

# L'infrastruttura di integrazione in ARIADNE

Achille Felicetti

VAST-LAB, PIN, Prato, Italia  
achille.felicetti@pin.unifi.it

**Abstract.** This paper describes the definition and implementation of the interoperability platform of the ARIADNE project for the integration of archaeological archives, the different components it includes and the various steps necessary for its creation and its deployment.

**Keywords:** ARIADNE, archaeology, semantics, interoperability, CIDOC CRM

## 1 Introduzione

L'integrazione dei dati creati dalla ricerca archeologica e, in generale, nel settore dei beni culturali, è un processo molto complesso. Tale circostanza nasce fondamentalmente dall'enorme eterogeneità delle collezioni digitali gestite ed utilizzate dalle varie istituzioni, con la conseguenza che, sebbene vi siano notevoli somiglianze nel modo di gestire le informazioni, il processo di integrazione di dati diventa spesso molto difficoltoso.

I dati in questione, infatti, sono strutturati, dalle differenti istituzioni, in modo differente, spesso mediante l'utilizzo di diversi linguaggi e schemi di metadati per la loro codifica. Il modo in cui l'informazione è organizzata, inoltre, riflette l'impostazione metodologica e gli obiettivi specifici legati ai luoghi e alle epoche oggetto di indagine. Tuttavia, la gestione integrata dell'informazione dei beni culturali è oggi diventata una necessità imprescindibile per garantire un uso efficiente di questi dati e per poterne sprigionare il pieno potenziale, così da fornire un contributo significativo al progresso della ricerca.

Tali condizioni possono crearsi solo in un ambiente integrato in cui dati di diversa provenienza possano divenire reciprocamente interpretabili e in grado di essere fruiti dagli utenti come se fossero parte di un unico grande archivio. Solo un'interoperabilità di questo genere potrà garantire il reperimento di informazioni significative e una loro distribuzione organica e sinottica su piani sia spaziali che temporali.

La creazione di un'infrastruttura di questo tipo, in grado di preservare al tempo stesso la natura e la provenienza originale dei dati, e delle potenti interfacce necessarie per la sua consultazione potrebbe, tuttavia, rivelarsi un compito arduo con la tecnologia attualmente disponibile. Nonostante ciò è di importanza capitale incamminarsi lungo la via della progettazione e dello sviluppo di un tale strumento per poter garantire un

accesso uniforme ad informazioni che, sebbene frammentate e disperse in diversi archivi, si mostrano sostanzialmente omogenee.

Questo è l'obiettivo ambizioso che anima il progetto ARIADNE, il quale ha il merito di aver individuato le diverse fasi di sviluppo necessarie al raggiungimento di tale scopo. Ogni passo verso l'interoperabilità è il risultato di una proficua collaborazione tra le istituzioni archeologiche e gli sviluppatori, le due anime del progetto. Una collaborazione che rappresenta la vera, solida base sulla quale potrà avvenire la materiale realizzazione dell'infrastruttura.

## **2 Percorsi di integrazione**

Le attività di integrazione prevedono un'analisi preliminare degli archivi, analisi necessaria per l'individuazione dei formati, degli standard e dei servizi già in uso presso i partner archeologici che forniscono contenuti al progetto. Le descrizioni di queste analisi sono state raccolte in vario modo, utilizzando meccanismi online e codificate mediante uno standard di metadati (ACDM) sviluppato da ARIADNE per produrre una rappresentazione dettagliata, formale e univoca delle informazioni.

Quando si parla di integrazione si intende di solito una serie di operazioni complesse che si svolge su molteplici piani ed avviene su multipli livelli di profondità. Caratteristica di ogni attività di questo tipo è l'individuazione di elementi chiave, di tratti comuni che possano identificare oggetti e argomenti a livello concettuale per descriverli poi utilizzando un linguaggio comune.

Partendo da un livello più generale, gli elementi fondamentali riconoscibili in ogni archivio sono riassumibili negli ormai noti "who, what, where, when", persone, oggetti, luoghi ed intervalli temporali che rivestono particolare importanza specialmente in archeologia. Un'analisi attenta degli archivi in ARIADNE ha dimostrato come una integrazione di profilo generale sia già possibile a livello della descrizione degli archivi mediante l'utilizzo di questi concetti, se preceduta da una opportuna riduzione dei concetti stessi ad un comune linguaggio condiviso. ARIADNE ha dedicato parte delle sue attività all'individuazione di tali elementi chiave e alla loro appropriata codifica per mezzo di strumenti terminologici standard esistenti e già ben attestati nel panorama internazionale.

### **2.1 Soggetti ed argomenti: il "cosa"**

Gli argomenti e i soggetti generali a cui si riferiscono i vari record negli archivi (ad esempio: scavi e ricognizioni archeologiche, monumenti, sepolture, ceramica e simili, che costituiscono il "what") sono descritti per mezzo di termini tratti dall'Art and Architecture Thesaurus (AAT) del Getty Research Institute. Lo stesso AAT è stato costituito come perno centrale su cui impostare tutto il framework terminologico di ARIADNE, non solo per quanto riguarda i soggetti ed i temi generali, ma anche per ogni altra descrizione tipologica, morfologica e funzionale di oggetti archeologici e di attività ad essi connesse nei livelli più profondi di integrazione. L'utilizzo di un

thesaurus condiviso richiede una mappatura preventiva di ogni risorsa terminologica utilizzata dai content provider per la digitalizzazione dei loro archivi sui concetti di AAT, un'attività concettualmente rilevante anche per dimostrare la similitudine logica e ontologica esistente fra i diversi archivi. Le attività di mappatura sono assistite da uno specifico strumento online sviluppato appositamente per stabilire corrispondenze fra concetti provenienti da vocabolari diversi.

## **2.2 Luoghi: il “dove”**

Le entità spaziali (il “where”) sono altrettanto semplici da integrare. Molti archivi contengono infatti dati spaziali in formato standard, molto spesso corredati da coordinate in grado di identificare univocamente specifici luoghi. In questo caso ARIADNE raccomanda l'utilizzo o la conversione nel formato WGS84 delle coordinate spaziali per consentire la consultazione degli archivi per mezzo di strumenti geografici. Specifiche funzionalità del gazetteer GeoNames, uno dei più popolari strumenti di questo tipo, possono essere utilizzate per ricavare coordinate spaziali a partire da semplici nomi di luoghi nel caso in cui queste siano le uniche indicazioni geografiche presenti in un dato archivio. Per quanto riguarda l'utilizzo dei nomi storici che una località può aver avuto nel passato, risulta preziosa la collaborazione col progetto Pelagios, un aggregatore di informazioni geografiche in grado di attingere dati da Pleiades (un thesaurus di luoghi del passato costruito su base bibliografica) e di distribuirle in formato Linked Open Data sotto forma di URI utilizzabili per identificare luoghi del passato senza ambiguità alcuna [3].

## **2.3 Periodi, epoche, date: il “quando”**

Di particolare interesse è il caso dell'integrazione su base temporale a partire dalle indicazioni di date, epoche, intervalli di tempo e periodi di cui gli archivi archeologici sono solitamente ricchi. La condivisione di date precise non pone alcun problema essendo la misura dell'asse temporale di natura numerica e quindi non ambigua. Va invece notato che ogni periodo o epoca identificati nei database esclusivamente con il loro nome, senza riferimenti temporali assoluti, possono far sorgere forti ambiguità in una prospettiva di integrazione più ampia: ad es., l'Età del Ferro in Anatolia ha un'estensione temporale ben diversa da quella nelle Isole Britanniche. Risulta, quindi, evidente che la determinazione temporale di un'epoca in senso assoluto è impossibile senza che vi sia un riferimento spaziale che contribuisca ad identificarla.

Un'ovvia ed immediata soluzione al problema della periodizzazione consiste nel convertire ogni periodo in date assolute specificandone inizio e fine. Ciò tuttavia non risolverebbe le sovrapposizioni semantiche derivanti dalla necessità di conservare comunque le indicazioni temporali come parte integrante della documentazione. Una soluzione alternativa è fornita dal progetto PeriodO, con cui ARIADNE collabora, il

quale, in modo analogo a quanto fa Pelagios per i luoghi, gestisce collezioni di periodi intesi come intersezioni di eventi documentabili su determinate aree geografiche, definiti per mezzo di specifica documentazione da parte di enti o altri soggetti (archeologi, storici ecc.) nel corso dei loro studi. Analogamente a Pelagios, PeriodO gestisce e fornisce identificativi univoci per ognuno di questi periodi secondo la filosofia Linked Open Data. Questa fase di integrazione, basata su entità di alto livello, ha prodotto il Registry, una risorsa importante per la comprensione dei dati archeologici e dei relativi metadati e servizi ad essi collegati. Il Registry e le descrizioni che esso contiene rappresentano il baricentro stesso dell'integrazione, fornendo il supporto necessario per la progettazione e l'implementazione dell'interoperabilità dei dati archeologici ad un livello più profondo.

### **3 Preparare gli archivi**

Dopo l'analisi dei dati, un altro passo fondamentale verso l'integrazione è costituito da una serie di operazioni preparatorie effettuate sugli archivi dei fornitori di contenuti al fine di facilitarne l'adeguamento come parte dell'infrastruttura globale di ARIADNE. Il primo e più ovvio passo è l'elaborazione e la trasformazione dei dati archeologici "locali" al fine di uniformarli secondo le linee generali descritte sopra per renderli accessibili tramite il Web. Infatti, sebbene molti archivi e collezioni già prevedano un accesso tramite internet e forniscano servizi per la loro consultazione online (ad esempio ADS, Arachne, ZENON, Fasti Online), altri risultano inaccessibili poiché risiedono in documenti stand alone (Microsoft Excel, Microsoft Word, PDF) o addirittura non sono ancora disponibili in formato digitale (molti rendiconti e diari di scavo, ad esempio, si trovano esclusivamente in formato cartaceo).

L'accessibilità online da sola non basta, tuttavia, a rendere i dati interoperabili. Una volta attuata, essa deve poi essere arricchita di altre funzionalità, come ad esempio divenire parte di una ricerca distribuita in grado di reperire informazioni attraversando e analizzando dati memorizzati su archivi diversi, al fine di raccogliere informazioni pertinenti ai criteri di ricerca. Una tale accessibilità si realizza attraverso la pubblicazione online dei dati in formati standard e la progettazione e l'implementazione di interfacce macchina consultabili dagli altri componenti dell'infrastruttura. La realizzazione di queste funzionalità in ARIADNE si avvale di protocolli standard, come OAI-PMH, e di formati condivisi quali METS, OAI-ORE e altri schemi di metadati codificati in XML, RDF o Linked Open Data per garantire il riutilizzo e l'interscambio. Un set di API create su misura per ogni archivio rende la comunicazione con la piattaforma ARIADNE ancora più flessibile ed efficiente. Il processo di standardizzazione e pubblicazione può essere esteso anche alle risorse terminologiche (vocabolari controllati, tassonomie, thesauri), da utilizzare in combinazione con i dati originali [1].

Un ulteriore e più avanzato stadio di condivisione consentirà la pubblicazione in formato semantico non soltanto delle descrizioni ma anche di ogni record presente negli archivi archeologici, rappresentati per mezzo delle classi e delle proprietà del CIDOC CRM, il modello concettuale di riferimento scelto dal progetto [2].

ARIADNE fornisce un importante contributo allo sviluppo di un modello archeologico globale basato sul CIDOC CRM anche per mezzo della definizione di una sua estensione (CRMarcheo) studiata per ottenere un'integrazione ancora più capillare delle informazioni archeologiche.

#### **4 Mappatura e conversione di singoli record**

Una fase avanzata dello sviluppo della piattaforma di interoperabilità prevede l'integrazione dei singoli record degli archivi attraverso un ampio lavoro di concettualizzazione, mappatura e conversione delle informazioni archeologiche e la costruzione di un repository con funzionalità semantiche in grado di eseguire query complesse su dati aggregati.

L'implementazione di questo livello più profondo di granularità si basa sulla definizione di mappature in grado di catturare ed esprimere la ricchezza semantica dei dati archeologici. Il progetto prevede una serie di operazioni di mapping attraverso strumenti specifici che permettono ad ogni partner di mantenere definire le diverse e complesse corrispondenze tra le entità contenute nei loro archivi e le classi concettuali forniti dal CIDOC CRM e dalle sue estensioni (CRMarchaeo in primis) [4]. Queste "mappe concettuali" così definite da ciascun partner, successivamente applicate ai dati reali, consentono la creazione di rappresentazioni semantiche codificate in RDF per ogni singolo elemento e concetto, in modo da definire un grafo complesso di relazioni pronto per essere visualizzato, interrogato, integrato con le tecnologie semantiche e pubblicato nel formato Linked Open Data (vedi Figura 1).

Le operazioni di mappatura concettuale (cioè tra lo schema concettuale di ogni archivio archeologico e il modello CIDOC CRM) sono in via di definizione all'interno del progetto anche per mezzo di specifici strumenti che semplificano il complesso lavoro di mappatura per mezzo di sofisticate interfacce utente. Moduli specializzati della piattaforma di ARIADNE sono stati predisposti per eseguire le operazioni di conversione e pubblicazione [5][6].

L'infrastruttura di ARIADNE include anche un Visual Media Service per la visualizzazione di media complessi (immagini ad alta risoluzione e multispettro, modelli 3D ecc.) attraverso un semplice web browser, e un Landscape Service, per la rappresentazione 3D di paesaggi virtuali. Il primo è un servizio automatico che permette all'utente di caricare file multimediali sul server di ARIADNE e di trasformarli in un formato web che li rende idonei alla visualizzazione via browser. Tutte le elaborazioni necessarie per trasformare i dati in un formato web-compatibile ed dinamico avvengono in modo automatico attraverso specifiche funzionalità fornite dal server di ARIADNE. Particolari funzioni offerte dal servizio consentono di cambiare il paradigma di navigazione e lo stile della pagina, nonché di effettuare operazioni più complesse quali misurazioni punto-punto.

Il Landscape Service è invece un insieme di servizi online per l'elaborazione, la gestione e la pubblicazione in multi-risoluzione di informazioni interattive relative a

spazi 3D di grandi dimensioni, particolarmente utili per le ricostruzioni di archeologia virtuale e per gli studi di archeologia del paesaggio. La gestione dei dati avviene attraverso un servizio cloud, che consente lavori collaborativi tra istituti di ricerca e professionisti e permette la generazione e la pubblicazione di set di dati 3D, offrendo diverse opzioni per il controllo e l'integrazione di formati, risoluzioni, geometrie/compressioni, texture e altro ancora.

## **5 Archivi semantici ed interfacce utente**

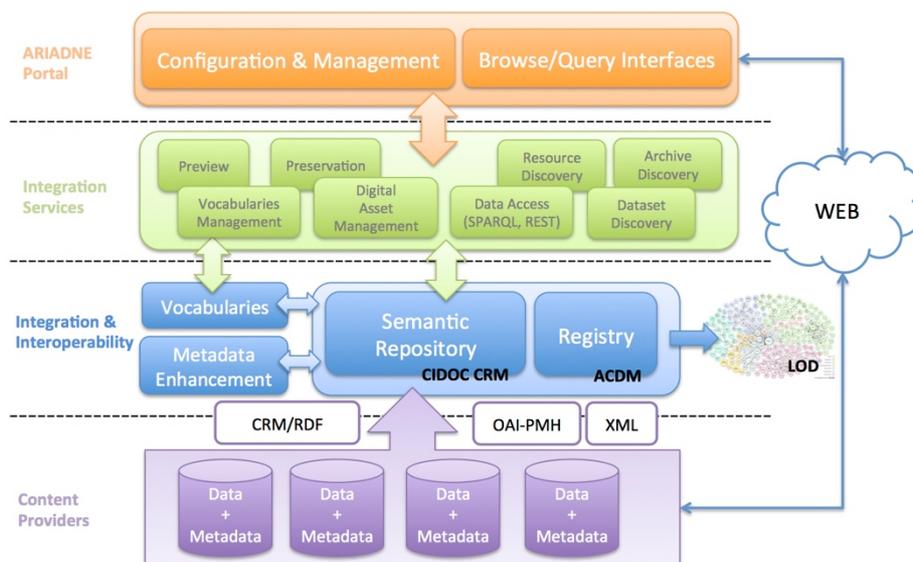
La piattaforma di integrazione così progettata da ARIADNE si presenterà, nella sua forma compiuta, alla fine del progetto, come un sistema modulare complesso, dotato di interfacce e funzioni avanzate e di un'architettura in grado di interagire con archivi distribuiti in modo trasparente. Il sistema sarà in grado di interrogare ed estrarre informazioni in qualsiasi tipo di formato (preferibilmente semantico) e di integrare i risultati in un unico grafo semantico, presentandoli all'utente in modo coerente e fornendogli tutti gli strumenti per visualizzarli, analizzarli ed eventualmente utilizzarli come parte di successive ricerche. L'infrastruttura sarà, inoltre, dotata di un sistema di aggiornamento in tempo reale capace di allinearsi alle eventuali modifiche degli archivi distribuiti e di sincronizzare le informazioni presenti in modo da fornire sempre la versione più aggiornata dei dati da interrogare.

Questa serie di operazioni comporta la costante interazione con il Registry, che contiene, oltre alle descrizioni dettagliate degli archivi originari, anche i dati relativi alla loro "digital provenance", inclusi gli "indirizzi" (end point per l'harvesting, sorgenti di dati condivise, access point e simili) da cui attingere le informazioni aggiornate. Queste informazioni del Registry sono usate per dirigere le query verso gli archivi più appropriati, che presumibilmente conterranno le informazioni a cui l'utente è interessato. L'interazione con i dati e i servizi terminologici è ugualmente importante, soprattutto in fase di formulazione delle query, per ottenere il supporto necessario a definire percorsi di ricerca coerenti con le strutture da consultare.

Il Portale di ARIADNE, che rappresenta il livello più alto nell'architettura del sistema, rappresenterà il punto di ingresso a tutto il meccanismo di consultazione e fruizione del contenuto. Attraverso di esso, gli utenti potranno estrarre, analizzare e utilizzare tutte le informazioni disponibili ed ottenere l'accesso attraverso i vari servizi integrati nel sistema.

Il Portale consentirà all'utente di interrogare, per mezzo di interfacce evolute, tutte le informazioni integrate nel grafo semantico, così da ottenere risultati rilevanti a partire da diversi punti di accesso tematici. Informazioni riguardanti oggetti, luoghi, eventi, persone e tipi possono essere recuperate e visualizzate in vari modi, per esempio su una linea temporale o su una mappa, elencate e filtrate per mezzo di viste e faccette sui campi più comuni ("what, where, when"), scaricate, condivise o riutilizzate in attività di comparazione con altri risultati simili. Riferimenti e URL indicanti la provenienza di ogni singolo frammento di conoscenza dagli archivi originari saranno costantemente disponibili per consentire agli utenti di risalire ai dati originari per consultarli direttamente alla fonte.

Questa fase avanzata di sviluppo, una volta raggiunta, rappresenterà la migliore manifestazione di quella «integrazione delle infrastrutture di dati derivanti dalla ricerca archeologica esistente», «attraverso nuove e potenti tecnologie», che è l'obiettivo principale del progetto ARIADNE.



**Fig. 1.** L'infrastruttura di integrazione definita da ARIADNE

## Ringraziamenti

Il presente lavoro è stato supportato dal progetto ARIADNE, finanziato dalla Commissione Europea (contratto 313193) nel contesto del VII Programma Quadro di Ricerca e Sviluppo, bando INFRA-2012-1.1.3. Le opinioni dell'autore non riflettono necessariamente quelle della Commissione Europea.

## Bibliografia

- [1] J. Richards, D. Tudhope and A. Vlachidis. 2015. Text Mining in Archaeology: Extracting Information from Archaeological Reports. In *Mathematics in Archaeology*, Juan Barcelo and Igor Bogdanovic (Eds.). CRC Press, Chapter 12.
- [2] M. Doerr. The CIDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability of Meta- data. *AI Magazine* 24, 3 (2003), 75–92.
- [3] C. Binding and D. Tudhope. 2016. Improving Interoperability using Vocabulary Linked Data. *International Journal on Digital Libraries* 17, 1 (2016), 5–21.

- [4] CRMarchaeo: the Excavation Model:  
[http://www.ics.forth.gr/isl/index\\_main.php?l=e&c=711](http://www.ics.forth.gr/isl/index_main.php?l=e&c=711)
- [5] 3M The Mapping Memory Manager. Disponibile: <http://www.ics.forth.gr/isl/3M>
- [6] A. Felicetti, T. Scarselli, M.L. Mancinelli, F. Niccolucci (2013). Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema, Vladimir Alexiev, Vladimir Ivanov, Maurice Grinberg (eds.): Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) Workshop, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), Valetta, Malta, September 26, 2013, CEUR-WS.org/Vol-1117, pp 11-22