

OPEN DIGITAL ARCHIVES IN ARCHEOLOGIA. GOOD PRACTICE

1. INTRODUZIONE

Questo articolo nasce dalle riflessioni scaturite durante il mio lavoro di dottorato, indirizzato allo studio archeologico di Pisa medievale da realizzare attraverso la creazione di un GIS urbano, ed in particolare dalle difficoltà incontrate nella raccolta di tutti i dati archeologici presenti e soprattutto nella ricerca di dati non interpretati sulla base dei quali poter fare nuove analisi.

Alla base del fenomeno open source risiede il principio che vede nella collaborazione di più sviluppatori la possibilità di ottenere un prodotto finale di maggiore complessità rispetto a quanto potrebbe ottenere un singolo gruppo di programmazione, secondo la formula:

proprietà collettiva + collaborazione = un aumento delle conoscenze

Se in campo strettamente informatico questo vuol dire libera fruibilità dei file sorgente dei programmi, in campo archeologico dobbiamo puntare all'accesso libero (open data) del dato primario, cioè alla libera condivisione dei dati grezzi (raw data), quelli che definiamo dati archeografici (in pratica la documentazione di scavo e del survey), che dovrebbero essere pubblicati su archivi on-line attraverso documenti ben riconoscibili. La costruzione di sistemi aperti che consentano lo scambio di dati e di informazioni e la condivisione di idee può innescare un circolo virtuoso, favorendo così il "lavorare insieme" a svantaggio del "lavorare separatamente", con un notevole beneficio globale per l'archeologia (PESCARIN 2006, 141).

2. OPEN ACCESS AND OPEN DATA

Negli ultimi anni anche l'archeologia si sta incamminando verso la sua versione 2.0 ed iniziano a manifestarsi i primi fenomeni di *open access* e di open data. Se da un lato l'open access si accontenta di proporre l'accesso gratuito all'informazione scientifica almeno nella sua versione elettronica (LASER 2005, 77-78), la filosofia open data aspira ad una libera fruibilità dei dati grezzi, soprattutto non testuali, e al loro riutilizzo, seppur controllato da apposite licenze. Esistono fundamentalmente due sistemi differenti di consentire l'accesso gratuito ad una pubblicazione: il primo, diretto, è effettuato dalle riviste stesse, che rendono immediatamente accessibili le pubblicazioni; il secondo, indiretto, è realizzato dagli autori che realizzano copie delle loro pubblicazioni e le rendono accessibili o su una homepage personale o su un

deposito istituzionale. Consultando il sito Directory Open Access Journal¹ si trovano elencate 21 riviste peer-review a carattere archeologico nel mondo², 4 tedesche e statunitensi, 3 croate, una italiana, il Web Journal on Cultural Patrimony, una sola francese, il Bulletin de l'Institut Français d'Études Andines pubblicato in Perù con licenza *Creative Commons*³, contro le 121 riviste di storia. All'elenco mancano ad esempio i Fasti FOLD&R⁴ che pubblicano relazioni di scavo, al fine di metterle a disposizione di tutti, Gradus, rivista dedicata all'archeologia dell'acqua e alle scienze applicate all'archeologia e al restauro⁵ e la stessa Archeologia e Calcolatori⁶. Tra i depositi istituzionali, invece, citiamo il progetto BibAr⁷ Biblioteca archeologica on-line di archeologia medievale ospitata all'interno del Portale di Archeologia Medievale del Dipartimento di Archeologia e Storia delle Arti dell'Università di Siena, che oltre ad articoli mette a disposizione intere pubblicazioni di scavo, e i volumi di *Medieval Archaeology* sul portale dell'*Archeology Data Service*⁸. Non è difficile prevedere, come recentemente ha fatto Nisha Doshi su PloS⁹ un incremento, in tempi brevi, delle pubblicazioni archeologiche su riviste open access. Due fattori, infatti, possono spingere ad un maggior utilizzo del web come mezzo di pubblicazione anche di intere edizioni di scavo¹⁰: l'ampiezza della "distribuzione" e l'abbattimento dei costi di stampa¹¹.

L'open access, quindi, offre la possibilità di mettere a disposizione i risultati finali delle ricerche proposti in forma di sintesi. Si tratta certamente di una conquista importante, ma non sufficiente; solo la libera fruibilità dei dati grezzi secondo il paradigma dell'open data consentirebbe di far fare

¹ <http://www.doaj.org/>.

² <http://www.doaj.org/doaj?func=subject&cpid=11> (accesso 4 aprile 2009).

³ <http://creativecommons.org/about/licenses/meet-the-licenses> (accesso 4 aprile 2009).

⁴ <http://www.fastionline.org/folder.php?view=home> I Fasti FOLD&R, basati sul sistema del *peer-review*, sono pubblicati a partire dal 2004 dall'Associazione Internazionale di Archeologia Classica in collaborazione con il Ministero per i Beni e le Attività Culturali.

⁵ http://www.cantierenavipisa.it/Pubblicazioni_Gradus.html, pubblicata a partire dal 2007, dal Cantiere delle Navi Antiche di Pisa.

⁶ <http://soi.cnr.it/~archcalc/index.htm>, che aderisce all'Open Archives Initiative (<http://www.openarchives.org/>).

⁷ <http://www.bibar.unisi.it/>.

⁸ http://ads.ahds.ac.uk/catalogue/library/med_arch/index.cfm?CFID=615449&CFTOKEN=57428789 (accesso 4 aprile 2009).

⁹ <http://www.plos.org/cms/node/455> (accesso 4 aprile 2009).

¹⁰ Un buon esempio è la pubblicazione degli scavi di Cosa curata da Elizabeth Fentress e Adam Rabinowitz – <http://www.press.umich.edu/webhome/cosa/home.html> (accesso 4 aprile 2009) – che offre anche parte dei dati grezzi scaricabili (lista delle unità stratigrafiche, con datazione, interpretazione; lista dei materiali; catalogo delle sepolture), mentre sono escluse sia le piante, sia le immagini.

¹¹ Chiaramente il passaggio a pubblicazioni completamente on-line, non avverrà in maniera gratuita, ma si può condividere la soluzione proposta (GRUPPO LASER 2005, 94) di far pagare chi scrive, partendo dal concetto che la pubblicazione serve anche per far conoscere il proprio lavoro e ottenere impatto e credito scientifico nella comunità, e che in genere i costi di stampa sono compresi in un buon budget di ricerca.

un ulteriore passo in avanti verso un'archeologia veramente 2.0. Il futuro della ricerca archeologica è, infatti, interconnesso alla conservazione dei dati ed al loro libero riutilizzo per ulteriori indagini ed analisi. Analizziamone i motivi:

– Eseguire uno scavo o un survey archeologico vuol dire essenzialmente documentare in maniera analitica ogni evidenza individuata, tradurre in documenti scritti, grafici e fotografici quanto emerge nelle operazioni di scavo o di ricognizione. Questo processo archeografico (MOBERG 1981; GUIDI 1994) permette sia di realizzare il processo archeologico, interpretativo, da parte di coloro che hanno direttamente eseguito l'indagine, sia di riutilizzare la documentazione raccolta nel corso delle indagini sul campo per formulare ulteriori ipotesi e ricostruzioni storiche da parte di altri studiosi (D'ANDREA 2006, 130). Se, infatti, non si possono fare buoni ragionamenti critici, senza una buona descrizione oggettiva dei dati di partenza, non si può fare archeologia senza archeografia (MANNONI 2000, 217).

– Lo scavo archeologico, ed in parte anche il survey, sono pratiche non ripetibili. L'unico elemento di riproducibilità e di rianalisi è costituito dalla continua possibilità di utilizzare i dati grezzi, archeografici. La loro condivisione non solo permette da parte della comunità scientifica la comprensione e l'eventuale rilettura del processo interpretativo, ma anche il riutilizzo dei dati su scala differente e in relazione a nuove indagini. I dati grezzi sono il vero ed unico codice sorgente dell'archeologia.

– La stragrande maggioranza¹² degli interventi archeologici in Italia non è legata in senso stretto alla ricerca, bensì alle varie coniugazioni delle pratiche di emergenza (interventi di emergenza veri e propri, preventivi e/o programmati) soprattutto per quanto riguarda i contesti urbani, che come ben sappiamo sono la palestra dell'archeologia professionale.

– La stragrande maggioranza degli interventi archeologici rimangono inediti (BROGIOLO 2000, 354)¹³ e questo risulta ancor più valido per tutti gli interventi che possiamo definire d'emergenza.

– È ormai pratica consueta da parte degli archeologi creare degli archivi digitali, spesso non sicuri, ai fini della preservazione dei dati, così come acquisire parte delle informazioni direttamente in formato digitale.

¹² Sebbene non esistano dati complessivi su tutto il territorio nazionale, cosa di cui peraltro ci sarebbe bisogno, i dati su Pisa, dove sono stati censiti (ANICHINI, PARIBENI 2005) tutti i singoli interventi a carattere archeologico conosciuti (554 differenti interventi archeologici di varia natura a partire dal 1520), mostrano come negli ultimi 20 anni l'83% degli interventi sia di scavo, sia di ricognizione siano stati eseguiti da archeologi professionali e comunque non collegati ai Dipartimenti universitari.

¹³ Anche in questo caso non esistono statistiche ufficiali, ma i dati raccolti per Pisa mostrano come solo il 20% degli interventi degli ultimi 20 anni sia edito e come solo poco più dell'1% sia stato pubblicato come edizione completa di scavo.

- La creazione di archivi digitali on-line consente di preservare i dati.
- Avere a disposizione archivi on-line contenenti dati grezzi facilita la ricerca, le politiche di pianificazione territoriale¹⁴, e l'allargamento della comunità scientifica superando le barriere locali e regionali.

3. LA PRATICA ARCHEOGRAFICA IN ITALIA

Manca una ricerca completa sulla pratica della documentazione archeologica¹⁵ in Italia: i dibattiti metodologici-procedurali sono stati considerati fundamentalmente chiusi con i primi anni '90, mentre i progressi tecnico-informatici e la crescita dell'archeologia professionale li rendono quantomai attuali. La presenza in Italia di una comunità archeologica che si va facendo sempre più variegata, diversificata tra organi ministeriali, organi di ricerca e archeologi professionali rende sempre più urgente un'analisi di questo tipo. Basti ricordare come la ricerca¹⁶ che diede il via al progetto dell'*Archaeological Data Service* (ADS)¹⁷, curata dall'Università di York, prevedesse, seppure in un quadro differente sia da un punto di vista normativo, sia professionale, il coinvolgimento delle varie e sfaccettate realtà archeologiche presenti. È, infatti, questa pluralità di soggetti che produce il dato archeografico e che deve attivamente partecipare alla creazione delle linee guida operative (good practice) e degli standard per la raccolta del dato archeologico, che, in pratica, rappresentano un consenso professionale su pratiche comuni e servono per controllare la coerenza di un processo lavorativo e/o organizzativo all'interno di una comunità professionale, almeno su scala nazionale. È dagli standard che dipende la possibilità di condivisione dei dati¹⁸.

La ricognizione sulla documentazione di scavo consegnata presso la Soprintendenza per i Beni Archeologici della Toscana (SBAT), Ufficio distaccato di Pisa, a partire dall'anno 2000, senza pretesa di essere paradigmatica, ma solo esemplificativa di situazioni comuni anche in altre aree della penisola, mostra un quadro composito che va dall'assenza di documentazione, a relazioni di scavo più o meno articolate, alla consegna completa di tutta la documentazio-

¹⁴ In particolare si fa riferimento alla Legge 25 Giugno 2005, n. 109 articoli 2 ter Verifica preventiva dell'interesse archeologico e 2 quater Procedura di verifica preventiva dell'interesse archeologico.

¹⁵ In questa sede ci si occuperà solamente della documentazione di scavo, per la quale nel corso degli anni '80 vennero fatti dei passi verso la standardizzazione (<http://www.iccd.beniculturali.it/Catalogazione/standard-catalografici/normative/nomative>, accesso 4 aprile 2009), tralasciando il discorso del survey, che non ha mai raggiunto livelli di standardizzazione equivalenti e condivisi dopo i timidi passi intrapresi tra la fine degli anni '80 e i primi anni '90 (per una disamina sulla confrontabilità e la standardizzazione dei dati da survey si veda TERRENATO 2006; GATTIGLIA, STAGNO 2005 e più in generale BERNARDI 1992 e CAMBI, TERRENATO 1994).

¹⁶ <http://ads.ahds.ac.uk/project/strategies/> (accesso 4 aprile 2009) *Strategies for Digital Data. Findings and recommendations from Digital Data in Archaeology: A Survey of UserNeeds*.

¹⁷ <http://ads.ahds.ac.uk/>; per una completa analisi del programma che ha dato il via progetto cfr. D'ANDREA 2006, 85 ss.

¹⁸ Ed anche, fattore non trascurabile, il costo di un intervento archeologico.

ne prodotta sia stampata, sia su supporto digitale comprendente schede US, piante di strato, fotografie di US, quantificazione dei reperti mobili, fotografie di tutti i reperti, relazione di scavo e matrix, e mostra un generale incremento nella qualità del prodotto archeografico, rispetto ai decenni precedenti.

Aspettando, quindi, un'analisi complessiva sullo stato dell'archeografia oggi in Italia, credo che si possano già evidenziare alcuni principi fondamentali:

- l'attività archeografica deve avere un atteggiamento il più possibile neutrale per consentire ad altri archeologi di riutilizzare in futuro la documentazione raccolta nel corso delle indagini per ulteriori approfondimenti e magari per la formulazione di nuove ipotesi (D'ANDREA 2006, 78);
- dati grezzi, analisi preliminari dei rinvenimenti, informazioni contestuali sui siti, dati da survey, cataloghi ed indici costituiscono fonti vitali di informazione per la ricerca e dovrebbero essere accessibili sia su supporto digitale, sia cartaceo¹⁹;
- la pratica archeografica, sia essa legata ad operatori professionali, sia alle strutture dedicate alla ricerca, è sempre e comunque un'attività di ricerca, dal momento che produce dei dati unici e irripetibili.

Da ciò discende che ogni intervento di scavo deve produrre e rendere fruibile almeno la seguente documentazione archeografica:

- elenchi di US;
- schede di US;
- piante di US;
- matrix;
- relazione preliminare;
- immagini;

alle quali non sarebbe male aggiungere:

- schede di quantificazione preliminari;
- immagini di reperti;
- eventuali diari di scavo (ZANINI, COSTA 2006).

4. OPEN ARCHIVES: PRESERVARE PER CONDIVIDERE

È ormai consuetudine della pratica archeologica creare degli archivi digitali, più o meno complessi, nei quali inserire i dati archeografici siano essi di tipo testuale (report, schede, diari di scavo), grafico (immagini, piante, sezioni) o digitale (database, CAD, GIS, etc.). Agli standard atti a normare l'attività di registrazione delle indagini sul campo bisogna, quindi, affiancare

¹⁹ <http://ads.ahds.ac.uk/project/strategies/> (accesso 4 aprile 2009) "Strategies for Digital Data. Findings and recommendations from Digital Data in Archaeology: A Survey of UserNeeds".

una serie di standard²⁰, di *good practice*, relativi alle procedure informatiche da adottare per salvaguardare la progettazione, implementazione e conservazione degli archivi digitali. Gli standard assumono quindi un ruolo centrale nell'agenda dell'archeologia computazionale e dell'archeologia in generale, ed assieme all'accessibilità degli archivi sono le parole chiave su cui deve fondarsi il futuro dell'archeologia (D'ANDREA 2006, 78). In primo piano nell'affrontare queste problematiche si pongono le esperienze dell'English Heritage Centre for Archaeology (CfA) e dell'Archaeology Data Service (ADS) descritte rispettivamente nel *Digital Archiving Strategy 2.0*²¹ e nel *Digital Archives from Excavation and Fieldwork. Guide to Good Practice, Second Edition*²². In estrema sintesi i due manuali si propongono di fornire uno strumento utile per chiunque, a livello pubblico e privato, avesse la necessità di creare archivi digitali, e affrontano i temi legati all'obsolescenza delle risorse prescelte, alle modalità di manutenzione e salvaguardia a lungo termine degli archivi e, infine, alla loro accessibilità. Appare evidente come le finalità principali di una strategia di archiviazione digitale risiedano nella preservazione dei record originali per qualsiasi uso futuro, partendo da una definizione chiara della natura del dato registrato, delle sue caratteristiche essenziali e della sua autenticità, e nel loro trasferimento all'interno di facilities dove possano essere resi disponibili tramite Internet²³. Il modello proposto prevede che la creazione di archivi digitali aperti sia realizzata in accordo ad appropriati standard nazionali ed internazionali, dettagliatamente documentati in modo da determinare un accreditamento utile alla certificazione di qualità degli archivi, e che la divulgazione dei dati sia supportata da istituzioni nazionali, determinando una piena integrazione e collaborazione tra agenzie nazionali e comunità scientifica internazionale. Come si può facilmente capire il mantenimento degli archivi digitali presuppone un vasto impiego di risorse umane necessarie a garantire un servizio di assistenza²⁴, un'ampia

²⁰ Gli standard assicurano il rispetto di vincoli relativi alle forme e ai contenuti imposti per la codifica e la normalizzazione di categorie di oggetti e attributi; le strategie digitali consentono la valorizzazione del processo di documentazione, inventariazione e classificazione rendendolo semplice e di facile utilizzo per assicurare il futuro accesso agli archivi e la conservazione a lungo termine dei record archeologici (http://www.english-heritage.org.uk/upload/pdf/dap_manual_archiving.pdf?1239806706, accesso 4 aprile 2009). La creazione di standard, inoltre, facilita lo scambio di informazioni, provocando lo sviluppo di una intensa collaborazione anche con istituzioni soprannazionali impegnate nello sviluppo di programmi finalizzati alla diffusione di standard e linee guida.

²¹ http://www.english-heritage.org.uk/upload/pdf/dap_manual_archiving.pdf?1239806706 (accesso 4 aprile 2009).

²² <http://ads.ahds.ac.uk/project/goodguides/excavation/> (accesso 4 aprile 2009) creato dall'*Archaeology Data Service* (ADS).

²³ <http://ads.ahds.ac.uk/project/goodguides/excavation/sect81.html> (accesso 4 aprile 2009).

²⁴ Il portale dell'ADS fornisce PATOIS: una serie di *tutorials* per apprendere come utilizzare i dati digitali presenti on-line <http://ads.ahds.ac.uk/project/patois/module2/index.html> (accesso 4 aprile 2009).

varietà di formati²⁵ per ogni *dataset*, atta a consentirne l'utilizzo da parte di una vasta platea di utenti, e la manutenzione dei dati²⁶. Per quanto concerne la preservazione del contenuto vanno considerate sia le caratteristiche fisiche dell'archiviazione digitale, sia quelle strutturali. I documenti dovrebbero essere salvati in caratteri ASCII, mentre le strutture dovrebbero contenere ogni informazione sul processo di codifica adottato. Non dovrebbero, inoltre, mancare dati sulla provenienza, sul contesto e sugli inventari. Una distinzione fondamentale viene fatta tra dati creati direttamente in formato digitale e quelli acquisiti attraverso scansioni²⁷. Tutti i dati digitali acquisiti da fonti esterne devono contenere una documentazione esplicativa sul formato adoperato e sulla compatibilità dei dati con standard, data-model e thesauri esistenti. A tutti i dati sono associati dei metadati, il cui obiettivo è descrivere ogni informazione relativa al tracciato del record memorizzato. Le strategie di preservazione dei dati si occupano dei requisiti necessari per il formato di memorizzazione e per i supporti (longevità, capacità, durata, obsolescenza, costo, predisposizione al danneggiamento), mentre le strategie di accessibilità individuano i criteri necessari per una gestione di lungo termine (valutazione, inserimento, migrazione²⁸, riposo).

5. UTOPIA?

Se la necessità di creare archivi digitali aperti secondo un'ottica veramente open data rappresenta il futuro dell'archeologia, quando riusciremo ad avere in Italia un archivio come quello dello scavo di Eynsham Abbey²⁹, che contiene tutti i documenti archeografici comprese le singole piante di

²⁵ I progetti di analisi degli archivi digitali nacquero anche sulla scia dei problemi legati alla proliferazione di soluzioni software "proprietarie", per loro stessa natura incapaci di garantire l'interoperabilità dei dati. L'indicazione che spesso viene fornita è, infatti, proprio quella di adottare il più possibili soluzioni open source nella creazione degli archivi digitali, infatti «quanto più si farà uso di formati di scambio aperti, tanto più i gruppi di lavoro saranno liberi di utilizzare i programmi, commerciali o open, che più si adattano alle proprie esigenze, e progetti e dati potranno "migrare" più semplicemente da un sistema ad un altro, elemento questo da non sottovalutare vista la continua evoluzione dell'informatica» (PESCARIN 2006, 144).

²⁶ Il modello prevede, ancora una volta, la partecipazione di tutta la comunità archeologica, che alla fine ne è la beneficiaria, per sviluppare strategie finalizzate al finanziamento di infrastrutture per l'informatica.

²⁷ «Documentary archives held safely in paper form need not be digitised unless there is a need to provide digital access to the information contained therein or unless there is specific need for secure copying in digital form» <http://ads.ahds.ac.uk/project/goodguides/excavation/sect11.html> (accesso 4 aprile 2009). Nell'acquisizione degli originali cartacei devono essere rispettate delle linee guida relative alle procedure da seguire per la scansione e la digitalizzazione di contenuti testuali e/o grafici.

²⁸ Il *Technology Watch* stabilisce quando i formati originariamente adottati sono divenuti obsoleti oppure se sono disponibili formati più appropriati. http://www.english-heritage.org.uk/upload/pdf/dap_manual_archiving.pdf?1239806706 (4 aprile 2009).

²⁹ http://ads.ahds.ac.uk/catalogue/projArch/eynsham_OAU/downloads.cfm (4 aprile 2009).

strato? Forse facendo nascere un movimento dal basso, che coinvolgendo l'intera comunità archeologica, come accaduto per i movimenti *Open*, costringa a modificare abitudini e ad imporre gli standard necessari, piuttosto che aspettando un'operazione istituzionale tra gli enti proposti alla salvaguardia e tutela del patrimonio archeologico che porti alla definizione di comuni standard e linee-guida che incoraggino i ricercatori verso il loro uso (D'ANDREA 2006, 78). Questo anche in virtù di un clima di generale immaturità nei confronti di queste istanze. Se la ricerca svolta dall'ADS³⁰ un decennio fa aveva messo in evidenza come gli enti preposti a mantenere gli archivi non avessero la tecnologia necessaria per immagazzinare, preservare e mantenere archivi digitali, temo che la situazione italiana sia ancora più complessa³¹. Se il traguardo finale è quello di costituire degli archivi digitali come quelli prodotti e gestiti dall'ADS, la proposta che si vuol fare è quella di iniziare a far circolare una grande quantità di dati grezzi³², attraverso vari archivi che siano rintracciabili tramite un portale esterno ad essi dedicato³³. Iniziare a condividere i dati significa anche iniziare a ragionare sulle procedure di standardizzazione, così come risulta importante dividerli in tempi brevi rispetto alla loro produzione; i dati invecchiano ed è meglio metterli a disposizione subito, senza cercare la perfezione, quando la comunità scientifica è maggiormente in sintonia metodologica con chi li ha prodotti. Il problema maggiore consiste nel far accettare un mezzo così rivoluzionario alla comunità accademica archeologica, spesso tesa a proteggere i propri dati per future, eventuali pubblicazioni. Sebbene i dati siano pubblici sappiamo come spesso restino chiusi nei cassetti delle Direzioni Scientifiche. Se si parte

³⁰ <http://ads.ahds.ac.uk/project/goodguides/excavation/> (4 aprile 2009).

³¹ A titolo esemplificativo cito la difficoltà che ha tuttora la Soprintendenza per i Beni Archeologici della Toscana a gestire i propri archivi cartacei e che ha un ricercatore nell'individuare i dati necessari, sempre che questi siano presenti (sappiamo quanta resistenza spesso esista da parte degli archeologi nel consegnare la propria documentazione di scavo per paura che venga utilizzata da altri). La gestione dell'archivio risponde ad una pratica che porta alla divisione della documentazione di scavo in tre differenti sottoarchivi: documentazione scritta; archivio fotografico; archivio grafico. Con forte rischio di depauperamento delle fonti conservate; basti pensare che la documentazione grafica conservata nell'archivio storico (interventi anteriori al 2000) di Pisa consiste in una unica pianta della metà degli anni '50.

³² I dati grezzi, infatti, non sono sottoposti a tutte le limitazioni della letteratura scientifica. È facile immaginare che gli archeologi abbiano decine di database in attesa di una qualsiasi URL presso la quale rendere disponibili i propri dati ad altri archeologi, altri ricercatori, e perché no al pubblico in generale; non possiamo dimenticare che si tratta di dati pubblici, spesso creati con i soldi dei contribuenti.

³³ «The Swain report looked into the usage of archaeological archives and concluded that they were a grossly under-utilised resource. Reasons for this under-use may include difficulty in locating information about the contents of archives and the dispersal of finds and documentary material to different archival repositories. The development of appropriate resource discovery tools is fundamentally important in helping potential users not only to find the material they require, enabling re-use of digital resources, but also to point them towards archive repositories containing the material of interest»: <http://ads.ahds.ac.uk/project/goodguides/excavation/sect13.html> (accesso 4 aprile 2009).

dalla considerazione che la pratica archeografica, sia essa legata ad operatori professionali, sia alle strutture dedicate alla ricerca, 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 che la condivisione dei dati grezzi deve essere considerata come una vera e propria pubblicazione scientifica, salvaguardando sia la conoscenza, sia il lavoro di chi tale dato ha raccolto, e che il web rappresenti il supporto ideale sul quale pubblicare questi dati; infatti, grazie alle attuali tecnologie, distribuire conoscenza costa meno che produrla (LASER 2005, 39). La pubblicazione di questi dati (archeografica) deve, quindi, essere garantita come la pubblicazione dei dati interpretati (archeologica) e dovrebbe essere citata come curatela divisa tra la Direzione Scientifica e chi ha raccolto e prodotto il dato archeografico, mantenendo un diritto di prelazione per un periodo limitato di tempo per la pubblicazione interpretativa (archeologica) (3 anni dal termine dell'indagine e/o della concessione di scavo³⁴), ma lasciando liberi i dati per altre analisi attraverso l'uso di licenze come ad esempio *Creative Commons Share Alike*³⁵.

6. RAW DATA NOW!

Il progetto di creare un archivio digitale archeologico con i dati inerenti Pisa (Medieval Pisa Open Archaeology Project)³⁶ nasce dalla raccolta di una cospicua quantità di dati³⁷ sia archeografici, sia archeologici, per le analisi su base GIS necessarie a sviluppare la ricerca di dottorato, ma si propone anche di garantire una continua implementazione con l'acquisizione dei dati di nuove ricerche. Il principio fondamentale che regola l'immissione dei dati è quello

³⁴ Se in generale il principale metro di valutazione della competitività nelle istituzioni scientifiche e accademiche è proprio la quantità di conoscenza che un qualsiasi ricercatore riesce a rendere pubblica attraverso articoli e libri nel minor tempo possibile (GRUPPO LASER 2005, 39), il mondo archeologico non sembra utilizzare lo stesso metro e al posto del copyright utilizza la non accessibilità dei dati.

³⁵ <http://creativecommons.org/licenses/by-sa/3.0/deed.it> (accesso 4 aprile 2009). L'uso di offrire i dati con questa licenza è stato scelto da alcuni archivi digitali come Open Context.

³⁶ Il sito è attualmente in creazione, pertanto verranno descritti i dati archeografici presenti e i formati.

³⁷ In estrema sintesi il lavoro preparatorio è consistito nell'acquisizione di tutti i dati archeologici disponibili su Pisa, partendo da un censimento complessivo di tutti gli interventi archeologici, di qualsiasi tipo ed epoca (ANICHINI, PARIBENI 2005), che sono stati georeferenziati e vettorializzati. Il passo successivo è stato l'acquisizione dei dati di scavo pubblicati, successivamente georeferenziati e vettorializzati. Sono stati acquisiti, georeferenziati e vettorializzati anche tutti i dati interpretati presenti, come le ricostruzioni proposte del tessuto urbano e i dati ambientali. Sono attualmente stati inseriti tutti i dati archeografici (piante di strato, immagini, relazioni, schede US, matrix, quantificazioni reperti) relativi agli interventi eseguiti direttamente dallo scrivente o per i quali si è avuta l'adesione al progetto (Scavo di Piazza S. Omobono; Via Consoli del Mare, Via Uffizi, Via Gereschi, ANICHINI, GATTIGLIA 2008, Palazzo Scotto, GATTIGLIA, MILANESE 2006).

che tutti gli interventi archeologici eseguiti in città, di qualsiasi tipologia o dimensione, siano importanti, così come tutti i dati “ricostruttivi”³⁸. L’archivio digitale sarà diviso in due sezioni: la prima contenente tutti i dati archeografici; la seconda i dati interpretati (come ad esempio le ipotesi ricostruttive)³⁹.

Ad ogni intervento di scavo saranno associati (in una specie di meta-database generale), in formato testuale, tutti i dati inerenti lo scavo stesso, la produzione archeografica, la struttura e il formato dei dati digitali, secondo lo schema riportato di seguito che rappresenta una semplificazione ed un riadattamento dello schema proposto dall’ADS⁴⁰:

Storia dell’indagine:

- scopo dell’indagine;
- tematiche dell’indagine;
- metodologia di indagine;
- tipologia della documentazione archeografica;
- localizzazione geografica;
- arco cronologico;
- Direzione Scientifica/équipe di scavo;
- altro...

Fonti utilizzate per la creazione dei dati:

- archivi interrogati;
- cartografia utilizzata per la georeferenziazione;
- precedenti indagini archeologiche nell’area indagata.

Metodo e struttura dei dati:

- creazione del dataset;
- georeferenziazione dei dati;
- cartografia utilizzata per la georeferenziazione;
- lista dei file presenti e loro contenuto;
- lista degli identificativi assegnati;
- lista dei codici e loro significato;
- thesauri;
- indicazione dei sistemi utilizzati nell’acquisizione dati;
- descrizione di eventuali conversioni ad altri formati;
- staff;
- altro...

³⁸ Tali dati risultano uno strumento di lavoro imprescindibile per chi si trovi ad operare su un GIS urbano, dal momento che ai fini della ricostruzione topografica rappresentare correttamente un edificio ipotetico o un percorso ipotetico di un tratto del circuito murario non è meno importante che riportarne su carta uno effettivamente esistente (NARDINI 2004, 366).

³⁹ In questo contributo analizziamo solo la parte inerente i dati di scavo.

⁴⁰ <http://ads.ahds.ac.uk/project/goodguides/excavation/>.

Relazioni:

- bibliografia;
- luogo di conservazione della documentazione archeografica;
- luogo di conservazione dei reperti.

6.1 I formati

I dati digitali sono stati realizzati per la maggior parte con software proprietari. Sebbene questa scelta possa essere discutibile, è anche vero che da un lato l'Open Source rappresenta la possibilità di adottare alternative possibili, piuttosto che una completa "riconversione" (PESCARIN 2006, 138), dall'altro i software proprietari rappresentano la maggioranza dei software impiegati, e che con questi bisognerà fare i conti nella creazione di archivi digitali. Il concetto base dell'*open data*, infatti, in maniera pragmatica, non è tanto, o non solo, come i dati siano stati realizzati, quanto piuttosto che siano condivisi e messi a disposizione della comunità archeologica. Pertanto la scelta dei formati tramite i quali effettuare questo scambio di conoscenza è stata fatta partendo dall'idea di raggiungere l'utenza più ampia possibile⁴¹.

6.1.1 Database

I database creati, in MSAccess 2003, saranno resi disponibili in formato .xml⁴². Ideato come meta-linguaggio adatto a rappresentare contenuti web e informazioni provenienti da basi di dati, esso permette la descrizione in forma gerarchica di contenuti testuali. La scelta di usare XML è dovuta al fatto che questo ben si adatta all'obiettivo di archiviare in modo trasparente e flessibile i dati archeologici (D'ANDREA 2006, 51) e che la sua trasparenza⁴³ rappresenta un vantaggio rispetto ai formati proprietari di archiviazione dei database e ai problemi connessi alle continue evoluzioni del software.

6.1.2 CAD

I file CAD sono stati creati nel formato proprietario di AutoDesk. Non esistono formati standard per scambiare dati CAD tra differenti software. Un buon suggerimento può essere quello di usare il formato più comune, dal mo-

⁴¹ Non è stata prevista, per un'eventuale futura implementazione dei dati, alcuna differenza tra il formato di consegna e quello di distribuzione cosa, invece, prevista da progetti più complessi e completi come quello dell'ADS, e che, probabilmente permette un maggiore più facile reperimento dei dati stessi. Lo stesso vale per la scelta di distribuire in un unico formato, ben consci delle avvertenze dell'ADS secondo cui «the more formats a dataset is copied into, the greater the chance of it surviving».

⁴² XML (eXtensible Markup Language), definito anche "ASCII del 2000" (SIGNORE, MISIKOFF, MOSCATI 2005, 299), ha una sintassi semplice, la sua elaborazione automatica è poco complessa e un documento XML resta comprensibile alla lettura diretta.

⁴³ "documents should be human-legible and reasonably clear" <http://www.w3.org/>.

mento che più file esistono in un dato formato maggiori saranno i convertitori scritti per quel formato; attualmente il software CAD più usato è AutoCAD di AutoDesk con i suoi formati .dwg e .dxf⁴⁴. Pertanto si è scelto di distribuire i file CAD proprio con questi due formati sebbene presentino problemi di compatibilità non solo con altri software, ma con differenti versioni del prodotto originario e debbano costantemente essere migrati alla versione più recente; questo vale anche per il formato .dxf (drawing exchange format), pur essendo un formato di output per trasferire file ad altri programmi. La OpenDWG Alliance⁴⁵ si batte per rendere il formato .dwg uno standard pubblico e non è detto che non riesca a vincere la sua battaglia.

6.1.3 GIS

I file GIS sono stati realizzati con ArcGIS ESRI. Si è scelto di distribuire i file GIS con il formato .shp. Sebbene si tratti di un formato proprietario di ESRI, rappresenta il formato vettoriale GIS più diffuso, condivisibile e rielaborabile da tutti i principali software open source (GRASS, OpenJump, Qgis, gvSIG). Tale scelta è condivisa dall'ADS⁴⁶, che però distribuisce anche nei formati .dwg e .dxf. Non viene utilizzato il formato di interscambio E00, perché oltre ad essere un formato proprietario, presenta problemi nella conversione e non è utilizzabile dalla maggior parte dei software open source.

6.1.4 Immagini

Per le immagini si è scelto di distribuire in formato JPEG⁴⁷, che non solo è open source, ma è anche lo standard di compressione delle immagini fotografiche più utilizzato.

6.1.5 Testi

I file di testo saranno distribuiti in formato PDF e in Open Document Format (ODF), formato standard per il salvataggio e lo scambio dei documenti da ufficio, grazie all'approvazione da parte dell'ISO avvenuta nel 2006. Attualmente è supportato, tra gli altri, dai programmi OpenOffice.org, Google Docs e Spreadsheets, IBM Workplace, Lotus Symphony e Sun's StarOffice. Microsoft ancora non supporta questo formato, ma si è impegnata a supportarlo nel prossimo service pack di Microsoft Office 2007, previsto per il 2009; nel frattempo è scaricabile il plugin ODF per Microsoft Office, non compatibile però con Microsoft Office 2007, messo a disposizione da Sun Microsystems.

⁴⁴ <http://ads.ahds.ac.uk/project/goodguides/cad/sect45.html>, accesso a aprile 2009 (ADS CAD Guide to good practice).

⁴⁵ <http://www.opendesign.com/>.

⁴⁶ <http://ads.ahds.ac.uk/project/goodguides/gis/> (accesso 4 aprile 2009).

⁴⁷ <http://www.jpeg.org/>.

Tipologia	Formato nativo	Distribuzione
Databases	MDB	XML
CAD	DXF, DWG	DXF, DWG
GIS	SHP	SHP
Immagini	JPEG, TIFF	JPEG
Foglio elettronico	XLS	XLS
Testi	ODF	ODF, PDF
Metadata	ODF	ODF, XML

Tab. 1 – Formativi nativi e formati di distribuzione scelti.

6.1.6 Fogli elettronici

Utili per la redazione dei matrix, saranno distribuiti in formato .xls, che pur essendo proprietario viene letto da OpenOffice.org Calc, Gnumeric, KSpread e NeoOffice

6.2 *Metadati*

Per i metadati si è scelto di adottare lo schema Dublin Core (DC)⁴⁸, nato in ambito bibliotecario e gestito dalla *Dublin Core Metadata Initiative* (DCMI). È un sistema semplice che si basa su un set di quindici elementi in grado di descrivere adeguatamente, anche in virtù della loro ripetibilità, risorse documentarie di ogni genere:

- title: il nome attribuito al file;
- creator: l'autore del file, con il contatto mail;
- subject: l'argomento trattato dal file espresso come keywords;
- description: scopo, obiettivi e breve descrizione del file;
- publisher: chi rende accessibile il file;
- contributor: indica chi ha contribuito in qualche modo alla realizzazione della risorsa;
- date: data di creazione e di pubblicazione;
- type: tipo di file;
- format: formato digitale del file;
- identifier: URL che identifica la fonte digitale;
- source: l'eventuale/i fonti da cui il file deriva;
- language: il contesto linguistico in cui si inserisce il file;
- relation: eventuali risorse collegate, compresi eventuali ulteriori archivi;
- coverage: area geografica di riferimento; sistema di coordinate di riferimento;
- rights: i diritti sull'utilizzo della risorsa.

⁴⁸ <http://dublincore.org/>.

7. CONCLUSIONI

Parafrasando: la ricostruzione del paesaggio archeologico è un processo dinamico che deve sottoporre ad analisi diversi tipi di dati e che dunque richiede il coinvolgimento di diverse professionalità. Un approccio open data, per i motivi sopra esposti, contribuirebbe a costruire una rete di ricerca attraverso la quale gruppi di lavoro a distanza potrebbero collaborare, mediante Internet, alla ricostruzione di un territorio (PESCARIN 2006, 145). Pubblicare i dati grezzi, oltre a consentire un aumento, probabilmente esponenziale, delle nostre conoscenze, consentirebbe di fare il punto sullo stato della ricerca archeologica in Italia sia da un punto di vista storico, sia da un punto di vista metodologico, e permetterebbe di valutare con maggior dettaglio tutte quelle operazioni di ricerca che potrebbero essere meglio finalizzate, con risparmio di tempo e di costi. Avere a disposizione dati grezzi non è un aiuto solo per la ricerca, ma anche per chi svolge il lavoro dell'archeologo da un punto di vista professionale (sono e saranno sempre la maggioranza) permettendo una migliore conoscenza del territorio in cui si troveranno ad operare, migliorando la qualità del proprio operato e di quelle operazioni di valutazione del potenziale archeologico che sempre più sono loro richieste e che sempre più saranno utili per la gestione del territorio e per il suo sviluppo consapevole. Infine, rendere accessibili i dati grezzi ha un aspetto sociale importante legato alla possibilità di trasmettere quelle informazioni che formano la memoria e consentono un corretto sviluppo culturale.

GABRIELE GATTIGLIA
Università di Pisa

BIBLIOGRAFIA

- ANICHINI F., GATTIGLIA G. 2008 (con contributi di M. FEBBRARO, M. GIORGIO), *Nuovi dati sulla topografia di Pisa medievale tra X e XVI secolo. Le indagini archeologiche di Piazza Santomobono, Via Uffizi, Via Consoli del Mare e Via Gereschi*, «Archeologia Medievale», 35, 113-142.
- ANICHINI F., PARIBENI E. 2005, *Il Gis archeologico della città di Pisa. Primi risultati per la tutela e la gestione del patrimonio archeologico pisano*, «Notiziario della Soprintendenza Archeologica per la Toscana», 1, 205-210.
- BERNARDI M. (ed.) 1992, *Archeologia del paesaggio*, Firenze, All'Insegna del Giglio.
- BROGIOLO G.P. 2000, *Urbana, Archeologia*, in R. FRANCOVICH, D. MANACORDA (eds.), *Dizionario di archeologia*, Roma-Bari, Laterza, 350-355.
- CAMBI F., TERRENATO N. 1994, *Introduzione all'archeologia dei paesaggi*, Roma, NIS.
- D'ANDREA A. 2006, *Documentazione archeologica, standard e trattamento informatico*, Budapest, Epoch, Archeolingua.
- GATTIGLIA G., MILANESE M. (eds.) 2006, *Palazzo Scotto Corsini. Archeologia e storia delle trasformazioni di un'area urbana a Pisa tra XI e XX secolo*, Pisa, Felici Editore.
- GATTIGLIA G., STAGNO A.M. 2005, *La documentazione scritta nella ricognizione archeologica sul territorio: un "vecchio" sistema di schedatura*, «Archeologia Medievale», 32, 453-459.

- GUIDI A. 1994, *I metodi della ricerca archeologica*, Roma-Bari, Laterza.
- LASER 2005, *Il sapere liberato: il movimento dell'open source e la ricerca scientifica*, Milano, Feltrinelli.
- MANNONI T. 2000, *Alcune considerazioni metodologiche*, in G.P. BROGIOLO, G. OLCESE (eds.), *Produzione ceramica in area padana tra il II secolo a.C. e il VII secolo d.C.: nuovi dati e prospettive di ricerca*, Convegno internazionale (Desenzano del Garda, 8-10 aprile 1999), Mantova, SAP, 217-219.
- MOBERG C.A. 1981, *Introduzione all'archeologia*, Milano, Mazzotta.
- NARDINI A. 2004, *L'applicazione della tecnologia GIS nella gestione dei contesti urbani*, in *Relazione del progetto archeologia dei paesaggi medievali (anno 2000-2004)*, 365-374 (<http://www.paesaggimedievali.it/>).
- PESCARIN S. 2006, *Open source in archeologia. Nuove prospettive per la ricerca*, «Archeologia e Calcolatori», 17, 137-155.
- SIGNORE O., MISSIKOFF O., MOSCATI P. 2005, *La gestione della conoscenza in archeologia: modelli, linguaggi e strumenti di modellazione concettuale dall'xml al Semantic Web*, «Archeologia e Calcolatori», 16, 291-319.
- TERRENATO N. 2006, *Le misure (del campione) contano! Il paradosso dei fenomeni globali e delle ricognizioni locali*, in N. MANCASSOLA, F. SAGGIORO (eds.) *Medioevo, paesaggi e metodi*, Mantova, SAP, 9-24.
- ZANINI E., COSTA S. 2006, *Organizzare il processo conoscitivo nell'indagine archeologica: riflessioni metodologiche ed esperimenti digitali*, «Archeologia e Calcolatori», 17, 241-264.

ABSTRACT

Raw data, the archaeological source code, that is, all the excavation and fieldwork recorded, should be published on digital on-line archives stored in recognizable format documents. This paper contributes to the discussion with the description of a project dedicated to the creation of an urban GIS about medieval Pisa.

