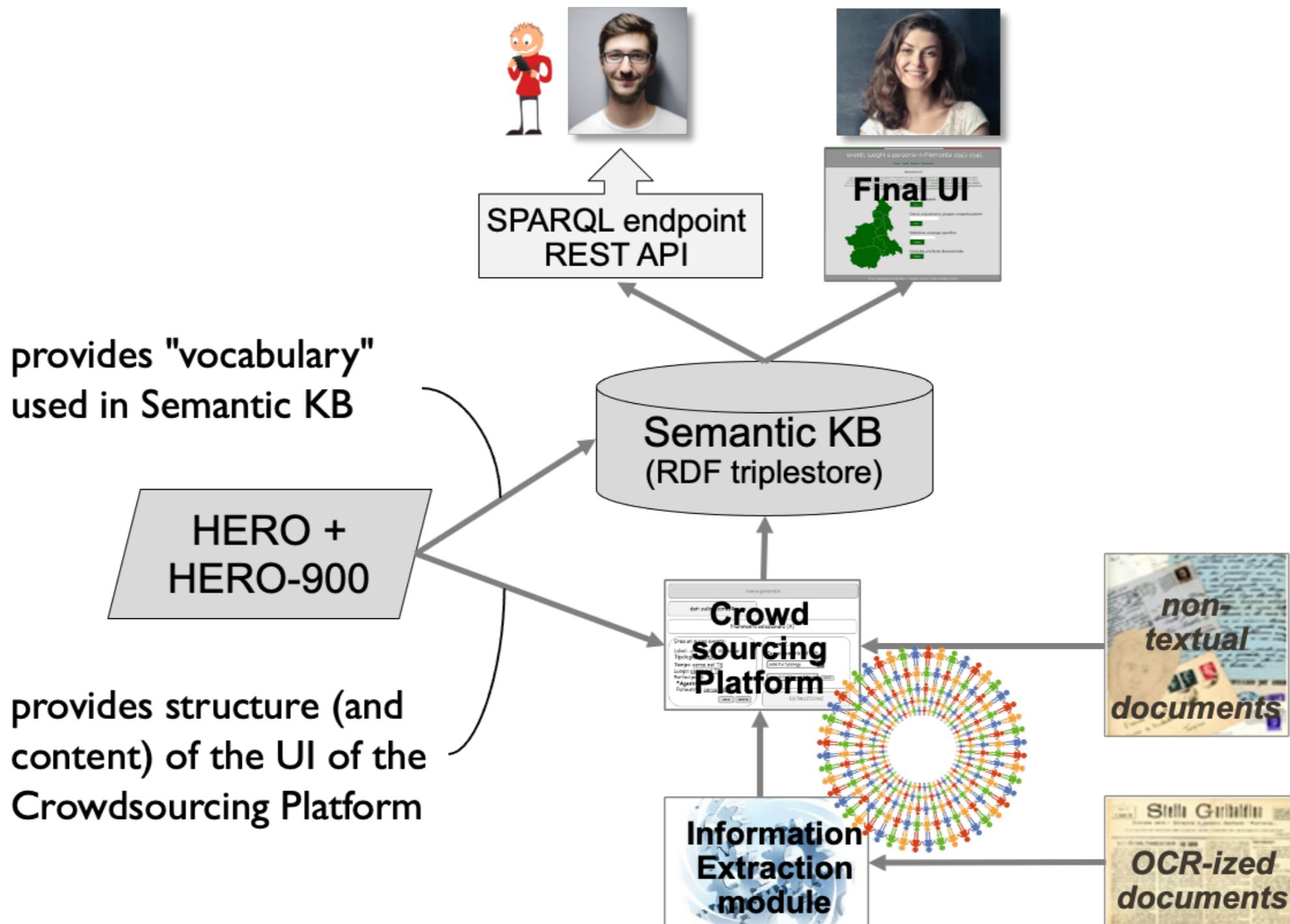




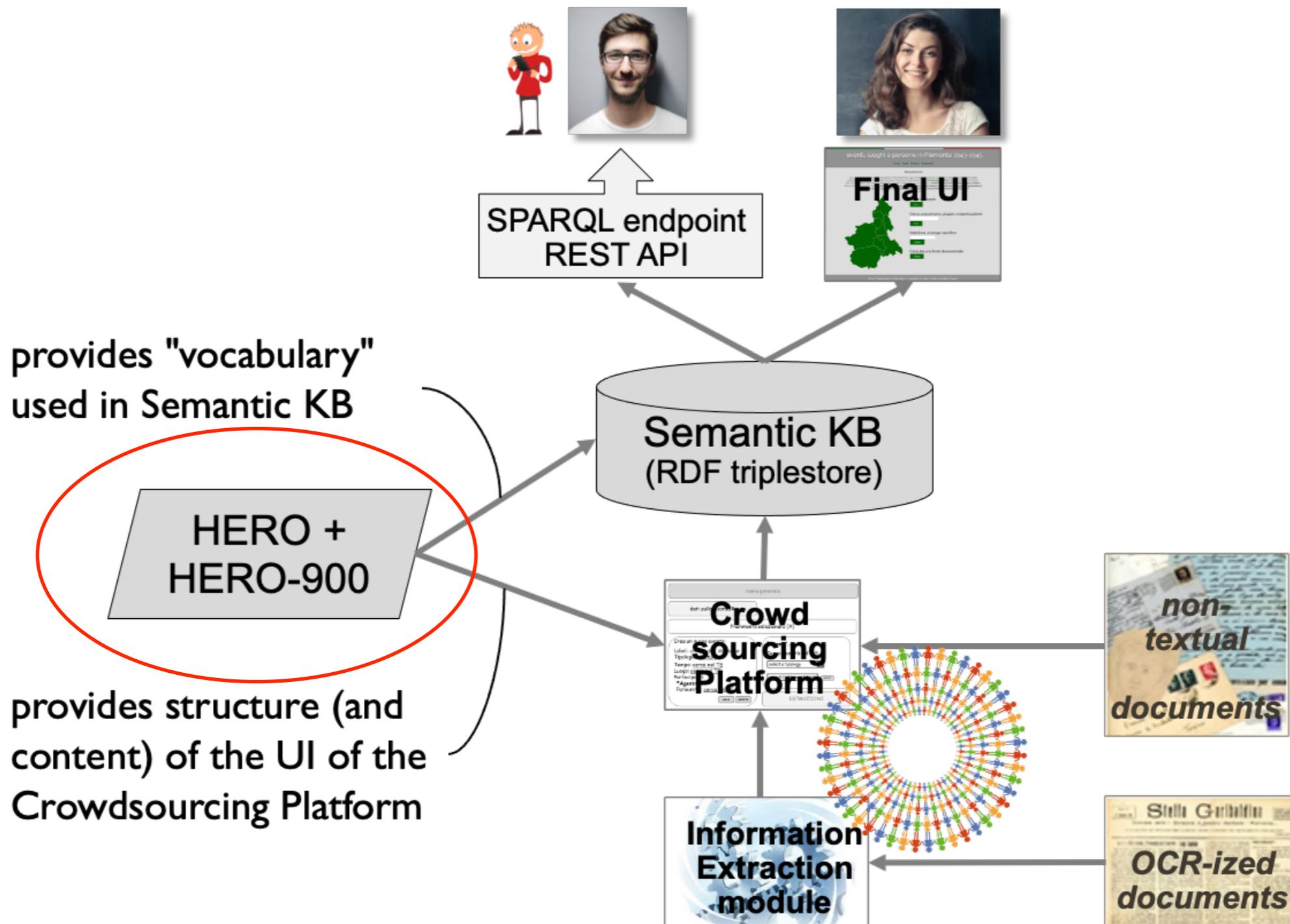
Le tecnologie a supporto dell'accessibilità degli archivi. Il caso di studio PRISMHA

Davide Colla
Dipartimento di Studi Storici
Università di Torino

ARCHITETTURA COMPLESSIVA



L'ONTOLOGIA *HERO*



L'ONTOLOGIA *HERO*

HERO (Historical Event Representation Ontology)

- è un **vocabolario comune**, condiviso tra
 - il **sistema**, gli **utenti** della piattaforma di **crowdsourcing** e gli **utenti finali** che cercano storie e documenti
 - **informatici/ontologi** (che progettano e realizzano il sistema) e **storici** (che forniscono la prospettiva storica sui documenti)
- è il risultato dell'**integrazione** di
 - un'analisi dei **modelli esistenti**
 - l'esito del **dialogo** tra **informatici** e **storici** sulla nozione di **evento**, con le sue proprietà (es. partecipazione, ruoli) e relazioni con altri eventi (es. causa, influenza)

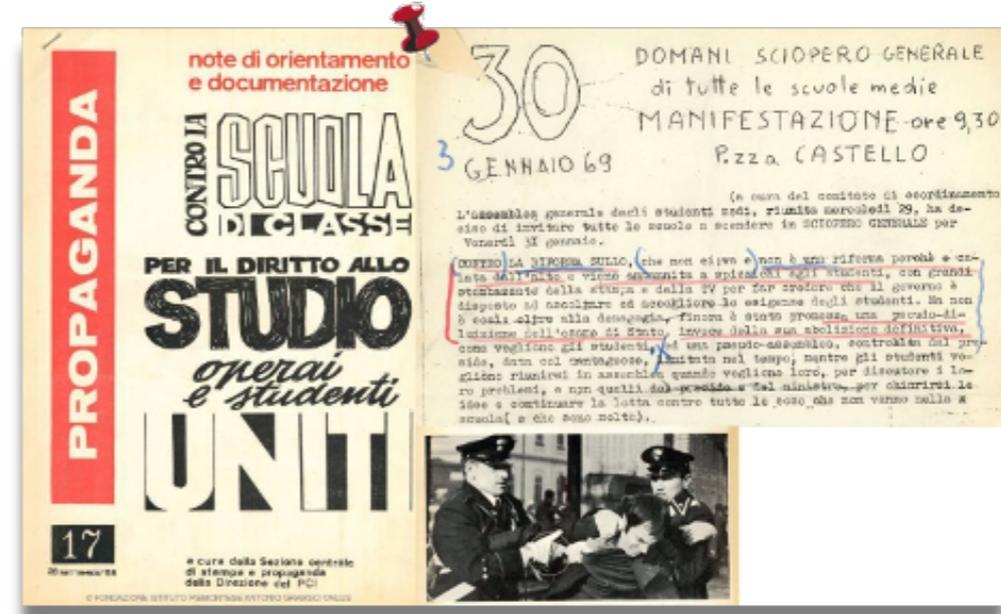
L'ONTOLOGIA *HERO*

Struttura di HERO (<https://w3id.org/hero/HERO>)

- **HERO-TOP**: classi e proprietà generali (entità astratta, oggetto (non)fisico, essere parte di, essere sotto-classe di, ...)
- **Moduli** con classi e proprietà utili per caratterizzare...
 - **HERO-EVENT**: [eventi](#) (evento, azione, partecipare in un evento, giocare un certo ruolo in un evento, causare, ...)
 - **HERO-ROCS**: [ruoli](#), [organizzazioni](#), [collettivi](#), [insiemi](#) (ruolo, organizzazione, giocare un ruolo (sociale), essere membro di un collettivo/insieme, ...)
 - **HERO-PLACE**: [luoghi](#) (pianura, montagna, fiume, luogo antropizzato, ...)
 - **HERO-TIME**: [intervalli di tempo](#) (giorni, anni, periodi, ...)

L'ONTOLOGIA HERO

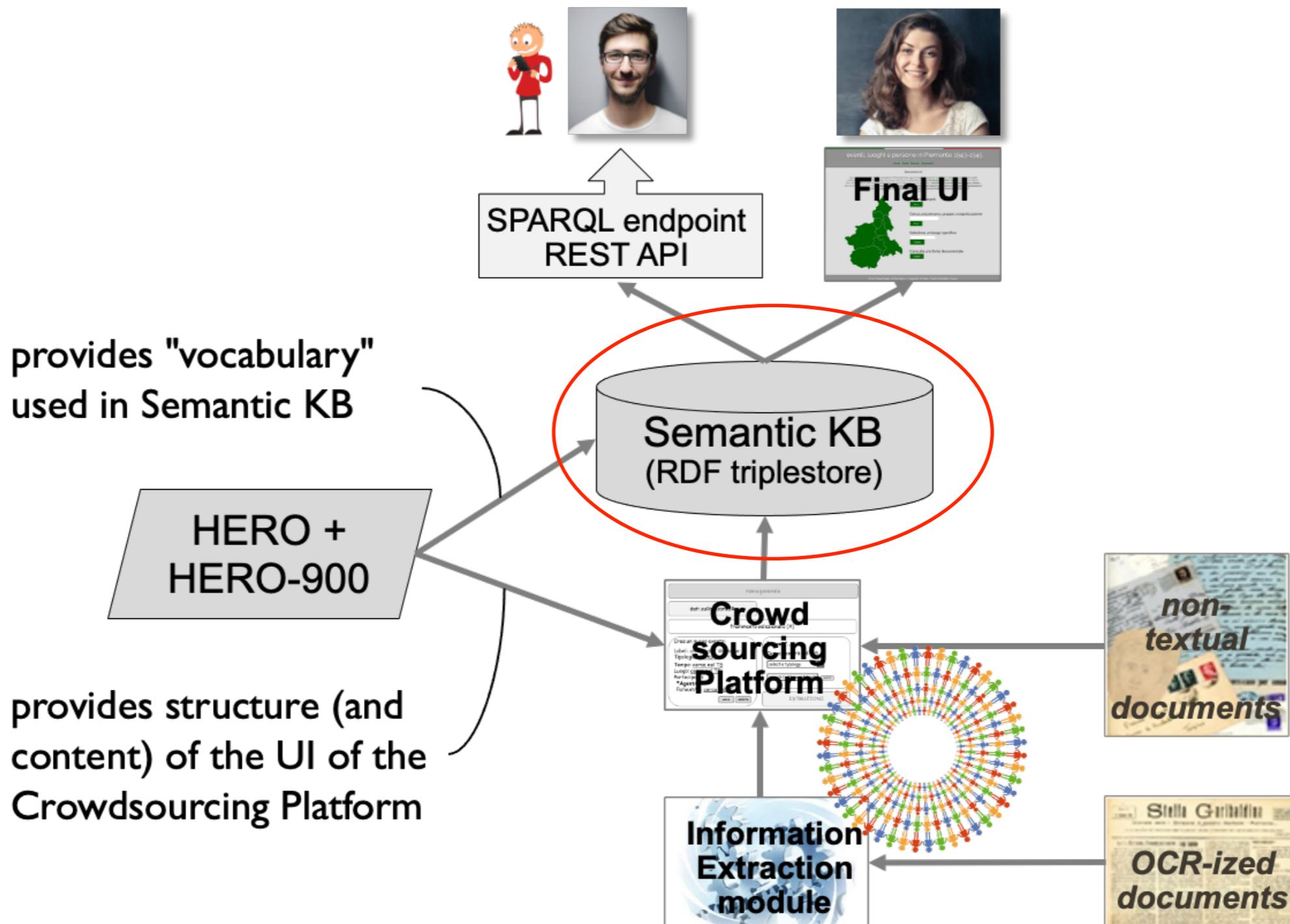
Il modello semantico top-level (HERO) ha guidato l'analisi e la selezione dei documenti dall'archivio dell'Ist. Gramsci (relativi alle proteste del '68): gli **storici** hanno analizzato i documenti e costruito delle **schede analitiche**, strutturate in base alle classi definite in HERO (eventi, persone, organizzazioni, collettivi, ...)



Le **schede**, insieme alla consulenza diretta degli **storici**, sono state usate per costruire l'**ontologia di dominio (HERO-900) = modello semantico** specifico che raffina HERO definendo concetti e proprietà specifici.

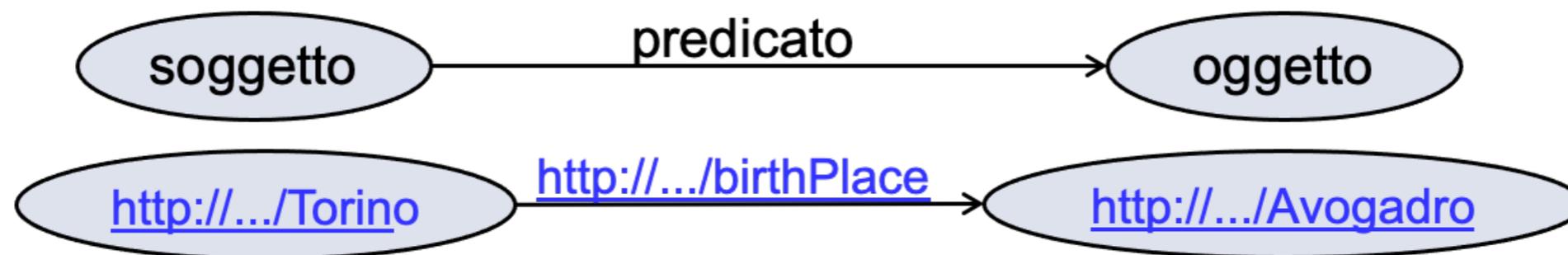
HERO e HERO-900 sono scritte in **OWL2**, una specifica del W3C.

La base di conoscenza semantica



La base di conoscenza semantica

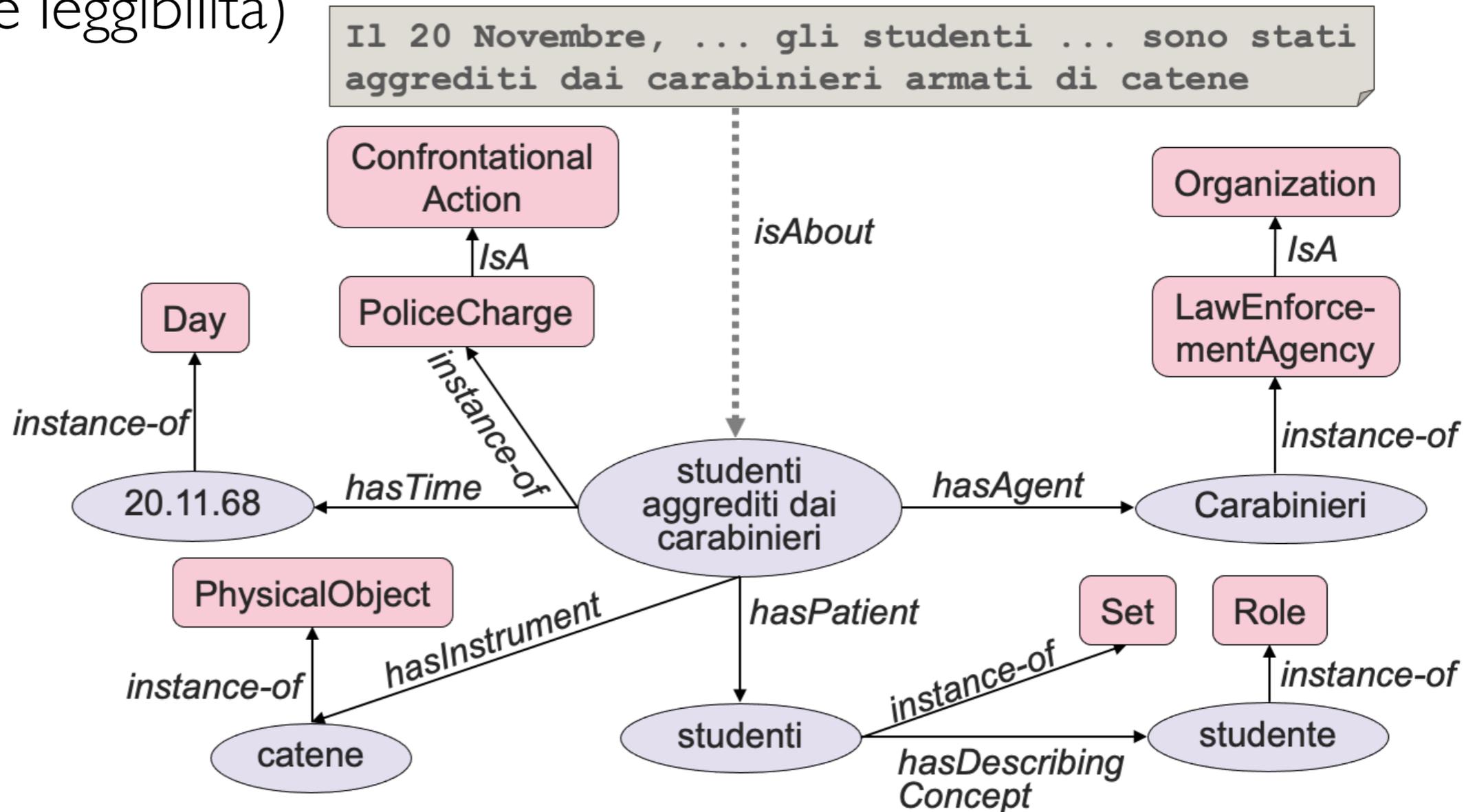
- la **Semantic KB** contiene le **rappresentazioni semantiche** del contenuto dei documenti d'archivio
- è implementata come un **RDF triplestore**
- **RDF** è un **modello** in cui i **dati** sono rappresentati da **triple** o "asserzioni" (*statements*) della forma $\langle \text{soggetto} \text{ predicato} \text{ oggetto} \rangle$, graficamente:



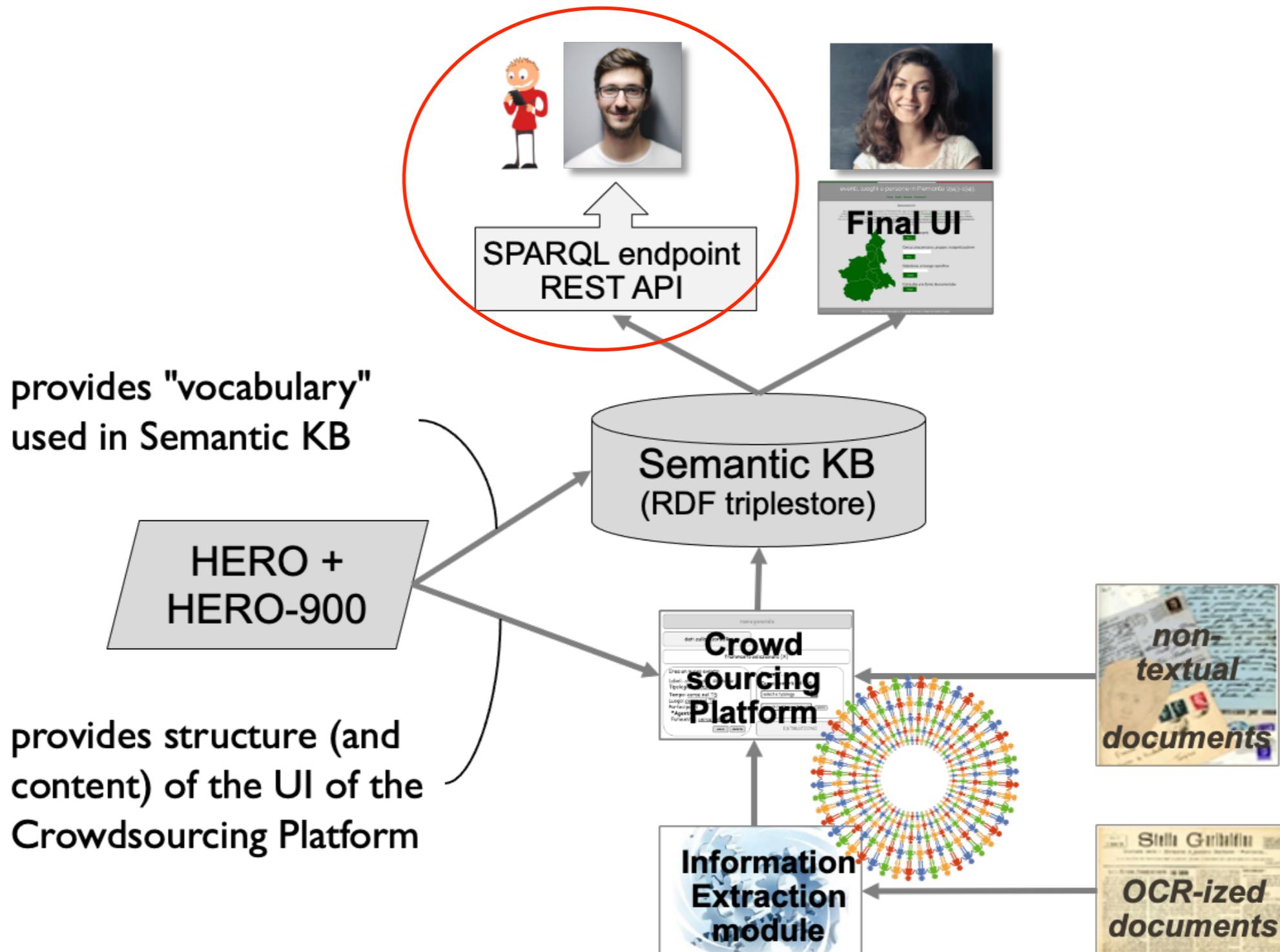
- soggetto, predicato e oggetto sono **risorse**
- le **risorse** sono univocamente **identificate da IRI**
- RDF è una specifica del W3C (www.w3.org)

La base di conoscenza semantica

Rappresentazione grafica (semplificata) di un **insieme di triple RDF** nella Semantic KB (IRI sostituiti da etichette per maggiore leggibilità)



ACCESSO DIRETTO AI DATI



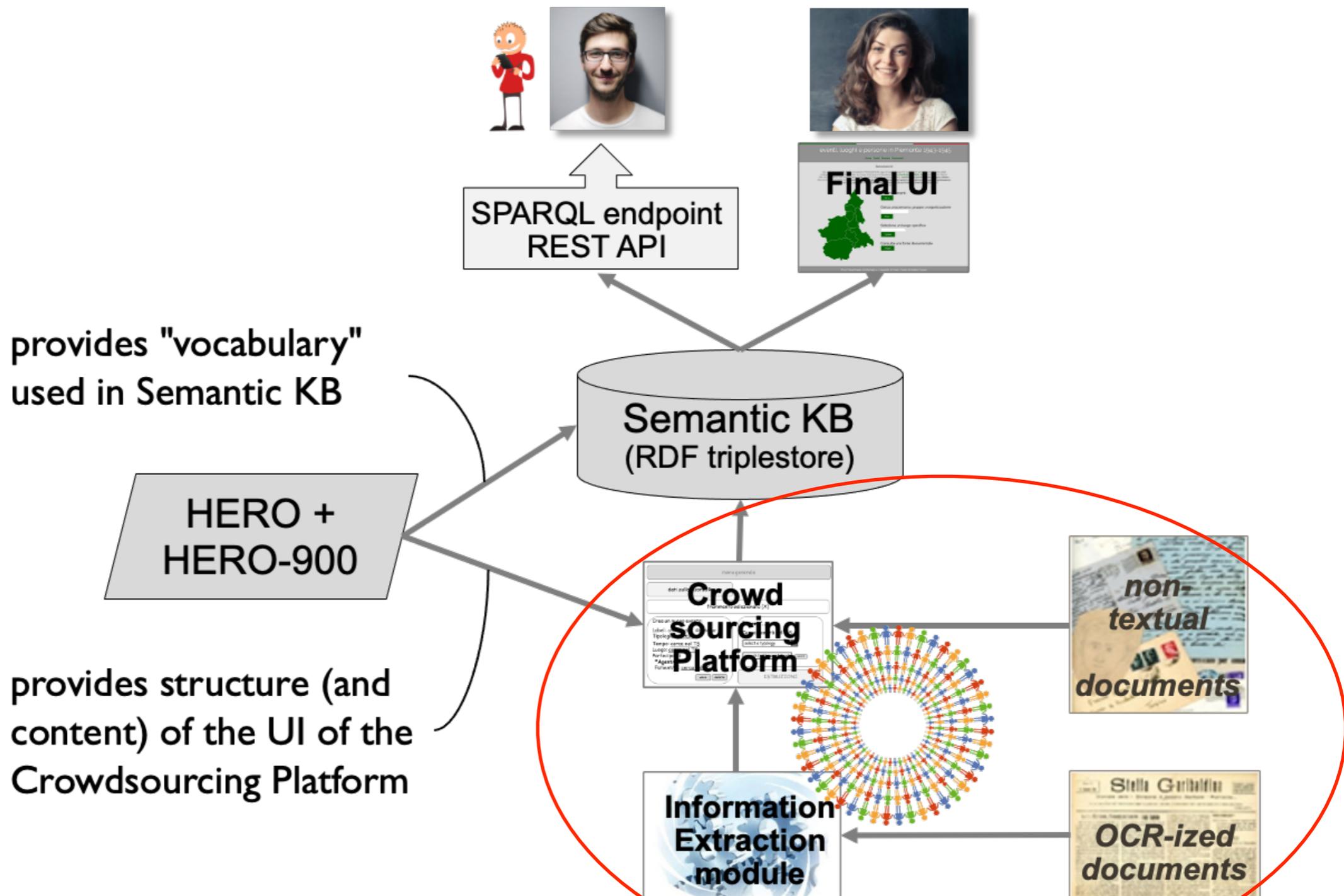
ACCESSO DIRETTO AI DATI

I dati nella **Semantic KB** saranno accessibili attraverso uno **SPARQL Endpoint** e **API REST**

- **SPARQL** è un **protocollo** e un **linguaggio di interrogazione e manipolazione di RDF triplestore**
- uno **SPARQL endpoint** è un **Server** che risponde a **richieste SPARQL**
- **REST** = insieme di principi "architetturali", basati su HTTP, per la progettazione e realizzazione di **interfacce di accesso a servizi e dati**

In questo modo è possibile **utilizzare i dati di PRiSMHA** (le rappresentazioni semantiche contenute nella Semantic KB) all'interno di **applicazioni di terze parti** (es: app **turistiche, educative, culturali**)

CROWDSOURCING



Prototipo online!



The screenshot displays the PRISMHA web application interface. At the top, there is a navigation bar with links for Home, Projects, Semantic Descriptions, Configuration, and Contact. The main heading is "1968 e dintorni: cosa accadeva e chi c'era" with a sub-heading "All included resources". Below this, there is a list of resources including "Biografia Sergio Garavini", "PSUP (Torino) Carabinieri e manifestanti (201)", and "Ciclostato Comitato Coordinamento Studenti Torino 2011/199".

The central part of the interface shows a complex network graph with nodes and connecting lines. A modal window titled "Link to an existing" is open, showing a list of "ALL TYPOLOGIES" and a search field for "SEARCH EXISTING ENTITY". Another modal window titled "Crea new entity" is also open, displaying a list of "Entry type" options such as "hasPart", "hasBeginningOfEvent", "hasLocation", "Influences", "hasTimeSpan", "hasPatient", "hasPurpose", "hasDamaged", "hasSupporter", "hasFinalLocation", "hasInRelationWith", and "hasActorTopic". Each option has a corresponding search field and a "Search on TS for" label.

CROWDSOURCING + NLP

- **Named Entity Recognition:** riconoscimento automatico di rif. a persone, organizzazioni, luoghi, date

Il pomeriggio dell'**8 settembre '43** ero in tram, diretta a **San Mauro**, dove c'erano **Renata** e altri amici. A un certo punto, in **corso Casale**, si vide un assembramento verso i bar da cui alta risuonava una voce dalla radio. Scesi e udii il proclama del maresciallo **Badoglio** che annunciava l'armistizio.

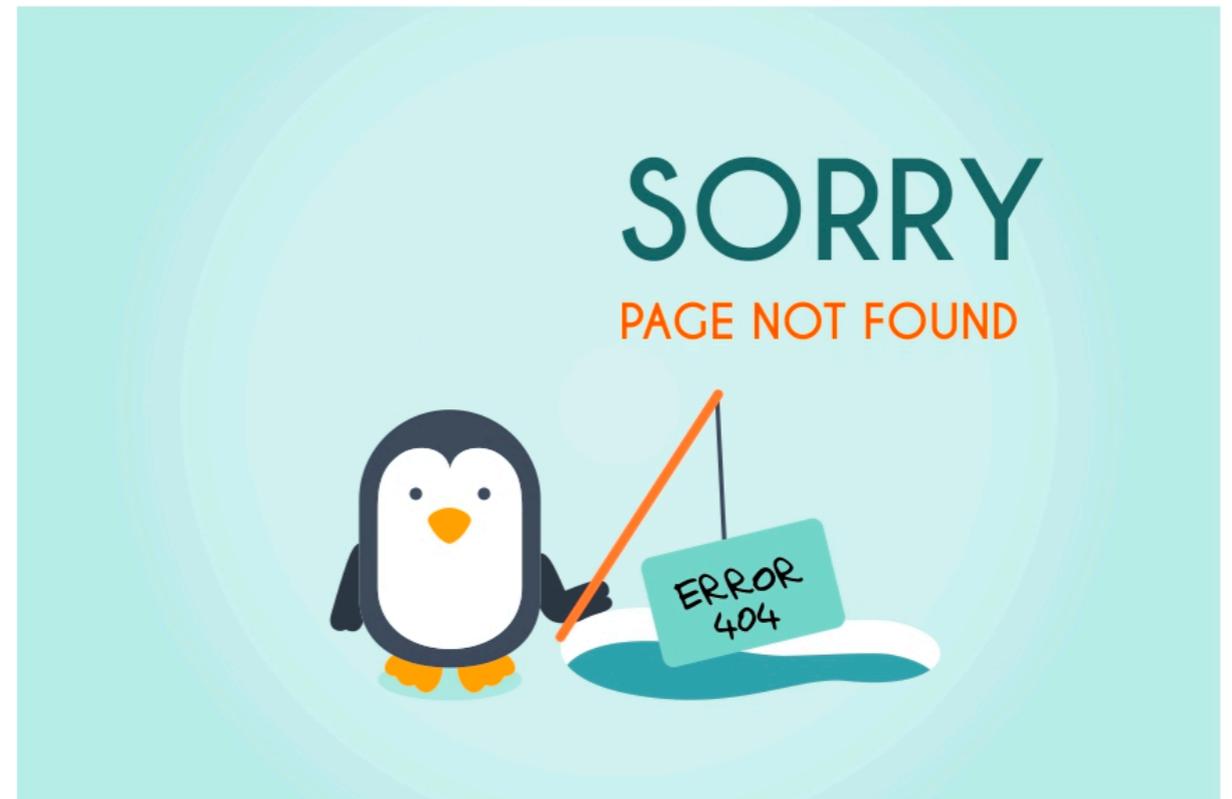
- **Event mining:** riconoscimento automatico di espressioni che denotano eventi

Gli effettivi del generale Schlemmer [...] Durante il cammino effettuano le ultime **stragi**: a Grugliasco **massacrano 66 persone**, garibaldini e civili; a Nichelino **uccidono 13 garibaldini** della Brigata Pisacane. Tenteranno di **proseguire** verso il Vercellese; **bloccati** però anche lì dai partigiani, finiranno per **arrendersi** il 3 maggio nelle mani degli Alleati.

Riflessione finale

Numerosi progetti coinvolgono ormai un **lato digitale**: l'aspetto di **disseminazione** è fondamentale (siti web, piattaforme, ...).

Tanti progetti però **svaniscono nel nulla** al termine dei finanziamenti.



Fonte: <https://www.scuolaecommerce.com/>



Urge un sistema standardizzato di **conservazione dei progetti**, per non perdere il lavoro fatto e continuare a reinventare la ruota!

Grazie per l'attenzione!

