

ARIADNE E GLI OPEN DATA: COME TRASFORMARE I DATI ARCHEOLOGICI DA OPEN A “FAIR”

1. INTRODUZIONE

Il 1 febbraio 2013 ha avuto inizio il progetto “ARIADNE – Advanced Research Infrastructure for Archaeological Dataset Networking in Europe” sull’integrazione dei dati digitali archeologici. Finanziato dalla Commissione Europea per quattro anni, e quindi fino al 31 gennaio 2017, il progetto ha avuto come scopo quello di superare la frammentazione degli archivi archeologici digitali, locali e nazionali, attraverso la creazione di un sistema unico di ricerca e un unico punto di accesso (NICCOLUCCI 2013; NICCOLUCCI, RICHARDS 2013a; NICCOLUCCI, RICHARDS 2013b). L’idea alla base di ARIADNE nasce dalla constatazione che le tecnologie informatiche sono state utilizzate in ambito archeologico sin dagli anni Ottanta per conservare e organizzare la documentazione di scavo, sia da parte di singoli ricercatori sia a cura di agenzie e uffici pubblici a ciò delegati, come ad esempio l’Istituto Centrale per il Catalogo e la Documentazione (ICCD) in Italia e l’Institut National de Recherches Archéologiques Préventives (INRAP) in Francia.

Si sono progressivamente allentate le restrizioni, non di rado immotivate, che impedivano l’accesso pubblico a tali dati; tuttavia lo sviluppo separato di queste banche dati e il fatto che le estensioni geografiche delle “culture” archeologiche del passato e i moderni confini politici sono di rado compatibili, rendono tuttora estremamente complicata, se non impossibile, un’analisi dei dati riferiti ad aree a cavallo dei confini moderni, come ad esempio uno studio sulle province romane. Anche la terminologia usata non è uniforme, al di là delle differenze linguistiche, e i periodi temporali sono necessariamente diversi: un mercante che attraversasse la Manica all’inizio della nostra era si muoverebbe non solo attraverso lo spazio, ma anche indietro nel tempo, iniziando il suo viaggio da un porto romano nella Gallia dell’età romana e terminandolo nel porto britannico di arrivo nell’Età del Ferro. Se quest’affermazione può far sorridere un lettore umano, viene presa invece sul serio da una macchina non addestrata a sapere che le definizioni dei periodi temporali hanno una traduzione in termini di tempo assoluto esclusivamente locale rispetto a una regione geografica più o meno ampia. Di conseguenza, come dimostrato dal progetto ARENA sin dal 2002, l’interoperabilità degli archivi archeologici richiede un lavoro non indifferente di organizzazione e di unificazione.

Lo strumento operativo individuato dalla Commissione Europea per affrontare questa problematica è quello dell’infrastruttura di ricerca (Research

Infrastructure). Nella terminologia europea, un'infrastruttura di ricerca è definita come il complesso delle strutture fisiche, delle risorse umane e finanziarie e dei servizi destinati a supportare la ricerca in un determinato settore o area scientifica¹. Si tratta dunque di un'estensione dell'uso tradizionale del termine "infrastruttura", che è comunemente utilizzato, anche in ambito di ricerca, per la sola componente materiale: un laboratorio, un grande telescopio, un acceleratore di particelle. Si pone, invece, l'accento sull'unitarietà dell'infrastruttura materiale, delle risorse (umane e finanziarie) destinate al suo funzionamento e ai servizi che essa può così erogare alla comunità scientifica. Accanto a questo concetto c'è l'idea tipicamente europea della transnazionalità, cioè del concorso di enti e istituzioni di più stati europei all'ideazione, alla progettazione, alla realizzazione e alla gestione dell'infrastruttura di ricerca.

Questo concetto e la sua realizzazione sono esclusivamente europei: in nessun'altra parte del mondo esistono i necessari strumenti normativi e finanziari, ma soprattutto non c'è la mentalità indispensabile per una collaborazione così stretta fra soggetti diversi e potenzialmente in concorrenza, cioè i partner dell'infrastruttura. I temi su cui realizzare infrastrutture di ricerca europee vengono selezionati in base alle indicazioni di un comitato di esperti chiamato ESFRI – European Scientific Forum on Research Infrastructures, dove tutti i paesi europei e tutti gli ambiti disciplinari sono rappresentati². Questo organismo pubblica a scadenze biennali una *roadmap* in cui evidenzia le criticità e le strategie della ricerca europea, individuando fra i vari ambiti scientifici quelli prioritari³. Per tali ambiti, il percorso che viene seguito per realizzare un'infrastruttura di ricerca inizia con un'attività, finanziata dalla Commissione Europea, cosiddetta di "integrazione", che realizza le pre-condizioni per implementare l'infrastruttura stessa, attraverso la creazione e il rafforzamento della rete scientifica che ne è alla base e la realizzazione delle ricerche necessarie a mettere in atto gli strumenti organizzativi e tecnologici necessari.

Quando la comunità scientifica impegnata in questo lavoro ha raggiunto un grado sufficiente di maturità e ha conseguito tali pre-condizioni, è il momento di verificare la disponibilità delle risorse finanziarie, che devono essere fornite dai governi nazionali secondo le rispettive priorità. Se entrambe le condizioni sono verificate, si può dar luogo agli accordi e alla realizzazione dell'infrastruttura di ricerca europea e alla creazione del soggetto che la gestisce, chiamato ERIC – European Research Infrastructure Consortium. In ogni caso, resta il lavoro fatto in termini di integrazione che può essere mantenuto nel tempo con un minimo d'investimento.

¹ https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=about.

² <http://www.esfri.eu/about/>.

³ <http://www.esfri.eu/roadmap-2016/>.

È chiaro che le ricerche che richiedono grandi infrastrutture, come l'astronomia o la fisica delle particelle, svolgono un ruolo preponderante in questa strategia globale, proprio perché richiedono grandi investimenti sulla componente materiale dell'infrastruttura. Un altro criterio di priorità riguarda quelle ricerche per le quali è indispensabile una collaborazione internazionale: è tipico il caso delle ricerche sul clima, le cui variazioni non si arrestano ai confini politici. Infine, tematiche di grande importanza sociale come la ricerca sul cancro costituiscono un altrettanto ovvio terreno prioritario su cui investire risorse. Questo tipo d'infrastrutture sono tipicamente basate sulle scienze naturali e comprendono grandi strutture o reti distribuite di rilevamento e di sperimentazione. A prima vista parrebbe che per le scienze umane e per l'archeologia e i beni culturali ci siano dunque poche possibilità di essere prese in considerazione. Non è stato così, invece, ed esse sono state inserite nella roadmap dell'ESFRI in un gruppo denominato "Scienze Umane e Sociali". Fra queste, dunque, ha trovato spazio anche l'archeologia, attraverso l'indicazione di un'attività d'integrazione focalizzata sulla componente digitale, fase preliminare alla futura costituzione di una possibile infrastruttura permanente di ricerca.

Questo approccio ha determinato l'apertura di un bando nel 2012, attraverso cui è stato selezionato appunto il progetto ARIADNE. Ad ARIADNE partecipano con i propri dati istituzioni pubbliche nazionali come il Ministero dei Beni e delle Attività Culturali e del Turismo (MiBACT) per l'Italia, l'INRAP per la Francia e il DANS - Data Archiving and Networked Services per l'Olanda; le Accademie delle Scienze austriaca, bulgara e ceca; il Museo Nazionale Ungherese; il Deutsches Archäologisches Institut (DAI); il Servizio Nazionale Archivi svedese; l'Archaeological Data Service dell'Università di York per il Regno Unito. Ci sono inoltre importanti centri di ricerca incaricati di sviluppare la tecnologia necessaria. Tutti insieme hanno messo a disposizione circa due milioni di archivi archeologici, in gran parte costituiti da rapporti di scavo non pubblicati, corredati da immagini, disegni, mappe e così via.

2. ARIADNE E OPEN DATA

Presupposto indispensabile perché i dati archeologici digitali siano presi in considerazione da ARIADNE è che siano presenti su Internet. Non rientra infatti fra i compiti del progetto trasferire on-line dati digitali archiviati su CD-ROM, hard disk personali o altri sistemi di archiviazione off-line; né tantomeno digitalizzare dati cartacei, operazione che richiederebbe tempo e risorse non disponibili o meglio utilizzabili in altro modo, a fronte della discutibile utilità di avere on-line qualche scansione di documenti scritti su carta, ma privi di metadati. Si è posto a questo punto il problema se i dati utilizzabili dal progetto debbano essere anche aperti. Per evitare d'impantanarsi in tante

discussioni quante sono le legislazioni europee e le relative interpretazioni, il progetto ha assunto in merito una posizione agnostica. Infatti si è scelto di privilegiare la “scopribilità” degli archivi, cioè di realizzare un catalogo integrato e quanto più possibile completo, rispetto alle modalità di accesso vero e proprio agli archivi, i cui criteri di apertura sono stati lasciati ai rispettivi proprietari. In altri termini, ARIADNE risolve il problema di individuare l'esistenza su Internet di uno o più archivi digitali archeologici corrispondenti a determinati criteri di selezione indicati dall'utente, ad esempio relativi a un determinato argomento, periodo storico o posizione geografica.

Tali archivi vengono anche descritti in un succinto sommario, mentre l'accesso ai dati veri e propri contenuti in tali archivi è soggetto alle regole stabilite dal proprietario dei contenuti, che in effetti variano dall'assoluta apertura, alla necessità di registrazione, fino al possesso di determinati requisiti personali da parte dell'utente, requisiti potenzialmente diversi da archivio ad archivio. Non si è infatti ritenuto di dover realizzare, in questo primo stadio, la cosiddetta identità federata, cioè un sistema unico di registrazione degli utenti che si interfaccia con i diversi fornitori di contenuti e verifica automaticamente il possesso dei requisiti di accesso, qualcosa di simile a Eduroam, il sistema internazionale di roaming dell'accesso a Internet utilizzato presso le università e i centri di ricerca in tutto il mondo. Un tale sistema è comunque nei piani per il futuro, in collaborazione con gli enti e le istituzioni in grado di organizzare tale accesso unificato in modo efficiente. Tanto meno si è ritenuto opportuno aprire una serie di confronti con i vari enti che possiedono dati archeologici per convincerli ad aprirli, ritenendo tale compito di carattere prevalentemente politico e comunque al di là degli obiettivi e compiti specifici del progetto.

Ciò non significa, tuttavia, che all'interno di ARIADNE non vi sia una posizione ferma sulla necessità di rimuovere le tante restrizioni inutili, anacronistiche, illegittime o meschine spesso opposte all'apertura degli archivi. A prescindere da quest'opinione, largamente condivisa in ARIADNE, va comunque considerato che la sola apertura dei dati non è sufficiente. Come per tutti i dati scientifici, valgono anche per quelli archeologici i principi riassunti nell'acronimo FAIR, una sigla azzeccata (*fair* in inglese significa infatti appropriato, giusto) che sta per *Findable – Accessible – Interoperable – Reusable*, cioè trovabili, accessibili, interoperabili e riutilizzabili (FORCE11⁴; WILKINSON *et al.* 2016). Per avere queste caratteristiche i dati devono essere descritti con metadati ricchi d'informazione, raccolti e indicizzati in un apposito registro come quello costruito da ARIADNE. Accessibilità significa possibilità di accesso attraverso una procedura di identificazione e autorizzazione, se necessario, preferibilmente federata a livello globale, come si è già detto, in modo

⁴ <https://www.force11.org/group/fairgroup/fairprinciples/>.

da evitare registrazioni multiple, tante quanti sono gli archivi: un sistema centrale deve quindi gestire gli accessi dialogando con i singoli depositi di dati in modo trasparente per l'utente.

L'interoperabilità dei dati richiede l'uso di standard di codifica sia a livello di archivi che, eventualmente, a livello di singolo record. Per questo scopo ARIADNE ha proposto un Data Model per la descrizione degli archivi (MEGHINI *et al.* c.s.). Tale modello è attualmente in corso di perfezionamento all'interno di un altro progetto denominato PARTHENOS (<http://www.parthenos-project.eu/>), in modo da adattarsi a tutte le necessità dell'archeologia digitale ma anche, più in generale, delle scienze dei beni culturali come la conservazione, l'archeometria, e così via.

Al contempo, il riuso dei dati in campo archeologico è un tema delicato. Se infatti i risultati di un esperimento scientifico sono generalmente considerati validi in assoluto senza limitazioni di tempo o di luogo, i dati di un'analisi scientifica in campo archeologico, o anche le stesse osservazioni svolte nel corso della ricerca, per essere validate richiedono una chiara descrizione delle circostanze in cui l'analisi si è svolta o le osservazioni effettuate. Si tratta quindi di aggiungere i cosiddetti paradata, le informazioni sulla provenienza scientifica di analisi quantitative o qualitative, che possono confermarne l'attendibilità e la pertinenza dei rispettivi dati di base anche a ricerche diverse da quelle per cui sono stati creati.

In conclusione, nell'ambito dell'archeologia digitale gli Open Data consentono di allargare lo spettro delle fonti su cui basare la ricerca, permettendo agli studiosi di avvalersi dei risultati di ricerche di altri a partire dalle fonti primarie, le loro osservazioni e i risultati delle loro analisi. Ciò, a patto che i dati stessi siano organizzati secondo i principi FAIR sopra ricordati.

3. ARIADNE E I PRINCIPI FAIR

ARIADNE ha realizzato un catalogo degli archivi archeologici digitali europei e ha raccolto sinora i metadati relativi a quasi due milioni di questi dataset. L'operazione di raccolta ha richiesto la standardizzazione della descrizione degli archivi stessi, dando luogo a un modello unico di metadati. Ciascun archivio ha conservato la propria fisionomia, struttura e collocazione on-line presso l'istituzione in cui è depositato. Si è poi realizzato un portale che permette di interrogare il catalogo secondo vari criteri (<http://portal.ariadne-infrastructure.eu/>). Questi possono essere definiti anche con un'interfaccia grafica, che permette di definire l'intervallo temporale d'interesse su una linea del tempo, o l'area geografica d'interesse su una carta geografica (Fig. 1). È anche possibile impostare come criterio di ricerca una parola chiave, oppure la tipologia di archivio che interessa. Il portale visualizza i risultati della ricerca effettuata con determinati parametri (Fig. 2; nell'esempio la ricerca è basata

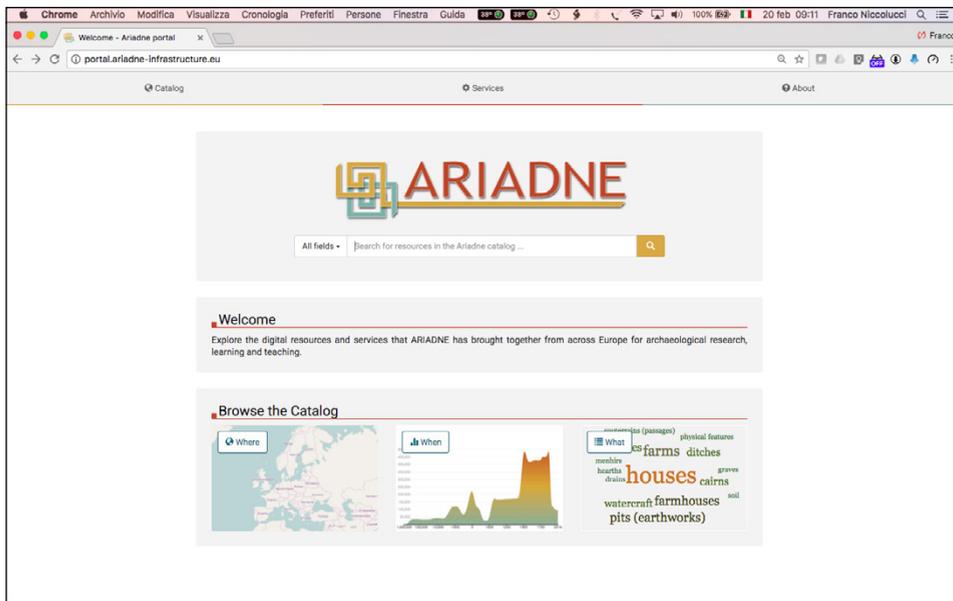


Fig. 1 – La pagina di accesso al portale di ARIADNE.

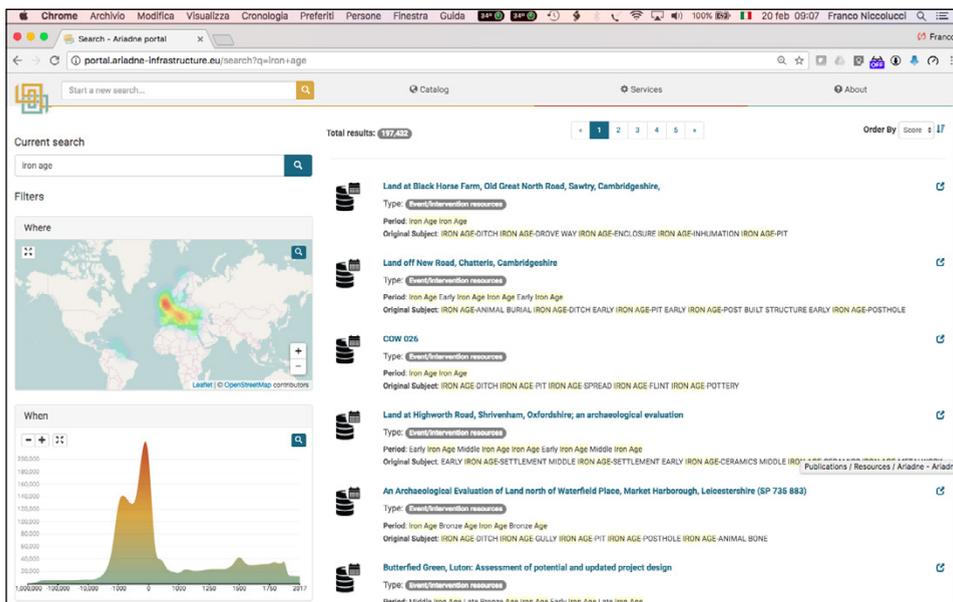


Fig. 2 – I risultati della ricerca con parola-chiave “Iron age”.

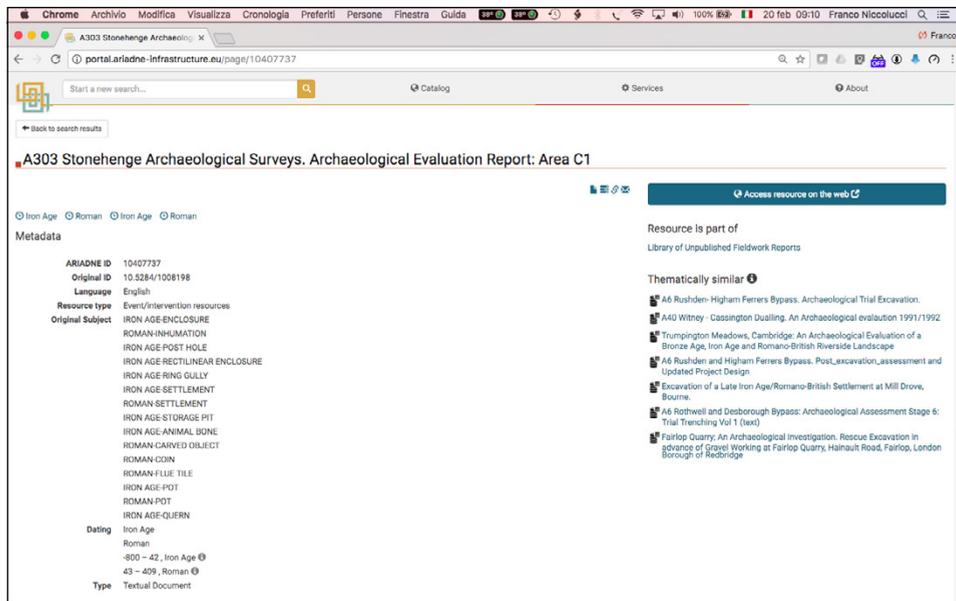


Fig. 3 – I dati di dettaglio dell’archivio selezionato.

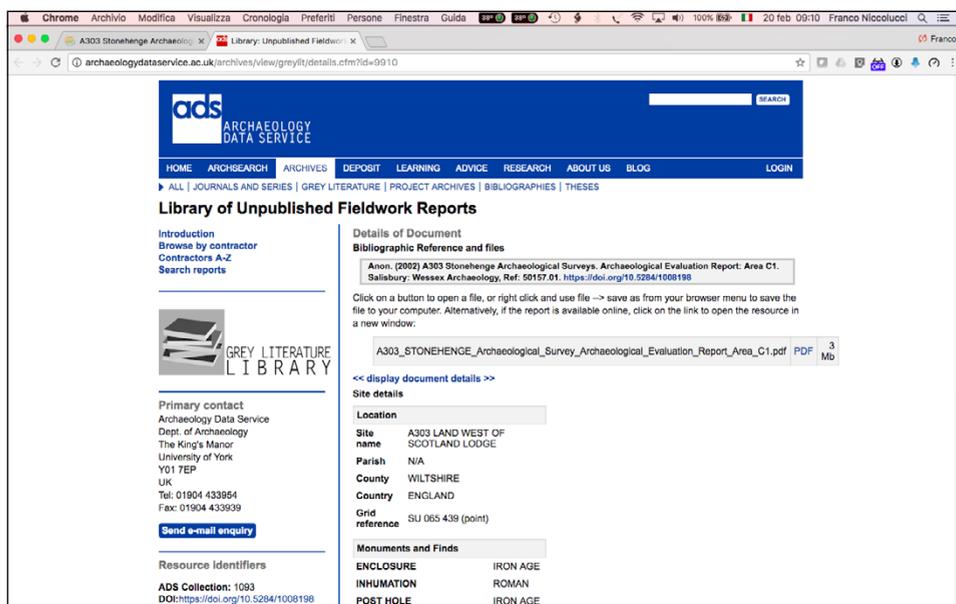


Fig. 4 – La risorsa originale presso ADS.

su “Iron age”) e, successivamente, mediante la loro descrizione sintetica. La ricerca può essere ulteriormente raffinata specificando ulteriori condizioni mediante le opzioni di ricerca presenti nel menù di sinistra e, una volta deciso il caso che interessa, viene mostrata la descrizione sintetica dell’archivio (Fig. 3) che dà infine accesso alla risorsa originale conservata presso il fornitore di contenuti (Fig. 4).

Con questo portale ARIADNE ha realizzato l’apertura degli archivi archeologici digitali secondo i principi FAIR: il portale consente di trovare (F) e accedere (A) agli archivi. Essi sono descritti in modo standardizzato e quindi interoperabile (I), pur conservando ciascuno la propria struttura, e questo ne consente il riuso (R) da parte di altri ricercatori. Il problema della qualità dei dati è stato risolto prendendo in considerazione, per ora, soltanto gli archivi forniti da istituzioni note e di alta reputazione, e includendo nella descrizione sommaria gli elementi atti a comprendere la finalità scientifica, o di altro tipo, per cui sono stati raccolti.

4. IL PORTALE DI ARIADNE: STATISTICHE D’USO

Nell’ultimo anno da quando è stato rilasciato in forma definitiva, il portale di ARIADNE ha avuto oltre 10.000 utenti distinti (il 90% dei quali dall’Europa), con oltre 15.000 sessioni di accesso e la visualizzazione di quasi 70.000 pagine. Si nota un tasso di incremento d’uso crescente nel tempo, com’è logico che sia, via via che il sistema viene migliorato e arricchito, gli eventuali errori corretti, e i potenziali utenti vengono a conoscenza dell’esistenza del portale.

Ipotizzando la stima di 33.000 archeologi in Europa effettuata nell’ambito del progetto “Discovering the Archaeologists of Europe” (DISCO 2014, 6), questo significa che il 27% di tutti gli archeologi europei hanno visitato il portale di ARIADNE, alcuni più di una volta. Considerando che la percentuale di archeologi interessata agli aspetti digitali non è certo totalitaria – non esistono stime a questo proposito, ma la riterremmo sicuramente inferiore al 50% – si può concludere che la maggioranza degli archeologi europei con interessi digitali ha utilizzato ARIADNE almeno una volta nel corso dell’ultimo anno.

5. ARIADNE: LA SOSTENIBILITÀ E IL LAVORO FUTURO

Con la fine del finanziamento europeo le attività di espansione di ARIADNE sono temporaneamente sospese. Se infatti il mantenimento del portale così com’è richiede risorse molto contenute che i partecipanti coinvolti, e in primo luogo il coordinatore PIN, si sono dichiarati disponibili a fornire, l’aggiunta di ulteriori dati necessita di risorse non eccezionali, ma comunque di qualche consistenza. Per ogni nuovo archivio censito nel

catalogo è necessario infatti acquisire le informazioni richieste dal Data Model – per lo più in modalità automatica, ma sempre con supervisione umana per i necessari aggiustamenti e correzioni – e integrarle nel catalogo. Simili considerazioni valgono anche per l’aggiornamento dei dati già censiti. In questo caso, tuttavia, si può ritenere che il lavoro necessario per l’inclusione nel catalogo degli aggiornamenti sia relativamente semplice e possa avvenire con la collaborazione dei fornitori di contenuti, che nella maggior parte dei casi si sono dichiarati disponibili a questo tipo di collaborazione volontaria.

Nel medio e lungo periodo, ARIADNE e il suo portale dovranno trovare fonti di sostentamento permanente. Una possibilità in tal senso è costituita dal nuovo ERIC denominato E-RIHS nel campo delle scienze dei beni culturali, che potrebbe ospitare il catalogo e provvedere al suo aggiornamento nell’ambito della propria attività. Resta comunque del lavoro da fare: aumentare la copertura del progetto, per ora limitata a metà dei paesi europei; migliorare le funzionalità di ricerca; integrare gli archivi in maniera più approfondita; promuovere l’uso del portale fra gli archeologi e, infine, svolgere attività di formazione sulle tematiche dei dati, un settore già avviato con grande successo nel progetto ARIADNE e nell’ambito del quale sono stati formati circa 100 ricercatori, in rappresentanza di un numero rilevante di progetti di ricerca o istituzioni culturali e governative interessate alla filosofia e alle realizzazioni di ARIADNE.

FRANCO NICCOLUCCI

PIN – Servizi Didattici e Scientifici per l’Università di Firenze
franco.niccolucci@gmail.com

BIBLIOGRAFIA

- DISCO 2014, *Discovering the Archaeologists of Europe 2012-2014: Transnational Report*, York Archaeological Trust, November 2014 (http://www.discovering-archaeologists.eu/national_reports/2014/transnational_report.pdf; ultimo accesso: 21/02/2017).
- MEGHINI C., SCOPIGNO R., RICHARDS J.D., WRIGHT H., GESER G., CUY S., FIHN J., FANINI B., HOLLANDER H., NICCOLUCCI F., FELICETTI A., RONZINO P., NURRA F., PAPATHEODOROU C., GAVRILIS D., THEODORIDOU M., DOERR M., TUDHOPE D., BINDING C., VLACHIDIS A. c.s., *ARIADNE: A Research Infrastructure for Archaeology*, «ACM Journal on Computer and Cultural Heritage», 10, 2, in corso di stampa.
- NICCOLUCCI, F. 2013, *Un’infrastruttura di ricerca per l’archeologia: il progetto ARIADNE*, «DigItalia», 2, Roma, MiBACT-ICCU, 154-161.
- NICCOLUCCI F., RICHARDS J.D. 2013a, *ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe. A new project to foster and support archaeological data sharing*, «The European Archaeologist», 39 (<http://e-a-a.org/TEA/TEA39.pdf>; ultimo accesso: 21/02/2017).
- NICCOLUCCI F., RICHARDS J.D. 2013b, *ARIADNE: Advanced Research Infrastructure for Archaeological Dataset Networking in Europe*, «International Journal of Humanities and Arts Computing», 7, 1-2, 70-88 (DOI: 10.3366/ijhac.2013.0082; ultimo accesso: 21/02/2017).
- WILKINSON M.D. et al. 2016, *The FAIR Guiding Principles for scientific data management and stewardship*, «Scientific Data», 3 (<https://www.nature.com/articles/sdata201618>).

ABSTRACT

In its four years of activity the ARIADNE project has created a catalogue of European digital archives which offers a portal that makes it possible to search the repository, where about two million datasets are recorded. The project has implemented an Open Data system applying the FAIR principles (Findable-Accessible-Interoperable-Reusable), and making available a concrete re-use of these important information sources, which otherwise would be difficult to access as most of the contents are unpublished.