

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

Franco Niccolucci (PIN), Carlo Meghini (CNR-ISTI), Achille Felicetti (PIN), Luca Pezzati (CNR-INO)

Abstract. Basato sull'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, che coprirà i settori della conservazione, del restauro e in generale della ricerca sui beni culturali. Questa comunicazione ne illustrerà gli aspetti principali, che realizzano i principi FAIR e pubblicano i dati relativi a un settore finora largamente inaccessibile on-line.

1. Introduzione

Il progetto ARIADNE (www.ariadne-infrastructure.eu), coordinato da PIN-Università di Firenze, ha realizzato un catalogo dei dataset archeologici in Europa che raccoglie i metadati di circa due milioni di dataset [1]. 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 trovabili, accessibili e interoperabili dati forniti da oltre trecento istituzioni e pubblicati da una ventina di centri di archeologia digitale. La ricerca nel catalogo è organizzata per luogo, tempo e contenuto, 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 stati affrontati con la creazione di vocabolari multilingue specializzati. Il sistema ha richiesto anche la creazione di un sistema di confronto 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.). 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) basato su DCAT. Successivamente, il progetto PARTHENOS (www.parthenos-project.eu), un cluster di progetti e iniziative digitali in campo umanistico, ha sviluppato un sistema più evoluto, chiamato PARTHENOS Joint Resource Registry Data Model (PDM) [2], basato sullo standard CIDOC CRM e 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 PDM, ha quindi prodotto un mapping da ACDM a PDM e ha realizzato la conversione dei dati, che sono ora catalogati secondo lo schema più aggiornato. Anche se PDM consente la gestione integrata dei dati relativi a tutte le scienze umane (linguistica, digital humanities, archeologia, architettura, storia, conservazione, e così via), è possibile aggiungere 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).

2. 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 PDM 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 *trovabili* 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 PDM, sia a livello di dati, ove possibile, proponendo uno schema basato su un profilo specializzato di CRMsci [3], l'estensione dello standard CIDOC CRM per la gestione dei dati scientifici, e compatibile anche con CRMba [4], l'estensione CRM per l'architettura, e CRMarcheo [5], l'analogica estensione per l'archeologia. Per quanto riguarda il *riuso*, DIGILAB propone un approccio basato sul Virtual Research Environment (VRE) fornito da D4Science [6], offrendo in ambiente cloud una serie di servizi utili ai ricercatori del settore. Questi 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 multispettrali 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.

3. 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 EOscipilot (www.eoscipilot.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.

Bibliografia

- [1] C. Meghini et al. *ARIADNE: A research infrastructure for archaeology*. Journal on Computing and Cultural Heritage 10(3):1-27 · August 2017.
- [2] N. Aloia et al *D5.2 Design of the Joint Resource Registry*. Available at http://www.parthenos-project.eu/Download/Deliverables/D5.2_Report_on_design_Joint_Resource_Registry.pdf
- [3] M. Doerr, A. Kritsotaki, Y. Rousakis *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>
- [4] P. Ronzino, F. Niccolucci, A. Felicetti, M. Doerr *CRMba a CRM extension for the documentation of standing buildings*. Int. J. on Digital Libraries 17(1): 71-78 (2016)
- [5] M. Doerr, A. Felicetti, S. Hermon et al. *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
- [6] M. Assante, L. Candela, P. Manghi, P. Pagano and D. Castelli *Providing research infrastructures with data publishing*. ERCIM News, Issue 100, January 2015