

Condivisione dei dati sui beni culturali: DIGILAB, l'esperienza di ARIADNE e di E-RIHS

Franco Niccolucci¹, Carlo Meghini², Achille Felicetti¹, Luca Pezzati³

¹ PIN, Università degli Studi di Firenze, ² CNR-ISTI, ³ CNR-INO

Abstract. Sulla base dell'esperienza del progetto ARIADNE, che ha realizzato un catalogo dei dataset archeologici in Europa, l'infrastruttura europea di ricerca E-RIHS (European Research Infrastructure on Heritage Science) sta realizzando un sistema integrato di gestione dei dati relativi alle scienze del patrimonio culturale, denominato DIGILAB, che coprirà i settori della conservazione, del restauro e in generale della ricerca sui beni culturali. Questa comunicazione illustrerà gli aspetti principali, che realizzano i principi FAIR e pubblicano i dati relativi a un settore finora largamente inaccessibile on-line.

Keywords. DIGILAB, Infrastrutture dati, Conservazione, Restauro, Beni Culturali

Introduzione

Il progetto ARIADNE (www.ariadne-infrastructure.eu), coordinato da PINUniversità di Firenze, ha realizzato un catalogo dei dataset archeologici in Europa che raccoglie i metadati di circa due milioni di dataset (Meghini et al. 2017). Questi comprendono report, immagini, database relativi alle ricerche archeologiche in tutti i paesi europei, indicizzati secondo uno schema dati comune. Il catalogo di ARIADNE rende così reperibili, accessibili e interoperabili dati forniti da oltre trecento istituzioni distribuite sul territorio europeo, e pubblicati da oltre venti centri di archeologia digitale. La ricerca nel catalogo è organizzata per luogo, periodo temporale e tipologia di contenuto dei dataset, oltre che a testo libero sui metadati di catalogo.

Le difficoltà linguistiche – i dati testuali che sono la grande maggioranza dei contenuti utilizzano una quindicina di lingue diverse – sono state affrontate con la creazione di vocabolari multilingue specializzati. Il sistema ha richiesto anche la creazione di un sistema di allineamento dei diversi periodi storici, collegato alla localizzazione geografica: com'è noto, ad esempio, l'età del ferro in Italia termina nel III sec. a.C. mentre in Gran Bretagna si conclude con la conquista romana (I sec. d.C.) e in Irlanda con le invasioni vichinghe (V sec. d.C.).

Il consorzio ARIADNE continua la sua attività anche dopo la fine del finanziamento europeo, sia raccogliendo nuovi elementi per il suo catalogo sia aggiornando il sistema di gestione. Infatti, il catalogo è stato inizialmente realizzato utilizzando uno schema dati apposito (ARIADNE Catalog Data Model, ACDM). Successivamente, il progetto PARTHENOS (www.parthenos-project.eu), un cluster di progetti e iniziative digitali in campo umanistico, ha riformulato il modello come estensione di un'ontologia standard, il CIDOC CRM, producendo così il PARTHENOS Entities Data Model (PEM) (Aloia et al. 2017). Il PEM è idoneo a gestire metadati relativi a dataset e servizi per un ampio spettro di discipline nel campo delle scienze umane. Il gruppo di lavoro originario di ARIADNE, che ha partecipato attivamente allo sviluppo di PEM, ha quindi prodotto un mapping da ACDM a PEM e ha realizzato la conversione

dei dati, che sono ora catalogati secondo lo schema più aggiornato. Anche se PEM è centrato sui dati relativi alle scienze umane (linguistica, digital humanities, archeologia, architettura, storia, conservazione, e così via), è estensibile con profili specializzati per i singoli settori che ne migliorano l'efficacia, mantenendo allo stesso tempo la compatibilità interdisciplinare.

L'esperienza di E-RIHS (www.e-rihs.eu), un'infrastruttura di ricerca a guida CNR già inserita nella roadmap ESFRI (European Strategy Forum on Research Infrastructures) e attualmente nella fase di preparazione di un ERIC (European Research Infrastructure Consortium), nasce invece da una serie di progetti per la ricerca sulle scienze della conservazione, inizialmente concepiti senza la componente digitale. Attraverso una sequenza decennale di progetti europei (EU-ARTECH, CHARISMA e infine IPERION CH, tuttora in corso), tutti coordinati dal CNR, quest'iniziativa si è sviluppata basandosi sull'accesso a laboratori mobili (MOLAB), fissi (FIXLAB) e archivi (ARCLAB) operanti nel campo della ricerca sulla conservazione del patrimonio; dopo l'incontro con ARIADNE e l'approvazione di E-RIHS comprende ora anche un'infrastruttura digitale (DIGILAB).

1. DIGILAB

DIGILAB è basato, come il catalogo di ARIADNE, sulla “federazione” di archivi locali, dove ricercatori e operatori del settore depositano i risultati digitali della propria attività. Il servizio principale è il catalogo, simile a quello di ARIADNE e basato su un profilo di PEM adattato alle specifiche esigenze della comunità scientifica di riferimento. La funzione di acquisizione (ingestion) dei metadati utilizza la stessa tecnologia di quelle di ARIADNE, come pure la funzionalità di ricerca nel catalogo che si basa su Elasticsearch.

DIGILAB è ispirato ai principi FAIR: rende i dati reperibili attraverso il suo catalogo; ne provvede l'accesso attraverso un sistema d'identità federata; ne supporta l'interoperabilità sia a livello di metadati organizzati secondo PEM, sia a livello di dati, ove possibile, proponendo uno schema basato su un profilo specializzato di CRMsci (Doerr et al. 2015), l'estensione dello standard CIDOC CRM per la gestione dei dati scientifici, e compatibile anche con CRMba (Ronzino et al. 2016), l'estensione CRM per l'architettura, e CRMarcheo (Doerr et al. 2016), l'analoga estensione per l'archeologia. Per quanto riguarda il riuso, DIGILAB propone un approccio basato sul Virtual Research Environment (VRE) fornito da D4Science (Assante et al. 2016), offrendo in ambiente cloud una serie di servizi utili ai ricercatori del settore. Questi servizi coprono una gamma di attività tipiche della ricerca, quali ad esempio la creazione di grafici e diagrammi su dati numerici di analisi chimiche o fisiche, l'elaborazione di immagini multi-spettrali e l'arricchimento dei metadati dei testi (ad esempio relazioni tecniche), utilizzando sistemi di Natural Language Processing (NLP) e Named Entity Recognition (NER) con l'uso di vocabolari specializzati. Alcuni di questi servizi sono attualmente già in fase di test. DIGILAB fornirà agli utenti anche servizi di corredo, quali un formulario per la compilazione on-line del Data Management Plan (DMP), ormai obbligatorio per i progetti finanziati dalla Commissione Europea, e un sistema chiavi in mano per la realizzazione di archivi locali, personalizzabile secondo le esigenze dell'utente ma basato sullo schema dati standard del progetto.

L'organizzazione decentrata degli archivi consente di non affrontare nell'immediato il problema dello storage dei dati, che è affidata ai nodi locali di E-RIHS o alle rispettive comunità di

ricercatori, mentre DIGILAB necessita di risorse modeste in quanto deve contenere soltanto i relativi metadati. Tuttavia, si pone il problema di organizzare in modo efficiente il nodo italiano, assicurando le necessarie risorse a livello nazionale per il deposito dei dati prodotti dalla ricerca nel settore.

DIGILAB è una realizzazione di PIN-Università di Firenze, già coordinatore di ARIADNE, e CNR, attraverso i suoi istituti ISTI per la parte informatica e INO, coordinatore di E-RIHS, che organizza la partecipazione di vari istituti di scienze dei beni culturali (ISTM, ICVBC e altri) e di istituti specializzati del MIBACT come l'OPD.

2. Piano di lavoro e successivi sviluppi

DIGILAB è attualmente in fase sperimentale; se ne stanno sviluppando le componenti e verificando le funzionalità su una serie di archivi di test. I moduli NLP e NER fanno parte di un dimostratore inserito in EOSCpilot (www.eosc-pilot.eu), un progetto pilota sulla realizzazione di EOSC (European Open Science Cloud).

I moduli necessari a DIGILAB richiedono tutta una gamma di attività differenti: alcune di tipo tecnologico, come l'implementazione del VRE; altre di tipo semantico, come la realizzazione dei vari schemi di dati che derivano e insieme impattano sui protocolli sperimentali adottati da E-RIHS; altre ancora di carattere organizzativo, come l'identità federata. I moduli tecnologici saranno rilasciati via via che verranno prodotti, e ci si aspetta di avere le prime funzionalità disponibili per la fine del 2017. Per gli inizi del 2018 sarà invece rilasciata la prima versione della parte semantica.

Una prima versione completa dell'intero sistema sarà disponibile entro il 2018, procedendo in parallelo con la creazione e l'adattamento degli archivi locali, di cui allo stato attuale solo alcuni sono in funzione. In ogni modo, è previsto che il sistema sia pienamente funzionante quando anche E-RIHS si costituirà come ERIC. Per ottenere quest'obiettivo, è indispensabile che le risorse necessarie a livello nazionale siano rese pienamente disponibili, in modo da offrire ai partner europei, e probabilmente a un'audience mondiale, il sistema avanzato di gestione dei dati della ricerca sopra descritto.

Riferimenti bibliografici

- Aloia N. et al. (2017), D5.2 Design of the Joint Resource Registry. Available at http://www.parthenosproject.eu/Download/Deliverables/5.2_Report_on_design_Joint_Resource_Registry.pdf
- Assante M., Candela L., Manghi P., Pagano P., Castelli D. (2015), Providing research infrastructures with data publishing. ERCIM News, Issue 100, January 2015.
- 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>
- Doerr M., Felicetti A., Hermon S. et al. (2016), Definition of the CRMarchaeo. An Extension of CIDOC CRM to support the archaeological excavation process. Available at: http://www.ics.forth.gr/isl/CRMext/CRMarchaeo/docs/CRMarchaeo_v1.4.pdf

Meghini C. et al. (2017), ARIADNE: A research infrastructure for archaeology. *ACM Journal on Computing and Cultural Heritage* 10(3):1-27, August 2017.

Ronzino P., Niccolucci F., Felicetti A., Doerr M. (2016), CRMba a CRM extension for the documentation of standing buildings. *Int. J. on Digital Libraries* 17(1): pp. 71-78.

Autori



Franco Niccolucci franco.niccolucci@gmail.com

Franco Niccolucci è il direttore del laboratorio VAST-LAB presso il PIN di Prato, e il coordinatore del progetto europeo PARTHENOS, cluster delle infrastrutture di ricerca nel settore Digital Humanities e Cultural Heritage. Ha coordinato l'infrastruttura ARIADNE nel campo dell'archeologia digitale, che ha realizzato un registry di oltre 2.000.000 di dataset archeologici. In passato ha coordinato vari progetti europei nel campo dei beni culturali. Matematico come formazione, è stato professore all'Università di Firenze fino al 2008 e successivamente, fino al 2013, direttore dello Science and Technology in Archaeology Research Center presso il Cyprus Institute a Nicosia, Cipro.

Carlo Meghini carlo.meghini@isti.cnr.it

Carlo Meghini è ricercatore capo presso il CNR-ISTI e coordinatore del gruppo Digital Libraries del laboratorio NeMIS di ISTI. La sua area di ricerca comprende la progettazione concettuale, le digital library e le infrastrutture di ricerca per le scienze umane. È coinvolto in progetti europei dal 1988, nelle aree delle digital library e della digital preservation. È stato il coordinatore della CSA PRELIDA per la preservation di Linked Data. Dal 2007 partecipa alla creazione di Europeana, la digital library europea, curando gli aspetti scientifici del progetto. A partire dal progetto ARIADNE, è coinvolto nella ricerca e nello sviluppo di infrastrutture di ricerca. Ha pubblicato oltre 90 articoli scientifici in riviste, libri e conferenze internazionali.



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 EOSCpilot, per la definizione di servizi e tecnologie di Natural Language Processing. È parte del team di sviluppo di CRMarchaeo, l'estensione CIDOC CRM per la codifica di informazioni di scavo.

Luca Pezzati luca.pezzati@cnr.it

Luca Pezzati è un fisico e uno specialista di ottica. Dal 1995 è con INO-CNR (l'Istituto Nazionale di Ottica del Consiglio Nazionale delle Ricerche d'Italia) dove è attualmente ricercatore senior. Ha avviato il Gruppo Beni Culturali (INA) nel 1998 e lo ha coordinato per 14 anni. Ha gestito molti progetti di ricerca per CNR nel campo delle tecnologie ottiche applicate al patrimonio culturale. È coordinatore di E-RIHS, l'Infrastruttura europea di ricerca per la scienza del patrimonio e del progetto d'integrazione IPERION CH. Ha coordinato il nodo nazionale di DARIAH ERIC, DARIAH-IT, dal 2013 al novembre 2016.

