

MOD (MAPPA OPEN DATA).
CONSERVARE, DISSEMINARE, COLLABORARE:
UN ARCHIVIO OPEN DATA PER L'ARCHEOLOGIA ITALIANA

1. L'APERTURA DEI DATI: LO SCENARIO DEGLI OPEN DATA

Negli ultimi mesi si è fatta sempre più accesa la discussione a livello internazionale – e fortunatamente anche nazionale – sugli open data e sull'importanza dell'accessibilità dei dati della pubblica amministrazione. Se in Europa, per promuovere la diffusione dei dati prodotti dalla ricerca finanziata con denaro pubblico sono stati avviati due distinti progetti, denominati rispettivamente OpenAIRE e OpenAIREplus (il primo è dedicato all'open access, il secondo agli open data), in Italia siamo ancora molto indietro, nonostante comincino ad arrivare i primi timidi segnali. Il "DL Crescita 2.0", approvato dal Consiglio dei Ministri il 18 ottobre 2012 n. 179, coordinato con la Legge di conversione 17 dicembre 2012, n. 221, prescrive infatti una rapida diffusione di archivi di open data in tutta la pubblica amministrazione, compreso il settore dei beni culturali. In campo archeologico, la necessità di aprire gli archivi si avverte con particolare urgenza in quanto il numero crescente di indagini (in parte dipendente dal diffondersi dell'archeologia preventiva) e l'affermarsi dell'archeologo come nuova figura professionale (specie nell'ultimo ventennio) stanno determinando una sempre maggiore esigenza di reperire informazioni in tempi brevi e da parte di un numero elevato di soggetti diversi: esigenza a cui non ha corrisposto finora una maggiore accessibilità delle informazioni.

Le dinamiche della rete, in particolare quelle legate all'idea dello "sharing", cioè della legittimità e dell'utilità della condivisione dei lavori digitali non commerciali per scopi culturali, rappresentano un potente incentivo verso l'apertura degli archivi. Per converso, chi vi si oppone lo fa in nome del principio di "proprietà intellettuale" che però – così come viene di norma [mal]intesa nel settore umanistico (e quindi anche nel mondo archeologico) – finisce con il costituire un freno allo sviluppo, non solo culturale, del Paese. La condivisione collaborativa, che nell'archeologia anglosassone ha trovato la sua formulazione più esauriente nel volume *Archaeology 2.0* (KANSA, WHITCHER KANSA, WATRALL 2011), sta finalmente facendo breccia – sia pure fra mille difficoltà e distinguo – anche in Italia, dove peraltro i dati archeologici sono prodotti da ricerche che, nella maggior parte dei casi, sono finanziate con denaro pubblico. È a questa filosofia che si riallaccia il Progetto MAPPA (Metodologie Applicate alla Predittività del Potenziale Archeologico), il cui scopo è realizzare un archivio archeologico di open data, premessa anche per quell'archeologia predittiva che rappresenta la risposta più innovativa alla necessità di far con-

vivere la tutela dei resti del nostro passato con le esigenze di vita del nostro presente e del nostro futuro.

Il termine open data ha un significato preciso. I dati aperti devono essere: completi, primari, tempestivi, accessibili, leggibili da computer, non proprietari, riutilizzabili, ricercabili, permanenti (ANICHINI, GATTIGLIA 2012, 51-52). In campo archeologico i dati primari o “grezzi”, ovvero i dati archeografici prodotti in un’indagine (documentazione grafica, fotografica, compilativa e quantificazione dei reperti mobili, debitamente collocati nello spazio geografico) devono essere messi a disposizione immediatamente, tutelando ovviamente la “paternità intellettuale” di chi quei dati ha prodotto e anche il suo diritto a pubblicarne l’interpretazione con i tempi necessari (che nel caso delle pubblicazioni archeologiche sono sempre inevitabilmente lunghi). Come in qualsiasi settore di ricerca, anche in archeologia la rapida circolazione delle informazioni e la possibilità di riutilizzarle liberamente per ulteriori nuove ipotesi e ricostruzioni storiche sono l’imprescindibile presupposto per un reale progresso degli studi. Senza di ciò, non può esserci progresso nella conoscenza e senza progresso nella conoscenza non può esserci neppure un’efficiente attività di tutela e valorizzazione dei beni archeologici.

M.L.G.

1.1 *Open data e open archives*

È quindi fondamentale che il dato archeografico sia affidabile, disponibile, aperto e riutilizzabile. Se, infatti, non si possono fare buoni ragionamenti critici senza una buona descrizione dei dati di partenza, non si può fare archeologia senza archeografia. Lo scavo archeologico – e in parte anche il survey – sono pratiche non ripetibili. L’unico elemento di riproducibilità e di ri-analisi è costituito dalla continua possibilità di utilizzare i dati grezzi. Pertanto la loro condivisione è l’unico modo che permette alla comunità scientifica la comprensione e l’eventuale rilettura del processo interpretativo. Ma non solo: la condivisione permette anche il riutilizzo dei dati su scale differenti, per fornire risposte a domande diverse in relazione ad altre o nuove indagini. I dati grezzi sono infatti il vero e unico “codice sorgente” dell’archeologia.

«La proprietà intellettuale è un male inutile», hanno scritto recentemente Boldrin e Levine (BOLDRIN, LEVINE 2012): si tratta di un’affermazione forte, ma è certo che, così come la si intende generalmente nelle scienze umanistiche – ovvero come il diritto, da parte di chi li ha prodotti, di appropriarsi dei dati, facendone un uso esclusivo per anni e talvolta per decenni (quando non disperdendoli, senza che nessuno abbia mai potuto prenderne visione) – la proprietà intellettuale rappresenta un freno all’innovazione e allo sviluppo della ricerca, impedendo, di fatto, l’apertura di nuovi filoni di studio e di nuove e più accurate sintesi. D’altra parte, questa prassi si basa

su una malintesa interpretazione del principio di “proprietà intellettuale” che non va confusa con la “paternità intellettuale”, che al contrario va tutelata e valorizzata più e meglio di quanto non si faccia oggi, attraverso quel sistema di corrette citazioni che la ricerca scientifica ben conosce. Se si parte dalla considerazione che la pratica archeografica, sia essa frutto del lavoro di professionisti o di ricercatori afferenti a strutture pubbliche, come le università e le soprintendenze, è sempre e comunque un’attività di ricerca (dal momento che produce dati unici e irripetibili) e che non vi è ricerca fino a quando non vi è pubblicazione del dato, appare evidente come la condivisione dei dati grezzi debba essere considerata come una pubblicazione scientifica a tutti gli effetti, salvaguardando sia le competenze che la capacità professionale e l’impegno, anche temporale, profuso da chi quel dato ha prodotto con il suo lavoro sul campo.

Il web rappresenta il supporto ideale per pubblicare i dati archeografici giacché consente, grazie alle attuali tecnologie, di distribuire conoscenza in modo capillare e a costi ridotti. La pubblicazione dei dati grezzi (archeografica) deve, quindi, essere garantita e riconosciuta nella sua “maternità” (ANICHINI, GATTIGLIA 2012, 53), così come la pubblicazione dei dati interpretati (archeologica) in forma integrale deve essere regolamentata da un diritto di prelazione, da parte di chi ha prodotto i dati, limitato e definito nel tempo (tre-cinque anni?), lasciando contestualmente liberi i dati per altre analisi attraverso l’uso di licenze come, ad esempio, la CC BY e l’uso di Digital Object Identifier (DOI). Per rendere i dati aperti ricercabili (e quindi fruibili), questi devono essere inseriti all’interno di archivi aperti (open archives). È ormai consuetudine della pratica archeologica creare archivi digitali, più o meno complessi, nei quali inserire i dati archeografici, siano essi di tipo testuale, grafico o digitale. Alla serie di standard, atti a normare l’attività di registrazione delle indagini sul campo (sui quali il Ministero per i Beni e le Attività Culturali, attraverso l’Istituto Centrale per il Catalogo e la Documentazione, lavora ormai da anni), occorrerà d’ora innanzi cominciare ad affiancare una serie di good practice circa le procedure informatiche da adottare per salvaguardare la progettazione, l’implementazione e la conservazione degli archivi digitali.

Gli standard assicurano il rispetto di vincoli connessi alle forme e ai contenuti imposti per la codifica e la normalizzazione di categorie di oggetti e attributi; le strategie digitali permetteranno la valorizzazione del processo di documentazione, inventariazione e classificazione, rendendolo semplice e di facile utilizzo e assicurando il futuro accesso agli archivi e la conservazione a lungo termine dei record archeologici. La creazione di standard, inoltre, facilita lo scambio d’informazioni, provocando lo sviluppo di un’intensa collaborazione anche con istituzioni sopranazionali, impegnate nello sviluppo di programmi finalizzati alla diffusione di standard e linee guida. Ne deriva

che gli standard assumono un ruolo centrale nell'agenda dell'archeologia computazionale e dell'archeologia in generale e, assieme all'accessibilità degli archivi, sono le parole chiave su cui deve fondarsi il futuro dell'archeologia.

G.G.

2. MAPPA OPEN DATA ARCHIVE (MOD): IN ITALIA È IMPORTANTE PARTIRE

Il MOD si propone come primo archivio archeologico open data italiano. Il progetto nasce utilizzando come primo set i dati, sia archeografici, sia archeologici, presenti nel comprensorio urbano di Pisa. L'obiettivo dichiarato non è realizzare un modello verificato ed esportabile per la creazione di numerosi altri archivi, bensì quello di proporre il MOD stesso come archivio italiano, luogo virtuale e struttura digitale entro la quale possano trovare spazio tutti quei dati fino ad oggi inaccessibili in rete. Crediamo prioritario, infatti, non disperdere energie e risorse nella creazione di mille archivi locali, ma mettere a disposizione di tutta la comunità scientifica il lavoro già fatto. A tale scopo, senza escludere un'evoluzione e un miglioramento che potrà nascere proprio grazie alla diffusione del suo utilizzo e agli input che verranno dagli utenti, la strutturazione del MOD è assai semplice. Il principio fondamentale che regola l'immissione dei dati è quello che tutti gli interventi archeologici, di qualsiasi tipologia o dimensione, abbiano la stessa visibilità. Ogni intervento viene identificato ("titolo") topograficamente con il nome della località dove è stato realizzato, la tipologia di intervento e l'anno di esecuzione. Una scansione per macro periodi storici e sintetiche tematiche archeologiche permette una prima sommaria classificazione che aiuta ad orientare l'utente nelle scelte di consultazione. Per affinare la ricerca è attiva una "ricerca avanzata" che aggiunge ai campi tematici e cronologici, l'anno di intervento, le parole note all'interno del titolo, il nome dell'autore e il formato del file ricercabile.

All'interno del singolo intervento è presente una sintesi descrittiva dell'indagine ("introduzione"), una pagina contenente tutta la documentazione disponibile ("dataset") suddivisa per tipologia (grafica, fotografica, compilativa), una pagina dedicata alla letteratura grigia ("relazione"). Rimangono costantemente visibili in tutte le pagine i codici DOI attribuiti alle varie sezioni ("identificatori"), il "contatto" principale al quale chiedere eventuali informazioni (solitamente corrispondente al nome dell'autore), il titolo, l'anno di esecuzione e il nominativo dell'autore (uno o più). È infatti proprio attraverso l'individuazione dell'autore e l'attribuzione di un codice DOI che si sancisce il riconoscimento della documentazione come effettiva pubblicazione scientifica. La differenziazione tra dataset e relazione viene ulteriormente rimarcata con l'apposizione di due codici DOI diversi; si stabilisce infatti che l'autore della relazione sia esclusivamente il firmatario della stessa, a differenza dell'autore o degli autori dei dataset di documentazione



Il progetto
Chi siamo
Pubblicazioni
Biblioteca
News
Multimedia
MAPPAGIS
MAPPAopenDATA
Servizi
Newsletter

REGIONE TOSCANA
FAS Fondo Area Sottoutilizzate 2007-2013
UNIVERSITÀ DI PISA
Dipartimento di Scienze Archeologiche
Dipartimento di Scienze della Terra
Dipartimento di Matematica

mappa
metodologie applicate alla predittività del potenziale archeologico

Il progetto Mappa
Chi
Il progetto si basa sull'integrazione di competenze diverse: archeologi, storici, geologi, geomorfologi, matematici, che lavoreranno insieme nel Laboratorio MAPPA, mettendo costantemente a confronto le proprie esperienze. Il coinvolgimento di studenti universitari nelle varie attività di ricerca fa di MAPPA anche un luogo di alta formazione professionale per i ricercatori del futuro.

Perche' Come **Chi** Quando Dove

Mappa Gis
Pubblicare, interrogare, analizzare, conoscere. Il WebGIS di MAPPA è Carta Archeologica e Carta del Potenziale Archeologico di Pisa: uno strumento interattivo e aperto, per la tutela, la governance e la ricerca.

Mappa Open Data
Conservare, disseminare, collaborare. L'archeologia libera e aperta è la nuova frontiera della ricerca.

Fig. 1 – Home page del portale del Progetto MAPPA (www.mappaproject.org).

che possono non coincidere, soprattutto quando, come nella maggior parte dei casi, la documentazione è il risultato di un lavoro di équipe. Questa specifica richiama la responsabilità scientifica dell'archeologo che redige la relazione di fine scavo (sempre più sottolineata nelle Valutazioni di Interesse Archeologico preventivo ed esplicitata, in generale, nel DPR 207/2010) e ne sancisce il processo creativo proprio dell'interpretazione archeologica; allo stesso tempo la pubblicazione dell'apparato archeografico di riferimento (dataset), oltre ad essere supporto imprescindibile all'interpretazione, ne aumenta immediatamente la verificabilità.

La consultazione e il download dei dati archeografici avviene dalla "pagina dataset" di ogni intervento, accedendo direttamente ai diversi file, singoli o raccolti in cartelle.zip. Siamo consapevoli che non tutti i formati di dati presenti (dwg, pdf, jpeg, csv, xml, xls) siano open o machine readable e che non otterranno le "cinque stelle" di T. Berners-Lee, ma per l'archeologia italiana, oggi crediamo che la cosa maggiormente urgente e necessaria sia di partire. Partire con ciò che abbiamo, nei formati disponibili, iniziare cioè a liberare dati, a far circolare informazioni, cercando di fare il primo e fondamentale passaggio: cambiare la mentalità degli archeologi, non ancora abituati

a condividere apertamente i dati da loro prodotti, perché possano scoprire l'utilità e le immense possibilità che lo share può offrire, senza averne paura. Il tempo per discutere, migliorare il sistema e trasformare i dati in formati più aperti può avvenire contemporaneamente; intanto iniziamo.

Non esistendo ancora una normativa specifica e confidando che si innesti un processo virtuoso promosso anche dallo stesso MiBAC, la pubblicazione dei dati all'interno del MOD avviene su base volontaria dei singoli autori, presupponendo quindi che ognuno possa scegliere quali dati pubblicare. I dati fino ad oggi caricati sono rilasciati con licenza CC-BY per consentirne il pieno riutilizzo anche come strumento professionale (si pensi ad esempio alla redazione di Valutazioni di Impatto Archeologico); è lasciata comunque all'autore la scelta tra la licenza CC-BY o CC-BY-SA.

Infine è stata elaborata una versione beta dello schema di metadati da applicare ad ogni dataset archeologico, che tenga conto delle peculiarità del dato in oggetto, esplicitando ad esempio il background storico e culturale nel quale il dato archeologico è stato raccolto, e dei campi del "core metadata" contenuto nello standard ISO 19115. Tale proposta è pubblicata all'interno del MOD per consentire a tutti di commentarla e suggerire possibili modifiche, così da rendere lo schema frutto di un'idea più ampiamente condivisa. Le modalità di immissione dei dati nell'archivio, esplicitate nella pagina "Come pubblicare nel MOD", sono frutto di un lavoro di confronto tra gli archeologi e i legali del gruppo MAPPA e i funzionari del MiBAC (Soprintendenza per i Beni Archeologici della Toscana e Direzione Regionale per i Beni Culturali e Paesaggistici della Toscana) e sono oggetto di contributi specifici (ANICHINI *et al.* 2013, 24-26; ANICHINI, CIURCINA, NOTI 2013; CIURCINA 2013).

F.A.

2.1 La struttura tecnologica

L'applicazione, ospitata su server GNU/Linux del Centro Interdipartimentale di Servizi Informatici per l'area Umanistica (CISIAU) dell'Università di Pisa, è stata progettata su piattaforma tecnologica open source LAMP con utilizzo di Apache Http Server, del linguaggio di scripting PHP 5.x e del database relazionale open source MySQL. Tale piattaforma, ampiamente collaudata a livello mondiale per la realizzazione di archivi informatici, garantisce stabilità, sicurezza e performance adeguate agli scopi del progetto. La struttura di database è concentrata sulla singola indagine archeologica definita da una serie di informazioni (titolo, autore, DOI, località, anno, contatto primario, introduzione, overview) e dal collegamento con fonti dati eterogenee come documenti testuali, immagini, oggetti multimediali, file di database, dati geografici, etc. È infatti associato all'archivio un repository di file, scaricabili dal lato utente, che gli operatori abilitati possono inserire attraverso una funzione di upload.

È inoltre presente la gestione standardizzata dei metadati riferiti al singolo dataset con la definizione di un set minimo di informazioni in modo da garantire una corretta usabilità del dato. Ad ogni intervento di scavo sono associati tutti i dati inerenti lo scavo stesso, la produzione archeografica, la struttura e il formato dei dati digitali, secondo uno schema che descriva la storia dell'indagine, le fonti utilizzate, il metodo e le relazioni con i dati fisici. Dal punto di vista operativo l'applicazione, consultabile attraverso qualunque browser compatibile con gli standard W3C, è composta da due segmenti distinti: una sezione amministrativa con accesso riservato ed una sezione pubblica. Il segmento amministrativo consente l'inserimento e la modifica dei contenuti del database da parte degli operatori con la possibilità di gestire categorie (per es. la cronologia) e sottocategorie (per es. "periodo romano", "periodo medievale", etc.) personalizzabili, di utilizzare un editor HTML evoluto per l'editing delle sezioni "introduzione" ed "overview", e di eseguire l'upload di file nel repository.

Il segmento pubblico, direttamente accessibile dal sito del Progetto "MAPPA", permette la consultazione dell'archivio, la visualizzazione della scheda del singolo dataset, la realizzazione di ricerche libere full-text o categorizzate sui contenuti del database, e il download dei file dal repository. L'interfaccia utente, tableless e realizzata con CSS, è stata pensata per garantire accessibilità ed usabilità con requisiti minimi necessari per il suo utilizzo.

Particolare importanza è stata attribuita alla scalabilità dell'applicazione in previsione della crescita dimensionale e alle problematiche riguardanti il ciclo di vita digitale dell'archivio. L'applicazione è stata progettata in modo da garantire un'agevole migrazione su altre piattaforme hardware o software cercando di ridurre quindi le future attività di adeguamento tecnologico. Il progetto prevede inoltre l'implementazione di appropriate strategie di backup, importanti anche per la presenza del repository associato al dataset. È prevista inoltre l'eventuale integrazione con un sistema documentale che garantisca una corretta gestione dell'archivio di file da parte degli amministratori per i controlli di versione, attività di backup e aggiornamento. Infine, particolare attenzione è stata posta verso la redazione di linee guida dettagliate per gli operatori con gli scopi di allineare i singoli dataset dal punto di vista tecnico e concettuale e di agevolare eventuali attività di riuso dei dati o di interfaccia con altri archivi.

Crediamo fortemente che conservare, disseminare, collaborare siano la nuova frontiera dell'archeologia.

V.N.

FRANCESCA ANICHINI, GABRIELE GATTIGLIA,
MARIA LETIZIA GUALANDI, VALERIO NOTI
Dipartimento di Civiltà e forme del sapere – LabMAPPA
Università di Pisa

BIBLIOGRAFIA

- ANICHINI F., GATTIGLIA G. 2012, #MappaOpenData. *From web to society. Archaeological open data testing*, in M.L. GUALANDI (ed.), *MapPapers 3-II/2012*, Roma, Edizioni Nuova Cultura, 51-56 (http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/Pre_atti_online3.pdf).
- ANICHINI F., CIURCINA M., NOTI V. 2013, *Il MOD: l'archivio open data dell'archeologia italiana*, in F. ANICHINI, N. DUBBINI, F. FABIANI, G. GATTIGLIA, M.L. GUALANDI, *MAPPA. Metodologie Applicate alla Predittività del Potenziale Archeologico*, vol. 2, Roma, Edizioni Nuova Cultura.
- ANICHINI *et al.* 2013 = ANICHINI F., CIURCINA M., FABIANI F., GATTIGLIA G., GHIZZANI MARCÌA F., GUALANDI M.L., NOTI V., SCIUTO C. 2013, *MOD (Mappa Archaeological Open Data archive): new ideas for new minds*, in F. ANICHINI, M. BINI, N. DUBBINI, F. FABIANI, G. GATTIGLIA, F. GHIZZANI MARCÌA, M.L. GUALANDI, *MapPapers*, 4-2013, Roma, Edizioni Nuova Cultura, 24-26 (http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/OP2013_pre_atti_rid.pdf).
- BOLDRIN M., LEVINE D.K. 2012, *Abolire la proprietà intellettuale*, Roma-Bari, Laterza.
- CIURCINA M. 2013, *Parere legale sul portale Mappa Open Data*, in F. ANICHINI, M. BINI, N. DUBBINI, F. FABIANI, G. GATTIGLIA, F. GHIZZANI MARCÌA, M.L. GUALANDI, *MapPapers*, 4-2013, Roma, Edizioni Nuova Cultura, 87-106 (http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/MapPapers_15_parere.pdf).
- KANSA E.C., WHITCHER KANSA S., WATRALL E. (eds.) 2011, *Archaeology 2.0: New Approaches to Communication and Collaboration*, Cotsen Digital Archaeology Series, 1, Cotsen Institute of Archaeology, UC Los Angeles (<http://www.escholarship.org/uc/item/1r6137tb/>).

ABSTRACT

An archaeological excavation is an unrepeatable practice. The only action that can be reproduced and re-analysed is the continuous use of raw data. Data sharing, therefore, is the only way to understand and re-examine the archaeological interpretative process and to answer questions regarding new surveys. Data do not circulate freely in Italian archaeology today. In order to make open data searchable (and, therefore, usable), they must be entered in open archives. Only the free access to raw data, based upon an Open Data approach, will allow a further step to be taken towards 2.0 archaeology. The idea of creating an Italian open digital archaeological archive, using data from the urban area of Pisa as case-study, originated from the MAPPA project (Methodologies Applied to Archaeological Potential Predictivity). The main objectives of the MAPPA Open Data (MOD) archive will be to allow simple and free access to all archaeological data (of any type or size) and to guarantee a digital lifecycle for as long as possible.