

Modelli e strumenti semantici per la documentazione dell'Heritage Science

Lisa Castelli¹, Achille Felicetti²

¹INFN, Istituto Nazionale di Fisica Nucleare, ²PIN, Università degli Studi di Firenze

Abstract. L'Heritage Science è una branca multidisciplinare delle scienze applicate che mira allo studio, la valorizzazione e la conservazione dei beni culturali. Le analisi di laboratorio condotte da questa disciplina producono informazioni digitali che necessitano di un'opportuna sistematizzazione, nell'ottica di una loro migliore reperibilità e di un loro efficiente riutilizzo. Questa comunicazione illustra un modello di codifica inteso a documentare tali informazioni per mezzo di ontologie, vocabolari e sofisticati strumenti di metadateazione. Il fine ultimo è la creazione di archivi semantici di conoscenza digitali che, uniti ad opportuni strumenti di ricerca, rendano i dati scientifici facilmente reperibili, accessibili e integrabili, in modo da consentirne il potenziale riutilizzo in ambiti diversi.

Keywords. Heritage Science, Ontologie, Modelli di metadati, Conservazione e restauro, Beni culturali

Introduzione

Che si tratti di stabilire la provenienza o la datazione di un reperto archeologico, di pianificare un restauro o la strategia di conservazione di un dipinto o di un monumento, la scienza è ormai divenuta indispensabile per lo studio dei beni culturali. Strumentazioni, tecniche e metodologie nate per la fisica, la chimica, la biologia, sono oggi utilizzate nei modi più svariati per espandere la conoscenza di oggetti e beni artistici (per comprenderne ad esempio la fattura o la provenienza) e per garantire la qualità della loro vita presente e futura. L'Heritage Science è ormai divenuta una branca interdisciplinare della ricerca, che mette in campo l'analisi scientifica fornendo un contributo insostituibile a supporto dello studio dei beni.

Esistono tuttavia problematiche legate alla gestione dei dati prodotti da questo tipo di ricerca che richiedono particolare attenzione. L'uso di strumenti digitali per le analisi scientifiche, ad esempio, determina la creazione di una mole notevole di informazioni digitali estremamente eterogenee (documenti, immagini, modelli 3D, ecc.) che, una volta utilizzate per il fine immediato della ricerca in atto, finiscono spesso per essere archiviate in modo poco accurato, tanto da risultare in seguito difficilmente reperibili, specialmente da ricercatori esterni all'istituzione che le ha prodotte.

Alcune iniziative europee stanno tentando di fornire possibili soluzioni al problema. In particolare, fra quelle finanziate dall'Unione Europea, ARIADNE (www.ariadne-infrastructure.eu), PARTHENOS (www.parthenos-project.eu) e E-RIHS (www.e-rihs.eu) hanno proposto diversi approcci al tema dell'organizzazione organica di dataset digitali e per la loro indicizzazione, soprattutto per mezzo dell'adozione dei principi FAIR, stabiliti per

promuovere un utilizzo proficuo dei dati: l'utilizzo di standard internazionali e di tecnologie semantiche sono infatti gli unici strumenti in grado di garantirne l'interoperabilità e la riusabilità. L'utilizzo di ontologie, vocabolari e altri strumenti per la creazione di metadati descrittivi è fondamentale in tal senso per conferire coerenza a informazioni per loro natura estremamente variegate, e per descriverne gli elementi essenziali in modo uniforme, tale da renderle poi facilmente reperibili.

1. Il modello semantico

Il modello qui proposto, sviluppato dalla collaborazione fra PIN, INFN e CNR, tutti partner di ARIADNEplus e E-RIHS, ha lo scopo di creare descrizioni standardizzate dei dataset prodotti dall'Heritage Science, indipendentemente dal loro formato digitale, al fine di indicizzarne e archivarne i contenuti in modo razionale ed efficiente per renderli accessibili e rintracciabili (anche online) con strumenti di query complessi. Si tratta, in realtà, di un meta-modello concettuale, ovvero un formato di scambio indipendente da ogni piattaforma o software, progettato per descrivere il contenuto dei dataset e le informazioni relative alle persone, le istituzioni, gli eventi e gli strumenti coinvolti nella loro produzione in un formato standard basato su ontologie e vocabolari controllati.

Da un punto di vista tecnico, ciò che viene generato per ogni dataset (per mezzo di interfacce utente o di procedure semi automatiche in grado di attingere a metadati esistenti) è una serie di moduli XML definiti secondo uno schema XSD appositamente predisposto. Lo schema implementa un meta-formato, lineare ed intuitivo sul lato utente, dal quale è poi possibile generare ulteriori contenuti, in formati relazionali anche complessi (Fig. 1), da utilizzare con qualunque archivio o base di conoscenza digitale, inclusi i database SQL e quelli semantici di ultima generazione, quali il DIGILAB di E-RIHS. Il livello più articolato, quello semantico, è implementabile attraverso un meccanismo di mappatura e conversione (già definito nel modello stesso) per la trasformazione dei metadati in formato CIDOC CRM (RDF), l'ontologia di alto profilo scelta per questa sperimentazione (<http://cidoc-crm.org>). I metadati descrittivi così generati sono consultabili con sistemi di query semantiche e perfettamente integrabili con metadati simili provenienti da ambiti legati all'Heritage Science (archeologia, storia dell'arte, conservazione beni culturali, ecc.). Questo consente la creazione di grafi semantici complessi e interdisciplinari, un risultato che è sempre stato fra gli obiettivi di PARTHENOS e che è alla base della creazione del DIGILAB di E-RIHS.

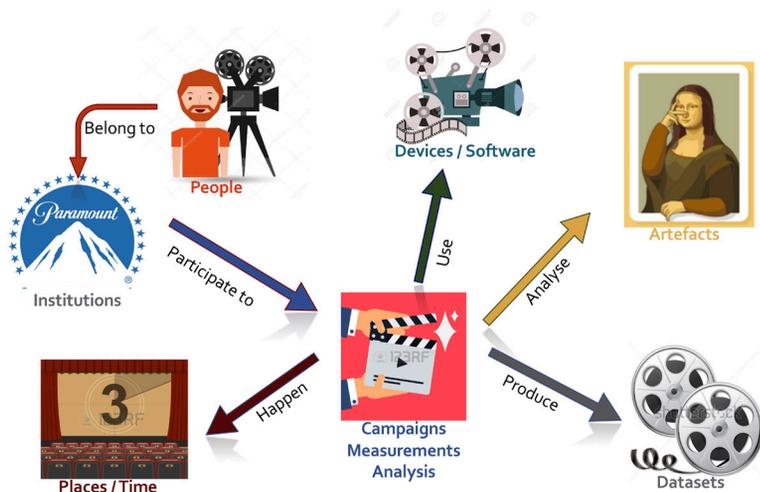


Fig. 1
Moduli e relazioni del modello semantico

2. Analisi e codifica di dati scientifici

Il modello è stato creato dopo un'attenta analisi di sistemi di metadati scientifici esistenti nel panorama internazionale, quali CERIF (<http://www.eurocris.org/cerif/main-features-cerif>) e OBOE (<http://semtools.ecoinformatics.org/oboe>), e una vasta ricognizione dei dataset contenenti risultati analitici di alcune importanti istituzioni scientifiche, impegnate nell'Heritage Science ed interessate alla definizione di un sistema di metadati funzionale. In particolare, per la sua creazione sono stati investigati dataset di laboratorio del CNR, dell'INFN, dell'Opificio delle Pietre Dure, dell'ISCR, ponendo attenzione al loro formato di codifica e al metodo di archiviazione utilizzato.

Le entità della prima release del nostro modello sono state utilizzate dai ricercatori INFN in stretta collaborazione con gli esperti informatici del PIN. INFN è attivo nell'ambito della Heritage Science tramite la sua rete per i beni culturali, INFN-CHNet, costituita da circa 20 laboratori diffusi sul territorio italiano, con competenze molto eterogenee nell'ambito della diagnostica sui beni culturali: fra le tecniche applicate e perfezionate ogni giorno dai laboratori di rete possono essere citate per esempio tecniche a raggi X, quali radiografia, tomografia e XRF, datazioni tramite radiocarbonio o termoluminescenza e analisi con fasci di particelle accelerate. La varietà di indagini diagnostiche disponibili presso i laboratori di rete rende i dati acquisiti da CHNet un buon banco di prova per testare l'efficacia del modello proposto.

Test preliminari sono stati eseguiti applicando il modello ad una serie di dataset di CHNet derivanti da analisi diagnostiche di tipo diverso, provando a individuare, per ogni tipo di analisi, quali fossero i metadati rilevanti ai fini di una ricerca successiva nel database. L'esito dei primi test ha dato esito decisamente positivo, tanto da incoraggiare il prosieguo della sperimentazione.

3. Conclusioni

Il modello, pur se ancora in fase di sviluppo, ha dimostrato di essere uno strumento versatile e sufficientemente capace di descrivere dataset scientifici già nella sua forma attuale. I test effettuati hanno reso inoltre possibile una valutazione del suo potenziale e degli sviluppi necessari per una sua ulteriore evoluzione. Nel prossimo periodo, una nuova serie di test verrà effettuata su nuove tecnologie diagnostiche, sia di INFN che di altri laboratori, al fine di consolidare il modello e di ampliarlo con la definizione di nuove entità e relazioni.

Il fine ultimo è la creazione di uno strumento flessibile in grado di catturare l'essenza di ogni tipo di dataset scientifico (di ambito Heritage Science ma non solo) e di fornirne una completa descrizione formale. L'ontologia CIDOC CRM e la sua estensione scientifica CRMsci (<http://cidoc-crm.org/crmsci>), unita all'utilizzo di strumenti terminologici utilizzati per la codifica dei metadati, consentirà una perfetta integrazione con informazioni dello stesso tipo, generate in altri domini. Particolare attenzione sarà dedicata all'implementazione di interfacce utente user friendly, in grado di garantire un'efficiente raccolta dei dati iniziali e una rapida consultazione degli archivi integrati. L'ultimo importante step riguarderà lo sviluppo della piattaforma digitale e del repository per l'archiviazione dei dati stessi, oltre che dei metadati: un processo che si concretizzerà nel DIGILAB, attualmente in fase di progettazione in E-RIHS.

Riferimenti bibliografici

Nicolucci F. et al. (2017), Condivisione dei dati sui beni culturali: DIGILAB, l'esperienza di ARIADNE e di E-RIHS. Proceedings of GARR Conference, pp.129–132, Venice (2017). DOI: 10.26314/GARR-Conf17-proceedings-24

Pezzati L., Felicetti A. (2017), DIGILAB: A New Infrastructure for Heritage Science, ERCIM NEWS 111, 2017

<https://ercim-news.ercim.eu/en111/special/digilab-a-new-infrastructure-for-heritage-science>

CHNet – La rete INFN dedicate ai beni culturali: <http://chnet.infn.it>

Doerr M., Kritsotaki A., Rousakis Y. (2015), Definition of the CRMsci. An Extension of CIDOC-CRM to support scientific observation. Available at: <http://www.ics.forth.gr/isl/CRMext/CRMsci/docs/CRMsci1.2.3.pdf>

Nicolucci F., Felicetti A. (2018), A CIDOC CRM-based Model for the Documentation of Heritage Sciences, Paper accepted at Digital Heritage Congress 2018, 26-30 October 2018, San Francisco, USA

Autori



Lisa Castelli - lisa.castelli@fi.infn.it

Lisa Castelli è una fisica impegnata dal 2013 nell'ambito delle tecnologie applicate ai beni culturali, in particolare nello sviluppo e applicazione di strumentazione portatile basata sui raggi X per la caratterizzazione elementare dei materiali nelle opere d'arte. Dal 2015 si occupa della gestione e organizzazione delle attività della rete per i beni culturali dell'Istituto Nazionale di Fisica Nucleare, INFN-CHNet, che raccoglie una ventina di strutture dell'Ente con competenze di vario tipo nell'ambito dell'archeometria.

Achille Felicetti - achille.felicetti@pin.unifi.it

Achille Felicetti è un archeologo attivo fin dal 2004 nello sviluppo e l'applicazione di nuove tecnologie per la codifica, la condivisione e l'integrazione di dati e nella definizione di ontologie e strumenti semantici per la modellazione di informazioni nell'ambito dei Beni Culturali. Ha partecipato a diverse iniziative europee e coordinato gruppi di ricerca e sviluppo in progetti quali ARIADNE, per l'interoperabilità di dati archeologici, PARTHENOS ed EO-SCpilot, per la definizione di servizi e tecnologie di Natural Language Processing. È parte del team di sviluppo delle estensioni CIDOC CRM CRMarchaeo, per la codifica di dati di scavo, e CRMtex, per l'integrazione di informazioni epigrafiche.

