

Sistema Archivistico Nazionale (SAN) e Open Data :

Il percorso intrapreso

<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
- intende far conoscere ad un pubblico non di soli specialisti **quali** risorse archivistiche esistano a livello nazionale, **chi** le ha prodotte, **dove** siano dislocate e **come** vi si acceda

Il SAN : principali caratteristiche e funzioni

- 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
- coopera con il **Portale europeo Apex**, **é interoperabile con CulturalItalia** il Portale della Cultura Italiana di MIBAC, aggregatore di risorse e documenti, provenienti da archivi, musei, biblioteche, soggetti pubblici e privati,
- Nell'ambito di un accordo di collaborazione, l'ICAR fornisce i metadati SAN utili a consentire l'**indicizzazione del SAN attraverso il motore di ricerca iStella**

Il SAN offre accesso alle seguenti tipologie di contenuti:

- **catalogo delle risorse archivistiche:** descrizioni dei soggetti 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 corrispondente 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** rilevanti per il dominio archivistico, con collegamento al Sistema Bibliotecario
- **percorsi, notizie, informazioni, testi e contenuti arricchiti** elaborati dalla redazione e da quanti aderiscono al SAN
- **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 **248.535 risorse**

185.845 Risorse Archivistiche, 2.510 Contenuti redazionali, 2.907 Risorse Bibliografiche, 40.441 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*** **40.441**
 - ***Totale*** **226.286**
-

Sistemi cooperanti

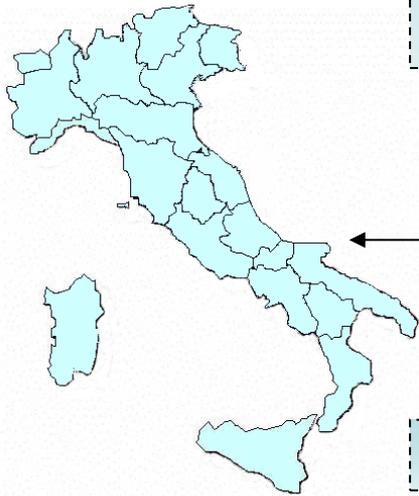


Regioni

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



SAN
SISTEMA ARCHIVISTICO NAZIONALE

CEI-AR

Portali tematici

Impresa
Per non dimenticare
Musica
Architetti
Moda

Archivio Storico Multimediale
del Mediterraneo
Territori
Antenati
...

Portali tematici

Il SAN integra, tra l'altro, a vario titolo, alcuni portali tematici dell'Amministrazione Archivistica:

- **Archivi di impresa:** fonti archivistiche, bibliografiche, testi, immagini, audio, video dal 1861 a oggi, provenienti da archivi di piccole, medie e grandi 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://www.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>
- **Territori :** fonti archivistiche catastali e cartografiche <http://www.territori.an.beniculturali.it>
- **Archivio Storico Multimediale del Mediterraneo ASMM** <http://www.archividelmediterraneo.org>



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

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**

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

1. Definire una **Ontologia di base** che rappresenti lo schema concettuale della struttura informativa espressa dai tracciati di scambio CAT SAN (standard di rappresentazione dei dati SAN) previa analisi delle esperienze nazionali ed internazionali che hanno affrontato o stanno affrontando, pur con sostanziali peculiarità legate ai rispettivi contesti di applicazione, analoghe problematiche descrittive
2. Definire 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) anche se non strettamente archivistiche
3. Realizzare **liste di authority, repertori, strumenti** su scala nazionale da mettere a disposizione della comunità archivistica
4. Rendere disponibili una serie di **strumenti di accesso ai dati LOD** prodotti (endpoint sparql, oai provider, set specifici di dati LOD, etc.)

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

5. Realizzare una sperimentazione prototipale – condotta inizialmente su dati provenienti dal Sistema Guida generale per ciò che attiene ai soggetti produttori e dal SIAS per ciò che attiene ai complessi archivistici – 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.

6. La strategia progettuale prevede di sperimentare in questa fase le possibili 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.

Cosa sono i Linked Open Data

Con l'espressione **Linked Open Data** (LOD) ci si riferisce ad 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).

Scopo dei Linked Open Data è di estendere il Web creando una rete di dati aperti e disponibili a tutti sfruttando il formato RDF per creare in automatico insiemi di dati provenienti da diverse sorgenti.

L'espressione **Linked Data** fu introdotta nel 2006 da Tim Berners Lee in una pubblicazione dedicata al Web semantico, con riferimento ad una serie di buone pratiche costituenti i diversi gradi di una scala di valutazione (da 1 a 5 *stelle*) dell'effettiva apertura e accessibilità dei dati.



LOD – i dati «5 stelle»

1. Il dato è **disponibile sul web** in qualsiasi formato purché con una licenza aperta
2. Il dato è disponibile in **un formato strutturato** che può essere interpretato da un software (es. excel)
3. Il dato è disponibile sul web in un formato strutturato **non proprietario** (es. csv anziché excel)
4. Il dato rispetta i criteri precedenti facendo uso di **standard aperti** definiti da W3C (quali **RDF** e **SPARQL**) per identificare e reperire le risorse
5. Il dato – che rispetta tutti i criteri precedenti - **contiene collegamenti ad altri dati** in modo che si possano scoprire ulteriori oggetti e rendere più visibili i propri

I principi LOD sono stati adottati da un numero sempre crescente di data provider pubblici e privati e riguardano ormai qualunque settore o dominio: dai dati governativi e statistici a quelli enciclopedici, dalle news giornalistiche ai dati sanitari, dai dati scientifici ai contenuti geografici. Specialmente dalla comunità del progetto Linked Data e dalla *task force* del W3C Linking Open Data, che indicizza tutti i dataset disponibili (esponendoli nella cosiddetta “nuvola LOD”).

Open Data e SAN : gli obiettivi del progetto

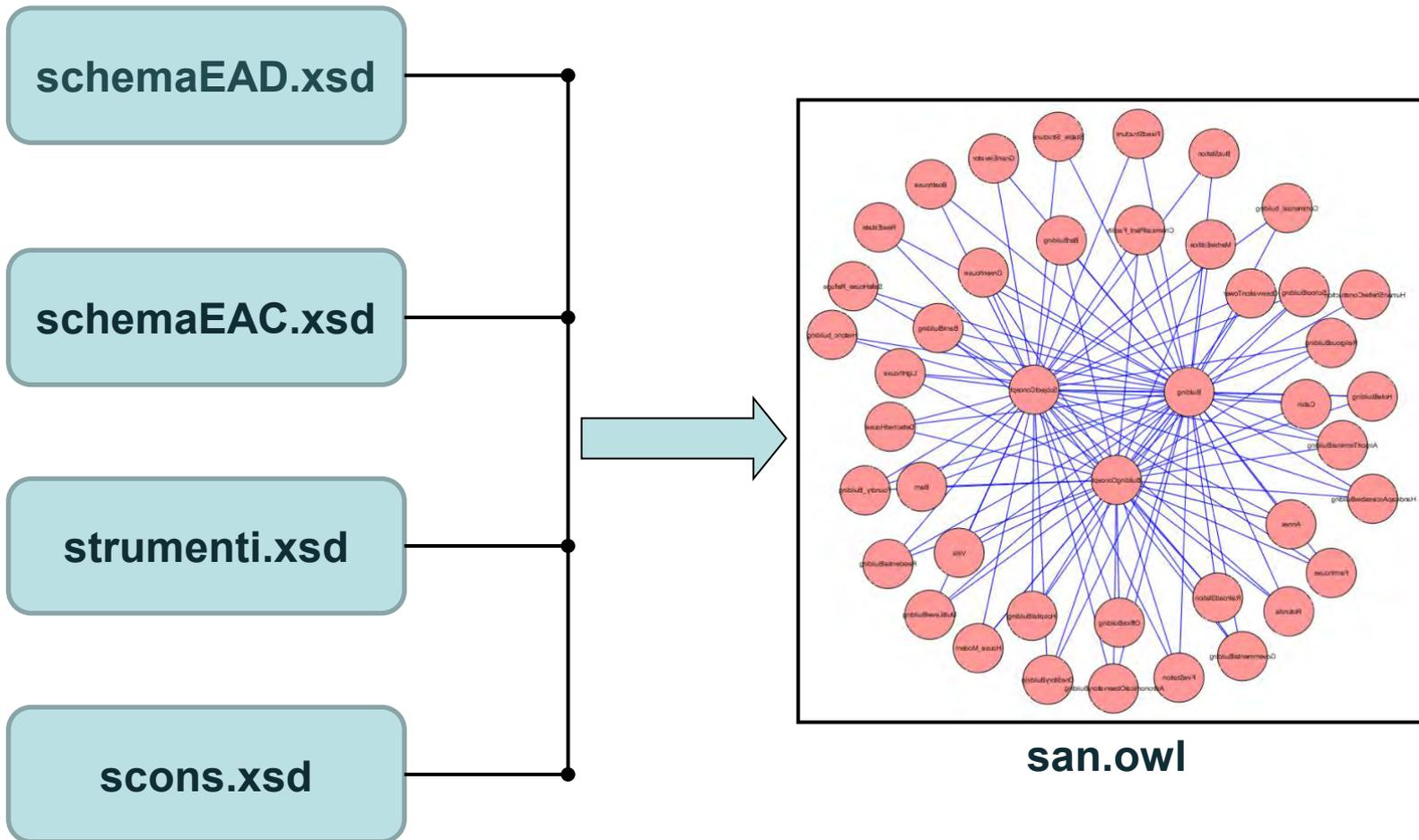
1. Creazione di un **modello concettuale del SAN** formalizzato in una ontologia espressa in linguaggio OWL corrispondente alla struttura informativa veicolata dagli schemi xsd definiti nell'ambito del SAN per la descrizione degli oggetti archivistici

Tali XML schema sono stati rilasciati nel 2009 da una apposita Sottocommissione tecnica per la definizione dei metadati relativi alle risorse archivistiche accessibili attraverso il Sistema Archivistico Nazionale come tracciato di scambio proposto ai sistemi aderenti per il conferimento al **Catalogo delle risorse archivistiche – CAT** dei propri dati relativi a complessi archivistici, soggetti produttori, strumenti di ricerca e soggetti conservatori.

La definizione della ontologia di base sulla scorta del modello rappresentato dai tracciati di scambio CAT SAN consentirà di ottimizzare al massimo gli sforzi già prodotti da tutti i soggetti a vario titolo coinvolti nel SAN; solo per gli oggetti e le relazioni non contemplati nei tracciati di scambio e definiti nelle “estensioni” della ontologia sarà infatti necessario prevedere una ulteriore attività di integrazione previa mappatura con lo schema concettuale definito.



Open Data e SAN : ontologia di base





Open Data e SAN : ontologia di base

Nella definizione dell'ontologia del Sistema Archivistico Nazionale verranno presi in considerazione tanto gli standard di dominio – peraltro adottati dalla Sottocommissione per la rappresentazione delle informazioni archivistiche e di contesto - 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#>

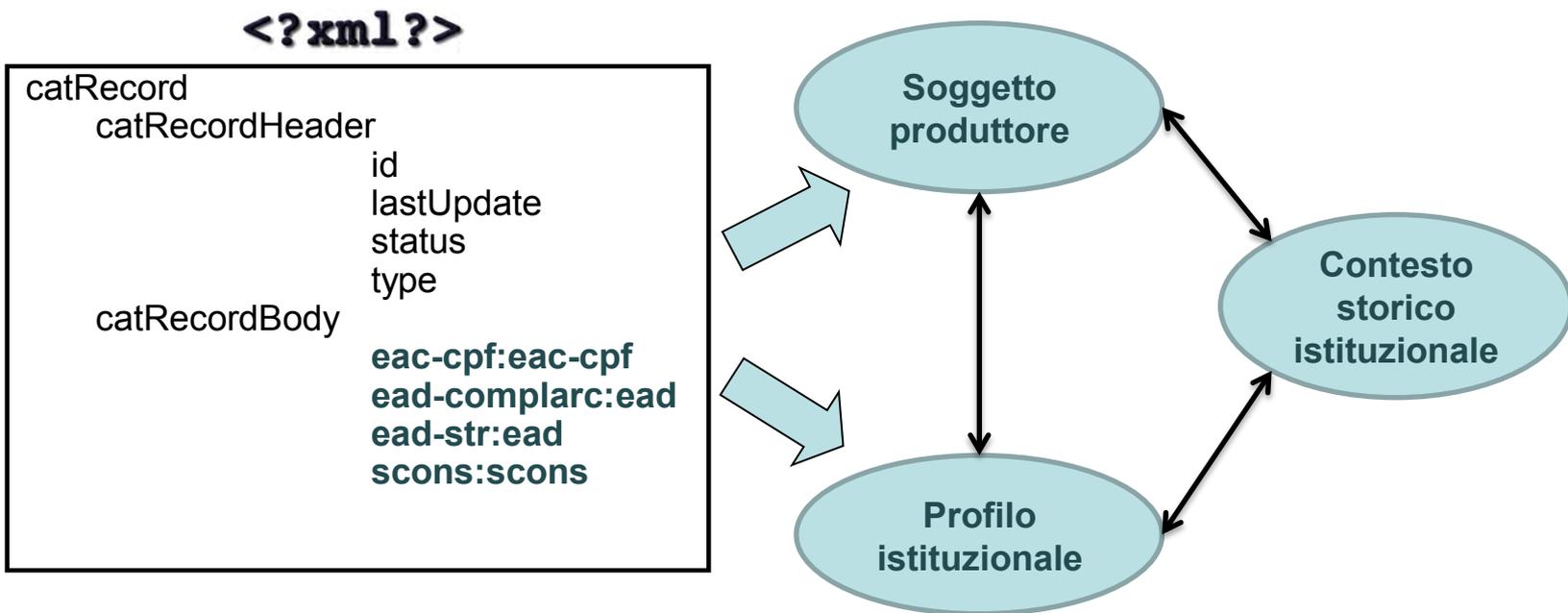
Open Data e SAN : gli obiettivi del progetto

- 
2. Sperimentazione prototipale, condotta su uno dei sistemi aderenti (Sistema Guida generale) per ciò che attiene ai soggetti produttori, di un metodo concettuale e operativo di **estensione di tale struttura informativa originaria** con classi di oggetti e relazioni tra risorse attualmente esterni al SAN

A partire da un rappresentativo subset di dati del Sistema Guida generale - che nel loro complesso impianto di relazioni alimentano l'Atlante storico - istituzionale (soggetti produttori, profili istituzionali e contesti storico istituzionali) - verranno definite delle estensioni alla ontologia SAN di base integrandola con entità e relazioni tra oggetti non previste in sede di elaborazione dei tracciati del CAT.

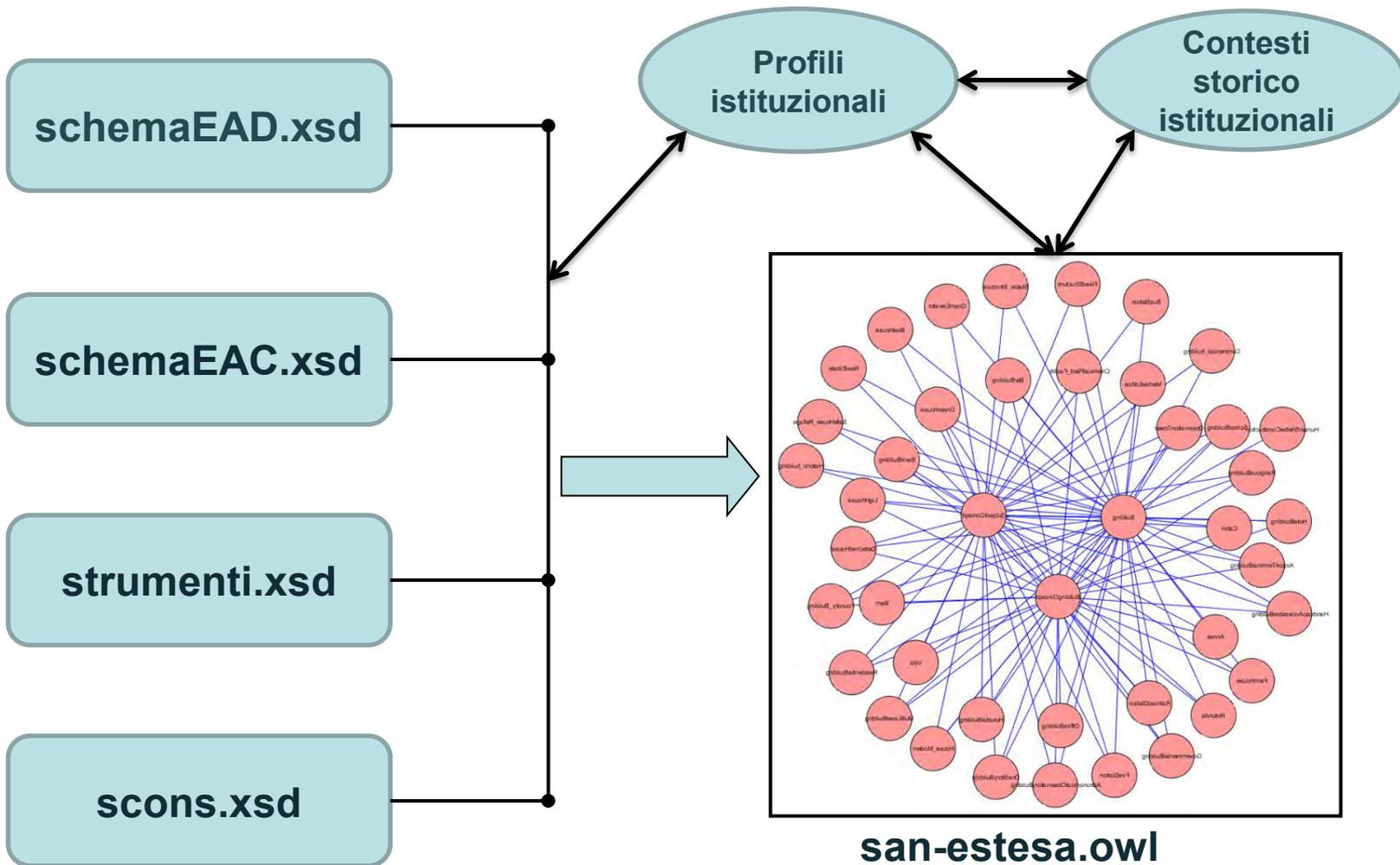
Open Data e SAN : l'ontologia "estesa"

Una prima estensione dell'ontologia SAN prevede il suo arricchimento con **due ulteriori classi di oggetti** (profilo istituzionale e contesto storico istituzionale) e con tutte quelle proprietà necessarie a descriverle e a metterle in relazione coi soggetti produttori.





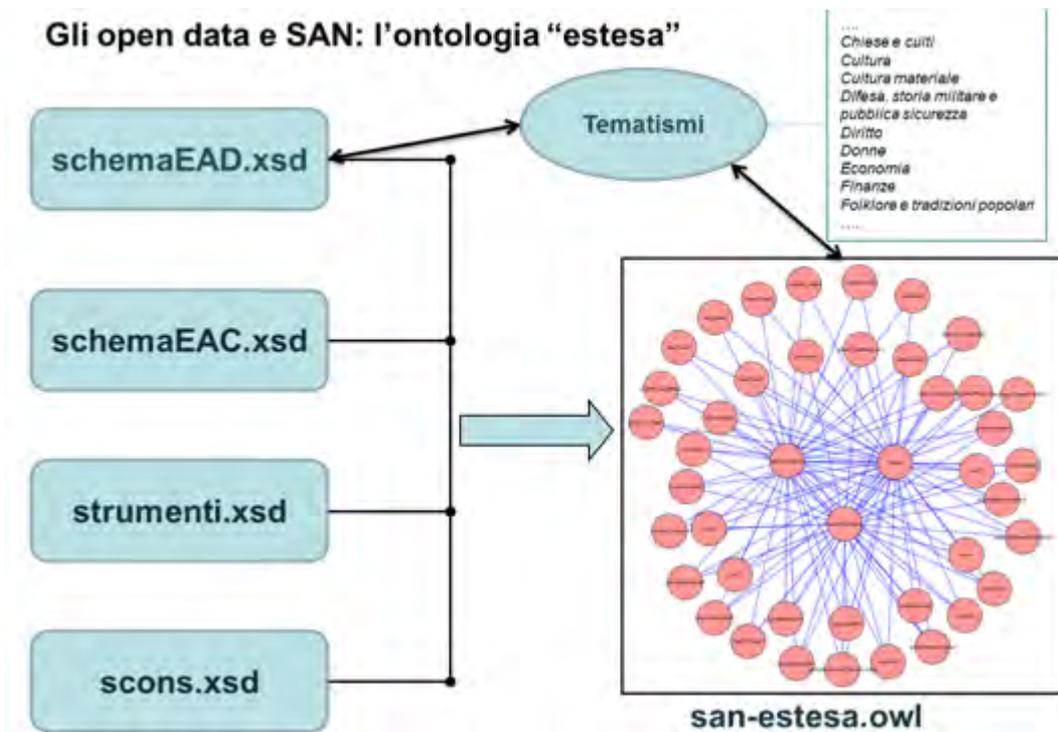
Open Data e SAN : l'ontologia "estesa"



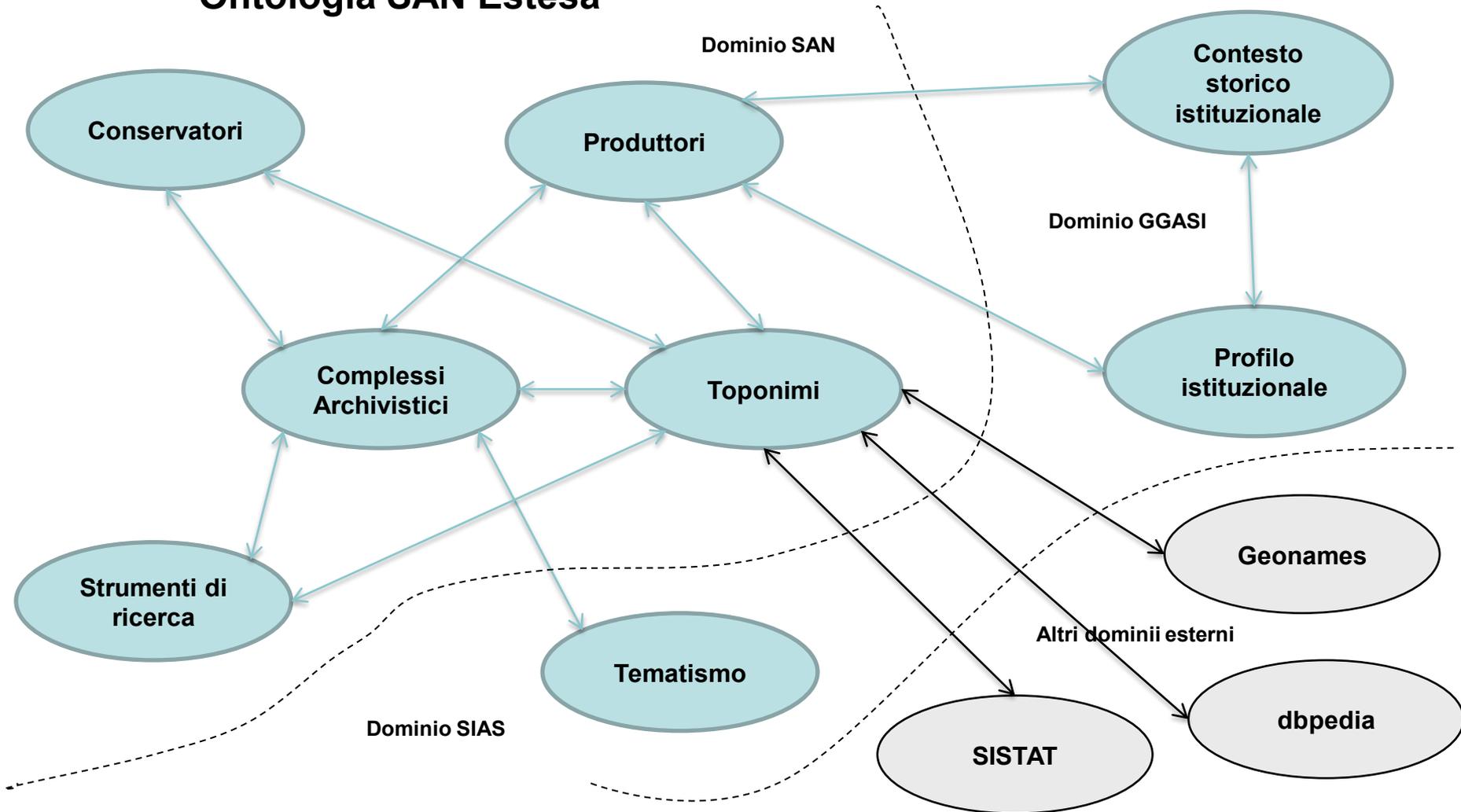


Open Data e SAN : l'ontologia "estesa"

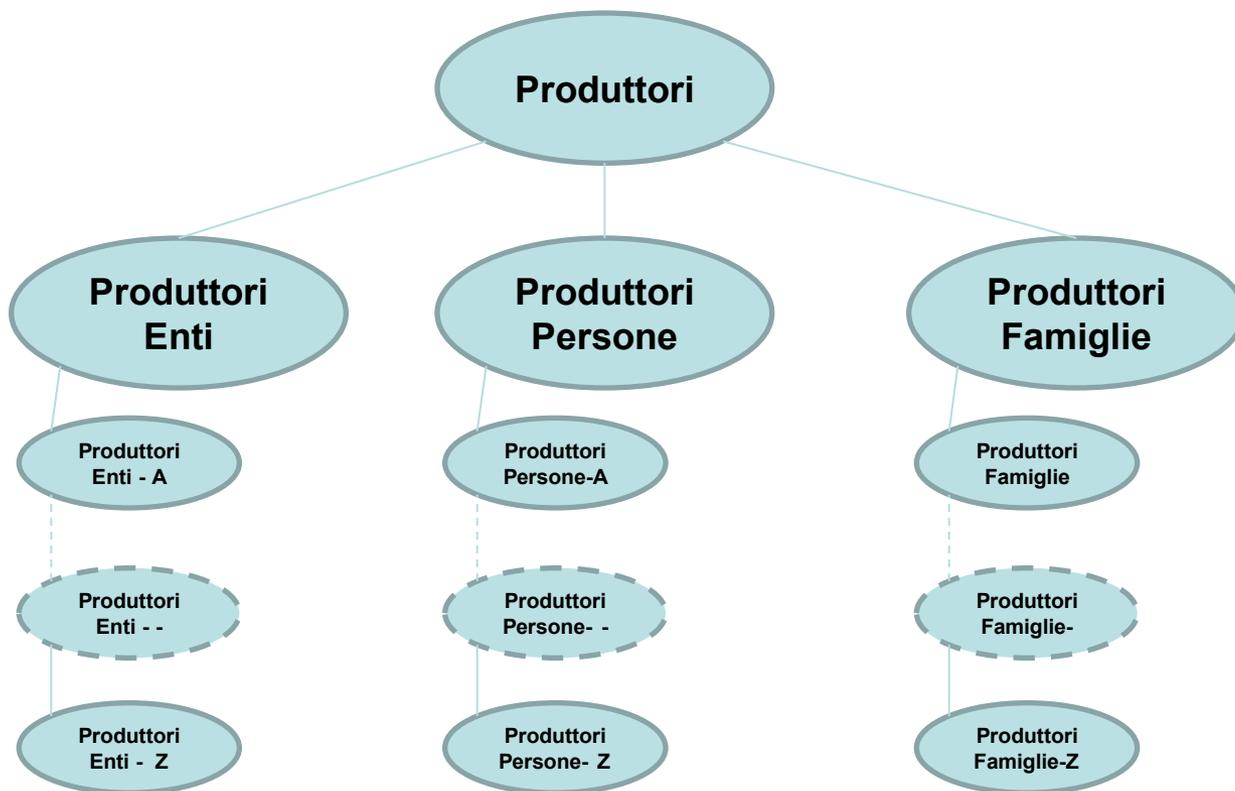
Una seconda estensione dell'ontologia SAN potrà riguardare l' **inclusione nel modello concettuale di quelle chiavi d'accesso** ai complessi archivistici organizzate per aree tematiche predisposte dal Sistema Informativo degli Archivi di Stato – SIAS, le quali potranno in prospettiva essere accolte e ricondotte anche a risorse archivistiche fornite da altri sistemi aderenti, potenziando e facilitando oltremodo i percorsi di ricerca.



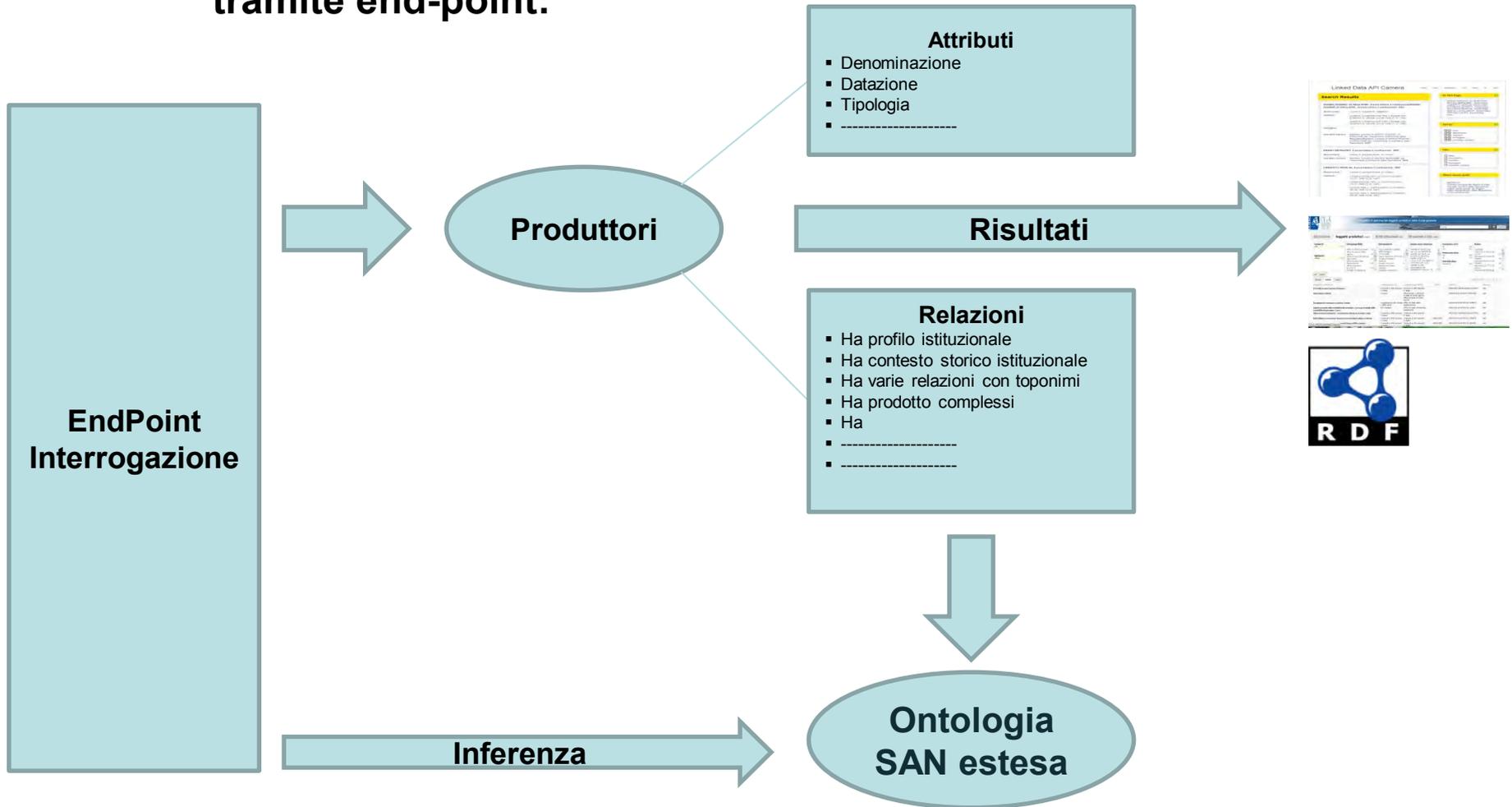
Ontologia SAN Estesa



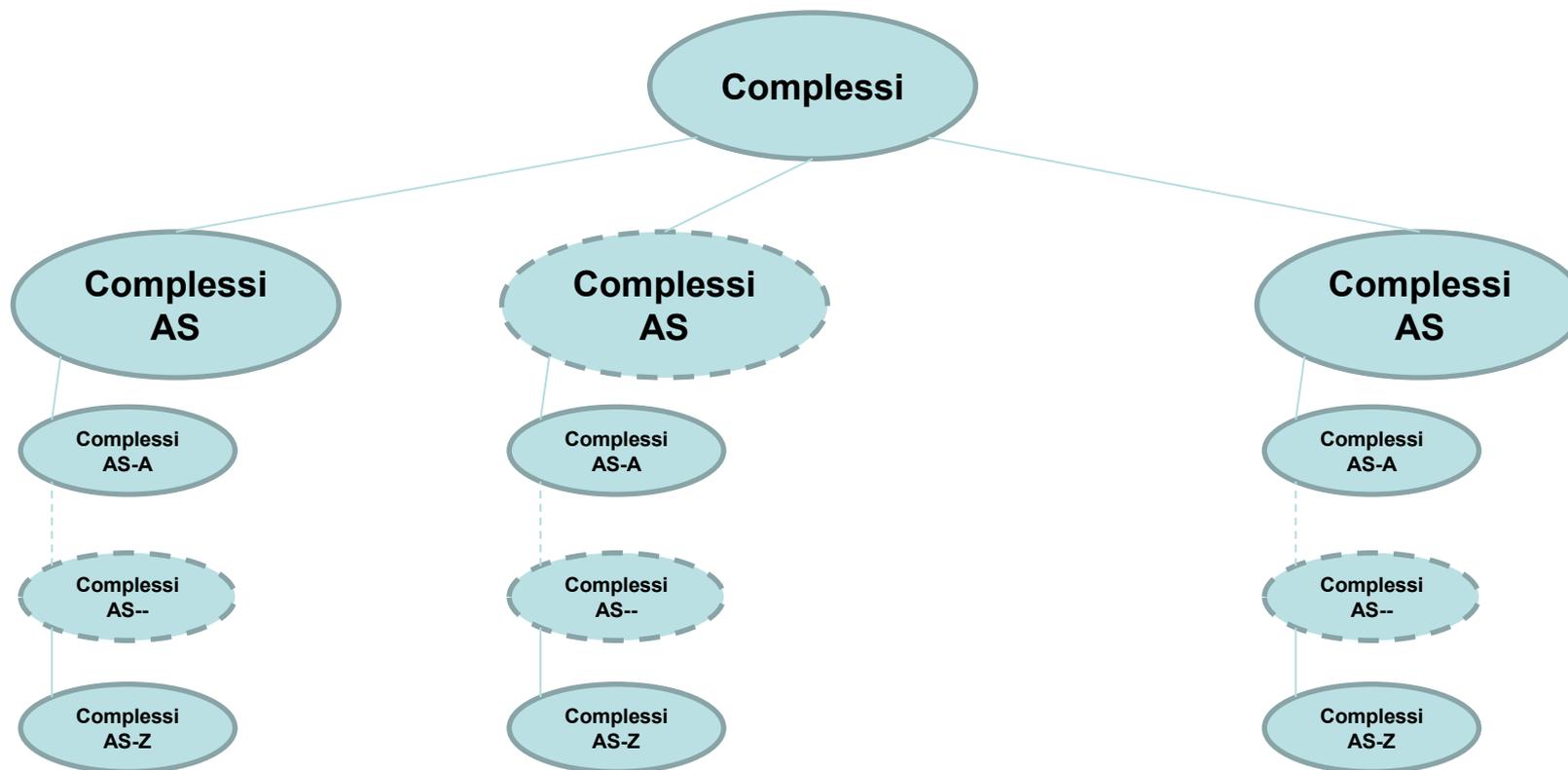
I SET LOD disponibili per i Soggetti Produttori per la navigazione gerarchica.



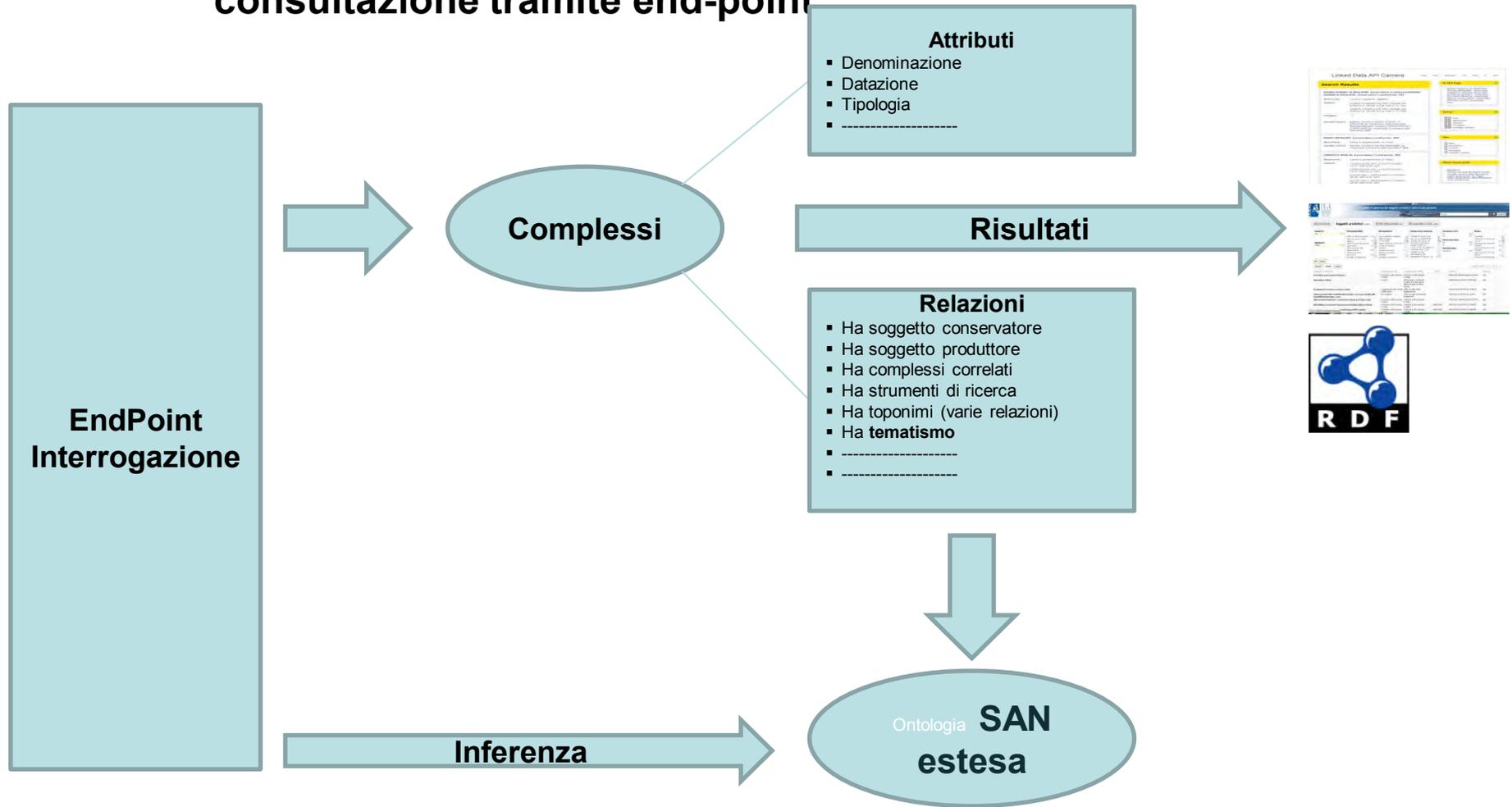
I SET LOD disponibili per i Soggetti Produttori per consultazione tramite end-point:



I SET LOD disponibili per i Complessi Archivistici suddivisi per Archivio di Stato di appartenenza.



I SET LOD disponibili per i Complessi Archivistici per consultazione tramite end-point:



Open Data e SAN : gli obiettivi del progetto

3. Elaborazione di una lista di autorità di **toponimi** su scala nazionale, basata principalmente sulla **rielaborazione in formato LOD** di dati istituzionali di fonte ISTAT : **Sistema Informativo Storico delle Amministrazioni Territoriali (SISTAT)** integrabili con fonti di particolare rilevanza storica, che sarà sperimentalmente raccordata coi dati SAN LOD di provenienza Guida Generale degli Archivi di Stato.

Anche questa operazione si configura come una estensione della ontologia di base SAN, che immette nel sistema nuova “conoscenza” attraverso l’integrazione di un repertorio di informazioni storico-geografiche di estrema rilevanza per l’ampliamento della portata informativa dei dati archivistici.



Open Data e SAN : toponimi e storia

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.

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)

Open Data e SAN : toponimi e storia

La fonte istituzionale ISTAT

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) con un aggiornamento dichiarato a marzo 2013.

Open Data e SAN : gli obiettivi del progetto

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

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

Disponibilità di un **OAI Provider** che distribuisca dati pubblici secondo gli schemi CAT – SAN.

Open Data e SAN : gli obiettivi del progetto

L'Architettura per l'erogazione dei servizi



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

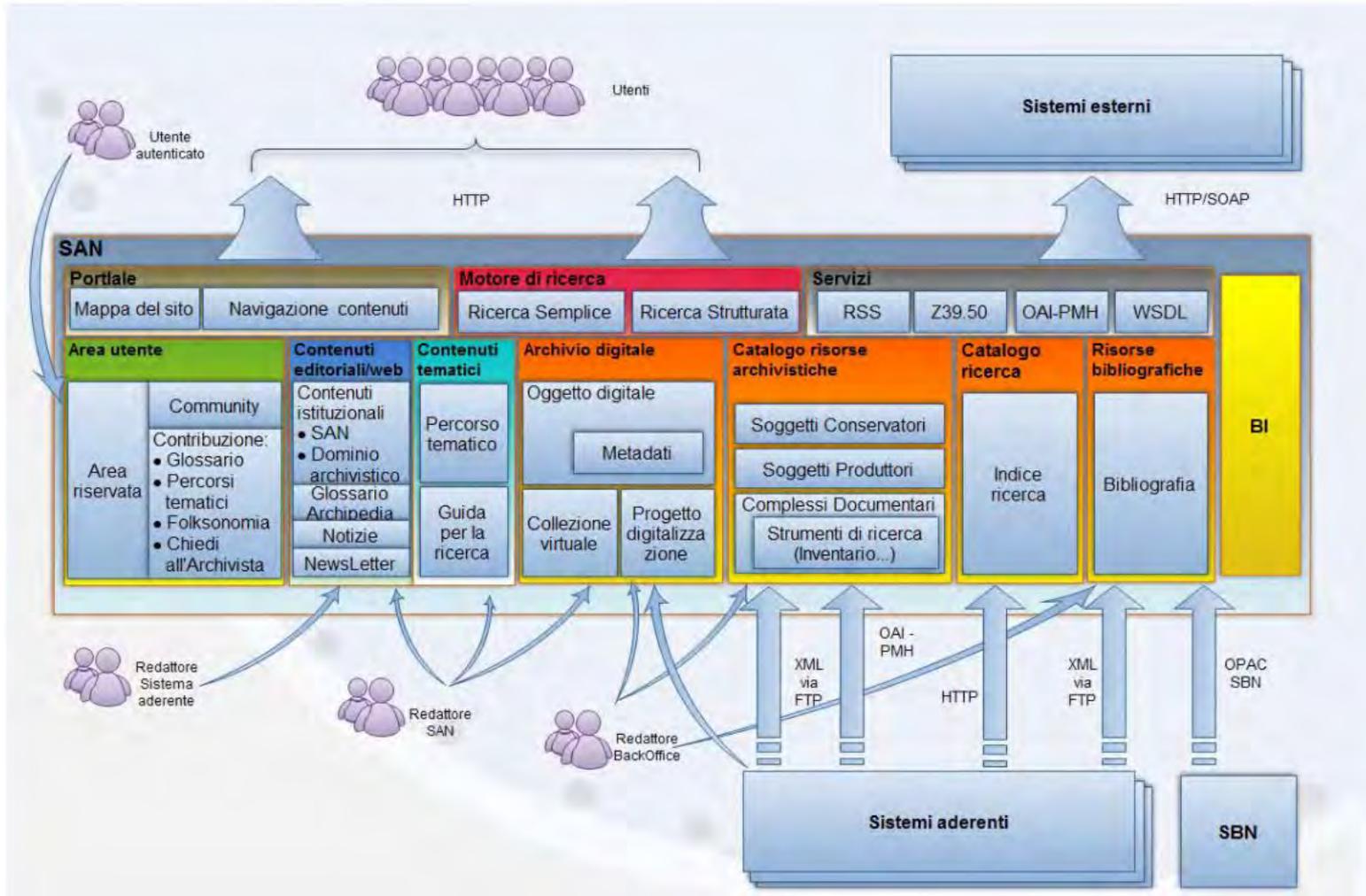


Un **OAI Provider** che distribuisca dati pubblici secondo gli schemi CAT – SAN.



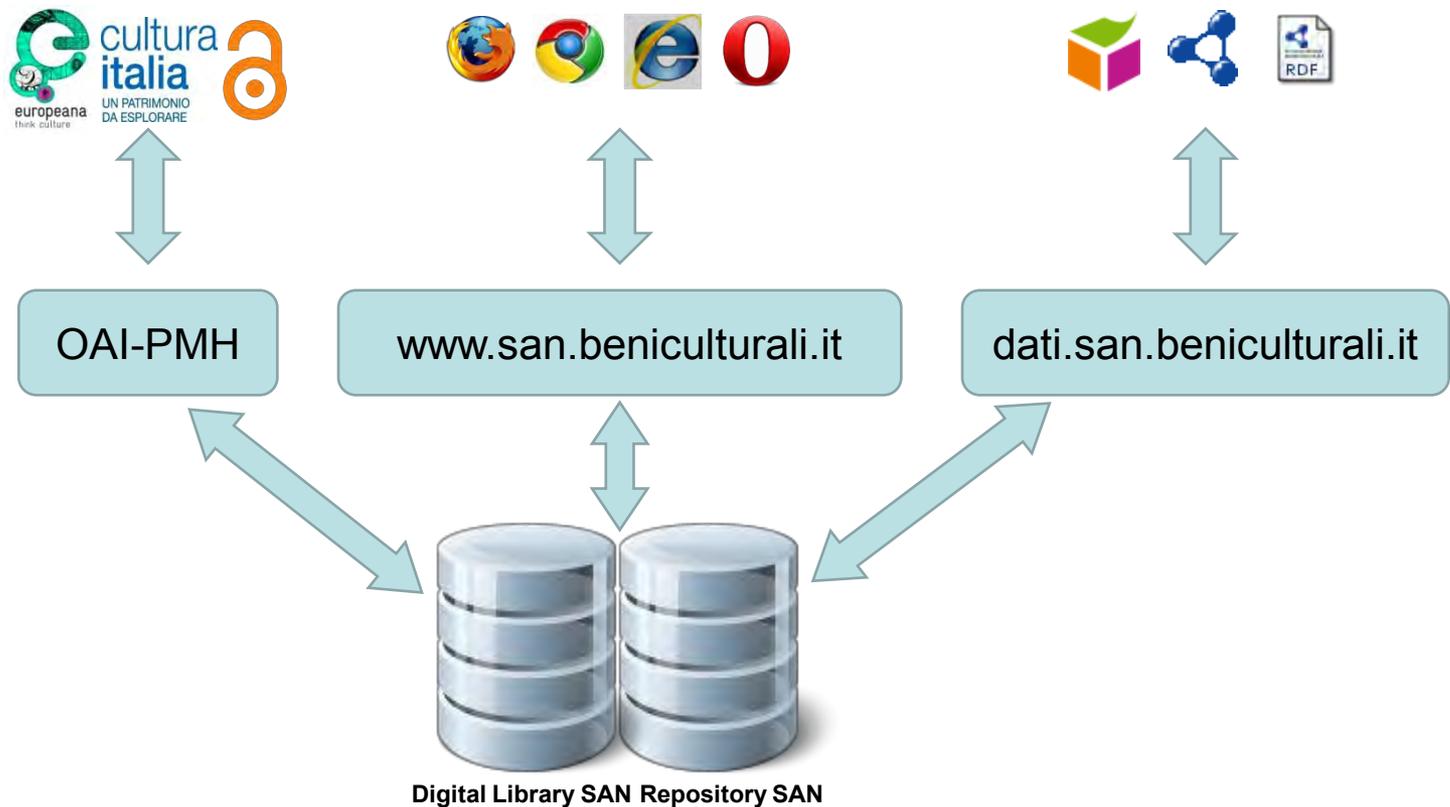
Una **area di download** dove poter scaricare i vari set di LOD resi disponibili dal SAN.

Modello SAN

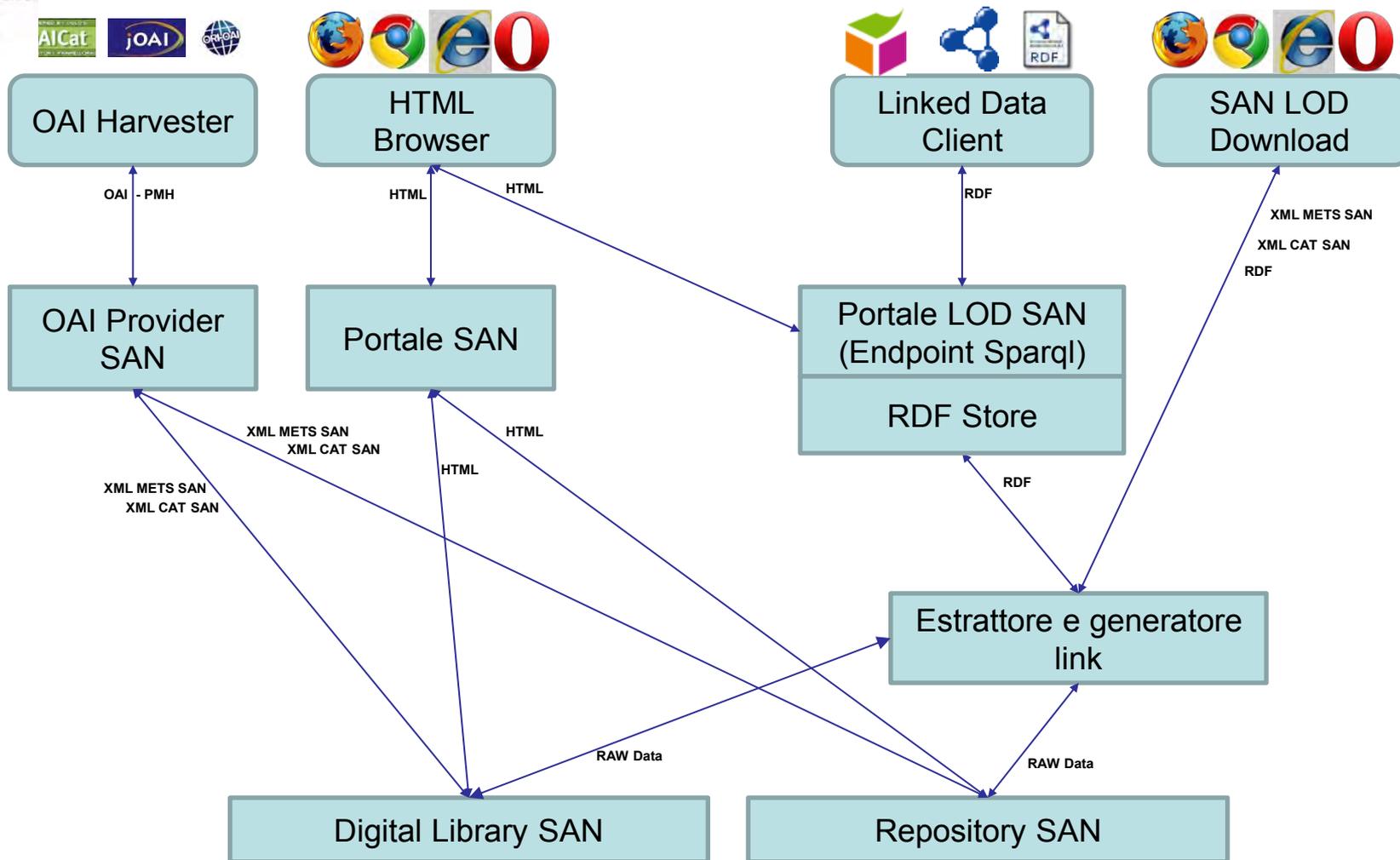




Strumenti di accesso ai dati LOD



Strumenti di accesso ai dati LOD : Flussi





Cronoprogramma per la realizzazione del percorso Open Data per il SAN

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 costituirà il punto di riferimento anche per le future scelte di presentazione dei dati SAN LOD (navigazione a faccette delle principali dimensioni informative, georeferenziazione, etc.)