	Profilo di esportazione CAT
01	Forma autorizzata del nome	obbligatorio	/scons/formaautorizzata formeparallele
02	Acronimo		/scons/acronimo
03	Forma parallela del nome in casi di bilinguismo	obbligatorio in caso di bilinguismo	/scons/formeparallele
04	Codice identificativo nei sistemi in cui è presente la scheda dello stesso soggetto e Link alla scheda descrittiva	obbligatorio	/scons/identifier/recordId /scons/identifier/sistemald /scons/@href
05	Tipologia	obbligatorio	/scons/tipologia
06	Localizzazione	obbligatorio ripetibile indirizzo e localizzazione geografica	/scons/localizzazione @paese @provincia @comune @cap
07	Sito web e altri siti pertinenti	Non obbligatorio Ripetibile	/scons/sitoweb @href
08	Servizio di consultazione al pubblico Si/no	Obbligatorio	/scons/consultazione "true" o "false"
09	Orario di apertura al pubblico (se 8 valorizzato sì)	Obbligatorio se valorizzato sì il n. 8	/scons/orario solo se <consultazione>="true"
10	Altre modalità di accesso (da valorizzare se è stato valorizzato non il n. 8)	Obbligatorio se valorizzato no il n 8	/scons/altroaccesso
11	Servizi al pubblico	Non obbligatorio	/scons/servizi
12	Descrizione	Non obbligatorio	/scons/descrizione

1 Sistema Archivistico Nazionale

2 Progetto Linked Open Data

3 Struttura dei dati CAT-SAN

4 Disegno di Ontologie

5 Ontologia SAN

6 SKOS e Thesaurus SAN

7 Openrefine

8 Conclusioni

Ontologia: Definizione di Borst e Gruber

“L'ontologia è una specificazione formale ed esplicita di una concettualizzazione condivisa”

- **concettualizzazione**: rappresentazione, in un modello astratto, di un determinato dominio del mondo reale attraverso i concetti (classi) rilevanti che lo caratterizzano;
- **esplicita**: i concetti, le relazioni fra questi, le proprietà del dominio debbono essere definiti in maniera non ambigua;
- **formale**: deve essere espressa in un linguaggio formale comprensibile da una macchina;
- **condivisa**: un'ontologia deve rappresentare conoscenza condivisa da un gruppo.

Un ontologia è quindi una rappresentazione (schema concettuale) esaustiva e rigorosa di un dato dominio; contiene tutte le classi (concetti) rilevanti, le relazioni fra di esse, gli individui (istanze) che popolano la realtà rappresentata e che sono inseriti nelle classi.

Per ogni classe dell'ontologia, quindi per tutti gli individui della classe, vengono definite proprietà (gli attributi) che ne descrivono caratteristiche ed eventuali restrizioni. Le classi possono essere primitive, se non derivano da altre presenti nell'ontologia, oppure derivate se sono ereditate da altre classi presenti.

Progettazione dell'ontologia

Per definire una ontologia dobbiamo:

- Definire le classi
 - Organizzare le classi in una gerarchia tassonomica (sottoclassi-superclassi)
 - Definire le proprietà e descrivere i valori leciti per ciascuna di esse
 - Attribuire i valori alle proprietà per tutte le istanze create.
1. Elencare i termini più importanti del dominio su cui vogliamo fare affermazioni e cerchiamo di rispondere alle seguenti domande:
 - Quali sono i termini di cui vogliamo parlare?
 - Quali sono le proprietà ad essi associate?
 - Cosa vogliamo dire a proposito di essi?
 2. Definire le classi e l'organizzazione gerarchica dell'ontologia:
 - Metodo Top-down: si parte dalla definizione dei concetti più generali e si specializzano tali definizioni in concetti più specifici.
 - Metodo Bottom-up: si parte dalla definizione delle classi più specifiche, e si cerca di raggrupparle in gruppi concettuali più astratti.

Spesso il metodo migliore è una combinazione dei due precedenti in una azione ciclica.

Progettazione dell'ontologia

3. Definire le proprietà delle classi che verranno modellate nelle classi relative:

- Proprietà “intrinseche” properties (il gusto di un vino)
- Proprietà “estrinseche” (il nome di un vino)
- Parti costituenti, se l'oggetto è strutturato (le parti del corpo)
- Relazioni ad altri soggetti (costruttore di ...)

Gli attributi sono ereditati dalle sottoclassi.

4. Definire i vincoli sulle proprietà (tipo dati, numeri e valori permessi, ecc.):

- Cardinality: quanti valori posso avere in un attributo
- Value type: tipo di dato
- Domain: a quali classi si applica questo attributo
- Range: classi permesse per gli attributi definibili sulle istanze.

5. Creare le istanze

- L'ultimo passo di sviluppo consiste nella creazione delle istanze delle classi presenti nell'ontologia.

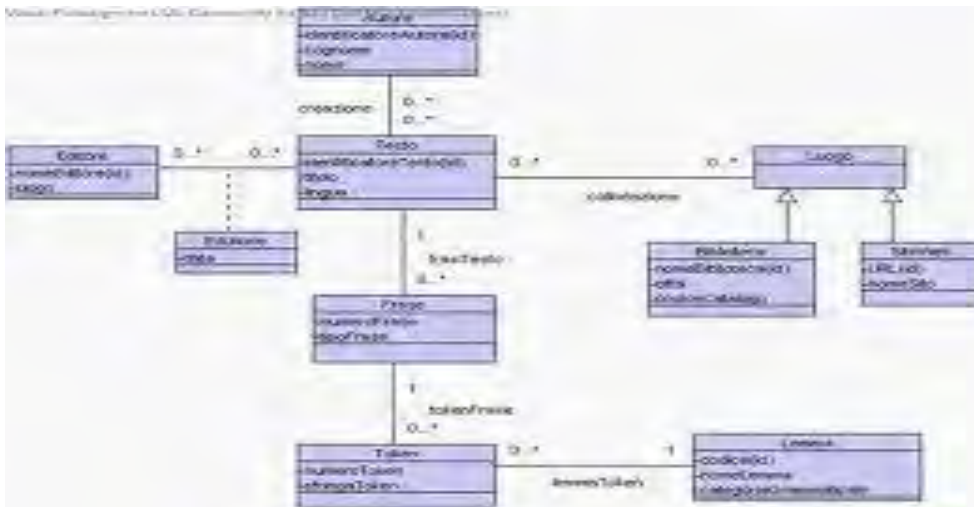
A questo punto spesso si scoprono errori di modellazione nei passi precedenti.

Differenze ontologie/database

Le ontologie ed i database sono entrambe metodologie di rappresentazione della conoscenza.

In un **database** l'attenzione è maggiormente centrata sui **dati** e sulle loro strutturazioni, mentre in un'**ontologia** si cerca di modellare concettualmente il **dominio con entità e relazioni**.

Ricordando la progettazione di un Database, possiamo pensare lo schema concettuale dei dati come un'ontologia della realtà del problema.



Revisione continua dell'ontologia

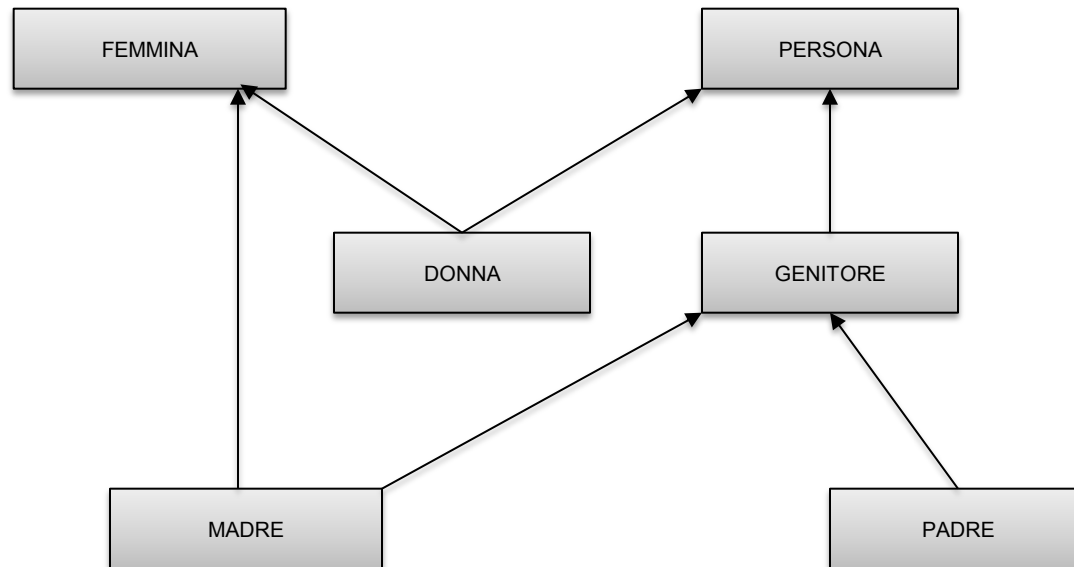
Un'ontologia deve riflettere la realtà del mondo che cerca di modellare nel miglior modo possibile.

Non vi è però un metodo corretto a priori per modellare un dominio di conoscenza e che la soluzione migliore dipende dall'applicazione prevista e dalle sue estensioni future, e che il processo di sviluppo di un'ontologia è un'opera continuamente in divenire.



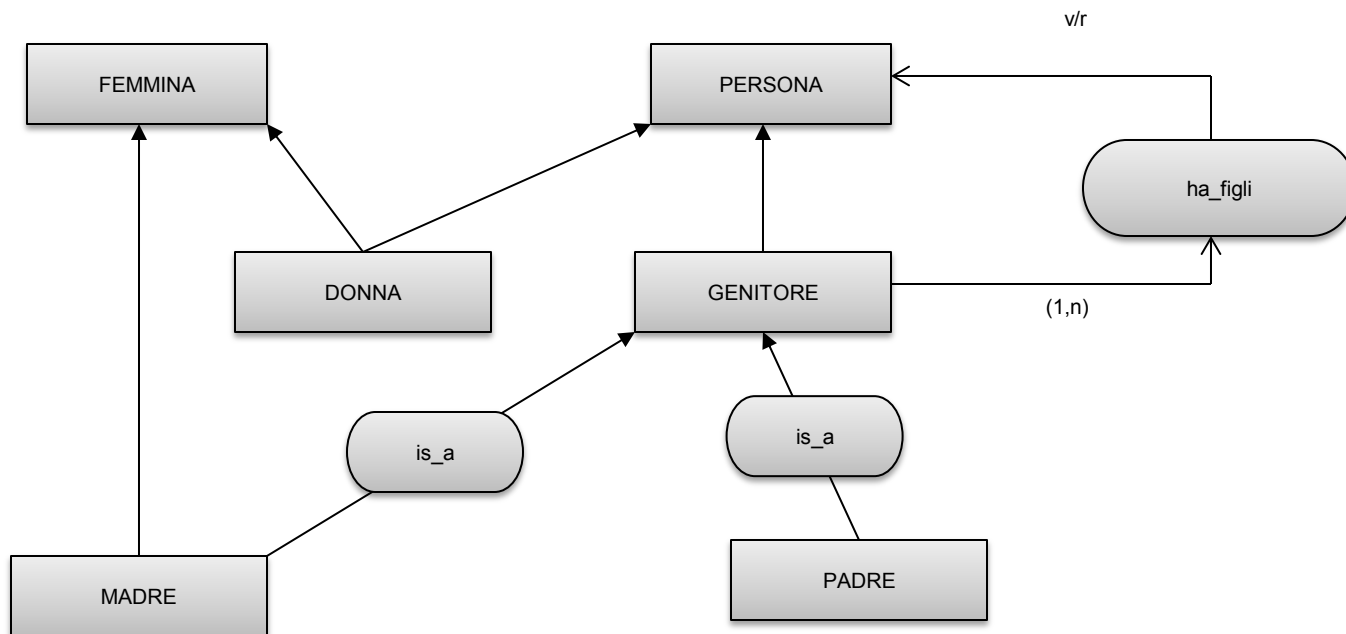
Elementi dell'ontologia: classi

La “classe” viene utilizzata per descrivere i concetti (le classi di individui) che si vuole rappresentare (es. PERSONA, GENITORE, MADRE ...)



Elementi dell'ontologia: proprietà

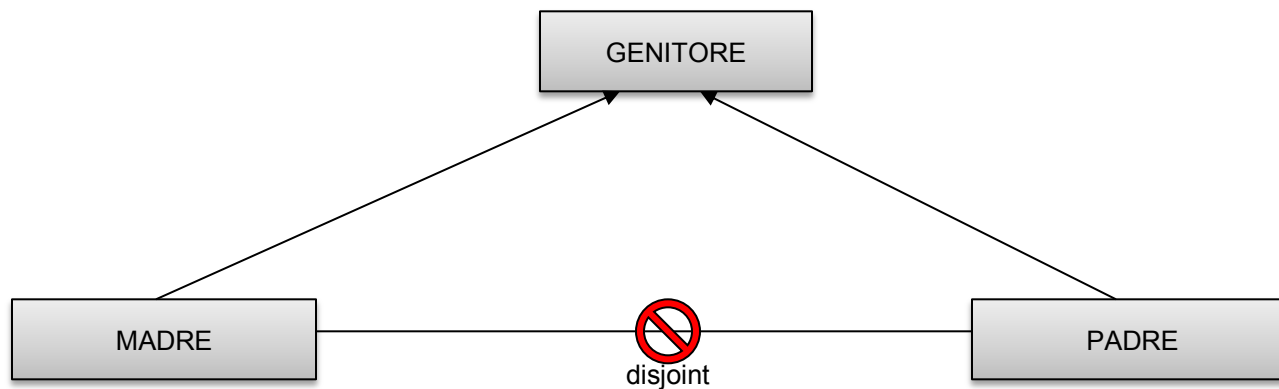
Rappresentano le relazioni esistenti tra le classi del dominio (es. “le MADRI sono GENITORI”, “i GENITORI hanno figli che sono PERSONE”...)



Elementi dell'ontologia: assiomi

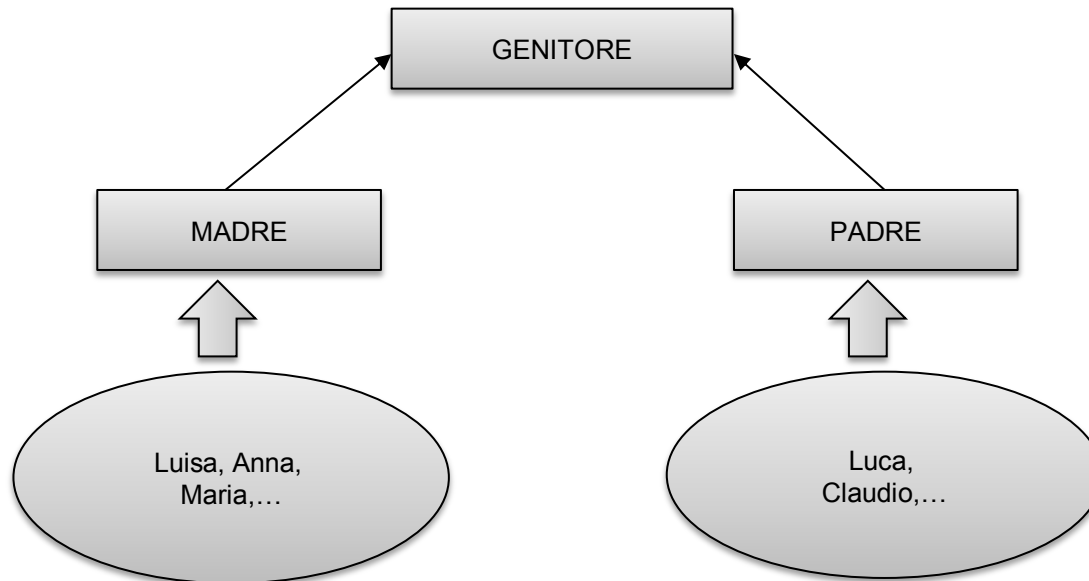
Gli assiomi vengono utilizzati per modellare in maniera esplicita espressioni in ogni caso vere

Possono essere utilizzati per diversi scopi: definire il significato dei vari componenti dell'ontologia, definire relazioni complesse, verificare la correttezza dell'informazione specificata o dedurre nuova informazione (es. *disjoint* (MADRE, PADRE) esprime il fatto che un elemento della classe PADRE non può mai essere anche un elemento della classe MADRE).



Elementi dell'ontologia: istanze

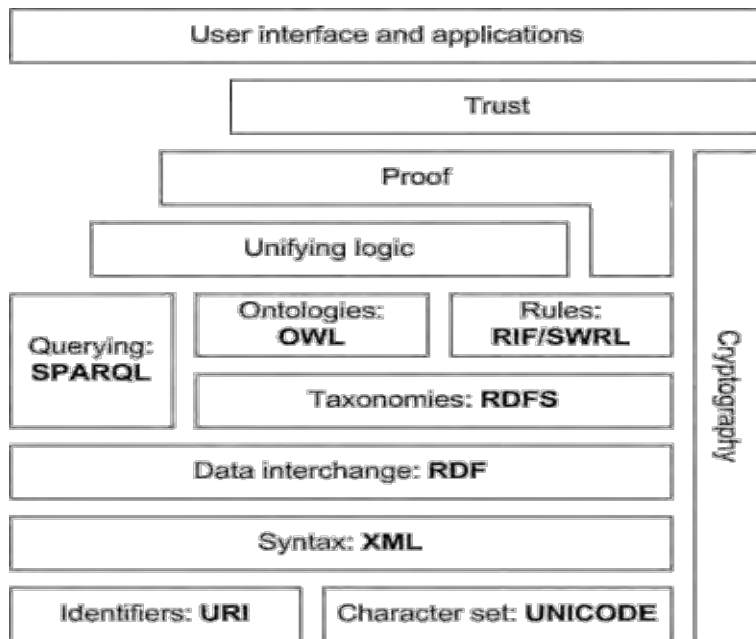
Sono i singoli oggetti contenuti in una classe, a vari livelli di generalità, a seconda dello scopo dell'ontologia



OWL - Web Ontology Language

Il linguaggio OWL (Web Ontology Language) è stato sviluppato dal World Wide Web Consortium (W3C Recommendation 10 February 2004) per la definizione di ontologie e di strumenti compatibili con l'architettura del Semantic Web.

Tramite OWL è possibile avere una maggiore comprensibilità della semantica di concetti da parte di agenti software rispetto a XML, RDF e RDF-Schema.



Semantic Web Stack Berners-Lee (2006)

OWL è stato costruito sulla base di RDF e RDF-Schema e rispetto ad essi aggiunge diversi costrutti per la definizione dei concetti e delle loro interrelazioni:

- **disgiunzione o equivalenza tra classi**
- **restrizioni di cardinalità per le proprietà**
- **relazioni di uguaglianza tra classi o proprietà**
- **classi enumerate**
- **nuovi tipi di proprietà e nuove caratteristiche per le proprietà.**

OWL – Sintassi di base

- definire una classe

```
<owl:Class rdf:about="http://san.beniculturali.it/SAN/ente">  
<rdfs:label rdf:datatype="string">ente</rdfs:label>  
<rdfs:subClassOf rdf:resource="Thing"/>
```

- dire che questa classe è disgiunta da un'altra

```
<owl:Class rdf:about="http://san.beniculturali.it/SAN/ente">  
<rdfs:label rdf:datatype="string">ente</rdfs:label>  
<rdfs:subClassOf rdf:resource="http://san.beniculturali.it/SAN/produttore"/>  
<owl:disjointWith rdf:resource="http://san.beniculturali.it/SAN/famiglia"/>  
<owl:disjointWith rdf:resource="http://san.beniculturali.it/SAN/persona"/>  
<owl:equivalentClass rdf:resource="corporateBoby"/>  
</owl:Class>
```

- dire che questa classe è equivalente ad un'altra

```
<owl:equivalentClass rdf:resource="Organization"/>
```

OWL – Sintassi di base

- Restrizione di proprietà nel definire le classi

Quantificatore universale: `allValueFrom`, tutti i valori di una proprietà devono provenire da quella classe

```
<owl:Class rdf:about="Courses" >  
<rdfs:Comment >i corsi possono essere tenuti solo da professori</rdfs:Comment >  
<rdfs:SubClassOf >  
  <owl:Restriction >  
    <owl:onProperty rdf:resource="#is_taught_by" / >  
    <owl:allValueFrom rdf:resource="#professor" / >  
  </owl:Restriction >  
</rdfs:SubClassOf >  
</owl:Class>
```


OWL – Sintassi di base

- Restrizione di proprietà nel definire le classi

Quantificatore esistenziale: esiste un corso del primo anno tenuto da un professore

```
<owl:Class rdf:about="Professors" >
```

```
<rdfs:Comment >i professori insegnano qualche corso universitario  
</rdfs:Comment >
```

```
<rdfs:SubClassOf >
```

```
<owl:Restriction >
```

```
<owl:onProperty rdf:resource="#teach" / >
```

```
<owl:someValueFrom rdf:resource="#Courses" / >
```

```
</owl:Restriction >
```

```
</rdfs:SubClassOf >
```

```
</owl:Class>
```

OWL – Sintassi di base

- Owl:minCardinality: specifica il valore minimo che può assumere una proprietà
<rdfs:subClassOf>
 <owl:Restriction>
 <owl:onProperty
 rdf:resource="http://san.beniculturali.it/SAN/formaAutorizzataProduttore"/>
 <owl:**minCardinality** rdf:datatype="nonNegativeInteger">1</owl:minCardinality>
 </owl:Restriction>
 </rdfs:subClassOf>
- Owl:maxCardinality: specifica il valore massimo che può assumere una proprietà
 <owl:Restriction>
 <owl:onProperty rdf:resource="http://san.beniculturali.it/SAN/descrizione"/>
 <owl:**maxCardinality** rdf:datatype="nonNegativeInteger">1</owl:maxCardinality>
 </owl:Restriction>

OWL – Sintassi di base

- Proprietà: Object property che collega un oggetto con un altro oggetto
<owl:ObjectProperty rdf:about="http://san.beniculturali.it/SAN/has_tipologia">
<rdfs:label rdf:datatype="string">ha tipologia</rdfs:label>
<rdfs:domain rdf:resource="http://san.beniculturali.it/SAN/conservatore"/>
<rdfs:range rdf:resource =
 "http://san.beniculturali.it/SAN/TesaurusSAN_1.0/TipologiaConservatore"/>
</owl:ObjectProperty>
- Proprietà: Data type property: collega un oggetto con un valore
<owl:DatatypeProperty rdf:about="http://san.beniculturali.it/SAN/titolo">
<rdfs:label rdf:datatype="string">titolo</rdfs:label>
<rdfs:comment> Tracciato schema SAN </rdfs:comment>
<rdfs:domain rdf:resource="http://san.beniculturali.it/SAN/strumento"/>
</owl:DatatypeProperty>

OWL – Sintassi di base

- Proprietà inversa definisce una proprietà come inversa di un'altra
<owl:ObjectProperty rdf:about="http://san.beniculturali.it/SAN/has_cpf">
<rdfs:label rdf:datatype="string">ha cpf</rdfs:label>
<rdfs:domain
 rdf:resource="http://san.beniculturali.it/SAN/contestoStoricoIstituzionale/"
<owl:inverseOf rdf:resource="http://san.beniculturali.it/SAN/has_CSI"/>

- Functional Property: definisce una proprietà che ha al massimo un valore per ciascuno soggetto (ad esempio, l'età)
<owl:Datatype Property rdf:ID="hasAge" >
<rdfs:domain rdf:resource="person">
<rdfs:range rdf:resource="www.w3c.org/xmlschema#nonNegativeInteger" >
<rdf:type rdf:resource="owl:FunctionalProperty" >
</ owl:ObjectProperty >

OWL – Sintassi di base

- Inverse Functional Property: definisce una proprietà per cui due differenti oggetti non possono avere lo stesso valore :

```
<owl:ObjectProperty rdf:ID="teaches" >  
<rdfs:domain rdf:resource="staff">  
<rdfs:range rdf:resource="course" >  
<owl:inverseOf rdf:resource="isTaughtBy"  
</ owl:ObjectProperty >
```

- Restrizione di proprietà: hasValue: specifica che valore deve avere una proprietà:

```
<owl:Class rdf:about="MathCourses" >  
<rdfs:Comment >I corsi di matematica sono insegnati da Rossi</rdfs:Comment >  
<rdfs:SubClassOf >  
  <owl:Restriction >  
    <owl:onProperty rdf:resource="#are_tought" / >  
    <owl:hasValue rdf:resource="Rossi" / >  
  </owl:Restriction >
```

OWL – Sintassi di base

- Operatori booleani
 - Complementarietà <owl:ComplementOf >
 - Unione tra classi <owl:UnionOf >
 - Intersezione <owl:intersectionOf” >

```
<owl:Class rdf:about="http://san.beniculturali.it/SAN/produttore">  
<rdfs:label rdf:datatype="string">soggetto produttore</rdfs:label>  
<owl:Class>  
  <owl:unionOf rdf:parseType="Collection">  
    <rdf:Description rdf:about="http://san.beniculturali.it/SAN/ente"/>  
    <rdf:Description rdf:about="http://san.beniculturali.it/SAN/famiglia"/>  
    <rdf:Description rdf:about="http://san.beniculturali.it/SAN/persona"/>  
  </owl:unionOf>  
</owl:Class>
```


OWL – Sintassi di base

- Enumerazione

```
<owl:OneOf rdf:ID="giorni" >  
  < owl:thing rdf:about="#Sabato">  
  < owl:thing rdf:about="#Domenica">  
</ owl: OneOf >
```
- Importare ontologie esistenti:

```
<owl: import rdf:resource="http://purl.org/dc/elements/1.1/">  
<owl: import rdf:resource="http://xmlns.com/foaf/0.1/">
```
- Etichette

```
<owl: label> Complesso archivistico </owl:label
```
- Istanze: le istanze in OWL si definiscono sempre attraverso l’RDF

```
<rdf:Description rdf:ID="9898765">  
  <rdf:type rdf:resource="professor">  
</rdf:Description>
```

1 Sistema Archivistico Nazionale

2 Progetto Linked Open Data

3 Struttura dei dati CAT-SAN

4 Disegno di Ontologie

5 Ontologia SAN

6 SKOS e Thesaurus SAN

7 Openrefine

8 Conclusioni

Open Data e SAN : ontologia di base

Nella definizione dell'ontologia del Sistema Archivistico Nazionale sono stati presi in considerazione tanto gli standard di dominio quanto il riferimento ai principali namespace in ambito LOD per le operazioni di riconciliazione necessarie alla loro integrazione in ambiente Linked Data:

Dublin Core: <http://purl.org/dc/elements/1.1/>

Biographical Vocabulary: <http://purl.org/vocab/bio/0.1/>

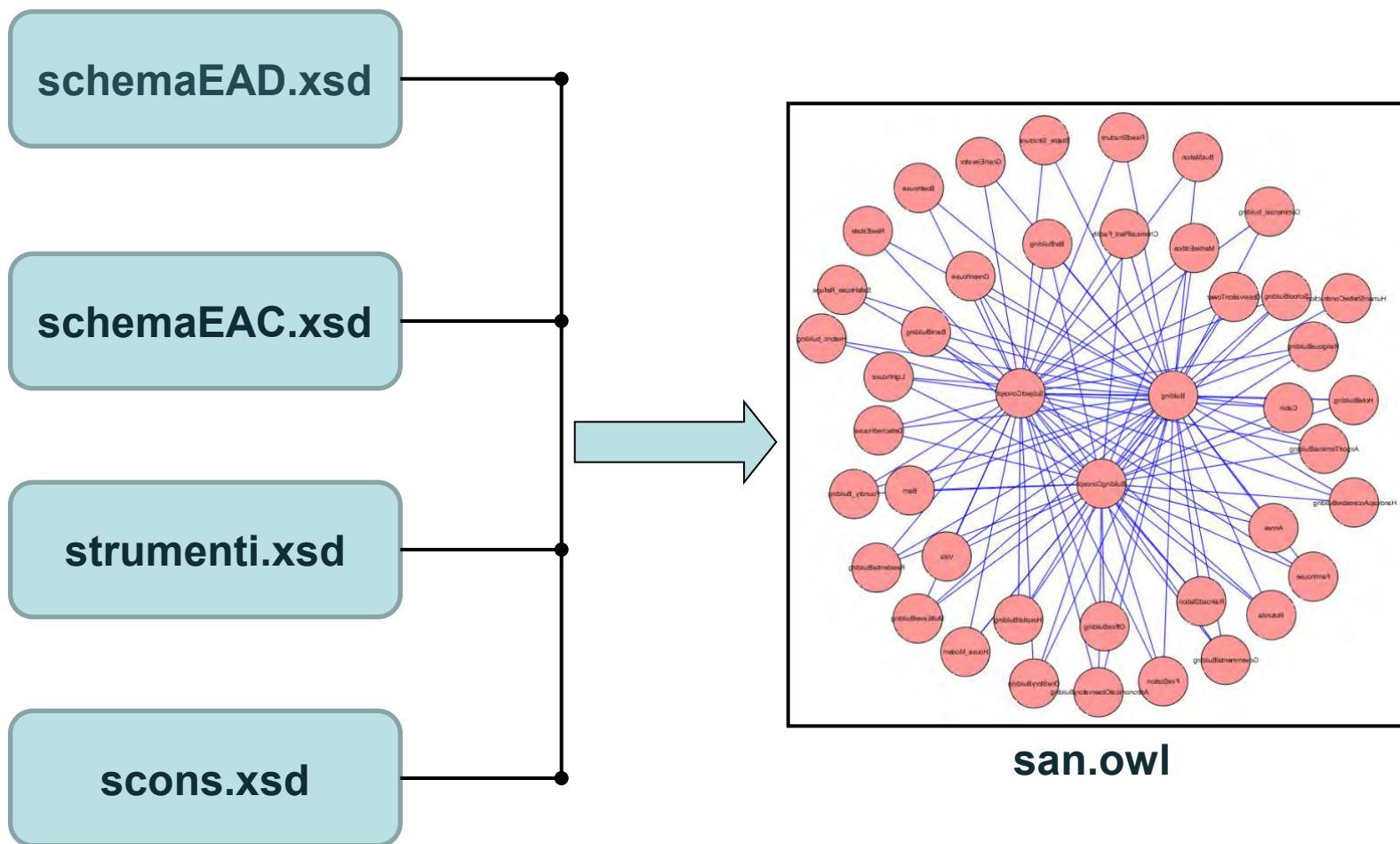
Foaf, Friend of a friend: <http://xmlns.com/foaf/0.1/>

Viaf , Virtual International Authority File: <http://viaf.org/ontology/1.1/#>

dbpedia: <http://dbpedia.org/resource/>

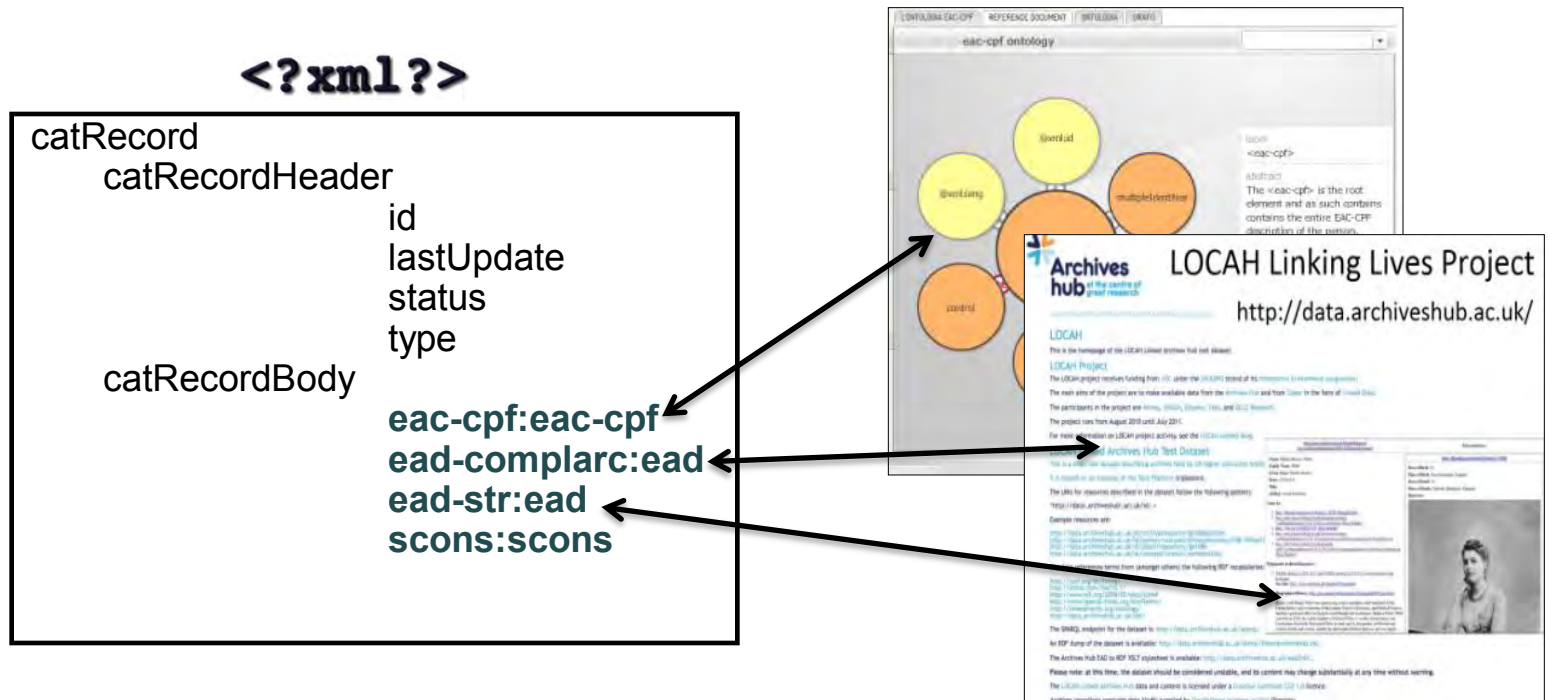
geoNames: <http://www.geonames.org/ontology#>

Open Data e SAN : ontologia di base

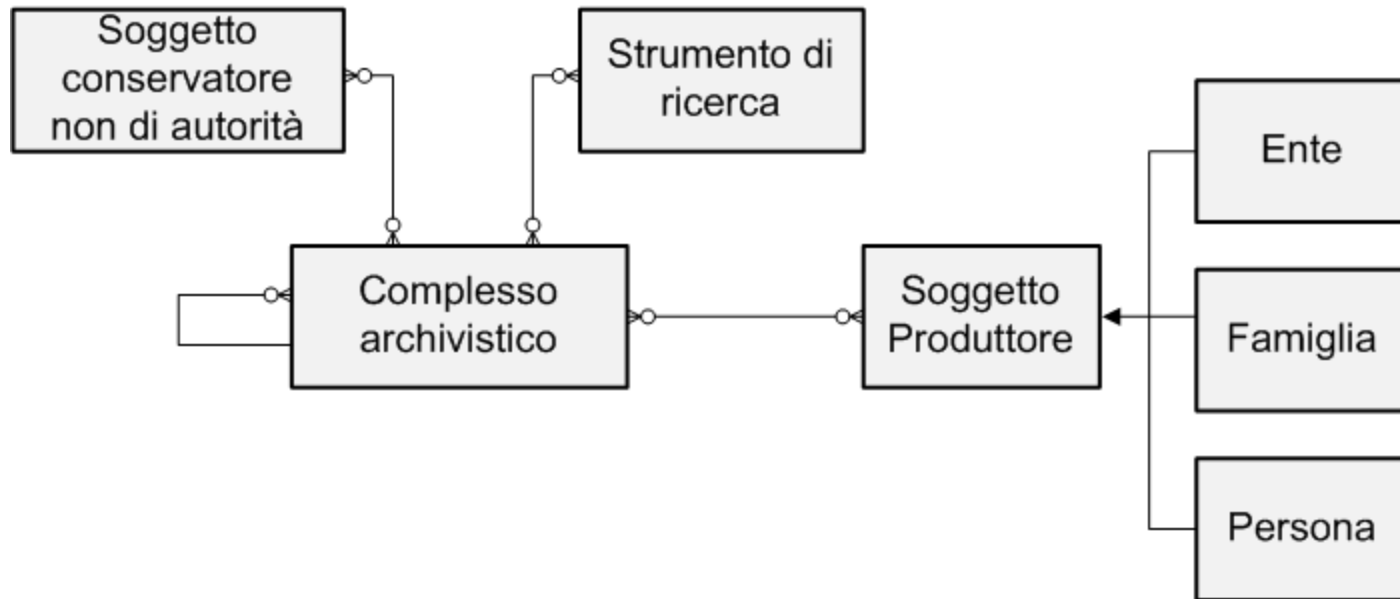


Open Data e SAN

Particolare attenzione è stata prestata a quelle esperienze nazionali (es. Progetto **RELOAD** - Repository for Linked open archival data) ed internazionali (es. progetti **LOCAH** e **Linking Lives**) che hanno affrontato o stanno affrontando – pur con sostanziali peculiarità legate ai rispettivi contesti di applicazione - analoghe problematiche descrittive.



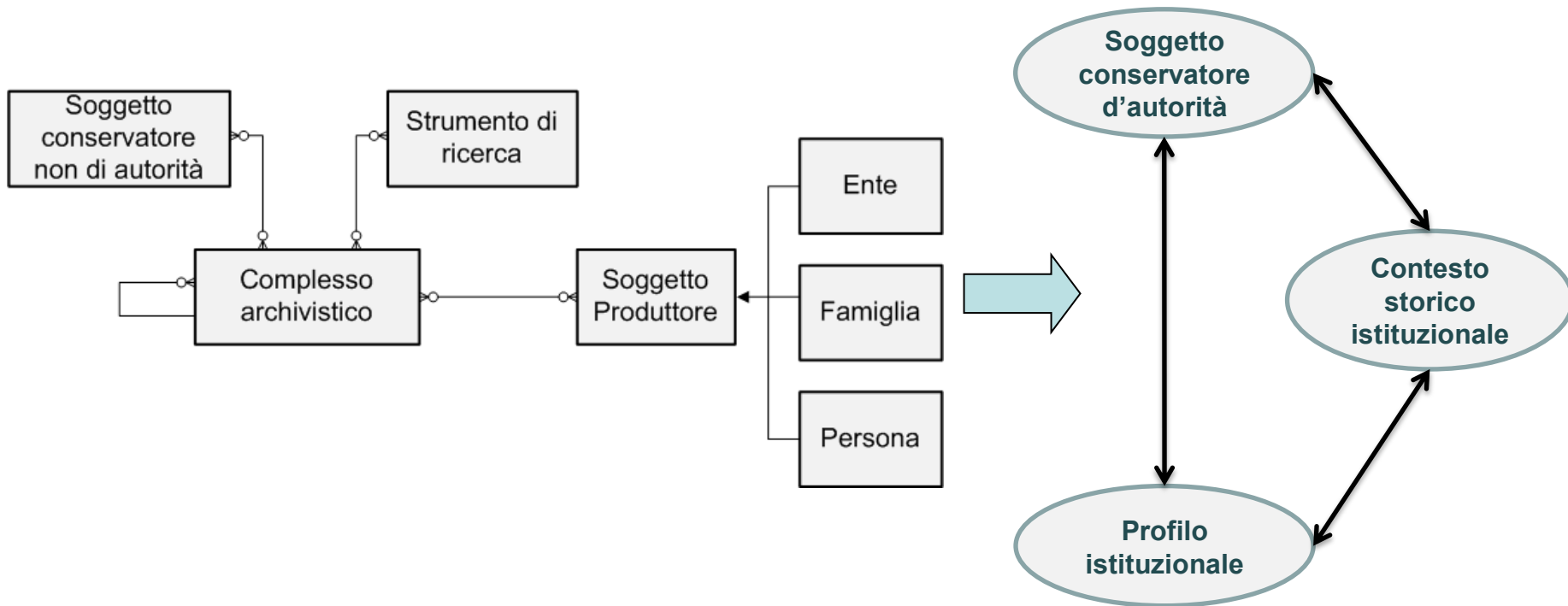
Schema concettuale limitato agli schemi CAT-SAN



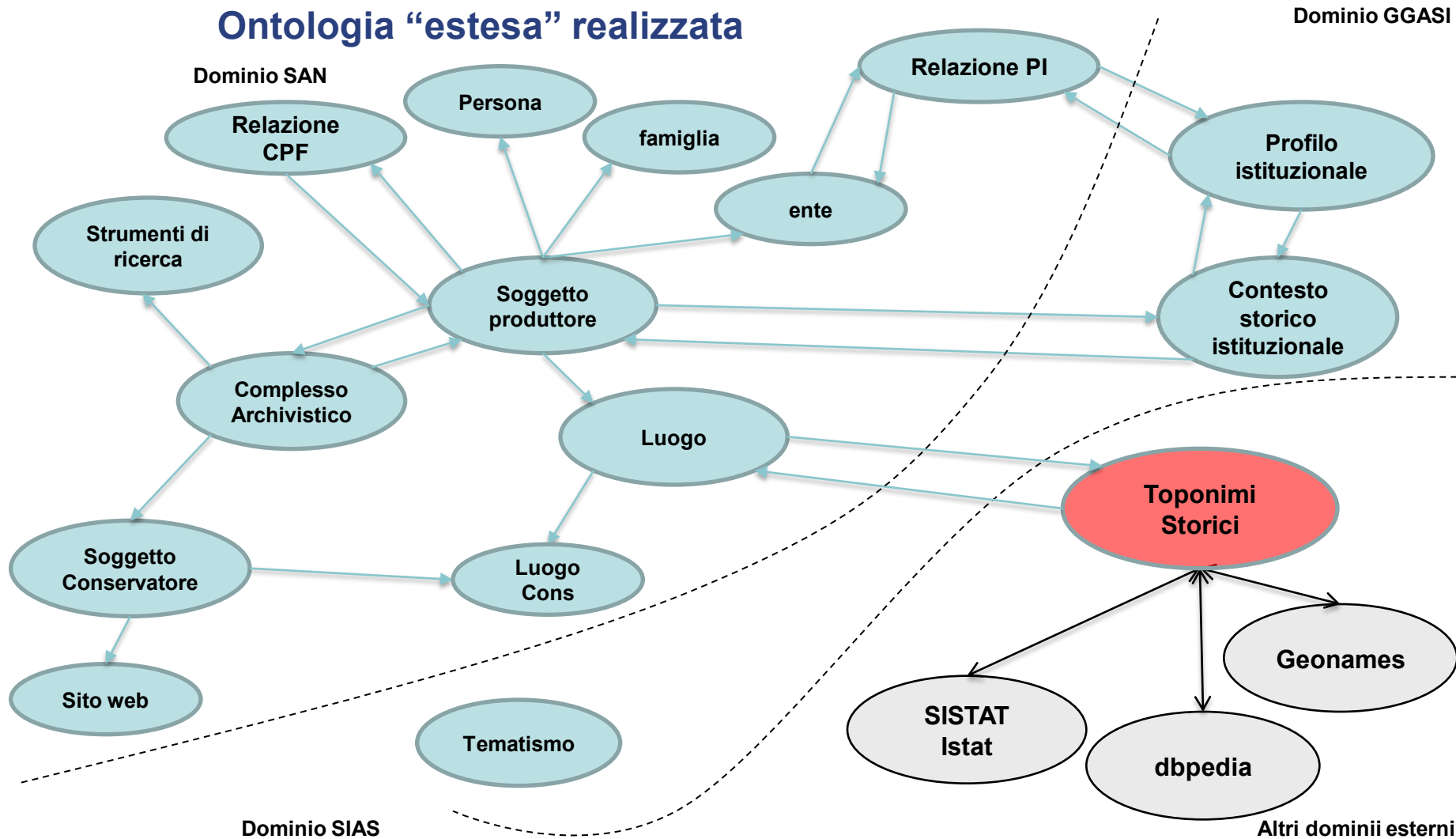
Si è partiti da uno schema concettuale limitato agli oggetti archivistici descritti nei tracciati CAT-SAN per poi poterlo estendere secondo il modello prescelto.

Open Data e SAN

La prima estensione dell'ontologia SAN ha previsto il suo arricchimento con **tre ulteriori classi di oggetti** (soggetto conservatore d'autorità, profilo istituzionale e contesto storico istituzionale) e con tutte quelle proprietà necessarie a descriverle e a metterle in relazione coi soggetti produttori.



Ontologia “estesa” realizzata



Definizioni introdotte nell'ontologia SAN (1/2)

soggetto conservatore

Il soggetto sul quale ricade la responsabilità della conservazione e valorizzazione dell'archivio

complesso archivistico

Locuzione generica per indicare un corpus documentale oggetto di descrizione.

strumento di ricerca:

Repertorio - in forma di inventario analitico o sommario, elenco, elenco di versamento, elenco di consistenza, elenco numerico o altro - che descrive in maniera più o meno dettagliata un fondo o parte di esso.

soggetto produttore

Soggetto (ente, famiglia o persona) che ha prodotto o acquisito per finalità amministrative, familiari o personali il complesso archivistico oggetto di descrizione. Può essere collegato ad altri soggetti produttori con relazioni gerarchiche o temporali.

Definizioni introdotte nell'ontologia SAN (2/2)

profilo istituzionale

Istituzioni, preunitarie o unitarie, descritte una volta per tutte, con riferimento al periodo in cui hanno operato e al contesto storico istituzionale o statale di appartenenza. Descrive una volta per tutte un particolare soggetto produttore, e costituisce una chiave di ricerca per tutti i soggetti produttori che appartengono al profilo descritto.

contesto storico istituzionale

Organizzazione istituzionale entro cui ha operato il soggetto produttore. Può trattarsi di uno Stato o di articolazioni territoriali o amministrative di esso. L'entità statale può essere individuata con la denominazione propria di un determinato periodo storico (per es. "Repubblica di Siena", "Granducato di Toscana") o con le denominazioni assunte in un più lungo arco cronologico (es. "Regno d'Italia poi Repubblica italiana").

Ontologia SAN: i numeri

L'ontologia SAN è composta da 28 tra classi e sottoclassi:

6 classi corrispondenti alle entità principali del SAN

- **complesso archivistico**
- **soggetto conservatore**
- **strumento di ricerca**
- **soggetto produttore (ente,famiglia,persona)**
- **profilo istituzionale**
- **contesto storico istituzionale**

5 riferibili a dimensioni informative generali e/o trasversali alle entità

- **Date (date complesso archivistico, date contesto storico istituzionale, date esistenza, date strumento, estensione date complesso archivistico)**
- **Luogo (luogo conservatore)**
- **sito web**
- **qualificazione**
- **PiRelation (Correlazione profili Istituzionali e soggetti produttori)**

Identificativi persistenti dell'Ontologia del SAN

La disponibilità di riferimenti stabili e certificati su Web per gli oggetti trattati dall'ontologia archivistica SAN è fondamentale per le applicazioni per identificare queste risorse in modo affidabile, ma anche per garantirne l'accessibilità nel tempo e la referenziabilità da altri domini ed ontologie.

Gli identificativi utilizzati sono in un forma standard del tipo:

- per il complesso archivistico
 - http://san.benculturali.it/SAN/complarc_GGASI_GG0840052147
 - http://san.benculturali.it/SAN/complarc_SIAS_IT-ASPA-F70010325
- per il soggetto produttore
 - http://san.benculturali.it/SAN/produttore_GGASI_SP081135
 - http://san.benculturali.it/SAN/produttore_SIAS_IT-ASPA-CC70010115
- per il soggetto conservatore
 - http://san.benculturali.it/SAN/conservatore_ASpalermo
 - http://san.benculturali.it/SAN/conservatore_AStorino
- per lo strumento di ricerca
 - http://san.benculturali.it/SAN/strumento_SIAS_IT-ASPA-SR70010327
- per i termini dei vocabolari skos
 - http://san.benculturali.it/SAN/TesauroSAN_1.0/Pergamene
 - http://san.benculturali.it/SAN/TesauroSAN_1.0/archivio_di_Stato

Prefissi utilizzati nell'ontologia SAN

Nell'ontologia SAN sono stati importati ed utilizzati una serie di namespace ed ontologie esterne:

SAN	http://san.beniculturali.it/SAN/
DC	http://purl.org/dc/elements/1.1/
eac-cpf	http://archivi.ibc.regione.emilia-romagna.it/ontology/eac-cpf#
foaf	http://xmlns.com/foaf/0.1/
terms	http://purl.org/dc/terms/
org	http://www.w3.org/ns/org#
gn	http://www.geonames.org/ontology#
Rdfs	http://www.w3.org/2000/01/rdf-schema#
ocsa	http://lod.xdams.org/reload/ocsa/
owl	http://www.w3.org/2002/07/owl#
xsd	http://www.w3.org/2001/XMLSchema#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
oad	http://lod.xdams.org/reload/oad/
skos	http://www.w3.org/2004/02/skos/core

Ontologia SAN. Classe Complesso Archivistico

IRI: <http://san.beniculturali.it/SAN/complessoArchivistico>

Locuzione generica per indicare un corpus documentale oggetto di descrizione.

is equivalent to

[oad:uod](#)

has super-classes

[owl:Thing](#)

has cardinality constrains

[consistenza^{dp}](#) **exactly** 1

[abstract^{dp}](#) **max** 1

[sistema provenienza^{dp}](#) **exactly** 1

[tipologia complesso^{dp}](#) **exactly** 1

[forma autorizzata complesso archivistico^{dp}](#) **exactly** 1

[ha livelloSuperiore^{op}](#) **exactly** 1

[scheda provenienza href^{dp}](#) **exactly** 1

[record provenienza id^{dp}](#) **exactly** 1

[ha statusProvenienza^{op}](#) **exactly** 1

is in domain of

[abstract^{dp}](#), [altra denominazione complesso archivistico^{dp}](#), [consistenza^{dp}](#), [forma autorizzata complesso archivistico^{dp}](#), [ha conservatore^{op}](#), [ha date complesso archivistico^{op}](#), [ha estensione date complesso archivistico^{op}](#), [ha livelloSuperiore^{op}](#), [ha produttore^{op}](#), [ha statusProvenienza^{op}](#), [ha strumento di ricerca^{op}](#), [ha tematismo^{op}](#), [ordinamento^{dp}](#), [record provenienza id^{dp}](#), [scheda provenienza href^{dp}](#), [sistema provenienza^{dp}](#), [tipologia complesso^{dp}](#)

is in range of

[ha complesso archivistico^{op}](#), [ha livelloSuperiore^{op}](#)

Ontologia SAN. Classe contesto storico istituzionale

IRI: <http://san.beniculturali.it/SAN/contestoStoricoIstituzionale>

Organizzazione istituzionale entro cui ha operato il soggetto produttore. Può trattarsi di uno Stato o di articolazioni territoriali o amministrative di esso. L'entità statale può essere individuata con la denominazione propria di un determinato periodo storico (per es. "Repubblica di Siena", "Granducato di Toscana") o con le denominazioni assunte in un più lungo arco cronologico (es. "Regno d'Italia poi Repubblica italiana").

has super-classes

[owl:Thing](#)

is in domain of

[denominazione contesto storico istituzionale^{dp}](#), [descrizione contesto storico istituzionale^{dp}](#), [ha cpf^{op}](#), [ha date contesto storico istituzionale^{op}](#)

is in range of

[ha contesto storico istituzionale^{op}](#)

Ontologia SAN. Classe date e sottoclasse dateComplarc

IRI: <http://san.beniculturali.it/SAN/date>

has super-classes

[owl:Thing](#)

has sub-classes

[date complesso archivistico^C](#), [date contesto storico istituzionale^C](#), [date esistenza^C](#), [date strumento^C](#), [estensione date complesso archivistico^C](#)

IRI: <http://san.beniculturali.it/SAN/dateComplarc>

has super-classes

[date^C](#)

has cardinality constrains

[date complesso archivistico normal^{dp} exactly 1](#)

[date complesso archivistico testo^{dp} exactly 1](#)

is in domain of

[date complesso archivistico normal^{dp}](#), [date complesso archivistico testo^{dp}](#)

is in range of

[ha date complesso archivistico^{op}](#)

Ontologia SAN. Sottoclasse dateCSI e sottoclasse dateEsistenza

IRI: <http://san.beniculturali.it/SAN/dateCSI>

has super-classes

[date^c](#)

has cardinality constrains

[date contesto storico istituzionale testo^{dp}](#) **exactly 1**

is in domain of

[date contesto storico istituzionale testo^{dp}](#)

is in range of

[ha date contesto storico istituzionale^{op}](#)

IRI: <http://san.beniculturali.it/SAN/dateEsistenza>

has super-classes

[date^c](#)

has cardinality constrains

[ha tipo date di esistenza^{op}](#) **exactly 1**

[date esistenza normal^{dp}](#) **exactly 1**

[date esistenza testo^{dp}](#) **exactly 1**

is in domain of

[date esistenza normal^{dp}](#), [date esistenza testo^{dp}](#), [ha tipo date di esistenza^{op}](#)

is in range of

[ha date esistenza^{op}](#)

Ontologia SAN. Sottoclasse dateStrumento e sottoclasse estensioneDateComplarc

IRI: <http://san.beniculturali.it/SAN/dateStrumento>

has super-classes

[date](#)^c

is in domain of

[date strumento testo](#)^{dp}

is in range of

[ha date strumento](#)^{op}

IRI: <http://san.beniculturali.it/SAN/estensioneDateComplarc>

has super-classes

[date](#)^c

has cardinality constrains

[estensione date normal](#)^{dp} **exactly** 1

is in domain of

[estensione date normal](#)^{dp}, [ha estensione date tipo](#)^{op}

is in range of

[ha estensione date complesso archivistico](#)^{op}

Ontologia SAN. Classe SoggettoProduttore

IRI: <http://san.beniculturali.it/SAN/produttore>

Soggetto (ente, famiglia o persona) che ha prodotto o acquisito per finalità amministrative, familiari o personali il complesso archivistico oggetto di descrizione. Può essere collegato ad altri soggetti produttori con relazioni gerarchiche o temporali.

is equivalent to

[ente^c](#) or [famiglia^c](#) or [persona^c](#)
[eac-cpf.entity](#)

has super-classes

[owl:Thing](#)

has cardinality constrains

[ha statusProvenienza^{op}](#) **exactly 1**
[abstract^{dp}](#) **max 1**
[forma autorizzata produttore^{dp}](#) **min 1**
[sistema provenienza^{dp}](#) **exactly 1**
[record provenienza id^{dp}](#) **exactly 1**
[scheda provenienza href^{fdp}](#) **exactly 1**

has sub-classes

[ente^c](#), [famiglia^c](#), [persona^c](#)

is in domain of

[abstract^{dp}](#), [altra denominazione produttore^{dp}](#), [forma autorizzata produttore^{dp}](#), [ha complesso archivistico^{op}](#), [ha contesto storico istituzionale^{op}](#), [ha date esistenza^{op}](#), [ha luogoProduttore^{op}](#), [ha qualificazioni relazioni Cpf^{op}](#), [ha statusProvenienza^{op}](#), [record provenienza id^{dp}](#), [scheda provenienza href^{fdp}](#), [sistema provenienza^{dp}](#)

is in range of

[ha cpf^{op}](#), [ha produttore^{op}](#), [ha relazione produttore^{op}](#)

Ontologia SAN. Sottoclasse Ente

IRI: <http://san.beniculturali.it/SAN/ente>

is equivalent to

[eac-cpf.corporateBoby](#)

has super-classes

[soggetto produttore^c](#)

is in domain of

[ha natura giuridica ente^{op}](#), [ha qualificazioni relazioni PI^{op}](#), [ha sottotipologia ente^{op}](#)

is in range of

[ha relazione^{op}](#)

is disjoint with

[famiglia^c](#), [persona^c](#)

Ontologia SAN. Sottoclassi famiglia e persona

IRI: <http://san.beniculturali.it/SAN/famiglia>

is equivalent to

[eac-cpf:family](#)

has super-classes

[soggetto produttore](#)^c

is disjoint with

[ente](#)^c, [persona](#)^c

IRI: <http://san.beniculturali.it/SAN/persona>

is equivalent to

[foaf:person](#)

[eac-cpf:person](#)

has super-classes

[soggetto produttore](#)^c

is in domain of

[professione](#)^{dp}, [titolo nobiliare](#)^{dp}

is disjoint with

[ente](#)^c, [famiglia](#)^c

Ontologia SAN. Classe luogo

IRI: <http://san.beniculturali.it/SAN/luogo>

is equivalent to

[gn:Feature](#)
[eac-cpf:place](#)

has super-classes

[owl:Thing](#)

has cardinality constrains

[luogo testo](#)^{dp} **exactly** 1

has sub-classes

[luogo conservatore](#)^c

is in domain of

[luogo testo](#)^{dp}

is in range of

[ha luogoProduttore](#)^{op}

Ontologia SAN. Sottoclasse luogoConservatore

IRI: <http://san.beniculturali.it/SAN/luogoConservatore>

has super-classes

[luogo^C](#)

has cardinality constrains

[luogo_provincia^{dp}](#) exactly 1

[luogo_paese^{dp}](#) exactly 1

[luogo_comune^{dp}](#) exactly 1

[luogo_CAP^{dp}](#) exactly 1

is in domain of

[luogo_CAP^{dp}](#), [luogo_comune^{dp}](#), [luogo_paese^{dp}](#), [luogo_provincia^{dp}](#)

is in range <http://san.beniculturali.it/SAN/luogoCAP>

[ha luogoConservatore^{op}](#)

Ontologia SAN. Classe profilo Istituzionale

IRI: <http://san.beniculturali.it/SAN/profiloIstituzionale>

Istituzioni, preunitarie o unitarie, descritte una volta per tutte, con riferimento al periodo in cui hanno operato e al contesto storico istituzionale o statutale di appartenenza.

Descrive una volta per tutte un particolare soggetto produttore, e costituisce una chiave di ricerca per tutti i soggetti produttori che appartengono al profilo descritto.

has super-classes

[owl:Thing](#)

has cardinality constrains

[ha statusProvenienza^{op}](#) **exactly 1**

[forma autorizzata profilo istituzionale^{dp}](#) **min 1**

[scheda provenienza href^{fdp}](#) **exactly 1**

[abstract^{dp}](#) **max 1**

[record provenienza id^{dp}](#) **exactly 1**

[sistema provenienza^{dp}](#) **exactly 1**

is in domain of

[altra denominazione profilo istituzionale^{dp}](#), [forma autorizzata profilo istituzionale^{dp}](#), [ha contesto storico istituzionale^{op}](#), [ha date esistenza^{op}](#), [ha natura giuridica ente^{op}](#), [ha qualificazioni relazioni PI^{op}](#), [ha sottotipologia ente^{op}](#), [ha statusProvenienza^{op}](#)

is in range of

[ha cpf^{op}](#), [ha relazione^{op}](#)

Ontologia SAN. Classi relazioni Cpf e relazioni PI

IRI: <http://san.beniculturali.it/SAN/qualificazioniRelazioniCpf>

has super-classes

[owl:Thing](#)

has cardinality constrains

[tipo qualificazioni relazioni Cpf^{dp}](#) **exactly** 1

[ha relazione produttore^{op}](#) **exactly** 1

is in domain of

[date qualificazioni relazioni Cpf^{dp}](#), [ha relazione produttore^{op}](#), [tipo qualificazioni relazioni Cpf^{dp}](#)

is in range of

[ha qualificazioni relazioni Cpf^{op}](#)

IRI: <http://san.beniculturali.it/SAN/qualificazioniRelazioniPI>

has super-classes

[owl:Thing](#)

has cardinality constrains

[tipo qualificazioni relazioni PI^{dp}](#) **exactly** 1

[ha relazione^{op}](#) **exactly** 1

is in domain of

[date qualificazioni relazioni Cpf^{dp}](#), [ha relazione^{op}](#), [tipo qualificazioni relazioni PI^{dp}](#)

is in range of

[ha qualificazioni relazioni PI^{op}](#)

Ontologia SAN. Classi sitoweb

IRI: <http://san.beniculturali.it/SAN/sitoWeb>

is equivalent to

[ocsa:contactArea](#)

has super-classes

[owl:Thing](#)

is in domain of

[sito web href^{dp}](#), [sito web testo^{dp}](#)

is in range of

[ha sito web^{op}](#)

Ontologia SAN. Classe soggetto conservatore

IRI: <http://san.beniculturali.it/SAN/conservatore>

Il soggetto sul quale ricade la responsabilità della conservazione e valorizzazione dell'archivio

is equivalent to

[org:Organization](#)

has super-classes

[owl:Thing](#)

has cardinality constrains

[servizi](#)^{dp} **max 1**

[consultazione](#)^{dp} **exactly 1**

[ha luogoConservatore](#)^{op} **min 1**

[forma autorizzata conservatore](#)^{dp} **min 1**

[tipologia](#) **exactly 1**

[acronimo](#)^{dp} **max 1**

[descrizione](#)^{dp} **max 1**

[codice ISIL](#)^{dp} **max 1**

is in domain of

[acronimo](#)^{dp}, [altro accesso](#)^{dp}, [codice ISIL](#)^{dp}, [consultazione](#)^{dp}, [descrizione](#)^{dp}, [forma autorizzata conservatore](#)^{dp}, [ha complesso archivistico](#)^{op}, [ha luogoConservatore](#)^{op}, [ha sito web](#)^{op}, [ha tipologia](#)^{op}, [indirizzo](#)^{dp}, [orario](#)^{dp}, [scheda provenienza href](#)^{dp}, [servizi](#)^{dp}

is in range of

[ha conservatore](#)^{op}

Ontologia SAN. Classe strumento

IRI: <http://san.beniculturali.it/SAN/strumento>

Repertorio - in forma di inventario analitico o sommario, elenco, elenco di versamento, elenco di consistenza, elenco numerico o altro - che descrive in maniera più o meno dettagliata un fondo o parte di esso.

is equivalent to

[oad:findingAid](#)

has super-classes

[owl:Thing](#)

has cardinality constrains

[scheda provenienza href](#)^{dp} **exactly** 1

[titolo](#)^{dp} **exactly** 1

[sistema provenienza](#)^{dp} **exactly** 1

[descrizione](#)^{dp} **max** 1

[autore](#)^{dp} **max** 1

[record provenienza id](#)^{dp} **exactly** 1

is in domain of

[autore](#)^{dp}, [casa editrice](#)^{dp}, [descrizione](#)^{dp}, [edito](#)^{dp}, [ha complesso archivistico](#)^{op}, [ha date strumento](#)^{op}, [luogo pubblicazione](#)^{dp}, [record provenienza id](#)^{dp}, [scheda provenienza href](#)^{dp}, [sistema provenienza](#)^{dp}, [strumento digitale link](#)^{dp}, [strumento digitale titolo](#)^{dp}, [titolo](#)^{dp}

is in range of

[ha strumento di ricerca](#)^{op}

Data Properties definite nell'ontologia SAN

abstract	date qualificazioni relazioni Cpf	luogo pubblicazione
acronimo	date strumento testo	luogo testo
altra denominazione complesso archivistico	denominazione contesto storico istituzionale	nome livello
altra denominazione produttore	descrizione	orario
altra denominazione profilo istituzionale	descrizione contesto storico istituzionale	ordinamento
altro accesso	edito	professione
autore	estensione date normal	record provenienza id
casa editrice	forma autorizzata complesso archivistico	scheda provenienza href
codice ISIL	forma autorizzata conservatore	servizi
consistenza	forma autorizzata produttore	sistema provenienza
consultazione	forma autorizzata profilo istituzionale	sito web href
date complesso archivistico normal	Indirizzo	sito web testo
date complesso archivistico testo	luogo CAP	strumento digitale link
date contesto storico istituzionale testo	luogo comune	strumento digitale titolo
date esistenza normal	luogo paese	tipo qualificazioni relazioni Cpf
date esistenza testo	luogo provincia	tipo qualificazioni relazioni PI
date qualificazioni relazioni Cpf		tipologia complesso
		titolo
		titolo nobile

Esempio - Bureau de conservation des hypothèques de Turin

```
<owl:NamedIndividual rdf:about="http://san.beniculturali.it/SAN/complarc_GGASI_GG0840052147">
  <rdf:type rdf:resource="http://san.beniculturali.it/SAN/complessoArchivistico"/>
  <rdfs:label rdf:datatype="string">Bureau de conservation des hypothèques de Turin</rdfs:label>
  <SAN:formaAutorizzataComplarc rdf:datatype="string">
    Bureau de conservation des hypothèques de Turin </SAN:formaAutorizzataComplarc>
  <SAN:abstract rdf:datatype="string">
    Costituito con il decreto del 5 termidoro a. IX (241u. 1801) che estendeva al Piemonte la legge
    11 brumaio a. VII (1 nov. 1798) </SAN:abstract>
  <SAN:recordProvenienzaId rdf:datatype="string">GG0840052147</SAN:recordProvenienzaId>
  <SAN:sistemaProvenienza rdf:datatype="string">GGASI</SAN:sistemaProvenienza>
  <SAN:tipologiaComplesso rdf:datatype="string">fondi</SAN:tipologiaComplesso>
  <SAN:schedaProvenienzaHref rdf:datatype="anyURI">
    http://purl.oclc.org/NET/GGASI/EAD/GG0840052147 </SAN:schedaProvenienzaHref>
  <SAN:consistenza rdf:datatype="string">regg. 197</SAN:consistenza>
  <dc:description>
    Costituito con il decreto del 5 termidoro a. IX (241u. 1801) che estendeva al Piemonte la legge
    11 brumaio a. VII (1 nov. 1798) </dc:description>
  <dc:date>(1798-1814)</dc:date>
  <dc:title>Bureau de conservation des hypothèques de Turin</dc:title>
  <SAN:has_livelloSuperiore
    rdf:resource="http://san.beniculturali.it/SAN/complarc_GGASI_GG0840052147"/>
  <SAN:has_conservatore rdf:resource="http://san.beniculturali.it/SAN/conservatore_Astorino"/>
  <SAN:has_dateComplarc rdf:resource="http://san.beniculturali.it/SAN/dateComplarc_1798-1814"/>
  <SAN:has_prodotto rdf:resource="http://san.beniculturali.it/SAN/prodotto_GGASI_SP038560"/>
  <SAN:has_statusProvenienza rdf:resource="http://san.beniculturali.it/SAN/TesoroSAN_1.0/scheda_publicata"/>
</owl:NamedIndividual>
```


Esempio – Luogo TORINO

```
<owl:NamedIndividual rdf:about="http://san.benculturali.it/SAN/Torino">
  <rdf:type rdf:resource="http://san.benculturali.it/SAN/luogo"/>
  <rdf:type rdf:resource="http://san.benculturali.it/SAN/luogoConservatore"/>
  <rdfs:label rdf:datatype="string">Torino</rdfs:label>
  <SAN:luogoCAP>10100</SAN:luogoCAP>
  <SAN:luogoPaese rdf:datatype="string">Italia</SAN:luogoPaese>
  <SAN:luogoProvincia rdf:datatype="string">TO</SAN:luogoProvincia>
  <SAN:luogoComune rdf:datatype="string">Torino</SAN:luogoComune>
  <SAN:luogoTesto rdf:datatype="string">Torino</SAN:luogoTesto>
  <dc:title>Torino</dc:title>
  <owl:sameAs rdf:resource="http://sws.geonames.org/3165525"/>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/Turin"/>
</owl:NamedIndividual>
```

Esempio – Soggetto conservatore Archivio di Stato di Palermo

```
<owl:NamedIndividual rdf:about="http://san.beniculturali.it/SAN/conservatore_ASpalermo">
<rdf:type rdf:resource="http://san.beniculturali.it/SAN/conservatore"/>
<SAN:schedaProvenienzaHref rdf:datatype="anyURI">http://www.archivi-sias.it/scheda_istituto.asp?Filtralstituto=70010000
</SAN:schedaProvenienzaHref>
<SAN:schedaProvenienzaHref rdf:datatype="anyURI">http://purl.oclc.org/NET/GGASI/SCONS/IC0570030601
</SAN:schedaProvenienzaHref>
<rdfs:label rdf:datatype="string">Archivio di Stato di Palermo</rdfs:label>
  <SAN:formaAutorizzataConservatore rdf:datatype="string">Archivio di Stato di Palermo
</SAN:formaAutorizzataConservatore>
<SAN:orario rdf:datatype="string">lun./ven. 08.00 - 18.30; sab. 08.00 - 13.30</SAN:orario>
<SAN:consultazione rdf:datatype="boolean">>true</SAN:consultazione>
  <dc:title>Archivio di Stato di Palermo</dc:title>
<SAN:descrizione rdf:datatype="string"> - - - - </SAN:descrizione>
<dc:description rdf:datatype="string"> - - - - </dc:description>

<SAN:indirizzo>Corso Vittorio Emanuele 31 - 90139, Palermo (Palermo)</SAN:indirizzo>
<SAN:has_luogoConservatore rdf:resource="http://san.beniculturali.it/SAN/Palermo"/>
<SAN:has_sitoWeb rdf:resource="http://san.beniculturali.it/SAN/sitoWeb_www.archiviodistatodipalermo.it"/>
<SAN:has_sitoWeb
rdf:resource="http://san.beniculturali.it/SAN/sitoWeb_www.icar.beniculturali.it/mw141/index.php?it/167/repertori-archivistici-non-
editi-degli-archivi-di-stato"/>
<SAN:has_tipologia rdf:resource="http://san.beniculturali.it/SAN/TesaurusSAN_1.0/archivio_di_Stato"/>
<SAN:has_complarc rdf:resource="complarc_SIAS_IT-ASPA-F70010325"/>
- - - - -
<owl:sameAs rdf:resource="http://spcdata.digitpa.gov.it/UfficioProtocolloAOO/m_bac-mbac-as-pa"/>
</owl:NamedIndividual>
```

1 Sistema Archivistico Nazionale

2 Progetto Linked Open Data

3 Struttura dei dati CAT-SAN

4 Disegno di Ontologie

5 Ontologia SAN

6 SKOS e Thesaurus SAN

7 Openrefine

8 Conclusioni

Il Thesaurus del Sistema Archivistico Nazionale

Il Thesaurus del Sistema Archivistico Nazionale è costituito dai diversi vocabolari controllati presenti negli schema CAT SAN come supporto ad una descrizione normalizzata delle diverse risorse - istituti di conservazione, complessi archivistici, soggetti produttori - provenienti dai diversi sistemi aderenti al SAN.

La decisione di dare ai vocabolari controllati una propria definizione autonoma come schema di classificazione SKOS, offerto in un formato aperto, flessibile e facilmente implementabile, si inserisce in una logica di produzione di ausili e strumenti per la standardizzazione delle voci descrittive che l'ICAR intende perseguire a vantaggio della comunità archivistica.

Vocabolari controllati adottati nel Sistema Archivistico Nazionale – SAN

- **Tipologia della qualifica associata al soggetto produttore**
- **Tipologia del luogo associato al soggetto produttore**
- **Tipologia della data di esistenza del soggetto produttore**
- **Tipologia del soggetto conservatore**
- **Tematismo**
- **Status della scheda nel sistema di provenienza**
- **Estensione alla data del complesso archivistico**

SKOS - Simple Knowledge Organization System

Skos è un modello dati per sistemi di organizzazione della conoscenza (thesauri, schemi di classificazione, tassonomie) .

I dati SKOS sono espressi come triple RDF e può essere esteso con altri vocabolari RDF come DCMI Metadata Terms e FOAF (Friend of a Friend).

SKOS possiede una serie di proprietà raggruppate in famiglie:

- Properties per le etichette lessicali
- Properties per le etichette per la documentazione
- Properties per le relazioni semantiche
- Properties per le etichette simboliche

Le properties all'interno di queste famiglie sono organizzate in maniera gerarchica in modo che sia possibile estendere quella più appropriata per le proprie necessità.

Il sito di riferimento è <http://www.w3.org/TR/skos-reference/>

SKOS – Classe Concept

SKOS considera "Concept"; l'unità fondamentale di qualunque concept-schema dalla quale modellare una determinata risorsa in maniera da esprimere questa come un concetto.

Un concetto può essere immaginato come un qualcosa che può essere definito o descritto e può avere una sola descrizione o meglio un solo "preferred term"; mentre può avere illimitate descrizioni alternative o "alternative tags";.

Il modo più semplice per definire un concetto in RDF è assegnargli un URI e utilizzare la forma skos:Concept:

```
<rdf:Description rdf:about="http://san.beniculturali.it/SAN/TesauroSAN_1.0/professione">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:prefLabel xml:lang="it">Professione del soggetto produttore persona</skos:prefLabel>
</rdf:Description>
```

SKOS – Etichette lessicali

Le etichette lessicali servono a denotare le risorse utilizzando il linguaggio comune.

- `skos:prefLabel`, indica il termine preferito o più comunemente usato
- `skos:altLabel`, indica un sinonimo del termine preferito
- `skos:hiddenLabel`: un etichetta "nascosta"; che normalmente non è visibile ma che permette l'accesso al contenuto ad applicazioni che fanno ricerche basate sul testo.
- `skos:prefSymbol`: un etichetta che permette di includere un simbolo o un'immagine di una certa risorsa
- `skos:altSymbol`: un simbolo o un immagine alternativa per una risorsa

E' possibile inserire etichette multilingua con l'attributo `xml:lang` come ad esempio in: `<skos:prefLabel xml:lang="it">`

SKOS – Proprietà descrittive

Esistono in skos 7 proprietà che si possono usare per aggiungere dell'informazione per la descrizione di un concetto.

Queste proprietà hanno come superclasse `skos:note`: essa può essere usata per fornire una qualunque tipo di descrizione utile a qualsiasi scopo.

Le sottoclassi invece sono:

- `skos:definition`: utile per dare una spiegazione completa e precisa della risorsa;.
- `skos:scopeNote`: utilizzato restringere od aumentare la descrizione de concetto;
- `skos:example`: utilizzato per fornire degli esempi dell'uso del termine o del concetto;
- `skos:historyNote`: utilizzato per indicare significativi cambiamenti di significato;
- `skos:editorialNote`: utilizzato per fornire informazioni di tipo amministrativo;
- `skos:changeNote`: utilizzato per tener traccia delle varie modifiche occorse;

SKOS – Relazioni semantiche

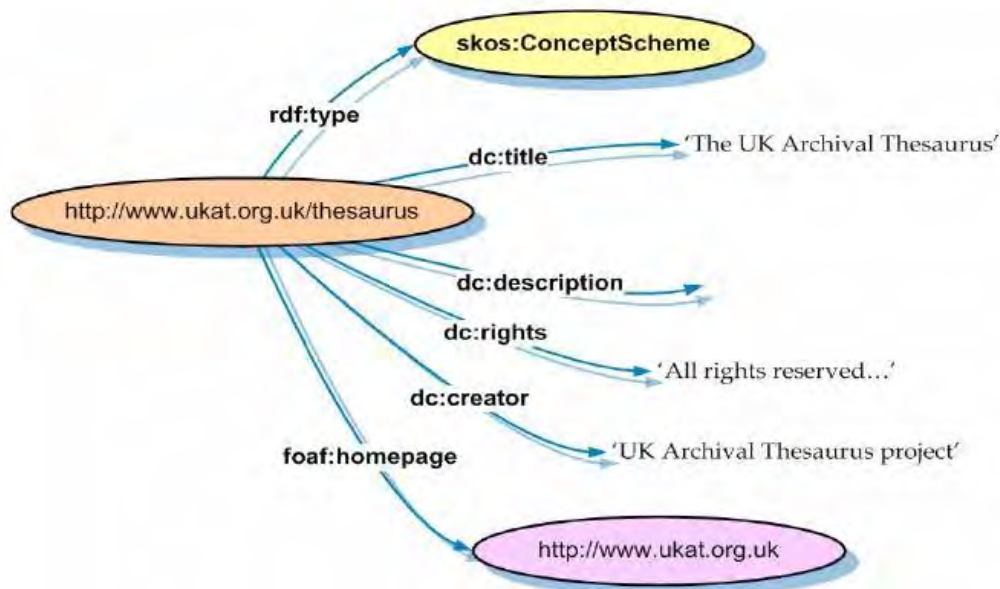
Le relazioni semantiche di SKOS permettono di costruire cosiddetti "concept schema"; cioè quelle strutture come thesauri, vocabolari, tassonomie che integrano all'interno della propria struttura un determinato schema con delle relazioni semantiche per classificare, ordinare, regolare diversi concetti.

Le proprietà di SKOS che permettono di definire relazioni semantiche tra i concetti, sono :

- skos:broader, per indicare che ha un significato più generale di un altro
- skos:narrower per indicare ha un significato più ristretto di un altro ed è una proprietà duale rispetto a skos:broader
- skos:related. si usa per dire che un determinato concetto ha una qualche relazione con un altro.

SKOS – Concept Scheme

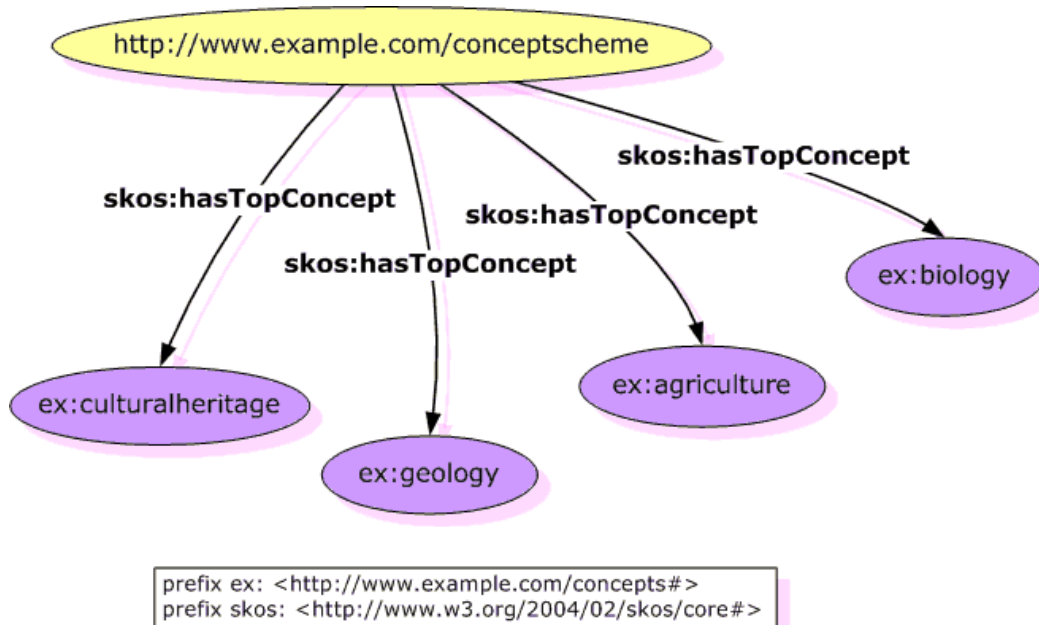
la classe `skos:Concept Scheme` permette di definire un concept-schema personalizzato, basato sulle nostre necessità e che, grazie al meccanismo degli URI, potremmo usare insieme con altre strutture già esistenti per descrivere un certo dominio essendo sicuri di non portare nessuna ambiguità.



Ecco presentato qui sopra un esempio di thesaurus che utilizza delle meta properties di DUBLIN CORE e FOAF per aggiungere dei metadati riguardanti lo schema: il titolo, l'autore e la descrizione.

SKOS – Proprietà `skos:hasTopConcept` e `skos:inScheme`

La proprietà `skos:inScheme` permette di dichiarare se un certo termine appartiene ad un particolare schema e ci permette di gestire entità esterne a se stanti.



La proprietà `skos:hasTopConcept` è molto importante in quanto, creando un collegamento tra concept schema e i vari concetti, permette di identificare in maniera immediata i termini padre di un dato termine e questo può essere molto utile in fase di consultazione o di ricerca nel thesauro.

SKOS: classe collection

La classe `skos:Collection` permette di modellare quelle particolari tipologie di categorie che organizzano i concetti tramite faccette (dall'inglese "facet") ed attraverso la proprietà `skos:member` è possibile aggiungere i vari componenti della schiera al nodo principale che descrive la categoria.

In alcune domini l'ordine con cui sono rappresentati i concetti è significativo, quando ad esempio vogliamo rispettare l'ordine alfabetico. La sottoclasse `skos:CollectionOrdered` attraverso la proprietà `skos:memberList` permette di definire collezioni di concetti che rispettano un certo ordine.

SKOS : subject indexing

Il processo di subject indexing o indicizzazione per soggetto, consiste in un'analisi ragionata del contenuto del documento al fine di identificare i concetti in esso presenti, ed assegnare loro una qualche etichetta testuale, per permettere successivamente un rapido ritrovamento dell'informazione per un uso futuro.

Subject Indexing properties è in insieme di proprietà messe a disposizione da SKOS che permette l'estensione della proprietà dc:subject di Dublin Core per definire relazioni tra risorse e concetti.

Le proprietà messe a disposizione da SKOS per il subject indexing sono:

- skos:subject
- skos:isSubjectOf
- skos:primarySubject
- skos:isPrimarySubjectOf.

Skos:subject è una sub-property di dc:subject. Il significato è lo stesso ma mentre dc:subject si utilizza con qualsiasi tipo di risorsa, il campo di azione di skos:subject è ristretto alla classe Concept. Essa è quindi utilizzata per dire che un certo concetto è il soggetto di una determinata risorsa.

Il thesaurus del Sistema Archivistico Nazionale: esempi in rdf

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:xsd="
http://www.w3.org/2001/XMLSchema#" xmlns:SAN="http://san.beniculturali.it/SAN" xmlns:dc="http://purl.org/dc/terms/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xml:base="
http://san.beniculturali.it/SAN">
  <rdf:Description rdf:about="http://san.beniculturali.it/SAN/TesauroSAN_1.0">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#ConceptScheme"/>
    <dc:creator>ICAR-MAAS</dc:creator>
    <rdfs:label xml:lang="it">Schema concettuale dei vocabolari controllati SAN</rdfs:label>
    <dc:title>Schema concettuale dei vocabolari controllati SAN</dc:title>
    <dc:description xml:lang="it">Il Tesauro del Sistema Archivistico Nazionale &#232; costituito dai diversi vocabolari controllati
    presenti negli schema CAT SAN come supporto ad una descrizione normalizzata delle diverse risorse - istituti di conservazione,
    complessi archivistici, soggetti produttori - provenienti dai diversi sistemi aderenti al SAN. Per sua stessa natura, il "Tesauro SAN"
    si presenta pertanto come uno strumento di classificazione aperto ad ulteriori evoluzioni ed integrabile con altri sistemi di
    organizzazione della conoscenza nazionali ed internazionali.</dc:description>
    <skos:hasTopConcept rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Vocabolari_controllatiSAN"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Vocabolari_controllatiSAN">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:topConceptOf rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0"/>
    <skos:prefLabel xml:lang="it">Vocabolari controllati adottati nel Sistema Archivistico Nazionale - SAN</skos:prefLabel>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Tipo_qualifica_CPF"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Tipo_luogo_CPF"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Tipo_data_esistenza_CPF"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/tipologia_Conservatore"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Tematismo"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Status_scheda_provenienza"/>
    <skos:narrower rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/Estensione_data_complesso"/>
  </rdf:Description>

```

Definizioni
generali

Elementi principali
del thesaurus

Il thesaurus del Sistema Archivistico Nazionale: esempi in rdf

```

<skos:broader rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/non_statali"/>
</rdf:Description>
<rdf:Description rdf:about="http://san.beniculturali.it/SAN/TesauroSAN_1.0/ente_di_culto_acattolico-associazione_acattolica">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0"/>
  <skos:prefLabel xml:lang="it">ente di culto acattolico/associazione acattolica</skos:prefLabel>
  <skos:scopeNote xml:lang="it">Sono raggruppati in questa voce gli enti relativi al culto o a funzioni religiose non cattoliche: ebraiche, valdesi, metodiste, etc. La costituzione repubblicana riconferma liberamente la propria fede religiosa in qualsiasi forma, individuale o associata, e di seguirne in pieno e pacatamente le conseguenze; non contrari al buon costume.</skos:scopeNote>
  <skos:broader rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/non_statali"/>
</rdf:Description>
<rdf:Description rdf:about="http://san.beniculturali.it/SAN/TesauroSAN_1.0/ente_culto_cattolico-associazione_cattolica">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0"/>
  <skos:prefLabel xml:lang="it">ente di culto cattolico/associazione cattolica</skos:prefLabel>
  <skos:scopeNote xml:lang="it">Sono raggruppati in questa voce gli enti ecclesiastici, quali diocesi, chiese, abbazie, parrocchie, mense, capitoli, tribunali ecclesiastici o seminari che, nel caso di fonti conservate presso gli Archivi di Stato, vi sono pervenute, di massima, frammentate agli archivi delle corporazioni religiose sopresse o di tribunali dell'antico regime. Gli enti ecclesiastici e le associazioni cattoliche conservano i propri archivi che, se estinti, possono confluire negli archivi di altri enti ecclesiastici o presso altre istituzioni. Per un proprio regolamento gli Archivi storici diocesani.</skos:scopeNote>
  <skos:broader rdf:resource="http://san.beniculturali.it/SAN/TesauroSAN_1.0/non_statali"/>
</rdf:Description>

```

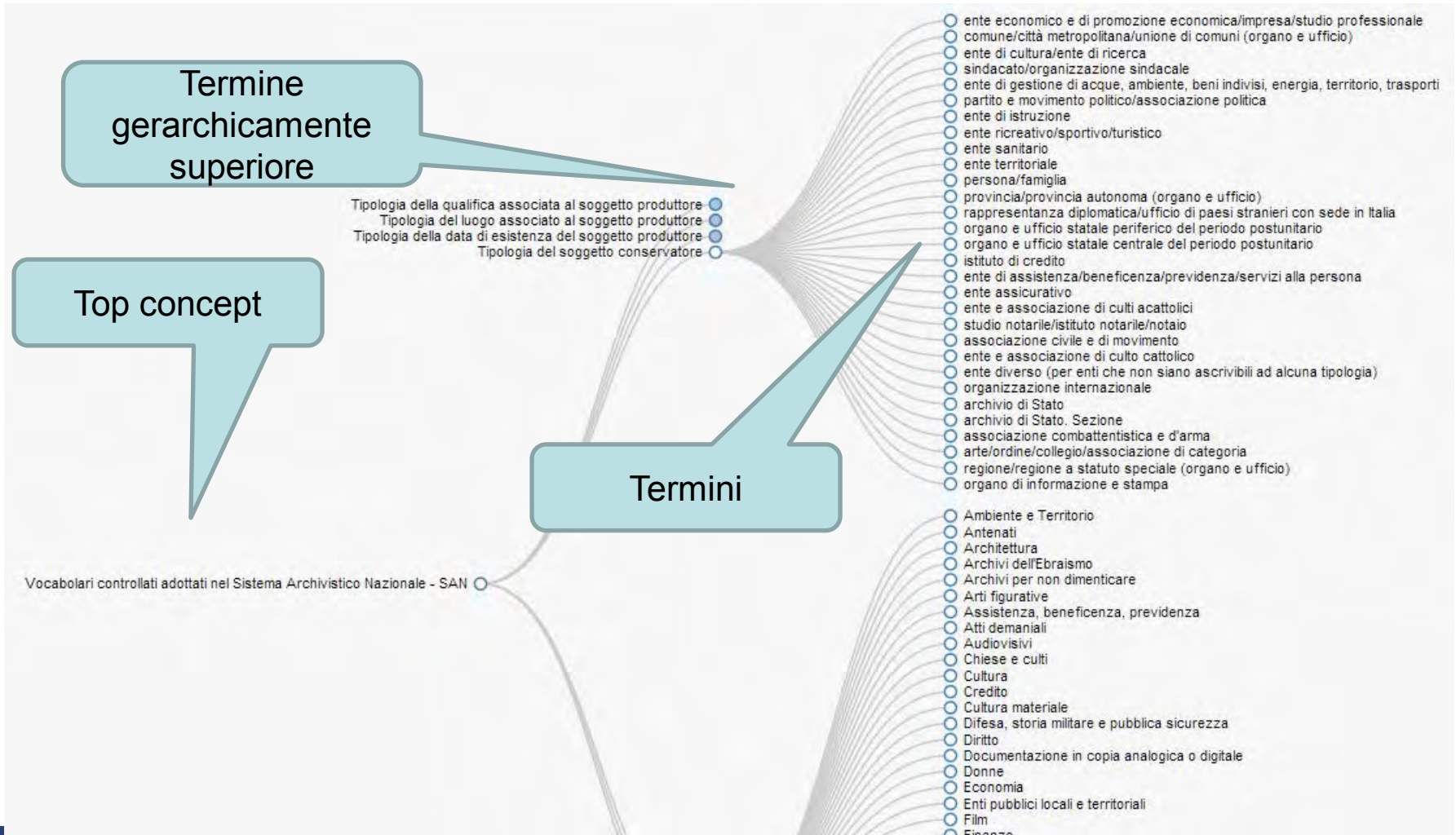
Termine gerarchicamente superiore

Concetto e riferimento

Descrizione estesa del concetto

Descrizione breve

Il thesaurus del Sistema Archivistico Nazionale: esempi



Il thesaurus del Sistema Archivistico Nazionale: esempi



1 Sistema Archivistico Nazionale

2 Progetto Linked Open Data

3 Struttura dei dati CAT-SAN

4 Disegno di Ontologie

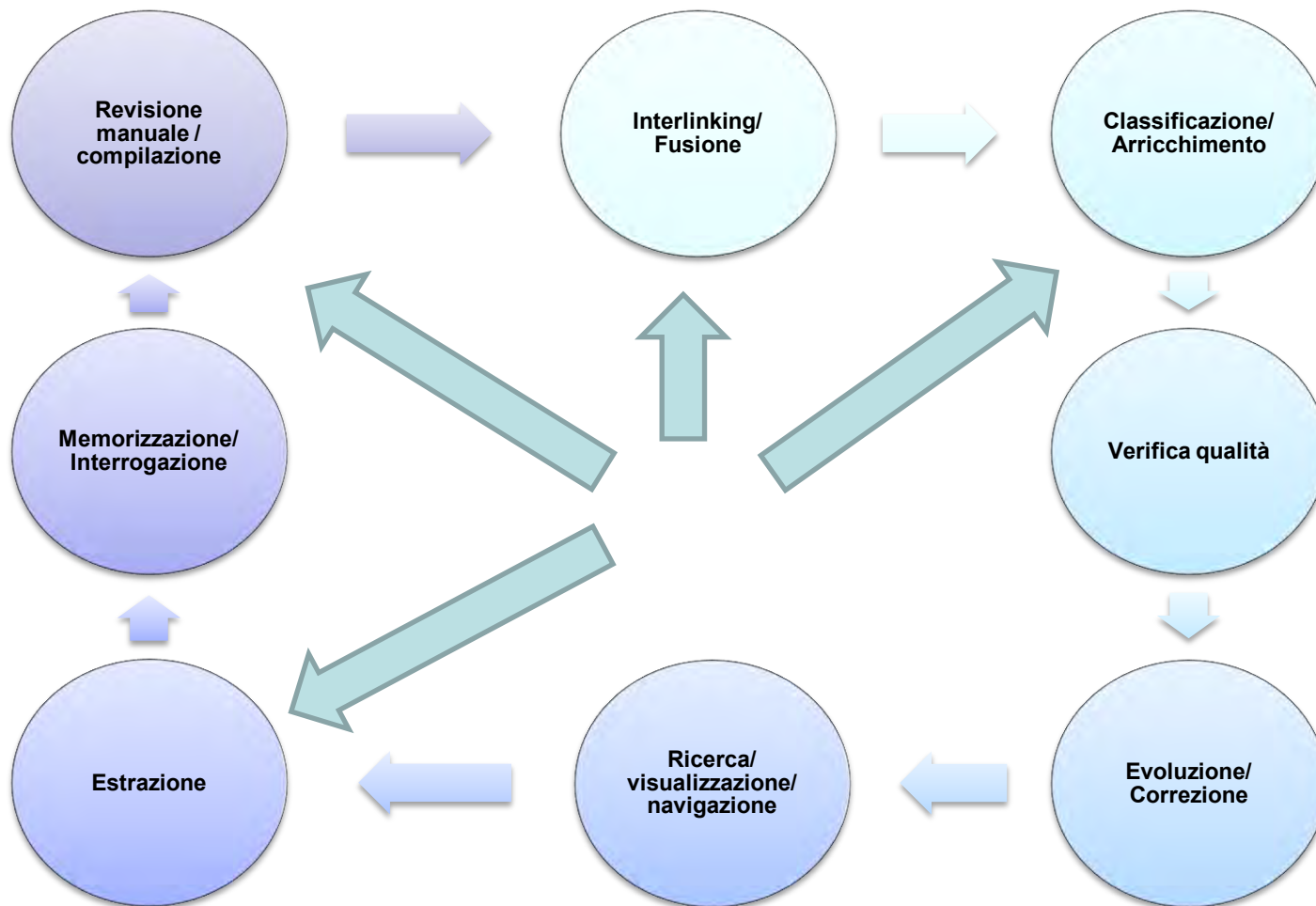
5 Ontologia SAN

6 SKOS e Thesaurus SAN

7 Openrefine

8 Conclusioni

Ciclo di vita dei Linked Data



Cit: lod2.eu

Ciclo di vita dei Linked Data

Per consentire la pubblicazione di determinati dati e metadati è fondamentale perseguire un appropriato livello di qualità e di minimi errori.

Questo vuol dire:

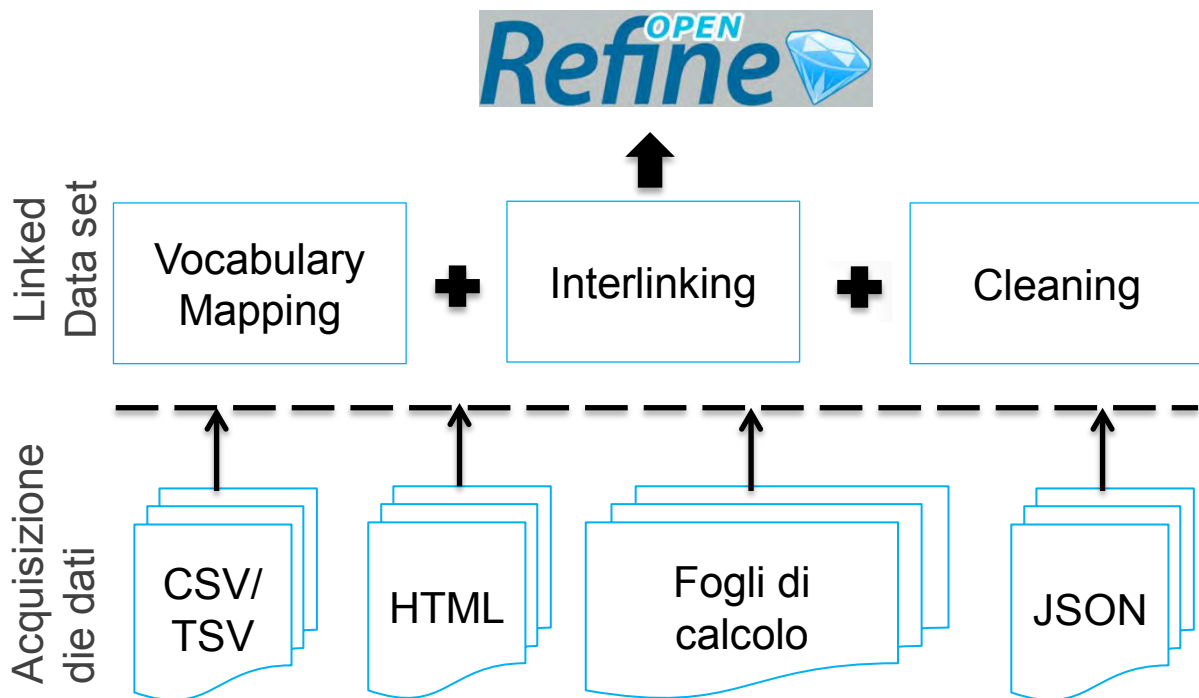
- Correzione degli errori.
- Trasformazione e omogeneizzazione formati.
- Allineamento incongruenze nei dati e metadati.
- Rimozione di informazioni duplicate / ridondanti.
- Aggiunta delle informazioni mancanti.
- Sicurezza che le informazione siano aggiornate.



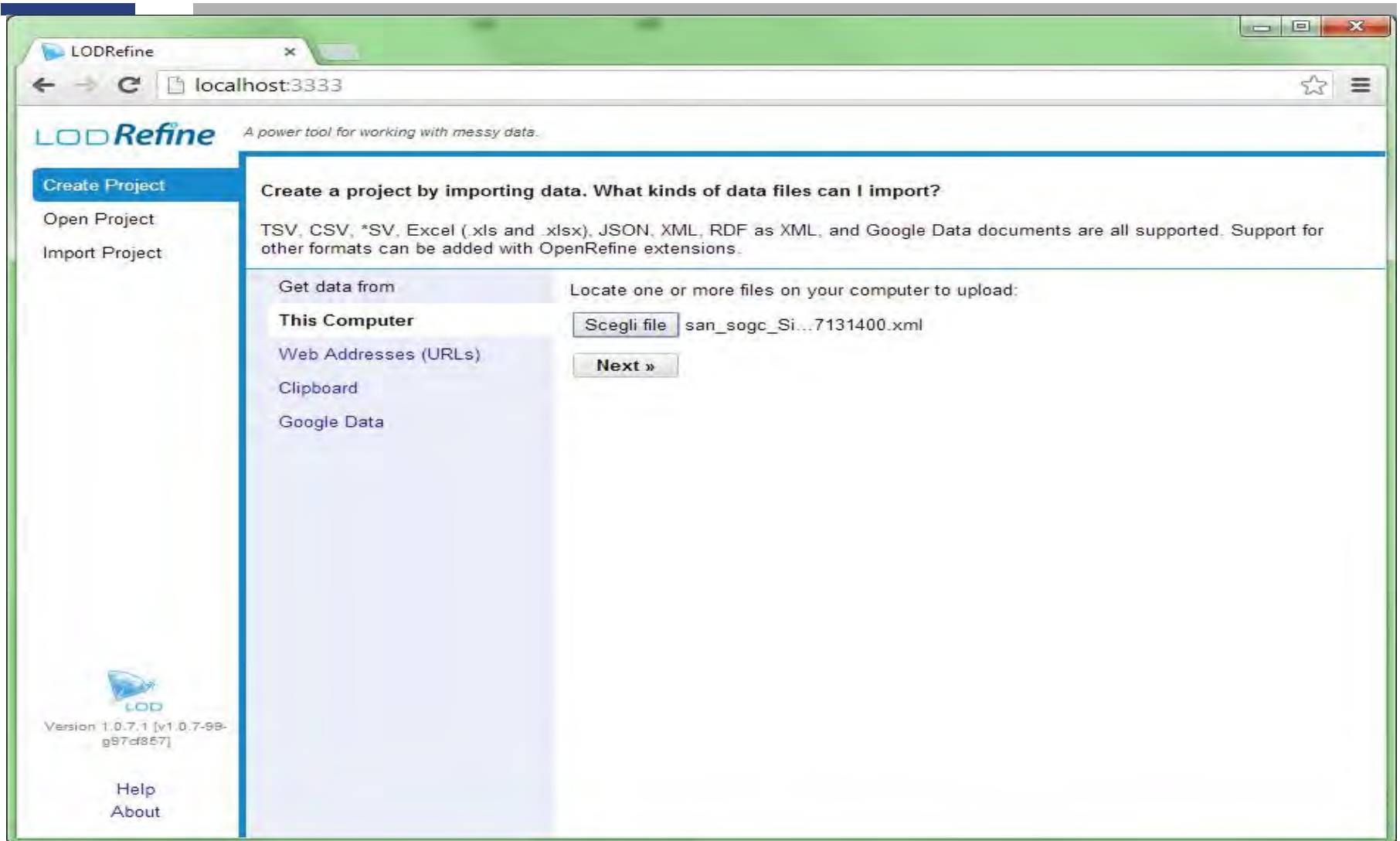
OpenRefine

E' uno strumento open source per la gestione di dati disorganizzati, per la trasformazione e la manipolazione di formati.

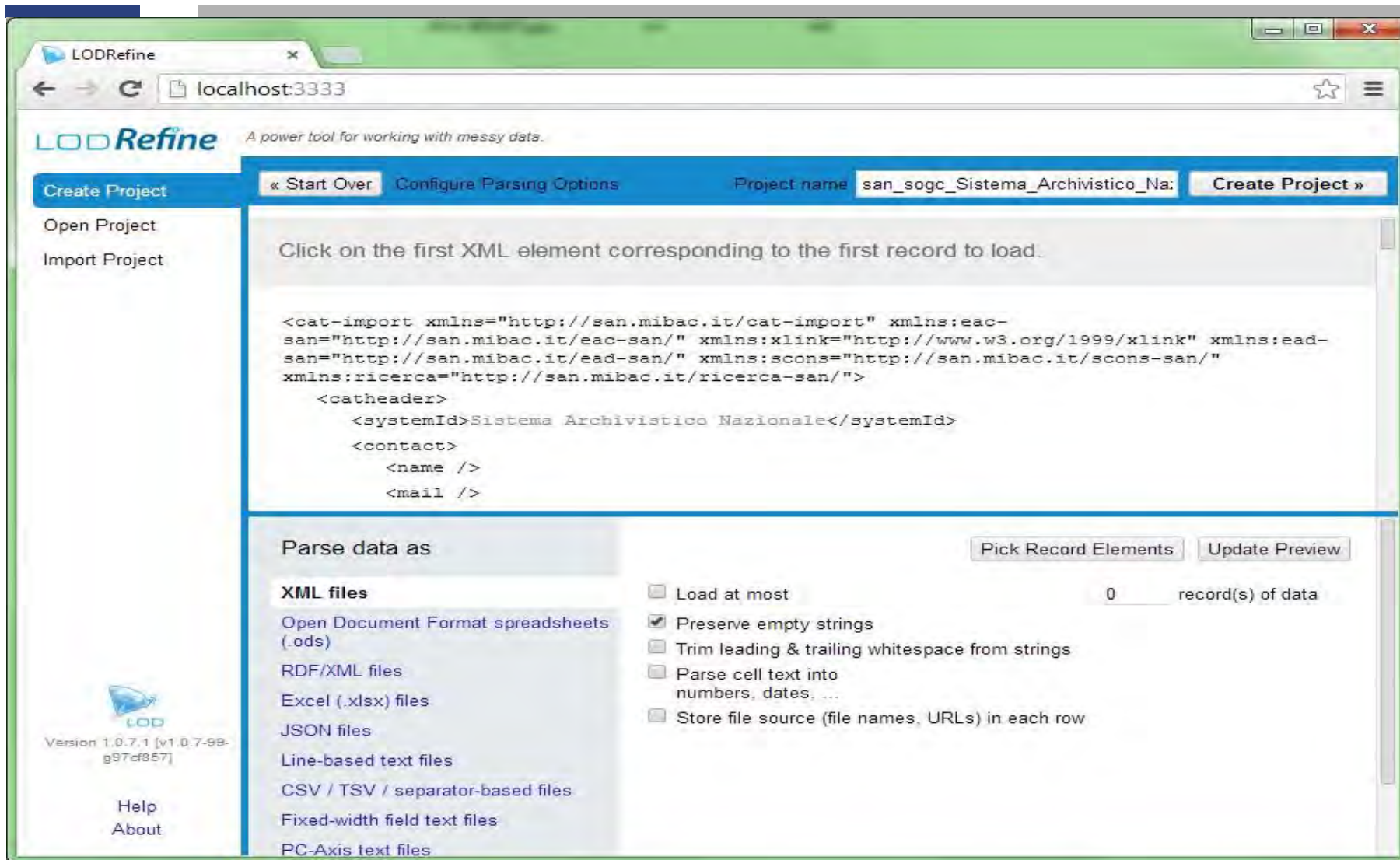
Permette la riconciliazione di entità attraverso endpoint SPARQL o dati RDF e può essere esteso con plugin per aumentarne le funzionalità



<http://openrefine.org/>



The screenshot shows a web browser window with the URL `localhost:3333`. The page title is "LODRefine" with the tagline "A power tool for working with messy data.". On the left, there is a navigation menu with "Create Project" selected, and other options: "Open Project" and "Import Project". The main content area is titled "Create a project by importing data. What kinds of data files can I import?". Below this, it lists supported formats: "TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions." There are four options for "Get data from": "This Computer", "Web Addresses (URLs)", "Clipboard", and "Google Data". The "This Computer" option is selected, and a file selection dialog is open, showing a file named "san_sogc_Si...7131400.xml" and a "Next »" button. At the bottom left, there is a version number: "Version 1.0.7.1 [v1.0.7-99-g97cf857]" and links for "Help" and "About".



The screenshot shows the LODRefine web application running in a browser. The browser address bar shows 'localhost:3333'. The application header includes the title 'LODRefine' and the tagline 'A power tool for working with messy data.'. Below the header is a navigation bar with buttons for 'Create Project', 'Start Over', 'Configure Parsing Options', 'Project name' (with the value 'san_sogc_Sistema_Archivistico_Na.'), and 'Create Project'. A sidebar on the left contains links for 'Open Project' and 'Import Project'. The main content area displays an instruction: 'Click on the first XML element corresponding to the first record to load.' followed by a snippet of XML code. Below this is a 'Parse data as' section with a list of file formats and a set of checkboxes for parsing options. The 'Pick Record Elements' button shows '0 record(s) of data'. The bottom left corner of the application shows the version '1.0.7.1 [v1.0.7-99-g87cf857]' and links for 'Help' and 'About'.

LODRefine A power tool for working with messy data.

Create Project « Start Over Configure Parsing Options Project name san_sogc_Sistema_Archivistico_Na. Create Project »

Open Project
Import Project

Click on the first XML element corresponding to the first record to load.

```
<cat-import xmlns="http://san.mibac.it/cat-import" xmlns:eac-san="http://san.mibac.it/eac-san/" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:ead-san="http://san.mibac.it/ead-san/" xmlns:scons="http://san.mibac.it/scons-san/" xmlns:ricerca="http://san.mibac.it/ricerca-san/">
  <catheader>
    <systemId>Sistema Archivistico Nazionale</systemId>
    <contact>
      <name />
      <mail />
    </contact>
  </catheader>

```

Parse data as Pick Record Elements Update Preview

XML files Load at most 0 record(s) of data

Open Document Format spreadsheets (.ods) Preserve empty strings

RDF/XML files Trim leading & trailing whitespace from strings

Excel (.xlsx) files Parse cell text into numbers, dates, ...

JSON files Store file source (file names, URLs) in each row

Line-based text files

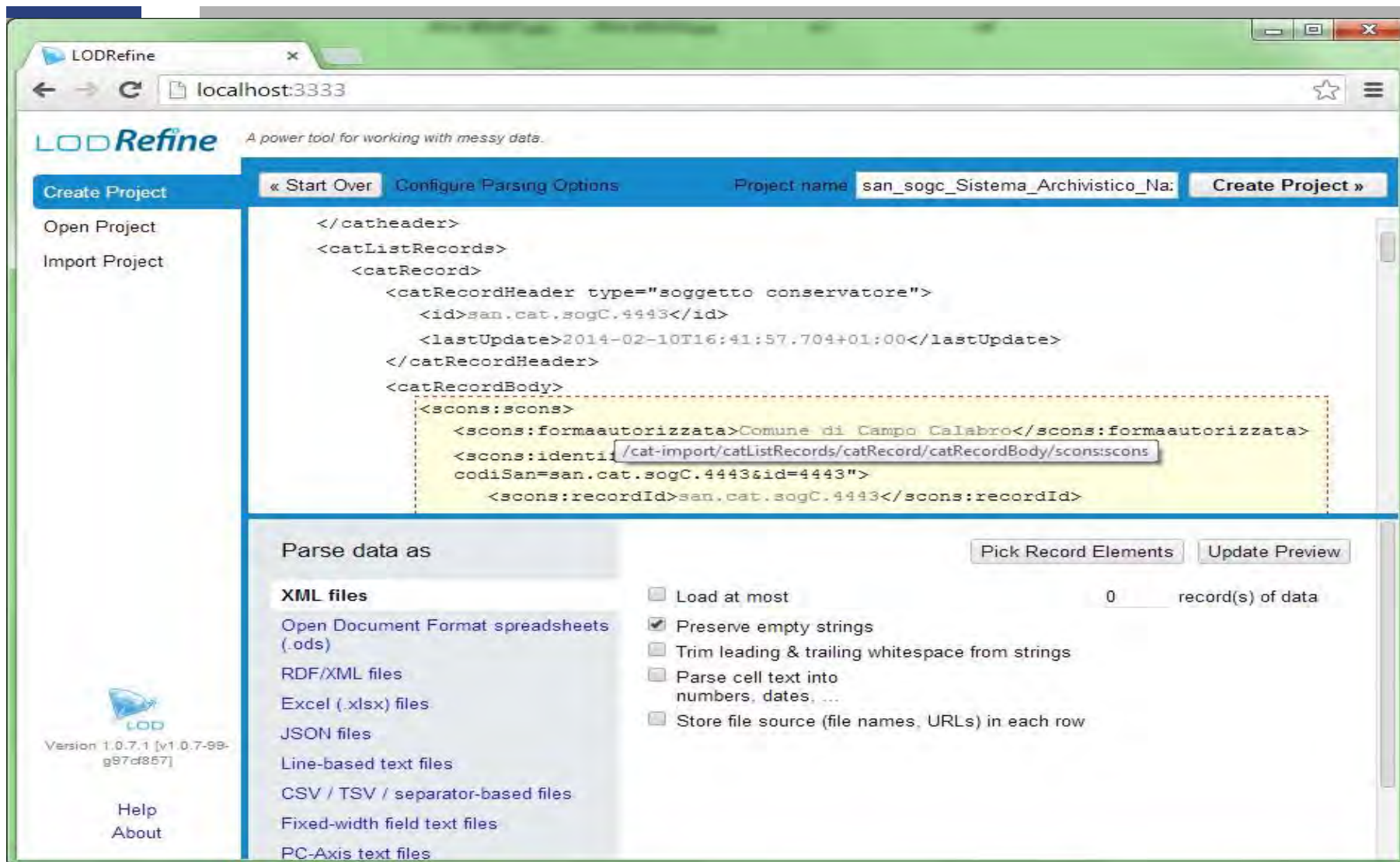
CSV / TSV / separator-based files

Fixed-width field text files

PC-Axis text files

Version 1.0.7.1 [v1.0.7-99-g87cf857]

Help
About



The screenshot shows the LODRefine web application in a browser window. The browser address bar shows 'localhost:3333'. The application header includes the title 'LODRefine' and the tagline 'A power tool for working with messy data.'. Below the header is a navigation bar with buttons for 'Create Project', 'Start Over', 'Configure Parsing Options', 'Project name' (set to 'san_sogc_Sistema_Archivistico_Na:'), and 'Create Project'. A sidebar on the left contains 'Open Project' and 'Import Project' options. The main content area displays XML data for a record, with a dashed box highlighting a specific section:


```

    <catRecordHeader type="soggetto conservatore">
      <id>san.cat.sogC.4443</id>
      <lastUpdate>2014-02-10T16:41:57.704+01:00</lastUpdate>
    </catRecordHeader>
    <catRecordBody>
      <scons:scons>
        <scons:formaautorizzata>Comune di Campo Calabro</scons:formaautorizzata>
        <scons:identif /cat-import/catListRecords/catRecord/catRecordBody/scons:scons
        codiSan=san.cat.sogC.4443&id=4443">
          <scons:recordId>san.cat.sogC.4443</scons:recordId>
      </scons:scons>
    </catRecordBody>
    
```

 Below the XML is a 'Parse data as' section with a list of file formats and checkboxes for parsing options. The 'Pick Record Elements' button shows '0 record(s) of data'.

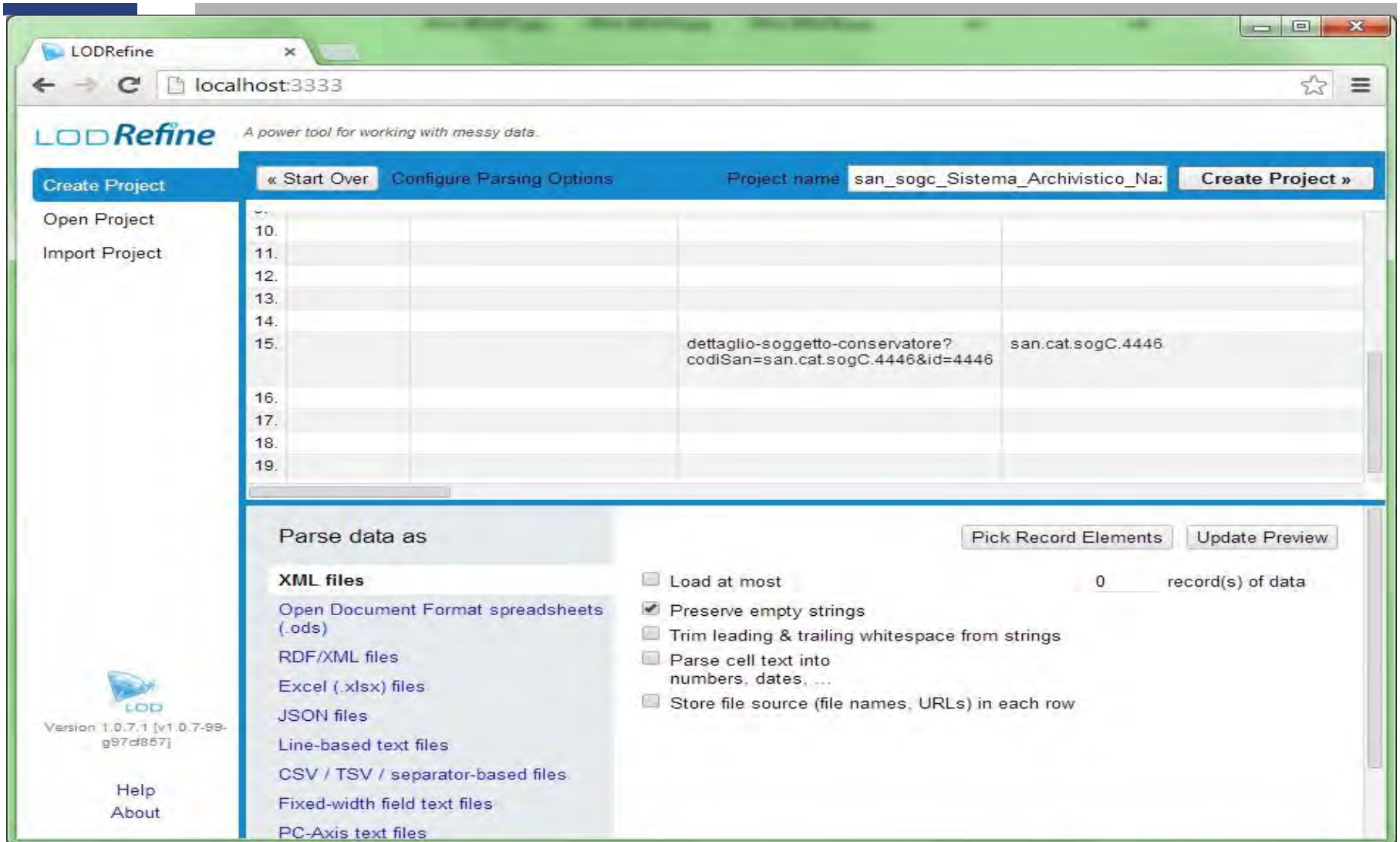
Parse data as Pick Record Elements Update Preview

XML files Load at most 0 record(s) of data

- Open Document Format spreadsheets (.ods)
- Preserve empty strings
- RDF/XML files
- Trim leading & trailing whitespace from strings
- Excel (.xlsx) files
- Parse cell text into numbers, dates, ...
- JSON files
- Store file source (file names, URLs) in each row
- Line-based text files
- CSV / TSV / separator-based files
- Fixed-width field text files
- PC-Axis text files

Version 1.0.7.1 [v1.0.7-99-g97cf857]

Help
About



LODRefine A power tool for working with messy data.

Project name: san_sogc_Sistema_Archivistico_Na. Create Project »

10.			
11.			
12.			
13.			
14.			
15.		dettaglio-soggetto-conservatore? codiSan=san.cat.sogC.4446&id=4446	san.cat.sogC.4446
16.			
17.			
18.			
19.			

Parse data as

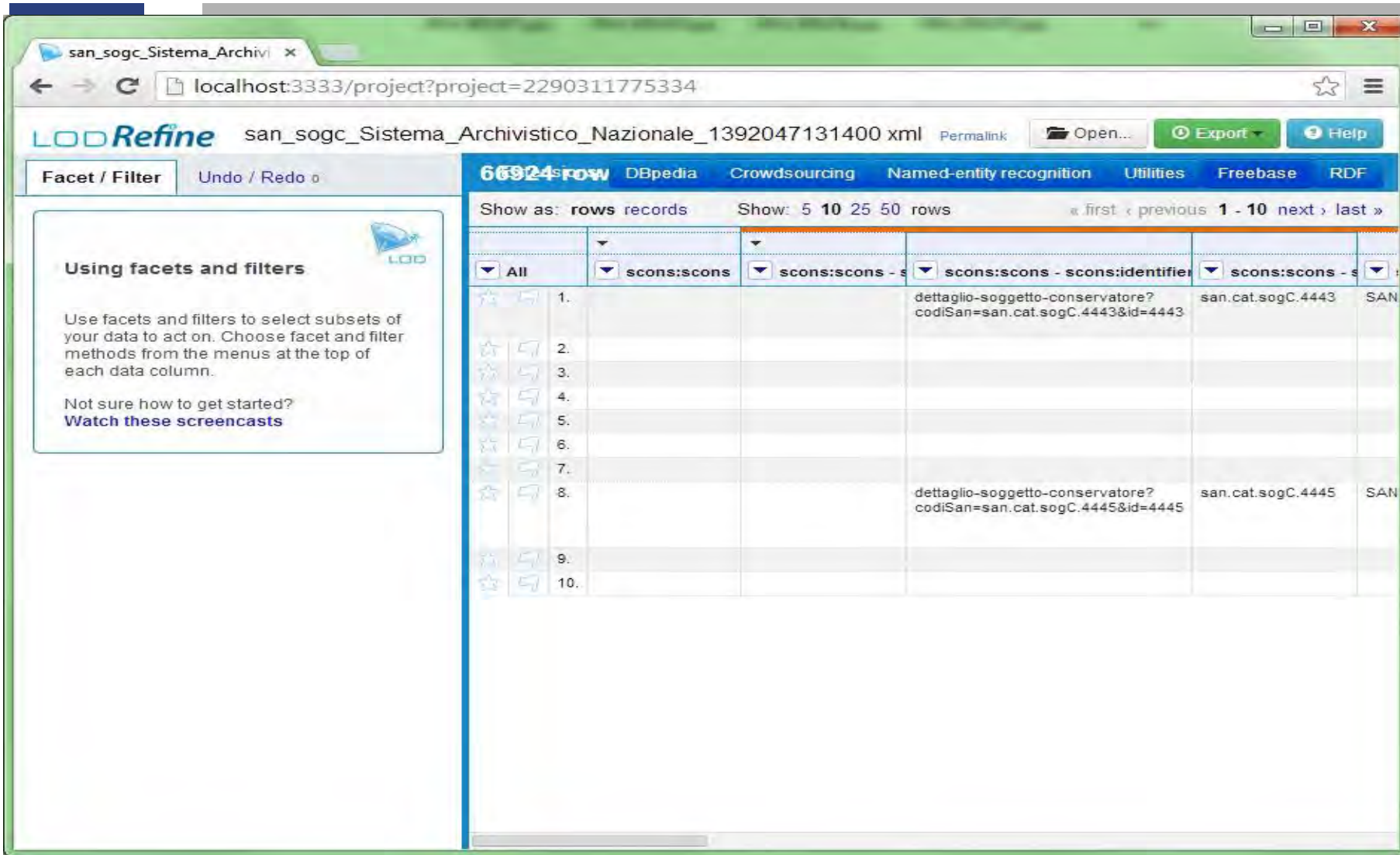
Pick Record Elements Update Preview

Load at most 0 record(s) of data

- Preserve empty strings
- Trim leading & trailing whitespace from strings
- Parse cell text into numbers, dates, ...
- Store file source (file names, URLs) in each row

Version 1.0.7.1 [v1.0.7-99-g97cf857]

Help About



san_sogc_Sistema_Archivi x

localhost:3333/project?project=2290311775334

LOD Refine san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml Permalink Open... Export Help

Facet / Filter Undo / Redo

66924 row DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

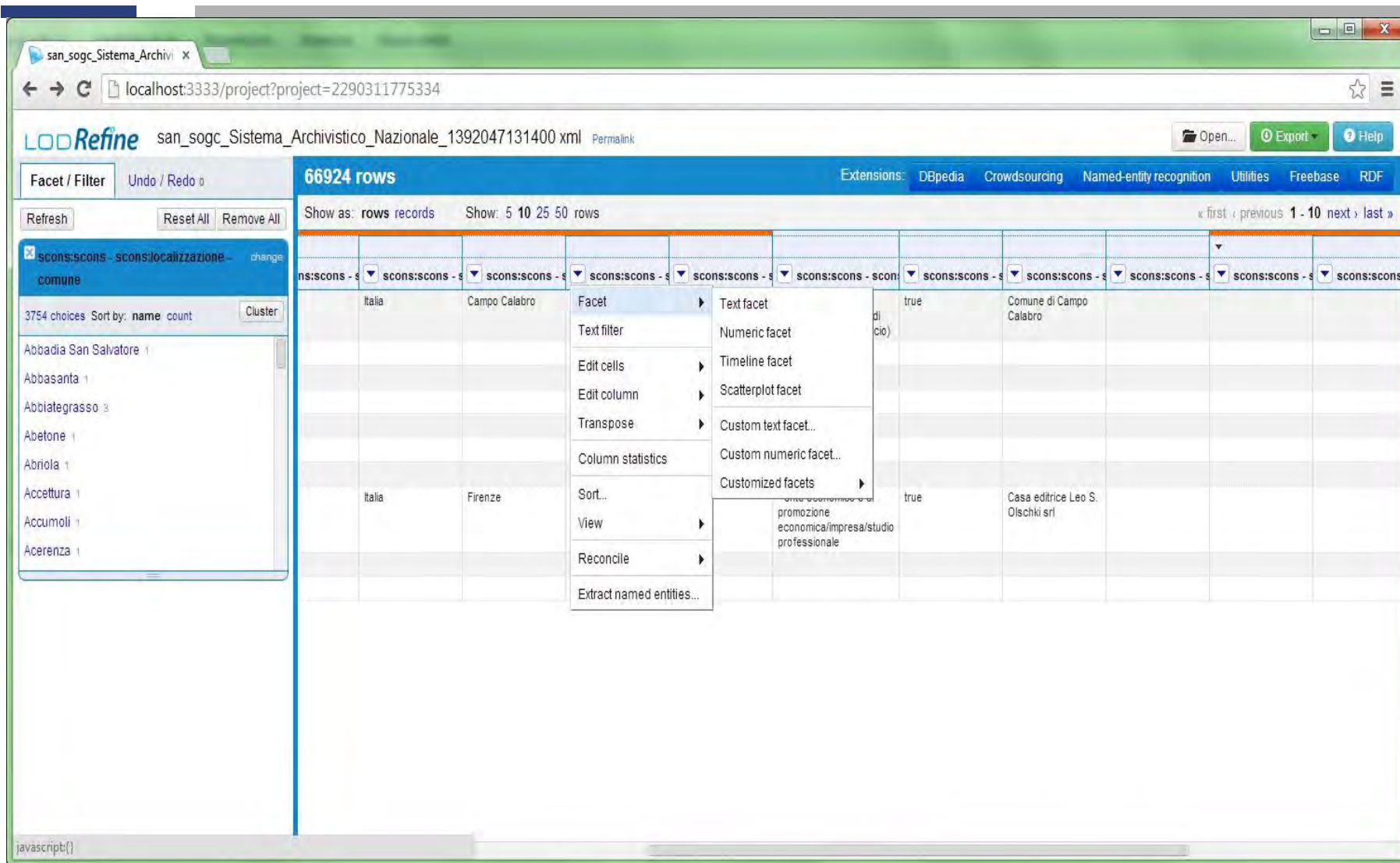
Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

All	scons:scons	scons:scons - s	scons:scons - scons:identifier	scons:scons - s	scons:scons - s
1.			dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.4443&id=4443	san.cat.sogC.4443	SAN
2.					
3.					
4.					
5.					
6.					
7.					
8.			dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.4445&id=4445	san.cat.sogC.4445	SAN
9.					
10.					

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)



The screenshot shows the LODRefine interface with the following elements:

- Browser Tab:** san_sogc_Sistema_Archivi
- Address Bar:** localhost:3333/project?project=2290311775334
- Page Title:** LODRefine san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml
- Toolbar:** Open... Export Help
- Facet/Filter:** Undo / Redo
- Table Header:** 66924 rows, Show as: rows records, Show: 5 10 25 50 rows, navigation: < first | previous 1 - 10 next | last >
- Table Content:**

Italia	Campo Calabro	di	true	Comune di Campo Calabro
Italia	Firenze	promozione economica/impresa/studio professionale	true	Casa editrice Leo S. Olschki srl
- Context Menu:**
 - Facet
 - Text facet
 - Text filter
 - Numeric facet
 - Edit cells
 - Timeline facet
 - Edit column
 - Scatterplot facet
 - Transpose
 - Custom text facet...
 - Custom numeric facet...
 - Customized facets
 - Sort...
 - View
 - Reconcile
 - Extract named entities...
- Left Panel:**
 - Facet: scons:scons - scons:localizzazione - comune
 - 3754 choices Sort by: name count Cluster
 - Abbadia San Salvatore 1
 - Abbasanta 1
 - Abbiategrosso 3
 - Abetone 1
 - Abriola 1
 - Accettura 1
 - Accumoli 1
 - Acerenza 1

san_sogc_Sistema_Archivi

localhost:3333/project?project=2290311775334

LODRefine san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml Permalink

Open... Export Help

Facet / Filter Undo / Redo

Refresh Reset All Remove All

66924 rows Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

ns:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s
Italia	Campo Calabro	Reggio Calabria	Piazza dei Martiri	- comune/città metropolitana/unione di comuni (organo e ufficio)	true	Comune di Campo Calabro				
Italia	Firenze	Firenze		- ente economico e di promozione economica/impresa/studio professionale	true	Casa editrice Leo S. Olschki srl				

Facet: comune
3754 choices Sort by: name count Cluster

- Milano 417
- Perugia 196
- Firenze 173
- Roma 164
- Reggio Di Calabria 97
- Genova 92
- Terni 87
- Foligno 83

javascript{}

san_sogc_Sistema_Archivi

localhost:3333/project?project=2290311775334

LODRefine san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml Permalink

Open... Export Help

Facet / Filter Undo / Redo

66924 rows Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

Refresh Reset All Remove All

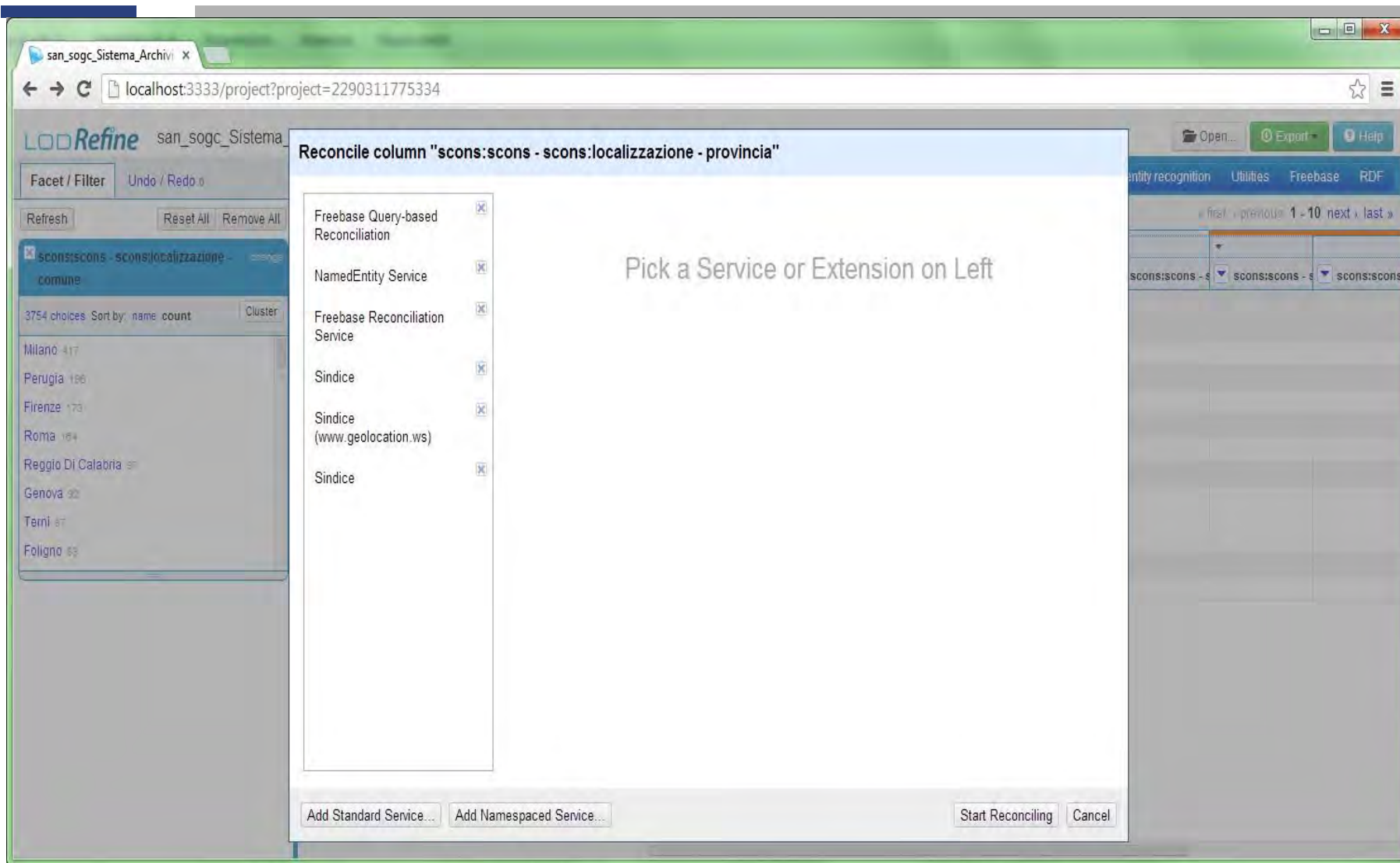
scons:scons - scons:localizzazione - change
comune

3754 choices Sort by: name count Cluster

- Milano 417
- Perugia 196
- Firenze 173
- Roma 164
- Reggio Di Calabria 97
- Genova 92
- Terni 87
- Folligno 83

scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s
Italia	Campo Calabro	Facet	el Martiri	- comune/città metropolitana/unione di comuni (organo e ufficio)	true	Comune di Campo Calabro				
Italia	Firenze	Text filter								
		Edit cells								
		Edit column								
		Transpose								
		Column statistics								
		Sort...		- ente economico e di promozione economica/impresa/studio professionale	true	Casa editrice Leo S. Olschki srl				
		View								
		Reconcile	Start reconciling...							
		Extract named entities...	Facets							
			QA facets							
			Actions							
			Copy reconciliation data...							
			Discover related RDF datasets...							

javascript({}



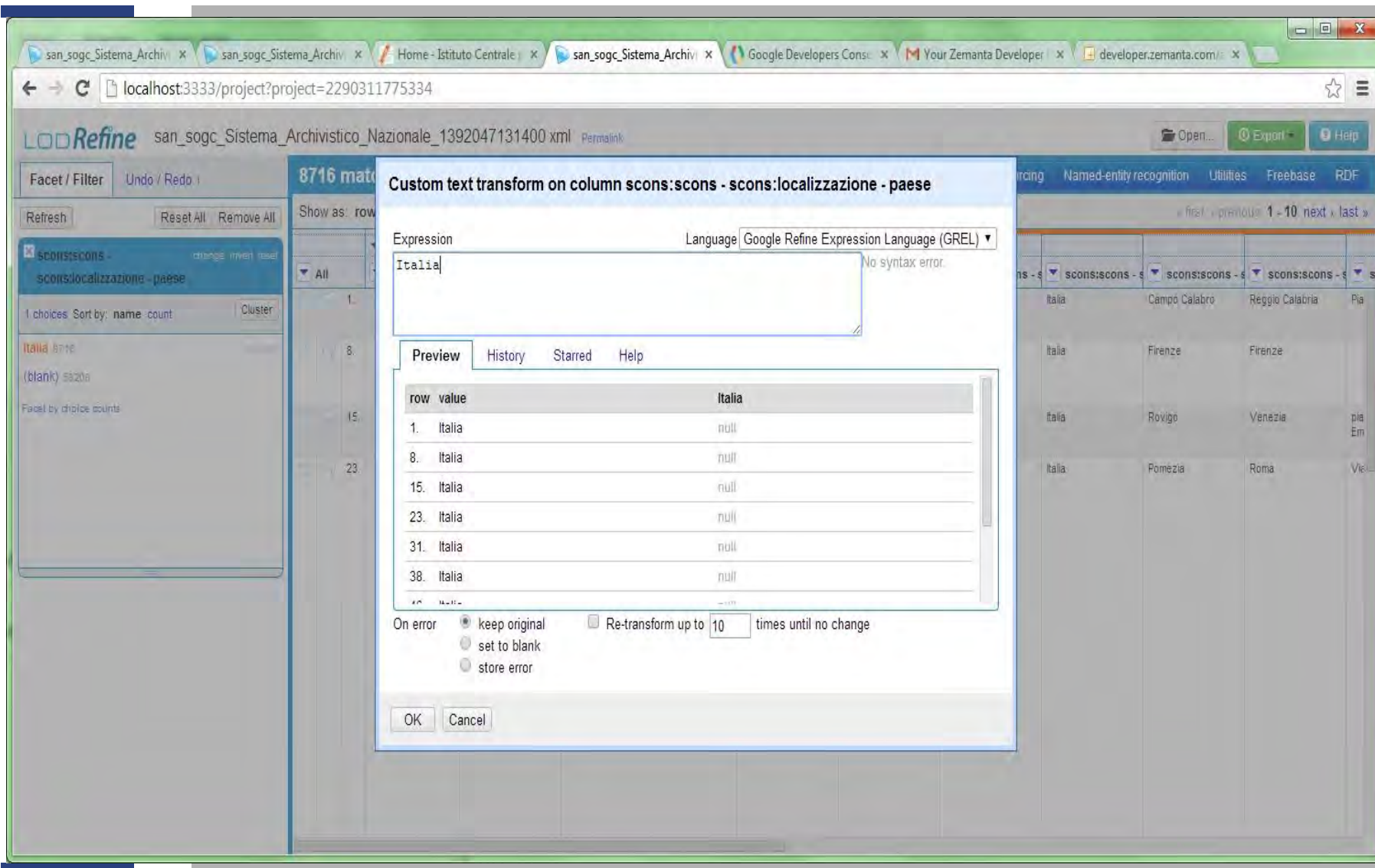
The screenshot shows the LODRefine web interface. A dialog box titled "Reconcile column 'scons:scons - scons:localizzazione - provincia'" is open. The dialog contains a list of services and extensions on the left, each with a small 'X' icon to its right:

- Freebase Query-based Reconciliation
- NamedEntity Service
- Freebase Reconciliation Service
- Sindice
- Sindice (www.geolocation.ws)
- Sindice

In the center of the dialog, the text "Pick a Service or Extension on Left" is displayed. At the bottom of the dialog, there are buttons for "Add Standard Service...", "Add Namespaced Service...", "Start Reconciling", and "Cancel".

The background interface shows a browser window with the URL "localhost:3333/project?project=2290311775334". The LODRefine interface includes a "Facet / Filter" section with a "Refresh" button and a list of facets. The selected facet is "scons:scons - scons:localizzazione - comune", showing 3754 choices sorted by name and count. The list includes:

- Milano 417
- Perugia 196
- Firenze 173
- Roma 164
- Reggio Di Calabria 98
- Genova 92
- Terni 87
- Foligno 83



The screenshot shows a web browser window with the URL `localhost:3333/project?project=2290311775334`. The page displays a LODRefine interface for the dataset `san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml`. A dialog box titled "Custom text transform on column `scons:scons - scons:localizzazione - paese`" is open, showing the expression `Italia` and a preview table.

Custom text transform dialog details:

- Title:** Custom text transform on column `scons:scons - scons:localizzazione - paese`
- Expression:** Italia
- Language:** Google Refine Expression Language (GREL)
- Status:** No syntax error.
- Preview Table:**

row	value	Italia
1.	Italia	null
8.	Italia	null
15.	Italia	null
23.	Italia	null
31.	Italia	null
38.	Italia	null
- On error options:**
 - keep original
 - set to blank
 - store error
- Re-transform:** Re-transform up to 10 times until no change

san_sogc_Sistema_Archivi x Home - Istituto Centrale x san_sogc_Sistema_Archivi x Google Developers Consc x Your Zemanta Developer x developer.zemanta.com x Understanding Expressio x

localhost:3333/project?project=2290311775334

LODRefine san_sogc_Sistema_Archivistico_Nazionale_1392047131400.xml Permalink

Open... Export Help

Facet / Filter Undo / Redo 7

66924 rows Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Refresh Reset All Remove All Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

scons:scons - scons:localizzazione - paese

case sensitive regular expression

All	scons:scons	scons:scons - s	scons:scons - scons:identifier	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s	scons:scons - s
1.			dettaglio-soggetto-conservatore? codiSan=san.cat.sogC.4443&id=4443	san.cat.sogC.4443	SAN	98052				
2.										
3.										
4.										
5.										
6.										
7.										
8.			dettaglio-soggetto-conservatore? codiSan=san.cat.sogC.4445&id=4445	san.cat.sogC.4445	SAN	50100				
9.										
10.										

Facet alabro Reggio Calabria Piaz

Text filter

Edit cells Transform...

Edit column Common transforms

Transpose Fill down

Column statistics Blank down

Sort.. Split multi-valued cells...

View Join multi-valued cells...

Reconcile Cluster and edit...

Extract named entities...

javascript{}

localhost:3333/project?project=1527800416607

LODRefine scheda_san_CL_20131004_test.csv Permalink

Facet / Filter Undo / Redo 41

6 rows Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Show as: rows records Show: 5 10 25 50 rows

	<san-dlhaSiste	san-dlhaProge	san-dlhaConse	ead-contextrep	ead-contextabb	ead-contextcor	san-dlhaComple	ead-contextunio	ead-contextuni	DirittiOD	DirittiOA
breapp/francobelli025QUi1.jpg	CASTORE	san.cat.prDgt.289		00440020501	GGASI:IC0440020501	Archivio di Stato di Lucca	san.cat.complArch.42146	GGASI:GG0440021079	Catasto Italiano	Progetto Castore - Regione Toscana	Agenzia del Territorio - Ufficio provinciale del Territorio di Lucca
breapp/francobelli008QUi1.jpg	CASTORE	san.cat.prDgt.277									
breapp/francobelli043_B03A.jpg	CASTORE	san.cat.prDgt.260		00300012901	GGASI:IC0300012901	Archivio di Stato di Firenze	san.cat.complArch.16688	GGASI:GG0300013708	Deputazione sopra il catasto poi Direzione generale del pubblico censimento di Firenze	Progetto Castore - Regione Toscana	Archivio di Stato di Firenze
breapp/francobelli186_Z03i.jpg	CASTORE	san.cat.prDgt.288	san.cat.sogC.4985	GGASI:IC0360017901	GGASI:IC0360017901	Archivio di Stato di Grosseto	san.cat.complArch.65129	GGASI:GG0440021079	Catasto	Progetto Castore - Regione Toscana	Archivio di Stato di Grosseto
breapp/francobelli046_D06i.jpg	CASTORE	san.cat.prDgt.261	san.cat.sogC.4985	GGASI:IC0440020501	GGASI:IC0440020501	Archivio di Stato di Lucca	san.cat.complArch.42156	GGASI:GG0440021077	Catasto Vecchio	Progetto Castore - Regione Toscana	Archivio di Stato di Lucca

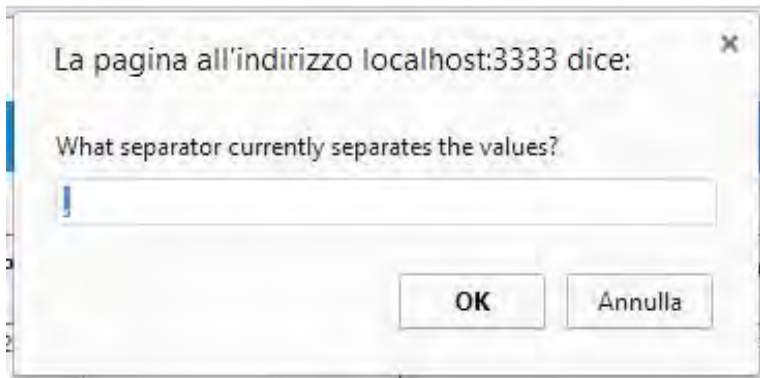
COMUNE change

6 choices Sort by: name count Cluster

Anghiari | Barga | Calenzano | Camaiore | Capalbio | Pietrasanta

Facet by choice counts

javascript:[]



localhost:3333/project?project=1527800416607

LOD Refine scheda_san_CL_20131004_test.csv

Facet / Filter Undo / Redo

6 rows

Show as: rows records

Refresh Reset All Remove All

Zenaida

6 choices. Sort by: name count Cluster

preapp/francobolli/025QU1.jpg CA

preapp/francobolli/008QU1.jpg CA

preapp/francobolli/043_B03A.jpg CA

preapp/francobolli/166_2031.jpg CA

preapp/francobolli/046_D061.jpg CASTORE

san.cat.gr.Dgt.201 san.cat.sog.C.4985 GGASI:CG440020501 GGASI:CG440020501 Archivio di Stato di Lucca san.cat.complArch.42158 GGASI:GG0440021077 Catasto Vecchio

Project: 1527800416607

Open... Export Help

dia Crowdsourcing Named-entity recognition Utilities Freebase RDF

frak | previous: 1 - 6 next | last

ldhaComple... ead-contextunit... ead-contextun... DirittiOD... DirittiOA

complArch.42148 GGASI:GG0440021079 Catasto Italiano Progetto Castore - Regione Toscana Agenzia del Territorio - Ufficio provinciale del Territorio di Lucca

complArch.38185 GGASI:GG0050003282 Catasto di Arezzo e zone limitrofe Progetto Castore - Regione Toscana Archivio di Stato di Arezzo

complArch.16668 GGASI:GG0300013706 Deputazione sopra il catasto per Direzione generale del pubblico censimento di Firenze Progetto Castore - Regione Toscana Archivio di Stato di Firenze

complArch.65129 GGASI:GG0440021079 Catasto Progetto Castore - Regione Toscana Archivio di Stato di Grosseto

preapp/francobolli/046_D061.jpg CASTORE san.cat.gr.Dgt.201 san.cat.sog.C.4985 GGASI:CG440020501 GGASI:CG440020501 Archivio di Stato di Lucca san.cat.complArch.42158 GGASI:GG0440021077 Catasto Vecchio Progetto Castore - Regione Toscana Archivio di Stato di Lucca

Add column based on column san-dl:haCompleso

New column name:

On error: set to blank store error copy value from original column

Expression: `'http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.' + split(value, ".") [3] + '&id=' + split(value, ".") [3]`

Language: Google Refine Expression Language (GREL) No syntax error.

Preview History Starred Help

row	value
	'http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.' + split(value, ".") [3] + '&id=' + split(value, ".") [3]
1.	san.cat.complArch.42146 http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.42146&id=42146
2.	san.cat.complArch.38185 http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.38185&id=38185
3.	san.cat.complArch.16668 http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.16668&id=16668
4.	san.cat.complArch.65129 http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.65129&id=65129

OK Cancel

Browser tabs: san_sogc_Sistema, Dettaglio notizza, henamesbarbara, The Programmio, Cleaning Data wit, LOD2 Webinar Sei, Google Refine 2.0, scheda_san_CL_20131004_test csv, 7% complete - sc, Dettaglio Conserv...

Address bar: localhost:3333/project?project=1527800416607

LODRefine scheda_san_CL_20131004_test csv Permalink

Facet / Filter Undo / Redo 44

6 rows

Show as: rows records Show: 5 10 25 50 rows

Create new column Url sogc based on column san-dl:haCompleso by filling 6 rows with
 grel:'http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.'+split(value, ".") [3]+'&id='+split(value, ".")[3] Undo

Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Facet by choice counts

url_franc	<san-dl:haSiste	san-dl:haProge	san-dl:haConse	Url sogc	ead-contextrep	ead-contextabb	e
http://www502.regione.toscana.it/castoreapp/francobolli/025QU11.jpg	CASTORE	san.cat.prjDgt.289	san.cat.sogC.4985	http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.42146&id=42146	GGASI:IC0440020501	GGASI:IC0440020501	Ar Lu
http://www502.regione.toscana.it/castoreapp/francobolli/006QU11.jpg	CASTORE	san.cat.prjDgt.277	san.cat.sogC.4963	http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.38185&id=38185	GGASI:IC0050003001	GGASI:IC0050003001	Ar Ar
http://www502.regione.toscana.it/castoreapp/francobolli/043_B03A.jpg	CASTORE	san.cat.prjDgt.280	san.cat.sogC.4944	http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.16666&id=16666	GGASI:IC0300012901	GGASI:IC0300012901	Ar Fir
http://www502.regione.toscana.it/castoreapp/francobolli/186_Z031.pdf	CASTORE	san.cat.prjDgt.286	san.cat.sogC.4985	http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.65129&id=65129	GGASI:IC0360017901	GGASI:IC0360017901	Ar Gr
http://www502.regione.toscana.it/castoreapp/francobolli/046_D06I.jpg	CASTORE	san.cat.prjDgt.281	san.cat.sogC.4985	http://www.san.beniculturali.it/web/san/dettaglio-soggetto-conservatore?codiSan=san.cat.sogC.42156&id=42156	GGASI:IC0440020501	GGASI:IC0440020501	Ar Lu

Browser: localhost:3333/project?project=1527800416607

LODRefine scheda_san_CL_20131004_test.csv

Facet / Filter Undo / Redo 39

6 rows

Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Show as: rows records Show: 5 10 25 50 rows

Facet / Filter menu:

- Facet
- Text filter
- Edit cells
- Edit column
- Transpose
- Column statistics
- Sort...
- View
- Reconcile
- Extract named entities...

	idmappa	TIPOMAPPA	enteconser	data_inf	data_sup	catasto	sign_arch	dimensione	scalaneum0	comunita	COMUNE	Zemanta	url_sc
1.	025QU1	Quadro d'unione	Agenzia del Territorio - Ufficio provinciale del Territorio di Lucca	01/01/1887	31/12/1897	Catasto post-unitario di Lucca	Vecchio Catasto Terreni (VCT), Contenitore 2 - Barga - Quadro unione	1050 x 690	1.25000			Barga	http://wv.layer-otr
2.	006QU1	Quadro d'unione	Archivio di Stato di Arezzo	01/01/1819	31/12/1825	Catasto Generale della Toscana - Continente	Catasto particellare - ANGHIANI - Mappe - Impianto - Quadro d'insieme	600 x 910	1.25000			Anghiani	http://wv.layer-otr
3.	043_B03A	Foglio	Archivio di Stato di Firenze	01/01/1819	31/12/1825	Catasto Generale della Toscana - Continente	Catasto Generale Toscano - Mappe - Calenzano - 8	777 x 608	1.5000			Calenzano	http://wv.layer-otr
4.	186_Z031	Foglio	Archivio di Stato di Grosseto	01/01/1824	31/12/1824	Catasto Generale della Toscana - Continente	Antico Catasto Toscano - Capalbio - Sezione Z - Foglio 3	766 x 600	1.5000	Manciano	Capalbio	Capalbio	http://wv.layer-otr
5.	046_D061	Foglio	Archivio di Stato di Lucca	01/01/1868	31/12/1868	Catasto Borbonico di Lucca	F. 15 P. 1931	572 x 729	1.2000	Camaioere	Camaioere	Camaioere	http://wv.layer-otr

Browser: localhost:3333/project?project=1527800416607

LODRefine scheda_san_CL_20131004_test.csv

Facet / Filter Undo / Redo

6 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: DBpedia Crowdsourcing Named-entity recognition Utilities Freebase RDF

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts

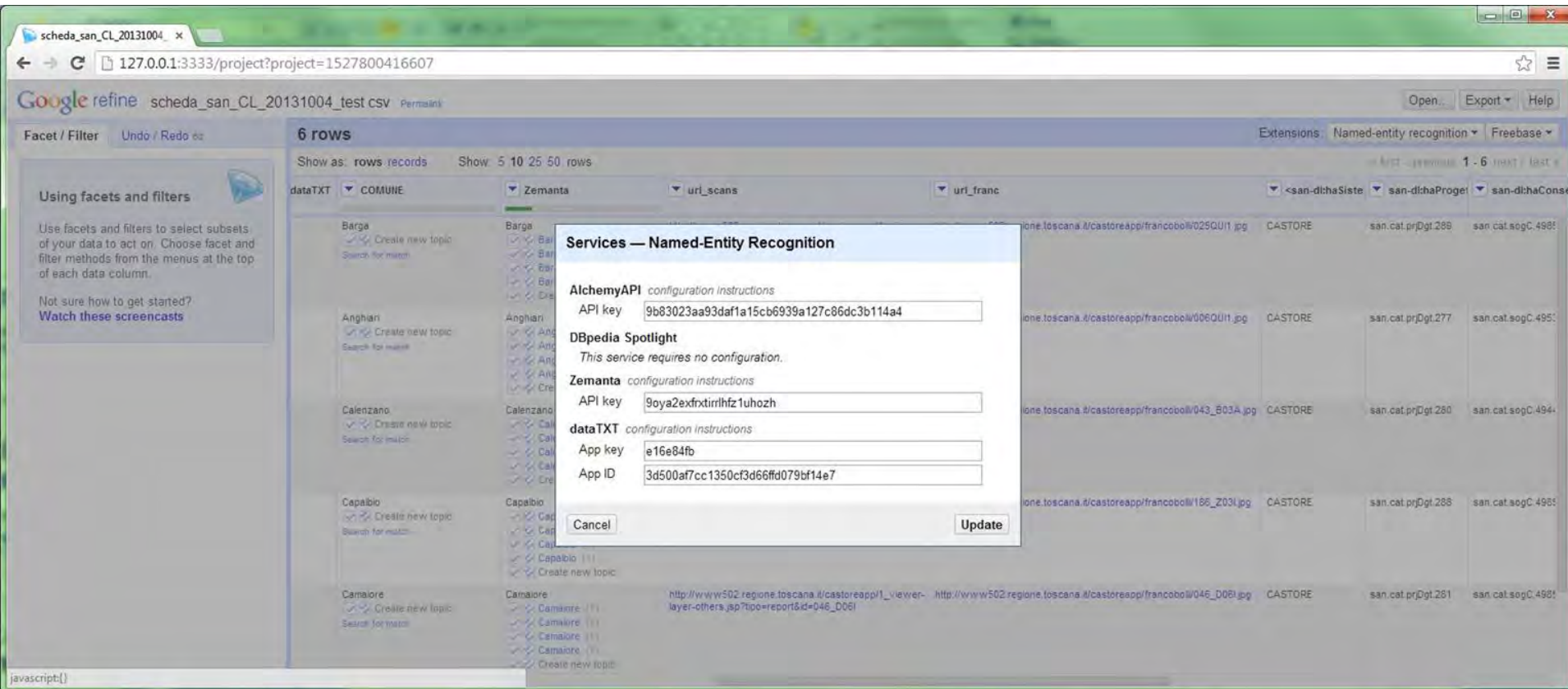
	idmappa	TIPOMAPPA	enteconser	data_inf	data_sup	catasto	sign_arch	dimensione	scalannumero	comunita	COMUNE	Zemanta	uri_sc
1.	025QUH	Quadro d'unione	Agenzia del Territorio - Ufficio provinciale del Territorio di Lucca	01/01/1887	31/12/1897	Catasto post-unitario	Vecchio Catasto Terreni	1050 x 690	1:25000	Barga	Barga Self Taken (1) Barga (1,204) Bagni di Lucca & Barga (1,227) Bagni di Lucca & Barga (1,227) Create new topic Search for matches	Barga Barga (1) Barga (1) Barga (1) Barga (1) Create new topic	http://wv.layer-otf
2.	006QUH	Quadro d'unione	Archivio di Stato di Arezzo						1:25000	Anghiari	Anghiari Choose new matches	Anghiari Anghiari (1) Anghiari (1) Anghiari (1) Anghiari (1) Create new topic	http://wv.layer-otf
3.	043_B03A	Foglio	Archivio di Stato di Firenze						1:5000	Calenzano	Calenzano Create new topic Search for matches	Calenzano Calenzano (1) Calenzano (1) Calenzano (1) Calenzano (1) Calenzano (1) Create new topic	http://wv.layer-otf
4.	186_Z03I	Foglio	Archivio di Stato di Grosseto	01/01/1824	31/12/1824	Catasto Generale della Toscana - Continente	Antico Catasto Toscano - Sezione Z - Foglio 3	786 x 600	1:5000	Manciano	Capalbio Create new topic Search for matches	Capalbio Capalbio (1) Capalbio (1) Capalbio (1) Capalbio (1) Create new topic	http://wv.layer-otf
5.	046_D06I	Foglio	Archivio di Stato di Lucca	01/01/1868	31/12/1868	Catasto Borbonico di Lucca	F. 15 P. 1931	572 x 729	1:2000	Camaione	Camaione Lido di Camaione (Temperale) (1,298) Tramonto a Lido di	Camaione Camaione (1) Camaione (1) Camaione (1) Create new topic	http://wv.layer-otf

Extract named entities in comunita

Named-entity recognition services
Use the following services to recognize and extract named entities in the comunita column:

- AlchemyAPI
- DBpedia Spotlight
- Zemanta
- dataTXT

Cancel Start extraction

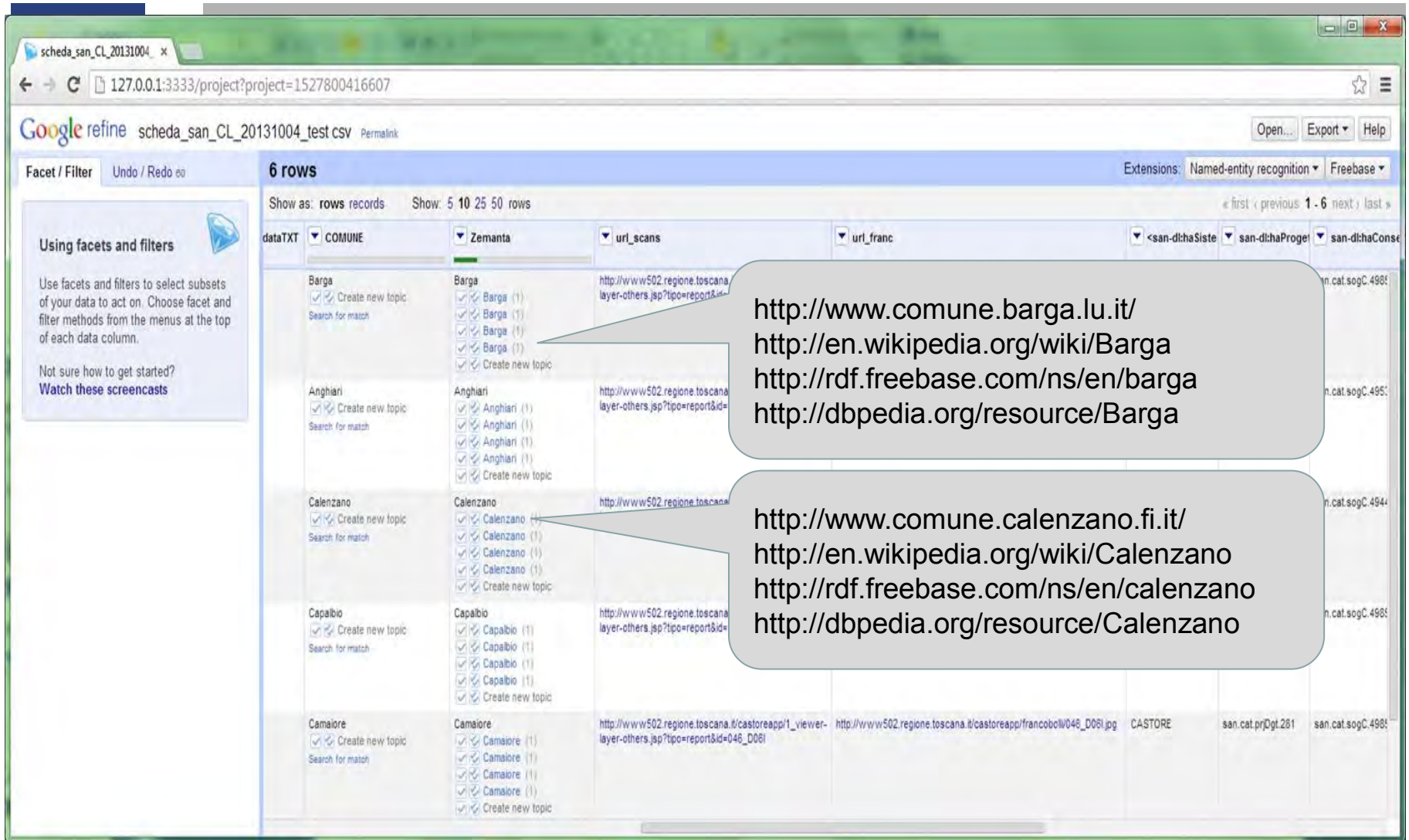


The screenshot shows a web browser window displaying a data table with 6 rows. A modal dialog box titled "Services — Named-Entity Recognition" is open in the center, allowing configuration of various services. The background table has columns for data types (dataTXT, COMUNE, Zemanta, uri_scans, uri_franc) and various identifiers.

Services — Named-Entity Recognition

- AlchemyAPI** *configuration instructions*
API key:
- DBpedia Spotlight**
This service requires no configuration.
- Zemanta** *configuration instructions*
API key:
- dataTXT** *configuration instructions*
App key:
App ID:

Buttons:



Google refine scheda_san_CL_20131004_test csv Permalink

Facet / Filter Undo / Redo 6 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Named-entity recognition Freebase

dataTXT COMUNE Zemanta url_scans url_franc <san-dlthaSiste san-dlthaProgel san-dlthaConse

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

Barga
 Create new topic
 Search for match

Anghiari
 Create new topic
 Search for match

Calenzano
 Create new topic
 Search for match

Capalbio
 Create new topic
 Search for match

Camaiore
 Create new topic
 Search for match

Barga
 Barga (1)
 Barga (1)
 Barga (1)
 Barga (1)
 Create new topic

Anghiari
 Anghiari (1)
 Anghiari (1)
 Anghiari (1)
 Anghiari (1)
 Create new topic

Calenzano
 Calenzano (1)
 Calenzano (1)
 Calenzano (1)
 Calenzano (1)
 Create new topic

Capalbio
 Capalbio (1)
 Capalbio (1)
 Capalbio (1)
 Capalbio (1)
 Create new topic

Camaiore
 Camaiore (1)
 Camaiore (1)
 Camaiore (1)
 Camaiore (1)
 Create new topic

Callout for Barga:
<http://www.comune.barga.lu.it/>
<http://en.wikipedia.org/wiki/Barga>
<http://rdf.freebase.com/ns/en/barga>
<http://dbpedia.org/resource/Barga>

Callout for Calenzano:
<http://www.comune.calenzano.fi.it/>
<http://en.wikipedia.org/wiki/Calenzano>
<http://rdf.freebase.com/ns/en/calenzano>
<http://dbpedia.org/resource/Calenzano>

Esempio – Luogo Calenzano

```
<owl:NamedIndividual rdf:about="http://san.beniculturali.it/SAN/Calenzano">
  <rdf:type rdf:resource="http://san.beniculturali.it/SAN/luogo"/>
  <rdf:type rdf:resource="http://san.beniculturali.it/SAN/luogoConservatore"/>
  <rdfs:label rdf:datatype="string">Calenzano</rdfs:label>
  <SAN:luogoCAP>10100</SAN:luogoCAP>
  <SAN:luogoPaese rdf:datatype="string">Italia</SAN:luogoPaese>
  <SAN:luogoProvincia rdf:datatype="string">FI</SAN:luogoProvincia>
  <SAN:luogoComune rdf:datatype="string">Calenzano</SAN:luogoComune>
  <SAN:luogoTesto rdf:datatype="string">Calenzano</SAN:luogoTesto>
  <dc:title>Calenzano</dc:title>
  <owl:sameAs rdf:resource="http://sws.geonames.org/6541087"/>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/Turin"/>
  <owl:sameAs rdf:resource="http://www.comune.calenzano.fi.it"/>
  <owl:sameAs rdf:resource="http://en.wikipedia.org/wiki/Calenzano"/>
  <owl:sameAs rdf:resource="http://rdf.freebase.com/ns/en/calenzano"/>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/Calenzano"/>
</owl:NamedIndividual>
```

Freebase



Popolamento
sameAs

Arricchimento
riconciliazione

Servizio di riconciliazione e relazione con domini esterni

Obiettivo finale è rendere disponibili tutti i dati correlati, normalizzati, verificati come servizio di riconciliazione.

Il SAN metterà a disposizione open data sia di tipo archivistico che di tipo geografico relativamente ai toponimi in modo che si possa utilizzare come servizio per relazionarli con la documentazione presente sul territorio.

In riferimento ai toponimi storici, questo servizio diventa fondamentale in quanto le informazioni sono garantite dalla fonte statale e diventano quindi elemento di continuo arricchimento per i sistemi collegati.

In parallelo si intraprenderanno i contatti necessari per la correlazione di servizi come dbpedia o freebase per far sì che possano inglobare loro le i collegamenti al Sistema archivistico Nazionale.



dati.san.beniculturali.it



- 1 Sistema Archivistico Nazionale
- 2 Progetto Linked Open Data
- 3 Struttura dei dati CAT-SAN
- 4 Disegno di Ontologie
- 5 Ontologia SAN
- 6 SKOS e Thesaurus SAN
- 7 Openrefine
- 8 Conclusioni**

Conclusioni

- **Attenzione al ciclo di vita dei dati**– creazione, aggiornamento e conservazione dei dati, e disattivazione dei dataset.
- **Attenzione al ciclo di vita dei metadati** sia come struttura dell'ontologia sia come contenuti
- Mettere in atto **processi** per assicurarsi che i dati ed i metadati abbiano un appropriato livello di **qualità**.
- Monitorare l'**uso** ed il **feedback** ai (potenziali) utilizzatori dei dati, in particolare se i dati che necessitano sono quelli corretti, in quale modo si possono utilizzare i dati e soprattutto cosa ne pensano della qualità.
- **Riuso delle esperienze**, dei modelli operativi e tecnologici
- **Proliferazione dei vocabolari e delle ontologie**

Documentazione

Sito Web dell'Istituto Centrale per gli Archivi:

<http://www.icar.beniculturali.it>

Portale del Sistema Archivistico Nazionale:

<http://www.san.beniculturali.it>

L'ontologia SAN è disponibile su :

<http://www.san.beniculturali.it/SAN/SAN-LOD.v1.0.owl>

Una rappresentazione HTML dell'Ontologia SAN LOD.1.0 è disponibile su

:<http://www.maas.ccr.it/SAN-LOD/lode/>

Il Thesaurus SAN è disponibile su:

<http://www.san.beniculturali.it/SAN/TesauroSAN.v1.0.rdf>

Una rappresentazione grafica del Thesaurus è raggiungibile su:

<http://www.maas.ccr.it/SAN-LOD/skos/>

Bibliografia

- Della Valle, Celino, Cerizza (2009) Semantic web. Dai fondamenti alla realizzazione di un'applicazione, ISBN 8871925785
- Di Noia, Di Sciascio, De Virgilio (2013) Semantic web. Tra ontologie e Open Data, ISBN 8850332149
- Ducharme (2011), Learning SPARQL: Querying and Updating With Sparql 1.1, ISBN 1449306594
- Allemang, Hendler (2011), Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, ISBN 0123859654

Contatti



ICAR (Istituto Centrale per gli Archivi)
Direttore: Marina Giannetto
via di Castro Pretorio 105, 00185 Roma



ic-a.direttore@beniculturali.it; mbac-ic-a@mailcert.beniculturali.it



(+39) 06 5190976



<http://www.icar.beniculturali.it/>

Gruppo tecnico SAN – LOD

Stella Di Fazio



ic-a@beniculturali.it; stella@maas.ccr.it

Maria Idria Gurgo



ic-a@beniculturali.it; mariaidria.gurgo@beniculturali.it



Costantino Landino

costantino.landino@gmail.com; ic-a.sistemiinformativi@beniculturali.it



(+39) 342 1957312



costantinolandino



www.costantinolandino.it

