



**ICCU – ICAR**

**Servizi e infrastrutture nazionali per la gestione del patrimonio archivistico e bibliografico**

## **Processi e modelli formali per la gestione integrata e il riuso del patrimonio informativo archivistico**

**Stella Di Fazio**

**Istituto centrale per gli archivi**

**Milano, Fondazione Stelline, 15 marzo 2019**

## Il patrimonio informativo archivistico

---

L'informazione relativa ai patrimoni documentari è sempre stata caratterizzata, rispetto a quella relativa al patrimonio bibliografico, da una evidente **frammentazione descrittiva**, principalmente conseguenza della stessa **natura eterogenea delle risorse** complessivamente riconducibili al concetto di fonte archivistica.

A questa realtà variegata e plurale e non facilmente governabile in modo unitario contribuiscono inoltre, localmente, **scelte di formalizzazione, rappresentazione e gestione** di tali risorse non sempre allineate ai modelli proposti in sede centrale.

## L'informazione nei sistemi archivistici nazionali

---

L'ICAR gestisce per mandato istituzionale sistemi informativi che detengono informazione relativa al patrimonio archivistico nazionale.

Allo stato attuale i **sistemi nazionali** si alimentano sia – direttamente – attraverso specifici moduli di *back-end*, sia in virtù di previste procedure di **riversamento** di risorse informative residenti in sistemi nativi locali.

La predisposizione di **strumenti e meccanismi di interoperabilità** sintattica e semantica tra le diverse realtà che nel loro complesso compongono il patrimonio informativo di dominio costituisce pertanto un tassello prioritario nella strategia operativa dell'Istituto.

---

## Modelli di gestione e valorizzazione organica del patrimonio archivistico

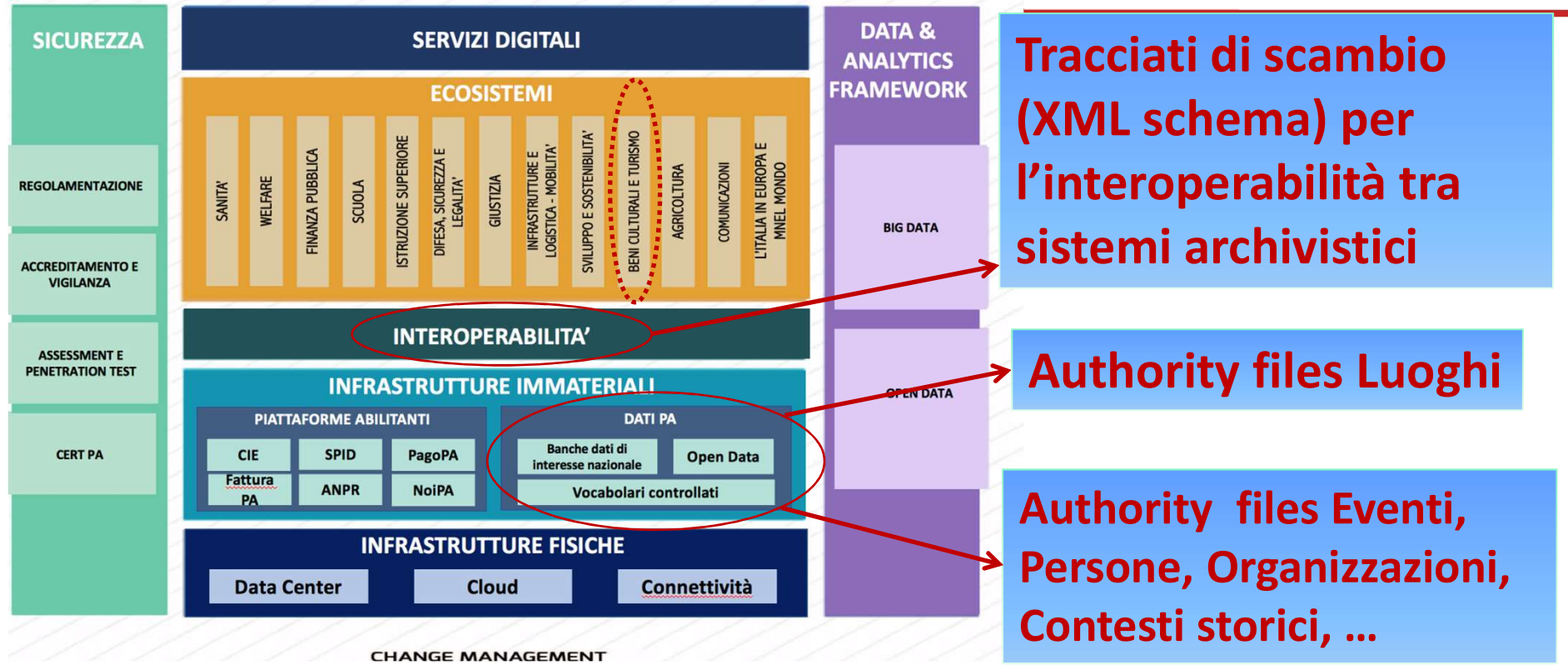
---

Si tratta di un **modello operativo** che ambisce principalmente a sostenere e a facilitare una **gestione organica del patrimonio informativo archivistico**, consentendone l'integrazione, la distribuzione, la fruizione e il riuso in **modalità open e standard**.

Alla base di tale modello si trovano:

- schemi di metadati definiti come **tracciati di interoperabilità** che consentono lo scambio di informazioni tra sistemi diversi;
- specifiche infrastrutture informative (**infrastrutture immateriali** nella terminologia adottata dal *Piano triennale per l'informatica nella PA*) che possono fungere da **authorities** di riferimento nelle attività di descrizione del patrimonio culturale.

# Strategia ICAR vs Piano triennale per l'informatica nella PA



## L'interoperabilità tra i sistemi: i tracciati di scambio

---

I tracciati di interoperabilità sono un **pacchetto integrato di formati di scambio** elaborati dall'ICAR allo scopo di consentire un elevato livello di **interoperabilità fra i principali applicativi e sistemi di descrizione archivistica** in uso in Italia e agevolare l'esportazione e l'importazione dei dati da e verso tali software e sistemi.

I tracciati definiscono la **struttura informativa delle entità** che nel loro complesso caratterizzano il dominio della descrizione archivistica individuandone puntualmente attributi e relazioni, **formalizzati in XML** conformemente agli XML schema dei principali standard descrittivi internazionali o a modelli dati XML elaborati per il SAN dall'ICAR.

## L'interoperabilità tra i sistemi: i tracciati di scambio

Elemento	Descrizione
Tipologia entità ( <b>Obbligatorio</b> )	<p><b>Path XML:</b>            eac-cpf/cpfDescription/identity/entityType</p> <p><b>Note d'uso:</b>            Per il soggetto produttore di tipo ente si introduce il valore "corporateBody".            Si veda quanto indicato nella Premessa per l'indicazione della tipologia di scheda; in questo caso l'attributo @localType assume il valore "soggettoProduttore".</p> <p><b>Esempio:</b>  <code>&lt;identity localType="soggettoProduttore"&gt;</code>  <code>&lt;entityType&gt;corporateBody&lt;/entityType&gt;</code></p>
Denominazione ( <b>Obbligatorio</b> )	<p><b>Path XML:</b>            eac-cpf/cpfDescription/identity/nameEntry/part</p> <p><b>Note d'uso:</b>            Si indica nell'attributo @lang il codice ISO 639-2 della lingua in cui è espressa la denominazione.</p> <p><b>Esempio:</b></p>

## L'interoperabilità tra i sistemi: i tracciati di scambio

---

### Entità rappresentate nei tracciati:

complessi archivistici, unità/sottounità archivistiche, unità documentarie, soggetti produttori, soggetti conservatori, strumenti di ricerca, entità/voci d'indice.

La struttura informativa di base è quella dei **tracciati CAT-SAN**, in quanto schemi di riferimento consolidati già adottati dai principali sistemi/software descrittivi per interoperare con i sistemi nazionali.

I tracciati CAT-SAN sono stati **rivisti e ampliati** per includervi metadati ritenuti necessari per garantire la massima interoperabilità fra i sistemi di descrizione archivistica e perseguire al meglio **l'integrità informativa originaria**.

---



## L'interoperabilità tra i sistemi: i tracciati di scambio

---

### Gli standard di metadattazione: EAD3, EAC-CPF, METS

Enfatizzando una scelta già propria dei tracciati CAT-SAN, la formalizzazione dei tracciati di interoperabilità fa perno su modelli dati in **formato XML** che costituiscono gli standard internazionali di riferimento per la descrizione delle entità peculiari del dominio archivistico nonché della documentazione in formato digitale.

- **EAD3**: qualunque livello di descrizione archivistica; strumenti di ricerca;
- **EAC-CPF**: produttori di archivio ed in genere tutte le entità coinvolte nel contesto di produzione della documentazione;
- **METS**: metadati strutturali, descrittivi e amministrativi per la gestione degli oggetti digitali

## L'interoperabilità tra i sistemi: i tracciati di scambio

---

Per quanto riguarda **EAD**, i tracciati sono stati allineati all'ultima versione dello standard caratterizzata dalla maggiore granularità descrittiva di numerose componenti informative (datazione e descrizione fisica in primis) e dall'obiettivo di avvicinarsi a una modalità di **rappresentazione formale dell'informazione archivistica** di tipo **risorsa-centrico**, svincolabile cioè dallo strumento di ricerca (inventario, guida, etc.) in cui la risorsa descritta si trova gerarchicamente posizionata e con la possibilità di **gestire esplicitamente le relazioni** interne ed esterne al dominio attraverso l'utilizzo di un nuovo tag **<relations>**, mutuato dallo schema **EAC-CPF**

→ **influsso del Semantic Web / Linked Open Data**

---

## Tracciati per l'interoperabilità (ma non solo)

---

Promuovendo l'adozione dei tracciati di interoperabilità all'interno dell'amministrazione archivistica l'ICAR intende anche favorire l'**adozione di standard aperti** per la codifica delle informazioni archivistiche, e ridurre così i rischi di perdita di dati o di loro difficile recupero in caso di obsolescenza tecnologica o abbandono dei software in uso, evitando il noto problema del *lock-in* e adeguandosi ai **principi di economicità ed efficienza** stabiliti da normativa e Linee guida:

- **art. 68 CAD:** "Analisi comparativa delle soluzioni"
- **art. 69 CAD:** "Riuso delle soluzioni e standard aperti"
- **Linee Guida** su acquisizione e riuso di software per le PA di **AgID**

## Le infrastrutture immateriali: authority files e vocabolari controllati

---

Authority files e vocabolari controllati costituiscono **architetture informative di supporto alla descrizione e alla contestualizzazione del patrimonio archivistico**, che agevolano la normalizzazione e il controllo d'autorità dei dati evitando la frammentazione degli interventi e quindi la duplicazione delle attività.

Per la loro **centralità nel processo descrittivo** è opportuno che vengano formalizzati in modo tale da risultare facilmente integrabili e/o referenziabili nei diversi contesti informativi.

L'ICAR sta realizzando progressivamente queste infrastrutture informative ricorrendo ai formalismi e alle tecnologie del **Semantic Web**.

---

## **Atlante storico istituzionale dell'Italia unita**

---

L'**Atlante storico istituzionale dell'Italia unita (ASI)** è stato realizzato dall'ICAR in collaborazione con il Centro MAAS attraverso varie fasi di intervento che hanno condotto alla pubblicazione nel maggio 2018.

**Obiettivo:** rendere disponibile e fruibile in un **contesto di uso aperto e interoperabile** un'ingente mole di informazione storico-territoriale reperita in fonti autorevoli, direttamente in formato semantico (**RDF OWL**), riconducendola complessivamente ad un **modello concettuale unitario**, capace di evidenziare nei dati sia la dimensione **diacronica** che una visione **sincronica**.

## Atlante storico istituzionale dell'Italia unita

---

Strumento di valenza generale che ricostruisce dinamicamente, registrandone le **evoluzioni nel tempo**, le **circoscrizioni territoriali** dell'Italia postunitaria, a partire dalle circoscrizioni **amministrative** (base informativa di questo primo esito pubblicato)

Il modello ontologico definito costituisce un **repertorio di autorità di nomi di luogo** attestati nel tempo (**storicizzati**) e **georeferenziati**, che può rappresentare un concreto e autorevole supporto per denominare, descrivere e localizzare correttamente **le istituzioni e le fonti**, e può costituire riferimento di base in qualunque operazione descrittiva nella quale risulti rilevante individuare aspetti legati all'**evoluzione storica del territorio**.

---

## Atlante storico istituzionale dell'Italia unita

---

La base di conoscenza è stata alimentata attraverso un lungo processo di studio, analisi e recupero dei **dati sulle variazioni territoriali** dal 1861 al 2013, pubblicati come open data dall'ISTAT (→ **SISTAT**). Questi dati sono stati **rivisti, incrementati** e poi **rimodulati in un'ontologia in formato OWL** in cui sono confluite informazioni relative ad altri livelli amministrativo-territoriali provenienti da dizionari storici accreditati.

Attualmente l'ontologia espone i **dati relativi a regioni, province e comuni dello Stato italiano**, dall'unità ai giorni nostri; i dati relativi a **circondari e mandamenti** sono stati invece inseriti solo per gli anni 1861-1862, con l'obiettivo di completarne progressivamente il quadro storico arrivando fino al 1927.

## Atlante storico istituzionale dell'Italia unita

---

### Sistema aperto e in espansione:

il modello è stato pensato per essere progressivamente esteso con entità, concetti e dati relativi ad **altre tipologie di circoscrizioni territoriali** (giudiziarie, ecclesiastiche, militari, ecc.), per realizzare una sempre più esaustiva rappresentazione dell'evoluzione storico istituzionale italiana rapportata al territorio.

L'ICAR sta per avviare una nuova fase operativa del progetto indirizzata allo studio e all'analisi delle **circoscrizioni territoriali di carattere giudiziario** e della loro evoluzione nel tempo:

- **estensione** dell'ontologia OWL
- **integrazione** in ASI dei dati storico territoriali riferibili a tale tipologia.



# Atlante storico istituzionale dell'Italia unita

---

## ASI – vantaggi della rappresentazione in formato RDF dei dati

- aver convogliato e reso immediatamente disponibili in un **unico punto di accesso online** informazioni complesse, in parte già fornite dal SISTAT, ma in un sistema chiuso e poco usabile
  - possibilità di riferirsi a ciascuna delle **unità territoriali** descritte dall'Atlante come **risorsa RDF referenziata da un URI**, richiamabile all'interno della descrizione di qualunque risorsa culturale (incamerando così tutte le informazioni storico territoriali presenti nell'Atlante per quella risorsa);
  - **interdisciplinarietà**: trattando dimensioni informative universali come Spazio, Tempo, e Eventi collocati su tali direttrici, realizza una intelaiatura informativa fondamentale di **supporto alla descrizione di qualunque tipologia di risorsa culturale**
-

# Applicativo di consultazione a faccette

## Istituto Centrale per gli Archivi - ICAR

### Atlante Storico Istituzionale dell'Italia unita

ricerca per parole  x ripristina

L'Atlante Storico Istituzionale rende disponibile un repertorio storico delle principali circoscrizioni amministrative dell'Italia postunitaria, descrivendone l'evoluzione nel tempo attraverso gli eventi e i provvedimenti normativi che ne hanno cambiato la denominazione, variato l'estensione territoriale, modificato la dipendenza gerarchica, determinato l'affluenza in altre circoscrizioni oppure la soppressione.

L'Atlante è basato su un modello concettuale espresso come ontologia in formato OWL che identifica le **entità** o **classi** rappresentate, le loro relazioni e le rispettive proprietà. La ricerca e la navigazione nel sistema mettono a frutto le potenzialità della classificazione a faccette, consentendo di esplorare le principali classi definite nel modello, secondo una logica pluridimensionale di accesso alle risorse informative, che, percorrendo le diverse proprietà o "facce" dell'entità descritta, permette di ottenere risposte e conseguire risultati attraverso percorsi alternativi.

Vai alla descrizione del progetto e all'ontologia

unità territoriali (12281)
r.s.t. (38076)
eventi (33342)

tipo unità territoriale	tipo evento	toponimo	fonte
+ Unità amministrativa 12252	preesistenza 9685		Dizionario dei comuni del Regno d'Italia 9683
+ Unità geografica 29	costituzione 3380	Abano 1	Sistat. Sistema storico delle
	cambio denominazione 2450	Abano Terme 1	amministrazioni territoriali 10389
	acquisizione territorio 4783	Abbadia 2	
	cessione territorio 4678	Abbadia Alpina 1	
	cambio unità territoriale sovraordinata 3191	Abbadia Cerreto 1	
	cambio codice 321	Abbadia Lariana 1	
	estinzione 3073	Abbadia San Salvatore 2	
	modifica in livello superiore 2339	Abbadia Sopra Adda 1	

periodo 1861 - 2012

Sintesi
Tabella
Timeline
Luoghi

1/1365
>
>>
>|
☰

Abbadia Cerreto

Abbadia San Salvatore

Abbadia poi Abbadia Alpina

## Cantalice

— <http://dati.san.beniculturali.it/ASI/UA05230>

<u>rdf:type</u>	<a href="#">geonames:Feature</a> <a href="#">owl:NamedIndividual</a> <a href="#">ASI:unitaTerritoriale</a>
<u>rdfs:label</u>	Cantalice
<u>sameAs</u>	<a href="#">dbpedia:Cantalice</a> <a href="http://datiopen.istat.it/odi/risorsa/territorio/comuni/Cantalice_57009">http://datiopen.istat.it/odi/risorsa/territorio/comuni/Cantalice_57009</a> <a href="http://dati.isprambiente.it/id/place/57009">http://dati.isprambiente.it/id/place/57009</a> <a href="http://it.dbpedia.org/resource/Cantalice">http://it.dbpedia.org/resource/Cantalice</a> <a href="http://spcdata.digitpa.gov.it/browse/Comune/B627">http://spcdata.digitpa.gov.it/browse/Comune/B627</a>
<u>dc:title</u>	Cantalice
<u>rdf:seeAlso</u>	<a href="#">Cantalice</a>
<u>ASI:ut2tpn</u>	<a href="#">Cantalice</a>
<u>ASI:ut2tut</u>	<a href="#">Comune</a>
is <u>ASI:ReST2ut</u> of	<a href="#">Cantalice (1861-03-17/1863-04-20)</a>

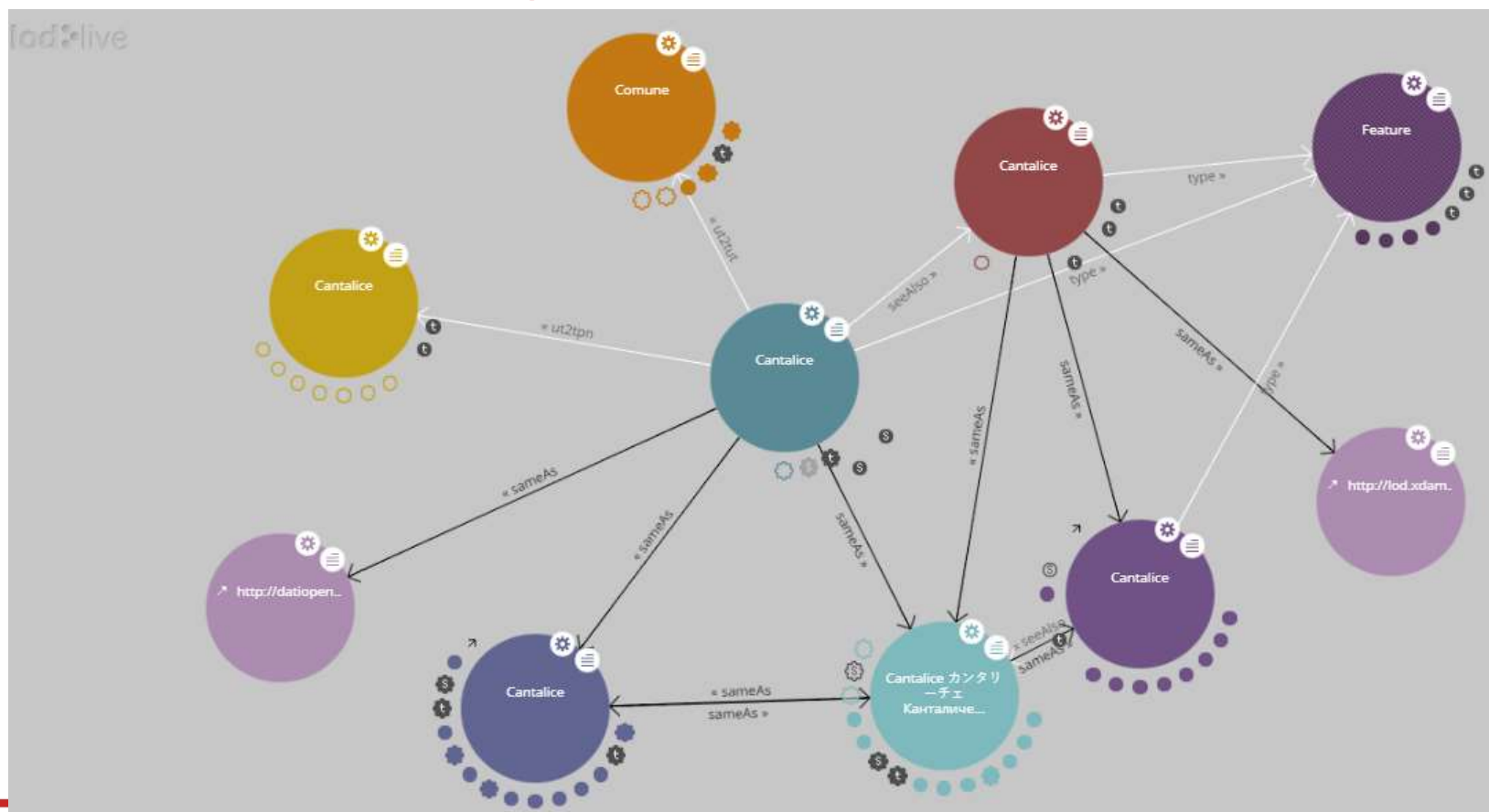
**ASI - Unità  
territoriale sul  
repository RDF**

## I Linked Open Data ASI

Nei dati ASI LOD sono stati inseriti **link semantici** ai corrispettivi dati delle principali ontologie geografiche disponibili (**GeoNames, ISPRA, ISTAT**, oltre ovviamente alla base di conoscenza di **DBpedia**), garantendo così forme di interoperabilità estesa col resto del dominio geografico.

Sono stati previsti anche collegamenti semantici tra le risorse RDF rappresentanti le **unità territoriali di ASI** e le corrispettive risorse nei **dati LOD del SAN** (corrispondenti alla sede del soggetto conservatore e del soggetto produttore dei complessi archivistici).

## Visualizzazione grafica dei dati LOD ASI



Stella Di Fazio

Fondazione Stelline, Milano, 15 marzo 2019

## Authority conservatori SAN (LOD)

---

Unitamente all'Atlante storico istituzionale, che costituisce una infrastruttura informativa per sua natura trasversale agli specifici domini (e sotto-domini) , l'ICAR rende disponibile - tra i numerosi dataset Linked Open Data corrispondenti alla versione LOD dei dati del SAN - i record di autorità di quasi 9.000 **soggetti conservatori**, ossia quei soggetti (enti, famiglie o persone) pubblici o privati che gestiscono e conservano il materiale documentario descritto nel Sistema Archivistico Nazionale e lo rendono accessibile al pubblico.

Tra le varie forme di riuso già attuate, si segnala che i soggetti conservatori LOD del SAN sono confluiti nell'ontologia Cultural-ON del MIBAC come risorse d'autorità che popolano la classe **CulturalInstituteOrSite** (Istituto e Luogo Della Cultura) sul versante archivistico.

## Authorities e Portali tematici

---

La definizione di specifici *authority files* assume particolare rilevanza anche nel **rinnovato modello di Portale tematico SAN** che l'ICAR sta definendo, a partire dai lavori di reingegnerizzazione del *Portale Rete Archivi per non dimenticare* attualmente in corso.

→ **nuovi Portali tematici**: l'**architettura concettuale** proposta sarà basata sulla individuazione delle **entità rappresentative del dominio di conoscenza**, implicitamente presenti nella struttura e nei contenuti attuali, di cui saranno definiti **in formato semantico** attributi descrittivi e mutue relazioni, nonché **link semantici con strumenti esterni**, in modo da garantire autonoma evidenza a tutte le dimensioni informative confluenti nello specifico Portale, ed elevati livelli di interoperabilità con altra conoscenza formalizzata.

---

## Authorities e Portali tematici

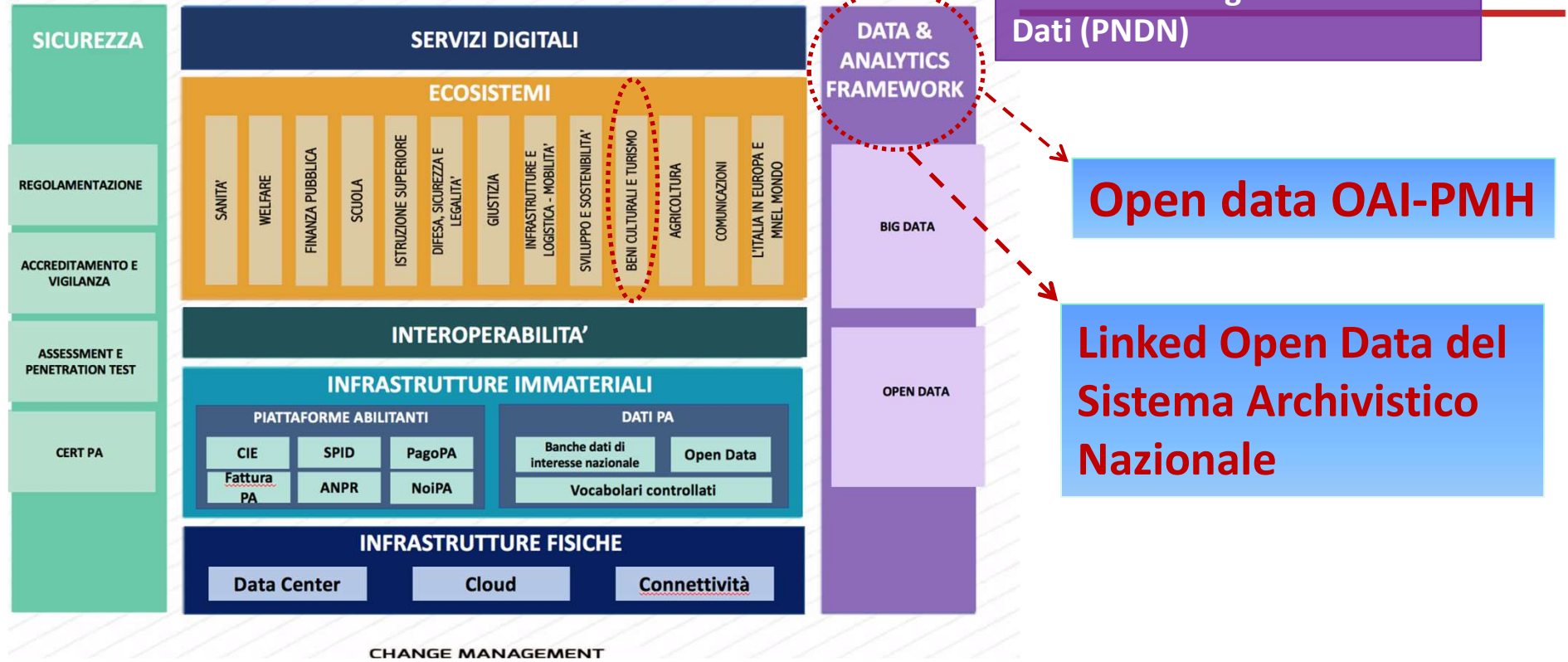
---

In particolare, nel **Portale Rete Archivi per non dimenticare** (che integrerà opportunamente dati dagli *authorities* per i Luoghi e i Conservatori) saranno realizzate delle liste di autorità relative ad **Eventi, Contesti storici, Organizzazioni e Persone** che definiscono concettualmente e contestualizzano le tematiche e gli eventi documentati dal materiale archivistico confluito e descritto nel Portale (fatti di terrorismo, violenza politica e mafia nell'Italia repubblicana).

### Tripla finalità:

- **controllo della descrizione;**
- **categorizzazione** dei contenuti destinate ad orientarne **accesso e fruizione**
- **formalizzazione di conoscenza di dominio d'autorità** esportabile e fruibile autonomamente





## Open data e Linked Open data del SAN

---

Il SAN ha aderito fin dall'inizio alle politiche di promozione della pubblicazione del patrimonio informativo pubblico sotto forma di **dati aperti**, principalmente attraverso **due canali**:

→ tramite un **OAI Provider**, che rende accessibili in formato **Open Data a 3 stelle** (non vincolati da piattaforme tecnologiche, espressi in formato non proprietario, distribuiti con licenze Creative Commons e IODL che ne garantiscono il riuso) tutti i contenuti descrittivi del SAN, in formato XML conforme ai tracciati CAT-SAN.

Oltre che attraverso **OAI-PMH**, i dati SAN sono resi disponibili anche in una codifica XML basata sul tracciato **PICO**, per garantire il riversamento delle informazioni del Sistema Archivistico Nazionale in **CulturalItalia**.

## Open data e Linked Open data del SAN

---

→ come **Linked Open Data** realizzati sulla base di un'ontologia che di base formalizza in linguaggio **OWL** l'architettura informativa dei **tracciati CAT-SAN**, integrando tuttavia alcune entità significative per la descrizione del patrimonio archivistico nazionale (contesti storici, profili istituzionali) non incluse originariamente nei tracciati CAT SAN, e quindi presenti esplicitamente solo nella versione LOD dei dati del SAN.

A corredo di questa operazione è stato definito in formato **SKOS** un **Tesauro del SAN**, che formalizza in modalità semantica e integrata tutti i **vocabolari controllati** adottati nel sistema per la normalizzazione delle informazioni descrittive.



# Grazie

**ISTITUTO CENTRALE PER GLI ARCHIVI**  
Viale Castro Pretorio 105 - 00185 Roma

**Direttore:** Stefano Vitali

**Centralino:** (+39) 06 5190.976 - (+39) 06 5196.0286

**Fax:** (+39) 06 4989.267

**E-mail:** [ic-a@beniculturali.it](mailto:ic-a@beniculturali.it)

**PEC:** [mbac-ic-a@mailcert.beniculturali.it](mailto:mbac-ic-a@mailcert.beniculturali.it)