Sistema **A**rchivistico **N**azionale (**SAN**) e **L**inked **O**pen **D**ata :

Marina Giannetto
Archivio Centrale dello Stato
14 gennaio 2014

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

Il SAN : principali caratteristiche e funzioni

- inaugurato il 17 dicembre 2011, è un **aggregatore nazionale di risorse archivistiche** realizzato a seguito di una consistente operazione di allineamento dei tracciati dei sistemi informativi archivistici aderenti.

- costituisce un **punto di accesso integrato** al patrimonio archivistico italiano, rendendo disponibili, attraverso il proprio **"Catalogo delle risorse archivistiche"** e la **Digital Library**, descrizioni di archivi e oggetti digitali di natura eterogenea, statali e non statali, pubblici e privati, finora consultabili separatamente

- **L'interoperabilità** costituisce uno degli **obiettivi prioritari** perseguiti dall'ICAR unitamente al progetto per lo sviluppo e la pubblicazione di **Linked Open Data di ambito archivistico – SAN LOD**

http://san.beniculturali.it/web/san/dettaglio-notizia-san?p_p_id=56_INSTANCE_X7Qi&articleId=2925424&p_p_lifecycle=1&p_p_state=normal&groupId=10704&viewMode=normal

Il **SAN offre accesso** alle seguenti tipologie di contenuti :

- **catalogo delle risorse archivistiche**: descrizioni dei **conservatori**, dei **produttori**, dei **complessi archivistici** e degli **strumenti di ricerca**, trasmesse al SAN attraverso standard (archivistici e NIERA), tracciati di scambio (CAT SAN) e protocolli siglati con i responsabili dei sistemi aderenti. Ogni scheda CAT ha un solo collegamento ipertestuale (tramite identificativo persistente) verso il sistema di provenienza delle descrizioni e tale collegamento punta alla scheda corrispettiva presente in quel sistema

- ***digital library***: risorse digitali riferite a documenti di diversa natura e tipologia (immagini, audio, video), descritti 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.

- **risorse bibliografiche, percorsi, notizie, informazioni, testi** e **contenuti arricchiti**

- **portali tematici** a diverso titolo integrati nel SAN

Gli Standard del SAN NIERA EPF; CAT SAN Metadati relativi alle risorse archivistiche accessibili attraverso il Sistema Archivistico Nazionale, METS SAN Metadati Oggetti Digitali nel Sistema Archivistico Nazionale, pubblicati in http://www.icar.beniculturali.it/index.php?it/150/archivio-news/59/gli-standard-di-san-niera-epf-metadati-risorse-archivistiche-metadati-oggetti-digitali

Attualmente si hanno 553.214 risorse

185.845 Risorse Archivistiche, 2.733 Contenuti Redazionali, 2.605 Risorse Bibliografiche, 361.931 Archivio Digitale, 16.832 Fonti Esterne

Nel Catalogo delle risorse archivistiche e nella Digital Library :

- ***Soggetti conservatori*** ***8.878***
- ***Soggetti produttori*** ***67.119***
- ***Complessi archivistici*** ***91.296***
- ***Strumenti di corredo*** *(inventari, guide, ecc.)* ***18.552***
- ***Oggetti digitali*** *(foto, video, audio, documenti pdf, ecc.) con relativi metadati* *361**.931***
- ***Totale*** ***553.214***

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 – SIASPG - SIASVE
Guida Generale Archivi Stato italiani
IBC ARCHIVI
Lombardia Beni Culturali (LBC)
INSMLI
CASTORE - Imago Tusciae
“Spazi della Follia”

SISTEMA ARCHIVISTICO NAZIONALE

CEI-AR

Portali tematici

Impresa
Per non dimenticare
Musica
Architetti
Moda
Archivio Storico Multimediale del Mediterraneo
Territori
Antenati
Verdi

SAN e Linked Open Data : Obiettivi

L' **Istituto Centrale per gli Archivi** ha reso 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

Un OAI Provider rende accessibili in formato Open Data i contenuti del SAN relativi alle seguenti entità: conservatori di archivi, produttori di archivi, complessi archivistici, strumenti di ricerca, oggetti digitali (metadati strutturati secondo schemi XML **ead-san, eac-san, ricerca-san, scons-san, mets-san)**

Alla prima necessaria 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 – SAN LOD**

Obiettivi del progetto di sviluppo e pubblicazione di **Linked Open Data di ambito archivistico - SAN LOD** :

1. Definire una **Ontologia di base SAN** in grado di traslare compiutamente in formato OWL (Ontology Web Language) lo schema concettuale della struttura informativa espressa dai tracciati di scambio CAT SAN (standard di rappresentazione dei dati all'interno del SAN), previa analisi delle esperienze nazionali ed internazionali che affrontano, pur con sostanziali peculiarità legate ai rispettivi contesti di applicazione, analoghe problematiche descrittive.

L'ICAR ha scelto di rappresentare con pienezza espressiva i tracciati CAT SAN nella ontologia, definendo i nomi delle classi di oggetti individuati e le relative proprietà sulla base della nomenclatura da essi adottata, poiché essi veicolano una **conoscenza consensuale**, frutto, nel 2009, del confronto tra i membri della *Sottocommissione tecnica per la definizione dei metadati relativi alle risorse archivistiche rese accessibili attraverso il Sistema Archivistico Nazionale.*

In questa ottica, si è deciso di **elaborare un Tesauro SAN in formato SKOS,** opportunamente integrato nell'ontologia OWL, costituito dai vocabolari controllati previsti dallo standard CAT SAN e dalle voci tematiche presenti in SIAS e SIUSA come attributo dei complessi archivistici

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

2. Definire una serie di **estensioni all'Ontologia di base** che permetta l'integrazione di entità, elementi informativi e relazioni tra oggetti – dei quali è emersa in corso d'opera la rilevanza - attualmente non presenti all'interno del SAN, perché non previsti nel tracciato CAT SAN, ma residenti nei sistemi di provenienza o in altre fonti individuate (ufficiali o rese disponibili come risorse LOD: geonames, dbpedia, etc.) non necessariamente di natura archivistica

3. Realizzare **liste di authority** di **toponimi, repertori, strumenti** su scala nazionale, basate principalmente sulla **rielaborazione in formato LOD** di dati istituzionali di fonte ISTAT : **Sistema Informativo Storico delle Amministrazioni Territoriali (SISTAT)**, che incamera dati a partire dal 1861, integrabili con i dati SAN LOD di provenienza Guida Generale degli Archivi di Stato.

4. Rendere disponibili una serie di **strumenti di accesso ai dati LOD** prodotti :

- **endpoint sparql** per la pubblicazione e l'interrogazione dell'ontologia e dei dataset SAN disponibili in formato LOD
- **OAI Provider** che distribuisca dati pubblici secondo gli schemi CAT – SAN
- **area di download** dove poter scaricare i set specifici di dati LOD resi disponibili dal SAN

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 – peraltro adottati dalla *Sottocommissione tecnica per la definizione dei metadati relativi alle risorse archivistiche rese accessibili attraverso il Sistema Archivistico Nazionale* - quanto il riferimento ai principali namespace in ambito LOD per le operazioni di riconciliazione necessarie alla loro integrazione in ambiente Linked Data.

Particolare attenzione è stata prestata inoltre ad identificare nell'ambito di ontologie (FOAF - Friend of a Friend) o formati di scambio (Dublin Core) estremamente diffusi, i punti di raccordo con gli elementi descrittivi essenziali del modello elaborato.

In questa fase operativa sono state introdotte anche corrispondenze semantiche con le ontologie pubblicate all'interno del *Progetto RELOAD - Repository for Linked open archival data*, iniziativa nazionale che ha affrontato – con sostanziali differenze rispetto all'esperienza condotta per il modello SAN, essenzialmente riconducibili ai diversi obiettivi dichiarati dai due progetti e ai rispettivi contesti di applicazione - analoghe problematiche descrittive in ambito archivistico.

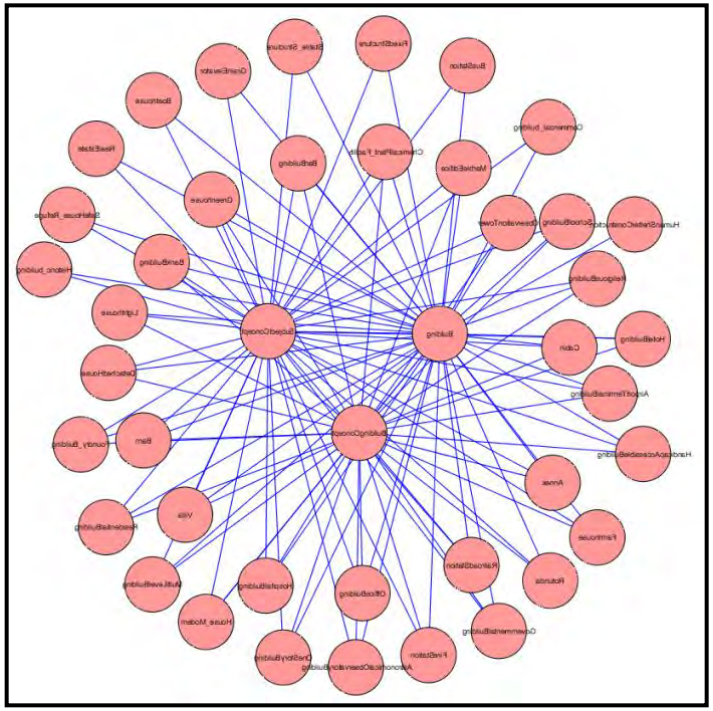
Ontologia di base : rappresentazione del modello concettuale struttura informativa SAN

schemaEAD.xsd

schemaEAC.xsd

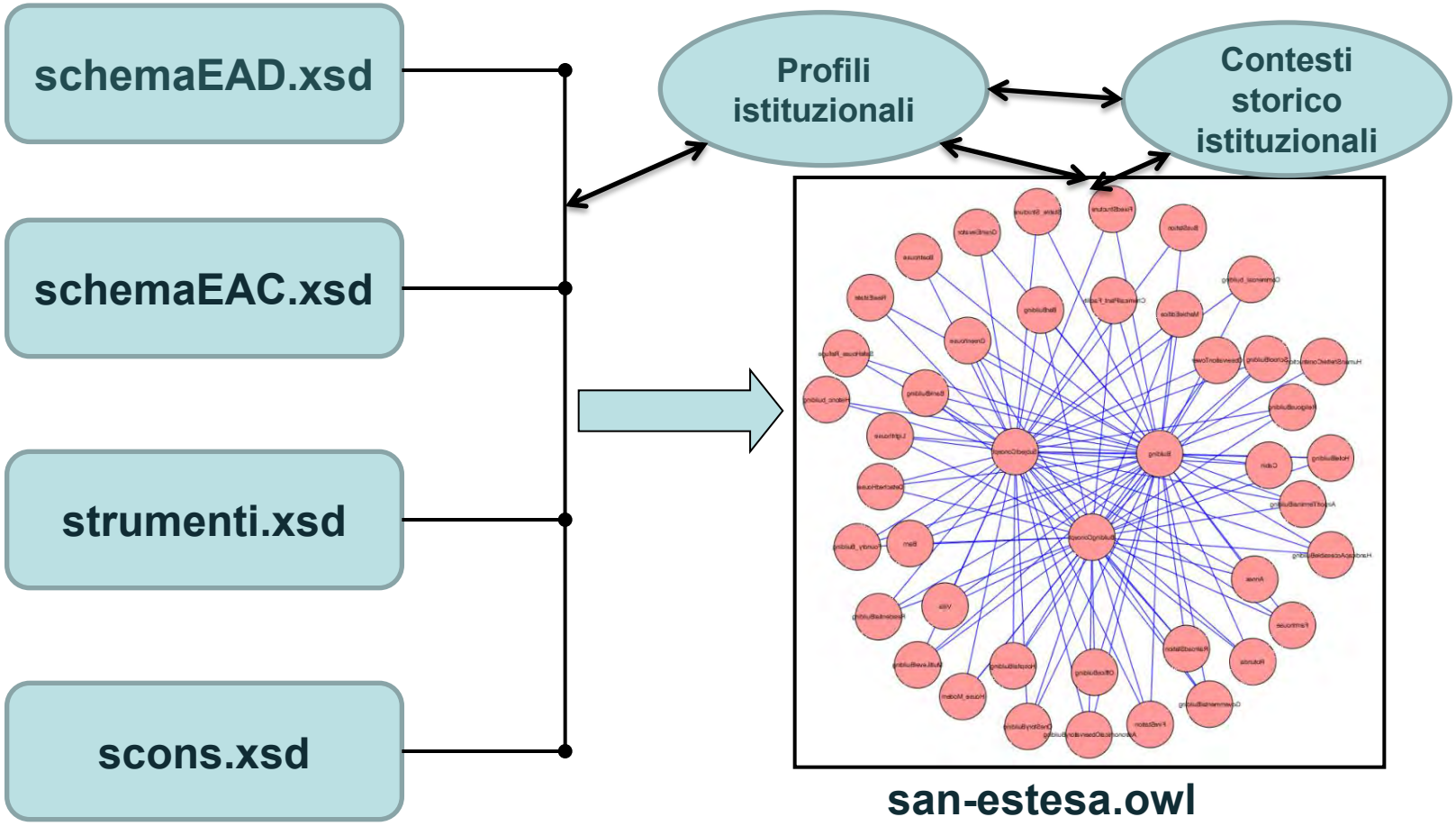
strumenti.xsd

scons.xsd



san.owl

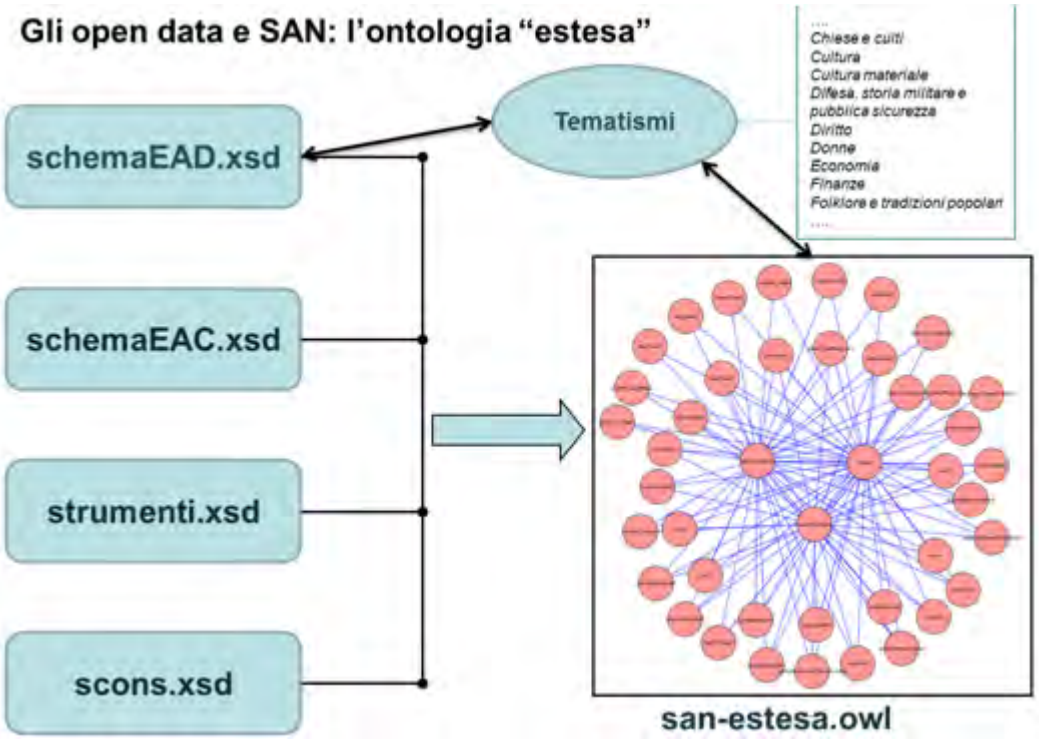
Ontologia "estesa" : una **prima estensione** ha comportato il suo arricchimento con **due ulteriori classi di oggetti** (profilo istituzionale e contesto storico istituzionale) e con quelle proprietà necessarie a descriverle e a metterle in relazione coi soggetti produttori.



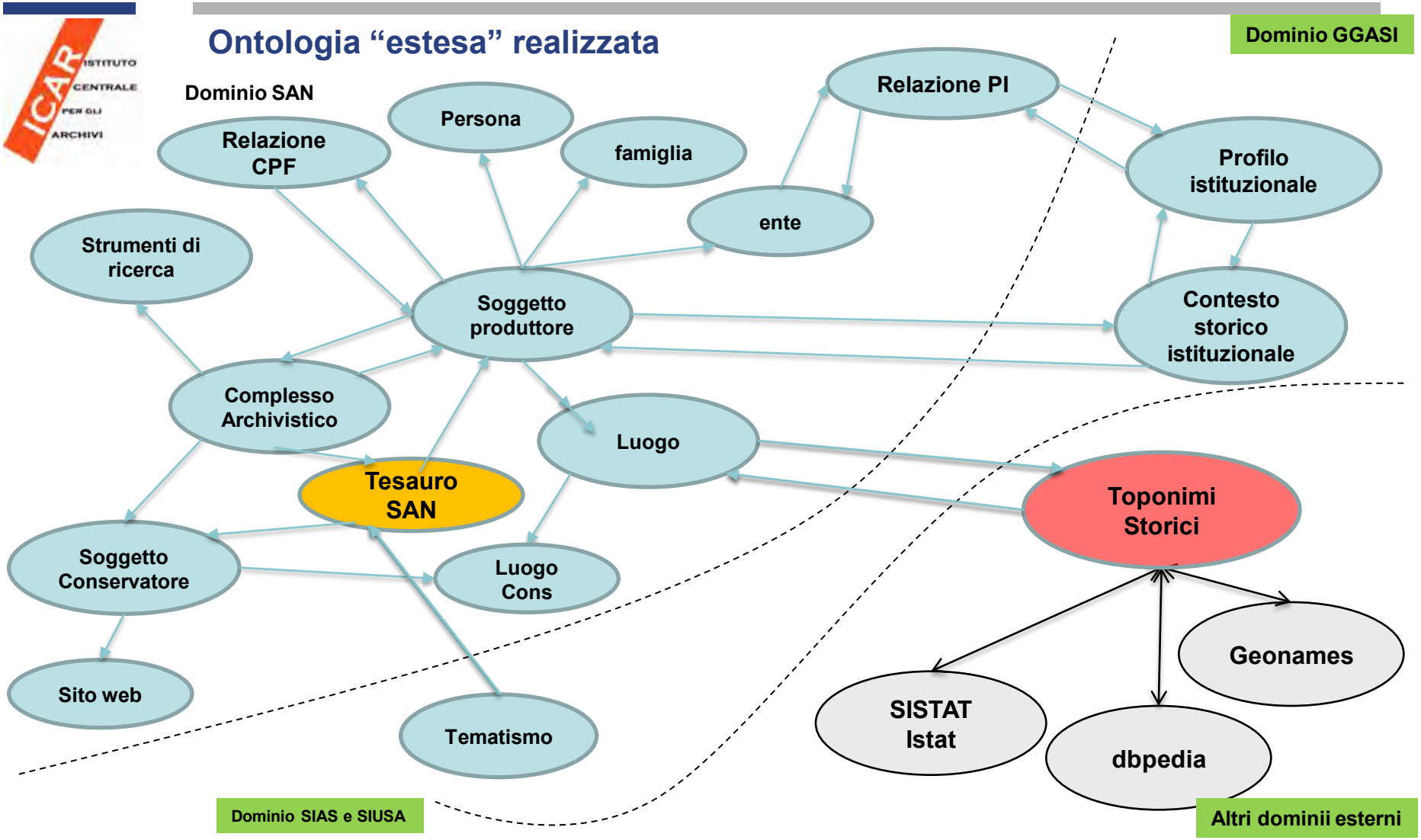
Open Data e SAN : l'ontologia "estesa"

Una **seconda estensione** dell'ontologia SAN riguarda l'**inclusione nel modello concettuale di chiavi d'accesso** ai complessi archivistici organizzate per **«Tematismo»**, composto da oltre quaranta voci tematiche presenti come attributo qualificante dei complessi archivistici in alcuni sistemi di provenienza (SIAS, SIUSA).

La classe «Tematismo» è stata introdotta nel SAN LOD come estensione del modello nativo, allo scopo di offrire una fondamentale chiave di classificazione e di accesso tematico alla documentazione, potenziando e facilitando oltremodo i percorsi di ricerca



Ontologia "estesa" realizzata



Ontologia SAN: i numeri

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

- **6 classi** corrispondenti alle entità principali del SAN_esteso (**complesso archivistico**, **soggetto conservatore**, **strumento di ricerca**, **soggetto produttore**, **profilo istituzionale**, **contesto storico istituzionale**);

- **5 classi** riferibili a dimensioni informative generali e/o trasversali alle entità (**date**, **luogo**, **sito web**, **qualificazione**, **PIrelation** [correlazione profili istituzionali e soggetti produttori])

- Una superclasse *skos:Concept* cui si riferiscono i 6 concetti principali (classi) in cui risulta articolato il Tesauro SAN (**Tipo_qualifica_CPF**, **Tipo_luogo_CPF**, **Tipo_data_esistenza_CPF**, **tipologia_Conservatore**, **Tematismo**, **Status_scheda_provenienza**, **Estensione_data_complesso**)

- **la classe soggetto produttore** articolato in **3 sottoclassi** (ente, persona, famiglia)
- la classe **date** articolata in **5 sottoclassi** (date complesso archivistico, estensione date complesso archivistico, date contesto storico istituzionale, date esistenza, date strumento)
- la classe **luogo** articolata in **1 sottoclasse** luogo produttore

Il Tesauro del Sistema Archivistico Nazionale

- è costituito dai diversi vocabolari controllati presenti negli schema CAT SAN come supporto ad una descrizione normalizzata delle risorse (conservatori, produttori, complessi) provenienti dai diversi sistemi aderenti al SAN :
  - Tipologia della qualifica associata al soggetto produttore e relative sottotipologie
  - Tipologia del luogo associato al soggetto produttore
  - Tipologia della data di esistenza del soggetto produttore
  - Tipologia del soggetto conservatore
  - Status della scheda nel sistema di provenienza
  - Estensione alla data del complesso archivistico

- I vocabolari controllati sono stati **elaborati come concetti SKOS** (Simple Knowledge Organisation System) e opportunamente integrati nell'ontologia OWL.

Alle voci originarie è stata aggiunto un ulteriore vocabolario "Tematismo" composto da oltre quaranta voci tematiche, presenti come attributo qualificante dei complessi archivistici in alcuni sistemi di provenienza (SIAS, SIUSA), allo scopo di offrire una chiave di classificazione e di accesso tematico alla documentazione

cfr. http://san.beniculturali.it/SAN/TesauroSAN_1.0

Il Tesauro del Sistema Archivistico Nazionale

Il Tesauro del Sistema Archivistico Nazionale è stato rappresentato in **SKOS/RDF**
Nell'ontologia è stata definita anche una superclasse *skos:Concept* cui si riferiscono i 7 concetti principali (classi) in cui risulta articolato il Tesauro SAN integrato con «Tematismo»:

- **Tipo_qualifica_CPF**
- **Tipo_luogo_CPF**
- **Tipo_data_esistenza_CPF**
- **Tipologia_Conservatore**
- **Tematismo**
- **Status_scheda_provenienza**
- **Estensione_data_complesso**

SKOS è l'acronimo di **Simple Knowledge Organisation System** ed è una famiglia di linguaggi formali creata per rappresentare glossari, classificazioni, tassonomie e qualsiasi tipo di vocabolario strutturato. SKOS si basa su RDF e RDFS e il suo obiettivo principale è di consentire una facile pubblicazione di vocabolari strutturati per il Semantic Web.

Il Tesauro del Sistema Archivistico Nazionale: esempi

```
<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:  abel>
  <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"/>
```

Definizioni generali

Elementi principali del tesauro

Il Tesauro del Sistema Archivistico Nazionale: esempi

```
  <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                                    #Concept"/>
  <skos:inScheme rdf:resour                                      l/TesauroSAN_1.0"/>
  <skos:prefLabel xml:lang="                                     azione acattolica</skos:prefLabel>
  <skos:scopeNote xml:lang="it">Sono raggruppat                  sta voce gli enti relativi al culto o a fun
associazioni religiose non cattoliche: ebraiche, valdesi, ec                  costituzione repubblicana ricon
liberamente la propria fede religiosa in qualsiasi forma, individua                  associata, e di seguirne in p
purch&#233; 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 conse                  te presso gli Archivi di Stato, vi sono
pervenute, di massima, frammiste agli archivi                  elle corporazioni religiose soppresse o di tribuna                  antico regime. Gli enti
ecclesiastici e le associazioni cattoliche                  servano i propri archivi che, se estinti, possono confluire                  archivi di altri enti
ecclesiastici o presso altre istituzio                  o un proprio regolamento gli Archivi storici diocesani.</skos:scop
  <skos:broad                                    ulturali.it/SAN/TesauroSAN_1.0/non_statali"/>
 </rdf:Descrip
```

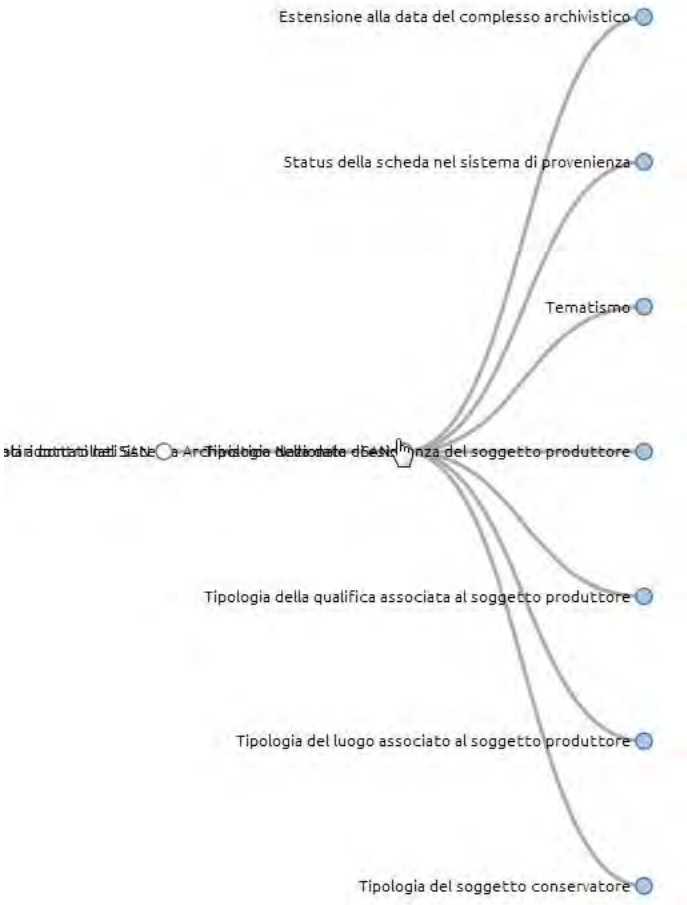
Termine gerarchicamente superiore

Concetto e riferimento

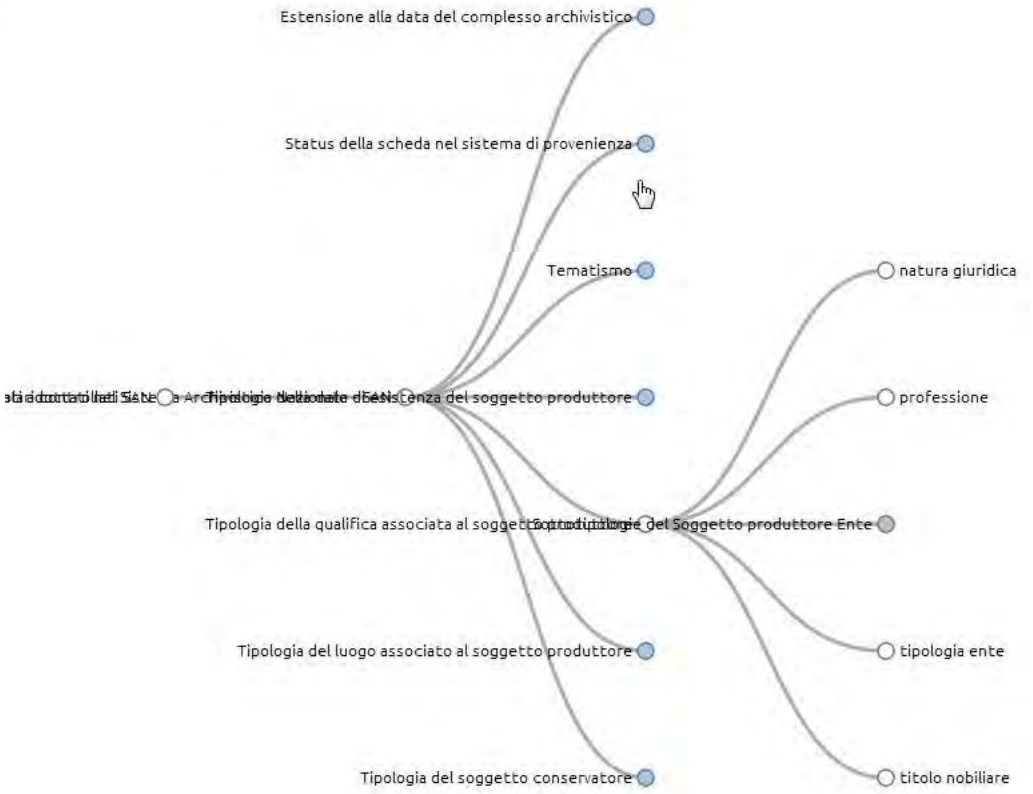
Descrizione estesa del concetto

Descrizione breve

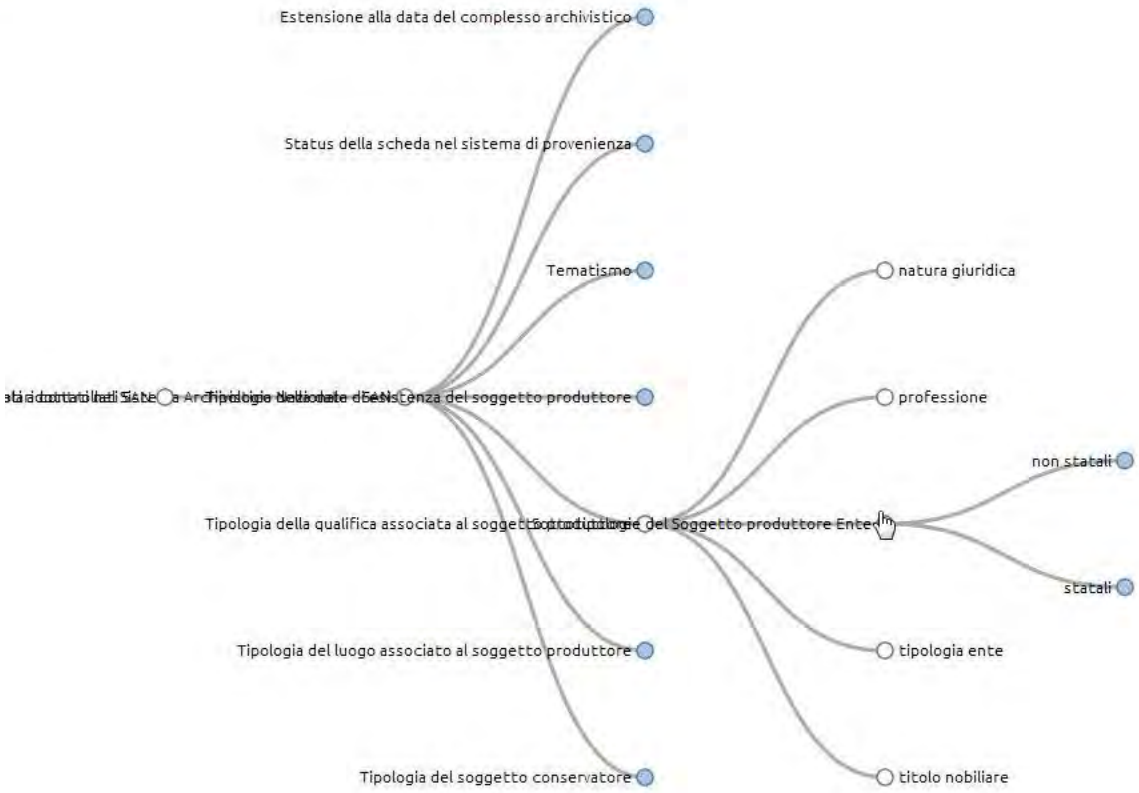
Il Tesauro SAN: modello di navigazione (2/4)



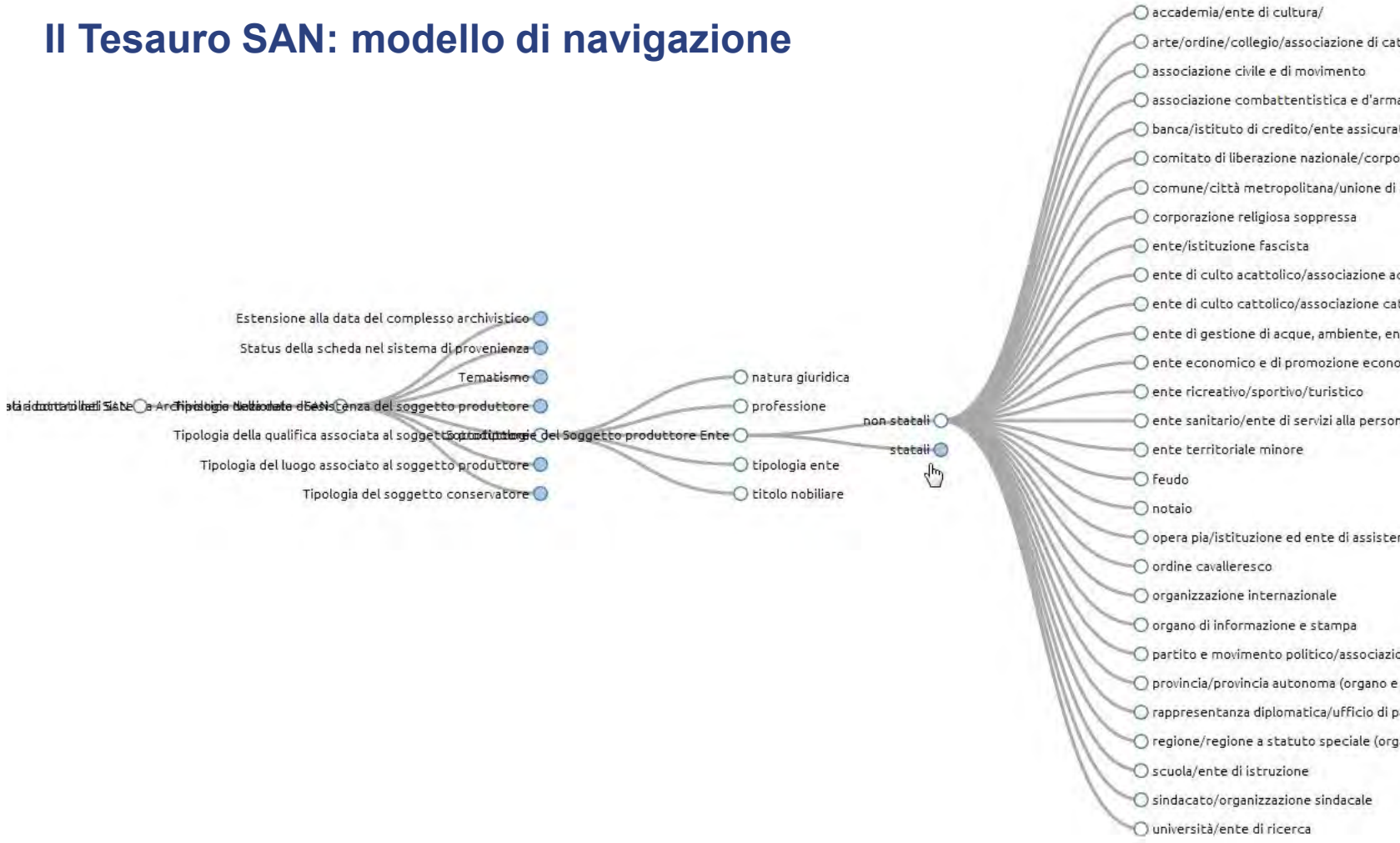
Il Tesauro SAN: modello di navigazione (3/4)



Il Tesauro SAN: modello di navigazione (4/4)



Il Tesauro SAN: modello di navigazione



Cronoprogramma realizzazione SAN Linked Open Data

Entro dicembre 2013

1. Ontologia di base
2. Estensioni all'Ontologia di base

Entro giugno 2014

3. Authority toponimi storici
4. Strumenti di accesso ai dati LOD

Per concludere: Il progetto di aggiornamento dei dati dei Soggetti Produttori e dei Profili istituzionali del Sistema Guida generale ha rappresentato per l'ICAR un primo passo verso l'adesione a questo modello tecnologico e concettuale e ha costituito il punto di riferimento anche per le scelte di definizione e presentazione dei **dati SAN LOD** (navigazione a faccette delle principali dimensioni informative, georeferenziazione, etc.)