

**JULIAN
BOGDANI**

ARCHEOLOGIA E TECNOLOGIE DI RETE

METODI STRUMENTI E RISORSE DIGITALI

Archeologia e tecnologie di rete

Metodi, strumenti e risorse digitali

di
Julian Bogdani

La collana “I cardini” di Groma è diretta da Enrico Giorgi

Questo volume è stato sottoposto a due revisioni paritarie a doppio cieco, secondo le stesse modalità previste per gli articoli scientifici della rivista *Groma Documenting Archaeology*.

Per maggiori informazioni <https://groma.unibo.it/publish>.

Roma 2019

ISBN: 978-88-98392-99-5

BraDypUS Editore. Roma

I contenuti della collana “I cardini” di Groma vengono diffusi nella versione cartacea ed elettronica secondo la licenza:

Creative Commons, Attribuzione – Non commerciale – Non opere derivate 3.0 Italia (testo completo all’indirizzo: <https://creativecommons.org/licenses/by-nc-nd/3.0/it/legalcode>).

Questo significa che i lettori sono liberi di: riprodurre, distribuire, comunicare ed esporre in pubblico quest’opera, a condizione che il suo contenuto non venga alterato o trasformato, che venga attribuita la paternità dell’opera all’autore e che infine l’opera non venga utilizzata per fini commerciali. Per tutti gli altri usi il copyright appartiene all’autore.

L’autore e l’editore difendono la gratuità del prestito bibliotecario e sono contrari a norme o direttive che, monetizzando tale servizio, limitino l’accesso alla cultura. Per questo motivo rinunciano a riscuotere eventuali *royalties* derivanti dal prestito bibliotecario di opere di questa collana. L’editore garantirà inoltre sempre il libero accesso ai contenuti dei volumi, senza limitazioni alla loro distribuzione in alcun modo.

A Fabiana e Greta

Indice

Prefazione di Enrico Giorgi.....	7
Introduzione.....	9
Capitolo 1. L'archeologia e i dati.....	11
Capitolo 2. Il Web come strumento di comunicazione, condivisione e pubblicazione.....	19
Breve storia del Web, a 30 anni dalla sua creazione.....	19
Il Web come deposito pubblico di dati.....	28
Le banche dati sul Web: il Web dinamico.....	33
Ritorno al futuro. Il Web semantico e le sue applicazioni in archeologia.....	38
Web semantico: ricette e ingredienti.....	50
Capitolo 3. Banche dati: archiviazione, gestione, pubblicazione.....	57
Il modello relazionale.....	57
Un linguaggio unico di interrogazione: SQL, e i suoi dialetti.....	68
Implementazione del modello relazionale. <i>Software</i> e formati.....	72
Database e Web, cambiamenti e prospettive.....	77
NoSQL: relazionale ma non solo.....	80
Capitolo 4. GIS e webGIS tra archiviazione, analisi e pubblicazione dei dati geografici.....	89
Standard per la pubblicazione dei dati geografici strutturati sul Web.....	97
Dietro le quinte, <i>server</i> cartografici.....	102
Sul <i>browser</i> : standard, librerie e strumenti webGIS.....	105
Pubblicare dati archeologicigeografici sul web.....	112
Capitolo 5. Free Software, Open Source e Open Data. Copyright e licenze nell'era della condivisione.....	123

Movimento Free Software.....	124
Open Source Initiative.....	126
Creative Commons.....	128
Open Data.....	130
Dati aperti... e collegati.....	132
Appendice. Casi di studio di applicazioni web e risorse <i>online</i> per la ricerca archeologica e antichistica.....	
Europeana.....	142
Ariadne.....	144
Nomisma.org.....	146
Kerameikos.org.....	148
Ancient World Mapping Center.....	150
Pleiades.....	153
Digital Atlas of Roman Empire.....	156
Digital Atlas of Roman and Medieval Civilizations (DARMC).....	159
OmnesViae e Vici.org.....	161
ORBIS: The Stanford Geospatial Network Model of the Roman World.....	164
ToposText.....	166
Pelagios Commons.....	167
FastiOnline.....	174
MAPPA Project.....	176
Archaeology Data Service (ADS).....	179
Open Context.....	182
SITAR.....	184
iDAI.objects arachne.....	186
PAThs. L'Atlante archeologico della letteratura Copta.....	188
Opere citate.....	193

Prefazione

di Enrico Giorgi.

Direttore della collana “I Cardini” di Groma

Sono trascorsi ormai vent’anni da quando, nell’ambito delle iniziative editoriali legate alla redazione della Rivista Groma, fu inaugurata la collana dei Cardini, che nel frattempo ha accolto alcuni importanti studi capaci di ampliare ulteriormente le prospettive di ricerca già presenti nella rivista stessa. Finalmente con il volume di Julian Bogdani sull’*Archeologia e le risorse digitali di rete* questa tradizione viene ripresa e rinnovata nel migliore dei modi.

Innanzitutto, giova ricordarlo, l’autore è uno dei più attivi ricercatori del *team* impegnato nella redazione della rivista, che in questi anni si è profondamente aggiornata proprio aprendosi verso il web, con l’*open access peer-reviewed e-journal*, online dal 2016 (groma.unibo.it).

Ma il volume di Julian Bogdani recupera una tradizione fondante del gruppo di ricercatori cresciuti in questi anni attorno all’attività della rivista, ossia quella dei saggi scientifici che permettono di avvicinare anche quegli archeologi che sono in una fase di strutturazione o di rinnovamento del loro percorso formativo, specialmente quelli interessati anche all’aggiornamento rispetto agli aspetti più attuali e tecnologici della disciplina. In questo senso andava, ad esempio, il fortunato volume *In profondità senza scavare*, pubblicato nel 2009, e dedicato alla diagnostica per l’archeologia, certamente oggi bisognoso di un aggiornamento, ma in quel momento perfettamente in linea con le tendenze della nascente archeologia preventiva. Già in quell’occasione l’autore si era dedicato alla riflessione sulle banche dati, che ora viene ripresa e attualizzata più compiutamente nell’ambito di questa nuova opera monografica. Oggi come allora, il tenore è quello del saggio scientifico, che tuttavia può essere utilizzato anche come strumento didattico, secondo la migliore tradizione della ricerca universitaria che sposa, sul campo e in laboratorio, l’acquisizione e la catalogazione con tec-

nologie innovative con la predisposizione e l'insegnamento di metodologie adeguate, nella convinzione che metodi e strumenti siano parte inscindibile della ricerca e della didattica. Se è vero che la disponibilità di nuove tecnologie ha profondamente rinnovato l'archeologia, questo processo è stato potentemente accelerato proprio dalla massiccia introduzione delle risorse digitali e della rete. Ormai da qualche decennio l'avvento dell'archeologia digitale ha aumentato esponenzialmente la quantità dei dati raccolti sul campo. Questa dinamica, in un primo momento, ha creato anche uno stallo, quasi l'incapacità di gestire una tale massa di informazioni con gli strumenti critici del passato. La possibilità di far dialogare tra loro, attraverso la rete, le banche dati che normalmente fanno capo ai vari gruppi di ricerca, apre ora nuove prospettive, ma pone anche nuovi problemi ai quali questo volume prova a rispondere, fornendo le conoscenze indispensabili per procedere in questa direzione e facendo riferimento ad alcuni casi di studio esemplificativi delle varie tendenze. L'organizzazione del volume e la modalità di presentazione dei temi più tecnici con un linguaggio accessibile agli archeologi rappresenta un valore aggiunto che, come abbiamo anticipato, riprende e rinnova nel migliore dei modi la tradizione di Groma.

Introduzione

Il presente volume ha come tema un aspetto ben definito di quella che viene comunemente chiamata *Digital Archaeology*¹, ovvero l'applicazione alla ricerca archeologica di metodi, tecniche e tecnologie di rete. Queste tecnologie sono nate, o hanno subito un profondo sviluppo, nel corso degli ultimi decenni, ossia da quando, alla fine degli anni Ottanta del secolo scorso, Internet e soprattutto il World Wide Web si sono capillarmente diffusi presso un numero altissimo di utenti. Va innanzitutto chiarito che il termine *Digital Archaeology* abbraccia un campo applicativo estremamente ampio e fluido, che difficilmente può essere circoscritto in modo preciso e onnicomprensivo. Gli archeologi hanno precocemente fatto impiego di metodologie di ambito *digitale*, anche quando l'uso dei calcolatori era relegato a una nicchia sperimentale², e da almeno tre decenni importanti iniziative editoriali e congressuali – nazionali e internazionali – ne hanno esplorato le potenzialità: basti pensare agli incontri annuali (e alle relative pubblicazioni) del CAA (Computer Applications and Quantitative Methods in Archaeology)³, o a progetti come “Archeologia e Calcolatori”⁴, rivista italiana che dal 1990 offre una panoramica aggiornata della ricerca, nazionale ed estera, in questo specifico campo.

Una trattazione esaustiva delle possibili *archeologie digitali* è un compito di gran lunga superiore alle competenze di chi scrive e produrrebbe probabilmente un risultato dispersivo e sproporzionato sia per un pubblico di

¹ Thomas L. Evans e Patrick T. Daly, *Digital Archaeology Bridging Method and Theory* (Abingdon, Oxon; New York: Routledge, 2006).

² Si veda a tal proposito il convegno di Lecce del 1986, edito l'anno successivo, Francesco D'Andria, a c. di, *Informatica e archeologia classica: atti del Convegno internazionale : Lecce, 12-13 maggio 1986* (Galatina, Lecce): Congedo editore, 1987).

³ <https://caa-international.org/>; le pubblicazioni delle conferenze in open access (dalla prima conferenza del 1973 a quella del 2011) sono disponibili all'indirizzo <https://proceedings.caaconference.org/>.

⁴ <http://www.archcalc.cnr.it/>

studenti (pur a vari livelli di avanzamento negli studi) sia per specialisti di ambiti disciplinari diversi, tuttavia curiosi di esplorare nuove possibilità di fare ricerca. Appare invece preferibile affrontare il tema mediante un testo più agile che, senza sminuirne la complessità, tratti in sintesi gli aspetti specifici della *Digital Archaeology* riguardanti il Web e offra gli strumenti di conoscenza basilari per il suo corretto inquadramento e, cosa fondamentale, per la sua contestualizzazione all'interno dello studio e della divulgazione del passato.

Questo volume è quindi concepito come un'analisi specifica sui temi dell'applicazione delle tecnologie di rete (principalmente Web) alla ricerca archeologica, aspetto che i manuali generali di informatica applicata alle scienze umane⁵ trattano, per mancanza di spazio, in maniera superficiale. Vengono qui peraltro del tutto trascurati, per i motivi sopra esposti, gli aspetti di base della scienza informatica, per i quali si rimanda alla relativa manualistica.

⁵ Per esempio il validissimo Marco Lazzari, *Informatica umanistica* (Milano: Mc Graw-Hill, 2014). Su gli aspetti più prettamente archeologici si vedano anche Giovanna Alvoni, Ulrich Rausch, e Legoprint, *Scienze dell'antichità per via informatica: banche dati, Internet e risorse elettroniche nello studio dell'antichità classica* (Bologna: CLUEB, 2002); Ada Gabucci, *Informatica applicata all'archeologia* (Roma: Carocci, 2005); Andrea D'Andrea, *Documentazione Archeologica, Standard e Trattamento Informatico*, vol. 2 (Budapest: Archaeolingua, 2006). Sullo specifico ambito delle applicazioni informatiche sullo studio e l'edizione dei testi, si veda Tito Orlandi, *Informatica testuale. Teoria e prassi* (Roma-Bari: Laterza, 2010).

Capitolo 1

L'archeologia e i dati

Siamo spettatori (e protagonisti) di un'epoca in cui i prodotti digitali pervadono ogni aspetto della quotidianità: nell'economia, sul lavoro, nell'intrattenimento, nel tempo libero, in famiglia, nei tanti servizi di istruzione e informazione offerti a qualsiasi scala. In tutta una serie di strumenti digitali, il confine tra sapere tecnico e conoscenza di base si sposta in maniera esponenzialmente rapida. Il rapporto quotidiano con un dispositivo tecnologico (dal telefono "smart" alla costosa *workstation*) o con la Rete, fino a qualche anno fa episodico o caratteristico solo di certe categorie professionali, è oggi capillare e gran parte di noi si serve di esso per acquistare un biglietto dell'autobus, produrre un certificato o accedere a gran parte dei servizi pubblici: strumenti digitali e di rete come lo SPID¹ sono sempre più l'unico mezzo possibile di dialogo tra Stato, amministrazione e cittadini.

Questa rivoluzione ha investito, in maniera sempre più totalizzante, anche gli aspetti della ricerca, compresa quella archeologica, che per tradizione è particolarmente attenta alle innovazioni della tecnologia tanto che, nel 2019, ci chiediamo se esista una ricerca archeologica che possa prescindere dalle risorse digitali. Dalle fotografie di contesti e reperti, alla documentazione grafica e topografica, fino alle relazioni e ai report di sintesi di fine campagna, tutto il materiale preso in esame sul campo e in laboratorio passa attraverso un calcolatore elettronico. Persino i dati che vengono raccolti sul campo in forma cartacea – descrizioni, appunti, diari, disegni, elenchi, promemoria – vengono in genere digitalizzati, in tempo reale contestualmente alla raccolta o in un momento immediatamente successivo. Per non parlare del fatto che la maggior parte della documentazione grafica e topografica in due e tre dimensioni viene oggi prodotta in formato già all'origine digitale, e sempre più di rado ne viene esportata una versione cartacea. Una vera e propria inversione di tendenza rispetto alle prime

¹ <https://www.spid.gov.it/>

applicazioni informatiche all'archeologia, quando l'impiego di strumenti e tecniche digitali di supporto al rilievo aveva come fine quello di produrre documentazione da stampare su supporti cartacei. La mole di dati che oggi è possibile raccogliere in maniera rapida, con dispositivi e metodologie sempre più facili da utilizzare, si può "costringere" solo a fatica entro formati cartacei, troppo riduttivi, semplitistici e sovente antieconomici², in primo luogo per il limite costitutivo della carta (e della conseguente stampa), supporto bidimensionale e piuttosto costoso. In seconda istanza, e sempre più spesso, va rilevato che l'elaborazione dei dati digitali e il loro trattamento richiedono costi viepiù crescenti in termini di tempo e risorse finanziarie. Come è semplice, economico e veloce acquisire e manipolare digitalmente dati complessi di alta qualità e precisione in tre o quattro dimensioni, così è lungo, difficile e quindi dispendioso tradurre in formato cartaceo questi prodotti.

Se l'adozione di strumenti e mezzi digitali nel processo della ricerca archeologica è sufficiente per etichettarlo come "digitale", allora si può senza timore affermare che non esistono più – alle soglie degli anni Venti del XXI secolo – archeologie non digitali. L'uso di strumenti digitali, dai programmi di videoscrittura ai dispositivi di ripresa ed elaborazione fotografica e di ricerca in rete, è ormai consolidato e diffuso in maniera così capillare che sarebbe fuorviante e improduttivo scinderlo dal più ampio contesto della *Digital Archaeology* e studiarlo nelle sue specifiche nicchie. Per applicazioni informatiche alla metodologia archeologica ci si riferirà quindi a quegli strumenti, *software* e algoritmi capaci di garantire un trattamento automatizzato digitale di raccolta, registrazione, analisi o pubblicazione dei dati e il cui uso ha, come vedremo, un impatto notevolissimo sulla metodologia della ricerca, sia in fase di acquisizione che in fase di analisi e sintesi delle informazioni. Sono applicazioni create di norma per scopi diversi da quelli delle scienze umane e adattati a queste ultime, oppure, sempre più spesso, studiate e realizzate *ad hoc*, da o per umanisti.

Due sono oggi le spinte principali che incentivano e favoriscono l'adozione di sistemi automatizzati digitali di gestione delle informazioni: la

² Per la complessità che raggiunge oggi il dato archeologico e per la ricchezza di formati e forme possibili di acquisizione e rappresentazione si vedano Stefano Campana, Gabriella Carpentiero, Marianna Cirillo e Roberto Scopigno, a c. di, *CAA2015. Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology* (Oxford: Archaeopress, 2016); Maurizio Forte e Stefano Campana, *Digital Methods and Remote Sensing in Archaeology* *Archaeology in the Age of Sensing* (Cham: Springer International Publishing, 2016).

moltiplicazione delle sorgenti di dati, da una parte e, dall'altra, lo sviluppo stesso di questi sistemi, che favorisce la sperimentazione e l'adozione di nuovi protocolli di gestione, analisi e visualizzazione, spalancando nuovi orizzonti su molti aspetti della ricerca, della fruizione e della valorizzazione dell'indagine sul passato.

Per procedere con ordine, la ricerca archeologica si è sempre posta a metà strada tra le discipline umanistiche e quelle scientifiche e da queste ultime ha ereditato diverse metodologie di intervento e di analisi. Lo stesso principio della stratigrafia non è altro che l'adattamento a contesti antropici di un metodo nato, per contesti naturali, nell'ambito della scienza geologica³. Attraverso l'applicazione di metodi scientifici, in archeologia è sempre più facile la raccolta dei dati sul campo, ed è aumentata e continua a crescere la quantità di informazioni da immagazzinare, interrogare e considerare in fase di sintesi. Si pensi per esempio a quanto hanno inciso gli sviluppi tecnologici nel campo della geomatica e del rilevamento, come pure la messa a punto di sistemi sempre più potenti, facili da usare e, non meno importante, economici sulla qualità della documentazione topografica di uno scavo o di una ricerca su scala territoriale. Lo stesso vale per la sempre più vasta scelta di analisi di laboratorio che oggi si offre all'archeologo: al tradizionale – ma sempre efficace – radiocarbonio si affiancano metodologie sempre più precise, grazie all'enorme esperienza maturata sull'indagine dei resti organici e inorganici, finalizzata alla ricostruzione fedele dell'ambiente, delle risorse alimentari e delle diverse tecnologie impiegate per realizzare i manufatti che si vanno a indagare. La stessa osservazione autoptica, tradizionalmente attuata sul campo, è diventata negli anni sempre più attenta, grazie al progressivo affermarsi del concetto – ovvio ma mai sottolineato a sufficienza – che il mestiere dell'archeologo non è quello di scoprire *cose*, bensì di tentare di ricostruire le storie di persone e luoghi non più esistenti attraverso le poche tracce materiali che la loro interazione con l'ambiente ha lasciato lungo il corso dei secoli. Queste tracce sono state a lungo identificate con le opere d'arte e gli oggetti di pregio, che gli archeologi pionieri della disciplina hanno voluto *salvare* dalle sabbie del tempo redigendo resoconti, elenchi e dettagliate descrizioni. Dopo una lunga battaglia, si è iniziato a dare valore anche ai manufatti artisticamente meno rilevanti (ma

³ Per una storia si veda Edward C Harris, *Principi di stratigrafia archeologica* (Roma: Carocci, 2017); Colin Renfrew e Paul Bahn, *Archeologia. Teoria, metodi e pratica*, trad. da A. Vezzoli e F. Piccarreta (Bologna: Zanichelli, 2018); per i limiti che questa eredità ha comportato per la ricerca archeologica, si veda Geoff Carver, «Archaeological Information Systems (AIS). Adapting GIS to Archaeological Contexts», *Archäologie Und Computer / 9*. 2004 9 (2004): 53.

non per questo meno densi di informazioni) in un secondo momento ogni tipo di evidenza, fosse essa una moneta, un'opera di artigianato artistico, un frustulo di ceramica àroma, un residuo carbonioso, il colore di uno strato, l'evidenza negativa di un taglio, il segno di un crollo, di una distruzione, di una frana o di un fenomeno alluvionale, è entrata nella sfera di interesse dell'archeologia e viene oggi minuziosamente osservata, classificata e descritta.

L'affinamento della metodologia d'indagine e l'articolazione della cornice teorica entro la quale la disciplina archeologica opera hanno accresciuto la nostra consapevolezza nel rapporto con l'antico, evidenziando come la nostra visione del passato sia influenzata e influenzi il presente. Ciò ha determinato un notevole aumento della quantità di dati e di collegamenti tra i dati che un'indagine archeologica è oggi in grado di generare, non solo per quanto riguarda i grandi progetti di scavo e ricerca condotti e coordinati da importanti istituzioni, talvolta (ma sempre meno) ben finanziati, ma anche per progetti di ricerca di piccole dimensioni, condotti su contesti minori e da parte di piccoli gruppi o ricercatori singoli, che riescono oggi a produrre quantità di dati sempre crescenti. La ricerca sui cosiddetti *big data*, con tutti i limiti del caso, investe oggi anche l'archeologia, proprio per la notevole complessità che caratterizza da sempre i *dataset* archeologici⁴.

A queste considerazioni va sommato un altro aspetto, sempre più evidente nell'archeologia contemporanea, ovvero la grande frammentazione delle carriere e l'iper-specializzazione dei profili. Da una parte questa dinamica costituisce una ricchezza indiscutibile per la disciplina, che trae linfa vitale e rinnovamento dal contatto, dall'ibridazione e dalla compenetrazione con altri àmbiti disciplinari. Dall'altra però rischia di produrre un effetto Babele, palesando il rischio che possa venir meno il minimo comune denominatore necessario alla sintesi e alla messa in atto di quei processi virtuosi (che potremmo definire "olistici", se il termine non fosse così abusato) che traggono il massimo del profitto dall'apporto reciproco delle varie discipline e portano a un risultato finale che costituisca qualcosa in più della mera sommatoria delle singole ricerche.

Questo discorso esula dalle finalità del presente lavoro, ma è utile tenerlo sullo sfondo per non dimenticare che i contesti che indaghiamo possono essere letti da molteplici punti di vista e soprattutto con occhi diversi, ciascuno dei quali aggiunge qualcosa di proprio. È quindi importante ripetere, quasi come un mantra, che i dati che produciamo attraverso il processo

⁴ Gabriele Gattiglia, «Think Big about Data: Archaeology and the Big Data Challenge», *Archäologische Informationen* 38 (2015): 12.

archeologico – per sua natura distruttivo e irreversibile – ci appartengono solo in minima parte, ossia per quel poco che saremo in grado di aggiungere attraverso la nostra analisi e dalla nostra prospettiva. Accogliere altri “occhi”, altre prospettive e altri punti di vista diventa dunque un obbligo morale, uno sforzo verso la conoscenza del mondo antico che non può che essere corale.

Per essere più chiari, un frammento ceramico fornisce alcune informazioni molto specifiche a chi si occupa di tecnologia antica, altre a chi studia l'economia e il commercio, altre ancora a chi, interessato all'epigrafia, ne legge il graffito o al chimico che ne analizza i resti organici sulle pareti interne o, ancora, al fisico che ne campiona l'argilla e la confronta con la propria banca dati di riferimento per determinarne la provenienza. Per non parlare dello stratigrafo, che usa quel frammento per dare un appiglio cronologico all'ordine sequenziale in cui ha disposto le varie azioni naturali e antropiche di cui ha trovato traccia nel contesto in esame, o infine dello studioso delle religioni, che può trovare conferma o smentita di un possibile uso rituale dalla posizione all'interno della sepulture in cui è stato rinvenuto. Compito dell'archeologo è quello di mettere insieme tutte queste informazioni, traendo da ciascuna il maggior potenziale informativo possibile, per offrire una narrazione convincente delle azioni degli individui che, in un dato momento, sono intervenuti su un fenomeno, lasciandovi tracce intenzionali o involontarie, minute o monumentali.

Al giorno d'oggi, la difficoltà di questo processo consiste nel mantenere sotto controllo la complessità derivante dall'accumulo di un'inedita messe di dati, scritti in “codici” che non sempre è possibile e immediato comprendere. Dati in codice che aumentano ogni giorno e diventano sempre più articolati, i cui processi di azione e risultati sono leggibili solo in epitome, senza la reale possibilità di approfondirne in dettaglio ogni aspetto. Le figure ibride di “arqueo-e-qualcosa”, tra le quali anche quella dell'arqueo-informatico⁵, servono proprio come “interpreti” tra questi codici diversi, essendo capaci di filtrare e far circolare l'informazione all'interno di sistemi informativi sempre più dettagliati.

Qual è l'apporto dell'informatica in questo contesto? È possibile un trat-

⁵ Guido Vannini, «Informatica per l'archeologia o archeologia per l'informatica?», in *Atti del I workshop Nazionale di Archeologia Computazionale, Napoli-Firenze 1999*, a c. di Andrea D'Andrea e Fabio Nicolucci, *Archeologia e Calcolatori* 11 (2000): 311–15. Si veda anche Andrea D'Andrea, «Dati Digitali e Metodologia della Ricerca Archeologica», in *Atti del workshop L'integrazione dei dati archeologici digitali - Esperienze e prospettive in Italia. Lecce, Italia, 1-2 Ottobre, 2015*, a c. di Paola Ronzino, *CEUR Workshop Proceedings*, 1634, 2016, 10–17.

tamento digitale automatizzato di queste fonti a fini di gestione, immagazzinamento, condivisione, analisi e pubblicazione? Per diversi di questi processi la modalità automatica è l'unica possibile, per altri, come l'analisi di sintesi, le procedure automatizzate facilitano il recupero dell'informazione, ma non la sua comprensione nella relazione con il contesto. Inoltre, gran parte delle discipline con le quali l'archeologia viene in contatto usano l'informatica come mezzo imprescindibile per l'analisi, il trattamento e la pubblicazione dei dati e quindi la loro assunzione comporta necessariamente l'integrazione degli strumenti, dei linguaggi e della terminologia specifica che le caratterizza.

Tra i vari aspetti pregnanti delle tecnologie informatiche, sono proprio quelli legati alla trasmissione, condivisione e pubblicazione dell'informazione che ultimamente hanno conosciuto un rapidissimo sviluppo, grazie alla capillare diffusione delle reti digitali, alla notevole riduzione dei costi di accesso e al parallelo aumento della velocità di connessione. Questa concomitanza di fattori ha favorito massicci investimenti nel campo dell'*hardware*, dove i dispositivi hanno conosciuto una notevole miniaturizzazione delle dimensioni fisiche e un rilevante aumento della potenza di calcolo⁶. Anche i *software*, i protocolli di comunicazione e le tecnologie di sviluppo hanno conosciuto importanti picchi di sviluppo, con strumenti (e spesso mode) che si susseguono a velocità impressionante.

La potenza di calcolo, la facilità di connessione, i bassi costi di configurazione e mantenimento di sistemi complessi, anche da parte di chi non è informatico di professione, è ben rappresentata dai modelli del cosiddetto *cloud computing*, un'architettura di gestione di risorse informatiche che, grazie al rapporto di scala, è in grado di offrire a utenti finali risorse di calcolo molto potenti, a prezzi davvero contenuti. Questo paradigma sta peraltro mutando la nostra vita quotidiana, invadendo quasi interamente il mondo dei servizi, offrendo nuove opportunità, ma anche – purtroppo – risolvendo vecchi problemi ancora irrisolti che è nostro compito di produttori, editori e consumatori di dati (anche archeologici) affrontare con lucida consapevolezza.

Quale sia il contributo delle tecnologie di rete, in particolare di quelle Web, alla ricerca archeologica e quali possano essere gli scenari del prossimo (e imprevedibile) futuro è uno dei temi affrontati nei seguenti capitoli.

⁶ In informatica questo aumento esponenziale è noto come Prima Legge di Moore, dal nome dello scienziato Gordon Moore che nel 1965 predisse che il numero dei transistor integrati nei chip sarebbe raddoppiato ogni 18 mesi. La legge è stata fino a oggi confermata dallo sviluppo tecnologico.

Per scelta consapevole, si sorvolerà su gran parte di quello che riguarda i singoli *software*, poiché si tratta di un mondo estremamente dinamico e facile da approfondire mediante guide e materiale didattico reperibile *online*. Maggiore attenzione, invece, verrà posta sui concetti chiave che hanno portato alla definizione delle varie tecnologie Web, degli standard e dei formati aperti. Questi argomenti presentano una significativa stabilità nel tempo e soprattutto sono i veri responsabili di quell'importante cambiamento che il World Wide Web ha determinato anche per l'archeologia. Il tema della fragilità e dell'instabilità degli archivi, creati con tanta fatica, e di quali possano essere le buone pratiche per dare ai nostri dati un'esistenza e un'accessibilità durevole, è di fondamentale importanza per chi opera sulle poche tracce rimaste del passato, e verrà quindi ampiamente trattato. Uno spazio sufficientemente ampio verrà dedicato all'esposizione di alcuni casi applicativi, ovvero una serie progetti nazionali e internazionali che fanno largo uso di tecnologie di rete. Questa esposizione è finalizzata non solo a dare concretezza espositiva ai metodi e strumenti esposti nei capitoli precedenti, ma anche a fornire risorse di ricerca ad accesso aperto, dati che possono essere facilmente riutilizzati e arricchiti.

Capitolo 2

Il Web come strumento di comunicazione, condivisione e pubblicazione

Breve storia del Web, a 30 anni dalla sua creazione

I tentativi, coronati da successo, di mettere in rete computer distanti tra loro risalgono alla prima metà del secolo scorso, ma furono gli anni Sessanta a vedere la concretizzazione degli sforzi profusi in questa direzione. Fu la tesi di dottorato di Leonard Kleinrock, pubblicata poi nel 1964¹, ad aprire la strada alla teoria e alla pratica delle tecniche di commutazione dei pacchetti, che, consentendo l'utilizzo dello stesso canale di comunicazione da parte di più flussi di dati, hanno rappresentato un passo fondamentale per la realizzazione di una rete globale di computer ad accesso non esclusivo. Kleinrock veniva dal MIT (Massachusetts Institute of Technology), così come Joseph Licklider, psicologo di formazione, il cui interesse scientifico si rivolse all'interazione tra macchine ed esseri umani². La sua carriera lo portò all'ARPA (Advanced Research Projects Agency), un'agenzia del Dipartimento della difesa statunitense dedicata ai progetti di ricerca avanzata, dove fu invitato a sviluppare la sua idea di Intergalactic Computer Network, una rete alla quale si potesse collegare chiunque da qualsiasi parte del globo, avendo quindi accesso a dati e programmi che si trovavano fisicamente anche a migliaia di chilometri di distanza. Un'idea questa che venne posta alla base di ARPANET, antenato di quello che oggi chiamiamo Internet³.

¹ Leonard Kleinrock, *Communication Nets; Stochastic Message Flow and Delay* (New York: McGraw-Hill Book Company, 1964).

² Joseph Carl Robnett Licklider, «Man-Computer Symbiosis», *IRE Transactions on Human Factors in Electronics*, *IRE Transactions on Human Factors in Electronics*, HFE-1, n. 1 (marzo 1960).

³ Peter H Salus, *Castling the Net: From ARPANET to Internet and Beyond* (Reading, Mass.: Addison-Wesley, 1995).

Sotto gli auspici dell'ARPA, Larry Roberts e Thomas Marill crearono nel 1965 la prima connessione di rete tra due computer situati in centri di ricerca distanti tra loro: un modello TX-2 installato al MIT, a Boston, e un modello Q-32 presente nei laboratori dell'UCLA di Santa Monica (California). Il mezzo di connessione era poco efficiente, la linea telefonica dedicata costosa, e durante la trasmissione uno dei due calcolatori andò in blocco, ma alla fine la connessione funzionò.



L. Kleinrock e l'Interface Message Processor in una scena del documentario "Lo and Behold - Internet: il futuro è oggi" di Werner Herzog, 2016."

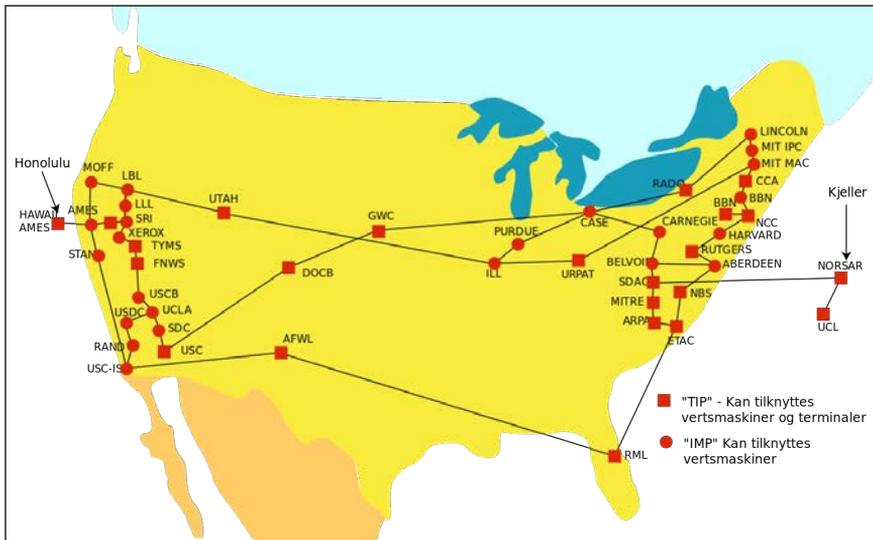
Quattordici anni dopo, nel 1989, ARPANET non esisteva più e aveva ceduto il posto ad altre reti a diversa estensione che erano diventate lo scheletro portante di quella "rete di reti" che oggi è Internet. Alla conferenza che ne celebrava i vent'anni dalla fondazione e allo stesso tempo la chiusura, Danny Cohen, uno dei suoi ingegneri più importanti, non senza una divertita ironia, ne riassunse la storia:

In the Beginning, ARPA created the ARPANET.
 And the ARPANET was without form and void.
 And darkness was upon the deep.
 And the spirit of ARPA moved upon the face of the network
 and ARPA said, 'Let there be a protocol,' and there was a protocol.
 And ARPA saw that it was good.
 And ARPA said, 'Let there be more protocols,' and it was so.
 And ARPA saw that it was good.
 And ARPA said, 'Let there be more networks,' and it was so.⁴

⁴ Matthew Lyon e Katie Hafner, *Where Wizards Stay Up Late: The Origins Of The*

Il passo più importante dello sviluppo di ARPANET, come si evince anche dalla citazione, è stata la messa a punto di protocolli condivisi: una rete virtualmente globale doveva essere in grado di collegare terminali con piattaforme e sistemi operativi diversi, che proprio per questa diversità di concezione e implementazione non possono interagire in maniera diretta. Il più famoso e diffuso dei protocolli condivisi è il TCP/IP (Transmission Control Protocol / Internet Protocol), che i singoli produttori di calcolatori devono adottare – e di fatto adottano – per mettere in connessione architetture e *hardware* differenti. Si tratta di un passaggio di fondamentale importanza, che svincola dalla struttura delle macchine il linguaggio che queste devono usare per comunicare tra loro, facendo sì che oggi possiamo continuare a utilizzare linguaggi (protocolli) definiti cinquant'anni fa con calcolatori di architettura più recente e notevolmente più evoluta.

La rapida e capillare diffusione delle reti in tutto il mondo aprì la strada a un altro evento che non sarebbe sbagliato definire epocale e che – caso raro – ebbe luogo in Europa per opera di uno scienziato inglese del CERN (Organizzazione europea per la ricerca nucleare) di Ginevra. Tim Berners-Lee scrive, nel marzo del 1989, una proposta sulla gestione e condivisione delle informazioni scientifiche tra colleghi di varie parti del mondo⁵, progetto che il suo superiore etichetta come «Vague, but exciting».



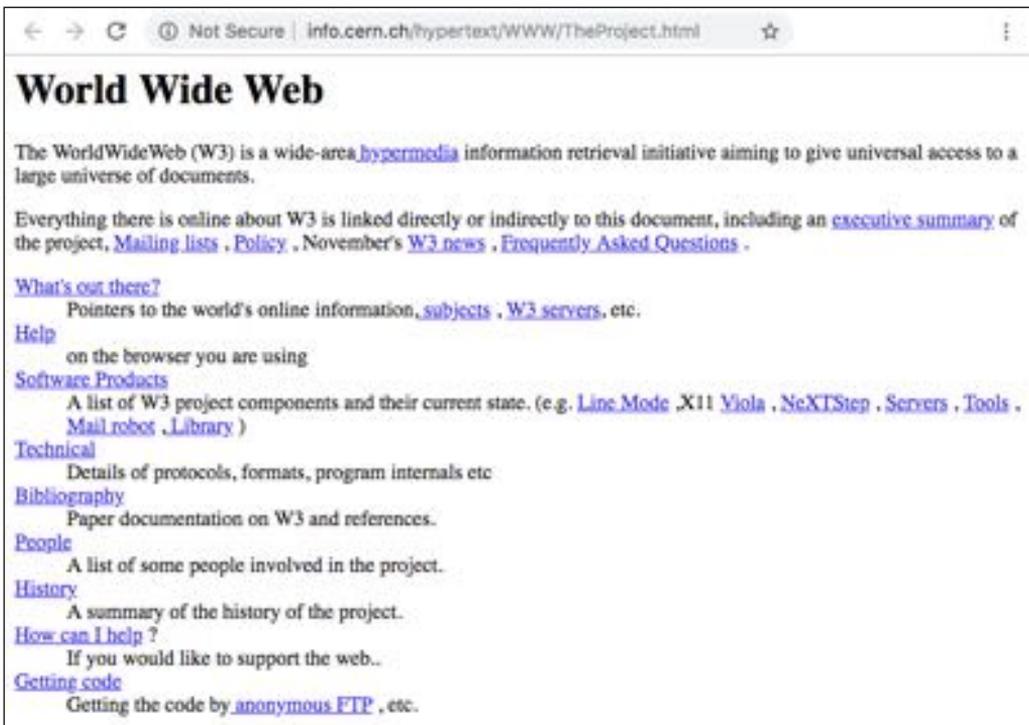
Rappresentazione semplificata di ARPANET, riferita al settembre del 1974 (Fonte: https://it.m.wikipedia.org/wiki/File:Arpanet_1974.svg, lic. Dominio Pubblico)

⁵ Tim Berners-Lee, «Information Management: A Proposal» (Geneva: CERN, 1989), <http://cds.cern.ch/record/369245/files/dd-89-001.pdf>.

L'idea che avrebbe cambiato Internet, introducendo il suo più famoso e più utilizzato servizio, appunto il World Wide Web (chiamato anche “www”, più semplicemente “Web” o in italiano “Rete”) venne quindi archiviata come una cosa interessante, ma un po' vaga. Nel documento, Berners-Lee teorizzava l'adozione del modello ipertestuale come principale struttura per la condivisione e la pubblicazione dei documenti, nonché lo sviluppo di protocolli e strumenti dedicati per la realizzazione materiale della struttura stessa.

Schermata del primo sito del World Wide Web, <http://info.cern.ch/hypertext/WWW/TheProject.html>

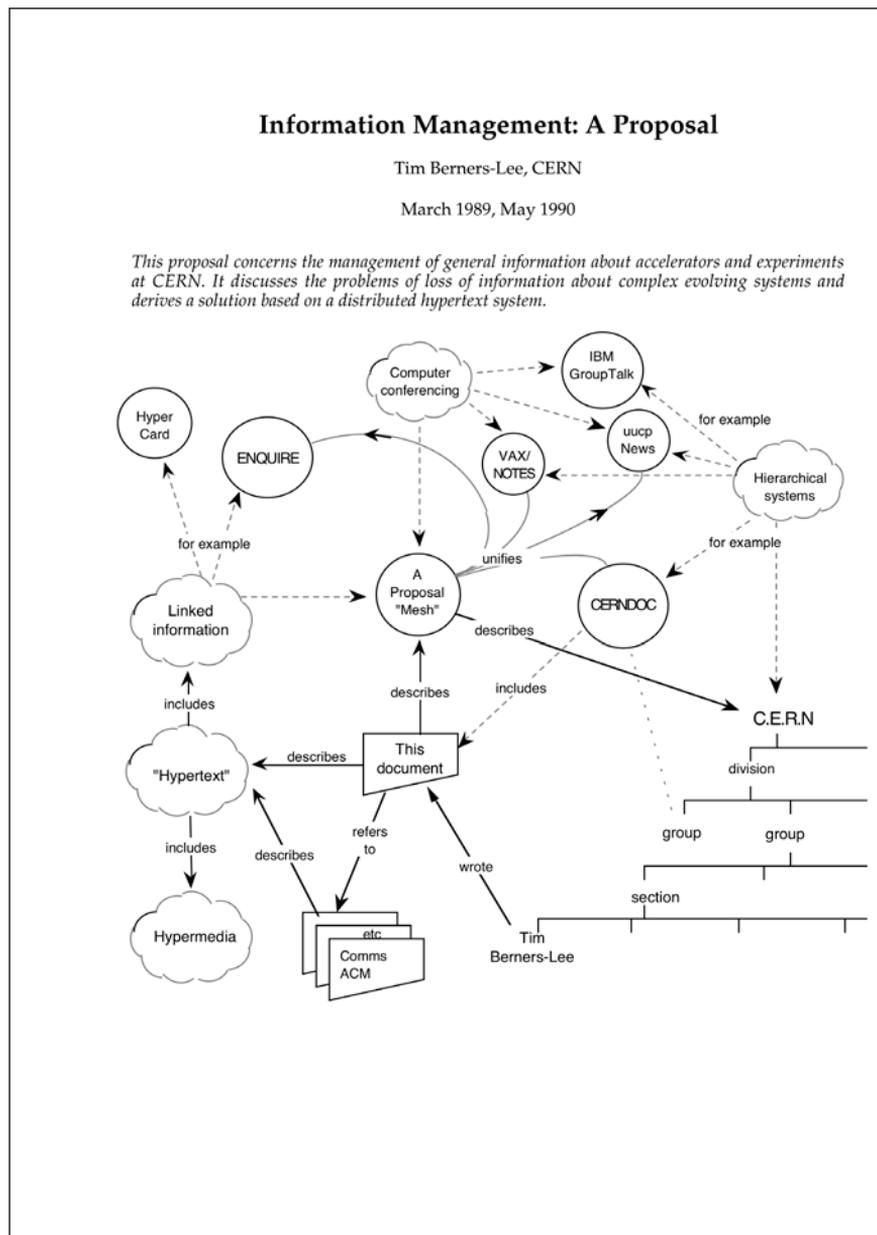
Tra il 1989 e il 1990 Berners-Lee scrisse il primo programma finalizzato alla pubblicazione di questi documenti, il primo *web server*, appunto, chiamato HTTPD, il primo *browser* (ed *editor*) chiamato WorldWideWeb, le prime specifiche di un linguaggio di *markup* che implementava il modello ipertestuale, chiamato HTML (HyperText Markup Language)⁶ e le basi dei protocolli URL (Uniform Resource Locator)⁷ e HTTP (HyperText Transfer



Internet (Simon and Schuster, 1999), 256.

⁶ La specifica più recente (versione 5.2) del HTML è disponibile all'indirizzo: <https://www.w3.org/TR/html52/>.

⁷ <https://www.w3.org/TR/url/>



Prima pagina della proposta di Tim Berners Lee del 1989, il documento che teorizza il World Wide Web

Protocol)⁸. Sono le fondamenta e i mattoni costitutivi del Web come lo conosciamo oggi. Il 6 agosto 1991 Berners-Lee pubblicò il primo sito web⁹ usando come *server* il suo calcolatore NeXT sul quale, con un pennarello rosso, aveva scritto «Questa macchina è un *server*. Non spegnere». Dal 2013, il CERN ha intrapreso un progetto per riportare *online*, usando gli stessi indirizzi web e gli stessi IP, il primo sito web, in una sorta di interessante museo digitale¹⁰.

Tim Berners-Lee lasciò poco dopo l'Europa per trasferirsi al MIT, dove fondò e tuttora dirige il World Wide Web Consortium (W3C)¹¹, un'organizzazione internazionale finalizzata a promuovere e sviluppare il Web e le tecnologie a esso connesse e a stabilire e aggiornare gli standard tecnici aperti utili al suo funzionamento, sia per quanto riguarda i linguaggi che per quanto concerne i protocolli.

Tralasciando per ora la parte più tecnica, si può affermare che la chiave del successo del Web, che nel giro di pochi anni si era esteso all'intera rete Internet diventandone lo strumento principale di comunicazione, sia nell'idea di fondo di apertura e libera condivisione e in una filosofia estrema-

Il calcolatore NeXT di Tim Berner Lee al CERN: il primo web server (fonte: https://commons.wikimedia.org/wiki/File:First_Web_Server.jpg).



⁸ <https://www.w3.org/Protocols/HTTP/>

⁹ <http://info.cern.ch/hypertext/WWW/TheProject.html>.

¹⁰ <https://first-website.web.cern.ch/>

¹¹ <https://www.w3.org/>.

mente pratica e democratica. Tim Berners-Lee¹² rilasciò tutti i *software* e i protocolli nel pubblico dominio¹³, perché pensava che nella collaborazione risiedesse l'unica vera chiave di successo. Il Web delle origini nacque su alcune idee fondamentali la cui forza ha oltrepassato i confini della tecnologia e ha raggiunto i più diversi ambiti sociali. Li riportiamo di seguito, in una versione sintetica¹⁴:

Decentralisation: No permission is needed from a central authority to post anything on the web, there is no central controlling node, and so no single point of failure ... and no “kill switch”! This also implies freedom from indiscriminate censorship and surveillance.

Decentralizzazione: non è necessario alcun permesso da nessuna autorità centrale per pubblicare qualcosa sul Web, non esiste un nodo centrale di controllo e quindi non esiste un unico punto debole ... e non esiste un “interruttore generale”! Questo implica anche la libertà da una censura indiscriminata e dalla sorveglianza.

Non-discrimination: If I pay to connect to the internet with a certain quality of service, and you pay to connect with that or a greater quality of service, then we can both communicate at the same level. This principle of equity is also known as Net Neutrality.

Non discriminazione: se io pago per essere connesso a Internet con una certa qualità di servizio e tu paghi per essere connesso con una qualità simile o superiore, possiamo comunque comunicare allo stesso livello. Questo principio di equità è conosciuto anche come “Neutralità della rete”.

¹² E soprattutto il CERN, che essendo il suo datore di lavoro, era il proprietario delle sue creazioni di quegli anni

¹³ Per lo statuto giuridico del “pubblico dominio” si veda il capitolo dedicato ai dati aperti e le licenze d’uso.

¹⁴ <https://webfoundation.org/about/vision/history-of-the-web/>

<p>Bottom-up design: Instead of code being written and controlled by a small group of experts, it was developed in full view of everyone, encouraging maximum participation and experimentation.</p>	<p>Progettazione dal basso: il codice non viene scritto e controllato da un gruppo ristretto di esperti, ma viene sviluppato sotto gli occhi di tutti, nel costante tentativo di incoraggiare la partecipazione e la sperimentazione.</p>
--	---

<p>Universality: For anyone to be able to publish anything on the web, all the computers involved have to speak the same languages to each other, no matter what different hardware people are using; where they live; or what cultural and political beliefs they have. In this way, the web breaks down silos while still allowing diversity to flourish.</p>	<p>Universalità: affinché chiunque sia in grado di pubblicare qualsiasi cosa sul Web, tutti i calcolatori coinvolti devono poter parlare la stessa lingua, indipendentemente dal diverso <i>hardware</i> che le singole persone utilizzano; indipendentemente da dove esse vivono, dalla loro cultura o fede politica. In questa maniera il Web abbatte muri e contemporaneamente stimola la diversità.</p>
---	---

<p>Consensus: For universal standards to work, everyone had to agree to use them. Tim and others achieved this consensus by giving everyone a say in creating the standards, through a transparent, participatory process at W3C.</p>	<p>Consenso: affinché uno standard universale possa funzionare, è necessario che tutti siano d'accordo a utilizzarlo. Tim (Berners-Lee) e gli altri hanno raggiunto questo consenso in seno al W3C, dando la possibilità a ciascuno di contribuire alla creazione degli standard, attraverso un processo partecipativo molto trasparente.</p>
---	---

Questi pochi punti riassumono l'idea che ha accompagnato la nascita della tecnologia web e chiariscono l'ambito sociale e culturale in cui essa è stata concepita e sviluppata. L'efficacia della filosofia del W3C ha probabilmente contribuito, più di ogni vantaggio o merito tecnico, alla rapidissima diffusione e al grandissimo successo di questo strumento. Molti di questi punti hanno avuto conseguenze assai importanti nell'implementazione del-

la Rete, benché tuttora esistano organizzazioni, paesi o gruppi di paesi che pongono ancora oggi veti e limitazioni di vario genere. A trent'anni dalla nascita, con metà della popolazione mondiale connessa, il World Wide Web è un luogo molto diverso da quello delle origini, ma quello che non è cambiato è lo spirito dei suoi fondatori e del W3C, che continuano a lavorare nel rispetto dei principi fondativi¹⁵ e mantenendo alto l'impegno a incidere positivamente sulla società. Ne è segno tangibile la recente iniziativa del cosiddetto "Contratto per il Web", una carta sui diritti e doveri compilata sulla falsariga della "Dichiarazione universale dei diritti umani", del "Diritto del Mare" o del "Trattato sullo spazio extra-atmosferico": un documento da scrivere in maniera collaborativa e da far approvare a governi e organizzazioni internazionali, per offrire una base giuridica unitaria che detti comuni linee guida legislative¹⁶.

In Italia, la ricerca archeologica ha precocemente mostrato un grande interesse per questo strumento di divulgazione e pubblicazione, come dimostrano incontri e convegni organizzati già nel corso degli anni Novanta. In occasione di uno dei primi, tenutosi nel 1995 nella Certosa di Pontecagnano (Siena) ed edito nel 1997¹⁷, si elencavano le varie iniziative avviate sull'onda dell'iniziale e comprensibile entusiasmo per lo strumento digitale. La sperimentazione a tutti i livelli – uno dei pilastri, come si è visto, della filosofia del Web – ha guidato gran parte di questa prima fase della ricerca, che ha coinvolto per lo più gli archeologi del comparto museale, i più interessati a esplorare nuove forme multimediali di narrazione. Meno partecipativo è stato il mondo dell'editoria archeologica, rimasto, come vedremo, a lungo legato a processi tradizionali di produzione e distribuzione editoriale, basati esclusivamente sul supporto cartaceo.

All'estero, invece, esperienze importanti come quella della rivista "Internet Archaeology"¹⁸ hanno segnato un'anticipazione davvero significativa di tendenze e pratiche che solo ora stanno trovando un'applicazione su larga scala, come ad esempio il concetto del libero accesso ai contenuti disponibili in Rete (in inglese *open access*). In Italia, "Archeologia e Calcolatori", la rivista fondata nel 1990 da Mauro Cristofani e Riccardo Francovich, è la sede editoriale che si è fatta perno della riflessione e del

¹⁵ <https://webfoundation.org/2019/03/web-birthday-30/>

¹⁶ <https://contractfortheweb.org/>

¹⁷ Antonio Gottarelli, a c. di, *Sistemi informativi e reti geografiche in archeologia: GIS-INTERNET. VII Ciclo di Lezioni sulla Ricerca applicata in Archeologia (Certosa di Pontignano 1995)* (Firenze: All'Insegna del Giglio, 1997)..

¹⁸ <http://intarch.ac.uk/>

dibattito sui temi delle metodologie, delle teorie e delle applicazioni della computazione automatica all'ambito archeologico¹⁹. Dal 2005 la rivista è disponibile, anche per i numeri precedenti, in *open access*.

Il Web come deposito pubblico di dati

La Rete, per com'è stata concepita e sviluppata, è stata a lungo un contenitore per la pubblicazione di contenuti statici. I documenti, scritti come ipertesto in formato HTML, eventualmente collegati a elementi grafici, venivano salvati in macchine sempre accese e sempre connesse, che li “servivano” a chi – macchina o individuo – ne avesse fatto richiesta. Da qui il nome di “servente”, detto più comunemente *server*, in riferimento alla macchina in cui risiedono i documenti.

Scrivere un documento in formato HTML – il formato per eccellenza del WWW – richiede una competenza informatica piuttosto bassa, resa ancora più semplice dalla presenza di *software* di compilazione che ne permettono la creazione e modifica, con l'aiuto di interfacce grafiche, anche da parte di chi non conosce le specifiche del linguaggio di marcatura. Inoltre i *browser*, che permettono la richiesta e la visualizzazione dei documenti HTML, dispongono tradizionalmente di un ottimo grado di adattamento, che consente di visualizzare correttamente anche documenti che presentano importanti errori di sintassi. Si potrebbe affermare che il successo dell'HTML – e quindi della Rete – sia andato in parallelo a quello del *globish*, la versione semplificata internazionale dell'inglese, che proprio la Rete ha contribuito a generare. Pur essendo entrambi per definizione linguaggi fortemente limitati nella grammatica e nel vocabolario, e spesso usati con errori sintattici rilevanti, grazie all'immediatezza espressiva e alla forza comunicativa sono riusciti a imporsi a livello planetario. Come può succedere in un convegno internazionale che nessuno tra relatori parli un inglese perfetto, bensì un ottimo *globish*, magari infarcito di termini specialistici, così accade che gran parte dei siti web pubblicati in Rete contenga codice HTML non valido, ma comunque leggibile. L'idea di fondo è che costruire una cosa imperfetta ma perfettibile è sempre e comunque meglio che seguire un ideale di perfezione difficilmente realizzabile o addirittura utopico.

La staticità del Web è un'idea che appartiene al passato, ma che ha grandemente segnato la storia del suo utilizzo in questi trent'anni. La neces-

¹⁹ <http://www.archcalc.cnr.it>, Mario Cristofani e Riccardo Francovich, «Editoriale», *Archeologia e Calcolatori* 1 (1990): 1–7.

sità primaria del CERN era quella di diffondere e condividere documenti scientifici per mezzo della ormai consolidata rete Internet, ed è questo che il Web è stato a lungo: un deposito dove chiunque, con minime conoscenze informatiche, fosse in grado di pubblicare liberamente contenuti, fossero essi un report scientifico o il diario delle vacanze. Per questo motivo – nel campo dell’archeologia, come già osservato – è stato il settore museale a essere stato attratto per primo e in misura maggiore dalla comunicazione web, che per definizione era in grado di raggiungere il grande pubblico. Meno interesse ha destato tra gli archeologi che si occupavano principalmente della raccolta dei dati, del loro immagazzinamento e della loro analisi, ambiti che hanno sempre avuto un precoce contatto con il mondo delle tecnologie e che ne hanno assorbito con grandissima facilità ed entusiasmo le novità, ma ai quali il Web dei primi anni aveva ancora sostanzialmente poco da offrire.

L’editoria scientifica, settore in cui ci si sarebbe aspettata una più naturale propensione all’adozione di tecnologie web, merita un discorso a parte. Il mondo editoriale infatti, soprattutto quello legato alla produzione accademica, ha avuto una reazione di estremo conservatorismo e di tenace resistenza ai nuovi strumenti di comunicazione, giudicati, talvolta a ragione, troppo “leggeri” e informali, più adatti alla divulgazione che a diventare sede stabile di prestigiose pubblicazioni.

I due fattori sopra esposti hanno chiuso la ricerca archeologica – e più in generale quella umanistica – in una morsa che ha bloccato a lungo sviluppi importanti. Chi aveva bisogno di strumenti dinamici che lo potessero aiutare nella raccolta dei dati o di una potenza di calcolo indispensabile all’analisi, non poteva trovare nel Web risorse utili. Chi per contro si occupava di pubblicare e divulgare non ha ritenuto interessante investire in nuove forme di edizione, perché non particolarmente redditizie e perché presentavano, come tutte le novità, una serie di problematiche di metodo che avevano bisogno di essere indagate, affrontate e quindi risolte prima di entrare nello stato dell’arte. Problemi legati al fattore estetico, alla durevolezza dei prodotti digitali *online*, al loro accesso e alla monetizzazione della loro produzione e distribuzione hanno fatto sì che, tranne pochi illuminati esempi di cui si è già parlato, l’editoria guardasse alla Rete con estrema diffidenza.

Un bell’esempio di uso pionieristico del Web per lo studio dell’archeologia e della storia antica è la “Rassegna degli Strumenti Informatici per lo Studio dell’Antichità Classica”, creata nel 1995 dall’Università di Bologna²⁰,

²⁰ Alessandro Cristofori, Giovanni Geraci, e Carla Salvaterra, «Rassegna Strumenti

la quale «si propone di offrire una guida ai nuovi strumenti a disposizione di studenti e studiosi del mondo classico, con brevi descrizioni e valutazioni e accesso diretto alle risorse disponibili *on line*». Le pagine di indice, accuratamente organizzate per argomento e per area geografica, sono tuttora mantenute *online*, anche se gran parte dei collegamenti risultano ormai orfani, visto che l'aggiornamento attivo della Rassegna è stato interrotto nel 2007.

Pagina iniziale della Rassegna degli Strumenti Informatici per lo Studio dell'Antichità Classica

Una seconda risorsa, questa volta internazionale, rimasta a lungo di grande interesse è stata ArchNet²¹, un portale che si proponeva di diventare una biblioteca virtuale sul Web dedicata all'archeologia. L'ultima versione del sito, datata 7 gennaio 2015 è oggi disponibile su Archive.org²².

Nel corso del primo decennio del XXI secolo l'entusiasmo per l'accesso libero e gratuito al Web ha portato singoli studiosi e appassionati a costruire

Rassegna degli Strumenti Informatici per lo Studio dell'Antichità Classica

fondata da Alessandro Cristofori, Giovanni Geraci e Carla Salvaterra

a cura di Alessandro Cristofori, con la collaborazione di Alice Benicivenni, Enrica Fontani, Davide Faoro, Roberta Mazza, Marina Milella, Paolo Monella e Carla Salvaterra

sotto l'Alto Patrocinio dell'Istituto Italiano per la Storia Antica

Ultimo aggiornamento: 19 gennaio 2007



Coppa attica a figure rosse
Kunsthistorisches Museum,
Vienna

L'UNIVERSITÀ DELLA CALABRIA

Dall'aprile 2005 la Rassegna è aggiornata grazie al sostegno finanziario del progetto Ex-MURST 60% su "Strumenti digitali per la ricerca e la didattica in Storia Antica, tra applicazioni pratiche e riflessioni metodologiche", stanziato dalla nuova sede del curatore, l'Università degli Studi della Calabria, con il supporto delle attrezzature del Laboratorio Multimediale per le Fonti Storiche attivo presso il Dipartimento di Storia di UniCal.



A partire dal mese di novembre 2004 la Rassegna è parte integrante della World Wide Web Virtual Library History Network (vedi WWW.VL.History.Central.Catalogue), a cura di Serge Noiret e Inaki Lopez Martin presso l'Istituto Universitario Europeo di Firenze.

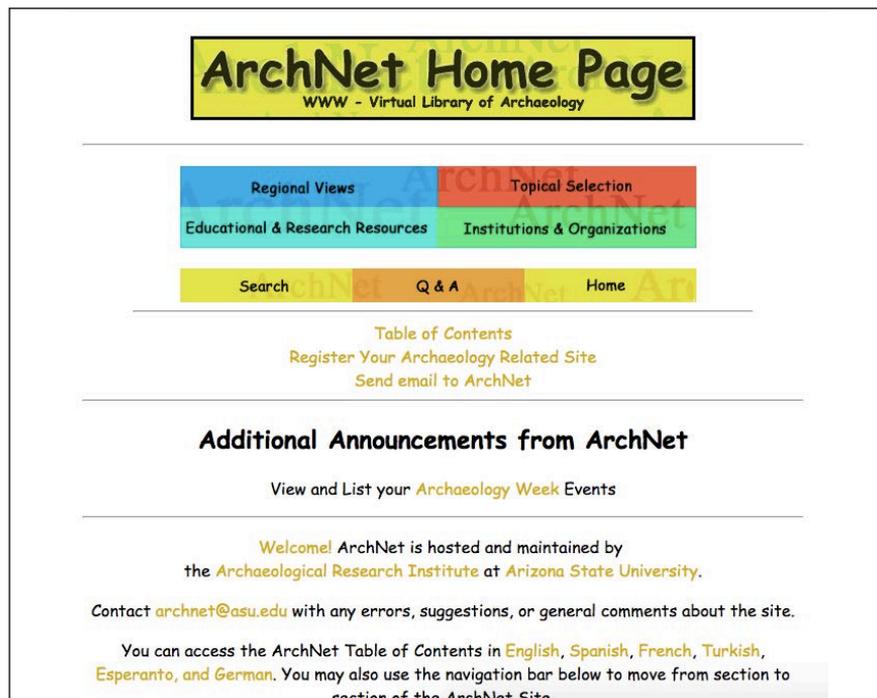


La Rassegna è anche un sito associato di [La Storia - Consorzio italiano per le discipline storiche online](http://www.la-storia.it).

informatici per l'Antichità Classica», 1998, <http://www.rassegna.unibo.it/>.

²¹ Department of Anthropology University of Connecticut e Archaeological Research Institute Arizona State University, *ArchNet: WWW virtual library of archaeology* (Tempe, Ariz.: Archaeological Research Institute, 1993), <https://web.archive.org/web/20150107082249/http://ari.asu.edu/archnet/>. Il portale ha subito vari spostamenti e oggi non è più disponibile.

²² Archive.org è una importantissima iniziativa mirata a costruire una biblioteca digitale di siti web e altre risorse culturali digitali, pubblicati in forma libera. Contiene centinaia di miliardi di pagine, milioni di libri e testi, video, audio, immagini, programmi, ecc. Parte di questa iniziativa è la cosiddetta *Wayback machine*, l'archivio del World Wide Web sotto forma di istantanee (*snapshot*) scattate a intervalli periodici. Su questo archivio è quindi possibile recuperare e consultare pagine web ormai scomparse o non più accessibili.



Il portale di ArchNet: WWW virtual library of archaeology, come compariva nel 7 gennaio 2015 (fonte: <https://web.archive.org/web/20150107082249/http://ari.asu.edu/archnet/>)

propri siti web su cui pubblicare immagini, informazioni o contributi originali su vari ambiti culturali e cronologici, determinando una rilevantissima proliferazione di contenuti di diversa natura. La comunità scientifica – e in primo luogo quella accademica – si è fatta carico di creare filtri, mettere ordine, compilare indici tematici e geografici e valutare le singole iniziative. Un compito difficile e dispendioso a livello di energie – considerati gli elevatissimi tassi di crescita quotidiani del Web – ma che ha rivestito un ruolo fondamentale per l’orientamento degli studenti in primo luogo, ma anche di un pubblico più vasto che volesse districarsi in un mondo sempre più labirintico, soprattutto per il diverso livello qualitativo dei contenuti. In questo, la comunità scientifica ha replicato un modello molto diffuso negli anni ‘90 del secolo scorso, quando il moltiplicarsi delle pagine web stimolò la nascita di indici tematici, compilati manualmente, che in breve tempo sarebbero diventati le “porte” del Web. È interessante a questo riguardo il caso dello sviluppo di Yahoo!, nato come indice personale di pagine web di due dottorandi della Stanford University e cresciuto poi a dismisura, fino a diventare un indice globale organizzato per argomenti e

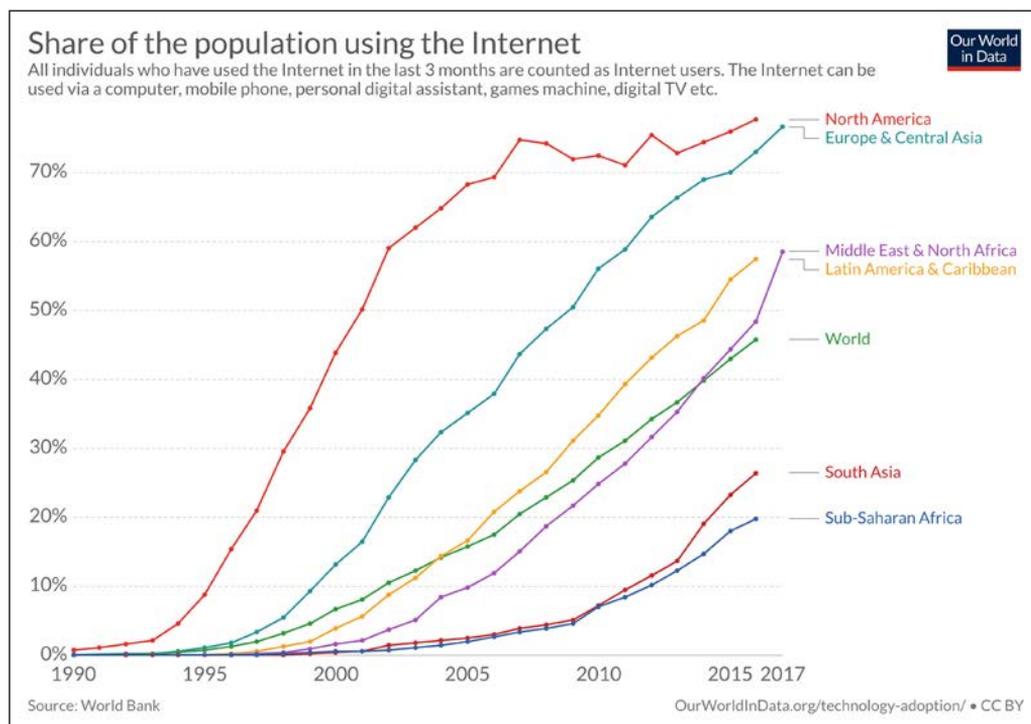
sottoargomenti, compilato manualmente da una fitta schiera di dipendenti e collaboratori dell'azienda e da un esercito ancor più numeroso di utenti privati, che segnalavano su base volontaria siti e risorse²³.

Quello dell'indice è però un modello destinato a scomparire in brevissimo tempo: il Web è un luogo estremamente vivo e dinamico, un "paese" con un tasso di nascite altissimo e in continuo aumento, ma con un altrettanto alto tasso di decessi²⁴. Tutto succede e cambia a grande velocità. Il frutto del lavoro di anni e di tante persone, come può essere un indice tematico, nato da una forte necessità di orientamento, diventa nel giro di pochi mesi una sorta di fossile, utile per capire il percorso fino a quel momento intrapreso dalla disciplina e la storia degli studi, ma non più come una guida per studenti e studiosi alla ricerca di bibliografia, risorse, confronti e spunti di ricerca. Il dinamismo portato alle estreme conseguenze è tuttavia una tendenza che non è possibile invertire e contrastare, una parte integrante della natura stessa dello strumento che, come tale, va tenuta in considerazione e analizzata, anche se si pone in netto contrasto con alcuni principi fondanti della ricerca scientifica, come la stabilità e la costante citabilità della letteratura di riferimento, nonché con la stessa base ontologica dell'archeologia, visto che, nella maggior parte dei casi, le risorse digitali scompaiono senza apparentemente lasciare alcuna traccia di sé.

La ricerca scientifica ha infatti bisogno di stabilità e durevolezza, due principi che il Web, per come è nato e si è sviluppato, non può garantire, soprattutto a causa della poca consapevolezza degli attori iniziali, non solo utenti privati ma anche – e soprattutto – enti e istituzioni pubbliche. Dal punto di vista archeologico, invece, l'apparente immaterialità del mondo digitale pone importanti sfide teoriche sul nostro presente e sulla capacità futura di poterlo leggere. È questo un punto importante, sul quale torneremo più volte.

²³ I contributi volontari erano stimolati dalla crescita stessa della piattaforma: non esservi indicizzati significava scomparire dalla mappa della rete – e conseguentemente dalla possibilità di essere raggiunti. Gli utenti erano quindi fortemente motivati a segnalare e tenere aggiornata la propria presenza.

²⁴ Interessante a questo proposito risulta uno studio condotto sul destino di 88 siti web di contenuto archeologico create sulla piattaforma Geocities di Yahoo!, "spenta" nel 2009 (Matt Law e Morgan Colleen, «The archaeology of digital abandonment: online sustainability and archaeological sites», *Present Pasts* 6, n. 1 (2014): Art.2, <https://doi.org/10.5334/pp.58>). Le considerazioni su benefici e rischi dell'utilizzo delle piattaforme web commerciali e l'esame dell'esperienza passata possono aiutare a meglio inquadrare la presenza attuale dell'archeologia sul web e il suo probabile destino futuro.



Tasso di crescita del World Wide Web, per numero totale di utenti, in varie parti del mondo (fonte dati World Bank – World Development Indicators, <http://data.worldbank.org/data-catalog/world-development-indicators>, elaborazione da <https://ourworldindata.org/internet>).

Le banche dati sul Web: il Web dinamico

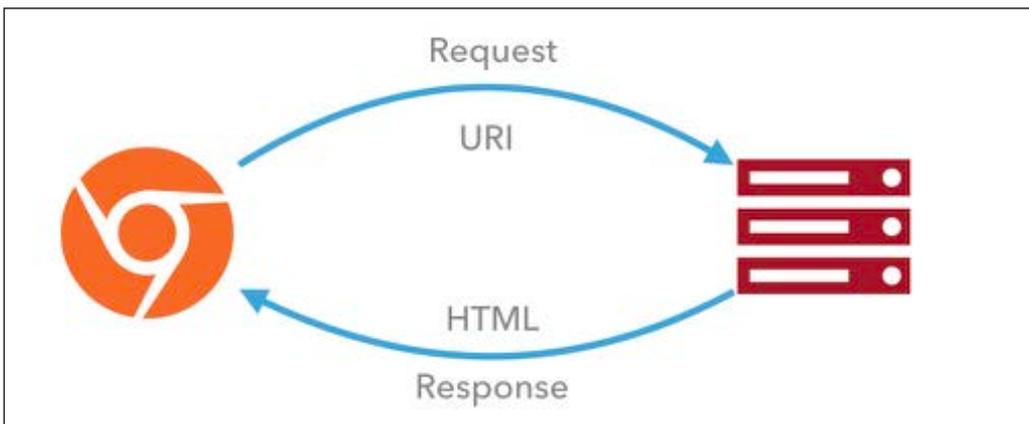
Nel Web cosiddetto statico, a ogni richiesta di una persona fisica eseguita attraverso un *browser* installato su un elaboratore (*client*) risponde un *software server* che, come un diligente bibliotecario, elabora la richiesta, recupera il documento in questione e lo invia al richiedente. Tutti i documenti sono scritti in un unico linguaggio, l'HTML, e tutte le richieste seguono lo stesso modello di indirizzario, lo Uniform Resource Locator (URL).

Il modello, inventato da Tim Berners Lee trent'anni fa, è stato aggiornato e migliorato, ma nella sostanza è rimasto invariato. Lo stesso non si può dire del Web. Sotto questa struttura semplice e basilare, che garantisce il funzionamento della comunicazione, è infatti attecchito e proliferato uno sviluppo tecnologico senza precedenti. Il primo considerevole cambiamento è stato innescato dall'esigenza di connettere alla Rete banche dati che agenzie e istituti di ricerca avevano raccolto e strutturato nei decenni

precedenti. È nato così il cosiddetto Web “dinamico”, che rende possibile la pubblicazione e l’interazione in tempo reale di questi *database*, in genere molto più articolati di una serie di documenti HTML preconfezionati. Dal punto di vista tecnico, il cambiamento è molto semplice da spiegare: il modello di richiesta e risposta (protocollo HTTP) non subisce sostanziali variazioni. Il *browser* continua a fare una richiesta secondo lo standard URL e il *server* continua a restituire un documento HTML. La differenza è che il documento restituito e visualizzato sullo schermo non è stato scritto da una persona, ma viene compilato “al volo” (ing. *on the fly*) a seguito di ogni richiesta, da un programma che fa uso di molte fonti, per lo più organizzate in banche dati relazionali. I dati strutturati vengono quindi separati dalla loro presentazione e vengono a trovarsi in una posizione centrale e privilegiata. “Data is the Next Intel Inside” è il nome di un capitolo del famoso libro di Tim O’Reilly²⁵ incentrato sulla definizione del cosiddetto Web 2.0, che nel 2009 portava all’attenzione di tutti quello che ormai, a dieci anni di distanza, è evidente: i dati, il loro possesso, la loro pubblicazione e distribuzione sono la vera sfida del presente.

Per rendere possibile questo processo viene usato uno standard chiamato CGI (Common Gateway Interface), sviluppato inizialmente per la gestione delle liste di distribuzione (*mailing list*), e successivamente utilizzato in gran parte delle istanze del Web, per eseguire sul *server* il programma che si occupa dell’interrogazione dei contenuti della banca dati, della loro automatica impaginazione e quindi della confezione del documento HTML da restituire all’utente. Questo tipo di programmi può eseguire programmi molto semplici, come comporre un documento mettendo insieme testi che, per esempio, corrispondono a una data ricerca, sia operazioni più impe-

Esemplificazione del protocollo http e del modello di comunicazione client-server

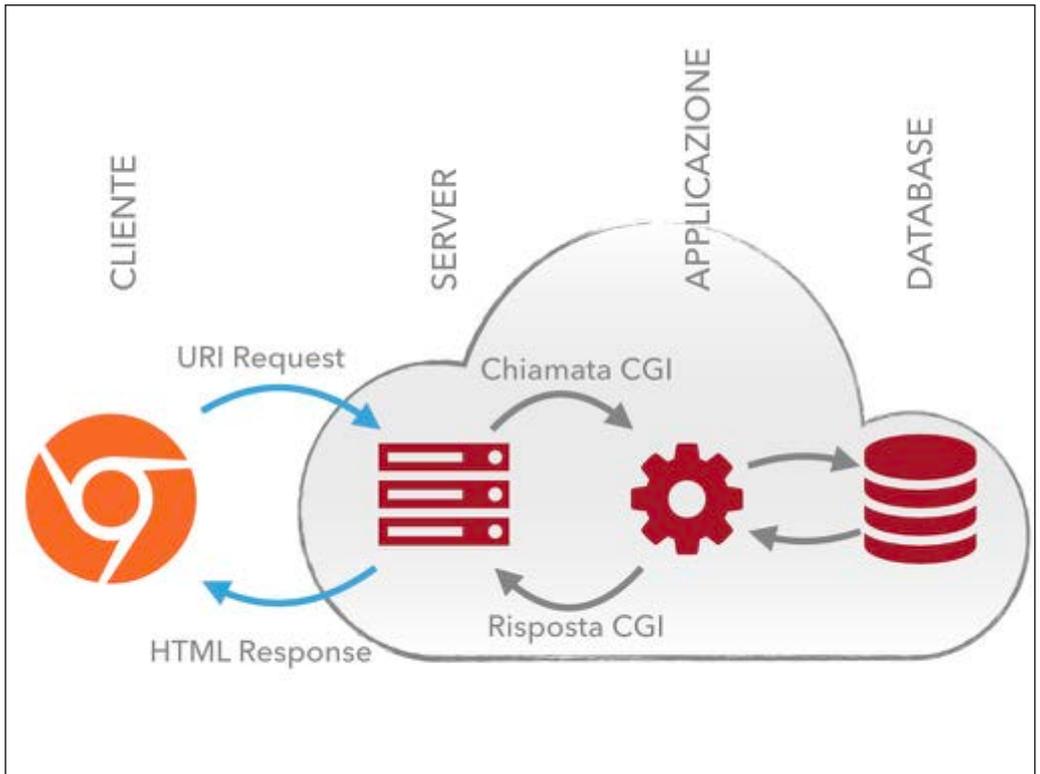


²⁵ Tim O’Reilly, *What Is Web 2.0* (Sebastopol: O’Reilly Media, 2009).

gnative, come lanciare interrogazioni complesse su banche dati, inserire nuovi dati, cancellare, fare analisi di vario genere e restituire i risultati a video. Questa semplice tecnologia ha potuto portare sul Web la potenza computazionale dei calcolatori, sui quali l'archeologia e altre discipline umanistiche facevano già da decenni sperimentazione. Da deposito statico di documenti semplici collegati tra loro da *link* (ipertesto), il Web è diventato uno spazio attivo, in cui è possibile programmare e creare. Si sono moltiplicati dunque i linguaggi di programmazione, i *framework* di sviluppo e le banche dati.

Per gestire i contenuti sul Web sono disponibili diversi sistemi, veri e propri programmi chiamati CMS (Content Management System, gestori di contenuti) che permettono anche a utenti non particolarmente esperti di informatica di creare siti e portali web di grande complessità e qualità, tecnica e grafica. Oggi il mercato offre una vasta scelta di questi strumenti, adatti sia a creare piccoli siti personali, piattaforme editoriali o di vendita, sia siti molto complessi e ricchi di contenuti, come i siti istituzionali e di servizi. Wordpress²⁶ è sicuramente il più celebre tra tantissimi, mentre lo

Modello client-server a protocollo CGI per collegare applicazioni esterne e banche dati



²⁶ <https://wordpress.com/>.

seguono per fama e diffusione: Drupal²⁷, Joomla²⁸, Mediawiki²⁹, Blogger³⁰, Google Sites³¹, ecc. Alcuni sono progettati per un uso generico, altri sono specializzati in particolari ambiti, come per esempio l'*e-commerce* (v. es. Magento³², Prestashop³³ o Shopify³⁴). Molti professionisti, inoltre, ne creano continuamente di personali o aziendali, tarati su esigenze talvolta molto precise e puntuali. Si tratta di un settore che negli anni futuri è destinato a crescere ulteriormente, in parallelo all'avanzamento tecnologico e al numero delle persone connesse alla rete.

Se il corredo tecnologica che sorregge il Web – HTTP, URL e HTML – non ha subito radicali cambiamenti negli ultimi 30 anni, ma solo importanti aggiornamenti, e soprattutto non ha visto la nascita o l'affermazione di progetti alternativi, rimanendo *de facto* l'unica via possibile per implementare il World Wide Web, i linguaggi di programmazione, i *database*, i *framework*, le piattaforme e le tecnologie che più in generale servono il Web hanno al contrario conosciuto uno sviluppo vertiginoso e fulmineo. Non ci sono regole prestabilite: se una tecnologia è veloce, sicura, facile da implementare, o più semplicemente alla moda e popolare, sicuramente verrà usata da un numero importante di progetti per un dato periodo. Linguaggi di programmazione disdegnati da informatici di professione, come il PHP³⁵, pur non essendo né perfetti né tra i più sicuri, hanno conosciuto una diffusione enorme grazie alla facilità di utilizzo e al successo di alcune piattaforme che li hanno impiegati. Come accade che il successo e la larga diffusione di un linguaggio favorisca la formazione di “dialetti”, più o meno distanti dall'originale, che risolvano problemi specifici o necessità particolari. Così dal PHP, il linguaggio di programmazione più usato e diffuso sul Web, impiegato da Google.com, Facebook.com, Yahoo!, Wordpress.org, Wikipedia (per citare solo i più celebri), deriva, grazie a un colosso quale

²⁷ <https://www.drupal.org/>

²⁸ <https://www.joomla.org/>

²⁹ <https://www.mediawiki.org/>

³⁰ <https://www.blogger.com/>

³¹ <https://sites.google.com/>

³² <https://magento.com/>

³³ <https://www.prestashop.com/>

³⁴ <https://shopify.com/>

³⁵ <https://www.php.net/>

Facebook.com, un “dialetto” (Hack³⁶), più veloce, efficiente e preciso, finalizzato a un uso interno del *social media* e rilasciato come progetto *open source* per incentivarne l’uso. Ma non sono solo i grossi nomi a lanciare le novità. Il *topos* mediatico dello studente di Stanford, Harvard o simili, che per noia o divertimento inventa uno strumento – un indice (Yahoo!), un algoritmo (Google), un portale (Facebook) o una piattaforma di sviluppo (Node.js) – che cambia radicalmente il modo di concepire e sviluppare il Web, con una serie di implicazioni sociali e culturali relevantissime, non fa altro che replicare se stesso a intervalli regolari.

Più stabile risulta essere il mondo dei *software* relativi alle banche dati, visto che la stragrande maggioranza di siti web e di applicazioni *online* di piccola e media grandezza fanno ampio utilizzo del modello relazionale e di *software* come MySQL, o del suo ramo di sviluppo parallelo MariaDB³⁷. Un’altra opzione *open source* è costituita da PostgreSQL, anch’esso *database* di tipo relazionale, ma maggiormente specializzato nel campo della gestione di dati geografici e applicazioni GIS, per i quali di fatto costituisce lo standard. Pur non mancando altre numerose alternative, dalla fine degli anni dieci del XXI secolo solo una piccola percentuale di siti e applicazioni web non utilizza un *database* MySQL/MariaDB o PostgreSQL per organizzare i propri dati.

Anche il mondo delle scienze umanistiche – e in primo luogo l’archeologia – trova questo nuovo Web molto più interessante per le sue possibilità di immagazzinare, analizzare e gestire in maniera semplice e intuitiva dati complessi, secondo il modello relazionale. In fondo, sono attività che costituiscono una parte fondamentale della ricerca archeologica sul campo e in laboratorio. Nel primo decennio degli anni 2000 non si registra tuttavia

³⁶ <https://hacklang.org/>.

³⁷ Fino al 2009 MySQL (<https://www.mysql.com/>) era la soluzione *open source* più adottata nel Web e non solo. In quell’anno la Sun Microsystems acquistò la società che gestiva MySQL e di conseguenza diversi collaboratori, tra i quali il fondatore e principale sviluppatore, lasciarono il progetto. Il timore era che la Sun Microsystems ne avrebbe ostacolato lo sviluppo, anche perché l’azienda è controllata da Oracle (<https://www.oracle.com/>), produttore dell’omonimo *software* per banche dati, principale concorrente di MySQL. È nato in questo modo un progetto parallelo (*fork*) chiamato MariaDB (<https://mariadb.org/>). Le rassicurazioni della Sun e la donazione dell’ormai vuoto progetto alla Apache Foundation (un importante incubatore per progetti *open source*) non hanno ricucito la frattura e MariaDB è oggi il *database* predefinito delle varie distribuzioni GNU/Linux. Si tratta di un fenomeno ricorrente nel mondo dell’industria *software*: un altro esempio famoso è quello della divisione tra OpenOffice.org e LibreOffice, sempre in conseguenza di azioni di mercato che vedevano coinvolto Sun Microsystems.

in Italia una particolare spinta verso la migrazione in Rete degli archivi digitali: gran parte degli archeologi sceglie di continuare a implementare i propri *database* di ricerca sui tradizionali sistemi *desktop*, ancora di gran lunga più comodi, pratici e intuitivi. L'istanza di portare in Rete i dati è un'esigenza molto recente, che si è manifestata sostanzialmente nel corso dell'ultimo decennio³⁸, e che nel caso specifico dell'Italia, è stata ostacolata dalla frammentarietà dell'infrastruttura della rete fisica, dagli alti costi di connessione e dalla ristrettezza della capacità di banda. Molti dei siti in cui si svolgono le ricerche si trovano in luoghi isolati, lontani dalle grandi infrastrutture, privi di un accesso immediato e diretto alla rete, e solo negli ultimi dieci anni è stato possibile disporre di una buona copertura di rete sull'intero territorio nazionale.

Ritorno al futuro. Il Web semantico e le sue applicazioni in archeologia

Il titolo di questa sezione è chiaramente un paradosso, ma pone l'accento sul fatto che quello che il Web del futuro *può* o *deve* diventare è già stato predetto nei primi anni Novanta del secolo scorso, ed è di nuovo frutto della mente di Tim Berners-Lee. Nella prima International World Wide Web Conference tenutasi al CERN di Ginevra nel 1994, la conferenza in cui fu annunciata la creazione del W3C, Berners-Lee constata come il Web dell'ipertesto (il Web dei documenti), benché diventato un luogo estremamente interessante per le persone, sia tutto sommato uno spazio piatto e poco significativo per le macchine, che, pur contenendo tutti i documenti, non sono in grado di comprenderne il significato profondo (semantica) e di conseguenza non sono capaci di farne un uso intelligente³⁹.

Perché si abbia una Rete meno piatta e più ricca di significato – il cosiddetto Web “semantico” – è necessario rivestire i documenti di nuovi “strati” informativi, descrizioni e metadati, leggibili ed elaborabili in maniera automatica dalle macchine, che potranno così “comprendere” i contenuti

³⁸ Gian Pietro Brogiolo, nel 2012, vedeva nella pubblicazione in liberi archivi online dei dati archeologici una delle vie per la “sopravvivenza” dell'archeologia (Gian Pietro Brogiolo, «Archeologia pubblica in Italia: quale futuro», *Post-Classical Archaeologies* 2 (2012): 269–78). Qualche anno prima, infatti, aveva segnalato come il 90% degli scavi d'emergenza fosse destinato a rimanere inedito, per la mancanza di una filiera che legasse in maniera stabile il processo di produzione con quello di pubblicazione e diffusione dei dati, parlando di una “archeologia senza risultati” (Gian Pietro Brogiolo, «Archeologia e istituzioni: statalismo o policentrismo?», *Archeologia medievale*, n. 24 (1997): 7–30).

³⁹ <https://www.w3.org/Talks/WWW94Tim/>.

e rispondere a ricerche intelligenti e non unicamente basate sul confronto di stringhe di testo.

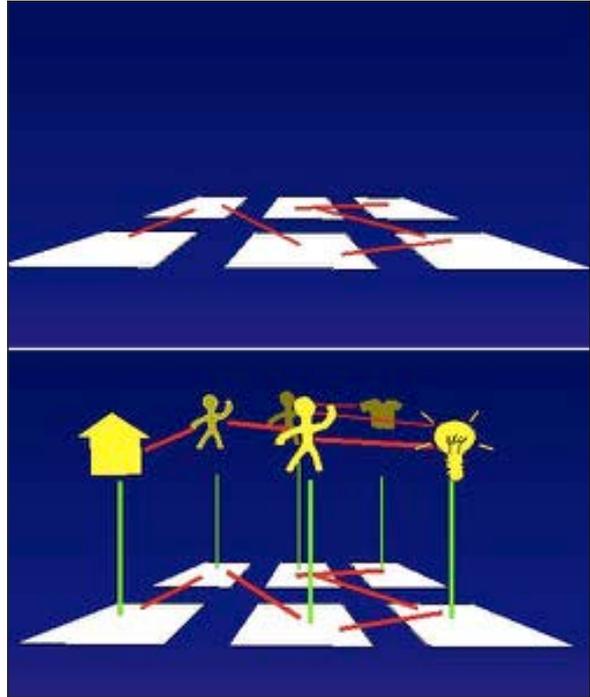
Un contratto di vendita e una scheda di catalogo di un reperto archeologico sono, per un *server web*, due documenti perfettamente comparabili, fatti di testo ed eventualmente immagini; se però ciascuno dei documenti venisse contrassegnato con “etichette” che ne esplicitassero l’appartenenza a una categoria specifica (“documento legale” o “reperto archeologico”, per esempio) – metadati appunto –, per la macchina sarebbe possibile distinguerli e catalogarli in maniera automatica. La creazione di metadati, dunque, è il primo grande passo da fare per arricchire di significato i documenti pubblicati sul Web.

Il Web non si compone infatti solo di documenti, ma anche di collegamenti tra questi: la struttura ipertestuale permette di stabilire le connessioni necessarie alla costruzione della rete, anche se nel Web attuale queste connessioni non sono “qualificate”, non vengono cioè definite per il loro significato.

Ad esempio, la pagina (il documento HTML) relativa al concetto di “archeologia” nella versione in italiano di Wikipedia, raggiungibile mediante l’URL <https://it.wikipedia.org/wiki/Archeologia>, contiene moltissimi collegamenti (*link*) ipertestuali, riconoscibili nel layout dell’enciclopedia libera grazie al colore blu del testo. Per un utente parlante italiano non è difficile distinguere i collegamenti che rimandano a concetti di approfondimento (“scavo stratigrafico”, “ricognizione archeologica”, “archeologia computazionale”, ecc.) da altri tipi di collegamento, quali ad esempio le diverse lingue in cui la pagina è disponibile (simile, ma mai uguale nei contenuti), i *link* sulle informazioni generiche dell’enciclopedia, le sezioni tecniche sulle procedure di modifica della pagina e discussione sui contenuti, o infine dai *banner*, magari temporanei, relativi a concorsi, donazioni, ecc. Un utente che non parla l’italiano ma che è abituato a usare Wikipedia riuscirà comunque a fare molte di queste distinzioni, anche se non sarà in grado di comprenderne le sfumature: per esempio, non saprà stabilire con certezza il rapporto tra l’argomento principale della pagina corrente, “archeologia”, e i concetti di “scavo stratigrafico”, “ricognizione archeologica” o “archeologia computazionale”. Un utente, poi, che non abbia mai visto Wikipedia e non conosca l’italiano, ma conosca il World Wide Web, vedrebbe solo un documento con tantissimi *link*, distinti dal resto del testo per il convenzionale colore blu, che rimandano ad altri documenti, dei quali però non sarebbe in grado di descrivere la natura senza prima accedervi e leggerli.

Alla possibile obiezione che non esistono utenti di quest’ultima tipologia

Immagini dalla presentazione di Tim Berners Lee del 1994 a Ginevra dove viene introdotta l'idea del web semantico



(o almeno non nel mondo occidentale contemporaneo, profondamente immerso nella Rete), si può ribattere che invece esistono, che sono tanti e che ognuno di noi ne “pratica” numerosi ogni giorno: tutti i nostri dispositivi, che definiamo intelligenti o “smart”, vedono infatti il Web esattamente come la terza tipologia di utente descritto sopra. I nostri *smart device*, di qualsiasi potenza o fascia di prezzo, vedono nelle pagine Wikipedia solo una serie di rimandi ad altre pagine, senza essere in grado di stabilire se queste siano utili o meno all’approfondimento di un determinato concetto. È per questo motivo che Berners-Lee definisce il Web come un “flat, boring world devoid of meaning”, un mondo piatto, noioso e privo di significato. Nel Web semantico invece i collegamenti sono tutti descritti attraverso il rapporto che creano tra le risorse che collegano e quindi possono essere a loro volta catalogati e interrogati. Per riprendere l’esempio di prima, nel Web tradizionale possiamo avere:

```
https://it.wikipedia.org/wiki/Archeologia [rimanda genericamente a] https://en.wikipedia.org/wiki/Archaeology
```

Chi conosce l’italiano e l’inglese o ha dimestichezza con Wikipedia sa che questo enunciato significa che la pagina italiana di Wikipedia ha un

The image shows a screenshot of the Italian Wikipedia page for 'Archeologia'. The page layout includes a top navigation bar with 'Voci', 'Discussione', 'Leggi', 'Modifica', 'Modifica wikitesto', 'Cronologia', and 'Cerca in Wikipedia'. Below the navigation bar is a banner for Wikimedia Italia with the text 'Libera la cultura. Dona il tuo 5x1000 a Wikimedia Italia. Scrivi 94039910156.' The main content area features the title 'Archeologia' and a sub-header 'Da Wikipedia, l'enciclopedia libera.' The text defines archaeology as the science that studies the civilization and the culture of the past and its relationship with the environment. It also mentions the traditional division of archaeology into classical, industrial, or paleontology, and specific fields like urban archaeology or theoretical archaeology. A table of contents is provided, listing sections such as 'Tecniche e metodi di indagine', 'Metodi di datazione', 'Discipline correlate e attuali indirizzi di studio', and 'Storia dell'archeologia'. The page also includes two small images: one showing students at an archaeological excavation in the Forum of Trajan, and another showing an archaeologist working in a trench in Turkey.

collegamento alla pagina in inglese di Wikipedia. Basta una generica esperienza del Web per sapere che questo collegamento non è necessariamente speculare: se il sito a ha un collegamento al sito b, non è detto che il sito b abbia il reciproco collegamento al sito a. Nel Web semantico, i collegamenti assumono una rilevanza maggiore e, se organizzati in un sistema organico di definizione della conoscenza (in una “ontologia”), permettono la deduzione (ing. *inference*) anche di informazioni non espressamente codificate. Per questo motivo i collegamenti del Web semantico hanno un valore di significato, una “etichetta” che viene chiamata “predicato”, parte di una struttura semplice tripartita costituita da soggetto-predicato-oggetto, ad esempio:

Pagina Wikipedia dedicata all'Archeologia (Fonte: <https://it.wikipedia.org/wiki/Archeologia>, 10 aprile 2019)

`https://it.wikipedia.org/wiki/Archeologia` → [è scritta in] → italiano
`https://en.wikipedia.org/wiki/Archaeology` → [è scritta in] → inglese
`https://it.wikipedia.org/wiki/Archeologia` → [è la

versione in altra lingua di] → <https://en.wikipedia.org/wiki/Archaeology>

In un simile sistema, dove i collegamenti sono qualificati, catalogati, interrogabili e organizzati in sistemi logici che permettono analisi dettagliate, anche una macchina potrebbe orientarsi e offrire risultati più pertinenti alle nostre ricerche. Tim Berners-Lee nel 1994 faceva l'esempio di un programma che potrebbe cercare per noi una casa che corrisponda alle nostre aspettative, negoziare il trasferimento di proprietà dal vecchio al nuovo proprietario e aggiornare il registro catastale: potremmo fare a meno di agenti immobiliari e notai.

Per fare un esempio in campo archeologico, se sostituissimo le pagine di Wikipedia con tipologie di reperti e contesti e i predicati possibili fossero [è stato rinvenuto in] (il cui speculare sarebbe [conteneva]) o [ha decorazione], [ha testo], [è datato al], ecc., potremmo fare ricerche molto specifiche e costruire *corpora* tematici in maniera immediata, dai più semplici (“tutte le forme ceramiche defunzionalizzate in antico provenienti da contesti sepolcrali”) ai più complessi (“tutti i testi romani che trattano il concetto giuridico di libertà”) e molto altro ancora, ponendoci come unico limite la fantasia e la complessità della banca dati di base.

Proprio questi dati collegati (*linked data*) sono uno dei fondamenti del Web semantico⁴⁰, argomento sul quale torneremo più volte; per ora basti tenere a mente che se il Web dei documenti ipertestuali richiese qualche mese per esplodere e diventare una realtà globale, dopo venticinque anni dalla conferenza del 1994 il Web semantico è tuttora in divenire e in fase di progressiva implementazione.

Dal punto di vista tecnico e tecnologico, il Web, per come lo conosciamo oggi, costruito sul protocollo HTTP, sul linguaggio di marcatura HTML e sul sistema di indirizzamento degli URL, è perfettamente in grado di supportare il passaggio al gradino successivo, ovvero alla semantica. Anche i protocolli con i quali si sta sviluppando il Web semantico sono già stati definiti e risultano molto stabili.

Uno degli elementi di base, forse difficile da comprendere a un primo approccio, è costituito dagli Uniform Resource Identifier (URI), ovvero un sistema integrabile – o meglio un insieme di sistemi integrabili – funzionale a identificare ogni risorsa in modo univoco e non ambiguo. Una risorsa

⁴⁰ Christian Bizer, Tom Heath, e Tim Berners-Lee, «Linked Data - The Story So Far», *International Journal on Semantic Web and Information Systems (ISWIS)* 5, n. 3 (2009): 1–22, <https://doi.org/10.4018/jswis.2009081901>.

è, nel Web (semantico), un qualsiasi elemento, materiale o immateriale: un documento HTML, un'immagine, un reperto archeologico, una forma ceramica, un colore, un luogo, una persona, ecc. Gli URI sono dunque un sistema di “nomi”, o più correttamente “identificativi”, potenzialmente estensibili e che possibilmente contengano in sé la via per raggiungere la risorsa stessa (anche se questo aspetto è meno importante)⁴¹. Naturalmente, è impensabile che possano esistere sistemi universali (come è impossibile che un ente o un'istituzione si impegni ad assegnare identificativi a qualsiasi oggetto o concetto esistente nelle nostre culture), ma è del tutto ragionevole che possano esistere (e di fatto ci sono) sistemi circoscritti collegabili o collegati tra loro.

Un esempio di sistemi di identificazione ben noto a molti è l'International Standard Book Number (ISBN), che attribuisce a ogni libro pubblicato un codice a 13 cifre (fino al 2007 era a 10 cifre, poi è stato esteso) e che è nato proprio dall'esigenza di avere una gestione digitale del catalogo dei libri. Benché non vi sia alcun divieto di pubblicare un libro senza il codice ISBN, è nel pieno interesse di editori, autori e distributori che ogni volume sia dotato di un proprio codice univoco, affinché la diffusione e la distribuzione ne sia agevolata. Il successo di questo sistema si deve alla sua utilità generale e per questo motivo si è rapidamente diffuso. Può inoltre accadere che un sistema di nomi diventi uno standard anche per altre categorie di oggetti, come è successo al numero di ISBN che è stato esteso da Amazon a tutti i prodotti disponibili nel proprio negozio *online*: il codice ASIN, che identifica ogni singolo articolo in vendita sul noto portale di *e-commerce*, non è altro che un codice numerico a 13 cifre che coincide con il codice ISBN nel caso dei libri e con un nuovo codice creato *ad hoc* per tutti gli altri prodotti.

Anche all'interno di uno scavo o di una ricerca archeologica vengono in continuazione create e mantenute liste di identificativi per ogni strato, reperto, pianta, sezione, fotografia, rilievo, ecc. Questi elenchi possono essere estremamente semplici (la lista delle piante create nel corso di una campagna) o anche molto complessi, contenenti riferimenti incrociati ad altre entità, ecc. Siamo di norma abituati a gestire questa complessità in maniera intuitiva,

⁴¹ Per esempio l'URI <https://www.wikidata.org/wiki/Q23498> può essere aperto con un *browser* e rimanda alla pagina di Wikidata relativa al concetto di “archeologia”. La stringa <http://example.com/archeologia/tecnologia-di-produzione-ceramica/vernice-nera>, è un URI valido e anche molto comunicativo, in quanto permette di capire facilmente l'oggetto che descrive, in quel determinato ambito linguistico che è l'italiano; tuttavia, se si tenta di inserirla nella barra degli indirizzi di un *browser*, non offre nessuna informazione sulla risorsa.

Inv.	US	Cassa	Reperto	Confronti	Campione	SABAPM
2011/00181	236	004	VN	Morel 2784	VN30	649705
2011/00678	248	024	TSI/TSNI	Conspectus 18	TSI13	214990
2011/00296	228	008	TSI/TSNI	Conspectus 4.6	TSI12	554744
2011/00039	76	001	TSI/TSNI	Conspectus 22.1.4	TSI1	936114
2011/00336	313	009	TSA	Hayes 61b	TSA6	406614
2011/00514	464	015	PS	Atlante II, tipo 71	PS11	246528
2011/00483	333	014	LUC	Atlante VIII, p. 195.	LUC7	841173

Un tipico esempio di una tabella relativa ai reperti archeologici, estratto dalla banca dati degli scavi nella città romana di Suasa, tabella dei reperti (fonte: <https://db.bradypus.net/#/suasa>). Nella tabella sono allineati sistemi d'identificazione relativi ad autorità diverse.

senza soffermarci a riflettere sul significato o le potenzialità che un documento all'apparenza molto semplice – come la lista dei reperti da catalogare che viene consegnata alla Soprintendenza a fine campagna – possa avere. In una lista del genere si trova di norma una mappa articolata, composta di identificativi appartenenti a vari sistemi, collegati insieme – appunto – dal documento in questione. Un primo sistema, ad esempio, è costituito dai numeri di inventario che vengono assegnati a ogni reperto, sul campo, al momento del rinvenimento, oppure in seguito in laboratorio. Il campo Inv nella tabella esemplificativa contiene un codice univoco formato dall'anno della campagna di scavo seguito da un numero progressivo (che ogni anno viene azzerato e ricomincia da uno). In un'eventuale trasposizione nel Web semantico della banca dati che stiamo esaminando, questo campo deve essere espresso sotto forma di URI, completo e possibilmente significativo; ad esempio, il primo reperto potrebbe avere il seguente identificativo:

<http://example.com/suasa/unibo/reperti/2011/00181>

Dove *suasa* indica il sito, *unibo* indica che si è all'interno del sistema informativo dell'Università di Bologna⁴² e non di altri soggetti che operano

⁴² Il sito romano di *Suasa*, si trova presso Castelleone di *Suasa* (Ancona) ed è da più di trent'anni oggetto di studio da parte dell'Università di Bologna, che ha indagato varie area urbane e periurbane e il territorio circostante. Negli ultimi dieci

nel sito (e che potrebbero avere altri sistemi di identificazione), reperti indica il catalogo dei reperti archeologici (e non, per esempio, di unità stratigrafiche, campioni, analisi o altro), infine 2011/00181 identifica il nostro frammento. È chiaro come questo URI sia unico e identifica la nostra risorsa (il frammento di ceramica a vernice nera) in maniera molto puntuale e senza possibilità di equivoci all'interno del sistema *Suasa*, ma anche a un livello globale. La stringa potrebbe essere anche meno autoevidente, purché venga garantita l'univocità. Un altro aspetto, che fa sorgere qualche perplessità, è che un URI non deve necessariamente essere un URL valido, vale a dire che un URI non deve per forza potersi aprire in un *browser*. L'URI sopra indicato è un URI perfettamente valido, benché non rimandi ad alcun contenuto sul Web⁴³. Il Web semantico definisce ogni risorsa con un URI di questo tipo, che spesso appare inutilmente lungo e complesso, ma che è estremamente efficace per identificare ogni elemento in maniera precisa e al di là di ogni possibile ambiguità

Allo stesso modo si potrebbe indicare l'US 236: <http://example.com/suasa/unibo/us/236> e la cassa dove il reperto è conservato: <http://example.com/sabapm/suasa/dep/004>. Si noti che nel secondo caso il modello è cambiato: in prima posizione si trova in questo caso l'indicazione della Soprintendenza Archeologia, Belle Arti e Paesaggio delle Marche, che evidentemente gestisce l'immagazzinamento e quindi il sistema di numerazione delle casse; si sta quindi facendo riferimento a un sistema di identificativi gestiti da un ente esterno.

Per chi studia i reperti ceramici di età romana, le sigle che descrivono il reperto non destano grossi problemi: VN sta per “vernice nera”, una tecnologia di produzione ceramica molto diffusa in età ellenistica in tutto il Mediterraneo. Non si tratta chiaramente di una specificità legata a un determinato sito archeologico – nel nostro caso *Suasa* – e probabilmente il vocabolario usato per classificare queste classi è condiviso da molti o tutti

anni è stato messo a punto un sistema informativo unificato per la gestione dei dati dello scavo, che fa uno sì banche dati *online* (<http://db.bradypus.net/#/suasa>), un sistema GIS e alcuni canali webGIS per solo uso interno. Per una sintesi sulle ricerche si veda Enrico Giorgi e Giuseppe Lepore, a c. di, *Archeologia nella valle del Cesano da Suasa a Santa Maria in Portuno: atti del Convegno per i venti anni di ricerche dell'Università di Bologna (Castelleone di Suasa, Corinaldo, San Lorenzo in Campo 18-19 dicembre 2008)* (Bologna: Ante Quem, 2012). Sugli aspetti relativi alla documentaione di scavo si veda Enrico Giorgi, Federica Boschi, Michele Silani e Julian Bogdani, «Documentare l'archeologia da Burnum (Sebenico, Croazia) a Suasa (Ancona): una tradizione rinnovata», *Archeologia e Calcolatori* 23 (2012): 261–82.

⁴³ È implicito che è preferibile che gli URI siano “risolvibili”, ovvero puntino a pagine web che in qualche modo ne rappresentino o descrivano i contenuti.

i progetti che si occupano del periodo classico dell'Università di Bologna. Quindi un URI come: <http://example.com/unibo/vocab/ceramica/vn> non avrebbe bisogno di particolari spiegazioni. Usare vocabolari standardizzati e condivisi è una raccomandazione importante per favorire l'interoperabilità tra i vari sistemi e aiuta enormemente ricerche trasversali su diversi *dataset*. Un ulteriore passo verso una maggiore interoperabilità potrebbe essere quello di usare vocabolari sempre più condivisi e, per esempio, definire la vernice nera utilizzando il seguente URI: <http://vocab.getty.edu/aat/300387491>. Si tratta della definizione di ceramica a vernice nera (*black-gloss ware*) all'interno del Getty Vocabularies Project, un'iniziativa finalizzata a produrre una terminologia strutturata per l'arte, l'architettura, le arti decorative, gli archivi, i documenti audiovisivi e il materiale bibliografico⁴⁴. Questo permetterebbe al nostro frammento ceramico di essere virtualmente messo in connessione con tutte le altre istanze di "vernice nera" in altre banche dati internazionali che utilizzano il vocabolario del Getty Museum. Gli archeologi conoscono molto bene l'importanza di usare terminologie condivise, ma anche le difficoltà che sorgono

Schermata di *PeriodO* con la definizione di "Alto Medioevo" come definito da *Ariadne*, a sua volta derivato da *FastiOnline*. Il periodo di riferimento assoluto è definito dall'intervallo 530-1071. Al periodo è attribuito l'URL <http://n2t.net/ark:/99152/p0qhb66pn9d>

The screenshot shows the PeriodO interface for the 'Alto Medioevo' period. The main content area displays the following information:

- Periods:** Viewing 1 - 8 of 8. Show 25 periods at a time.
- Filters:** Time range (Bar chart showing distribution from -8,000 to 1000 BP, with 'Hide outliers?' checked).
- Text:** Match string: alto.
- Source:**
 - 5 David G. Anderson, *Digital Index of North American Archaeology (DNAIA)*, 2012.
 - 2 ARIADNE Consortium, *ARIADNE Data Collection*, 2015.
 - 1 César Pascero Oubiña and César González-Pérez, *SIA+ Chrono-Cultural Categories*, 2014.
- Language:**
 - 8 English
 - 2 Spanish
 - 1 Italian
- Spatial coverage:**
 - 2 Florida
 - 1 Iowa
 - 1 Kentucky
 - 1 Spain
 - 1 Sicily
 - 1 Missouri
 - 1 Galicia and neighbouring areas in NW Spain

The main definition for 'Alto Medioevo' is as follows:

- Original label:** Alto Medioevo (Ita-Istr)
- Alternate labels:** Early Medieval (eng-Istr)
- Spatial coverage des...:** Sicily
- Start:** 1470 BP (2000) (SSD value: 530) 109 BP (2000)
- Stop:** (SSD value: 1071)
- Editorial notes:** Derived from FASTI, but no match in 2004 dataset.

Below the definition is a table of related periods:

Period	Earliest start	Latest stop
Alto Romano	-50	200
Alto Romano	-50	190
Dalton	-8000	-7000
Dalton	-8500	-7000
Dalton	-6550	-7550
Dalton-Henderson	-8500	-8000
Fl. Walton A.D. 1000-1500	1000	1500

da tentativi di uniformazione.

Tra i molti tentativi in atto di creare vocabolari condivisi è di articolare interesse in ambito archeologico e storico il progetto PeriodO, un'iniziativa che mantiene *online* una lista articolata e dotata di identificativi univoci (URI) di periodi cronologici di interesse storico, archeologico o relativi alla storia dell'arte, per come sono stati definiti in pubblicazioni scientifiche. È in questo modo possibile riferirsi a una definizione cronologica in modo inequivocabile e quindi collegare dati attraverso la datazione, anche questa è espressa usando sistemi di riferimento diversi⁴⁵.

In Italia, l'Istituto Centrale per il Catalogo e la Documentazione (ICCD) è l'ente ministeriale preposto alla definizione degli standard nazionali catalografici⁴⁶ e dei vocabolari (*thesauri*) da utilizzare nella compilazione delle schede descrittive dei Beni Culturali. Molti vocabolari di enti di ricerca o di singoli progetti si rifanno idealmente a questi schemi, la cui complessità, però, ne ha fortemente limitato la diffusione e l'adozione.

Gli standard del Web semantico, mediante la struttura degli URI, permettono a ciascun gruppo di ricerca di definire e utilizzare i propri schemi – i più adatti all'obiettivo di ricerca –, ma offrono al contempo l'opportunità di una maggiore interoperabilità grazie alla possibilità di mescolare tra loro vari schemi di diversa provenienza, ad esempio:

```
http://example.com/suasa/unibo/reperti/2011/00181
è parte della classe
http://vocab.getty.edu/aat/300387491
```

O anche di stabilire collegamenti di uguaglianza, similitudine, derivazione, ecc., tra concetti in diversi schemi, ad esempio:

```
http://example.com/unibo/vocab/ceramica/vn
è uguale a
http://vocab.getty.edu/aat/300387491
ed è uguale a
https://collection.britishmuseum.org/resource/
```

⁴⁵ <http://perio.do/>, Adam Rabinowitz, Ryan Shaw, Sarah Buchanan e Patrick Golden, «Making sense of the ways we make sense of the past: The periodo project», *Bulletin of the Institute of Classical Studies* 59 (1 dicembre 2016): 42–55, <https://doi.org/10.1111/j.2041-5370.2016.12037.x>; Adam Rabinowitz, Ryan Shaw, e Patrick Golden, «Making up for Lost Time: Digital Epigraphy, Chronology, and the PeriodO Project», in *Crossing Experiences in Digital Epigraphy. From Practice to Discipline*, a c. di Annamaria De Santis e Irene Rossi (Berlin: De Gruyter, 2018), 202–15.

⁴⁶ <http://www.iccd.beniculturali.it/it/standard-catalografici>.

thes:x14737

ed è uguale e

<https://www.wikidata.org/wiki/Q908930>

Il penultimo elemento è l'URI del concetto "vernice nera" nel *database* semantico del British Museum, mentre l'ultimo fa riferimento alla banca dati semantica collaborativa di Wikidata.

Collegare insieme risorse attraverso connessioni significative, fondamento del Web semantico, è anche una necessità costante e concreta del nostro lavoro quotidiano. È, si potrebbe dire, un modo molto "umano" di conoscere e descrivere la realtà. Le liste autoritative (*authority files*), elenchi e vocabolari condivisi ai quali fare riferimento in maniera più o meno puntuale, sono strumenti di fondamentale importanza per creare connessioni su una scala più ampia, superando le barriere linguistiche: nel vocabolario semantico del Getty, l'URI <http://vocab.getty.edu/aat/300387491> è legato alle varie forme in lingua inglese (il vocabolario è disponibile solo in inglese) che descrivono "vernice nera", come "*black gloss ware*", "*black-gloss ware*", "*ware, black-gloss*", "*black-gloss wares*", "*black-glaze ware (black-gloss ware)*", mentre altri *database* più generici, come Wikidata, danno accesso alle "etichette" (*label* è il termine tecnico) in altre lingue, tra le quali anche l'italiano "Ceramica a vernice nera".

Il Web semantico, dunque, richiede un elevato grado di chiarezza, anche quando si tratta di concetti molto diffusi e ben conosciuti all'interno di un dominio di conoscenza, come potrebbero essere le sigle "VN" o "TSA" o i confronti "Morel 2784" o "Hayes 61b" per chi studia la ceramica ellenistico-romana. Il sistema degli URI assicura che ogni possibile ambiguità (anche semplicemente linguistica) venga risolta alla radice.

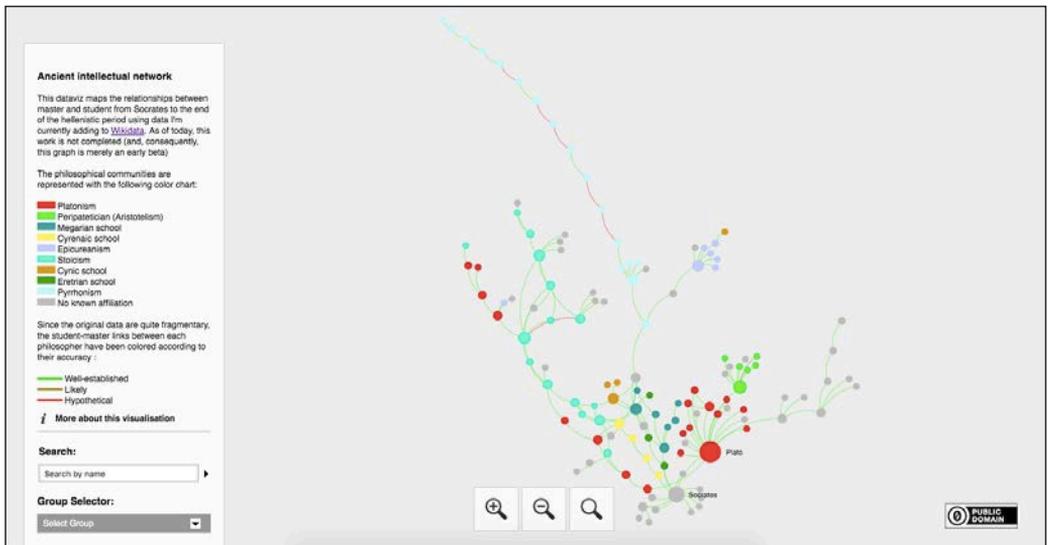
Un altro ambito – molto frequentato dagli archeologi – che presenta non poche ambiguità (da risolvere) è quello della toponomastica. Motivi storici, linguistici, grafici, politici, economici, religiosi, ecc., hanno fatto sì che una stessa località potesse essere indicata con toponimi diversi, tra i quali è spesso difficile orientarsi. In questo campo si ha a disposizione l'indice geografico (*gazetteer*) Geonames⁴⁷, un *database* di toponimi (associati a coordinate) che viene popolato a partire da una moltitudine di fonti diverse. A ogni luogo è attribuito un identificativo numerico univoco, al quale vengono riferite le varianti in altre lingue o dialetti. Uno strumento che serve dunque a unire forme diverse, ma anche a disambiguare e a distinguere forme omografe che si riferiscono a luoghi distinti. Per esempio, "Rome"

⁴⁷ <https://www.geonames.org/>.

si può riferire alla forma inglese del toponimo alla città del Lazio in Italia⁴⁸ oppure a una località in Georgia, Stati Uniti⁴⁹. Al contempo, la capitale d'Italia può essere indicata dai toponimi (etichette) “Rome”, “Roma”, “Urbs”, “Ρώμη”, “אמור”, “Romë”, “Рим”, “□—ㄣ” o anche “00100” secondo il sistema dei Codici di Avviamento Postale (CAP).

L'ultimo esempio di lista che potrebbe dirsi infinita, nata della necessità di collegare *record* di autorità (*authority file*) diversi nel tentativo di dare una mappa chiara della realtà, al di là di ogni possibile variante culturale o linguistica, è costituito dal Virtual International Authority File (VIAF, <http://viaf.org/>). Questo progetto internazionale, sostenuto da molte biblioteche, cerca di collegare le liste di persone, luoghi, istituzioni, opere, ecc., che ciascuna istituzione partecipante mantiene, fornendo quindi un indice, una “lista delle liste”, dove tutto viene definito da URI e messo in correlazione. Così sappiamo che all'identificativo di VIAF 96994048⁵⁰ corrisponde lo scrittore inglese William Shakespeare, il cui nome in libri editi in tempi e luoghi diversi potrebbe comparire in maniera differente. Con questo identificativo si ha la possibilità di risalire in maniera diretta (e soprattutto automatica) a tutte le istanze diverse in cui questo personaggio (non più una forma specifica del suo nome) è menzionato, almeno per quanto riguarda le biblioteche che partecipano al VIAF.

Esempio di una visualizzazione automatica di una possibile Rete filosofica antica, prodotta dai dati aperti e collegati (LOD) presenti in Wikidata (fonte: Ancient intellectual network, <http://projetjournal.org/network/>)



48 <http://www.geonames.org/3169070>.

49 <http://www.geonames.org/4219762>.

50 URI <http://viaf.org/viaf/96994048>.

Anche il già menzionato Wikidata è una banca dati semantica che registra una moltitudine straordinaria di risorse e grande parte dei concetti trattati in Wikipedia, definendo URI stabili e una mappa molto articolata di etichette in varie lingue e di collegamenti significativi. Basta visitare la pagina dedicata a Roma (la città del Lazio, in Italia)⁵¹ per rendersi conto della quantità di possibili connessioni che una risorsa può avere. Queste relazioni possono essere interrogate, esattamente come possono essere interrogati i dati collegati, producendo ricerche ad alto grado di attendibilità e significato.

Web semantico: ricette e ingredienti

Le cause per cui una tecnologia dal potenziale così elevato non è stata ancora adottata su vasta scala sono complesse e molteplici, e sono dovute sia ai limiti della tecnologia, sia a imposizioni della legislazione vigente, che non sempre permette la libera disponibilità di dati da usare e collegare (sulle problematiche legate ai Linked Open Data – LOD – si tornerà diffusamente più avanti).

Dal punto di vista tecnologico, il W3C ha rilasciato uno strumento basilare, lo standard chiamato Resource Description Framework (RDF)⁵², nato per definire modelli di metadati per le risorse web e divenuto rapidamente uno standard unico per la codifica di relazioni semantiche tra dati e risorse, in modo da renderli interpretabili e interrogabili anche dalle macchine. In altre parole, il RDF è la tecnologia che consente di esprimere in maniera formale quei collegamenti tra risorse fin qui trattati. Come il modello relazionale di *database* permette di organizzare i dati usando formati tabellari fatti di righe e colonne collegati tra loro, così il modello RDF offre la possibilità di descrivere ogni risorsa (oggetto concreto o concetto astratto) usando frasi dalla sintassi estremamente semplice, costituita dai tre elementi di soggetto-predicato-oggetto. Questi enunciati (*statement*) sono chiamati, proprio a causa di questa struttura, triple (*triple*). Un esempio molto semplice potrebbe essere l'informazione "Lewis Carroll, il cui vero nome è Charles Lutwidge Dodgson, è l'autore del libro *Alice nel paese delle meraviglie*". Questa informazione si potrebbe dividere nelle seguenti *triple* (l'uso delle parentesi quadre per dividere le tre parti è introdotto solo per chiarezza):

[Charles Lutwidge Dodgson] [ha come nome di penna]

⁵¹ <https://www.wikidata.org/wiki/Q220>.

⁵² <https://www.w3.org/RDF/>.

```
[Lewis Carroll]
[Charles Lutwidge Dodgson] [è l'autore di] [Alice nel
paese delle meraviglie]
```

Un calcolatore che si trova davanti a questi enunciati non sa che Charles Lutwidge Dodgson è una persona e neppure che *Alice nel paese delle meraviglie* è un libro. Allora potremmo essere ancora più espliciti nei nostri enunciati e introdurre due nuovi concetti, quello di “persona” e quello di “libro”:

```
[Charles Lutwidge Dodgson] [è] [una persona]
[Charles Lutwidge Dodgson] [ha come nome di penna]
[Lewis Carroll]
[Charles Lutwidge Dodgson] [è l'autore di] [Alice nel
paese delle meraviglie]
[Alice nel paese delle meraviglie] [è] [un libro]
```

A questo punto abbiamo bisogno di definire le persone, le azioni e le idee in maniera astratta, facendo ricorso agli URI. Se ne potrebbero introdurre di propri, oppure si potrebbero usare identificativi condivisi, tratti da *authority file* più diffusi rispetto a quelli che potremmo creare noi. Per le persone, soprattutto se si tratta di personaggi legati alla letteratura, viene in soccorso il già menzionato VIAF. Da una semplice ricerca scopriamo che VIAF già ha unito i nomi di Charles Lutwidge Dodgson e Lewis Carroll sotto un unico identificativo, ovvero <http://viaf.org/viaf/66462036>. Allo stesso modo, sempre VIAF attribuisce un identificativo all'opera, <https://viaf.org/viaf/180646266>. In questo modo possiamo scrivere:

```
[http://viaf.org/viaf/66462036] [è] [una persona]
[http://viaf.org/viaf/66462036] [ha nome] [Charles
Lutwidge Dodgson]
[http://viaf.org/viaf/66462036] [ha pseudonimo]
[Lewis Carroll]
[http://viaf.org/viaf/66462036] [è l'autore di]
[https://viaf.org/viaf/180646266]
[https://viaf.org/viaf/180646266][ha titolo] [Alice
nel paese delle meraviglie]
```

Per poter codificare in maniera corretta queste informazioni abbiamo bisogno di trovare un modo di istruire il computer sul significato dei predicati è, ha nome, ha pseudonimo, è autore di, ha titolo, ecc. Questo

è compito delle ontologie, le rappresentazioni formali della conoscenza, articolate in una rigida codifica di concetti, nomi, relazioni, proprietà, relativi a un dominio specifico della conoscenza. Un'ontologia definisce, in un linguaggio formale, la lista dei predicati possibili, i loro significati e le conseguenti regole di deduzione. Per fare un esempio, in un'ipotetica ontologia che descrive i rapporti tra le persone, alla relazione [è padre di] corrisponde, nell'altra direzione, la relazione [è figlio/a di].

Abbiamo quindi bisogno di un'ontologia e di vocabolari di riferimento che descrivano le persone e le relazioni autorialità/produzione intellettuali, oppure dobbiamo utilizzare contemporaneamente più schemi, ciascuno specializzato in aspetti diversi, selezionando in questi le relazioni e le proprietà che maggiormente soddisfano le nostre esigenze. Per la descrizione di persone e relazioni tra persone possiamo usare Friend Of A Friend (FOAF)⁵³, un'ontologia estremamente semplice di termini relativi a persone:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
<http://viaf.org/viaf/66462036>
  a foaf:Person;
  foaf:name "Charles Lutwidge Dodgson"
  foaf:nick "Lewis Carroll"
.
```

L'esempio sopra citato è un caso di utilizzo di RDF formale standard, esplicitato (serializzato) secondo la sintassi Turtle⁵⁴. Il prefisso “foaf:” definisce il *namespace* di pertinenza, ovvero indica che il termine a seguire è definito nell'ontologia FOAF, che è disponibile all'indirizzo <http://xmlns.com/foaf/0.1/>. La quinta riga, invece, stabilisce che la risorsa che si sta descrivendo fa parte (è una istanza) della classe di Person di FOAF, ovvero stiamo parlando di una persona.

Segue l'etichetta “name” per indicare il nome proprio e quella “nick”, una forzatura per definire il nome di penna. Per stabilire le relazioni sull'auto-

⁵³ <http://www.foaf-project.org/>.

⁵⁴ Turtle è un formato per RDF facilmente leggibile ed è uno standard W3C (<https://www.w3.org/TR/turtle/>). La tecnologia RDF prevede vari formati alternativi di serializzazione, più o meno espressivi e più o meno di facile lettura. Alcuni dei più comuni sono RDFa (<https://www.w3.org/TR/rdfa-primer/>), RDF/XML (<http://www.w3.org/TR/rdf-syntax-grammar/>), N-Triples (<http://www.w3.org/TR/n-triples/>), JSON-LD (<http://json-ld.org/>).

rialità abbiamo bisogno di usare una seconda ontologia, dal momento che FOAF non li prevede, per esempio la diffusissima Dublin Core⁵⁵:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix wdata: <http://www.wikidata.org/entity/> .

<http://viaf.org/viaf/66462036>
  a foaf:Person;
  foaf:fame "Charles Lutwidge Dodgson"
  foaf:nick "Lewis Carroll"
  dc:creator <https://viaf.org/viaf/180646266>
  .
<https://viaf.org/viaf/180646266>
  a wdata:Q47461344
  .
```

Allo schema sopra è stato aggiunto un ulteriore prefisso, quello relativo al già menzionato Wikidata, dove sono già presenti e descritte le nostre risorse. Tra le risorse disponibili troviamo anche il concetto di “opera scritta” (ingl. *written work*), definito dall’URI <https://www.wikidata.org/wiki/Q47461344>. A questo punto lo schema sopra rappresenta (quasi) perfettamente la nostra frase “Lewis Carol, il cui vero nome è Charles Lutwidge Dodgson è l’autore del libro *Alice nel paese delle meraviglie*”, dove “libro” viene specificato nella sua accezione di opera dell’intelletto e non in quella di oggetto materiale. La complessità di questa sintassi e la difficoltà a orientarsi in ontologie e vocabolari sono il prezzo da pagare per una codifica che, a questo punto, è indipendente da ogni linguaggio naturale, sia esso l’italiano, l’inglese o qualsiasi altra lingua. Questi dati inoltre, essendo collegati a banche dati esterne – nel presente caso VIAF e Wikidata – possono ricevere da queste in maniera perfettamente automatizzata tutte le altre informazioni via via rese pubbliche. Non solo: altri collegamenti permettono di allargare ancora di più il numero dei risultati, fino a giungere a una serie di informazioni dalla trama virtualmente infinita come quella del WWW, ma con la differenza che ogni connessione è qualificata da un significato ben preciso. Proprio per distinguere questo modello

⁵⁵ <http://dublincore.org/>.

informativo dal WWW, Berners-Lee ha coniato il termine di “Giant Global Graph” o GGG⁵⁶.

Il Resource Description Framework, l’ossatura di questa trama, prevede anche una serie di strumenti tecnologici adeguati che consentono alla sua architettura di funzionare. In particolare, è stato sviluppato un linguaggio standard di interrogazione (*query*), chiamato SPARQL⁵⁷, che permette di “navigare” e di interrogare questa rete di collegamenti in maniera facilitata, usando un linguaggio molto simile a un inglese semplificato e derivato dal più diffuso SQL, impiegato nei *database* relazionali e trattato diffusamente più avanti. Anche dal punto di vista *hardware* sono necessari strumenti dedicati: dai *server* specializzati nella gestione e interrogazione degli RDF, a *database* specifici per la gestione delle triple, chiamati *triplestore*. Si tratta di *software* altamente specializzati, funzionali alla gestione di questa diversa rappresentazione e codifica della conoscenza, in continua evoluzione e che conosceranno nel futuro un’applicazione sempre più capillare.

Per una trattazione più esaustiva e approfondita degli strumenti tecnologici relativi al complesso argomento del Web semantico si rimanda a pubblicazioni specifiche⁵⁸ e alle tantissime risorse disponibili *online*.⁵⁹ Di norma sono richieste competenze informatiche avanzate per costruire applicazioni di Web semantico, non facili da acquisire per un umanista che si interessi di informatica. È però verosimile che, con la diffusione del Web semantico e il progressivo allargamento dei produttori e dei gestori di dati che decideranno di pubblicare RDF, nascano e si sviluppino piattaforme e *software* di più semplice utilizzo, come è successo ad esempio per i *database* relazionali. È un cambiamento auspicabile, dato che l’inaccessibilità di questa tecnologia da parte degli umanisti rappresenta un serio limite, che attualmente blocca la diffusione e la sperimentazione di applicazioni semantiche

⁵⁶ Tim Berners-Lee, «Giant Global Graph», *DIG* (blog), 21 novembre 2007, <https://web.archive.org/web/20160713021037/http://dig.csail.mit.edu/breadcrumbs/node/215>.

⁵⁷ SPARQL Protocol and RDF Query Language, <https://www.w3.org/TR/rdf-sparql-query/>.

⁵⁸ Per una introduzione completa e dettagliata, con molti esempi e alla portata anche di un pubblico generico senza particolari competenze informatiche si veda Tommaso Di Noia et al., *Semantic web: tra ontologie e open data* (Santarcangelo di Romagna (RN): Maggioli editore, 2018).

⁵⁹ La pagina web di riferimento è quella ufficiale del W3C (<https://www.w3.org/2001/sw/>) dove vengono pubblicate le ultime novità dei gruppi di lavoro e una serie importante di risorse e approfondimenti.

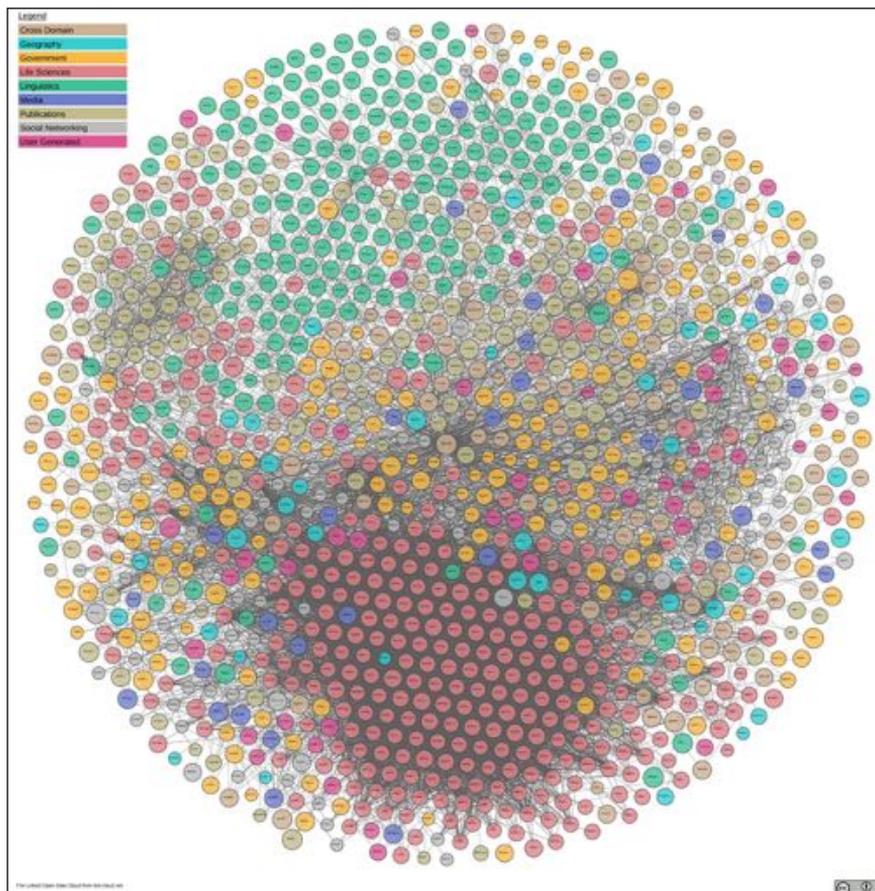


Diagramma della nuvola dei dati aperti collegati (LOD), al 29/03/2019 (fonte: <https://lod-cloud.net/>)

relative agli ambiti specifici dell'archeologia o delle scienze umane⁶⁰.

Nel campo dei beni culturali, l'ICOM (International Council of Museums) ha formalizzato un modello concettuale di riferimento chiamato CIDOC-CRM⁶¹. Si tratta di un'ontologia che fornisce definizioni, vocabolari e struttura formale per descrivere concetti impliciti ed espliciti nel campo dei beni culturali. Sviluppata all'interno di quello specifico comparto chiamato GLAM e che comprende gallerie, biblioteche (ingl. *libraries*), archivi e musei, dispone di strumenti per la rappresentazione anche di concetti

⁶⁰ Uno studio importante in questo ambito è costituito dalla tesi dottorale Leif Isaksen, «Archaeology and the Semantic Web» (University of Southampton. School of Electronics and Computer Science, 2011), <https://eprints.soton.ac.uk/206421/>. Gran parte dell'esperienza di questo lavoro è stato assorbito dal progetto Pelagios Commons, esaurientemente trattato nell'appendice.

⁶¹ CIDOC è la sigla che identifica il comitato internazionale per la documentazione all'interno del ICOM, mentre CRM sta per Conceptual Reference Model, <http://www.cidoc-crm.org/>.

prettamente archeologici⁶². Una caratteristica di questo modello è quella di incentrarsi sugli eventi e non sugli oggetti in sé, per cui gli oggetti vengono definiti e descritti da una seriazione di eventi che implicano la loro creazione, uso, manutenzione, ecc. Nulla vieta, comunque, il nascere di iniziative che da una parte siano mirate ad approfondire aspetti specifici⁶³ e dall'altra mirino a una semplificazione o parziale adozione delle ontologie più complesse. Il contributo “dal basso” dei singoli studiosi porterà verosimilmente a un progressivo arricchimento delle risorse disponibili nel prossimo futuro.

⁶² Øyvind Eide, Achille Felicetti, Christian-Emil Ore, Andrea D'Andrea e Jon Holmen, «Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping», in *Open Digital Cultural Heritage System*, a c. di David B. Arnold, Franco Niccolucci, Daniel Pletinckx e Luc Van Gool (Budapest: Archeolingua, 2011).

⁶³ Si vedano in appendice gli esempi di Nomisma e Kerameikos.

Capitolo 3

Banche dati: archiviazione, gestione, pubblicazione

Come si è visto, le banche dati sono alla base del Web dinamico e rivestono un ruolo fondamentale nel Web semantico, chiamato proprio “Web dei dati” (“*Web of data*”)¹ per l’importanza che assume l’accesso diretto, automatico e intelligente alle informazioni in esse contenute. I *database*, strumenti digitali capillarmente diffusi nelle scienze umanistiche e soprattutto nell’archeologia, sono un campo di applicazione molto stabile, dal momento che l’adozione pressoché esclusiva del modello relazionale ha portato a una base tecnologica estremamente affidabile. A parte poche eccezioni, di cui si farà menzione alla fine di questo capitolo, il modello relazionale è pressoché l’unico modello disponibile in campo archeologico, e verrà diffusamente trattato nei paragrafi seguenti. Una particolare attenzione verrà dedicata anche al linguaggio per eccellenza di gestione di banche dati relazionali, l’SQL, che è oggi divenuto uno dei principali linguaggi del Web.

Diversa la situazione in altri ambiti della ricerca umanistica, in particolare in quelli legati allo studio dei testi, dove i modelli di riferimento sono di tutt’altro genere. Anche in questo caso, e per le discipline più affini all’archeologia, si cercherà di offrire una panoramica generale, senza la pretesa di essere esaustivi.

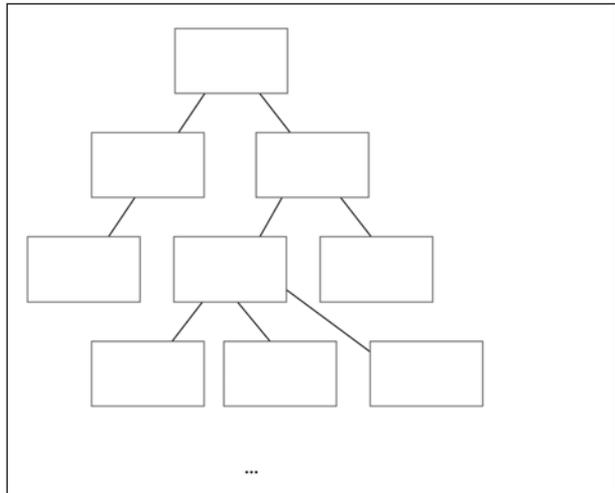
Il modello relazionale

Già in un articolo del 1998 dedicato alle applicazioni informatiche in archeologia, si concludeva che il cuore di un sistema informativo elettronico, in un qualsiasi ambito archeologico su qualsiasi scala, non potesse

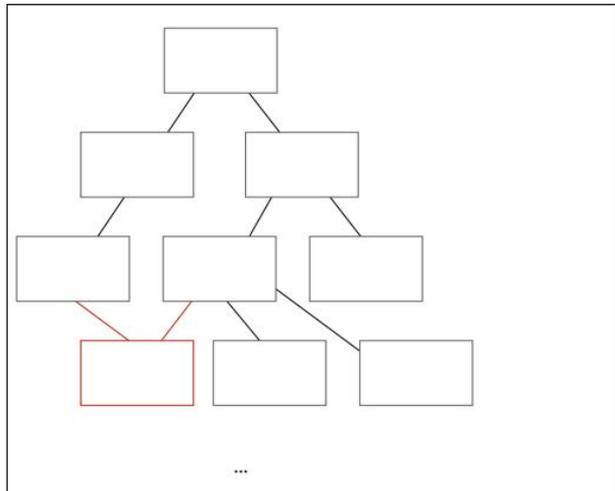
¹ <https://www.w3.org/2013/data/>.

essere altro che un flessibile gestore di *database* relazionali². Il modello relazionale è peraltro un concetto noto – in maniera più o meno consapevole – anche a chi non si occupa quotidianamente di gestire una banca dati, perché lo utilizziamo cercando un contatto nella rubrica del telefono o consultando una pagina di un sito web: dietro la maggior parte delle applicazioni *online* che hanno a che fare con dati strutturati, c'è probabilmente un *database* relazionale.

Esempio schematico del modello gerarchico di dati, esemplificato con un grafico ad albero: tutti gli oggetti hanno un'unica origine, chiamata radice (root) e ogni elemento ha un unico genitore



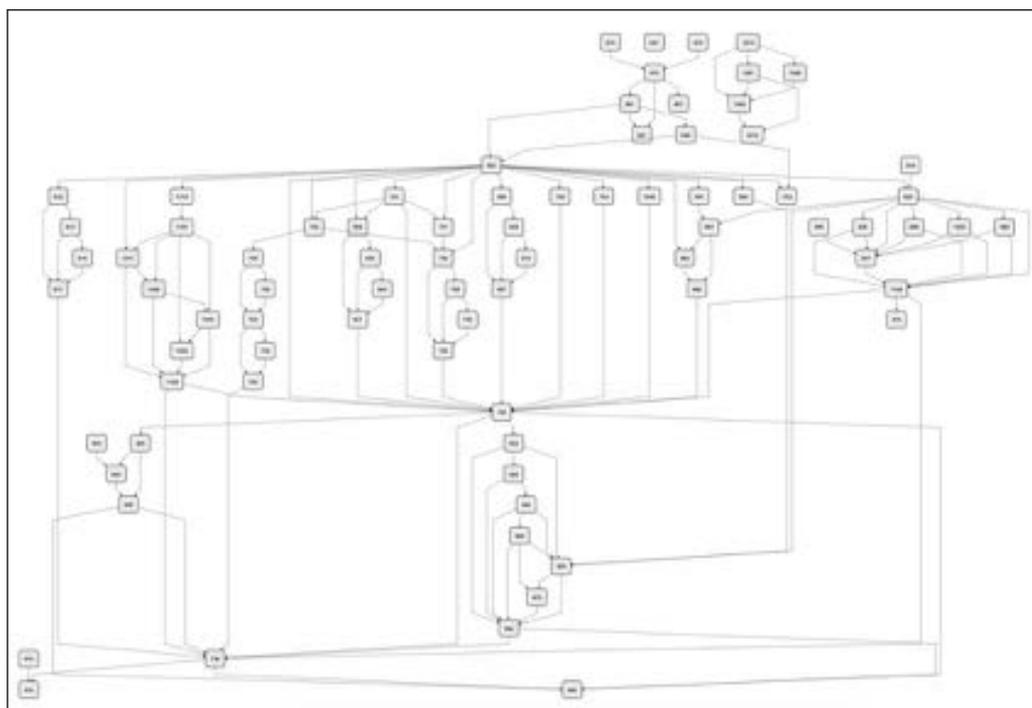
Esempio schematico del modello di dati a rete (network): ogni elemento può essere connesso a più di un genitore



² Albertus Voorrips, «Electronic Information Systems in archaeology. Some notes and comments», in *Methodological Trends and Future Perspectives in the Application of GIS in Archaeology*, a c. di Paola Moscati, *Archeologia e Calcolatori* 9 (Firenze: All'Insegna del Giglio, 1998), 251–67.

Il modello relazionale venne teorizzato da Edgar F. Codd, impiegato di IBM, nel 1969³ e negli anni '90 del secolo scorso era già di fatto il modello per eccellenza anche in ambiti periferici di applicazione, come può essere l'archeologia. Data la relativa facilità di implementazione, almeno per quanto riguarda progetti con strutture informative semplici, esso non richiedeva di norma competenze professionali esterne (informatici di mestiere) e poteva essere realizzato e mantenuto dagli stessi archeologi. A differenza di altri modelli di gestione dei dati, come ad esempio quello gerarchico o quello a rete, è molto più elastico e duttile nello stabilire complesse reti di relazioni tra le informazioni.

In realtà, il modello relazionale non è l'unico correntemente impiegato in ambito archeologico. Se ne potrebbe citare almeno un altro, quello a grafo, una modalità di organizzazione e rappresentazione dei dati comunemente utilizzato per la rappresentazione delle relazioni stratigrafiche e basata sul cosiddetto *matrix* di Harris.



Esempio del *matrix* di Harris, un grafo orientato comunemente usato per codificare e rappresentare le relazioni fisiche e dunque la sequenza cronologica relativa delle unità stratigrafiche (fonte: <https://db.bradyopus.net/#/suasa>).

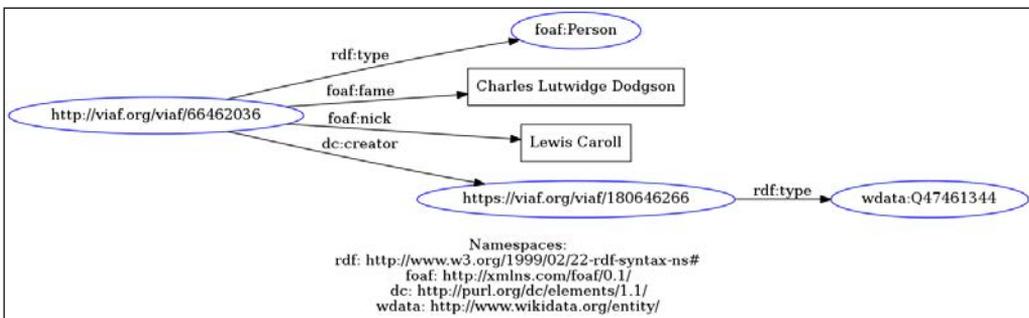
³ Edgar F. Codd, «Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks» (IBM, 1969); Edgar F. Codd, «A relational model of data for large shared data banks», *Communications of the ACM archive*. 13, n. 6 (1970): 377–87.

Nel modello a grafo – che qui non verrà trattato nelle sue originarie implicazioni matematiche e topologiche – i dati sono strutturati in “nodi”, legati tra loro da relazioni dette “archi”. Gli archi vengono definiti “orientati” quando la loro direzione ha un significato ben preciso e univoco e “non orientati” quando due dati sono connessi da una relazione biunivoca. Il *matrix* di Harris è un grafo orientato, poiché la direzione delle relazioni ha un significato specifico e univoco (eccezion fatta per la relazione di uguaglianza e quindi di contemporaneità tra due azioni). Per fare un esempio, la relazione copre (= è posteriore) non è biunivoca ma ha il suo corrispettivo opposto in è coperto (= è anteriore).

Il modello a grafo, per la sua struttura tripartita, nella sua forma più elementare di due nodi legati da un arco, è inoltre particolarmente adatto a rappresentare le triple (soggetto – predicato – oggetto) degli RDF e per questo motivo il suo utilizzo ha conosciuto recentemente una maggiore diffusione.

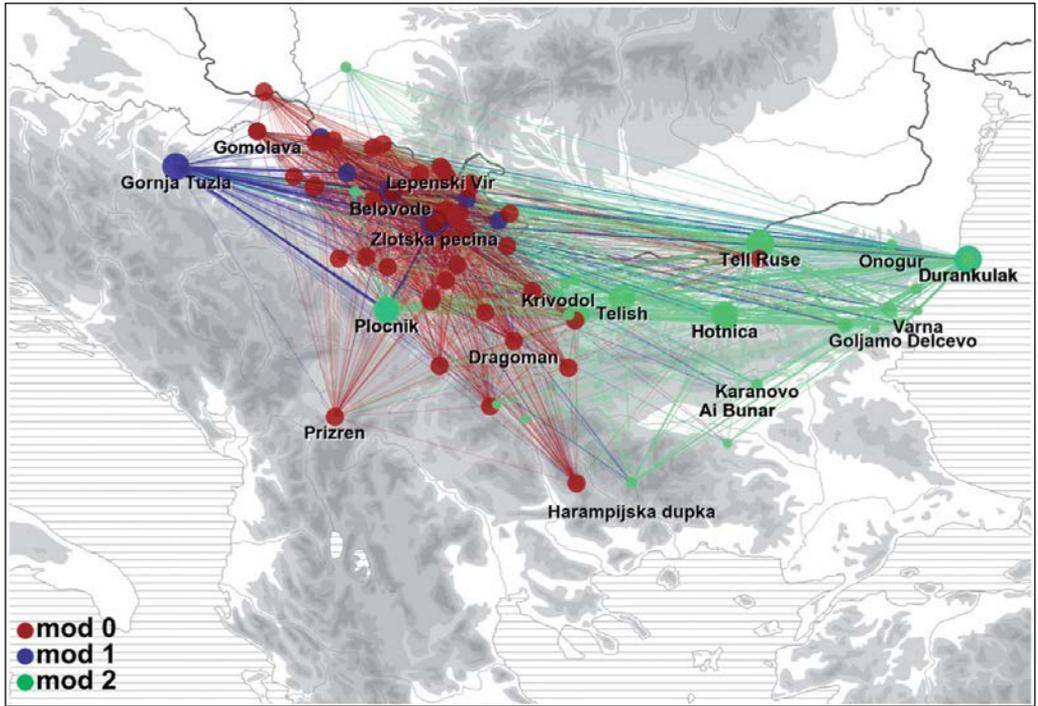
È infine indicato per quelle particolari analisi dei dati raggruppate sotto l’etichetta di “analisi delle reti sociali” (*Social Network Analysis*), applicate e diffuse ormai anche nell’ambito dell’archeologia⁴.

Il modello relazionale organizza invece i dati in tabelle (dette anche “relazioni”) a loro volta divise in righe (chiamate “tuple” o “ennuple”) ordinate in colonne (chiamate anche “attributi”).

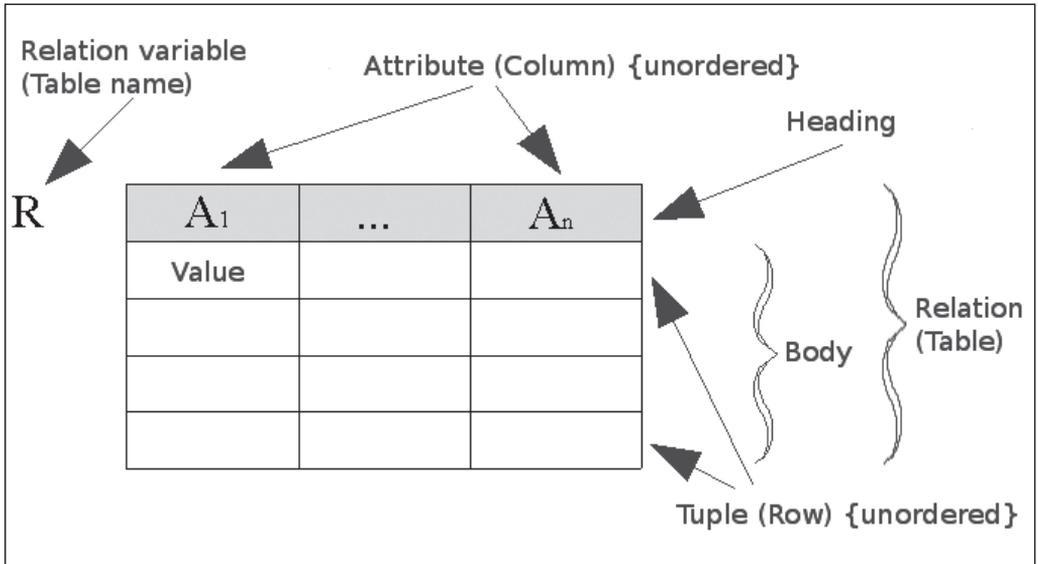


Rappresentazione in forma di grafo orientato dei dati in formato RDF relativi all’enunciato “Lewis Carol, il cui vero nome è Charles Lutwidge Dodgson è l’autore del libro *Alice nel paese delle meraviglie*”, visti in dettaglio nel capitolo precedente. Il grafo è stato creato in automatico utilizzando il servizio <http://www.ldf.fi/service/rdf-grapher>

⁴ Carl Knappett, *Network Analysis in Archaeology. New Approaches to Regional Interaction* (Oxford: Oxford University Press, 2013); Barbara J. Mills, «Social Network Analysis in Archaeology», *Annual Review of Anthropology* 46, n. 1 (2017): 379–97, <https://doi.org/10.1146/annurev-anthro-102116-041423>.



Analisi modulare di rete che rivela tre comunità che producono e scambiano rame nell'area balcanica tra il 6000 a.C e il 3200 a.C., da M. Radivojević e J. Grujić, «Community structure of copper supply networks in the prehistoric Balkans: An independent evaluation of the archaeological record from the 7th to the 4th millennium BC», *Journal of Complex Networks* 6, n. 1 (24 luglio 2017): 114, <https://doi.org/10.1093/comnet/cnx013>, fig. 4.



Schema esemplificativo del modello relazionale

Un esempio concreto può aiutare a chiarire questo aspetto meglio delle definizioni astratte.

nr	tipo	colore	composizione	consistenza
1	US	7.5BG 7/2	limoso	100
2	US	8.5BG 7/2	argilloso	20
3	USN	NULL	NULL	NULL
4	US	5BG 8/2	sabbioso	4

Esempio semplificato di una tabella di unità stratigrafiche

Quella nell'esempio qui sopra è una tabella molto semplice che elenca le unità stratigrafiche, le descrive usando una serie di attributi predefiniti (numero, tipo, colore, composizione, consistenza) e definisce delle regole molto precise riguardo ai valori che questi attributi possono avere. Basta una rapida occhiata (e un po' di esperienza) per capire in modo intuitivo che in questo progetto:

- le unità stratigrafiche vengono identificate da un valore numerico unico;
- la tipologia è una sigla a scelta tra US e USN (unità stratigrafica negativa);
- per il colore dello strato viene usata la codifica del sistema Munsell⁵;
- per la composizione una lista di stringhe di valori convenzionali, selezionati da un vocabolario predefinito;
- per la consistenza un sistema di valori numerici per i quali possiamo intuire la presenza di una scala graduata, ma da cui non si riesce a capire chiaramente la corrispondenza con il dato reale.

Questa serie di regole, fondamentali per organizzare coerentemente il lavoro di un cantiere o di una ricerca, possono essere codificate in maniera

⁵ Il sistema di colori Munsell è un codifica basata sui tre parametri di tonalità (*hue*), luminosità (*value* o *lightness*) e saturazione (*chroma* o *purity*), adottato dal Dipartimento dell'Agricoltura degli Stati Uniti come standard ufficiale per la descrizione dei colori del suolo. Il *Munsell soil color charts* è una pubblicazione cartacea che permette di confrontare facilmente il colore di un campione di terreno con una serie di colori di riferimento, identificati da stringhe alfanumeriche. Si tratta di un vocabolario condiviso indispensabile nelle operazioni sul campo per un descrizione non equivoca del colore dei suoli o anche degli impasti ceramici.

rigida nel modello relazionale, che può garantire in questo modo un controllo formale molto efficiente dei dati immessi.

Una interessante considerazione è che la tabella (o relazione) può essere descritta come un insieme di enunciati semplici, aventi ancora una volta una forma tripartita di soggetto, predicato e oggetto. Ogni colonna rappresenta una serie di caratteristiche fisse, che ogni oggetto può avere. La basilare tabella mostrata sopra, secondo quanto definito dal modello relazionale, è leggibile anche come la seguente lista di predicati, dalla struttura semplicissima:

```
Esiste una US che ha numero "1",  
  che ha colore "7.5BG 7/2"  
  che ha composizione "limoso"  
  che ha consistenza 100.
```

Ogni attributo costituisce una relazione elementare e la collezione di tutti gli attributi compone la tabella. Normalmente i *database* relazionali prevedono una serie di controlli sui dati in immissione, importanti per garantire la consistenza delle informazioni; una rigida pianificazione di tali controlli garantisce una migliore qualità del dato e più efficienti possibilità di interrogazione. Un controllo obbligatorio in tutti i principali programmi di *database* relazionali è costituito dalla tipologia del campo. Al momento della creazione della tabella è necessario infatti stabilire che tipo di informazioni ogni campo dovrà contenere: stringhe (e quanto lunghe), numeri interi (e quanto lunghi), numeri decimali (quanto lunghi e con quante cifre dopo lo zero), geometrie (e che tipo di codifica usare), dati binari, e così via. È fondamentale quindi avere un'idea molto chiara e puntuale dei dati che si andranno a trattare già al momento della progettazione del futuro *database*, possibile solo se si dispone di un'ottima conoscenza dell'ambito puntuale di ricerca e delle domande che lo studio si pone.

È poi possibile impostare un'altra serie di controlli a garanzia dell'integrità e dell'interrogabilità del dato: si è già visto, per esempio, che è utile per la corretta registrazione delle unità stratigrafiche stabilire una lista di valori per il campo *composizione*, affinché sia impedito al compilatore di inserire valori non previsti o errati. Abbiamo già visto come la condivisione di vocabolari permette una maggiore interoperabilità anche tra contesti e siti diversi: i vocabolari sono dunque una risorsa importante, ma costituiscono anche un limite oggettivo, in quanto stabiliscono una griglia rigida di opzioni dalle quali non è possibile discostarsi.

Non c'è una regola specifica e universalmente valida: è l'esperienza stessa,

la buona conoscenza del contesto che si sta indagando e della natura delle domande che la ricerca si pone a guidare verso la soluzione migliore, in equilibrio tra l'estremo formalismo – come vorrebbe la teoria informatica – e un eccessivo apporto interpretativo soggettivo nell'apparato descrittivo, che renderebbe vano ogni tentativo di definire raggruppamenti di dati o di individuare modelli ricorrenti.

Inoltre, le relazioni non avvengono solo all'interno di una stessa tabella, ma anche – e soprattutto – tra tabelle diverse. Un esempio molto semplice è quello dei reperti rinvenuti nelle diverse unità stratigrafica. Dal momento che le informazioni sui reperti archeologici sono sostanzialmente diverse da quelle sulle unità stratigrafiche, è necessario destinare loro una tabella separata, dove salvare il numero di inventario dato al reperto sul campo o in laboratorio, il numero di registro del magazzino, nonché una serie di altri dati utili alla ricerca archeologica. Perché ci possa essere un collegamento tra unità stratigrafiche e reperti in esse rinvenuti, è necessario segnalare per ogni reperto l'unità stratigrafica di provenienza:

inventario	us	magazzino	tipologia	...
1	2	2019/505	contenitore ceramico	
2	2	2019/506	contenitore ceramico	
3	2	2019/507	moneta	
4	4	2019/508	carbone	

Esempio semplificato di una tabella di reperti archeologici

Sia nella tabella us che in quella reperti i campi nr e inventario ricoprono un ruolo fondamentale, poiché rappresentano gli identificativi primari degli oggetti descritti e in quanto tali sono unici: ci si riferisce a una us con il suo numero, a un reperto archeologico con il suo numero di inventario, ed entrambi questi valori devono essere univoci, pena un sostanziale blocco dell'intero sistema. Se due unità stratigrafiche ricevono lo stesso nome, infatti, il contesto di riferimento diventa ambiguo, e la documentazione prodotta perde di significato.

Questi campi speciali, destinati a contenere gli identificativi di ogni elemento che popola il *database*, si chiamano chiavi primarie (*primary key* o PK) e sono un elemento necessario in ogni tabella di un *database* rela-

zionale. È buona norma usare un valore di tipo numerico e attivare una funzione di assegnazione automatica e progressiva. Quasi tutti i *database* relazionali (o per lo meno quelli di largo utilizzo) dispongono di strumenti interni per la gestione della chiave primaria e permettono la generazione automatica di valori univoci e incrementali non modificabili dall'utente. Nel nostro esempio, se si desidera che il *database* assegni automaticamente il nome delle unità stratigrafiche, è consigliabile rendere il campo *nr* una chiave primaria; altrimenti, se si ha la necessità di inserire manualmente il numero delle unità stratigrafiche (perché assegnato precedentemente sul campo, in forma non consecutiva, ecc.) allora è necessario aggiungere alla tabella un altro campo che svolga la funzione della chiave primaria. Si avrebbero in questo caso due identificativi, uno per uso interno del *database* e uno per l'uso sul campo e di studio.

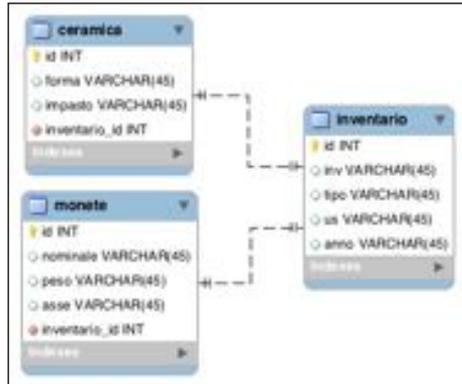
Per poter collegare due tabelle, oltre alle chiavi primarie bisogna anche definire quella che viene chiamata una chiave esterna (*foreign key* o FK). La regola vuole che la chiave esterna della seconda tabella contenga il riferimento alla chiave primaria della prima, alla quale questa deve collegarsi. Nel nostro elementare *database* di esempio, il campo *us* della tabella *reparti* (*reparti.us*)⁶ è a tutti gli effetti una chiave esterna, in quanto contiene la chiave primaria della tabella *us*. Queste due tabelle sono collegate secondo il modello chiamato “uno-a-molti” (o “1-n”) poiché a un elemento di una tabella (*us*) corrispondono molti elementi nella seconda (*reparti*). Esistono vari tipi di relazioni, che coprono la stragrande maggioranza delle casistiche che si riscontrano e che ci accingiamo a esaminare.

Relazione “uno-a-uno” (“1-1”)

La relazione “uno-a-uno” avviene quando a un *record* della prima tabella non corrisponde nessuno o, corrisponde un solo *record* della seconda tabella. Entrambe le tabelle hanno una propria chiave primaria e una delle due ha una chiave esterna che contiene la chiave primaria dell'altra; è su questo campo che viene normalmente eseguito un controllo di unicità. In archeologia è comune trovare questo tipo di relazione nella descrizione di entità complesse e fortemente eterogenee come possono essere i reperti archeologici, che hanno bisogno di essere considerati come unità dal punto di vista inventariale ma che necessitano di norma di protocolli descrittivi molto

⁶ Di norma un campo viene chiamato semplicemente con il suo nome (es. *us*), più comunemente l'identificativo viene posposto al nome della tabella, separando i due da un punto (es. *reparti.us*). In questo modo è possibile riferirsi in modo univoco a campi di tabelle diverse.

Esempio di relazione 1:1 tra la tabella inventario e le tabelle ceramica e monete. Il collegamento è stabilito tra la chiave primaria inventario.id e le rispettive chiavi esterne ceramica.inventario_id e moneta.inventario_id. Su queste ultime due è attivato un controllo di unicità, per evitare errori di inserimento, dal momento che a un unico numero di inventario non può corrispondere più di un elemento nelle tabelle collegate.

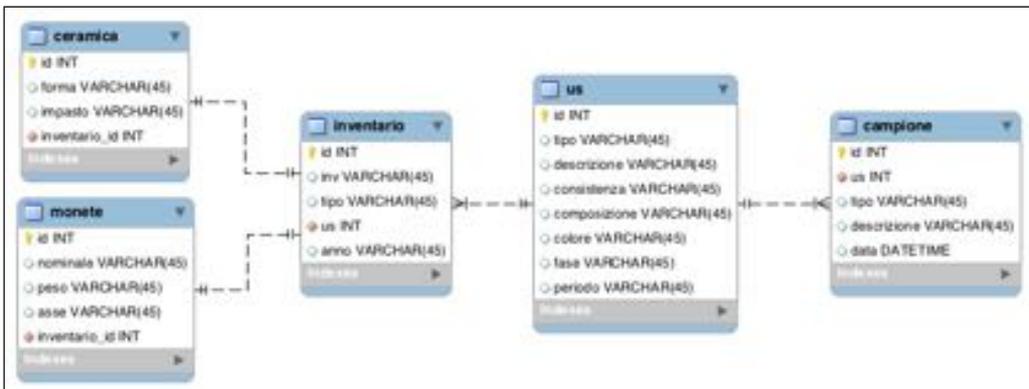


Esempio di relazioni 1-a-molti tra la tabella us e le tabelle inventario e campione. Il PK us.id è riferito dai FK inventario.us e campione.us, che permettono naturalmente valori duplicati.

differenziati a seconda della tipologia. In questi casi si crea una tabella chiamata inventario che tiene conto in modo centralizzato di numero di inventario, tipologia di reperto, contesto topografico di rinvenimento, data di rinvenimento, persone responsabili, ecc. A questa vengono affiancate tabelle con strutture dedicate che descrivono in maniera indipendente e specifica le varie tipologie di reperto, come la ceramica, le monete, i resti organici, i frammenti architettonici, ecc.

Relazione “uno-a-molti” (“1-n”)

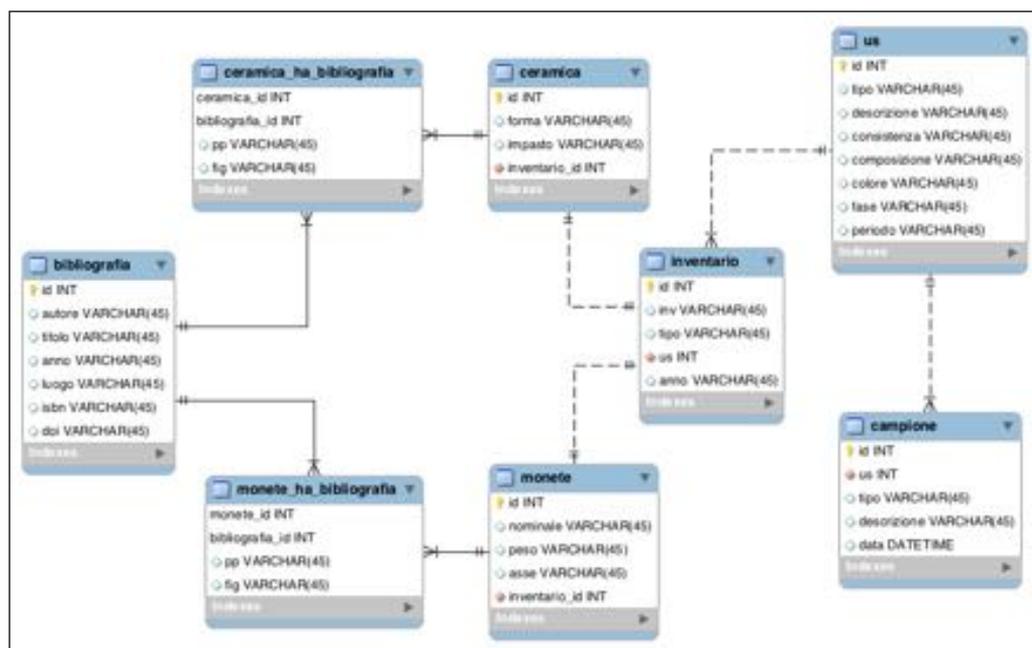
La relazione “uno-a-molti” avviene quando a un elemento della prima tabella non corrisponde nessuno, o corrispondono uno o più record nella seconda. Dal punto di vista della struttura del database non ci sono differenze rispetto alla relazione “uno-a-uno”: ciascuna tabella ha la sua chiave primaria e una ha una chiave esterna popolata con la chiave primaria della prima. È il caso più frequente di collegamento, come abbiamo già avuto modo di approfondire, che viene ampiamente usato in archeologia, ad esempio per definire le vari fasi cronologiche di un sito (un sito ha molte



fasi), i reperti di uno strato (una US ha 0/1/molti reperti), le analisi di laboratorio, gli eventuali interventi di restauro eseguiti su un reperto, ecc.

Relazione “multi-a-molti” (“n-n”)

Nella relazione “multi-a-molti” a un elemento della prima tabella può non corrispondere nessuno, o possono corrispondere uno o molti elementi nella seconda tabella; allo stesso tempo a un *record* della seconda tabella possono corrispondere nessuno, uno o molti *record* nella prima. È questo il caso tipico dei *database* bibliografici: prendendo ad esempio la bibliografia riguardante i reperti archeologici, una forma ceramica può avere riscontri in più di una pubblicazione (*record* bibliografico) e allo stesso tempo una pubblicazione può trattare diverse forme ceramiche. Stesso discorso può essere fatto per i collegamenti tra schedature di monumenti, edifici e tipologie di decorazione architettonica: un edificio può fare uso (simultaneo o diacronico) di diverse tipologie di decorazione architettonica (capitelli, basi di colonna/pilastro), e una certa tipologia di decorazione architettonica



Schema relazionale, che comprende anche i collegamenti n-n, tra monete e bibliografia e tra ceramica e bibliografia. Nel primo caso, l'associazione è data dalla tabella monete_ha_bibliografia, composta da una chiave primaria (id), da due chiavi esterne (monete_id e bibliografia_id) e da due altri campi che aggiungono informazioni aggiuntive per ogni relazione, in particolare le pagine in cui l'oggetto è menzionato (pp) e l'eventuale numero di figura (fig). Questi due campi non sono strettamente necessari dal punto di vista tecnico per garantire il collegamento, ma lo sono dal punto di vista del sistema di citazione bibliografica.

ca (capitello, base di colonna/pilastro) può essere impiegata in monumenti diversi, distinti nello spazio e nel tempo. Gran parte del lavoro dell'archeologo, e dell'umanista in generale, è infatti basato sul confronto, che nei termini di questa riflessione equivale all'individuazione e la registrazione di relazioni "n-n". Dal punto di vista del *database* relazionale è necessario introdurre nello schema fin qui delineato una tabella aggiuntiva, chiamata tabella di associazione (*association table*, *join table*, *junction table*, ecc.), la quale registra le istanze di ogni connessione. Generalmente questa tabella intermedia ha una struttura molto semplice, di soli tre campi: la propria chiave primaria, una chiave esterna relativa alla prima tabella e una seconda chiave esterna relativa alla seconda tabella. Normalmente queste connessioni non sono qualificate e il loro significato è dichiarato esternamente o è esplicitato dalla struttura generale. Una connessione tra una tabella *edifici* e una tabella *tipologie_di_frammenti_architettonici*, per esempio, non ha bisogno di spiegazioni articolate: il significato di ogni collegamento è evidente in entrambe le direzioni. Meno chiaro potrebbe essere il collegamento tra un *contenitore_ceramico* e un *luogo*, in quanto il luogo potrebbe essere quello di rinvenimento, di attestazione, di produzione, di deposito, ecc.: si tratta cioè di un'informazione che, se non viene specificato per ciascun collegamento nella tabella di associazione, lascia la connessione muta. La scelta tra creare collegamenti etichettati e qualificati o lasciare che il significato venga esplicitato dalla struttura generale dipende naturalmente dalle necessità concrete di ogni progetto di ricerca: entrambi i casi possono comunque essere facilmente codificati mediante il modello relazionale.

Un linguaggio unico di interrogazione: SQL, e i suoi dialetti

La rapida diffusione e l'ampia adozione del modello relazionale è legata al fatto che la comparsa dei *database* relazionali abbia portato in parallelo allo sviluppo di un linguaggio unificato di interrogazione, lo Structured Query Language o SQL. Anche se il nome fa riferimento solo all'interrogazione (*query*), l'SQL è una sorta di "coltellino svizzero" delle banche dati relazionali, in quanto viene usato anche per la creazione delle strutture, per l'impostazione della politica degli accessi, per la definizione dei controlli sui dati, per l'inserimento e la modifica degli stessi e infine anche per il loro recupero e lettura. Nonostante l'SQL sia uno standard ANSI (American National Standard Institute) e ISO (International Organization for Standardization), e quindi in linea di principio utilizzabile indipenden-

temente dal gestore di *database*, la prassi dei produttori di *software* è quella di far uso di “dialetti” SQL, che in un modo o nell’altro deviano dallo standard vero e proprio. Si tratta di scelte commerciali, oppure di esigenze specifiche, per cui un produttore può sviluppare e implementare alcune funzioni non previste nello standard e tralasciarne altre che non reputa utili alla sua politica commerciale o che non sono significative per la sua utenza di base. Questo fa sì che, di norma, non ci sia una perfetta compatibilità tra le varie versioni del “dialetto” e che alcune *query* che funzionano bene su una piattaforma possano invece generare errori o non essere supportate su un’altra. Si tratta però in genere di funzionalità avanzate e specialistiche, che non vanno a incidere sul linguaggio di base, largamente adottato in maniera uniforme: solitamente un utente medio non nota differenze sostanziali tra un *software* e l’altro, mentre uno avanzato potrebbe dover ricorrere all’utilizzo di forme dialettali specifiche per eseguire compiti particolari.

La conoscenza del linguaggio SQL è superflua per un utente di base, in quanto le interfacce grafiche di gestione dei *database* coprono quasi tutte le azioni fondamentali di creazione, gestione e manutenzione della struttura delle banche dati, nonché la maggior parte delle operazioni sui dati. Diventa invece indispensabile quando si cerca di agire in maniera più approfondita o si ha a che fare con strutture complesse di dati, che richiedono una maggiore consapevolezza delle operazioni da svolgere. La conoscenza dell’SQL è peraltro requisito fondamentale (almeno per quanto riguarda la lettura dei dati) per chi si trova a gestire quei *database*, particolari e specifici che sono i Sistemi Informativi Geografici (GIS): questi ultimi, infatti, pur disponendo (talvolta) di interfacce grafiche che aiutano a compilare e validare gli enunciati SQL, richiedono comunque all’operatore una conoscenza di base di questo linguaggio per poter interagire con i dati.

L’SQL è un linguaggio formale facilmente leggibile, basato sull’uso di alcune parole chiave espresse in lingua inglese, che definiscono l’ambito generale di intervento, ciascuna delle quali prevede una sintassi specifica⁷. Si tratta di un linguaggio relativamente semplice, ma piuttosto potente. Per le attività di *routine* di un archeologo sono necessari pochi concetti base, soprattutto se le banche dati vengono affiancate alle piattaforme GIS, che di norma implementano per l’interrogazione interfacce grafiche poco dettagliate. Quando la complessità e soprattutto la quantità dei dati aumenta

⁷ Esistono molti manuali e guide allo SQL. In italiano, si veda Marco Ferrero, *SQL: la guida tascabile al linguaggio di interrogazione dei database* (Milano: Apogeo, 2004); in lingua inglese Thomas Nield, *Getting Started with SQL* (Sebastopol: O’Reilly Media, 2016); Alan Beaulieu, *Learning SQL* (Sebastopol: O’Reilly Media, 2009).

in modo importante, sono richieste competenze più specializzate. Va detto, però, che raramente i dati archeologici raggiungono livelli significativi di complessità o numeri molto grandi (nell'ordine delle centinaia di migliaia di *record* per tabella).

Alcune delle domande più frequenti che si pongono quando ci si trova davanti all'esigenza di progettare un sistema informativo archeologico sono:

- quali dati codificare e salvare nel *database* e quali tralasciare. Bisogna, infatti, essere consapevoli del fatto che, per quanto attenta possa essere la nostra ricerca sul campo, si raccoglie e registra solo parte dei dati che in termini assoluti è possibile recuperare. Cosa viene registrato e cosa trascurato dipende da una molteplicità di fattori (tempo a disposizione, disponibilità di strumenti, formazione dell'operatore, interesse generale dell'intervento, peso informativo dei dati tralasciati, ecc.); è importante accertarsi che il sistema informativo in uso rispecchi gli obiettivi dell'intervento e che si adegui alle domande di ricerca e non viceversa.
- Qual'è il metodo migliore di codifica dei singoli dati, ovvero se un certo campo deve essere di tipo numerico, booleano (che accetta solo un valore tra sì e no), testuale, binario, ecc. Qual è il formato migliore per registrare informazioni particolari quali le date, le geometrie, le cronologie, ecc. Gli RDBMS ci aiutano, mettendoci a disposizione diversi *data types* tra i quali poter scegliere, ma questa scelta non è mai di tipo tecnico. La struttura dipende in massimo grado dagli obiettivi della ricerca, che determinano le interrogazioni che sarà possibile fare. Ad esempio, se si prevede che il nostro *dataset* debba fare nel futuro da base ad analisi statistiche e quantitative, è di fondamentale importanza cercare di rappresentare in forma numerica anche informazioni che normalmente si rappresenterebbe in forma testuale.
- Quali vocabolari usare: un punto fondamentale, che fa sì che un *dataset* possa essere interrogato in maniera efficiente. Il mancato uso dei vocabolari, o un loro utilizzo troppo frammentario e personalizzato, rende il nostro *dataset* inutile, un ricco contenitore dal quale non è però possibile estrarre modelli ricorrenti (*pattern*). Il segreto è di trovare un buon compromesso tra una rappresentazione troppo astratta (vocabolari semplici e rigidi) e una rappresentazione troppo poco astratta (assenza di vocabolari a favore di una più verbosa e puntuale descrizione della realtà fisica). La scelta dei vocabolari determina, inoltre, il grado di interoperabilità del nostro sistema in-

formativo con altri *dataset*.

- Quali sistemi di validazione dei dati attivare: è un nodo centrale, che solitamente marca un confine molto netto tra il punto di vista di un informatico e quello di un archeologo. Dal punto di vista informatico, per avere un *database* coerente ed efficiente tutti i dati devono essere validati secondo politiche molto rigide, che ne verifichino la validità formale e semantica; dal punto di vista dell'archeologo, ogni imposizione strutturale è un limite oggettivo a una rappresentazione fedele del contesto osservato. Entrambi hanno ragione, entrambi hanno torto, e ancora una volta non esiste una regola universale. Ogni sistema informativo stabilisce la propria, che dovrebbe essere fondata su un ragionamento chiaro e soprattutto dovrebbe essere condiviso anche con i fruitori esterni ed esplicitato nella documentazione relativa.

L'elenco potrebbe protrarsi per molte pagine, ma ci si augura sia chiaro da queste righe che quando si tratta di indagine archeologica bisogna avere sempre ben presente che i dati che confluiscono nei nostri *database* sono sempre il risultato di più livelli di interpretazione personale. Anche informazioni apparentemente "oggettive" come il colore di uno strato, i suoi limiti topografici precisi, alcuni rapporti stratigrafici, l'appartenenza di un oggetto a una determinata classe, ecc. dipendono in gran parte dalla sensibilità e dall'esperienza dell'operatore/ricercatore. Non bisogna mai dimenticare che il contesto archeologico presenta una complessità molto più elevata di quanto la nostra documentazione sarà mai in grado di rappresentare. Un approccio eccessivamente scientifico, che pretenda di rappresentare in un sistema informativo la realtà in quanto tale, pone dunque delle serie questioni di metodo. Un approccio troppo generico, d'altro canto, che manchi di segnalare le specificità dei singoli elementi (associazioni, oggetti, dinamiche), rischia di appiattirne il potenziale informativo globale. Alla soglia degli anni Venti del ventunesimo secolo il migliore strumento tecnologico che abbiamo a disposizione per la comprensione di un contesto archeologico è – e rimane – la nostra capacità di osservare, capire e porre domande. Un buon sistema informativo è in grado di sintetizzare e assecondare questa capacità, senza la pretesa di poter sostituire in forma digitale la realtà concreta andata irrimediabilmente distrutta.

Non esistono, in conclusione, un sistema informativo o una struttura di *database* perfetti, ma tante possibili forme diverse che dovremmo fare lo sforzo di mettere in relazione. Le tecnologie semantiche ci insegnano che è giusto prenderci una grande libertà nell'analisi di un caso specifico, in-

ventarci, se non esistono, gli strumenti descrittivi e di analisi più adatti al nostro scopo, ma rimane fondamentale porre la nostra ricerca all'interno di una rete più ampia e sforzarci di collegarla ad altre iniziative, in modo da restituire un contesto più ampio e ricco al nostro oggetto di ricerca, che sia un saggio di scavo, un'area urbana, una classe di reperti, un'iconografia, una porzione di territorio, ecc.

Implementazione del modello relazionale. *Software e formati*

Nelle pagine precedenti si sono affrontate i temi del funzionamento del modello relazionale e del suo principale linguaggio di gestione (SQL). Si è anche accennato a quelle implementazioni del linguaggio SQL fatte dai produttori di *software*, diverse e mai pienamente compatibili l'una con l'altra, che abbiamo chiamato "dialetti". Le possibili alternative *software*, infatti, sono tante e la scelta dipende non tanto da questioni scientifiche quanto dalla scala del progetto e dal *budget* a disposizione. Neppure il fattore tecnico è un discriminante significativo: per quanto articolata e complessa possa essere una ricerca di tipo archeologico (e si potrebbe allargare il campo alle scienze umane in generale), dal punto di vista tecnico della progettazione e gestione di una banca dati, il programma gestionale di un negozio di ferramenta (con vendite in negozio e *store online*, gestione del magazzino, clienti e fornitori, fatturazione, pagamenti, ecc.) presenta mediamente una struttura più articolata di quella di un importante progetto internazionale. Un discorso analogo può essere fatto per la potenza di calcolo: è molto difficile che la mole e complessità di dati prodotti da una ricerca nel nostro ambito possa mettere in crisi un qualche prodotto *software* (sempre nell'ambito dei RDBMS), per quanto limitato esso possa essere.

Questo non vuol dire che creare una banca dati per una ricerca archeologica sia un compito facile o di poco conto, anzi è vero il contrario. Come si è già avuto modo di considerare, la tenace e tradizionale resistenza dei

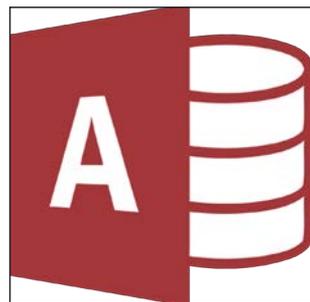
professionisti e ancor di più dei ricercatori dei Beni Culturali agli standard, l'alto grado di interpretazione personale che ogni forma di raccolta dati implica, la specificità e il particolare taglio di ricerca che ogni iniziativa adotta rendono particolarmente difficile la progettazione di sistemi informativi e banche dati.

Tradizionalmente, sono stati soprattutto due i *software* che si sono divisi, in modo tutto sommato pacifico, il mercato degli ultimi decenni, nell'ambito dell'archeologia. Filema-

Logo di
FileMaker
Pro



ker Pro, nato in ambiente Apple Macintosh, è stato adottato da un numero considerevole di studiosi per la sua estrema facilità di creazione e gestione dei *database*, dalla quasi totalità di coloro che erano abituati all'uso del Macintosh, ma anche da moltissimi utenti di Windows. Il secondo, diffusissimo prodotto, è Microsoft Access, distribuito dalla casa produttrice all'interno del suo pacchetto Office, fatto che gli ha garantito una presenza ubiqua. È stato, e per molti versi è tuttora, il gestore personale di *database* più diffuso, anche per la sua facilità di utilizzo. Si tratta di un gestore di *database* relazionali che permette di usare anche fonti di dati esterne, vale a dire che è possibile usare il *software* come interfaccia di gestione e consultazione di dati immagazzinati in altri formati, quali SQL Server o Oracle (o altri ancora) grazie all'interfaccia ODBC⁸.



Logo di
Microsoft
Access

In archeologia il successo indiscusso di MS Access è dovuto anche al fatto che il suo formato era perfettamente integrato nelle piattaforme GIS create dalla marca ESRI, cosa che permette di creare un ambiente ideale di lavoro, comprendendo quasi in un'unica piattaforma dati tabellari di tipo alfanumerico e dati geografici. Il limite più importante di questo *software* è quello di essere disponibile solo per ambiente Windows, cosa che ne preclude l'utilizzo, anche in sola lettura, in altri ambienti come Linux o MacOS. Alternative *open source*, come per esempio OpenOffice Base, pur essendo multipiattaforma, non hanno avuto di fatto ampia diffusione, pur non essendo mancate le sperimentazioni.



Logo di
LibreOffice
Base

Se questi strumenti rimangono tuttora molto diffusi, una decina di anni fa il paradigma è cambiato soprattutto grazie alla diffusione delle reti – di Internet e del World Wide Web – che hanno favorito e incentivato la diffusione del modello *client-server* e hanno in breve portato, come si è visto, a connettere al Web le banche dati.

Anche in archeologia è ormai molto diffusa la gestione centralizzata e da remoto delle banche dati. Questo modello da una parte allontana fisicamente l'utente dal *database* (il quale si può trovare in qualche *server* lontano

⁸ L' Open DataBase Connectivity (ODBC) è un'interfaccia per l'accesso ai sistemi di gestione di banche dati in maniera indipendente dal *software* utilizzato o dal sistema operativo. Si tratta dunque di uno strato intermedio tra la banca dati e l'applicazione che la gestisce, garantendone quindi la connessione. Questo tipo di soluzione permette l'utilizzo di una applicazione per diversi tipi di database dei quali sia disponibile l'interfaccia ODBC.

talvolta anche centinaia di chilometri dall'utilizzatore) e richiede una maggiore competenza informatica per la configurazione, dall'altra permette di allargare la base di coloro che vi possono avere accesso: non più il singolo studioso o ricercatore ma anche i suoi colleghi, collaboratori, studenti, o il pubblico generico in caso di banche dati ad accesso libero.

<i>Software</i>	<i>Server</i>	<i>Open source</i>	<i>Piattaforma</i>
Amazon Aurora	sì	no	(distribuito come servizio attraverso Amazon RDS)
FileMaker Pro	anche	no	MacOS, Windows
IBM DB2	sì	no	Linux, Unix, Windows
MS Access	anche	no	Windows
Microsoft SQL Server	sì	no	Windows, Linux
MySQL	sì	sì	Linux, Unix, Windows
MariaDB	sì	sì	Linux, Unix, Windows
OpenOffice.org Base / Libre Office Base / HSQLDB	no	sì	Linux, Unix, Windows
Oracle RDBMS	sì	no	Linux, Windows
PostgreSQL	sì	sì	Linux, Unix, Windows
SQLite	anche	do - minio pub - blico	Tutti

Tabella di confronto di alcuni dei DRBMS più diffusi, elencati in ordine alfabetico

Per la scelta di un determinato *software*, dunque, è necessario definire la metodologia di lavoro, ovvero stabilire se si abbia bisogno di un RDBMS a uso personale oppure se l'utilizzo che se ne intende fare sia condiviso. Nel primo caso la scelta tra FileMaker Pro, MS Access, Base o SQLite è detta-

ta dalla preferenza individuale e determinata principalmente dall'estetica dell'interfaccia o dalla piattaforma utilizzata (Access, come si è detto, è disponibile solo per Windows). FileMaker Pro e Access presentano sicuramente interfacce molto curate e strumenti di gestione dei dati di grandissima semplicità e molto intuitive.

Se invece il futuro *database* deve essere usato da un numero maggiore di utenti, è necessario fare una scelta *server*. Gli strumenti elencati sopra sono tutti molto diffusi e sicuramente adatti a gestire dati di rilevante complessità. Il supporto per la piattaforma è infatti quasi irrilevante: dovendo attivare il servizio su un *server* remoto, questo sarà quasi sicuramente Linux, e comunque in generale tutti i *database server* più diffusi supportano anche Unix, MacOS e Windows. I *database server*, essendo progettati per l'uso condiviso in rete, implementano strumenti avanzati e sicuri per una gestione efficiente e molto dettagliata delle politiche di accesso ai dati, garantendo in genere un ottimo livello di sicurezza.

I *database server* vengono distribuiti senza interfacce grafiche o con interfacce grafiche per la sola gestione e non per il popolamento e l'utilizzo da parte degli utenti generici; per questo, sono disponibili una serie di soluzioni *software*, a vario costo, che permettono di costruire maschere di inserimento e di gestione dei dati di facile usabilità. Anche alcuni *software desktop* come MS Access o LibreOffice (o OpenOffice) Base possono essere usati come interfacce per accedere ai dati gestiti da un *database server*. Di norma però, anche in ambito archeologico, la scelta più comune è quella di crearsi di volta in volta le proprie interfacce di gestione, anche per adattarsi in maniera ottimale alla struttura sottostante dei dati.

Questa tipologia di configurazione (*database server* + interfaccia di gestione / pubblicazione creata *ad hoc*) si è molto diffusa con l'avanzamento del Web dinamico e con l'impiego su larga scala di gestori di contenuti per la pubblicazione di siti e portali web (CMS). In fondo, queste applicazioni – che permettono la gestione di archivi web di dati archeologici complessi e molto personalizzati – non sono altro che dei CMS fortemente specializzati nella raccolta e visualizzazione dei dati. Spesso includono forme di esposizione non testuali (immagini, disegni, *matrix*), visualizzazioni aggregate e funzioni statistiche (grafici di vario genere) che aiutano la lettura delle informazioni. Talvolta includono interfacce geografiche, fondamentali in gran parte della ricerca archeologica. Ci sono stati diversi tentativi di creare uno strumento che fosse adattabile a contesti e progetti diversi, ma l'ostacolo principale è quello della struttura estremamente fluida della ricerca archeologica, difficilissima da costringere entro schemi rigidi che si

possano universalmente esportare e modellare *ad hoc* con facilità.

Ultima nota da tenere in conto quando si effettua una scelta tecnologica è la necessità di considerare non solo il presente o futuro prossimo, dove l'efficienza del sistema, la facilità di implementazione, i costi e l'usabilità dell'interfaccia sono i requisiti fondamentali: è indispensabile, soprattutto per chi si occupa di ricostruire il passato a partire dalle tracce materiali che di esso sopravvivono, pensare anche al futuro e fare in modo che la nostra attività lasci un'eredità quanto più cospicua e fruibile possibile.

Queste considerazioni assumono un significato ancora maggiore nel mondo della tecnologia, in cui la velocità di cambiamento è rapidissima e il futuro più lontano si colloca non oltre 2-5 anni dal presente: la scelta tecnologica deve prevedere una stabilità nel tempo e una durezza che nessun produttore di *software* o azienda, per quanto grande e importante sia, può attualmente garantire. È quindi necessario pensare di offrire ai dati una vita più lunga di quella del *software* con il quale oggi li gestiamo ed ecco perché è fondamentale – nella scelta di un *software* – la politica commerciale della casa che lo produce. Alcune *software house*, infatti, aggiornano annualmente i formati di codifica dei *file* che gestiscono, introducendo incompatibilità con le versioni precedenti, in modo da tenere vivo un mercato in cui piazzare periodicamente la versione più recente o l'aggiornamento. In una tale dinamica – del tutto legittima per un'azienda il cui scopo primario è la vendita, ma deleteria per chi invece si preoccupa della stabilità e accessibilità dei dati – le alternative *open source* offrono una garanzia maggiore. Non essendo la vendita del prodotto una priorità, viene a mancare l'eterna corsa all'aggiornamento del formato, che in questo caso avviene solo e unicamente per migliorare le funzionalità e l'usabilità del *software*. Inoltre, un *software open source* utilizza formati di codifica aperti, le cui specifiche sono pubbliche, ben note e non coperte da vincoli commerciali di utilizzo o diffusione. Chiunque sia capace e interessato a farlo, può sviluppare un programma che legge e scrive il formato SQLite⁹, inventato da Dwayne Richard Hipp e subito rilasciato al pubblico dominio, ovvero regalato alla comunità senza alcun tipo di *copyright* o restrizione. Per questo motivo SQLite è divenuto in breve tempo il *database* più diffuso in assoluto. La stessa cosa non succede con il formato di MS Access: anche chi ne possiede le competenze, non può liberamente creare un programma in grado di leggere e scrivere in questo formato, non perché sia tecnicamente difficile, ma perché per farlo è necessario avere accesso

⁹ <https://www.sqlite.org/index.html>.

alla documentazione del formato – non pubblica – e pagare i diritti di utilizzo alla casa madre, ovvero Microsoft.

Non è però sempre possibile usare solo e unicamente *software* e formati aperti, e ciò per motivi molto diversi. In tal caso rimane un imperativo morale (che lentamente sta diventando anche un imperativo legale, nel caso in cui la ricerca che produce i dati sia finanziata con fondi pubblici) cercare di avere un piano di recupero dei dati in grado di esportarli automaticamente e a intervalli regolari verso formati aperti, semplici da leggere e processare anche con strumenti terzi. L'obiettivo, quindi, è di mettere sempre in primo piano i dati, quelli che si raccolgono con la faticosa e costosa ricerca sul campo o quelli che si producono con il confronto, lo studio e il ragionamento, e di garantire loro la maggiore longevità possibile. Nel caso delle banche dati relazionali, quasi tutti i *software* fin qui visti prevedono procedure per l'esportazione dei dati in formati aperti quali il CSV¹⁰ oppure lo stesso SQL. Si tratta di formati testuali leggibili con facilità anche da un umano (a differenza di un formato binario) usando *software* basilari e universalmente diffusi. La semplicità di questi formati ne garantisce dunque le leggibilità e l'accesso negli anni a venire, al di là dei *software* con i quali sono stati prodotti.

```
"nr","tipo","colore","composizione","consistenza"  
1,"US","7.5BG 7/2","limoso",100  
2,"US","8.5BG 7/2","argilloso",20  
3,"USN","NULL","NULL","NULL"  
4,"US","5BG 8/2","sabbioso",4
```

Esempio di esportazione in formato CSV della tabella esemplificativa relativa alle unità stratigrafiche già vista in precedenza

Database e Web, cambiamenti e prospettive

Si è visto sopra, con la breve rassegna dei *database server*, che Internet e il World Wide Web hanno introdotto profondi cambiamenti nel mondo delle banche dati, facendoli diventare il cuore pulsante del WWW, anche

¹⁰ Comma Separated Values (valori separati da virgole), un formato di testo per dati tabellari: i campi vengono separati da caratteri di controllo (le virgole, come da nome del formato, ma anche punto e virgola, tab, o qualsiasi altro carattere) e le righe da a-capo, <https://tools.ietf.org/html/rfc4180>. Si tratta quindi di un formato facilmente leggibile mediante un qualsiasi editor di testo e da tutti i principali *software* di *database*. Normalmente viene visualizzato con programmi di fogli di calcolo (MS Excel o LibreOffice/OpenOffice Calc).

se nella stragrande maggioranza dei casi i loro dati non sono direttamente accessibili sulla rete, ma solo attraverso il filtro di siti e applicazioni web. Nel mondo dell'archeologia la riservatezza e la restrizione degli accessi è stata – ed è tuttora – un prerequisito fondamentale e una *conditio sine qua non* per l'utilizzo delle tecnologie di rete quale mezzo di *data entry*, gestione o archiviazione dei dati della ricerca, fatto questo che ha importanti implicazioni sull'accessibilità dei dati¹¹.

Nell'elenco dei *software* riportato nel capitolo precedente al primo posto è stato menzionato Amazon Relational Database Service (RDS), un prodotto molto diverso dagli altri. Non si tratta di un “*software*” nel senso tradizionale del termine, del quale sia possibile comprare la licenza (se commerciale), e che si possa scaricare, installare, usare ed eventualmente aggiornare. Si tratta di un programma offerto come servizio (*Software as a Service*, *SaaS*, paradigma tipico del cosiddetto Web 2.0), acquistabile attraverso un abbonamento. L'utente finale non si deve preoccupare di configurare o mantenere un'infrastruttura, ma può semplicemente utilizzare un servizio già fortemente ottimizzato. Si tratta di un modello che sta conoscendo un rapidissimo sviluppo e diffusione e che investe quasi tutti i settori: per menzionare quelli che un archeologo utilizza più spesso, basti pensare alla *suite* da ufficio di Google (G Suite)¹², che permette non solo di estendere la memoria del proprio PC con risorse aggiuntive in rete, ma anche di usare programmi di videoscrittura, fogli di calcolo, presentazioni, disegni, moduli per la raccolta dati e molto altro. Lo stesso si può dire per il disegno assistito da computer (CAD)¹³ o anche per settori specifici come ad esempio la fotogrammetria¹⁴. All'alta qualità delle funzionalità offerte corrisponde, dietro le quinte, una potenza di calcolo che difficilmente può essere eguagliata dalle risorse di un singolo ricercatore, di un progetto o di un'istituzione di piccole-medie dimensioni. Il *cloud computing*, o nuvola informatica, sta ultimamente conquistando anche grandi settori dell'industria, che basano il loro modello di *business* sui servizi Web e che fino a poco fa erano costretti a investire risorse considerevoli per sviluppare strumenti informatici propri. Recentissimamente stanno nascendo anche

¹¹ Si veda più diffusamente il capitolo 5.

¹² <https://gsuite.google.it/>.

¹³ Autodesk, la casa produttrice di AutoCAD, il più importante *software* commerciale di disegno assistito, ha da qualche anno lanciato una versione web, <https://web.autocad.com>.

¹⁴ L'utilizzatissimo servizio web Autodesk 123D Catch è stato sospeso, come anche Hypr3d. È ancora attivo Arc3D, <https://www.arc3d.be/>.

soluzioni *cloud* personalizzate specificatamente per il settore pubblico¹⁵. In questo contesto è facile prevedere come servizi di banche dati pronte all'uso (*SaaS*), come quello di Amazon e altri, saranno sempre più interessanti per i bassi (spesso nulli) costi di avvio, i limitati costi di gestione, gli alti livelli di sicurezza e infine la cessata necessità di manutenzione e aggiornamento. Il prezzo da pagare (oltre ai costi di abbonamento quando previsti) è la totale rinuncia al possesso fisico dei dati, che vengono affidati a un fornitore (*provider*) terzo, con il quale si istaura un fondamentale rapporto di fiducia¹⁶. Si tratta di un tema molto delicato, di cui si sta occupando anche il Legislatore, ponendo vincoli e restrizioni che talvolta rendono particolarmente difficili e costose le fasi iniziali di avvio. L'auspicio è che lo Stato, attraverso il Ministero per i Beni e le Attività Culturali (MiBAC) e quello dell'Istruzione, dell'Università e della Ricerca (MIUR), possa sviluppare e quindi offrire come servizio infrastrutture tecnologiche basate sul modello della nuvola e informatiche, a norma della legge vigente, per la gestione e la sperimentazione di archivi archeologici *online*.¹⁷ Fino ad allora, sarà onere dei singoli ricercatori mettere in atto le politiche più efficaci per la protezione e la salvaguardia dei propri dati digitali, ma la "fame" di risorse digitali, tecnologie e servizi di rete aumenta giornalmente, così come le spese per la messa a punto e la manutenzione di sistemi informativi centralizzati e le risorse di *cloud computing* all'interno di un quadro legislativo – giustamente – molto rigido. Mentre aumentano le dimensioni, le forme e i formati della nostra produzione digitale e mentre i rischi sulla sicurezza innalzano giornalmente il livello degli standard di garanzia che

¹⁵ Si vedano, per esempio, le soluzioni proposte dalla Microsoft specificatamente per l'amministrazione pubblica statunitense (<https://azure.microsoft.com/it-it/global-infrastructure/government/>). In Italia è l'Agenzia per l'Italia Digitale (<https://www.agid.gov.it/>) della Presidenza del Consiglio dei Ministri l'ente che detta le linee guida e traccia le caratteristiche organizzative, di sicurezza, di performance e scalabilità, interoperabilità, portabilità e conformità legislativa a cui dovranno uniformarsi i fornitori di servizi e di infrastrutture per le pubbliche amministrazioni.

¹⁶ Non è possibile in questa sede affrontare con competenza le questioni etiche e legali legate al possesso dei dati di terzi e degli usi legali, para-legali o illegali che le grandi società dell'informatica ne fanno. Il problema ha dimensioni ormai enormi e sempre più spesso viene a galla nella stampa, nelle aule dei tribunali e anche nei banchi parlamentari di molti paesi. Il caso attualissimo dello scandalo di Facebook e Cambridge Analytica è solo una spia di quello a cui può portare il ritardo culturale e legislativo sulle tecnologie di rete.

¹⁷ Andrea Carandini parlava più di dieci anni fa di «sistemi informativi archeologici» unitari, statali, regionali e universitari, capaci di condividere un minimo di procedure essenziali.», Andrea Carandini, *Archeologia classica: vedere il tempo antico con gli occhi del 2000* (Torino: Einaudi, 2008), 19.

ogni nodo della rete deve giocoforza implementare, è chiaro che le soluzioni “fai da te” non siano più percorribili, in quanto poco performanti, assolutamente poco sicure e difficilmente aderenti alla legislazione vigente. È quindi fondamentale, per l’avanzare delle ricerche sulle applicazioni digitali e di rete alla ricerca archeologica, che istituzioni pubbliche quali le università, i grandi enti di ricerca, o altre realtà preposte alla gestione dei Beni Culturali intervengano tempestivamente nel fornire servizi tecnologici di alta qualità, finalizzati alla raccolta, gestione e pubblicazione dei dati. Fino ad allora, ci si dovrà rivolgere ai grandi *provider*, con tutte le implicazioni morali, etiche e legali che ne conseguono.

NoSQL: relazionale ma non solo

Come si è già visto, il modello relazionale è senza dubbio il più utilizzato e diffuso, ma non è stato l’unico nella storia dei *database*. Non è qui il caso di aprire una parentesi esaustiva sugli altri modelli in uso in vari campi della ricerca, e spesso creati per venire incontro a esigenze molto specifiche. È utile, però, fare un breve accenno a quei modelli di organizzazione dei dati che vengono recentemente raggruppati sotto il termine generico di banche dati NoSQL. Si tratta di modelli che si distinguono e differenziano da quello relazionale o che talvolta lo estendono a tal misura da diventare qualcosa di completamente diverso¹⁸. Si è già fatto cenno ai cosiddetti *triplestore*, banche dati specializzate alla gestione delle triple RFD (ma non solo) e quindi specificatamente concepite per le esigenze del Web semantico. Un altro modello che ultimamente sta conoscendo (nuovamente) una grande diffusione è quello orientato ai documenti (*document oriented database*), che si focalizza sulla gestione di dati semi-strutturati presenti in contenitori genericamente chiamati documenti. Invece delle righe di un *database* relazionale, dunque, l’unità è il “documento”, che può essere strutturato in modo indipendente. Se gli elementi di un RDBMS (le righe di una tabella) presentano tutti per definizione la stessa struttura (gli stessi campi), i singoli documenti possono anche essere organizzati in maniera molto diversa l’uno dall’altro. Ne deriva una grandissima libertà descrittiva, ma anche una minore possibilità di controllo e validazione dei dati e una grandissima difficoltà al momento della ricerca e dell’analisi.

Questi modelli non sono nuovi come concezione, in quanto esistono da vari decenni, ma recentissimamente stanno conoscendo una rapida diffu-

¹⁸ L’etichetta “NoSQL” è nata proprio per marcare una visione diametralmente opposta ai *database* relazionali, e al loro linguaggio fondamentale, l’SQL; talvolta, però, il termine viene inteso come un meno avversativo Not Only (non solo) SQL.

sione proprio grazie allo sviluppo della Rete, sempre più alle prese con grandi quantità di dati complessi (*big data*) che difficilmente possono essere gestiti dalle piattaforme *hardware* tradizionali. La necessità crescente di scalabilità orizzontale¹⁹ ha messo in crisi i *database* relazionali che talvolta sono stati sostituiti da altri modelli, normalmente appositamente costruiti²⁰. Di solito solo le grandi aziende informatiche dispongono della reale capacità tecnica ed economica – e hanno sentito la necessità – di promuovere iniziative di questo tipo. Google, ad esempio, ha sviluppato per i propri prodotti Bigtable²¹, un sistema proprietario di gestione di dati usato per quasi tutti i suoi prodotti, GMail, YouTube, Google-Maps e GoogleEarth, Google Books, ecc. Anche Facebook ha sviluppato un *database* NoSQL, inizialmente per risolvere un problema di ricerca molto specifico della sua piattaforma *social*. Il progetto è stato adottato e utilizzato da moltissime altre aziende e progetti importanti, dopo essere stato aperto e diventato *open source* sotto il nome di Apache Cassandra²².

Logo di
Apache
Cassandra



XML, TEI, EpiDoc: banche dati document oriented per l'archeologia e settori affini

Nell'ambito dell'archeologia, le soluzioni *document oriented* non hanno avuto fino a oggi larga diffusione, ma in ambiti a essa molto vicini sono stati e sono tuttora lo strumento più idoneo e utilizzato²³. Ci si riferisce a tutte le discipline umanistiche – antichistiche o meno – il cui oggetto

¹⁹ Ovvero la possibilità di distribuire lo stesso *dataset* su più macchine diverse, sia perché una sola non è in grado di fisicamente contenerlo e gestirlo, sia per creare la ridondanza necessaria per una maggiore sicurezza e velocità di trasmissione.

²⁰ I RDBMS normalmente non presentano problemi di scalabilità verticale: con l'aumento degli elementi le loro prestazioni non ne dovrebbero risentire, anche se sono necessarie molte accortezze per gestire banche dati con centinaia di milioni o qualche miliardo di *record*. Si tratta comunque di cifre molto lontane da quelle che siamo abituati ad affrontare in campo archeologico.

²¹ Ora offerto anche servizio web: <https://cloud.google.com/bigtable/>.

²² <http://cassandra.apache.org/>.

²³ Per un'eccezione di veda Andrea D'Andrea, Achille Felicetti, Sorin Hermon, Franco Niccolucci, Tommaso Zoppi e Roberto Bagnara, «I linguaggi standard del W3C e gli strumenti Open Source per la gestione dei dati archeologici», in *Open Source, Free Software e Open Format nei Processi di Ricerca Archeologici: Atti del I Workshop (Grosseto, 8 maggio 2006)*, a c. di Roberto Bagnara e Giancarlo Macchi Janica (Firenze: Centro editoriale toscano, 2007), 51–66.

principale di studio è il testo: dall'epigrafia alla filologia, passando anche per gli ambiti maggiormente incentrati sulla materialità dei supporti per la scrittura, come la codicologia, la papirologia, l'epigrafia, ecc. La grandissima varietà che gli oggetti relativi a testi e i testi stessi presentano ha impedito l'utilizzo di banche dati fortemente strutturate come quelle relazionali a favore di banche dati di documenti, in particolar modo di documenti XML. L'Extensible Markup Language (XML) è una raccomandazione del World Wide Web Consortium (W3C), datata 1998²⁴, che fissa regole precise per la definizione dei linguaggi di marcatura; un metalinguaggio per normare l'annotazione (marcatura) di documenti di testo, nato per la specifica esigenza di pubblicare documenti elettronici in forma semi-strutturata sul Web. Una delle sue più importanti caratteristiche è quella di essere un documento di testo, quindi facile da scrivere, leggere e comprendere sia da un operatore umano che da una macchina, senza particolari necessità di trasformazione, decodifica o ulteriore lavorazione. L'XML permette in tal modo di arricchire un qualsiasi testo di molti e sovrapposti livelli informativi, significativi per un umano come per una macchina. Il testo annotato può essere quello di una rubrica telefonica:

```
<contatto>
  <nome>Mario</nome>
  <cognome>Rossi</cognome>
  <telefono>+39 123 456 789</telefono>
  <email>mario.rossi@example.com</email>
  <im>skype:mario.rossi</im>
  <im> whatsapp:+39123456789</im>
</contatto>
```

quello un'epigrafe:

```
<ab>
  <lb n="1" />
  <supplied reason="lost">Εὐμοριτο</supplied>
  Ἐλλάδιος ἔζησε ἔτι <num value="50" ></num>
  <lb n="2" />
  <supplied reason="lost" >πλέον
  <expan>
    <abbr>ἔλατ</abbr><ex>τ</ex><abbr>οῦ</abbr>
```

²⁴ <https://www.w3.org/XML/>.

```

    </expan>
  </supplied>
  <supplied reason="lost" >μῆ</supplied>v<supplied
reason="lost">εξ</supplied>
  <supplied reason="lost">
    <num value="3">γ</num>
  </supplied>
  ἡμέρας
</ab>

```

oppure il contenuto di una banca dati relazionale (nell'esempio seguente la tabella delle unità stratigrafiche, già vista negli esempi precedenti):

```

<elenco_us>
  <us>
    <nr>1</nr>
    <tipo>US</tipo>
    <colore>7.5BG 7/2</colore>
    <composizione>limoso</composizione>
    <consistenza>100</consistenza>
  </us>
  <us>
    <nr>2</nr>
    <tipo>US</tipo>
    <colore>8.5BG 7/2</colore>
    <composizione>argilloso</composizione>
    <consistenza>20</consistenza>
  </us>
  <us>
    <nr>3</nr>
    <tipo>USN</tipo>
  </us>
  <us>
    <nr>4</nr>
    <tipo>US</tipo>
    <colore>5BG 8/2</colore>
    <composizione>sabbioso</composizione>
    <consistenza>4</consistenza>
  </us>
</elenco_us>

```

L'XML permette di definire per ogni documento la possibile lista dei marcatori (chiamati comunemente "tag", le parti di testo racchiuse tra parentesi uncinata), dei loro attributi, della loro struttura (quale tag può contenere altri tag, ecc.), e così via fino a ottenere un modello descrittivo molto preciso²⁵. In ambito umanistico esiste un'iniziativa importante, la Text Encoding Initiative (TEI)²⁶, un consorzio che sviluppa e mantiene in maniera collettiva uno standard per la rappresentazione dei testi in forma digitale. Il consorzio pubblica una serie di linee guida che disciplinano i metodi di codifica dei testi (perché siano leggibili dalle macchine) principalmente nel campo degli studi umanistici, delle scienze sociali e della linguistica. Il TEI ha sviluppato delle specifiche molto articolate e complesse, che coprono quasi tutti gli ambiti di applicazione, ma che possono disorientare un neofita. Per ovviare a questo problema è stata elaborata e pubblicata una versione semplificata, definita TEI Lite, progettata per soddisfare il "90% dei bisogni del 90% della comunità di utilizzatori del TEI"²⁷. Grazie alla semplicità di utilizzo e di apprendimento, è stato adottato su amplissima scala²⁸.

TEI Lite è comunque un'iniziativa rivolta a un pubblico generico; dal momento però che gli ambiti applicativi del TEI sono così vasti, sono nati diversi sottoprogetti per lo sviluppo di standard da applicare a settori più ristretti. Tra questi, di particolare importanza è l'EpiDoc²⁹, un'iniziativa che promuove e coordina lo sviluppo di linee guida per la codifica in TEI XML di testi relativi all'ambito dell'epigrafia e della papirologia³⁰. Grazie a questi importanti sforzi di coordinamento è stato possibile codificare in formato elettronico e pubblicare *online* importanti *corpora* di testi epigrafici e papirologici relativi al mondo antico.

Dal punto di vista pratico, è possibile creare e modificare un documento XML facendo uso di un semplice *editor* di testi, considerata la sua sintassi

²⁵ Queste "istruzioni" sono contenute in documenti separati usati per definire lo schema generale e la validazione dei dati. Nel tempo si sono succeduti vari standard per la loro redazione, tra i quali si possono menzionare Document type definition (DDT), XML schema (XSD), RELAX NG, ecc.

²⁶ <https://tei-c.org/>.

²⁷ <https://tei-c.org/guidelines/customization/lite/>.

²⁸ Per un manuale introduttivo in italiano su TEI Lite si veda Lou Burnard, C. M. Sperberg-McQueen, e Fabio Ciotti, *Il manuale TEI Lite: introduzione alla codifica elettronica dei testi letterari* (Milano: Silvestre Bonnard, 2005).

²⁹ <https://sourceforge.net/p/epidoc/wiki/Home/>.

³⁰ La versione più recente delle Guidelines è la 9, disponibile all'indirizzo: <http://www.stoa.org/epidoc/gl/latest/>.

semplice. Esistono comunque vari strumenti, gratuiti e a pagamento, che offrono un validissimo aiuto alla scrittura in formato XML, suggerendo i marcatori (*tag*) da usare e soprattutto fornendo, in fase di modifica, una validazione in tempo reale del documento, sia dal punto di vista sintattico³¹ che dal punto di vista semantico (una volta fornito un documento che definisce lo schema da usare)³².

Esistono inoltre strumenti *software* specifici per l'organizzazione e la gestione dei documenti in formato XML, i *database XML*, tra cui il diffusissimo programma *open source* eXist-db³³. Queste banche dati integrano strumenti molto potenti di ricerca che utilizzano un linguaggio di interrogazione standard chiamato XQuery³⁴, che si potrebbe descrivere come l'equivalente per l'XML dell'SQL per i *database* relazionali.

Alcuni produttori di RDBMS hanno inoltre aggiunto ai loro programmi il supporto per la gestione (archiviazione e interrogazione) dell'XML. Supportano il formato XML, ad esempio, IBM DB2, Microsoft SQL Server, Oracle Database e PostgreSQL. È quindi possibile, con questi *software*, combinare più linguaggi di interrogazione (*query*) per ottenere ricerche estremamente mirate. Queste soluzioni ibride sono però ad oggi poco o per nulla usate in campo umanistico.

L'adozione dei *database XML* corrisponde dunque a una necessità molto specifica, ovvero quello di descrivere in maniera particolareggiata un testo sotto vari punti di vista³⁵, cosa che risulta difficilmente realizzabile attra-

³¹ Alcuni esempi di violazione della sintassi sono l'omissione di apici per gli attributi, l'omissione del *tag* di chiusura, la sovrapposizione non corretta dei *tag*, ecc.

³² Aspetto importantissimo per l'interoperabilità, come si è già visto. La validazione semantica controlla che non si faccia utilizzo di *tag* non previsti dallo schema, che la struttura dei *tag* non contrasti con specifiche definite nello schema, che non si utilizzino valori non permessi per gli attributi, ecc.

³³ <http://exist-db.org/>.

³⁴ Anche XQuery è uno standard W3C, <https://www.w3.org/TR/xquery-31/>.

³⁵ Un testo epigrafico può quindi essere descritto in TEI XML sotto l'aspetto fisico della forma, misura e stato di conservazione del materiale, dell'aspetto altrettanto fisico dell'impaginazione del testo e di vari espedienti grafici che informano sul lavoro del lapicida. Questo tipo di marcatura permette, ad esempio, la produzione automatica di edizioni di varia tipologia. Ma il documento può essere descritto anche dal punto di vista dei contenuti del testo, dello scioglimento delle sigle e delle abbreviazioni, degli errori del copista, della trasposizione delle date e della loro versione in altri formati. Dell'onomastica, delle relazioni tra persone, della toponomastica, di mestieri, incarichi e carriere, ecc. È inoltre possibile marcare il testo dal punto di vista linguistico, individuando le varie parti della frase (analisi PoS, ovvero *Parto of Speech*), documentando costrutti particolari o forme linguistiche dialettali

verso un modello relazionale.

Questi due paradigmi – relazionale e su base XML – continueranno verosimilmente a coesistere a lungo, in quanto entrambi soddisfano pienamente le esigenze particolari dei loro utilizzatori; a questi potranno auspicabilmente affiancarsi nel prossimo futuro anche dati *triplestore* per la pubblicazione dei dati semantici. Attualmente la tendenza generale è quella di creare nuovi strumenti *software* specifici, i quali a partire da *dataset* relazionali o orientati ai documenti, producano triple RDF corrispondenti a uno schema di corrispondenze attentamente studiato.

Con questi tre modelli, lo spettro delle necessità per un archeologo o in generale umanista è più che soddisfatto.

Eventuali innovazioni e sperimentazioni potrebbero essere innescate dalla diffusione dei cosiddetti *database* in tempo reale (*real-time database*). È improbabile che siano le esigenze di ricerca a spingere in questa direzione, quanto piuttosto la curiosità per la sperimentazione e l'influenza della più generale evoluzione del Web. Si tratta di modelli di dati orientati ai documenti, di semplicissimo utilizzo e dipendenti, ancora una volta, da implementazioni *SaaS* (*software* come servizio). Fattori legati alla tendenza del momento, alla semplicità di sviluppo e alla necessità di allargare sempre di più la platea dei fruitori (anche attivi) della ricerca archeologica – obiettivo importante dell'archeologia pubblica³⁶ – potranno spingere nel prossimo futuro a fare sperimentazione su una tecnologia che sembra principalmente mirata allo sviluppo per dispositivi mobili, anche per via della progressiva diffusione dell'uso proprio di quest'ultimi soprattutto nei processi di raccolta dati.

Un panorama futuribile si prospetta grazie a un progetto fortemente orientato allo sviluppo di applicazioni *mobile* e all'utilizzo di tecnologie di rete come ArchAIDE, dell'Università di Pisa³⁷. Il suo obiettivo è quello di creare un'applicazione per dispositivi mobili (app) che, sfruttando la tecno-

e altre particolarità. Di norma, questi strati informativi non vengono usati contemporaneamente sullo stesso documento; un testo può ricevere vari tipi di marcature diverse, corrispondenti a una specifica esigenza di ricerca.

³⁶ Sui primi passi dell'archeologia pubblica in Italia si veda Chiara Bonacchi, «The development of Public Archaeology in Italy: a review of recent efforts», *Public Archaeology* 12, n. 3 (2013): 211–16; Cinzia Dal Maso e Francesco Ripanti, *Archeostorie: manuale non convenzionale di archeologia vissuta* (Milano: Cisalpino, 2015). Oggi in Italia è attiva anche una rivista scientifica ad accesso libero, in inglese, dedicata all'argomento "Archeostorie, Journal of Public Archaeology", <https://ar-cheostoriejpa.eu/>.

³⁷ <http://www.archaide.eu/>.

logia del riconoscimento automatico delle immagini (*automatic image recognition*), sia in grado di riconoscere le tipologie delle forme ceramiche a partire da foto o disegni effettuati sul campo. In casi come questo, quando la platea degli utenti attivi si allarga enormemente, in cui l'aggiornamento in tempo reale della banca dati di riferimento è un requisito indispensabile e dove un sistema di riscontro (*feedback*) sulla qualità della risposta automatica è uno strumento fondamentale, è possibile che nuovi modelli di strutturazione dei dati possano dare un contributo importante.

Capitolo 4

GIS e webGIS tra archiviazione, analisi e pubblicazione dei dati geografici

I sistemi informativi geografici (ingl. *Geographical Information System*, GIS) possono essere definiti come banche dati altamente specializzate nell'archiviazione, interrogazione, analisi e rappresentazione dei dati spaziali e geografici. Benché la rappresentazione grafica rivesta un ruolo importante all'interno di questi strumenti, sono le loro capacità di codifica rigorosa e analisi approfondita dei dati spaziali a differenziarli in maniera netta dagli altri sistemi di archiviazione e gestione fin qui considerati.

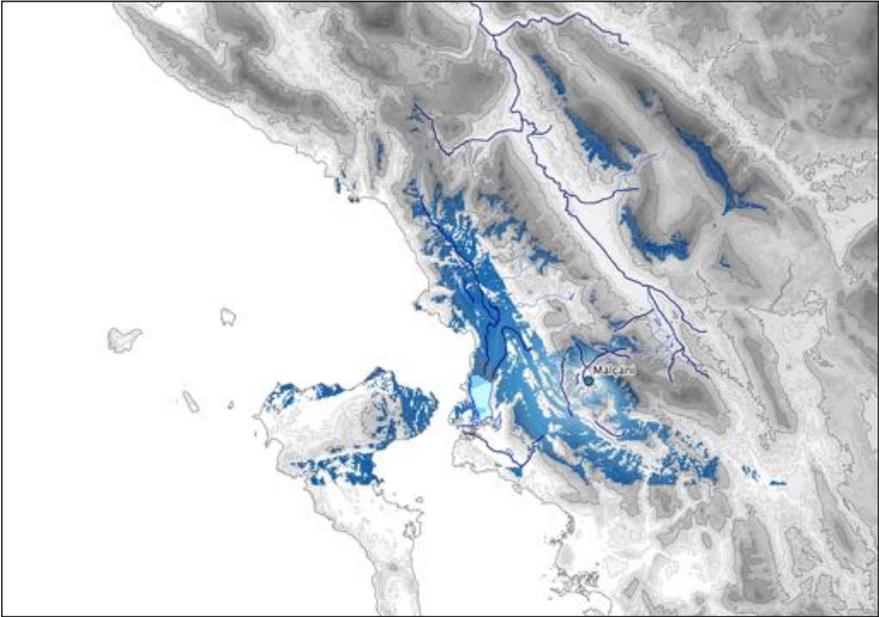
Se una banca dati testuale (relazionale o *document oriented*) permette di trovare tutte le occorrenze e fare confronti rispetto a una stringa o una sua parte, un numero, una data o un modello, un Sistema Informativo Geografico (GIS) permette di ricercare elementi che sono in un qualche rapporto spaziale con un termine di paragone. Un elemento dotato di una dimensione spaziale può quindi intersecare, contenere (o essere contenuto), toccare, sovrapporsi, incrociare, trovarsi a una certa distanza, ecc. da un altro elemento o da una serie di questi. Attraverso una *query* spaziale è possibile ad esempio evidenziare porzioni di territorio che rispondono ad alcune caratteristiche definite, come un certo grado di inclinazione e/o pendenza, una determinata esposizione al sole o alle correnti ventose, una particolare copertura di vegetazione o uno specifico uso del suolo.

Le piattaforme GIS permettono di gestire diverse fonti di dati spaziali, codificati secondo vari modelli e formati di *file*. I formati dei dati geograficamente rappresentabili vengono generalmente raggruppati in due grandi famiglie, i formati *raster* e i formati vettoriali. Un *file* in formato *raster* è determinato da una griglia di piccoli elementi quadrati chiamati *pixel* (sinonimi dei due termini inglesi *picture element*), che compongono l'immagine digitale disponendosi in righe e colonne (griglia, ingl. *raster*).

Ciascun *pixel* è associato a uno o più valori numerici che possono rappre-

Esempio di analisi di visibilità. Data una rappresentazione tridimensionale del terreno, è possibile evidenziare (in toni di blu) le aree visibili da un determinato punto di osservazione (il sito di Malçani).

Fonte:
Sistema Informativo Archeologica della Regione Caona (SI-TARC).



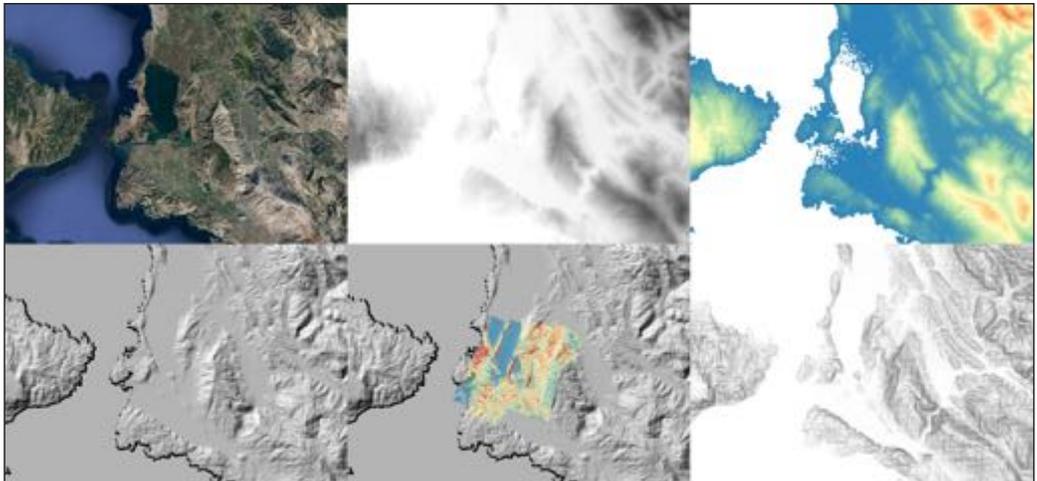
sentare fenomeni diversi: nel caso del formato *raster* più comune, la fotografia, il *pixel* esprime il colore (anche fuori dall'ambito GIS), ma in altre tipologie di immagini *raster* può indicare l'altimetria, ad esempio, espressa di norma in metri sopra il livello medio del mare. Altre informazioni semplici che possono essere codificate nei *pixel* sono la pendenza (indicata in gradi), l'esposizione al sole (a partire da una posizione stabilita del sole), la distanza oppure la visibilità da un dato punto di riferimento. Le immagini *raster* presentano modelli di codifica più complessi, poiché in esse ogni *pixel* contiene tre, quattro o più valori numerici (detti "bande") in combinazione tra loro a seconda del profilo colore adottato. Ci sono poi *raster* ancora più densi di informazioni, che rappresentano realtà articolate, come le immagini multispettrali, che in un unico *file* codificano – ampliando la serie numerica associata a ogni *pixel* – informazioni relative allo spettro del visibile o a frequenze non percepibili dall'occhio umano (invisibile), come l'infrarosso, l'ultravioletto, ecc.

Le informazioni numeriche contenute nei *pixel* vengono usate nelle piattaforme GIS per l'analisi e l'interrogazione dei dati *raster*, ed è quindi fondamentale, per poter eseguire una qualsiasi interrogazione, la conoscenza dei sistemi di codifica di ogni dato rappresentato nell'immagine.

Si presenta di norma in formato *raster* tutta la documentazione fotografica che si produce in un progetto archeologico, le acquisizioni digitali di dati cartacei attraverso scanner, macchine fotografiche o dorsi fotografici, i risultati dei procedimenti di fotopianamento e le ortofoto prodotte in vario



Immagine satellitare (Google) a sinistra e ingrandimento della parte evidenziata a destra, fino alla visualizzazione dei singoli pixel. Fonte: Sistema Informativo Archeologica della Regione Caona (SITARC), a cura dell'Autore

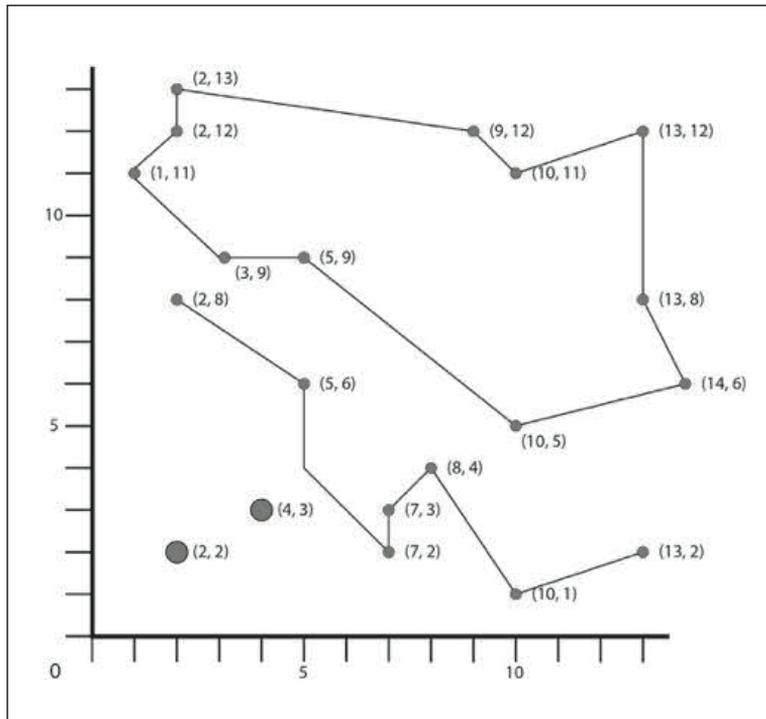


Diverse tipologie di immagini raster, riferite allo stesso territorio, area di Butrinto, Albania. In alto a sinistra, una fotografia satellitare a colori naturali (true color). A ogni pixel corrisponde una tripletta di numeri che codificano la rappresentazione di una sfumatura di colore nel formato RGB. In alto al centro, una rappresentazione dell'elevazione del terreno, in scala di grigi. A ogni pixel corrisponde il valore numerico della quota, espressa in metri sul livello medio del mare. In alto a destra, la stessa rappresentazione presente al centro, ma visualizzata in pseudocolore, con gradienti cromatici personalizzati e arbitrari. In basso a sinistra, una rappresentazione dell'esposizione al sole, espressa in scala di grigi. A ogni pixel corrisponde il valore dell'angolo del sole ("collocato" in un determinato azimut e altezza). Al centro, sull'immagine precedente sono sovrapposte avio-riprese multispettrali, rappresentate in falso colore (false color). Alle lunghezze d'onda invisibili all'occhio umano (ad esempio, lo spettro dell'infrarosso) sono assegnati dei colori arbitrari nel formato RGB. L'ultima immagine, in basso a destra, rappresenta le pendenze del terreno. Ogni pixel riporta il valore del coefficiente di massima differenza di elevazione rispetto ai pixel confinanti. Più questa differenza è grande, più il terreno è acclive. La pendenza è rappresentata in scala di grigi. Fonte: Sistema Informativo Archeologica della Regione Caona (SITARC), a cura dell'Autore.

modo, ossia mediante fotogrammetria (terrestre, aerea o UAV) o rilievi laser (scanner terrestri, aviotrasportati, ecc.).

Gli elementi costitutivi della grafica vettoriale sono, invece, primitive geometriche semplici (punti, linee o poligoni) descritte matematicamente tramite coppie di coordinate (o terne nel caso di immagini a tre dimensioni). Ciascuno di questi elementi costitutivi può essere associato a una serie di valori testuali, chiamati “attributi”, contenenti informazioni sulla visualizzazione (spessore del tratto, colore, trasparenza, riempimento, gradienti di colore, ecc.), oppure – come solitamente avviene nel caso dei vettori GIS – informazioni utili alla descrizione strutturata degli elementi. La progettazione e la corretta gestione degli attributi da associare a ogni elemento vettoriale è un passaggio importantissimo, poiché dalla struttura di questi dipendono tutte le possibilità di interazione (interrogazione/*query*) con i vettori. Quanto detto in precedenza a proposito della progettazione dei *database* relazionali vale infatti anche per i dati spaziali, soprattutto se consideriamo i dati vettoriali come dati tabellari organizzati in *database* relazionali dotati di colonne speciali per la memorizzazione delle geometrie. Ed effettivamente, pur esistendo moltissimi formati *file* specifici per la memorizzazione dei dati vettoriali, sono sempre più utilizzati per la loro archiviazione i RDBMS con estensioni e funzioni spaziali, i cosiddetti “ge-

Esemplificazione della grafica vettoriale, come uno spazio cartesiano all'interno del quale gli elementi sono definiti da coppie di coordinate (punti nello spazio bidimensionale) o da serie di coppie di coordinate (linee e poligoni nello spazio bidimensionale).



odatabase”, o *database* spaziali (*spatial database*), un termine generico che designa varie possibilità di codifica. Alcune implementazioni di *geodatabase* permettono anche l’archiviazione dei dati *raster*, diventando in tal modo punti di riferimento unici per la gestione strutturata di gran parte delle tipologie di dati prodotti dalla ricerca archeologica¹.

Esiste una vasta bibliografia sull’utilizzo del GIS in archeologia, alla quale si rimanda per le questioni più puntuali², mentre per gli aspetti operativi la via migliore di approfondimento è costituita dai manuali d’utente dei vari *software*. Nei paragrafi seguenti ci si concentrerà sugli aspetti della tecnologia GIS in rapporto a Internet e al Web, in linea con le finalità di questo volume.

L’avvento del World Wide Web ha infatti introdotto importanti novità nella creazione, gestione, fruizione e persino pubblicazione dei dati geografici e delle piattaforme GIS, creati e implementati tradizionalmente in locale. Come vedremo, l’approdo in rete delle piattaforme GIS è stato ritardato da due ostacoli principali: le grandi dimensioni che di solito hanno i *file* che le costituiscono (soprattutto i *raster*) e il monopolio della già menzionata, ESRI³, che per più di vent’anni ha dominato il mercato. Dal 1991 infatti, anno del rilascio del *software* ArcView (1991), piattaforma GIS concepita per il largo utilizzo, i prodotti commerciali ESRI non hanno avuto rivali e sono stati anche adottati dagli uffici tecnici degli enti pubblici, fattore che, se da una parte ha avuto il merito di imporre standard tuttora di largo utilizzo (lo *shapefile*, ad esempio), dall’altro ha impedito la libera sperimenta-

¹ È possibile archiviare dati *raster* nel formato *geodatabase* di ESRI (<http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/creating-raster-datasets-in-a-geodatabase.htm>), in PostGIS (https://postgis.net/docs/RT_reference.html) e anche in Spatialite (https://www.gaia-gis.it/fossil/spatialite_gui/wiki?name=world-rasters). È però sconsigliato l’uso di RDBMS per la gestione di dati *raster*, per le grandi e grandissime dimensioni che questi tipi di *file* possono raggiungere.

² Si veda in italiano Maurizio Forte, *I Sistemi Informativi Geografici in archeologia* (Roma: MondoGis, 2002); Julian Bogdani, «GIS per l’Archeologia», in *Groma 2. In profondità senza scavare*, a c. di Enrico Giorgi (Bologna: BraDypUS, 2009), 421–38. e in inglese James Conolly e Mark W Lake, *Geographical Information Systems in Archaeology* (Cambridge: Cambridge University Press, 2006); Andrea Scianna e Benedetto Villa, «GIS Applications in Archaeology», *Archeologia e Calcolatori* 22 (2011): 337–63. Per la bibliografia più recente si veda Julian Bogdani e Eugenio De Mitri, «A Bibliography on the Application of GIS in Archaeology and Cultural Heritage», *Groma. Documenting Archaeology*, 21 dicembre 2017, <https://doi.org/10.12977/groma12>.

³ Environmental Systems Research Institute (<https://www.esri.com/>), azienda leader nell’industria *software* dei GIS.

zione da parte degli utenti, legati alle scelte commerciali della casa produttrice. Le alternative aperte a ESRI non sono mai mancate: GRASS GIS⁴, per esempio, sviluppato dall'esercito americano e rilasciato per la prima volta nel 1984, è uno strumento estremamente versatile e potente, ma che è stato per lungo tempo troppo difficile da usare per un utente senza buone conoscenze informatiche. Una svolta importante è avvenuta nel 2000, con lo sviluppo e il rilascio, da parte di Frank Warmerdam, di GDAL⁵, una libreria⁶ che permetta la facile lettura e scrittura di una numerosa serie di formati *file*, sia *raster* che vettoriali (attraverso OGR Simple Features Library). Rilasciata con licenza *open source* è oggi alla base della maggior parte dei *software* GIS, anche quelli dei grandi produttori commerciali come ESRI. La disponibilità di uno strumento del genere, non soggetto a restrizioni commerciali, ha aperto la strada a un più ampio utilizzo dei GIS anche fuori dai tradizionali programmi *desktop*.

Oggi, per quanto riguarda l'ambiente *desktop*, ESRI ArcMap è ancora uno dei programmi più utilizzati, ma anche uno dei più "pesanti" e sicuramente il più costoso. Le alternative aperte, *in primis* lo stesso GRASS sopra menzionato (ormai dotato di interfacce grafiche che ne facilitano l'utilizzo), sono diverse: tra le più diffuse ricordiamo QGIS⁷, gvSIG⁸, SagaGIS⁹, seguite da altre più specifiche per utilizzi di settore. L'immediatezza d'uso, l'accesso facilitato a gran parte dei formati, la possibilità nativa di interfacciarsi con i principali RDBMS ha fatto dei GIS un ausilio fondamentale al lavoro dell'archeologo. Da strumento avanzato per le analisi spaziali, tanto apprezzate dall'archeologia quantitativa, è oggi comunemente utilizzato anche come strumento di interfaccia unico per piattaforme informatiche piuttosto complesse.

La presenza sempre più pervasiva della Rete ha d'altro canto innescato una dinamica di lenta sperimentazione sulla pubblicazione *online* dei dati

⁴ Geographic Resources Analysis Support System, <https://grass.osgeo.org/>.

⁵ Geospatial Data Abstraction Library, <https://www.gdal.org/>.

⁶ In informatica, il termine italiano "libreria" è un'errata traduzione dell'inglese *library* (ovvero biblioteca), ormai largamente diffuso, ed indica un insieme di funzioni specifiche che vengono sviluppate e distribuite in maniera indipendente e vengono successivamente collegate ai *software*. Permettono quindi di velocizzare lo sviluppo dei programmi, facendo ricorso a funzionalità già disponibili e pronte all'uso.

⁷ <http://qgis.org>.

⁸ <http://www.gvsig.com/>.

⁹ Particolarmente orientato alle analisi spaziali, <http://www.saga-gis.org/>.

geografici da parte di ricercatori e progetti di ricerca, per scopi di condivisione interna e con accesso limitato dall'esterno. Più importante è stata la pubblicazione *online* da parte degli enti pubblici di *dataset* geografici di grande consistenza e ricchezza di contenuti: il Geoportale Nazionale a cura del Ministero dell'Ambiente¹⁰, i portali cartografici delle singole Regioni e il sito dell'Istituto Geografico Militare¹¹ sono strumenti imprescindibili per chi opera sul suolo italiano. Anche se prodotti con scopi diversi da quelli della ricerca archeologica, la facile reperibilità di dati cartografici di alta qualità – foto aeree di annate diverse e in alta risoluzione, coperture geologiche e geomorfologiche, cartografia storica, dati LiDAR e *database* toponomastici – ha segnato un avanzamento importante in questo campo e ha diffuso e allargato la base di utilizzo sia dei prodotti digitali disponibili telematicamente sia dei *software* GIS, incentivando la produzione di nuovi tematismi.

Si deve però alle grandi aziende dell'informatica il definitivo ingresso del Web 2.0 nell'ambito della rappresentazione geografica. Nel giugno del 2001 Google lanciava il suo progetto GoogleEarth¹², mettendo profondamente in crisi l'idea del *desktop* GIS: con GoogleEarth gli utenti si sono trovati a disposizione uno strumento semplicissimo, leggero e veloce che, cosa più importante, fornisce gratuitamente una copertura di foto satellitari ad altissima risoluzione per l'intera superficie del globo. Le foto satellitari, scattate da Landsat 7 e, quelle più recenti, da Landsat 8¹³, vengono memorizzate nei *server* dell'azienda, che le fornisce come servizio ai propri utilizzatori attraverso il Web. Queste immagini sono state una fonte inestimabile di informazioni per una serie di studi sui paesaggi archeologici, sia come fonte "tradizionale" per la ricognizione e l'individuazione di tracce, sia come base di algoritmi e *software* finalizzati al riconoscimento automatico delle anomalie¹⁴. Le immagini satellitari sono accompagnate da un

¹⁰ <http://www.pcn.minambiente.it/>.

¹¹ <https://www.igmi.org/>.

¹² <https://www.google.it/earth/>.

¹³ I programmi Landsat sono stati promossi dalla NASA (National Aeronautics and Space Administration, <https://www.nasa.gov/>) e i dati vengono distribuiti dalla United States Geological Survey (USGS).

¹⁴ Le applicazioni di Remote Sensing nella ricerca archeologica sono un vasto campo di studi, particolarmente vivo in Italia, Stefano Campana e Maurizio Forte, *Remote sensing in archaeology: XI Ciclo di lezioni sulla ricerca applicata in archeologia, Certosa di Pontignano (Siena), 6-11 dicembre 1999*. All'Insegna del Giglio: Firenze, 2001; Stefano Campana, e Maurizio Forte, a c. di, *From Space to Place:*

modello tridimensionale del terreno, frutto delle misurazioni del progetto SRTM¹⁵ della NASA. Dal 2009, inoltre, il programma si è arricchito della quarta dimensione, grazie all'introduzione di Historical Imagery, una funzione che permette di vedere coperture satellitari dello stesso luogo compiute in periodi diversi. L'utente può facilmente creare elementi vettoriali utilizzando un formato inventato *ad hoc*, il KML (Keyhole Markup Language)¹⁶, pensato espressamente per la diffusione di dati geografici attraverso la rete Internet. Google Earth non fornisce strumenti di analisi spaziale come i tradizionali *software desktop* (se si escludono le funzioni elementari di misurazione lineare e areale), ma permette la ricerca per nome di località (*geocoding*) e l'estrazione di coordinate a video. Google Earth, in breve, anticipa e riassume già nel 2001 quello che sarebbe stato lo sviluppo dei webGIS nei vent'anni successivi.

Nel 2005 è stato inoltre lanciato un altro servizio, GoogleMaps, che alle foto satellitari ad alta risoluzione accompagna una copertura cartografica globale, alla compilazione e correzione della quale sono invitati a partecipare tutti gli utenti. L'anno precedente era stato lanciato un altro progetto simile, OpenStreetMaps (OSM)¹⁷, che, rifacendosi al concetto di implementazione collaborativa di Wikipedia, si proponeva di creare una copertura cartografica su base volontaria. Inizialmente il progetto doveva interessare solo l'area della Gran Bretagna, ma di seguito si è diffuso a livello mondiale e oggi vi si trovano integrati non solo i dati creati dagli utenti, ma anche dati geografici aperti e non coperti da *copyright* o licenze restrittive, pubblicati da enti pubblici di vari livelli¹⁸. La copertura car-

2nd International Conference on Remote Sensing in Archaeology: Proceedings of the 2nd International Workshop, CNR, Rome, Italy, December 4-7, 2006 (Oxford: Archeopress, 2006); James Wiseman e Farouk El-Baz, *Remote Sensing in Archaeology* (New York: Springer, 2007); Sarah H Parcak, *Satellite remote sensing for archaeology* (London; New York: Routledge, 2009); Maurizio Forte, Stefano Campana, e Claudia Liuzza, a c. di, *Space, Time, Place Third International Conference on Remote Sensing in Archaeology: 17th-21st August 2009, Tiruchirappalli, Tamil Nadu, India* (Oxford: Archeopress, 2010); Maurizio Forte e Stefano Campana, *Digital Methods and Remote Sensing in Archaeology Archaeology in the Age of Sensing* (Cham: Springer International Publishing, 2016).

¹⁵ Lo Shuttle Radar Topography Mission, <https://www2.jpl.nasa.gov/srtm/>.

¹⁶ Il formato era stato sviluppato dalla Keyhole Inc., poi acquisita da Google. Dal 2008 è diventato uno standard aperto della Open Geospatial Consortium (OGC), <http://www.opengeospatial.org/standards/kml>.

¹⁷ <https://www.openstreetmap.org/>.

¹⁸ Una lista di enti terzi che hanno contribuito con *open data* all'implementazione di OpenStreetMap è pubblicata e tenuta aggiornata all'indirizzo: <https://wiki>.

tografica di OSM rimane ad oggi disomogenea, nonostante molti sforzi siano stati profusi recentemente per equilibrare le disparità più importanti, dovute principalmente al divario tecnologico tra le diverse parti del mondo e alle diverse licenze d'utilizzo con cui vengono distribuiti i dati geografici ufficiali nei vari paesi. Si pensi per esempio che negli Stati Uniti qualsiasi dato prodotto dal governo federale esula per definizione dalla legge del *copyright* ed è di pubblico dominio e si può integrare senza restrizioni in questo tipo di piattaforme. In tempi recenti anche in Italia è in atto un grande sforzo pubblico verso la diffusione dei dati aperti, ma la strada per l'effettivo libero accesso è ancora lunga e piena di ostacoli.

La disponibilità di banche dati geografiche sempre più ricche (come OpenStreetMaps) e il paradigma della cartografia come servizio Web, introdotto da GoogleMaps, hanno incentivato lo sviluppo di nuovi protocolli e *software* per la facile e immediata messa *online* dei dati geospaziali.

Nei prossimi paragrafi verranno presentati brevemente i principali standard e strumenti oggi a disposizione per la condivisione e pubblicazione *online* di dati geospaziali, *raster* e vettoriali.

Nell'appendice di questo volume, infine, si trova una rassegna esemplificativa di alcuni progetti di grande respiro, relativi al mondo antico, che costituiscono modelli validissimi da utilizzare ed eventualmente arricchire.

Standard per la pubblicazione dei dati geografici strutturati sul Web

Quando si parla di codifica e rappresentazione di dati geospaziali non si può non menzionare l'OGC¹⁹, un'organizzazione internazionale che, dal 1994, coordina gli sforzi di definizione e diffusione di standard aperti per dati e servizi geospaziali. Dal 2006, inoltre, è attiva una seconda organizzazione, OSGeo²⁰, maggiormente orientata alla promozione e allo sviluppo di strumenti e formati geospaziali *open source*. Entrambe le associazioni concordano sull'apertura degli standard, un nodo importante sia per favorire e incoraggiare il facile interscambio dei dati (interoperabilità), sia per mettere in atto strategie mirate ad allungare la vita delle informazioni digitalmente codificate. Dalla fine degli anni Novanta del secolo scorso gli sforzi dell'OGC si sono concentrati nel tentativo di portare sul Web i

openstreetmap.org/wiki/Contributors. Vi si trovano anche alcune Regioni italiane (Emilia-Romagna, Piemonte, Toscana, Puglia), qualche Provincia e vari Comuni.

¹⁹ Open Geospatial Consortium, <http://www.opengeospatial.org/>.

²⁰ Open Source Geospatial Foundation, <https://www.osgeo.org/>.

dati GIS, compito arduo a causa della loro grande complessità e peso (in termini di memoria occupata), limiti questi assai rilevanti se messi in relazione a quella che era l'infrastruttura di rete più di vent'anni fa. I problemi sono stati però affrontati e risolti, con una serie di specifiche diverse, da parte delle due organizzazioni sopra menzionate, mediante la creazione di alcuni standard che vengono continuamente aggiornati e sono ancora oggi utilizzati da ormai tutte le applicazioni webGIS. Si tratta di standard di trasmissione e pubblicazione secondo il modello *client-server*: a una richiesta specifica da parte del *client* (un'applicazione di visualizzazione geografica, oppure un *software GIS desktop*), il *server* cartografico risponde restituendo una porzione di mappa, in vari formati a seconda della specifica.

La capillare diffusione di questi protocolli ne ha reso necessaria la conoscenza, anche solo sommaria, da parte degli archeologi, anche qualora non fossero intenzionati a pubblicare i propri dati *online*, ma solo, ad esempio, integrare nei propri progetti *dataset* resi disponibili da altri. Il loro utilizzo nei progetti GIS, di qualsiasi scala è, infatti, ormai generalizzato e molto comune.

WMS

Il Web Map Service è uno standard OGC, rilasciato nel 2000²¹, per la trasmissione attraverso il protocollo Web HTTP di metadati e immagini in formato di restituzione *raster*. L'immagine restituita corrisponde alla porzione della mappa richiesta e può essere ampiamente personalizzata: è possibile impostare *layer* specifici, stili, trasparenze, filtri, ecc. Si possono inoltre ottenere informazioni dettagliate (metadati) sul *dataset* geografico in uso²², sui *layer*, sulla legenda e anche eseguire *query* spaziali in base a una posizione geografica fornita (definita di norma da una coppia di coordinate), esattamente come in un *software GIS desktop*. Trattandosi di un servizio Web, la risorsa WMS è identificata da un URL, un indirizzo web che punta al *server* geografico. Ad esempio, il Portale Cartografico Nazionale del Ministero dell'Ambiente pubblica la copertura IGM in scala 1:100.000 relativa a tutto il territorio italiano all'indirizzo: http://wms.pcn.minambiente.it/ogc?map=/ms_ogc/WMS_v1.3/raster/IGM_100000.map. Questo URL può essere usato in un tradizionale *software GIS* (ArcMap o QGIS, ad esempio) come base per un progetto più ampio, oppure può

²¹ <http://www.opengeospatial.org/standards/wms>.

²² Questo tipo di informazioni si chiamano *capabilities* e vengono fornite in formato XML.

essere integrato in piattaforme web dinamiche. Si ha comunque bisogno di un *software* compatibile con il protocollo WMS per visualizzare i dati geografici, perché se si tenta di accedere all'URL con un *browser* web si riceveranno una serie di errori dovuti ai parametri mancanti. Esistono comunque *online* molti servizi generici, che permettono di visualizzare dati geografici distribuiti attraverso il protocollo WMS a partire da un URL.

Il WMS è tuttora il servizio più diffuso e utilizzato per la condivisione e pubblicazione di dati *raster*, come le basi cartografiche di varia tipologia e scala, foto satellitari e aeree, modelli digitali del terreno, dati LiDAR, ecc.

WFS

Il Web Feature Service è anch'esso uno standard dell'OGC²³ rivolto esclusivamente alla gestione dei dati vettoriali, che sono chiamati nel linguaggio GIS con il termine inglese *features*. Anche in questo caso ogni servizio è identificato da un URL, il quale richiama uno o più temi (*layer*) vettoriali. A differenza del WMS, il *server* non risponde con porzioni di mappa, ma di norma restituisce l'intero tema vettoriale, codificato per il trasposto in formato XML (o più specificatamente in GML²⁴).

Rispetto al WMS, il WFS permette, una maggiore interazione da parte dell'utente, in quanto non vi è nessuna differenza tra un *layer* vettoriale disponibile localmente da un *layer* WFS (fatte salve limitazioni introdotte dall'editore): un utente può analizzare, applicare filtri, simbologia e fare analisi sugli attributi o sulle proprietà spaziali. Le specifiche WFS permettono anche la possibilità di interagire attivamente con i dati salvati nel *server* geografico, consentendo la modifica da remoto (creazione di nuovi vettori, modifica di geometrie e attributi, eliminazione di vettori esistenti), anche se questo servizio è solitamente disabilitato negli archivi di dati geografici finalizzati alla sola pubblicazione e diffusione. Per fare un esempio, all'indirizzo http://wms.pcn.minambiente.it/ogc?map=/ms_ogc/wfs/Linea_costa_2009.map il Portale Cartografico Nazionale pubblica in modalità di sola lettura (ma con possibilità di scaricamento e di modifica della copia locale) la linea di costa della Repubblica Italiana, come è stata rilevata nel 2009. Come per il WMS, anche per il WFS sono necessari visualizzatori specifici per rappresentare il dato, che traducano il formato

²³ <http://www.opengeospatial.org/standards/wfs>.

²⁴ Il Geography Markup Language (GML) è un formato di rappresentazione di dati geografici basato sul metalinguaggio XML. Anche il GML è uno standard OGC, <http://www.opengeospatial.org/standards/gml>.

GML in elementi grafici, ma anche in questo caso tutte le applicazioni moderne GIS *desktop* e gran parte delle applicazioni web dispongono di strumenti dedicati di interazione con il protocollo WFS.

WCS

Il Web Coverage Service è lo standard OGC più recente, che estende notevolmente le funzionalità dei WMS²⁵. Nel linguaggio GIS, i *coverage* sono rappresentazioni di fenomeni complessi che variano nello spazio o nel tempo²⁶. A differenza del WMS, il WCS non restituisce *raster* statici, bensì dati complessi che possono essere visualizzati come immagini *raster* ma anche analizzati in dettaglio per le loro caratteristiche intrinseche (valori numerici associati ai *pixel*). Semplificando molto, si può affermare che il WMS è un protocollo adatto per la trasmissione di temi cartografici, mentre il WFS e il WCS sono protocolli per la trasmissione di dati cartografici. I temi si possono solo visualizzare, mentre i dati si possono anche analizzare e ulteriormente elaborare²⁷.

Il WCS restituisce quindi dati (*raster*) in formato XML GML e offre la possibilità di trattare il tema pubblicato come un *file* locale, permettendo dunque l'analisi, la personalizzazione della resa grafica (simbologia), e anche la conversione di formato e quindi lo scaricamento.

Come gli altri due protocolli, è possibile utilizzare il WCS attraverso i principali *software* GIS e applicazioni web. Ancora una volta è utile un esempio dal Portale Cartografico Italiano, dal quale si raggiunge all'URL: http://wms.pcn.minambiente.it/wcs/dtm_20m il modello digitale del terreno (DTM) della Repubblica Italiana, con una risoluzione a terra di 20m. Caricato su un visualizzatore, questo servizio risulta molto simile all'omologo WMS: http://wms.pcn.minambiente.it/ogc?map=/ms_ogc/WMS_v1.3/raster/DTM_20M.map, ma, a differenza di quest'ultimo, il tema può essere interrogato, analizzato e reso graficamente con stili e simbologia

²⁵ <http://www.opengeospatial.org/standards/wcs>.

²⁶ Il termine *coverage* non va confuso con l'ESRI *coverage*, un formato *file* avanzato per dati vettoriali (<http://desktop.arcgis.com/en/arcmap/10.3/manage-data/coverages/what-is-a-coverage.htm>).

²⁷ Per fare un ultimo esempio, la differenza tra WMS e WCS per un tema geografico è simile a quella tra un *file* PDF impaginato e un *file* MS Word (o LibreOffice Writer). Se la finalità è la sola lettura, non vi è grande differenza tra i due, anzi il formato PDF rispecchierà con maggiore fedeltà le intenzioni dell'autore, al sicuro da variabili locali (mancanza di font, differenza di versione del programma di visualizzazione, ecc.). Se però il fine è quello di fornire dei dati da riutilizzare, ovvero un testo modificabile, un formato Word è di gran lunga preferibile.

diversa da quella predefinita.

Il WMS è dunque ideale per la pubblicazione finale di piattaforme GIS concluse, con finalità per lo più illustrative; il WCS, invece, al pari del WFS, è più adatto alla condivisione dei dati geografici nell'ottica della collaborazione e del riutilizzo.

W/TMS

Il Tile Map Service (TMS) è una specifica di OSGeo²⁸ che offre un servizio simile al WMS, in quanto su richiesta del *client* vengono restituiti estratti *raster* di una mappa precompilata, ma di implementazione estremamente più semplice, in quanto non è richiesta la configurazione di un *server* geografico. La mappa complessiva, già esportata in un formato statico *raster*, viene suddivisa in moduli (*tile*) a vari livelli di zoom e a intervalli di coordinate prestabilite, che vengono numerati (indicizzati) e quindi offerti in risposta a ogni richiesta. La creazione dei *tile* viene realizzata in maniera automatica e semplificata, attraverso *plugin* dedicati di cui spesso i *software* GIS sono attrezzati.

Questo protocollo è stato assunto come base dall'OGC per un suo standard, chiamato Web Map Tile Service (WTMS)²⁹, molto simile come concezione a quello appena descritto. La differenza più importante tra due servizi è la diversa numerazione dell'asse Y: in TMS aumenta dal basso verso l'alto, mentre in WTMS aumenta dall'alto verso il basso. Si tratta di una differenza concettualmente trascurabile, ma che può rilevarsi importantissima nei casi applicativi.

Essendo i *tile* pre-costruiti, la risposta del *server* è molto veloce (a differenza dei servizi WMS, WFS e WCS che devono costruire dinamicamente a ogni richiesta la mappa dal progetto GIS); non sussiste però la possibilità di ricavare dalle mappe informazioni aggiuntive, come i metadati che riguardano i vari elementi rappresentati, né di interrogarle o analizzarle, anche in modo semplice. Si tratta dunque di un servizio meramente illustrativo, non adatto all'analisi e alla rielaborazione. Ma la facilità di implementazione e usabilità ne hanno decretato una grandissima diffusione nelle applicazioni web, dove è diventato di fatto il protocollo più utilizzato, mentre non sempre viene implementato nei *software desktop*, dove è più frequente impiegato tramite opzionali estensioni di funzionalità (*plugin*).

²⁸ https://wiki.osgeo.org/wiki/Tile_Map_Service_Specification.

²⁹ <https://www.opengeospatial.org/standards/wmts>.

XYZ

Non si tratta di una specifica vera e propria, ma di uno sviluppo generico di W/TMS, in quanto offre uno schema per l'indicizzazione di *tile* precostruiti. Si tratta più che altro di un'etichetta di varie implementazioni possibili, non necessariamente compatibili tra loro.

Dietro le quinte, *server* cartografici

Una rapida panoramica sul tema dell'infrastruttura necessaria per la pubblicazione di questi servizi potrebbe essere utile a chi è intenzionato a pubblicare dati cartografici, provenienti sia da studi personali, sia da progetti di ricerca più articolati e multidisciplinari. Si tratteranno qui gli aspetti generali senza scendere nei dettagli tecnici, poiché le implementazioni concrete richiedono una conoscenza informatica piuttosto dettagliata dei singoli programmi e protocolli. Per chi fosse interessato, le guide e i manuali dei singoli pacchetti *software* costituiscono una buona base di partenza.

Per pubblicare un qualsiasi documento sul Web, come si è già visto, è necessario disporre di un *server*, una macchina che materialmente contenga i dati, dotata di *software* specifici per rispondere alle richieste degli utenti (*client*). Tutti i servizi sopra elencati utilizzano il protocollo HTTP per la ricezione della richiesta e la restituzione dei dati, e si inquadrano, dunque, a pieno titolo come servizi Web. Per distribuire *tile* precostituiti (XYZ o W/TMS) – e quindi disponibili come *file* statici – è sufficiente un normale *server web*³⁰. Per i servizi dinamici, invece, dove la mappa che viene inviata al client viene creata “al volo” a ogni richiesta, è necessario disporre di programmi specifici in grado di elaborare dati cartografici sulla base delle richieste dell'utente. Si tratta di veri e propri *software* GIS, senza interfacce grafiche, sviluppati e ottimizzati per l'interazione con un *server web* e che di norma implementano i protocolli dell'OGC sopra descritti.

MapServer³¹ è certamente il *server* geografico più longevo oggi in circolazione, dal momento che i lavori per il suo sviluppo sono cominciati nel 1994 presso l'Università di Minnesota, con il supporto della NASA, e il rilascio di una versione stabile (1.0) è avvenuto già nel 1997. MapServer opera con i protocolli WMS, WFS, WCS ecc. e, facendo uso di GDAL/OGR, riesce a leggere un grande numero di formati *raster* e vettoriali. È

³⁰ I *software* oggi più diffusi sono Apache HTTP Server, Nginx, Lighttpd, Apache Tomcat, Microsoft IIS, ecc.

³¹ <https://www.mapserver.org/>, <https://www.osgeo.org/projects/mapserver/>.

```

# The annotated map file (sort of)
# Created by Pericles S. Nacionales for the MapServer tutorial
# 20050408
#
# MapServer map file uses the pound sign (#) to denote the start of a line
# comment--each line that needs to be commented has to be prepended with a "#".
#
# Map files begin with MAP keyword to signify the start of the map object.
# Well, the entire map file is THE map object. Enclosed between MAP and END
# at the very bottom of this map file, are keyword/value pairs and other
# objects.
MAP
  IMAGETYPE      PNG
  EXTENT         -97.238976 41.619778 -82.122902 49.385620
  SIZE           400 300
  SHAPEPATH      "/ms4w/apps/tutorial/data"
  IMAGECOLOR     255 255 255

# Layer objects are defined beneath the map object. You need at least one
# layer defined in your map file before you can display a map... You can
# define as many layers as you'd like although a limit is typically hard-coded
# in map.h in the MapServer source. The default limit is set at 100. You'd
# have to have a very specialized application to need more than 100 layers in
# your application.

# Start of LAYER DEFINITIONS -----
LAYER # States polygon layer begins here
  NAME          states
  DATA         states_ogr
  STATUS        OFF
  TYPE          POLYGON

# The class object is defined within the layer object. You can define as
# many classes as you need (well, there are limits as with layers, but it's
# senseless to define more than ten on a "normal" layer. There are
# situations, however, where you might have to do it.)
CLASS
  NAME          "The Upper Great Lakes States"

# There are styles in a class, just like there are classes in a layer.
# just like there are layers in a map. You can define multiple styles in
# a class just as you can define multiple classes in a layer and multiple
# layers in a map.
STYLE
  COLOR         232 232 232
  OUTLINECOLOR 32 32 32
END
END
END # States polygon layer ends here
# End of LAYER DEFINITIONS -----
END # All map files must come to an end just as all other things must come to...

```

Esempio di file di configurazione per la pubblicazione di uno shapefile con Mapserver. Fonte: <https://mapserver.org/tutorial/example1-1-map.html#example1-1-map>.

possibile interagire con il *server* usando diversi linguaggi di programmazione generici (PHP, Python, Perl, Ruby, Java, .NET, ecc.), il che permette il suo utilizzo in molte applicazioni web. È stato a lungo uno dei *software* più diffusi e rimane tutt'oggi uno dei più veloci. È rilasciato con licenza *open source* (MIT), è un progetto di OSGeo ed è disponibile per tutte le principali piattaforme (Linux, Windows, MacOS). Tutti i servizi del Portale Cartografico Nazione Italiano sono implementati utilizzando questo *software*.

GeoServer³² è anch'esso un progetto di OSGeo ed è oggi uno strumento in rapida diffusione. Rilasciato con licenza GNU GPL e scritto in Java, è disponibile per tutte le principali piattaforme ed è caratterizzato da una forte adesione ai protocolli definiti dall'OGC. A differenza di MapServer,

³² <http://geoserver.org/>, <https://www.osgeo.org/projects/geoserver/>.

GeoServer è dotato di un'interfaccia web molto semplice e intuitiva per il caricamento e la gestione dei dati geografici da pubblicare e per la configurazione del *server* stesso. Il suo successo è dovuto alla facilità di pubblicazione dei dati e alla conformità agli standard aperti.

The screenshot shows the GeoServer web interface. At the top left is the GeoServer logo. The top right shows 'Logged in as admin.' and a 'Logout' button. The main content area is titled 'Layer Preview' and contains a list of layers. The list has columns for 'Type', 'Title', 'Name', 'Common Formats', and 'All Formats'. The layers listed are all of type 'OpenLayers KML' and have a 'Select one' dropdown for 'All Formats'. The layers are named 'glacier_peak_2015_dtm_...' with various dtm values. The sidebar on the left contains navigation options: 'About & Status', 'Data', 'Services', 'Settings', and 'Tile Caching'. The 'Data' section is expanded, showing 'Layer Preview', 'Workspaces', 'Stores', 'Layers', 'Layer Groups', and 'Styles'. The 'Services' section shows 'WMTS', 'WCS', 'WFS', and 'WMS'. The 'Settings' section shows 'Global', 'Image Processing', and 'Raster Access'. The 'Tile Caching' section shows 'Tile Layers', 'Caching Defaults', and 'Gridsets'.

Interfaccia web di gestione dei dati di GeoServer.

Anche il noto programma GIS *desktop* QGIS³³ ha sviluppato una funzionalità *server*, che permette la facile pubblicazione dei progetti creati in versione locale. In particolare, è possibile sfruttare tutte le capacità di filtro, analisi, simbologia, ecc., e pubblicare interi progetti o singoli *layer* semplicemente copiando la cartella di lavoro sul *server* remoto. Si tratta, anche in questo caso, di una soluzione molto elastica e performante, che permette di tenere in sincronia il lavoro quotidiano con la versione pubblicata sul Web e rappresenta la scelta ideale per singoli ricercatori o piccoli progetti che vogliono dare un accesso pubblico e immediato ai propri *dataset* geografici.

Ci sono naturalmente molte altre possibilità, tra cui le soluzioni *server* offerte da ESRI, ma per quanto riguarda il mondo dell'*open source*, probabilmente quello più facilmente accessibile al ricercatore singolo o al piccolo-medio gruppo di ricerca, queste sono a oggi, nel 2019, le alternative più diffuse.

33 <https://qgis.org/>.

Sul *browser*: standard, librerie e strumenti webGIS

In questo panorama di estremo dinamismo, tipico del Web, esistono però alcuni progetti tecnologici che hanno negli anni fornito una tale stabilità e continuità di sviluppo e implementazione da essere considerati dei punti fermi: nella giungla di tecnologie, linguaggi di programmazione, paradigmi di sviluppo che caratterizza il mondo del World Wide Web, è emerso e si sta imponendo oggi – tra i tanti – un linguaggio per la creazione di quelle che vengono definite Rich Web Applications (tra le quali possiamo annoverare tutte le applicazioni di banche dati web e webGIS); ovvero il JavaScript³⁴. Si tratta di un linguaggio di *scripting*, creato nel 1995 da Brendan Eich e che è oggi supportato e utilizzato da tutti i *browser* in circolazione, il che ne fa uno dei linguaggi di programmazione più diffusi³⁵. Altre tecnologie usate in passato per ottenere nei *browser*, funzionalità tipiche dei *software desktop*, quali Adobe Flash, applet Java, Microsoft Silverlight, ecc., sono oggi ampiamente superate, il che rende spesso necessarie operazioni di tipo “archeologico”, per il recupero dei dati e applicazioni basati su quest’ultime.

Per quanto riguarda l’ambito molto specifico della rappresentazione cartografica, si presentano di seguito alcuni strumenti molto diversi da loro, tutti implementabili in JavaScript, che nel loro insieme coprono la quasi totalità delle applicazioni GIS destinate al Web (webGIS).

Google Maps

Lanciato come servizio nel 2005 dal colosso dell’informatica Google, non è solamente una piattaforma di distribuzione di immagini satellitari e cartografia di base (nonché di varie soluzioni miste) per il pubblico generico: alle funzionalità di navigazione, Google ha difatti aggiunto potenti strumenti di ricerca, integrati in Google Search, che trasformano questo

³⁴ Douglas Crockford, «JavaScript: The World’s Most Misunderstood Programming Language», 2011, <https://www.crockford.com/javascript/javascript.html>.

³⁵ Il JavaScript è un linguaggio lato *client*, a differenza per esempio di PHP, che è un linguaggio lato *server*. Ciò significa che le istruzioni del programma vengono eseguite da parte del *client* (ovvero vengono eseguite dal *browser*) mentre i linguaggi *server* vengono interpretati nella macchina remota (*server*). Dal momento che non vi è una implementazione standard del linguaggio sui vari browser, si ha spesso il paradosso che la stessa istruzione venga interpretata in maniera differente da *software* diversi. Internet Explorer, il cui sviluppo è stato ormai abbandonato, è famoso per impiegare una implementazione non standard del JavaScript, il che porta molte applicazioni moderne a non essere compatibili con questo *browser*.

prodotto in un vero e proprio motore di ricerca con interfaccia geografica. La casa madre ha inoltre reso disponibile una specifica API (Application Programming Interface) Javascript che consente agli sviluppatori web di incorporare (*embed*) Google Maps nelle proprie pagine, arricchendone le basi cartografiche con i propri dati. Punti, linee e poligoni possono essere facilmente aggiunti, collegando la pagina a una fonte di dati statica in KML o anche a fonti di dati più strutturate, come il *geodatabase* del proprio progetto GIS *desktop*. Questo permette di pubblicare su pagine HTML, per mezzo di poche linee di codice JavaScript, mappe e temi cartografici anche piuttosto complessi. Fino al 2018 il servizio era gratuito; successivamente, l'attività dei visitatori è sottoposta a puntuali misurazioni e il servizio prevede un pagamento, una volta superata una soglia predefinita di utilizzo.

OpenLayers

OpenLayers³⁶ è una libreria JavaScript, rilasciata per la prima volta nel 2006, che, in modo simile a Google Maps, semplifica la creazione e la visualizzazione di mappe dinamiche su pagine web. A differenza di Google Maps, OpenLayers è rilasciato con licenza *open source*³⁷ e quindi non pone limiti (presenti o futuri) al suo utilizzo; inoltre, può essere scaricato e usato anche in applicazioni *offline*, mentre Google Maps, permette l'utilizzo della sua libreria solo come servizio web. Si tratta di una soluzione professionale, modulare e molto ricca di funzionalità, adatta sia per progetti semplici, di piccola scala e ridotte capacità, sia per soluzioni GIS complete e piuttosto avanzate. Permette di lavorare con un gran numero di formati vettoriali (KML, GML, geoJSON, TopoJSON, GPX, WKT, ecc.), protocolli (WMS, WFS, WCS, (W)TMS, XYZ, le basi di Google Maps, Bing Maps, ecc.), e immagini *raster*. Permette inoltre di compiere analisi statistiche di base sia sui temi vettoriali che su quelli *raster* e soprattutto consente di estendere la proprie funzionalità attraverso librerie di terze parti, per

Esempio di una semplice analisi spaziale realizzata con Turf.js e OpenLayers: lungo una linea sono visualizzati dei punti ogni 200 metri. Fonte <https://openlayers.org/en/latest/examples/turf.html>



³⁶ <https://openlayers.org/>.

³⁷ FreeBSD o 2-clause BSD, <https://opensource.org/licenses/BSD-2-Clause>.

svolgere analisi statistiche, spaziali e topologiche piuttosto complesse. Si veda a titolo esemplificativo la possibilità di integrare in OpenLayers Turf.js, una libreria *open source* per complesse analisi geospaziali³⁸, o l'integrazione con Proj4js³⁹, per la trasformazione di coordinate e la riproiezione “al volo” di dati *raster* e vettoriali tra molti sistemi diversi, un'opzione quest'ultima fondamentale per chi lavora con *dataset* GIS articolati e di diversa provenienza.

Allo stato attuale, OpenLayers è certamente la soluzione più avanzata e completa per la produzione di piattaforme GIS destinate all'utilizzo sul Web.

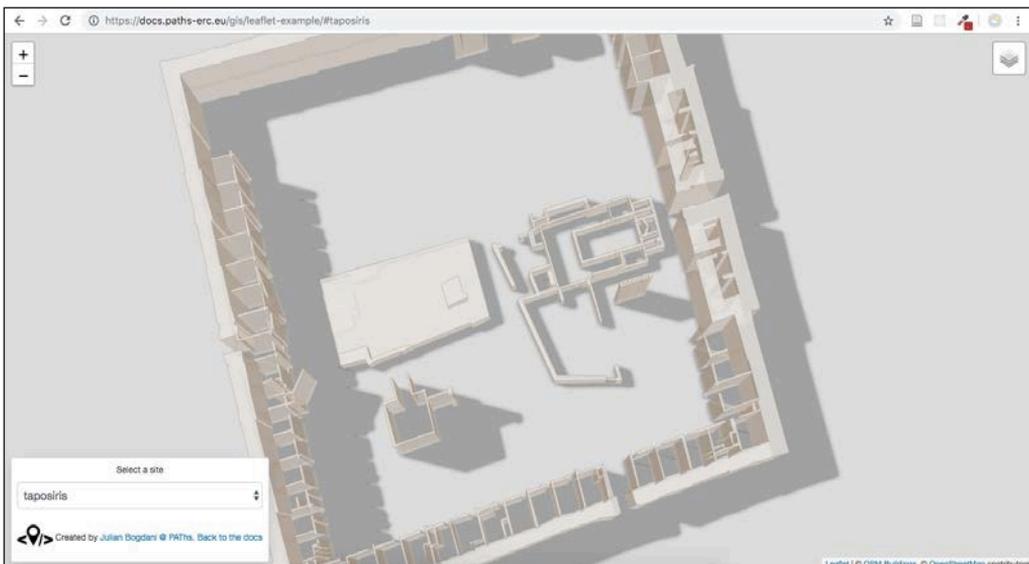
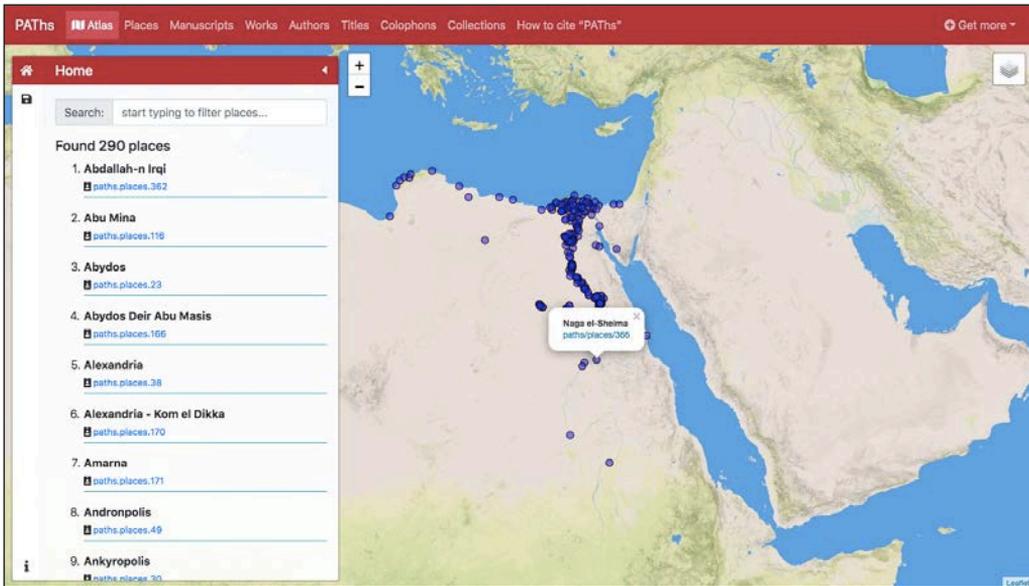
Leaflet

Leaflet⁴⁰, come OpenLayers, è una libreria JavaScript *open source* (stessa tipologia di licenza) nata per opera di Vladimir Agafonkin come alternativa più semplice e leggera allo stesso di OpenLayers, rivolta in particolare all'utilizzo sui dispositivi mobili. La versione 2 di OpenLayers infatti era piuttosto pesante e si era orientata verso un uso quasi esclusivamente professionale: attenta a supportare anche i *browser* più vecchi, non aveva un supporto ottimale per i dispositivi mobili più avanzati. Leaflet tende invece alla semplicità: supporta pochissimi formati geografici (solamente GeoJSON, (W)TMS/YXZ e WMS), ma può essere esteso con *plugin* creati dagli stessi utenti. Nonostante non sia adatto all'analisi spaziale dei dati e alla creazione di temi avanzati, e nonostante rimanga molto limitato nei formati di utilizzo, Leaflet ha conosciuto una rapidissima diffusione, anche nello sviluppo GIS più professionale, traendo vantaggio – ma al contempo condizionando – una maggiore standardizzazione dei formati vettoriali, tra i quali il GeoJSON è divenuto il più utilizzato.

³⁸ Per un esempio di integrazione delle due librerie si veda <https://openlayers.org/en/latest/examples/turf.html>. Turf.js è disponibile all'indirizzo <http://turfjs.org/>.

³⁹ <http://proj4js.org>.

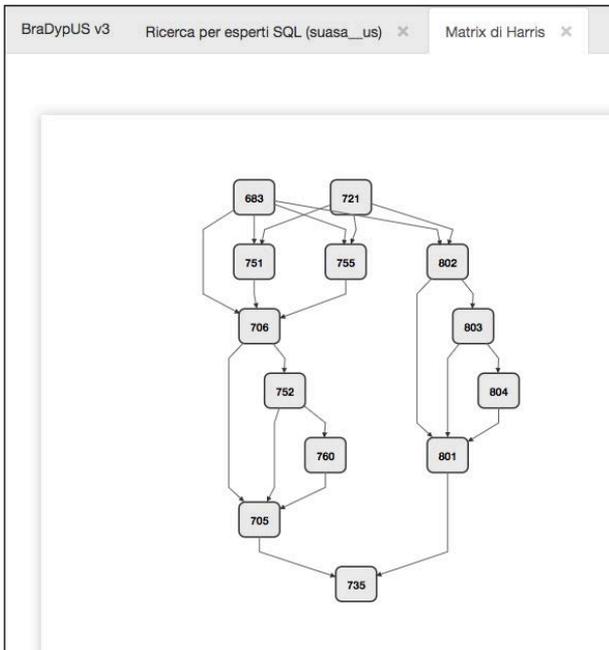
⁴⁰ <https://leafletjs.com/>.



In alto, vista estratta dall'Atlante archeologico della letteratura copta del progetto PATHs (<https://atlas.paths-erc.eu>), diretto da P. Buzi, Sapienza Università di Roma. La parte geografica dell'applicazione web è stata realizzata con Leaflet da chi scrive. In basso, esempio di visualizzazione di dati 3D attraverso l'integrazione in Leaflet di altre librerie; chiesa copta all'interno del temenos del sito di Taposiris Magna (fonte Atlante PATHs, <https://docs.paths-erc.eu/gis/leaflet-example/#taposiris>).

D3.js

A differenza delle altre librerie/servizi finora trattati, D3.js⁴¹ non è dedicato alla creazione e alla visualizzazione di mappe; è una libreria *open source*⁴² per la manipolazione e produzione di documenti basati sui dati e viene utilizzato nei tanti ambiti che pongono al centro della loro attenzione i dati (*data science*). Grafici di ogni sorta, grafi, *network*, analisi statistiche e anche mappe sono alcuni dei prodotti che si possono ottenere con D3.js. Non essendo rivolto a un pubblico generico ma a specialisti di analisi di dati, la sua applicazione richiede competenze avanzate, come avanzate sono le sue possibilità espressive. È infatti possibile impiegarlo su varie tipologie di mappe, per rappresentare dati quantitativi di varia natura e, se abbinato a OpenLayers⁴³, le rispettive peculiarità si valorizzano reciprocamente: OpenLayers per gestione della base cartografica e D3.js per la visualizzazione dei dati complessi.



Matrix di Harris relativo alle tombe T505 e T506 della Necropoli Orientale di Suasa (Castelleone di Suasa, AN) creato in maniera automatica utilizzando D3.js (fonte web-database degli scavi dell'Università di Bologna a Suasa, curato da chi scrive, <https://db.bradypus.net/#/suasa>).

ArcGIS API for JavaScript

Non può mancare, in una panoramica sulla cartografia, la casa produttrice di *software* GIS più importante, ESRI, che ha rilasciato una propria API JavaScript⁴⁴ per consentire l'integrazione su pagine web di mappe e carto-

⁴¹ <https://d3js.org/>, primo rilascio nel 2011.

⁴² Licenza BSD, <https://opensource.org/licenses/BSD-3-Clause>.

⁴³ Per un esempio applicativo si veda <https://openlayers.org/en/latest/examples/d3.html>.

⁴⁴ <https://developers.arcgis.com/javascript/>.

grafie create con i prodotti del marchio, in due e tre dimensioni. Si tratta però di un ecosistema chiuso, in quanto la libreria è pensata principalmente per interfacciarsi con il *server* geografico creato da ESRI⁴⁵ e quindi trova poche o nessuna applicazione in altri contesti.

Non si tratta dunque di un'alternativa di utilizzo generico, ma di una soluzione rivolta a una porzione molto specifica del mercato: per chi già utilizza il *software* e i servizi di ESRI, ArcGIS API è senz'altro uno strumento molto valido per la pubblicazione e l'esposizione dei propri dati geografici in rete.

Mapbox

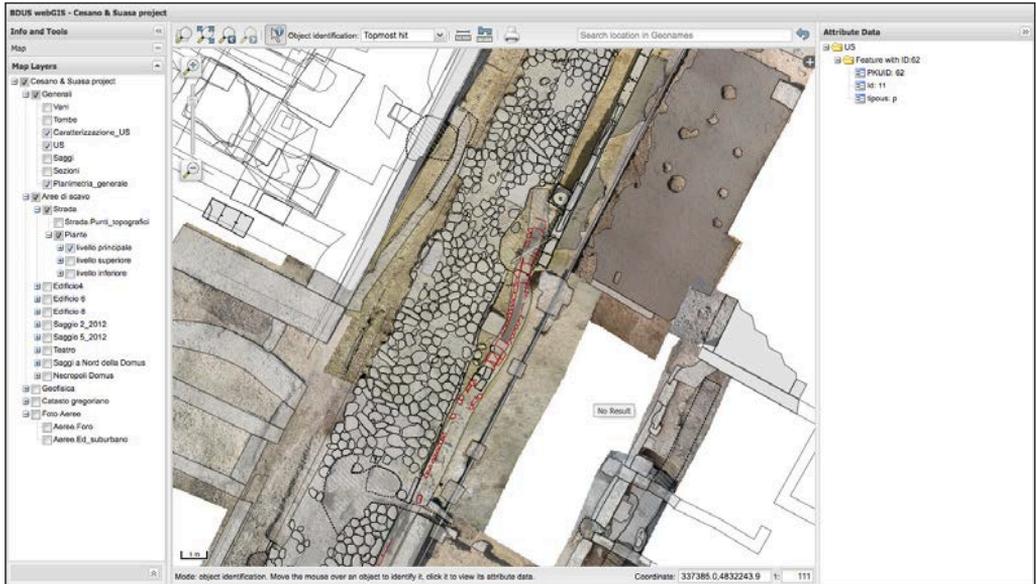
Mapbox⁴⁶ è una società fondata nel 2010, che ha basato il proprio modello di business nella fornitura di basi cartografiche altamente personalizzate e su misura. Un servizio di *hosting*, in definitiva, con costi scalari. Nella costruzione della propria infrastruttura Mapbox ha creato diversi strumenti, poi rilasciati con licenze aperte, e alcuni formati (come MBTiles) che hanno trovato larghissima diffusione e utilizzo. Come altre società⁴⁷, l'attenzione è rivolta a professionisti del settore Web, senza particolari esperienze GIS, che attraverso strumenti *online* di facile e intuitivo utilizzo vengono messi in condizione di creare cartografia destinata a essere pubblicata sul Web. Mapbox usa una API basata su Leaflet, estesa per aggiungere funzionalità dedicate. Per la semplicità di utilizzo e per la convenienza del servizio di *hosting*, diversi progetti di piccola e media scala (ma non solo) hanno scelto questo modello. Si tratta di un servizio particolarmente adatto anche a progetti personali, che per le loro piccole dimensioni rientrano facilmente nella fascia gratuita.

La necessità di sintesi ci impone di terminare qui la panoramica, che tuttavia rimarrebbe in ogni caso parziale perché le iniziative – sia per quanto riguarda la creazione di librerie e *software*, sia per le soluzioni che rientrano nella dinamica del *Software-as-a-Service* (*SaaS*) – sono tantissime. Numerosi sono anche i *framework* (strutture di supporto per la creazione di un *software*) che a partire da queste librerie – o più raramente imple-

⁴⁵ ArcGIS Server, <http://enterprise.arcgis.com/en/server/>.

⁴⁶ <https://www.mapbox.com/>.

⁴⁷ Ad esempio Carto (ex CartoDB), <https://carto.com/> che ha introdotto il CartoCSS, un linguaggio cartografico, simile al CSS, per la creazione di mappe a partire da dati strutturati.



mentando le proprie – permettono di creare vere e proprie applicazioni webGIS avanzate, spesso comprendendo nello stesso pacchetto anche la parte *server*. Non è possibile offrire in questa sede una trattazione né un elenco di questi strumenti, largamente adottati anche in archeologia, perché il mercato è estremamente dinamico e in continuo cambiamento. Alcune soluzioni webGIS piuttosto complete, *open source* e molto diffuse, abbastanza longeve e stabili nel tempo sono GeoExt⁴⁸, Lizmap⁴⁹, QGIS Web Client⁵⁰, Extended QGIS Web Client⁵¹, G3W Suite⁵², GISQUICK⁵³ e molte altre. Comune caratteristica di questi progetti è quella di facilitare la pubblicazione *online* di dati geografici, già strutturati in una piattaforma GIS, mediante interfacce utente molto gradevoli, che integrano, attraverso il *browser*, strumenti avanzati per l'interrogazione dei dati, la loro analisi, la possibilità di fare misurazioni di vario genere e infine la funzionalità di creare tavole da esportare per la presentazione e la stampa⁵⁴. La loro

Esempio di una piattaforma webGIS creata con QGIS Web Client e relativa agli scavi nella città romana di Suasa (Castelleone di Suasa, AN) dell'Università di Bologna. GIS e webGIS sono curati da chi scrive.

⁴⁸ <https://geoext.github.io/geoext3/>.

⁴⁹ <https://www.3liz.com/lizmap.html>.

⁵⁰ <https://github.com/qgis/QGIS-Web-Client> e la sua versione più aggiornata <https://github.com/qgis/qwc2-demo-app>.

⁵¹ <https://github.com/uprel/gisapp>.

⁵² <https://g3wsuite.gis3w.it>.

⁵³ <http://gisquick.org/>.

⁵⁴ Di norma queste soluzioni complete professionali basano la visualizzazione

installazione e configurazione richiede nella maggior parte dei casi qualche conoscenza informatica avanzata, ma sempre più spesso il processo di pubblicazione viene semplificato e talvolta integrato sotto forma di *plugin* negli stessi *software GIS desktop*. È facile immaginare che nel prossimo futuro la progressiva semplificazione delle procedure di edizione porti a un effettivo incremento delle banche dati geografiche *online*, anche per mezzo di soggetti che, finora, pur essendo produttori di dati, non sono stati in grado di pubblicarli. Studenti, dottorandi o anche piccoli progetti, che per mancanza di conoscenze informatiche o risorse finanziarie sono stati finora solo utenti passivi della Rete, potranno verosimilmente contribuire attivamente all'incremento delle risorse geografiche liberamente disponibili *online*.

Publicare dati archeologici geografici sul web

Come è stato già detto, il mondo dell'archeologia ha precocemente sperimentato le tecnologie di Rete, anche per quello che riguarda la pubblicazione dei dati spaziali e/o geografici, fondamentali nella ricerca sul campo e in particolare in quella su scala territoriale. Il numero delle sperimentazioni di GIS negli ultimi trent'anni è stato altissimo – basta sfogliare una rivista *open access* come “Archeologia e Calcolatori”, i suoi numeri speciali e supplementi (<http://www.archcalc.cnr.it>) per averne un'idea – e molti sono stati anche i tentativi di portare questa preziosa mole di informazioni in Rete, con il duplice scopo di offrire una migliore chiave di lettura alle proprie attività (a scopo illustrativo, si direbbe) e di condividere i dati a disposizione, favorendone l'uso e il riutilizzo. Non pochi e tutt'altro che trascurabili, per l'impatto che hanno avuto sullo sviluppo della metodologia della ricerca, sono inoltre i casi di pura sperimentazione, mirati a sondare i limiti dell'applicazione di questi strumenti e i motivi per cui alcuni obiettivi non sono ancora stati raggiunti (aporie tecnologiche, lacune nella formazione specialistica oppure semplicemente mancanza di interesse di ricerca). Meno incisivo – ma pur sempre esistente – è il limite economico, considerata l'ampia disponibilità di soluzioni *open source* di altissimo livello qualitativo e semplicità di implementazione. Infine, un ostacolo che purtroppo ci si trova ad affrontare è il divario tecnologico che separa geograficamente la ricerca scientifico-accademica e istituzionale condotta in aree del globo ad altissimo dinamismo da quella sviluppata in aree più

dei dati geografici su OpenLayers.

problematiche⁵⁵. La carenza di connessione a banda larga costituisce un impedimento notevole e la mancanza di infrastrutture moderne e adeguate per la ricerca ne costituisce un secondo.

Un incentivo importante è costituito da INSPIRE⁵⁶, una direttiva della Comunità Europea per la creazione di un'infrastruttura di dati spaziali (SDI, *Spatial Data Infrastructure*) di supporto alle politiche strettamente ambientali o alle attività che, più in generale, potrebbero avere un impatto sull'ambiente. Questa infrastruttura permette la condivisione di informazioni spaziali relative all'ambiente tra organizzazioni del settore pubblico, ne facilita il libero accesso all'interno della Comunità Europea e aiuta i processi di messa a punto di programmi dedicati, in un'ottica di supera-

The screenshot displays the INSPIRE Geoportal interface. At the top, there is a navigation bar with the European Commission logo, the text 'INSPIRE GEOPORTAL', and the tagline 'Enhancing access to European spatial data'. Below this, a breadcrumb trail reads 'European Commission > INSPIRE > Geoportals > Priority Data Sets Viewer'. A secondary navigation bar includes links for 'Home', 'Priority Data Sets Viewer', 'INSPIRE Thematic Viewer', 'Harvesting status', and 'Find out more about'. The main content area features a map titled 'Priority Data Sets - EU & EFTA Country overview' showing Europe with various countries highlighted in yellow. To the right of the map, there are controls for 'Show:' with 'Downloadable' and 'Viewable' options, and a statistics section titled 'INSPIRE Geoportals Data Set Statistics' showing '1259 Metadata records' and '482 Downloadable Data Sets'. The footer of the map area includes 'Leaflet | Credits: © OpenStreetMap contributors | EC-GISCO, © EuroGeographics for the administrative boundaries (©ecismef)'.

Pagina del geoportale del progetto INSPIRE, <http://inspire-geoportal.ec.europa.eu/overview.html?view=pdEu-Overview&legislation=all>

⁵⁵ Quello del divario tecnologico e infrastrutturale è un problema relevantissimo anche all'interno di ambiti dei quali ci si aspetterebbe una maggiore uniformità, quale potrebbe essere la Comunità Europea. Nello specifico campo della gestione dei dati archeologici si è notato una differenza importante tra paesi dotati da infrastrutture di gestione e ricerca e paesi che ne erano del tutto sprovvisti; in questi ultimi i dati archeologici digitali sono sotto rischio di definitiva scomparsa, Holly Wright e Julian D. Richards, «Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures», *Journal of Field Archaeology* 43, n. S1 (2018): S64, <https://doi.org/10.1080/00934690.2018.1511960>.

⁵⁶ Infrastructure for Spatial Information in the European Community, <https://inspire.ec.europa.eu/>.

mento dei confini nazionali⁵⁷. Attualmente, il portale ufficiale comprende anche un geoportale⁵⁸ che raccoglie tutti i dati e i metadati finora pubblicati dai vari paesi membri, liberamente accessibili e consultabili attraverso vari formati e tipo di servizi. Una pagina specifica tiene conto dello stato di aggiornamento della banca dati principale da parte delle istituzioni pubbliche dei paesi partecipanti⁵⁹.

Questa direttiva è di estremo significato, poiché il suo recepimento ha portato alla nascita del Portale Cartografico Nazionale (PCN), di cui si è già detto, che costituisce il nodo italiano della rete europea. Benché concentrato prevalentemente su tematiche ambientali, i dati che grazie a questa iniziativa sono stati resi liberamente accessibili attraverso i protocolli OGC sono di fondamentale importanza per il nostro lavoro, come si è già avuto modo di vedere. L'auspicio che si può avanzare per l'Italia è che anche a livello regionale si arrivi a un livello di disponibilità di dati simile a quello nazionale. Allo stato attuale, la mappa italiana è piuttosto disomogenea, con alcune Regioni significativamente avanzate nel processo di libera pubblicazione e altre decisamente più indietro. Di certo, comunque, iniziative europee come INSPIRE sono di grande ausilio per stabilire una base comune, legislativa e tecnica, e per stimolare realtà locali meno attive.

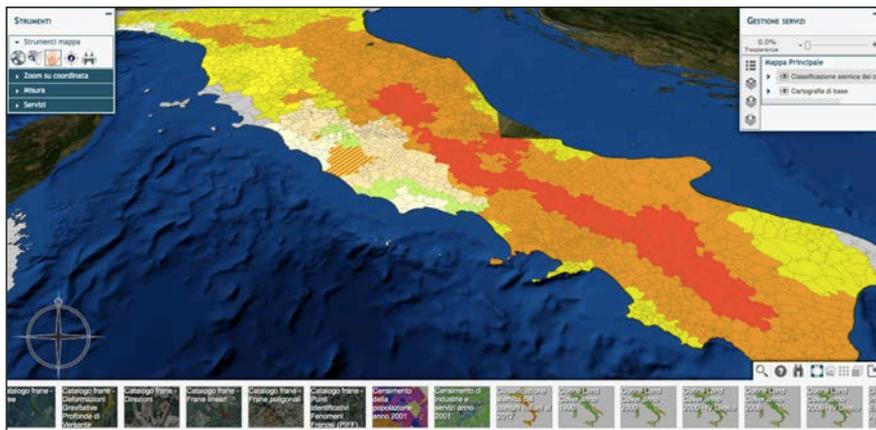
Chi volesse pubblicare dati GIS, ha quindi oggi a disposizione un'ampia gamma di standard e formati, mantenuti dall'OGC e OSGeo. Le diverse opzioni sono solo in parte sovrapponibili, come si è già avuto modo di vedere. I progetti di ricerca che utilizzano dati forniti da terze parti hanno necessariamente bisogno di protocolli di trasmissione e comunicazione che permettano l'analisi e l'accesso più approfondito ai dati, quindi preferibilmente WFS e WCS. Il WMS offre buone basi grafiche precompilate, ma è molto limitato nella ricerca e impedisce qualsiasi tipo di analisi; ancora più limitati sono i più semplici W/TMS o XYZ. Un esempio concreto potrebbe essere utile per chiarire l'importanza della pubblicazione dei dati a più livelli di elaborazione.

Consideriamo il caso abbastanza tipico, di una ricerca archeologica che faccia uso di strumenti GIS per la creazione di un modello di occupazione del territorio in un determinato periodo cronologico e l'eventuale successiva applicazione di questo modello ad altre aree, per le quali non si dispongano

⁵⁷ La mission del progetto è pubblicata all'indirizzo <https://inspire.ec.europa.eu/about-inspire/563>.

⁵⁸ <http://inspire-geoportal.ec.europa.eu/>.

⁵⁹ http://inspire-geoportal.ec.europa.eu/harvesting_status.html.



Portale Cartografico Nazionale, visualizzatore 3D con il tema della Classificazione sismica dei comuni italiani al 2012, fonte: <http://www.pcn.minambiente.it/viewer3D/>

informazioni sufficienti (una tipica operazione di archeologia predittiva⁶⁰). La mappa (o meglio il modello) dell'effettiva e probabile presenza umana nel dato territorio è frutto dell'attenta analisi ricostruttiva di dati provenienti da una pluralità di fonti diverse. La ricerca bibliografica e le indagini sul campo portano alla formazione di *geodatabase* di siti con dettagli su tipologia, estensione e cronologia. Se i dati bibliografici non sono disponibili già in forma strutturata (ad esempio una banca dati), è necessario portare a termine un preliminare lavoro di schedatura. Tutto va poi integrato con le informazioni sull'ambiente, che ha determinato, influenzato o è stato plasmato dall'attività dell'uomo, e con lo studio geologico e geomorfologico (geomorfologia, idrografia, acclività del terreno, ecc.) del territorio. I dati ambientali non sono mai frutto di rilievo diretto da parte dell'archeologo, il quale si avvale generalmente di fonti d'informazione esterna (geologia, geomorfologia, ecc.). Fino a poco tempo fa queste informazioni erano disponibili solo in formato cartaceo (o nella migliore delle ipotesi in formato digitale *raster*), e i processi di georiferimento e vettorializzazione, indispensabili per rendere utilizzabili questi tematismi, erano i più dispendiosi in termini di tempo. Oggi, il Portale Cartografico Nazionale o i portali cartografici delle singole Regioni costituiscono fonti importanti di dati strutturati, i quali, grazie al servizio WFS, possono essere facilmente archiviati per ulteriori lavorazioni. Al fine del nostro ipotetico studio pre-

⁶⁰ Sull'archeologia predittiva si veda H. Kamermans, M. Leusen, e Ph. Verhagen, *Archaeological Prediction and Risk Management: Alternatives to Current Practice* (Amsterdam: Leiden University Press, 2009). Per una sintesi sullo stato della disciplina nell'ambito italiano si veda Valeria Boi e Mirella Serlorenzi, «Archeologia preventiva, predittiva, potenziali archeologici. Una breve introduzione al panorama italiano», in *Proceedings of ArchoFOSS: Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica: VIII Edizione, Catania 2013*, a. c. di Filippo Stanco e Giovanni Gallo (Oxford: Archaeopress, 2016), 115-121..

dittivo, la rete idrografica è importante, ma è ancora più importante poterla “riadattare”, sistemandola per rappresentare una situazione più vicina alla fase antica che viene fatta oggetto di studio, quando verosimilmente i fiumi non avevano lo stesso corso che hanno oggi. Immaginiamo che un collega abbia compiuto studi approfonditi su questo specifico problema e oltre ai lavori di sintesi abbia pubblicato anche i dati vettoriali, magari usando protocolli aperti di Rete quali i WFS. In tal caso, non dovremmo far altro che “puntare” al suo URL e importare nel nostro progetto l’idrografia antica; in alternativa, dovremmo procedere ad acquisire digitalmente le sue pubblicazioni cartacee, georiferirne la cartografia e vettorializzare manualmente la rete idrica, un lavoro molto più lungo e oltremodo più impreciso. Queste considerazioni valgono anche per altri temi utili a questo tipo di ricerca, come i modelli di elevazione del terreno (DEM) variamente acquisiti e i modelli di andamento dei piani d’uso antichi.

A questo punto sarebbe possibile procedere all’analisi e alla modellizzazione della documentazione, alla ricerca delle costanti nei dati in nostro possesso, utili alla creazione di un modello di riferimento. Il progetto GIS così ottenuto potrebbe poi essere utile per creare delle belle tavole a corredo della nuova pubblicazione e infine archiviato. Oppure, i nostri dati potrebbero a loro volta essere pubblicati, non solo per illustrare su una pagina web e in modo dinamico i risultati del lavoro, ma anche per rendere nuovamente disponibili i *dataset* di partenza e di arrivo, così da consentire ad altri studiosi di leggere, riusare e arricchire il materiale di partenza senza ripercorrere di nuovo tutto il procedimento.

L’adozione di formati aperti e la pubblicazione dei dati geografici, grezzi e strutturati, non innescano, allo stato attuale, meccanismi di “premiatura” accademica paragonabili alle pubblicazioni tradizionali, un limite questo peculiare delle scienze umanistiche. I dati, tuttavia, sono un prerequisito fondamentale per l’avanzamento dello stato delle conoscenze, in uno scenario come quello di oggi caratterizzato dall’iperproduzione e dal sempre difficile reperimento di dati specialistici (ad esempio archeologici) grezzi, strutturati e a libero accesso. Esistono, a tal proposito, alcune iniziative editoriali interessanti, come il “Journal of Open Archaeology Data”⁶¹, una rivista scientifica, che prevede il processo della doppia revisione paritaria (*double blind peer review*) e che pubblica contributi che descrivono *dataset* archeologici ad alto potenziale di riutilizzo. Si tratta di uno strumento di grande utilità per approfondire la tematica delle possibili modalità e forme

⁶¹ <https://openarchaeologydata.metajnl.com/>.

di condivisione dei dati archeologici.

Cosa succederebbe, ad esempio, se per legge dello stato o per regolamento intero degli atenei italiani a tutti i neo-dottorati venisse posto l'obbligo di pubblicare sul Web, a libero accesso, i dati relativi alle proprie ricerche, così come viene loro imposto di rendere pubblico l'elaborato scritto? E cosa accadrebbe se tale obbligo venisse esteso a tutta la ricerca realizzata con finanziamento pubblico? Nei corsi, nei manuali e nelle dispense si potrebbe a quel punto fare riferimento anche a questi archivi (*repository*) e servizi, oltre a quelli generici forniti dagli enti pubblici; tali fonti potrebbero inoltre fare da base per sperimentazioni di visualizzazione e analisi incentrate sui dati, *mashup* di vario genere e livello, e potrebbero persino stimolare nuove ricerche. Anche il confronto tra ipotesi ricostruttive diverse e parallele ne trarrebbe giovamento, così come riceverebbero nuovi spunti la riflessione teorica e metodologica.

Un discorso diverso riguarda la presentazione di questi dati geografici e dei dati strutturati, in generale. È molto importante tenere ben distinto quello che è il processo di produzione, archiviazione e pubblicazione dei dati da quella che è la loro presentazione, anche se è difficile per gli archeologi separare questi due momenti, a causa dell'inveterato rifiuto a considerare i dati grezzi come un risultato da condividere o da pubblicare. Di norma, la forma principale di pubblicazione dei risultati di una ricerca archeologica è quella della sintesi, che può essere accompagnata o meno da un supporto documentario sotto forma di liste ed elenchi, più spesso in forma aggregata. Questo per limitare i costi di una pubblicazione "totale", che nel caso di un prodotto cartaceo potrebbero risultare davvero proibitivi. Peraltro, la riproduzione a stampa di lunghi elenchi e descrizioni di unità stratigrafiche, reperti archeologici e analisi di ogni sorta sarebbe in fondo pressoché inutile persino per il più motivato dei lettori. Una possibile soluzione al problema è stata quella di accompagnare le pubblicazioni con supporti digitali dedicati, inizialmente forniti su dispositivi ottici allegati (CD-ROM, DVD, ecc.) e più recentemente ospitati in Rete, su banche dati strutturate o applicazioni web più o meno complesse e più o meno aperte. Questi supporti presentano talvolta funzioni di filtraggio e di aggregazione dei dati che consentono varie tipologie di ricerca. Il problema di queste applicazioni, però, è la loro scarsa longevità, caratteristica che accomuna tutte le risorse digitali: un'applicazione distribuita su un supporto ottico (come un CD-ROM, DVD, Blue-Ray, ecc.) segue la vita del supporto, ma anche quella delle tecnologie, applicazioni e formati con i quali è stata cre-

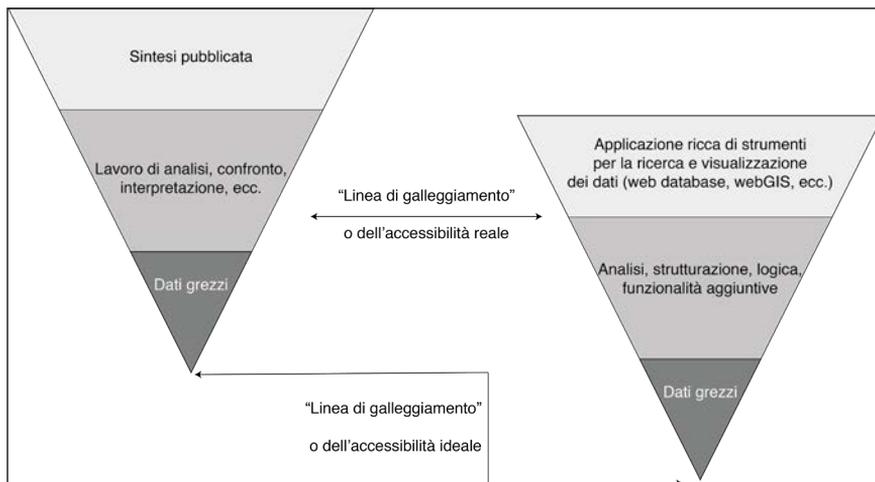
ata. Si pensi ad esempio alla tecnologia Adobe (prima Macromedia) Flash, ideata per aggiungere animazioni e contenuti multimediali per il Web e molto usata anche nella creazione di applicazioni da distribuire su supporti ottici. Oggi questa tecnologia è stata ampiamente superata dall'avvento di nuovi standard (principalmente HTML 5, CSS 3) più attenti all'accessibilità e all'apertura, e tutti i prodotti realizzati con Flash sono condannati all'obsolescenza. Inoltre, un'applicazione Flash una volta distribuita non può più essere aggiornata, a meno non si disponga – cosa che difficilmente accade – del codice sorgente, cosa che ne rende ancora più problematico il recupero dei dati⁶².

In generale, più il formato di distribuzione è semplice, più l'esperienza dell'utente è limitata e più è assicurata la durata nel tempo dei contenuti. Un *file* HTML creato agli albori del Web è ancora oggi facilmente leggibile da qualsiasi *browser*, e lo stesso dicasi per un *file* di testo semplice (*plain text*) o per una tabella di dati in formato CSV.

Questa dinamica si potrebbe rappresentare con l'immagine di una piramide rovesciata. La punta, in basso, corrisponde ai dati grezzi (*raw data*), più o meno strutturati, che di solito rimangono invisibili o comunque poco accessibili, mentre la base, in alto, rappresenta la sintesi che viene pubblicata. In questo caso non è alla "quantità" numerica a cui l'immagine della piramide si riferisce, bensì al potenziale informativo: se la sintesi è per definizione una forma concisa e ristretta – sintetica appunto – dei dati di partenza, dal punto di vista "qualitativo" rappresenta un arricchimento enorme, perché reca in sé un bagaglio importante di confronti e riferimenti esterni, ma anche – spesso ce ne dimentichiamo – un apporto personale fondamentale, fatto di esperienza, metodo e interpretazione.

Più si sale verso l'alto, dalla punta verso la base, più aumenta l'apporto dello studioso o del ricercatore che ha raccolto, analizzato, interpretato

⁶² Un esempio spesso citato quando si parla della fragilità dei formati e supporti digitali è quello del Domesday Project a cura della BBC, che negli anni Ottanta del secolo scorso raccolse frammenti di vita quotidiana da circa un milione di persone, con un ricchissimo apparato multimediale (immagini, video, audio, mappe interattive, dati statistici), per creare un'istantanea del Regno Unito, a 900 anni dalla compilazione del Domesday Book, il censimento dell'XI secolo fatto compilare da Guglielmo il Conquistatore. Negli anni Novanta la tecnologia d'avanguardia utilizzata per codificare le informazioni era diventata vecchia, i supporti particolari sviluppati *ad hoc* non erano più supportati da nessun *hardware* in circolazione e le aziende che avevano implementato la piattaforma non erano più attive. Nel 2011 alcuni contenuti furono faticosamente estratti e pubblicati su un sito web dedicato, oggi archiviato (Daniel J Cohen e Roy Rosenzweig, *Digital history: a guide to gathering, preserving, and presenting the past on the web* (Philadelphia: University of Pennsylvania Press, 2006), 224–25.).



Schema del progressivo aumento del potenziale comunicativo e informativo dei dati archeologici risultanti da un progetto di ricerca, a sinistra. A destra, il relativo processo di pubblicazione digitale di un dataset, attraverso la costruzione di applicazioni web ricche (database, portali webGIS). La linea di galleggiamento rappresenta il livello di accessibilità applicato agli utenti.

e rivestito i dati grezzi di nuovi significati. Analogo discorso per il processo di pubblicazione di dati e ricerca attraverso strumenti multimediali: più ci si allontana dai dati grezzi (*raw data*) e dai formati semplici, più il prodotto editoriale si articola e si dota di funzionalità aggiuntive che servono a fini specifici, come la possibilità di eseguire in modo semplice una ricerca complessa, di comporre una mappa o di visualizzare un grafico che restituisca un'idea più immediata e precisa delle relazioni quantitative dei dati numerici. Come la nostra ricerca aggiunge confronti, interpretazioni e significato ai dati, così librerie e *software* arricchiscono di funzioni l'applicazione finale. In entrambi i casi il risultato costituisce una entità molto più complessa dei dati di partenza e, per certi versi, più fragile, in quanto ogni apporto esterno introduce un possibile punto di errore. Un confronto sbagliato, una datazione incerta o una interpretazione poco supportata dai dati mettono in forte crisi l'intera struttura interpretativa. Allo stesso modo, una libreria problematica o un formato non supportato o un *software* obsoleto mettono a rischio la piattaforma di edizione.

Nello sviluppo di applicazioni web complesse – e quelle destinate alla pubblicazione di dati geografici lo sono sempre – è importante essere consapevoli del fatto che difficilmente i prodotti editoriali ad alto contenuto interattivo potranno avere una vita più lunga di qualche anno. La loro costruzione dipende da librerie e *framework* realizzati e mantenuti da terze parti, che ne semplificano il lavoro di realizzazione e permettono di aggiungere, con una certa facilità, funzionalità che ormai non è più pensabile sviluppare *ex novo*⁶³. Questa dipendenza è però un'arma a doppio taglio,

⁶³ Si pensi all'esempio già menzionato dell'importanza dello sviluppo della libreria GDAL/OGR per tutti i *software* GIS.

perché i prodotti delle terze parti seguono le logiche del mercato e la loro diffusione – oltre che dalla qualità dei prodotti – dipende anche da fattori incidentali, talvolta anche semplicemente dalla moda del momento. L'ampia diffusione è di norma in rapporto diretto con la qualità, perché un prodotto di largo impiego viene sviluppato in modo più attivo e riceve costanti aggiornamenti che ne migliorano la funzionalità e la sicurezza. Tuttavia, non vi è alcuna garanzia che un prodotto o una tecnologia oggi molto diffusa non venga soppiantata nell'arco di uno o due anni da una più nuova, meglio documentata o semplicemente più *trendy*. Queste considerazioni sono alla base della sostenibilità di ogni progetto di pubblicazione digitale, anche e soprattutto sul Web, che si tratti dell'applicazione di pubblicazione di un *database*, di un webGIS o di prodotti ibridi.

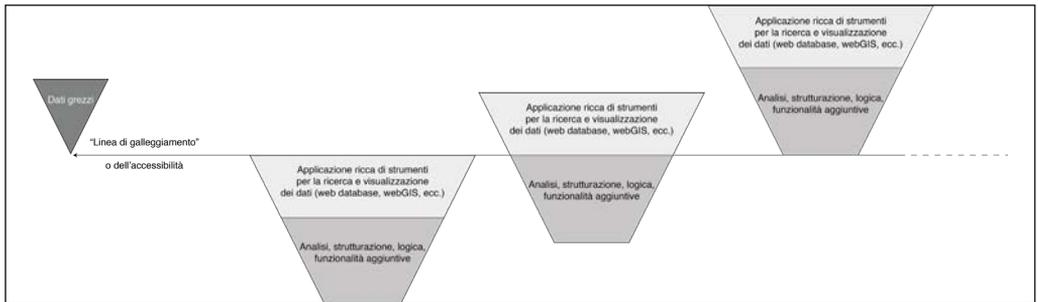
Non vi è soluzione, quindi, a questo problema? Se si cerca il *software* o il protocollo migliore per la pubblicazione sul Web di dati strutturati, tabellari o geografici o di altro genere, allora la risposta è negativa. La soluzione è nel capovolgere la domanda e al tempo stesso la nostra metodologia di ricerca, di produzione di conoscenza e della sua condivisione, ovvero pubblicazione.

Un primo passo potrebbe essere quello di separare il destino dei dati da quello delle piattaforme di pubblicazione, pubblicandoli in formati semplici, diffusi e aperti. Questo tipo di pubblicazione incrementa la durevolezza nel tempo degli archivi in quanto ne facilita il riversamento e trasformazione di formati e il rinnovo dei supporti. Questi ultimi dovrebbero essere per noi dei concetti famigliari, in quanto è grazie alla continuo trasformazione dei formati e supporti e grazie e cicliche operazioni di copiatura che la maggior parte dei testi antichi sono giunti fino a noi. Per i formati digitali, infatti, vale lo stesso principio, in quanto solo il periodico rinnovo dei supporti (da scheda perforata a banda magnetica, a disco ottico, a disco magnetico, a disco allo stato solido, ecc.) e il continuo rinnovo dei formati riescono a garantire una disponibilità più lunga nel tempo. Tra i vari formati quelli aperti garantiscono una più ampia accessibilità, in quanto non sono legati a particolari *software* o loro specifiche versioni⁶⁴.

In parallelo e a partire da questi archivi, possono essere costruite applicazioni e servizi web ricchissime di funzionalità, che rispondano a specifiche richieste di analisi e ricerca o che più semplicemente rendano fruibili dati complessi in maniera guidata e ragionata. La pubblicazione della metodologia seguita nella creazione dell'applicazione, infine, aggiunge un ulteriore elemento, che permette di scindere la logica dell'implementazione dai

⁶⁴ Sui risvolti etici relativi ai formati aperti, si veda il capitolo successivo.

mezzi tecnici impiegati. Questo passaggio assicura che l’invecchiamento di una tecnologia non metta in crisi tutta l’impalcatura e che dati e funzionalità possano essere “confezionate” in una nuova applicazione che fa uso di linguaggi e tecnologie diverse, più nuove, più efficienti o più performanti. Una metodologia di lavoro per compartimenti stagni permette, inoltre, di meglio integrare nel gruppo di lavoro esperti di settore, come gli informatici puri, certamente più qualificati nella gestione degli aspetti tecnici e tecnologici, ma di gran lunga meno titolati alla raccolta e strutturazione dei dati e nella metodologia della loro analisi e interrelazione.



Da una banca dati aperta (“linea di galleggiamento” posta sulla punta del triangolo) è possibile giungere a modelli applicativi diversi con vari gradi di apertura, anche paralleli e alternativi. L’invecchiamento o la “scomparsa” di uno di questi dovuto all’obsolescenza tecnologica non implica la scomparsa dei dati sottostanti. Si tratta di un modello di sviluppo molto più sostenibile sul lungo periodo del precedente.

Capitolo 5

Free Software, Open Source e Open Data. Copyright e licenze nell'era della condivisione

Uno degli aspetti “collaterali” più importanti dell’impatto che lo sviluppo tecnologico legato a Internet e al World Wide Web ha avuto sul nostro presente è sicuramente la ridefinizione del concetto di *copyright* per come l’abbiamo a lungo concepito, almeno a partire dall’inizio del XVIII secolo¹. È un processo che affonda le radici negli anni Ottanta del secolo scorso, ma che è letteralmente esploso con la diffusione della Rete, di Internet e quindi del Web².

La questione ha investito in maniera potente anche il mondo della ricerca archeologica, guadagnando negli anni una certa attenzione da parte della comunità scientifica e molto probabilmente diventerà un argomento sempre più pressante nel prossimo futuro³. Per questo motivo, e per il suo

¹ Il cd. “Statuto di Anna” del 1709, entrato in vigore l’anno successivo, è considerato la prima regolamentazione completa sul *copyright*, Laura Moscati, «Lo Statuto di Anna e le origini del copyright nell’Inghilterra del Settecento», in *Fides humanitas ius : studi in onore di Luigi Labruna*, a c. di Luigi Labruna, Cosimo Cascione, e Carla Masi Doria (Napoli: Editoriale scientifica, 2007). Per l’Italia si veda la sintesi di Maria Iolanda Palazzolo, «Per una storia del diritto d’autore in Italia: concetti, interessi, giurisprudenza», *Fabbrica del libro*, n. 2 (2011): 5–11.

² Per una sintesi esaustiva sull’argomento si veda Simone Aliprandi, *Copyleft & opencontent: l’altra faccia del copyright* (Lodi: PrimaOra, 2005).

³ Diverse sono ormai le iniziative volte a promuovere l’adozione di *software*, standard, formati e procedimenti aperti nel processo archeologico. Dal 2006 vengono, ad esempio, organizzati i convegni annuali di ArcheoFOSS (<http://www.archeofoss.org/>) i cui atti vengono regolarmente pubblicati: Roberto Bagnara e Giancarlo Macchi Jánica, a c. di, *Open source, free software e open format nei processi di ricerca archeologici: atti del 1. Workshop (Grosseto, 8 maggio 2006)* (Firenze: Centro editoriale toscano, 2007); Stefano Costa e Giovanni Luca Pesce, a c. di, *Open Source, Free Software e Open Format nei processi di ricerca archeologica Atti del II Workshop (Genova, 11 maggio 2007)*. (London: Ubiquity Press] u[, 2013); Luca Bezzi, Denis Francisci, Piergiovanna Grossi e Damiano Lotto, a c. di,

stretto collegamento con lo sviluppo della Rete, si è ritenuto opportuno dedicare alcune pagine di questo volume ai dati aperti, ai movimenti legati al *software* libero e all'*open source* e al loro impatto sull'universo dell'industria editoriale tradizionale. L'avvento di metodi di pubblicazione così rivoluzionari, resi grazie al Web, non poteva d'altro canto non mettere in seria discussione strumenti e modelli di business tradizionali, nati nell'era della stampa a caratteri mobili.

Movimento Free Software

Alla metà degli anni Ottanta Richard Stallman pubblicava il manifesto GNU⁴, con il quale si dava inizio a un movimento – divenuto presto sociale – basato su un'esigenza molto concreta: creare un sistema operativo del tutto simile a Unix, ma che fosse libero in tutte le sue parti e componenti. Chiunque doveva essere libero di usarlo, accedere al codice sorgente, modificarlo, migliorarlo e redistribuirlo, cosa non consentita da Unix, il cui sviluppo era centralizzato e controllato da licenze chiuse. A questo fine fu creata una fondazione chiamata Free Software Foundation (FSF)⁵, la cui

Open source, free software e open format nei processi di ricerca archeologica: atti del 3. Workshop (Padova, 8-9 maggio 2008) (Roma: Quasar, 2012); Paolo Cignoni, Augusto Palombini, e Sofia Pescarin, a c. di, *ArcheoFOSS: open source, free software e open format nei processi di ricerca archeologica. Atti del IV Workshop (Roma, 27-28 aprile 2009)* (Firenze: All'Insegna del Giglio, 2010); Giuliano De Felice e Maria Giuseppina Sibilano, a c. di, *ArcheoFOSS: open source, free software e open format nei processi di ricerca archeologica. Atti del V Workshop (Foggia, 5-6 maggio 2010)* (Bari: Edipuglia, 2011); Francesca Cantone, a c. di, *ArcheoFOSS: Open Source, Free Software e Open Format nei processi di ricerca archeologica. Atti del VI Workshop (Napoli, 9-10 giugno 2011)* (Pozzuoli: Naus Editoria, 2012); Mirella Serlorenzi, a c. di, *ArcheoFOSS free, libre and open source software e open format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)* (Firenze: All'Insegna del Giglio, 2013); Filippo Stanco e Giovanni Gallo, a c. di, *Proceedings of ArcheoFOSS: Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica: VIII Edizione, Catania 2013, 2016*; Patrizia Basso, Alessandra Caravale, e Piergiorgiana Grossi, a c. di, *ArcheoFOSS: free, libre and open source software e open format nei processi di ricerca archeologica: atti del IX Workshop, Verona, 19-20 giugno 2014* (Firenze: All'Insegna del Giglio, 2016); Stefano Campana, Roberto Scopigno, Gabriella Carpentiero e Marianna Cirillo, CAA2015. *Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, 1003–46. A questi eventi fissi sono stati aggiunte anche altre iniziative recenti, come per esempio Marco Arizza, Valeria Boi, Alessandra Caravale, Augusto Palombini e Alessandra Piergrossi, a c. di, *I dati archeologici. Accessibilità, proprietà, disseminazione* (Roma, CNR, 23 maggio 2017) (Firenze: All'Insegna del Giglio, 2018).

⁴ <https://www.gnu.org/gnu/manifesto.html>.

⁵ <https://www.fsf.org/>.

missione principale era (ed è) quella di promuovere e diffondere l'utilizzo del *software* libero.

Definire in maniera sintetica cosa sia il *software* libero non è semplice, soprattutto se si tiene conto delle tante sfumature possibili del concetto di libertà. Viene però fornita una lista molto sintetica, costituita da soli quattro punti, che illustrano le quattro “libertà” che un *software* deve garantire per essere definito libero⁶:

- libertà di eseguire il programma come si desidera, per qualsiasi scopo (libertà 0).
- libertà di studiare come funziona il programma e di modificarlo in modo da adattarlo alle proprie necessità (libertà 1). L'accesso al codice sorgente ne è un prerequisito.
- libertà di ridistribuire copie in modo da aiutare gli altri (libertà 2).
- libertà di migliorare il programma e di distribuirne pubblicamente i miglioramenti apportati (e le versioni modificate in genere), in modo tale che tutta la comunità ne tragga beneficio (libertà 3). L'accesso al codice sorgente è, anche in questo caso, un prerequisito.



Il logo del progetto GNU.

Strumento principale dello sviluppo del progetto GNU è una licenza d'uso, nota come GNU General Public License (o più semplicemente GNU GPL)⁷, che ricalca in maniera molto chiara le quattro libertà fondamentali. Inoltre viene sancito il meccanismo di *copyleft*⁸ o *share alike* (“condividi allo stesso modo”) come viene chiamato in alternativa. La norma prevede che tutti i *software* e programmi applicativi creati a partire da programmi GNU GPL (dette opere derivate) devono a loro volta essere distribuiti con la stessa licenza, o licenze equivalenti.

Un errore frequente è quello di confondere la licenza d'uso con il diritto di autore e di sovrapporre il concetto di “libero” con quello di “gratuito”: semplificando all'estremo, non è detto che un programma distribuito gratuitamente sia effettivamente un programma libero (perché, ad esempio, non viene dato accesso al codice sorgente)⁹, come non è detto che un program-

⁶ <https://www.gnu.org/philosophy/free-sw.it.html>.

⁷ La versione più recente di questa licenza è la 3 e il suo testo è disponibile all'indirizzo: <https://www.gnu.org/licenses/gpl-3.0.en.html>.

⁸ Un neologismo formato con un gioco di parole che sfrutta l'ambivalenza di significato della parola inglese “right”: “diritto” e “destra”.

⁹ Non sono considerati liberi, ad esempio, i *software* cosiddetti *freeware*, e neppure gli *shareware*.

ma libero non si possa vendere; anzi, a rigor di logica, negare la possibilità di vendita rende un prodotto non libero! Spesso infatti la distribuzione in forma non gratuita di *software* libero è una delle principali fonti di introito per il suo sviluppo e in quanto tale è consigliata dalla stessa FSF¹⁰.

Dal punto di vista della proprietà intellettuale, rilasciare un programma con una licenza libera, come la GNU GPL, non annulla il suo autore e i diritti morali che egli detiene sull'opera. Le licenze aperte prevedono esplicitamente il riconoscimento della paternità della stessa opera e obbligano anche gli autori delle successive modifiche a “firmare” i loro apporti, in modo da conservare traccia concreta dei contributi originali di ciascuno. Infine, non è infrequente che un autore rilasci una sua opera d'ingegno (es. un programma) con licenza GNU GPL e preveda la possibilità di accordi economici separati per usi commerciali che non sono contemplati da questo tipo di licenza.

Open Source Initiative

Logo del
progetto
Open Source
Initiative.



È chiaro da quanto detto sopra che, affinché un *software* venga riconosciuto come libero, deve dare sempre e comunque accesso al proprio codice sorgente: la *libertà 1* e la *3* indicano la libera disponibilità di questo codice come un prerequisito. Per questo motivo ogni *software* libero è anche un *software open source*, letteralmente a sorgente aperta. Negli anni si è costatata l'efficacia di questo paradigma di sviluppo, che allarga esponenzialmente la platea degli sviluppatori che possono intervenire a risolvere un problema oppure ad aggiungere o migliorare una specifica funzionalità. I progetti basati sul modello di sviluppo *open source*, infatti, tendono ad avere un ritmo di aggiornamento più veloce e quindi raggiungono, in minor tempo, migliori livelli di prestazioni e sicurezza rispetto ai *software* sviluppati da una sola persona o da *team* chiusi. Da questa constatazione parte la Open Source Initiative (OSI)¹¹, nata nel 1997 per promuovere la diffusione del paradigma *open source*, spogliandolo però dagli imperativi etici della Free Software Foundation. In

¹⁰ È molto frequente che privati o enti sponsorizzino la creazione di una nuova funzionalità o miglioria di un programma *open source*, per il proprio specifico interesse. La funzionalità o la miglioria diventano quindi disponibile per tutti gli altri utilizzatori e l'investimento fatto dal singolo ha un riscontro più ampio sull'intera comunità.

¹¹ <https://opensource.org/>.

breve, all'OSI non interessa garantire alcun principio etico sulla libertà degli utenti, ma solo il rapido ed efficace processo di sviluppo. Viene a cadere in quest'ottica l'imperativo del *copyleft* (o *share alike*) che si è visto sopra e vengono fortemente incentivate anche licenze di distribuzione che non pongono alcuna limitazione.

Ad esempio, una licenza come la Free Public License (nota anche come la BSD 0-clause)¹² non solo non impone l'applicazione delle stesse condizioni alle copie o ai prodotti derivati (*copyleft*), ma tralascia anche la menzione dell'autore (diritto morale), della stessa licenza. Di fatto, l'opera viene a trovarsi nel *pubblico dominio*, come le opere il cui termine di protezione legale è decorso e che possono essere utilizzate da chiunque senza dover corrispondere compensi o valersi di autorizzazioni preventive. L'OSI pubblica sul proprio sito¹³ una lista di licenze, accompagnate dal testo legale, che sono considerate a pieno titolo come *open source*.

Permission to use, copy, modify, and/or distribute this software for any purpose with or without fee is hereby granted.	La presente licenza dà diritto a usare, copiare, modificare e/o distribuire questo <i>software</i> con o senza pagamento.
---	---

The software is provided “as is” and the author disclaims all warranties with regard to this software including all implied warranties of merchantability and fitness. In no event shall the author be liable for any special, direct, indirect, or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of this software.	Questo <i>software</i> è fornito “così com'è” e l'autore disconosce qualsiasi garanzia espressa o implicita, che includa, ma che non sia limitata a garanzie implicite di commerciabilità e idoneità a scopo particolare. In nessun caso l'autore sarà ritenuto responsabile per qualsiasi danno diretto, indiretto o conseguente o per qualsiasi danno risultante dalla perdita di dati o profitti, come da contratto, responsabilità oggettiva, o torto derivante in qualsiasi modo dall'utilizzo di questo <i>software</i> .
---	---

Testo originale della più aperta licenza open source, la FPL (o BSD-0).

¹² <https://opensource.org/licenses/OBSD>.

¹³ <https://opensource.org/licenses/>.

Creative Commons

Questi movimenti, i concetti e le iniziative che li sorreggono non sono rimasti limitati al solo ambito della comunità degli *hacker*¹⁴, ma grazie alla diffusione del World Wide Web hanno investito anche, per estensione, il mondo delle opere dell'ingegno, non solo *software*, ma anche audio, video, grafica, testi di ogni sorta e, tra questi ultimi, quelli scientifici e accademici. Applicando quindi i concetti sviluppati negli anni Ottanta e Novanta del secolo scorso a tutte le risorse che il Web veicola, si è cominciato a studiare licenze utili anche per questi documenti, in particolare per quelle risorse "tradizionali", cartacee per lo più, la cui diffusione si vuole svincolare dalle restrizioni talvolta eccessive del *copyright*.



Logo di
Creative
Commons.

Una delle iniziative più importanti in questa direzione è senza dubbio costituito da Creative Commons (CC)¹⁵, un'organizzazione di diritto americano senza fini di lucro. Creative Commons aiuta i creatori di contenuti (o più genericamente gli "autori") a condividere in maniera legale le proprie opere, in modo che possano essere più facilmente divulgate e riutilizzate da parte di terzi, secondo una serie di condizioni facilmente combinabili tra loro a seconda delle proprie necessità.

In particolare, Creative Commons permette di comporre una licenza da applicare alle proprie opere originali ponendo due semplici domande:

- è possibile permettere la condivisione da parte di terzi di versioni adattate (derivate) della propria opera?
- è possibile permettere usi commerciali da parte di terzi della propria opera?

Alla prima domanda si può rispondere in maniera affermativa o negativa, mentre la seconda prevede una terza opzione, secondo la quale la condivisione è ammessa a condizione che l'opera derivata venga condivisa con la stessa licenza di quella originale (corrispondente al concetto di *copyleft* o *share alike* sviluppato in seno alla FSF). Ne consegue che le possibili combinazioni delle licenze d'uso sono le seguenti:

- Creative Commons Attribuzione (o nella forma abbreviata CC BY)
- Creative Commons Attribuzione – Condividi allo stesso modo (for-

¹⁴ Il termine viene qui usato nella sua accezione neutra di esperto informatico che ha nel mondo *open source*, non in quello negativo di "pirata" con cui viene più comunemente usato.

¹⁵ <https://creativecommons.org/>.

ma abbreviata CC BY-SA)

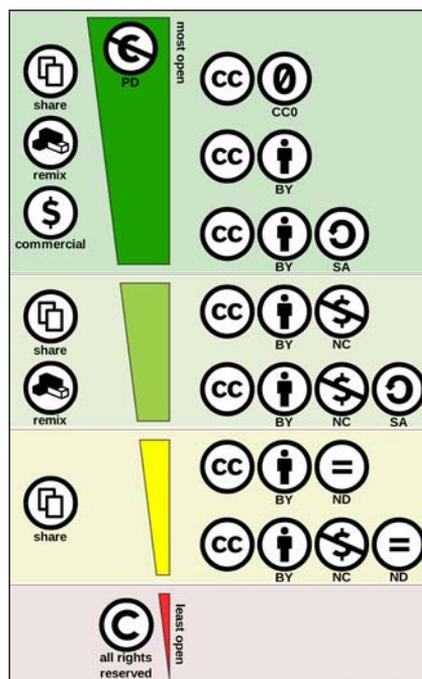
- Creative Commons Attribuzione – Non commerciale (forma abbreviata CC BY-NC)
- Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo (forma abbreviata CC BY-NC-SA)
- Creative Commons Attribuzione – Non opere derivate (forma abbreviata CC BY-ND)
- Creative Commons Attribuzione – Non commerciale – Non opere derivate (forma abbreviata CC BY-NC-ND)

Le combinazioni delle varie voci dell'elenco appena esposto sono in ordine di "apertura", dall'opzione più aperta a quella più ristretta. In verità esiste un'altra scelta ancora più aperta, chiamata CC0, che viene usata per attribuire un'opera al pubblico dominio, in modo analogo alla FPL o BSD-0, già analizzate. Le licenze Creative Commons possono riferirsi alla legislazione di un Paese specifico o essere riferite al diritto internazionale. Inoltre, i testi delle licenze sono stati aggiornati nel tempo, in seguito a sentenze particolari o a cambiamenti della legislazione vigente. Per questo motivo sono "versionate", ossia contrassegnate da un numero di versione che indica in maniera puntuale il testo normativo di riferimento. Attualmente la versione internazionale è la 4.0¹⁶ e quella italiana è la 3.0¹⁷. Infine, l'indicazione semplificata della licenza deve essere accompagnata dal testo legale completo (chiamato *Legal Code*) oppure da un collegamento a esso. L'indicazione del testo semplificato (*Common Deed*), che spiega in maniera chiara e in un linguaggio non giuridico e di facile comprensione cosa è o non è possibile fare con l'opera, è opzionale. Il già citato sito italiano o quello internazionale riportano i collegamenti ai testi più aggiornati. Il modo corretto di indicare la licenza, dunque, comprende tutte queste parti e una dicitura corretta potrebbe essere:

Attribuzione 3.0 Italia [Legal Code: <http://creativecommons.org/licenses/by/3.0/it/legalcode>].

¹⁶ <https://creativecommons.org/choose/>.

¹⁷ <https://creativecommons.it/Licenze>.



Le possibilità di combinazione delle clausole della CC, dalle alternative più restrittive (in basso) a quelle più aperte (in alto). Fonte: https://commons.wikimedia.org/wiki/File:Creative_commons_license_spectrum.svg.

Se si ha intenzione di divulgare la propria opera con una licenza Creative Commons, è molto importante riportare esplicitamente quest'indicazione, in quanto, in mancanza di segnalazioni, un'opera viene automaticamente a trovarsi sotto la forma più restrittiva possibile di *copyright*, ovvero "Tutti i diritti riservati". Per questo motivo Creative Commons si impegna costantemente a rendere gli autori sempre più coscienti di questi meccanismi fondamentali per la libera circolazione delle opere creative.

È interessante vedere come questi dispositivi legali vengano incontro alle nuove esigenze introdotte dal Web. Infatti, la platea dei produttori di opere di ingegno si è enormemente allargata nell'era del Web 2.0, tanto che è stato introdotto il concetto di *prosumer*, sincrasi delle parole inglesi *producer* (produttore) e *consumer* (consumatore), a indicare un utilizzo sempre più attivo e non solo passivo della Rete e dei "prodotti" che in essa trovano spazio. Un modello che è stato definito anche Produzione Paritaria, ovvero *Common-Based Peer Production*¹⁸, nato nell'ambito della tecnologia, ma ormai capillarmente diffuso nella società grazie al massiccio utilizzo di Internet.

Open Data

Alla luce di queste iniziative, che affondano le loro radici nella cultura *hacker* degli anni Ottanta, ma che sono tuttora vitali e attive e hanno aperto un dibattito molto fertile anche nel mondo accademico, è più facile comprendere cosa si intenda quando si parla di dati aperti o *open data*, per usare la forma inglese ben più diffusa. Il paradigma della sorgente aperta come presupposto indispensabile per uno sviluppo etico (punto di vista morale), efficace ed efficiente (punto di vista pratico), si è esteso ad altri ambiti – pressoché tutti – della nostra vita quotidiana, soprattutto in considerazione del fatto che viviamo ormai nell'era dell'informazione, dove la produzione e il consumo di dati è un'attività centrale e dove la raccolta e il possesso di quegli stessi dati è diventata un'attività economica di primaria importanza. *Open data*, quindi, è una definizione che indica uno o più *dataset* distribuiti con licenze aperte, che consentono il loro utilizzo e riutilizzo senza particolari restrizioni, eccetto forse quella di citare la fonte e/o l'eventuale *copyleft*, ovvero l'obbligo di condividere copie e opere derivate con le stesse modalità, anche a fini di lucro. Per venire incontro alle esigenze molto specifiche delle banche dati (diverse rispetto ai *software* e alle altre opere

¹⁸ Yochai Benkler e Helen Nissenbaum, «Commons-based peer production and virtue», *Journal of political philosophy* 14, n. 4 (2006): 394–419, <https://doi.org/10.1111/j.1467-9760.2006.00235.x>.

creative), sono stati elaborati strumenti legislativi dedicati, atti a incentivare la diffusione degli *open data*. In particolare, l'associazione Open Data Commons ha creato una serie di licenze specificatamente pensate per le banche dati e i loro contenuti, molto utili a chi fosse interessato alla loro distribuzione libera. Il modello seguito è quello delle licenze *copyleft* (principalmente GNU) per i *software* e quello delle Creative Commons per le opere di ingegno¹⁹.

Un forte incentivo alla diffusione degli *open data* è venuto dalla presidenza Barack Obama negli Stati Uniti, che ha molto insistito sull'idea di un'amministrazione aperta, il cui lavoro fosse trasparente e accessibile alla cittadinanza. Nel 2009 è stato reso disponibile un portale dedicato alla pubblicazione di dati aperti del governo americano²⁰, a cui ha fatto seguito nel 2012 il portale della Comunità Europea²¹. Anche l'Italia ha cominciato a pubblicare i propri dati governativi *online* a libero accesso a partire dalla fine del 2011²² e diversi portali regionali sono oggi disponibili per la pubblicazione dei dati di loro pertinenza.

Più difficoltosa è la via della pubblicazione libera dei dati archeologici, anche a causa di una consuetudine secondo la quale spetta ai funzionari del Ministero l'esclusiva del diritto di sfruttamento scientifico dei risultati di un'azione di tutela. La consuetudine è tuttora molto diffusa anche se non è supportata dal quadro legislativo²³, rispecchiando una visione piuttosto restrittiva e in aperto conflitto con politiche di apertura. Di fatto l'accesso alla documentazione archeologica in custodia presso gli enti pubblici è oggi possibile solo attraverso una richiesta esplicita, e i regolamenti non offrono chiarimenti sulla possibilità di sfruttamento economico di questi dati. Oggi, comunque, il Ministero per i Beni e le Attività Culturali ha

¹⁹ La Open Data Commons Open Database License (ODbL, <https://opendatacommons.org/licenses/odbl/index.html>) per i *database* e la Database Contents License (DbCL, <https://opendatacommons.org/licenses/dbcl/index.html>) per i contenuti delle banche dati sono due strumenti legislativi molto utili e con buon supporto legale nella Comunità Europea per la pubblicazione di *open data*. Un altro strumento particolarmente indicato per approfondire la questione è costituito da Open Data Handbook (<http://opendatahandbook.org/>), un sito volto a fare da guida e a suggerire casi di studio e buone pratiche per la pubblicazione di *open data*.

²⁰ <https://www.data.gov/>.

²¹ <http://data.europa.eu/euodp/>.

²² <https://www.dati.gov.it/>.

²³ Mario Trabucco, «Pubblico ma non pubblico: prospettive normative sulla proprietà intellettuale dei dati archeologici», in *ArcheoFOSS: Open Source, Free Software e Open Format nei Processi di Ricerca Archeologica: Atti del IV Workshop (Roma, 27-28 aprile 2009)*, a c. di Paolo Cignoni, Augusto Palombini, e Sofia Pescarini (Firenze: Edizioni all'Insegna del Giglio, 2010), 65–70.

attivato un punto di accesso aperto ai propri dati amministrativi²⁴, resi disponibili con licenza CC BY-SA-3.0. Vi si trovano pubblicati anche vari vocabolari (*thesauri*) per la descrizione dei reperti archeologici e diversi *dataset* regionali contenenti le schede dei beni culturali relativi ai reperti archeologici e non solo²⁵.

Dati aperti... e collegati

Si è visto fin qui quanto sia importante per la diffusione e il riutilizzo dei dati accompagnarli con licenze aperte e anche come la descrizione dei dati (ovvero la produzione di metadati) e loro pubblicazione in formati accessibili e standard sia presupposto fondamentale per il facile reperimento delle informazioni. Per questo motivo sono nati alcuni importanti progetti aggregatori anche nel campo dell'archeologia e dei beni culturali (alcuni dei quali descritti nell'appendice di questo volume) che richiedono un'adeguata strutturazione dei metadati da parte dei produttori ed editori di informazioni perché il loro *dataset* sia facilmente reperibile e usabile.

Questi accorgimenti permettono di condividere la propria ricerca e aprirla a una più ampia comunità, in uno sforzo collaborativo che è in atto da anni in altri ambiti del sapere e dell'industria. Esiste infine un ultimo aspetto che bisogna tenere in considerazione quando si parla di dati aperti, più tecnico ma non per questo meno importante. Si tratta della forma e del formato di condivisione, che rivestono un ruolo di primaria importanza nell'effettiva apertura e usabilità dei *dataset* da pubblicare.

Tim Berners-Lee ha affrontato questo tema nel 2010 con la proposta di un sistema cumulativo di valutazione attraverso il quale è possibile stabilire il grado di apertura e usabilità dei dati. Si tratta del sistema di valutazione a 5 stelle o *5-Star Open Data*²⁶.

★: rendere i dati liberamente disponibili sul Web. L'utilizzo di una licenza aperta è l'unico requisito per ottenere la prima stella. Il formato dei dati non è particolarmente importante.

★★: una stella aggiuntiva è assegnata se i dati vengono resi disponibili in

²⁴ <http://dati.beniculturali.it/>.

²⁵ Sul quadro legislativo europeo e italiano e sul difficile equilibrio da libero accesso e tutela del diritto di autore si veda anche Riccardo Pozzo, «Digital Humanities, Digital Cultural Heritage e l'istanza open», in *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico. Atti del IV Convegno di Studi SITAR (Roma, 14 ottobre 2015)*, a c. di Mirella Serlorenzi e Ilaria Iovine (Firenze: All'Insegna del Giglio, 2017), 47–52.

²⁶ <https://www.w3.org/DesignIssues/LinkedData.html>.

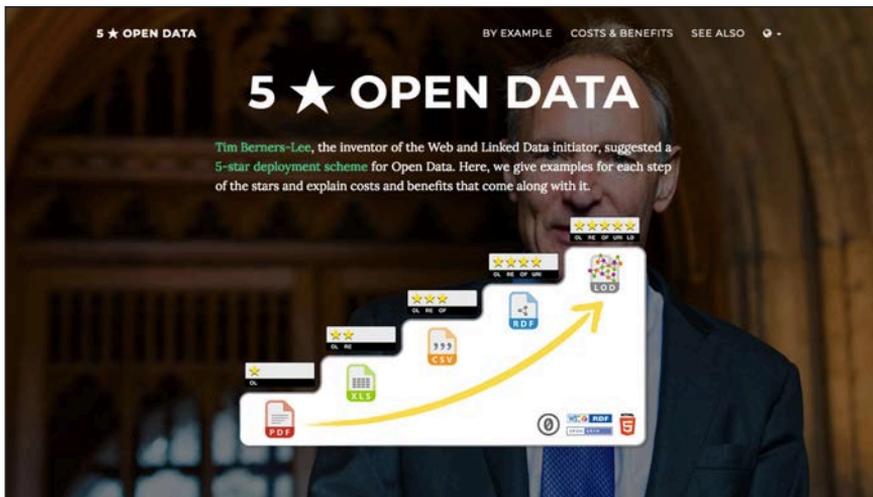
un formato strutturato, come potrebbe essere una tabella Microsoft Excel per la pubblicazione di dati tabellari.

★★★: un livello più alto è assegnato ai dati strutturati che vengono pubblicati facendo uso di formati aperti, non proprietari; una tabella in formato CSV, ad esempio, assicura una stella in più di rispetto alla tabella Excel, a parità di licenza di utilizzo.

★★★★: l'utilizzo di URI (Uniform Resource Identifiers) per riferirsi ai propri oggetti di studio dà accesso al livello successivo, perché gli URI permettono di assegnare nomi stabili e univoci agli oggetti eliminando l'ambiguità e consentendo a terzi di farvi riferimento e collegarli ai propri *dataset*. Lo standard RDF per rappresentare la conoscenza su un dominio è quindi il requisito principale per ottenere la quarta stella.

★★★★★: il massimo punteggio è assegnato se il *dataset* presenta collegamenti a risorse esterne, fatto che rende gli *open data* dei *linked open data*. Questo passaggio consente che il proprio *dataset* non rimanga una monade isolata nel Web, ma che grazie ai collegamenti possa entrare a far parte di una rete virtualmente globale.

Un esempio archeologico di questo schema potrebbe aiutare a far chiarezza. Si prenda l'esempio, abbastanza tipico, dei dati raccolti da una campagna archeologica, un progetto su piccola scala che lavora per di più in maniera tradizionale con documentazione cartacea. Al termine della campagna si sarà costituito un nutrito *dataset*, composto – nel caso più semplice – da elenchi e schede di unità stratigrafica e da elenchi ed eventuale schedatura dei reperti archeologici rinvenuti. Sicuramente si disporrà anche di



Pagina principale del sito <https://5stardata.info/>, nato per promuovere la conoscenza del sistema di classificazione a 5 stelle, proposto da Tim Berners-Lee.

documentazione grafica, digitale (fotografie, rilievi strumentali) e cartacea (rilievi manuali). Prima della fine della campagna, di questa documentazione viene fatta una copia digitale attraverso una semplice e rapidissima scansione, in formato PDF, che costituisce una copia di sicurezza e più facilmente trasportabile di quella cartacea.

Si ipotizzi che la direzione del progetto ottenga un accordo con le istituzioni centrali e locali preposte alla tutela e valorizzazione del patrimonio culturale che permetta la pubblicazione di questi *file* PDF sul Web, nella pagina ufficiale del progetto, in quella del Comune di riferimento o dell'ente che ha promosso le ricerche, su una pagina personale o su una *repository* specializzata (es. ADS o MOD²⁷). Se questa pubblicazione venisse collocata in un punto accessibile della Rete (non ad esempio su una pagina protetta da autenticazione) ed esplicitando una licenza aperta, come potrebbe essere CC BY, CC BY-SA o ODbL, l'archivio potrebbe fregiarsi della prima stella (★). I dati sarebbero infatti disponibili e raggiungibili da tutti grazie alla loro pubblicazione sul Web e la licenza ne permetterebbe il riutilizzo da parte di soggetti terzi, istituzionali (la Regione o il Comune che devono programmare interventi sul territorio) o privati (uno specialista che deve redigere un rapporto di archeologia preventiva o un gruppo archeologico che deve preparare una conferenza di sensibilizzazione della comunità locale sul patrimonio archeologico invisibile).

Va sottolineato che, dopo l'era cartacea e fino a qualche anno fa, questo era l'unico formato di documentazione accettato dalle allora Soprintendenze Archeologiche, sia per gli scavi di cui erano responsabili scientifici, che per quelli in concessione.

Tuttavia la scansione del materiale cartaceo in formato digitale (PDF) richiede uno sforzo importante di lettura, perché si tratta di acquisizioni *raster* e il testo contenuto in elenchi, tabelle e schede non è riconosciuto dalla macchina come testo e quindi non è, ad esempio, ricercabile. L'archivio diventa a due stelle (★★) se invece del PDF viene pubblicato un documento Microsoft Excel o un documento Numbers contenente la versione digitale degli elenchi e delle schede, i cui dati siano strutturati in righe e

²⁷ L'Archaeological Data Service, del quale si tratterà in maniera diffusa in appendice, richiede standard specifici per i *file* PDF, che devono essere nel formato PDF/A, specificatamente pensato per la lunga conservazione. Si tratta di un formato "auto-contenuto", ovvero che include nel documento tutti gli elementi per la sua visualizzazione (font, immagini, testi, colori), senza nessun tipo di crittografia e con restrizioni molto specifiche su determinati elementi tecnici quali *copyright* dei font, spazi colore, uso di interattività, e altri elementi che possono invalidare la visualizzazione del documento nel futuro.

colonne, in formato testuale, facilmente ordinabile, ricercabile e filtrabile. Sia PDF che MS Excel sono formati digitali, ma il primo contiene dati non strutturati mentre il secondo permette di incasellare ogni informazione in una griglia concettuale molto ben definita, grazie alla quale diventa possibile ordinare i siti per tipologia, cronologia, posizione, ecc. Lo stesso vale per il materiale topografico, che può essere vettorializzato e salvato in un qualche formato CAD (es. DWG) o GIS (es. ESRI personal geodatabase). Questo è a oggi il formato che le Soprintendenze Archeologia Belle Arti e Paesaggio richiedono per gli scavi da loro condotti e per quelli in concessione.

Il raggiungimento della terza stella (★★★) è un fatto meramente tecnico e può sembrare a prima vista un passaggio inutile o addirittura svantaggioso. Richiede l'adozione di un formato aperto (es. DXF e CSV) al posto di uno proprietario (es. DWG, XLSX o NUMBERS), processo che può essere visto come un'inutile complicazione. Molto spesso sono gli stessi *software*, durante il processo di esportazione, sconsigliano questo passaggio e segnalano la probabile incompatibilità o addirittura la perdita di informazione. Eppure si tratta di un passaggio di fondamentale importanza, se si vuole garantire ai propri dati una vita più lunga del programma con il quale sono stati creati o di una sua specifica versione. Un formato CSV può essere facilmente utilizzabile con qualsiasi *suite* da ufficio (Microsoft Office, OpenOffice, LibreOffice, iWorks, Google Documents, ecc.) e in casi limite anche da un programma basilare di gestione di testi semplici: i dati e le informazioni saranno sempre raggiungibili, qualsiasi sistema operativo si utilizzi, anche su architetture più vecchie e meno performanti. Non accade lo stesso per i formati proprietari, che a ogni aggiornamento rendono i *file* incompatibili con le versioni precedenti, solitamente più per motivi commerciali, che di miglioria tecnica.

Se si pubblica *online*, dunque, un foglio CSV contenente le informazioni sulle unità stratigrafiche e un *file* DXF o, meglio ancora, un *file* GIS correttamente georiferito, come ad esempio un ESRI *shapefile* o (preferibilmente) un GeoJSON, allora l'archivio può fregiarsi della terza stella. I dati che sono disponibili in un formato elettronico facilmente usabile e interrogabile sono liberi da dipendenze di natura commerciale legate a un certo specifico *software*.

L'idea che qui si sostiene non riguarda l'invito a utilizzare *software* libero al posto di programmi proprietari (invito che potrebbe comunque considerarsi valido in molti casi): la scelta dello strumento da usare viene lasciata all'utente, e talvolta accade che non ci siano alternative libere in grado di

competere con profitto con programmi professionali proprietari, e addirittura che sia necessario crearsi autonomamente programmi per eseguire compiti specifici. Il punto è piuttosto che la copia condivisa e/o quella archiviata per la lunga conservazione dovrebbero essere il più possibile libere da ogni condizionamento commerciale. Indipendentemente dalle scelte degli strumenti di lavoro, gli archivi di pubblicazione e quelli di sicurezza dovrebbero sempre cercare di raggiungere la terza stella, per garantirsi una maggiore durata e stabilità nel tempo.

Fino a oggi, questo punto non è stato percepito nella sua rilevanza da parte delle Soprintendenze che raccolgono e archiviano le informazioni archeologiche prodotte nei loro territori. I motivi sono diversi: mancanza di archivi digitali strutturati, carenza di formazione archivistica sui dati digitali e sulla conservazione a lungo termine, scarsa propensione (o più spesso decisa avversione) a pubblicare e rendere disponibili i *dataset* digitali, sono solo la punta dell'iceberg. La responsabilità non è solo dei singoli organi periferici, ma principalmente dello Stato, che continua a non dotarsi di una legislazione adeguata e manca di una visione politica proiettata sul medio e lungo termine, fattori che contribuiscono in modo determinante nell'innalzare una barriera contro l'accessibilità delle informazioni. Infrastrutture come l'ADS, come si vedrà nell'appendice, danno un peso rilevante ai formati forniti e penalizzano (economicamente) i progetti che depositano formati poco diffusi o legati a piattaforme o *software* particolari.

La quarta stella (★★★★) richiede, nella pubblicazione dei dati, l'adozione di standard semantici. È più complessa da implementare ed è difficile immaginare un prossimo futuro nel quale una qualche Soprintendenza richieda una documentazione espressa in triple RDF. Rimane compito degli archeologi definire sistemi univoci di identificazione delle proprie risorse (siti, unità stratigrafiche, reperti, iconografie, ecc) ed esplicitarli in maniera chiara e aperta, affinché collaboratori, colleghi o più in generale interlocutori possano farvi riferimento in maniera lineare e inequivocabile. Che si tratti di un certo sito della carta archeologica, di una qualche forma ceramica, di un modello iconografico o di un rituale funerario, la definizione e la pubblicazione di liste autoritative e vocabolari è un elemento di trasparenza metodologica che aiuta gli operatori esterni a interagire in modo più appropriato con i dati pubblicati e a fare confronti più puntuali. Esprimere queste liste e vocabolari usando il sistema degli URI è una procedura tecnica che può essere facilmente demandata ad algoritmi e processi automatici. La questione, quindi, rimane interamente metodologica, incentrata sulle scelte che l'archeologo compie nella strutturazione del proprio sistema in-

formativo o *database* e su come queste scelte vengano comunicate all'esterno. Una banca dati o sistema informativo archeologico ben strutturato – e che risponde a domande di ricerca metodologicamente corrette – può essere facilmente dotato, in un secondo momento, di un'interfaccia per la pubblicazione di dati semantici, in un qualsiasi formato RDF.

Per certi versi, per un archeologo è più semplice “guadagnare” la quinta stella di quanto lo sia ottenere la quarta. Guardare fuori dal proprio saggio di scavo o dalla propria cassetta di cocci alla ricerca di confronti e modelli ricorrenti fa infatti da sempre parte integrante e fondamentale del suo mestiere. I risultati della ricerca, le sintesi e le pubblicazioni sono piene di quei collegamenti che Tim Berners-Lee vorrebbe per gli *open data*. Eppure, gli archivi archeologici a cinque stelle (★★★★★) non sono poi così diffusi: perché? Oltre ai problemi elencati sopra in relazione alla quarta stella, vi è la questione che le banche dati non vengono ancora considerate come un obiettivo di ricerca, ma solo come uno strumento. Ciò rappresenta un grosso limite per una disciplina che basa la propria epistemologia sulla corretta raccolta, registrazione e interpretazione dei dati, mentre sarebbe invece fondamentale rendere questi ultimi interrogabili e fruibili, investendo risorse nella loro pubblicazione integrale e strutturata.

Non è chiaro perché la pubblicazione di una banca dati ben strutturata, aperta e magari collegata con rimandi e confronti ad altre similari, non possa essere considerata un risultato scientificamente pari a un articolo testuale descrittivo, che strutturalmente non è in grado di veicolare una simile densità informativa. Si potrebbe ribattere che è proprio la sintesi operata dal testo scritto che sposta l'ago della bilancia verso una maggiore considerazione di quest'ultimo da parte di colleghi e accademici. Neppure questo è però convincente: un tema cartografico ben elaborato, una carta di distribuzione di una data tipologia di anfora, una mappa di associazione tra un tipo di reperto e una tipologia di sepoltura all'interno di una necropoli, o ancora una carta delle potenzialità archeologiche, sono tutti elaborati che sintetizzano una serie talvolta impressionante di dati, ricerche e ipotesi. Nonostante ciò, questi elaborati difficilmente vengono considerati prodotti della ricerca se non accompagnati da lunghi testi, che nella maggior parte dei casi spiegano malamente quanto è già ben evidente nelle immagini, non tanto per colpa della cattiva prosa dell'autore, quanto per l'incapacità del *medium* testuale di veicolare altrettanto efficacemente informazioni di questo tipo²⁸.

²⁸ Un esempio in ambito contemporaneo di un prodotto editoriale scientifico maggiormente incentrato sulla rappresentazione dei dati in forma grafica, è co-

Il testo rimane ancora, ad oggi, l'unica forma accreditata dalla nostra cultura per la trasmissione del sapere, anche quando si lavora in ambiti dove una tabella, un grafico, una fotografia, un grafo o una mappa hanno un potenziale espressivo e informativo di gran lunga superiore. In quest'ottica culturale, una banca dati rimane un mero strumento, un mezzo per arrivare a una più nobile sintesi narrativa. Una complessa piattaforma GIS viene considerata una mera appendice (in certi periodi di gran moda), talvolta forzata a svolgere analisi statistiche avanzate e a costruire modelli matematici complessi sulla base di un campione di dati di partenza inadeguato, scarno, casuale e poco rappresentativo di realtà eterogenee e multiformi. La sua capacità di essere un contenitore strutturato di dati di diversa provenienza, origine e formato, di costituire un nodo di connessione e uno strumento di composizione di sintesi grafiche e infine una piattaforma di pubblicazione di dati ragionati, viene relegata a livello di tecnicismo da parte dell'archeologia più tradizionalista e a livello di diletterantismo da parte della *GIScience*, troppo affascinata dalle applicazioni matematiche.

Il Web corre il rischio di seguire la stessa sorte, e di essere usato solo per la creazione del sito "vetrina" di un progetto o per applicazioni avanzate di difficile accesso e comprensione sia per lo studente che per lo studioso, mentre potrebbe essere ben più utilmente impegnato per la veloce pubblicazione di dati strutturati, aperti e collegati. Un tale obiettivo sarebbe alla portata anche del lavoro di singoli studiosi o piccoli gruppi di ricerca, che vedano nei dati non solo il mezzo, ma anche il fine della ricerca stessa.

Oggi è possibile accedere facilmente a dati d'archivio o pubblicazioni per l'utilizzo dei quali non è più necessario chiedere difficili autorizzazioni²⁹. Queste risorse potrebbero così acquisire nuova vita nell'era digitale, non solo come copie digitali anastatiche a una stella bensì come ricche basi informative, possibilmente a cinque stelle, sulle quali costruire strumenti di ricerca e di sintesi.

Anche in Italia, per quanto riguarda la ricerca archeologica, si è ormai entrati nell'epoca dei dati nati digitali, sia perché gran parte della raccolta viene fatta attraverso strumenti di questo tipo, sia perché tutti i progetti prevedono un riversamento digitale, più o meno in tempo reale, dei pochi

stituito dalla rivista *online* dell'Università di Bologna "Cartografare il presente" (<https://cartografareilpresente.org/>). Qui studi complessi vengono pubblicati sotto forma di cartografie tematiche, infografiche e prodotti multimediali, corredati da un apparato testuale minimale.

²⁹ Il servizio offerto da <http://outofcopyright.eu/> aiuta a comprendere se un'opera è protetta o meno da *copyright*. Naturalmente, il servizio non è autoritativo e non sostituisce i correnti strumenti legali disponibili in ogni paese.

dati raccolti ancora su supporto cartaceo. Talvolta si producono copie di sicurezza (*backup*) su supporti cartacei, ma il più delle volte negli archivi delle Soprintendenze vengono depositati solo dati digitali. Nei prossimi dieci anni il Legislatore dovrà affrontare d'urgenza un problema prevedibile e previsto, ma che si è preferito ignorare: quello dell'obsolescenza tecnologica riguardante supporti, *software* e formati. La minaccia del Medioevo digitale (*Digital Dark Age*), come lo si è talvolta chiamato, potrebbe spingere a risolvere in maniera definitiva il problema del libero accesso dei dati archeologici. La moltiplicazione di copie liberamente disponibili, sotto gli occhi di tutti e non chiuse o protette da improbabili *caveau*, è l'unica via che abbiamo sempre avuto (e talvolta anche seguito) per la reale protezione dei nostri archivi, come ci insegna la lezione di Thomas Jefferson contenuta nella lettera (del 1791!) a E. Hazard³⁰:

To Ebenezer Hazard

Philadelphia Feb. 18. 1791.

Sir

*I return you the two volumes of records, with thanks for the opportunity of looking into them. They are curious monuments of the infancy of our country. I learn with great satisfaction that you are about committing to the press the valuable historical and state-papers you have been so long collecting. Time and accident are committing daily havoc on the originals deposited in our public offices. The late war has done the work of centuries in this business. **The lost cannot be recovered; but let us save what remains: not by vaults and locks which fence them from the public eye and use, in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.** This being the tendency of your undertaking be assured there is no one who wishes it a more complete success than Sir Your most obedient & most humble servt.,*

Th: Jefferson

³⁰ «From Thomas Jefferson to Ebenezer Hazard, 18 February 1791» Founders Online, National Archives, versione del 18 gennaio 2019, <https://founders.archives.gov/documents/Jefferson/01-19-02-0059>. [Fonte originale: *The Papers of Thomas Jefferson*, vol. 19, 24 January–31 March 1791, a c. di Julian P. Boyd. (Princeton: Princeton University Press, 1974), 287–289.]

Appendice

Casi di studio di applicazioni web e risorse *online* per la ricerca archeologica e antichistica

Questa appendice è dedicata all'esposizione di alcuni casi di studio, esempi di progetti e applicazioni fortemente orientate all'utilizzo delle tecnologie web, semantico e non, nell'ambito dell'archeologia e più in generale dell'antichistica. Particolare attenzione verrà riservata ai progetti che sviluppino maggiormente gli aspetti geospaziali e che adottano un approccio aperto e collaborativo nella pubblicazione dei dati. L'intenzione è di fornire delle esemplificazioni molto puntuali e concrete di quanto fin qui affrontato, ma anche e soprattutto di offrire strumenti e informazioni utili ai propri casi di studio e applicazioni, nella speranza di incentivare una più dinamica, condivisa e aperta pubblicazione dei dati della ricerca archeologica¹.

L'elenco non è certo rappresentativo della relazione tra tecnologie web e ricerca archeologica; manca, infatti, tutta la lunghissima serie di casi di piattaforme webGIS e banche dati *online* relative a singoli progetti di ricerca. La maggior parte di questi casi applicativi, infatti, rispecchia quanto già descritto a proposito delle principali tecnologie relative a *database* e GIS; si è pertanto ritenuto che produrne un elenco aggiungerebbe poco o nulla a quanto già detto.

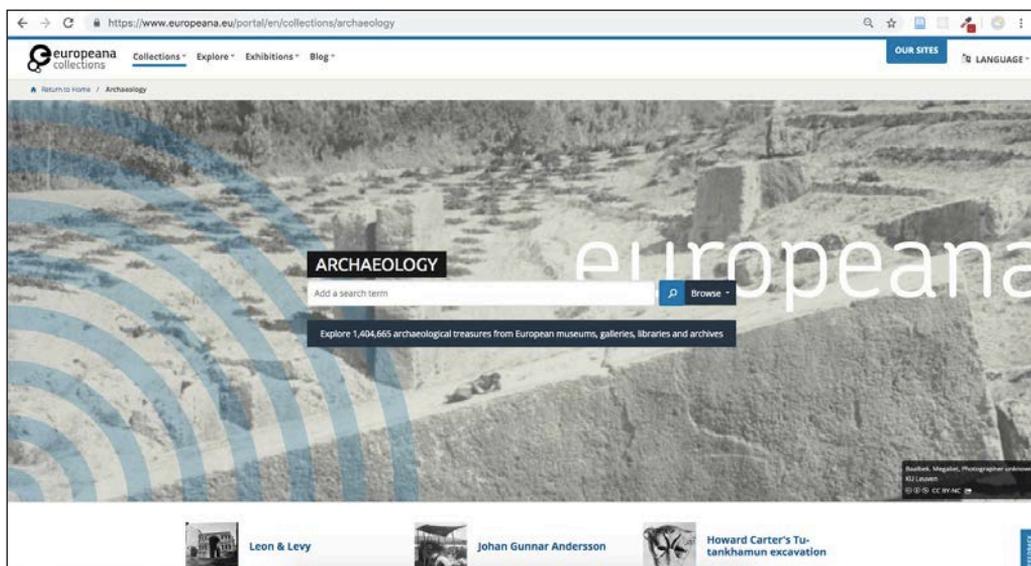
I casi seguenti sono stati selezionati in modo da offrire un ampio ventaglio di possibilità ed esigenze diverse, dalle sperimentazioni con il web semantico, all'elaborazione e condivisione di dati GIS, alla pubblicazione dinamica di temi di ricerca ben precisi (viabilità, collegamenti e trasporti, economia monetale, insediamenti, ecc.), alla creazione di archivi stabili e aperti dove pubblicare, conservare e condividere i dati, fino alle soluzioni che rispondono alle necessità di amministrazione e tutela del patrimonio culturale.

¹ Per una messa a punto delle risorse digitali riguardanti un ambito di studi molto vicino, quello degli studi romani, si veda Roger S. Bagnall e Sebastian Heath, «Roman Studies and Digital Resources», *Journal of Roman Studies* 108 (2018): 171–89.

Europeana

Europeana è un progetto della Comunità Europea che a oggi (aprile 2019) raccoglie più di 58 milioni di opere d'arte, manufatti, libri, film, musica, mappe, archivi, ecc., contenuti in musei, gallerie, archivi (GLAM) di 28 paesi europei², con una nutrita sezione archeologica. Europeana fornisce una struttura base unitaria per tutti i dati che vi vengono pubblicati, chiamata Europeana Data Model (EDM)³, sviluppata, con molte migliorie, su un precedente modello (chiamato ESE). Le singole istituzioni degli stati membri della CE che partecipano al progetto, creano ed esportano dei *set* di metadati per ciascun *dataset* digitale che intendono pubblicare, che sia un'immagine, un video, un *file* audio, una mappa o cartografia, un testo, la riproduzione di un manoscritto, ecc. Al di là di questi *set* dalla struttura condivisa, ciascuna istituzione segue gli schemi descrittivi più adeguati alle proprie necessità, senza restrizioni di sorta. Europeana raccoglie queste esportazioni e le rende disponibili alla ricerca centralizzata (federata) da parte degli utenti. Queste informazioni sono disponibili in formato semantico RDF, distinte tra metadati, descrizioni di oggetti e loro rappresentazioni, con riferimenti contestuali, secondo il modello dei dati aperti collegati, ovvero *Linked Open Data* (LOD). Molti importanti *database* na-

Vista della sezione archeologica di Europeana, <https://www.europeana.eu/portal/en/collections/archaeology>



² <https://www.europeana.eu/>.

³ <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>.

zionali dei vari paesi membri, soprattutto per quanto riguarda i dati librari, stanno implementando o hanno implementato la pubblicazione automatica di tutti i dati e metadati su Europeana, rendendolo di fatto un fondamentale strumento di ricerca bibliografica. Un altro aspetto fondante di questo progetto è l'autenticità, la robustezza e l'affidabilità dei dati contenuti, dal momento che questi vengono creati e aggiornati da istituzioni accreditate.

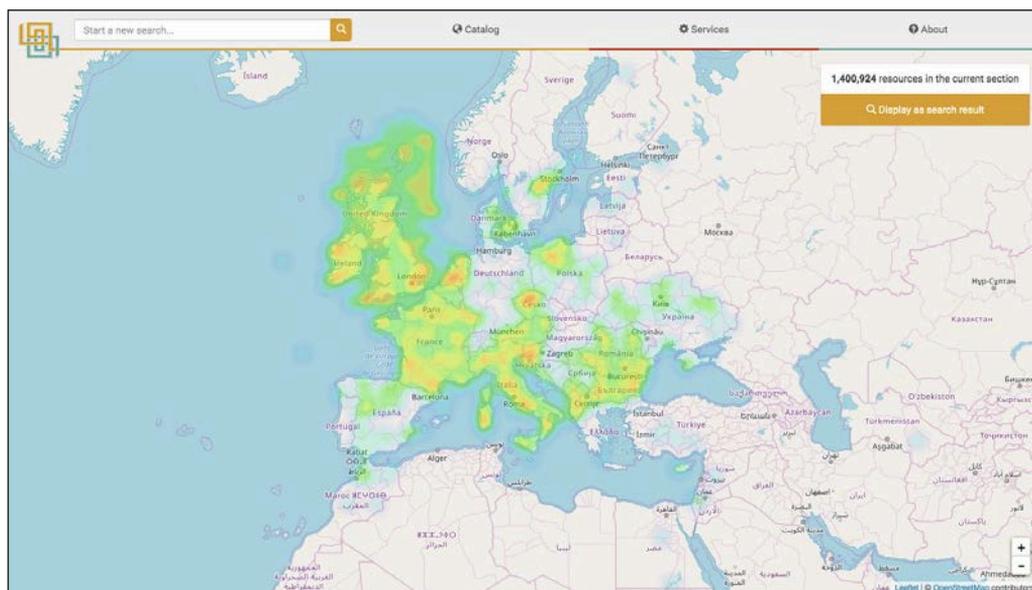
The screenshot shows the Europeana Collections interface for a specific record. At the top, the navigation bar includes 'europeana collections', 'Collezioni', 'Esplora', 'Esposizioni', and 'Blog'. A search bar is visible with the text 'Cerca per parole chiave qui'. The main content area features a large image of a relief sculpture titled 'Diogene e Alessandro Magno, rilievo, Arte Romana del I sec. a.C., Villa Albani, Roma'. Below the image, it states 'Created by Alinari, Fratelli'. There are buttons for 'CONDIVIDI' and 'SCARICA', and a 'POSSO USARLO?' section with a 'No' button and a Creative Commons license icon. A 'View more at: Alinari Archives' link is also present. The 'Persone' section lists 'Creatore: Alinari, Fratelli'. The 'Classificazioni' section includes 'Tipo: Fotografia' and a detailed 'Soggetto' list with terms like 'Rilievo', 'Reperto', 'Timpano', 'Elemento decorazione e parte d'edificio', 'Tempio', 'Non cristiana', 'Religiosa', 'Struttura', 'Antico', 'D'epoca', 'In costume', 'Abbigliamento', 'Moda e abbigliamento', 'Maschile', 'di personaggio famoso', 'Veduta di città', 'Veduta', 'Archeologia', 'Architettura', 'Ritratto', and 'Vedutismo'. The 'Extended Information' section is expanded to show 'Proprietà' with the format 'Formato: Black and white photo, N Gelatina ai sali di argento, N Silver salt gelatin'. On the right side, there are two vertical panels: 'Archeologia' with a definition of the field and a 'Subject of this object' link, and 'Ritratto' with a definition of a portrait and another 'Subject of this object' link. A 'FEEDBACK' button is located at the bottom right of the page.

Vista di una immagine di interesse archeologico, disponibile su Europeana e di provenienza dagli archivi Alinari, https://www.europeana.eu/portal/it/record/2024901/photography_Provide-dCHO_Alinari_Archives_ACA_F_027613_0000.html

Ariadne

A livello europeo, il progetto più importante che riguarda nello specifico i dati archeologici è senza dubbio ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe)⁴, una rete di 24 partner che cerca di rispondere all'esigenza di sviluppare infrastrutture per la gestione e l'integrazione di dati archeologici a livello comunitario⁵. Il progetto ha visto la realizzazione di un portale⁶ che integra informazioni tratte da moltissime fonti preesistenti (tra cui ADS e FastiOnline, per menzionare due iniziative descritte in questa appendice), ricercabili attraverso un'interfaccia geografica (una mappa web realizzata con Leaflet e OpenStreetMaps), per cronologia o per tipologia (viene data una *tag cloud* delle tipologie presenti). Viene inoltre offerta la possibilità di eseguire ricerche testuali e di usufruire di una serie di servizi, tra i quali anche la possibilità di caricare e condividere, attraverso una semplice interfaccia web, immagini complesse e modelli tridimensionali. La rete è sostenuta da tecnologie

Interfaccia geografica per la consultazione dei dataset presenti e indicizzati in ARIADNE. Fonte: <http://ariadne2.isti.cnr.it/index.php/browse/where>.



⁴ <https://ariadne-infrastructure.eu/>, Nicola Aloia, Ceri Binding, Sebastian Cuy, e Martin Doerr, «Enabling European Archaeological Research: The ARIADNE E-Infrastructure», *Internet Archaeology*, 43 (6 marzo 2017), <https://doi.org/10.111141/ia.43.11>.

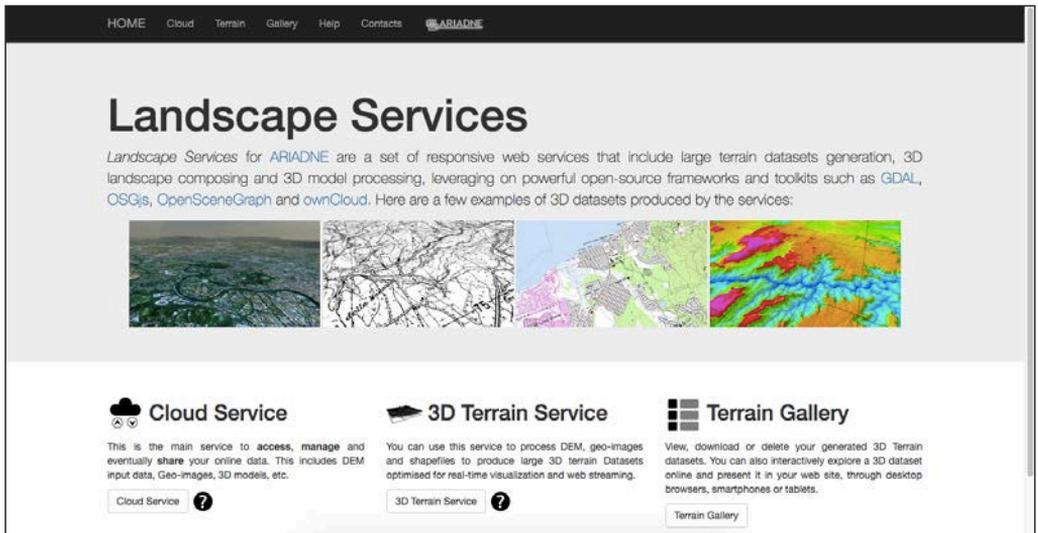
⁵ Julian D. Richards, «Digital Infrastructures for Archaeological Research: A European Perspective», *CSA Newsletter XXV*, n. 2 (2012), <http://csanet.org/newsletter/fall12/nlf1202.html>.

⁶ <http://portal.ariadne-infrastructure.eu/>.

di Web semantico basate sui Linked Open Data. La prosecuzione di questo progetto, chiamata ARIADNEplus (2019-2022), punterà a potenziare l'integrazione dei *dataset* già presenti e a una prima implementazione dell'EOSC (European Open Science Cloud) nell'ambito dell'archeologia e più in generale dei beni culturali.

È chiaro, però, che la mappa dell'Europa che fa da punto di accesso geografico ai dati di ARIADNE⁷ non rispecchia la reale situazione archeologica, bensì lo stato di implementazione del progetto a livello delle singole nazioni e il grado di disponibilità ad accesso libero degli archivi: questo fa sì che mentre il Regno Unito sia ben coperto e rappresentato (grazie alla miniera di informazioni costituita dall'ADS), l'Italia abbia una densità di informazione di molto inferiore (ben al di sotto della Slovenia, per intenderci) e che la Grecia, addirittura, non sia per nulla rappresentata. Come anche nel caso di un portale aggregatore ben più piccolo quale Peripleo (si veda più avanti), questi strumenti sono una chiara indicazione di quale sia la direzione da intraprendere in futuro, ma mettono anche in luce quanto sia ancora lunga la strada da percorrere.

Ariadne Landscape Service è un servizio web dedicato alle risorse geospaziali principalmente in tre dimensioni. Fonte: <http://landscape.ariadne-infrastructure.eu/>



⁷ <http://ariadne2.isti.cnr.it/index.php/browse/where>.

Nomisma.org

Nomisma.org è un progetto collaborativo che mira a fornire rappresentazioni stabili di concetti relativi alla numismatica, implementando il modello dei Linked Open Data⁸. I dati di base sono pubblicati in vari formati comuni (JSON-LD, TTL, RDF/XML). Vengono forniti URI stabili per ciascuna classe di dati, ovvero per ciascun concetto rappresentato, es. collezioni, zecche, autorità di emissione, contesti, conservazione, nominale, asse, tesoretti, materiali, monogrammi, ecc. È definita anche una serie molto ricca di proprietà che descrivono queste classi⁹. I dati di base sono implementati mediante un sistema collaborativo, coordinato e diretto da un comitato internazionale permanente. Classi e proprietà sono anche chiavi di ricerca molto elastiche ed efficaci, che permettono di creare un ampio ventaglio di ricerche e analisi statistiche.

Nomisma è un progetto estremamente aperto, non solo per quanto riguarda i contributi che raccoglie e pubblica: l'intero dataset è infatti disponibile per lo scaricamento gratuito¹⁰, e sempre gratuitamente è possibile ottenere il *framework*, ovvero il *software* di presentazione, visualizzazione e anali-

The screenshot shows the Nomisma.org interface for the term 'quinarius'. The left sidebar contains a list of properties and their values, such as 'skos:prefLabel' with values in multiple languages (Greek, English, French, Italian, Russian, Bulgarian) and 'skos:definition' with descriptions in English and French. The right side features an 'Export' section with options for 'Linked Data' (GitHub File, RDF/XML, RDF/TTL, JSON-LD) and 'Geographic Data' (KML, geoJSON (mints), geoJSON (hoards), geoJSON (finds)). Below this is a map of Europe with markers indicating the locations of mints (blue), hoards (red), and finds (green). A legend at the bottom of the map identifies these categories.

Un esempio di una ricerca per il nominale Quinarius (quinario): definizione (in diverse lingue), una breve descrizione e un inquadramento generale e URI corrispettivi in altre banche dati; a destra una mappa di distribuzione (fonte: <http://nomisma.org/id/quinarius>)

⁸ <http://nomisma.org/>.

⁹ La descrizione dell'ontologia in uso è disponibile all'indirizzo: <http://nomisma.org/ontology>.

¹⁰ I dati grezzi si possono scaricare, oltre che dal sito, anche dall'indirizzo <https://github.com/nomisma/data>. Tutti i dati sono rilasciati con licenza CC BY (si veda oltre).

si, che ognuno può personalizzare e utilizzare con *dataset* propri¹¹.

Il sistema accetta le *queries* in SPARQL, funzionalità non alla portata di tutti ma facilitata – dettaglio molto interessante – da una collezione di esempi di possibili interrogazioni che possono essere copiate, incollate ed eseguite.

Questo breve elenco offre una buona panoramica delle potenzialità di questo strumento¹²:

- Elenco degli imperatori romani (estratto dalle monete presenti)
- Elenco dei tesoretti rinvenuti in un raggio di x km da una data località (indicata con le coordinate geografiche)
- Elenco delle zecche localizzate in un raggio di x km da una data località (indicate con coordinate geografiche)
- Elenco dei pesi dei denari augustei
- Peso medio dei denari augustei
- Diametro medio di una data emissione (es. RIC Augustus 1A).

Type	Type Series	Example
<p>RRC 326/2</p> <p>Mint: Rome Denomination: Quinarus Date: 101 B.C. - 101 B.C.</p>	Roman Republican Coinage (Crawford 1974)	
<p>RRC 44/6</p> <p>Mint: Rome Denomination: Quinarus Date: 211 B.C. - 211 B.C.</p>	Roman Republican Coinage (Crawford 1974)	
<p>RRC 340/2e</p> <p>Mint: Rome Denomination: Quinarus Date: 90 B.C. - 90 B.C.</p>	Roman Republican Coinage (Crawford 1974)	
<p>RRC 85/1b</p> <p>Mint: Italy Denomination: Quinarus Date: 211 B.C. - 210 B.C.</p>	Roman Republican Coinage (Crawford 1974)	
<p>RRC 489/4</p> <p>Mint: Gallia Cisalpina Denomination: Quinarus Date: 43 B.C. - 42 B.C.</p>	Roman Republican Coinage (Crawford 1974)	

Nella stessa pagina dedicata al quinario è possibile consultare un elenco di esemplari, con una descrizione di base, immagine e collegamento al database di riferimento. Si noti che le immagini vengono caricate dal database di provenienza e non sono ospitate in Nomisma.org

¹¹ Il framework è disponibile all'indirizzo <https://github.com/nomisma/framework>. Purtroppo non viene fornita una documentazione adeguata per l'installazione e l'implementazione.

¹² L'elenco è disponibile all'indirizzo <http://nomisma.org/documentation/sparql>.

Kerameikos.org

Kerameikos.org, progetto dedicato alla codifica di categorie concettuali sulla produzione ceramica nel mondo greco¹³, utilizza la stessa strategia di Nomisma. Kerameikos ha sviluppato una propria ontologia specifica, tuttora in fase di evoluzione e arricchimento, composta da un numero limitato di classi (luogo di produzione, forma, stile, tecnica, tipologia [ing. *ware*])¹⁴ e da una serie di vocabolari che rendono possibili, anche in questo caso, ricerche piuttosto complesse. A differenza di Nomisma.org, il progetto è in fase di implementazione e dispone di un *dataset* ancora limitato, attualmente composto da esportazioni di dati aperti forniti da importanti musei internazionali (es. Harvard Art Museums, J. Paul Getty Museum, British Museum, The Fralin Museum, Ure Museum of Greek Archaeology). Il progetto non si propone di diventare un “contenitore” di dati, bensì uno strumento di raccordo di *database* indipendenti, che grazie al Web semantico è possibile unire e interrogare in maniera facilitata.

Anche in questo caso è possibile visualizzare i risultati delle ricerche su base geografica, eseguire analisi statistiche su quantità, distribuzione, tipologie ecc., e infine esportare tutti i dati per utilizzo proprio.

Kerameikos.org utilizza lo stesso *framework* di Nomisma.org, personalizzato per adattarsi alle proprie finalità specifiche.

Esempio di visualizzazione della categoria “ceramica con tecnica a figure nere” (URI: http://kerameikos.org/id/black_figure) su Kerameikos.org: In alto si offre un veloce inquadramento, con indicazione del nome in varie lingue; segue una visualizzazione geografica della distribuzione degli elementi di questa tipologia presenti nei database di riferimento, una serie di immagini dal database del British Museum (con collegamento alle rispettive schede) e infine una semplice analisi quantitativa sulla produzione delle singole forme attestate (nel database di riferimento) per la tipologia scelta (fonte: http://kerameikos.org/id/black_figure)

¹³ <http://kerameikos.org>, Ethan Gruber e Tyler Jo Smith, «Linked Open Greek Pottery», in CAA2014: 21st Century Archaeology: Concepts, Methods and Tools: Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, a c. di François Giligny (Oxford: Archaeopress, 2015).

¹⁴ <http://kerameikos.org/ontology>.

Ancient World Mapping Center

Schermata della mappa dinamica della Geografia di Strabone, a cura dell'AWMC. Le singole località sono interrogabili cliccando sulla mappa o anche attraverso una ricerca testuale integrata.

Se si parla di dati geografici relativi al mondo antico, una delle più importanti iniziative editoriali da menzionare è certamente l'*Atlante del mondo antico* di Richard J. A. Talbert e Roger S. Bagnall¹⁵. Il prezioso formato cartaceo, frutto del lavoro realizzato nell'arco di dodici anni da circa 200 tra studiosi e cartografi, ha però un limite intrinseco: l'opera ha cominciato ad invecchiare lo stesso giorno di pubblicazione e diventa sempre più obsoleta con il passare del tempo. Si tratta di un processo inevitabile per un ambito di ricerca così ampio come il Mediterraneo antico (e l'Asia centrale), sia dal punto di vista geografico che da quello cronologico, dove la produzione di nuova conoscenza e di nuovi dati procede con un ritmo quasi quotidiano. Per questo motivo il Classical Atlas Project, che ha prodotto il volume, è stato preso in carico dall' Ancient World Mapping Center (AWMC), un centro di ricerca interdisciplinare della University of North Carolina at Chapel Hill (USA)¹⁶. Il *dataset* originario del *Barrington Atlas*, ovvero la base cartografica, viene continuamente migliorata, estesa e arricchita, sia per quanto riguarda la copertura fisica (rappresentazione del territorio fisico in diverse fasi dell'antichità: idrografia interna e esterna, linee di costa, modelli di terreno in tre dimensioni, ecc.), sia per quanto riguarda i temi culturali (centri abitati, canali, centuriazione, rete viaria,



¹⁵ Richard J. A. Talbert e Roger S. Bagnall, *Barrington atlas of the Greek and Roman world* (Princeton, N.J. : Princeton University Press, 2000).

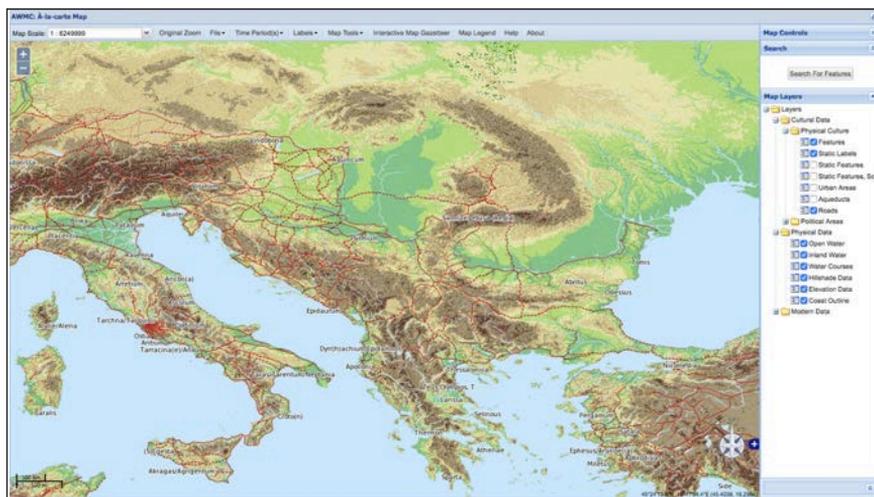
¹⁶ <http://awmc.unc.edu/>.

acquedotti e altre infrastrutture, ecc.). Il progetto produce e distribuisce liberamente varie elaborazioni cartografiche, sia statiche (immagini *raster* dei vari temi relativi al mondo antico) sia dinamiche (ad esempio la mappa della *Geografia* di Strabone, <http://awmc.unc.edu/awmc/applications/strabo/>). Tutte le mappe disponibili, raggruppate ed indicizzate per area geografica¹⁷ vengono distribuite con licenze aperte che permettono la loro facile riutilizzo e rielaborazione.

Il centro inoltre sviluppa e mantiene un vero e proprio webGIS, un'applicazione web (Rich Web Application) chiamata *Antiquity À-la-carte*, proprio per rimarcare l'accesso aperto, disponibile all'indirizzo: <http://awmc.unc.edu/awmc/applications/alacarte/>. Si tratta di un visualizzatore piuttosto potente, che permette di comporre “al volo” mappe complesse, a partire da tutti i dati geografici a disposizione del centro di ricerca.

L'applicazione dispone di un motore di ricerca per dati geografici e testuali e permette l'esportazione delle mappe create nonché la loro condivisione tramite URL personalizzati. Creata con GeoExt, *Antiquity À-la-carte* fa ampio uso dei protocolli WMS, (W)TMS e XYZ per la pubblicazione delle mappe.

Infine, tutti i dati geografici di partenza sono liberamente scaricabili in formati GIS aperti (CSV e shapefile, principalmente) e rilasciati con una licenza d'uso aperta¹⁸, che rende questo progetto una fonte di dati geografici specializzati (relativi esclusivamente al mondo antico) molto importante e preziosa. Di grandissimo interesse è anche una serie di dati *raster* in



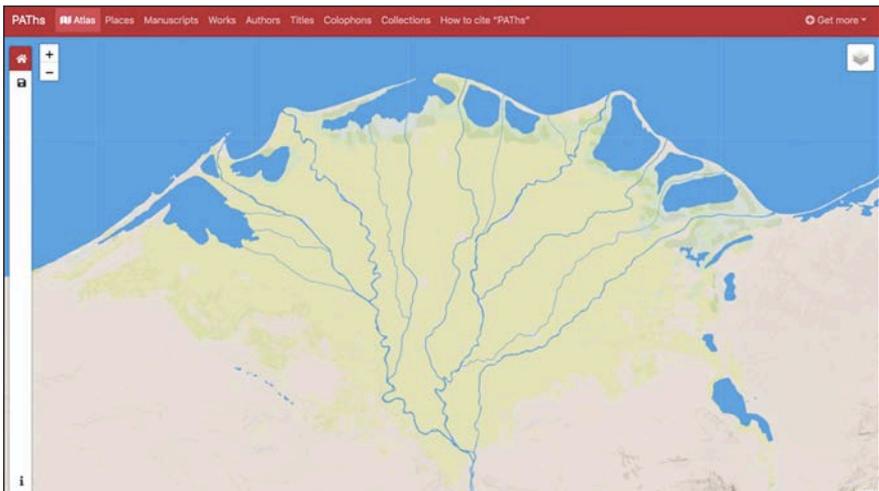
Schermata di *Antiquity à-la-carte*, a cura dell'AWMC, un visualizzatore webGIS che dà accesso ai dataset cartografici del centro.

¹⁷ <http://awmc.unc.edu/wordpress/free-maps/>.

¹⁸ Open Database License Any, <http://opendatacommons.org/licenses/odbl/1.0/>.

formato MBTiles che rappresentano fonti importanti per la ricostruzione della geografia antica, di facilissimo utilizzo e grande qualità grafica. Questi dati, disponibili anche come servizio web (protocollo XYZ mediante Mapbox) comprendono temi come la rete idrologica in vari periodi storici (età arcaica, classica, ellenistica, romana, ecc.), la copertura della rete stradale di età romana, ecc.

La pubblicazione, inclusa nel progetto, di dettagliati metadati per ciascun tema e dato pubblicato è uno strumento imprescindibile per capire e ricostruire la fonte di ciascuna elaborazione e i metodi che si sono seguiti per la digitalizzazione ed eventuale vettorializzazione. Si tratta di elementi determinanti quando si tratta di produzione cartografica e ancora di più quando questa produzione riguarda un territorio che sostanzialmente non c'è più ed è riproducibile solo attraverso difficili e lunghe ricerche. L'operato dell'AWMC è un punto di partenza fondamentale per ogni contributo originale, che viene vivamente stimolato da parte del centro: ogni ricercatore può scaricare i dati, migliorarli e arricchirli con i risultati della propria ricerca ed eventualmente restituirli, per tenere aggiornata la fonte principale. Un approccio collaborativo di grande efficacia, imprescindibile per un ambito cronologico e culturale così ampio.

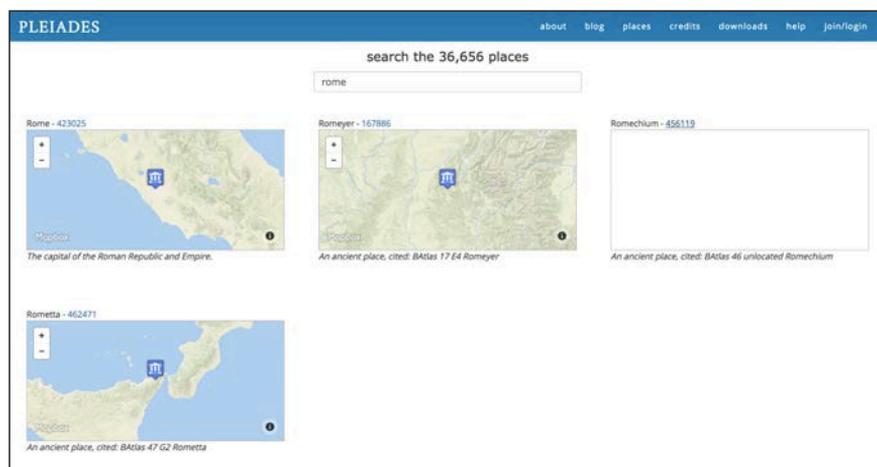


Dettaglio della morfologia del delta del Nilo in età romana. Servizio web a cura dell'AWMC, integrato nel PATHs Archaeological Atlas of Coptic Literature (<https://atlas.paths-erc.eu/map>) elaborato dall'autore.

Pleiades

Sempre in tema di condivisione, va sottolineato come i lavori dell'AWMC si avvalgono del supporto di un altro progetto importante, internazionale e collaborativo, Pleiades, finalizzato alla mappatura di luoghi e posizioni geografiche relative al mondo antico¹⁹. Pleiades presenta oggi una copertura molto estesa per il mondo greco e romano ed è in rapida espansione per quanto riguarda il Vicino Oriente antico, la geografia bizantina, celtica e quella del primo Medioevo²⁰. Si tratta di un progetto di forte interesse per la ricchezza del *database* in costante aggiornamento, sia in termini di quantità dei dati presenti (più di 36.000 luoghi, più di 32.000 toponimi e quasi 40.000 localizzazioni) che in termini di qualità della definizione della toponomastica e dello sforzo di fornire coordinate geografiche. Si tratta di un elenco geografico (*gazetteer*) che offre per ogni luogo una serie importante di informazioni, accessibili (*open*) e connesse (*linked*). Come fa Geonames²¹ per la geografia contemporanea, Pleiades espone identificativi stabili che riuniscono sotto un unico numero (URI) varie forme toponomastiche che sono mutate nella lingua, nella grafia e nel tempo.

L'URI <https://pleiades.stoa.org/places/423025>, ad esempio, corrisponde a Roma, “Ρώμη” in greco, “Roma” in latino (e italiano) e “Rome” in inglese. Per le località con posizione nota, vengono fornite anche le coordinate geografiche, e per ogni elemento è presente la bibliografia di base,



Ricerca in Pleiades per la parola “rome”. Di Romechium (ID 456119) non sono note coordinate geografiche.

¹⁹ <https://pleiades.stoa.org>.

²⁰ Tom Elliott e Sean Gillies, «Digital Geography and Classics», *Digital Humanities Quarterly* 003, n. 1 (26 febbraio 2009), <http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html>.

²¹ <https://www.geonames.org/>.



Pagina di Pleiades dedicata a Roma, ID: 423025, <https://pleiades.stoa.org/pleiades/423025>.

relativa al posizionamento dei siti e all'attestazione dei toponimi. Per ogni luogo è anche disponibile una lista di altri luoghi ai quali il *record* "Roma" è collegato (*makes connections with*), come il fiume Tevere, la via Appia e le altre vie consolari e anche una lista di luoghi che sono connessi a essa (*receives connections from*), come la tribù Claudia, l'arco di Costantino, la porta Aurelia, ecc. Roma, per esempio, presenta più di 200 connessioni in entrata. Inoltre, vengono fornite le occorrenze equivalenti in altre banche dati, più o meno specializzate, anche non necessariamente da riferire al mondo antico, come il *Digital Atlas of Roman and Medieval Civilizations* (DARMC, vedi sotto), il *Digital Atlas of Roman Empire* (DARE, vedi sotto), VIAF, Geonames, Wikidata e molti altri.

Si tratta di una risorsa strategica, proprio perché collaborativa e quindi facilmente implementabile anche da parte di gruppi di ricerca che si focalizzano su una data area geografica o su un periodo cronologico particolare. È possibile contribuire su singole voci e vengono forniti diversi strumenti di controllo della qualità delle informazioni, seguendo il modello di implementazione collettiva (*crowdsourcing*) caratteristico del progetto Wikipedia. Il portale di Pleiades, infine, permette di espandere la rete dei dati collegati anche utilizzando strumenti sviluppati non specificamente per lo studio delle antichità, come i *social network*. È infatti possibile illustrare un luogo anche con le immagini digitali prodotte dagli utenti, se in esse sono presenti tag speciali che riportano l'identificativo del luogo. Sempre nella pagina dedicata a Roma sono visualizzabili anche circa 300 immagini caricate dagli utenti in Flickr e contrassegnate con il tag `pleiades:depicts=423025` per le immagini della città di Roma e con il tag `pleiades:-`

studio delle antichità, come i *social network*. È infatti possibile illustrare un luogo anche con le immagini digitali prodotte dagli utenti, se in esse sono presenti tag speciali che riportano l'identificativo del luogo. Sempre nella pagina dedicata a Roma sono visualizzabili anche circa 300 immagini caricate dagli utenti in Flickr e contrassegnate con il tag `pleiades:depicts=423025` per le immagini della città di Roma e con il tag `pleiades:-`

findspot=423025 per gli oggetti rinvenuti a Roma (ma non necessariamente oggi collocati o esposti a Roma).

Tutti i dati del progetto sono liberamente scaricabili nei diversi formati aperti che si sono già incontrati: JavaScript Object Notation (JSON), Comma Separated Values (CSV), Keyhole Markup Language (KML), Resource Description Framework (RDF). Le esportazioni (*dump*) dei dati, adatte a diversi utilizzi – per una piattaforma GIS (KML), per un'applicazione web o simili (JSON) o per gli strumenti del Web semantico (RDF) – sono rese pubbliche a cadenza giornaliera in una pagina dedicata²².

Un ultimo aspetto che rende Pleiades un nodo importante nella rete dedicata alla geografia del mondo antico è la definizione di alcuni vocabolari usati internamente al progetto per la descrizione dei dati, e che, essendo pubblici e ben documentati, possono essere adottati anche da altri progetti. Si tratta di vocabolari molto generici, l'unica via possibile per un progetto di tale estensione geografica e cronologica, che coprono vari ambiti, come la certezza di associazione (*association certainty*), il livello di sicurezza nell'attribuzione cronologica (*level of confidence in temporal attestation*), le lingue e le scritture (*Pleiades Vocabulary Language and Script*), i periodi temporali (*Pleiades Vocabulary Time Periods*), le tipologie di luogo (*Pleiades Vocabulary Feature (or Place) Categories*), le categorie di luoghi (*Pleiades Vocabulary Location Categories*), e molti altri²³. Questi vocabolari sono sempre accompagnati da metadati che ne dichiarano i creatori, i curatori, le collaborazioni e la data dell'ultima modifica: trattandosi di un progetto che per definizione è in continuo mutamento, la dimensione temporale assume infatti un rilievo importante.

Pleiades espone anche una ben documentata API²⁴ che può essere usata come servizio Web per estrarre dalla banca dati informazioni in maniera programmata e automatica e creare in questo modo applicazioni geografiche sempre aggiornate, senza dover scaricare e mantenere la banca dati principale.

Pleiades non distribuisce cartografia di propria creazione e utilizza nel proprio sito le basi cartografiche relative alla topografia antica create dall'AWMC e distribuite attraverso Mapbox.

²² <https://pleiades.stoa.org/downloads>.

²³ <https://pleiades.stoa.org/vocabularies>.

²⁴ Application Programming Interface, <http://api.pleiades.stoa.org/>.

Digital Atlas of Roman Empire

Il DARE (*Digital Atlas of Roman Empire*)²⁵, nato per opera di Johan Åhlfeldt e ospitato all'Università di Lund (Svezia), è un'iniziativa sorta nel 2012 in collaborazione con il progetto Pelagios (vedi sotto) e poi sviluppata su piattaforma GIS adottando il *Barrington Atlas* come tema cartografico di base e unendo gli sforzi di diversi progetti²⁶. DARE, in particolare, fa largo uso di cartografia nazionale, immagini satellitari e contributi scientifici pubblicati, per una più precisa e puntuale descrizione e posizionamento dei siti di età romana. Attualmente (aprile 2019), il progetto è giunto a comprendere oltre 9.000 siti in più rispetto al *Barrington Atlas*. Anche il DARE assegna identificativi univoci per ogni sito (ad esempio, a Roma corrisponde l'identificativo 1438 e il corrispettivo URI: <http://dare.ht.lu.se/places/1438>), i quali vengono "legati" e riferiti con rimando speculare a quelli di Pleiades. Anche altri progetti di atlanti *online*, di mappatura di forme architettoniche, di dati testuali, ecc., vengono riferiti e collegati a ogni singola scheda quando si verificano delle occorrenze.

A titolo meramente esemplificativo, dal *record* "Roma" è possibile connettersi a un progetto che si occupa del censimento *online* degli acquedotti

Schermata (parziale) della scheda dedicata a Roma (ID: 1438) in DARE, con alcuni dei collegamenti possibili verso altri progetti e database online.

The screenshot displays the DARE interface for the site 'Roma'. On the left is a map of the Roman Empire with Rome highlighted. On the right is an information panel with the following content:

Digital Atlas of the Roman Empire
 version 12, Wikidata:Q108112, Roma

Roma

Information about the ancient place known as Roma, located at Roma

id	1438
URI	dare.ht.lu.se/places/1438
spatial coverage	Roma
LAD 2 Code	018091
LAD 2	Comune Roma
NUTS-1	Province Roma
country ISO 3166	IT
historical names	Roma
place type	city → display all
temporal coverage	
start date	-750
end date	840

Additional metadata and related external resources:

TWPlace	1203	whc	91
archtheatrum.de	1372	whidata	Q174769
brl.it-newspaly	rome-17023930	wikipedia	Q270
genavies	3184070	wikipedia	Ancient_Rome
name.de	Rom	wikipedia	Rome
name.en	Rome		
name.es	Roma		
name.fr	Rome		
name.gr	Ρώμη		
name.it	Roma		
name.pl	Rzym		
name.pt	Roma		
name.ru	Рим		
name.sv	Rom		
name.uk	Rome		
name.us	Rome		
name.zh	羅馬		
pleiades	433025		
romaq.org	429		
romaq.org	430		
romaq.org	431		
romaq.org	432		
romaq.org	433		
romaq.org	434		
romaq.org	435		

external maps:
 See further:
[OpenStreetMap | Google Maps](#)
[Panoramio Photos](#)

external resources:
 Pelagios 2
 Pelagios 3 (experimental)
 Geometry, additional built-in and external maps
 WGS 84 position: 42° 40' 40" N, 12° 48' 40" E
 Center: Zoom

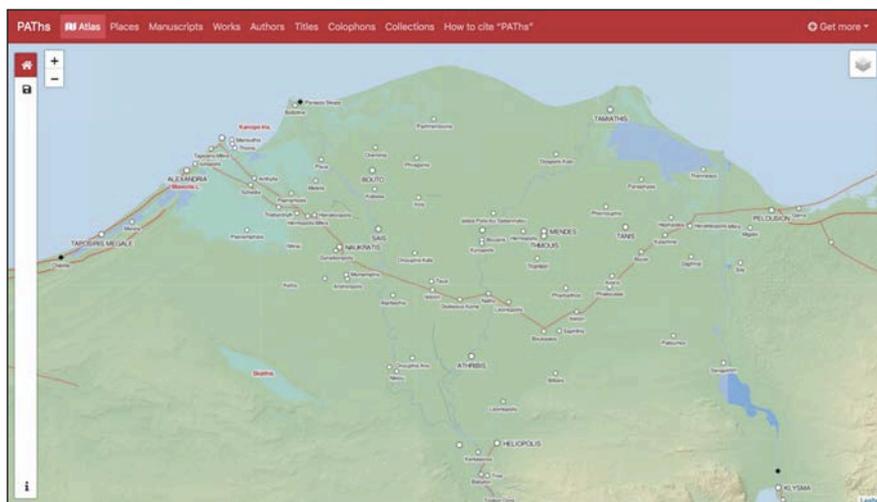
²⁵ <http://dare.ht.lu.se/>.

²⁶ Tra i quali il DARMC. Per un approfondimento su questo progetto si veda la scheda dedicata in questa appendice.

antichi²⁷, a un altro che mappa i teatri²⁸, gli anfiteatri²⁹, fino al collegamento con una ricerca automatica della parola “roma” in quel fondamentale *database* testuale relativo alle antichità classiche che è Perseus Digital Library³⁰, nel più recente ToposText³¹ o nei portali istituzionali come il World Heritage Convention dell’UNESCO³² e molto altro ancora.

Il progetto offre principalmente due servizi diversi, il primo è costituito da una mappa di base, che può essere utilizzata liberamente grazie alla licenza aperta con la quale è distribuita (Creative Commons BY-SA), sotto forma di *web tiles*. Disponibile a una scala massima di 1:250.000 è facilmente integrabile in applicazioni Web attraverso il protocollo XYZ, in modo simile a quanto è possibile fare con la cartografia generica di Google Maps, OpenStreetMaps e simili³³.

Se quindi l’AMWC fornisce dati strutturati GIS di alta qualità pronti all’uso, in formato *raster* e vettoriale, per applicazioni *desktop*, DARE interviene a colmare il vuoto del segmento Web. Usando i suoi *map tile* è semplicissimo creare applicazioni webGIS dove sia possibile rappresentare come



Dettaglio del Delta del Nilo. Servizio web a cura del DARE, integrato nel PATHs Archaeological Atlas of Coptic Literature (<https://atlas.paths-erc.eu/map/>).

²⁷ Atlas Project of Roman Aqueducts, <https://www.romaq.org/>.

²⁸ <http://www.theatrum.de/>.

²⁹ <http://amphi-theatrum.de/>.

³⁰ <http://www.perseus.tufts.edu/>.

³¹ <https://topostext.org/>, si veda la scheda dedicata.

³² <http://whc.unesco.org/>.

³³ Maggiori informazioni su questo servizio, con esempi pratici di implementazione si possono trovare all’indirizzo <http://commons.pelagios.org/2012/09/a-digital-map-of-the-roman-empire/>.

mappa di base non solo la cartografia attuale (o le foto satellitari o un'altra copertura del presente), ma anche una situazione geomorfologica e culturale (centri abitati e altri interventi umani sul territorio) più vicina all'antichità classica, principalmente quella greco-romana. La mappa presenta un'articolata legenda che tiene conto della complessità delle informazioni rappresentate e che offre una chiave di lettura molto efficiente.

Il secondo servizio offerto è una API web³⁴, attraverso la quale è possibile interrogare l'intero *database* geografico in base a diversi parametri di ricerca e ricevere la risposta in formato GeoJSON, un formato geografico aperto molto utilizzato sia in applicazioni GIS *desktop* che webGIS (dove è ormai uno standard *de facto* per i dati vettoriali). Il servizio permette di eseguire interrogazioni geospaziali (ad esempio tutti i siti compresi in una certa cornice di coordinate date, oppure tutti i siti vicini a una specifica coordinata e/o entro un certo raggio) e interrogazioni testuali (tutti i siti di una particolare tipologia, ad esempio le ville, o i siti che contengono una determinata stringa/parola nella loro descrizione, ecc., ma anche una combinazione delle due). Si tratta di uno strumento di estrema utilità in caso si vogliano creare prodotti web derivati (*mashup*), come possono essere atlanti tematici o rappresentazioni statiche e dinamiche di un fenomeno geografico o culturale.

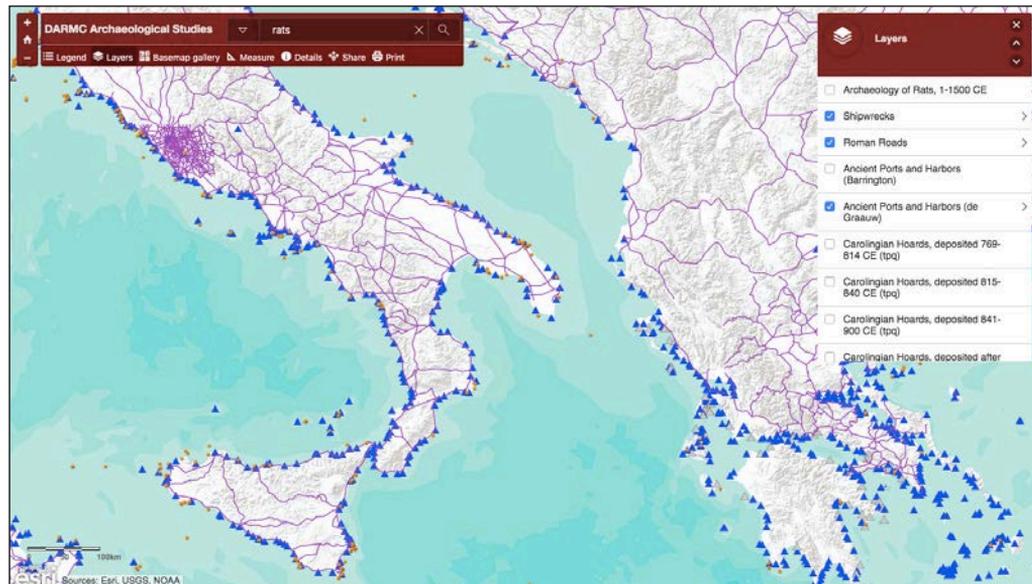
È infine doveroso tenere sempre presenti i dichiarati limiti di questo progetto (come di altri simili), per lo più determinati dalla sua impostazione generale. Si tratta della mancata indicazione della cronologia dei vari siti (e quindi l'impossibilità di eseguire ricerche per cronologie specifiche) e della scarsa precisione del posizionamento geografico di molti luoghi e attestazioni di toponimi. Quest'ultimo problema dipende fondamentalmente dallo stato degli studi, benché siano in atto sforzi costanti per la correzione degli errori e l'aggiunta di nuove coordinate. Chiaramente esiste un ampio divario sulla quantità di informazioni disponibili per ciascuna area geografica: tale disomogeneità non rispecchia lo stato reale della geografia antica quanto piuttosto lacune nella nostra conoscenza, certamente colmabili attraverso un comune impegno di aggiornamento e ricerca.

³⁴ http://dare.ht.lu.se/print.php?doc=info_api.

Digital Atlas of Roman and Medieval Civilizations (DARMC)

Il DARMC (*Digital Atlas of Roman and Medieval Civilizations*)³⁵ è un progetto dell'Università di Harvard che vanta «il miglior materiale disponibile per un approccio GIS alla mappatura e analisi spaziale del mondo Romano e Medioevale», come riporta la prima pagina del sito, per tutta l'area euroasiatica e per un periodo di tempo che comprende i primi 1500 anni della nostra era. Ancora una volta, la base di partenza è il *Barrington Atlas*, la cui precisione geografica viene costantemente aggiornata e raffinata. Il DARMC opera in strettissima connessione con Pleiades e si prefigge lo scopo di diventare un'interfaccia geografica unitaria di interrogazione dei dati da questo raccolti e mantenuti. L'infrastruttura geografica è realizzata con tecnologie ESRI, su piattaforme ArcGIS e ArcGIS Server (menzionate sopra). L'ambito geografico mediterraneo e centroasiatico è in fase di espansione verso l'Africa e l'Estremo Oriente, con l'integrazione dell'*Historical Atlas of China*.

Il DARMC pubblica i propri dati con licenza Creative Commons BY-NC-



Schermata della sezione archeologica del DARMC. Mappa che rappresenta la rete viaria di età romana, i relitti noti e i porti antichi (secondo il già menzionato studio di de Graauw). Fonte: <https://harvard-cga.maps.arcgis.com/apps/View/index.html?appid=e097dbcc4ee04a34ab0d5b-5169b8feae>.

³⁵ <https://darmc.harvard.edu/>.

SA 4.0, il che offre la possibilità agli utenti di scaricarli e di riutilizzarli in tutti i modi possibili, a condizione di citarne la fonte di provenienza. È possibile creare prodotti derivati (altre banche dati o applicazioni web) e ridistribuirli liberamente. Tutti i dati sono pubblicati nel loro formato di origine e accompagnati da dettagliati metadati, che offrono una completa descrizione delle risorse edite e tengono traccia del processo di lavorazione (progetti, persone e metodologia) e della consistenza degli archivi³⁶. Oltre al formato originale, i dati sono pubblicati anche come servizio WMS, sia per l'utilizzo *online* che per quello *desktop*.

I temi cartografici coprono un'ampia varietà di ambiti diversi, come:

- rete stradale di età romana
- rete portuale di età romana³⁷
- attestazioni e posizionamento dei relitti navali
- temi relativi alla monetazione (produzioni, rinvenimenti di tesoretti, ecc.)
- dati economici di età romana (preziari)
- dati climatici e ambientali
- dati sugli insediamenti anglosassoni
- dati sulle diocesi francesi intorno all'anno 1000, ecc.

Alcuni di questi *dataset* sono particolarmente adatti per un uso generico, altri – molto specialistici – offrono approfondimenti su ambiti di ricerca specifici.

Infine è importante sottolineare che il DARMC (come gli altri esempi visti) non offre solo dati, ma anche una metodologia e una strategia di ricerca e pubblicazione dei dati stessi, fornendo agli utenti importanti spunti di riflessione. In linea di principio, ogni lavoro di ricerca – dalla tesi di laurea di uno studente o di uno specializzando, fino a una tesi di dottorato e oltre – può essere reso pubblico mediante questa dinamica e reso accessibile da parte di una comunità più ampia, nonché diventare così risorsa disponibile per ulteriori studi.

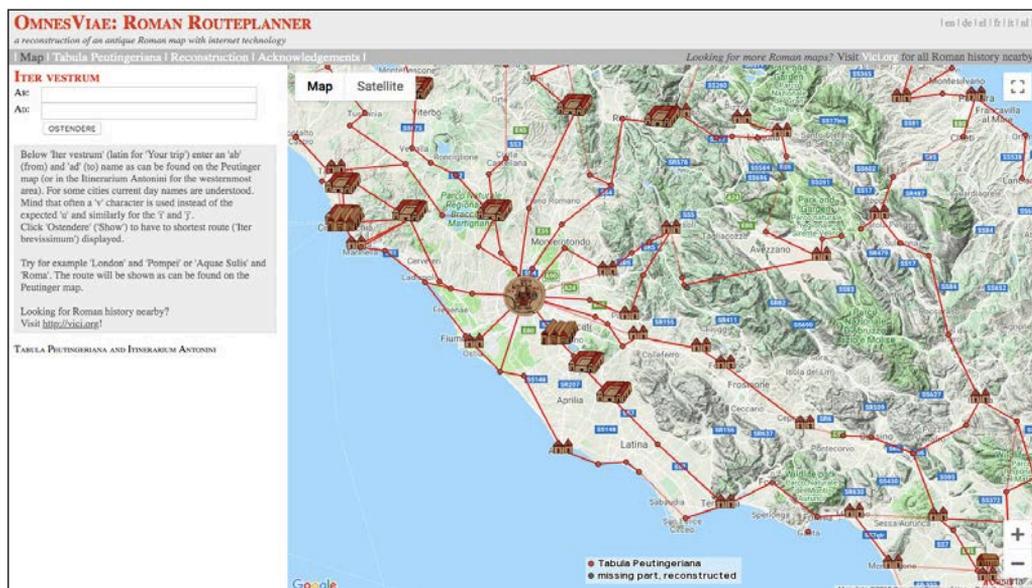
³⁶ Per una lista dei dataset disponibili si veda <https://darmc.harvard.edu/data-availability>.

³⁷ Questi dati, curati da Arthur de Graauw, sono stati pubblicati anche in un sito dedicato, per opera dello stesso autore <http://www.ancientportsantiques.com/>.

OmnesViae e Vici.org

Il lavoro di Richard Talbert sulla *Tabula Peutingeriana*, ha prodotto, oltre all'edizione cartacea³⁸, un sito web che ha accolto una grossa mole dei dati relativi a quello studio³⁹. Il sito ha stimolato nel 2011 la creazione, da parte di René Voorburg, di <https://omnesviae.org>, un progetto che, utilizzando l'opera di Talbert e le API pubbliche di Google Maps, ha messo in rete e ricostruito la mappa romana. I vari luoghi della *Tabula Peutingeriana* sono riportati sulla base cartografica contemporanea di Google, con la simbologia di rappresentazione originaria e collegati tramite tratti rettilinei. L'applicazione web integra anche un motore di ricerca che, rigorosamente in latino, offre un servizio di orientamento simile a quello di Google: dato un luogo di partenza (*ab*) e un luogo di arrivo (*ad*) è possibile visualizzare (*ostendere*) le tappe da percorrere, secondo il risultato di un'analisi di rete sui dati geografici disponibili. Strumento molto utile e istruttivo, questo *route planner* riporta per ogni luogo i toponimi nelle diverse varianti (quelli della *Tabula* e quelli normalizzati), tutti ricercabili e anche il relativo estratto della *Tabula Peutingeriana*.

Schermata dell'area di Roma e Lazio in OmnesViae.



³⁸ Richard John Alexander Talbert, *Rome's World: The Peutinger Map Reconsidered* (Cambridge: Cambridge University Press, 2014).

³⁹ Il sito, mantenuto dall'Università di Cambridge, è disponibile all'indirizzo <https://www.cambridge.org/us/talbert/>.

Lo stesso Voorburg ha realizzato, a partire dal 2012, un altro portale dedicato al mondo romano, Vici.org⁴⁰, un progetto che punta a creare un *gazetteer* georiferito, relativo al periodo romano, i cui dati sono distribuiti con licenza Creative Commons BY-SA 3.0. La differenza fondamentale rispetto ad altri progetti – come Pleiades, al quale si collega – è la modalità di implementazione, plasmata sulla pratica di Wikipedia.

Il portale prevede infatti la possibilità, previa autenticazione, di modificare o aggiungere contenuti geografici, grafici (immagini) e testuali in più lingue (inglese, tedesco, francese e olandese). La forma di implementazione – estremamente agile – ha fortemente incentivato il popolamento di questa banca dati, che opera in parallelo agli altri *database* menzionati riportando sempre i collegamenti relativi. Ad esempio: <http://vici.org/vici/1692> (Roma) è collegata a Pleiades 423025, a DARMC 30635, a DARE 1438 e così via. Anche Vici.org prevede un accesso dinamico ai propri dati attraverso una ben documentata API⁴¹, che permette di interrogare il *database* sia geograficamente che testualmente. In modo simile all'API di Pleiades, anche questo strumento può essere utilizzato per la creazione di servizi e applicazioni web.

Vista della pagina principale di Vici.org.

Vici.org utilizza OpenLayers per la visualizzazione dinamica delle mappe e fa uso di varie cartografie di base, relative sia al mondo antico (AWMC, DARE) che a quello attuale (OpenStreetMap e alcune basi Mapbox).



⁴⁰ <https://vici.org/>.

⁴¹ <https://vici.org/dataservices.php>.

È infine prevista anche la possibilità di importare da Vici.org e pubblicare sulle proprie pagine web piccoli riquadri cartografici, analogamente a come si fa con Google Maps, utilizzando un *widger*⁴² JavaScript molto simile a quello ben noto per incorporare (*embed*) i video da YouTube. Con questo *widger* è possibile creare carte in modo dinamico e pubblicarle facilmente come illustrazioni nel proprio sito. Questa funzione è usata in modo estensivo dal sito Livius.org⁴³, un ricco portale di contenuti relativi alla storia del Mediterraneo antico.



Livius.org con al centro il widget di rappresentazione geografica derivato da Vici.org, incentrato sulla Via Egnatia e la scheda relativa. Fonte <https://www.livius.org/>

⁴² Un *widger* (da *window gadget*) è un componente grafico che permette di aggiungere funzionalità specifiche al programma o applicazione.

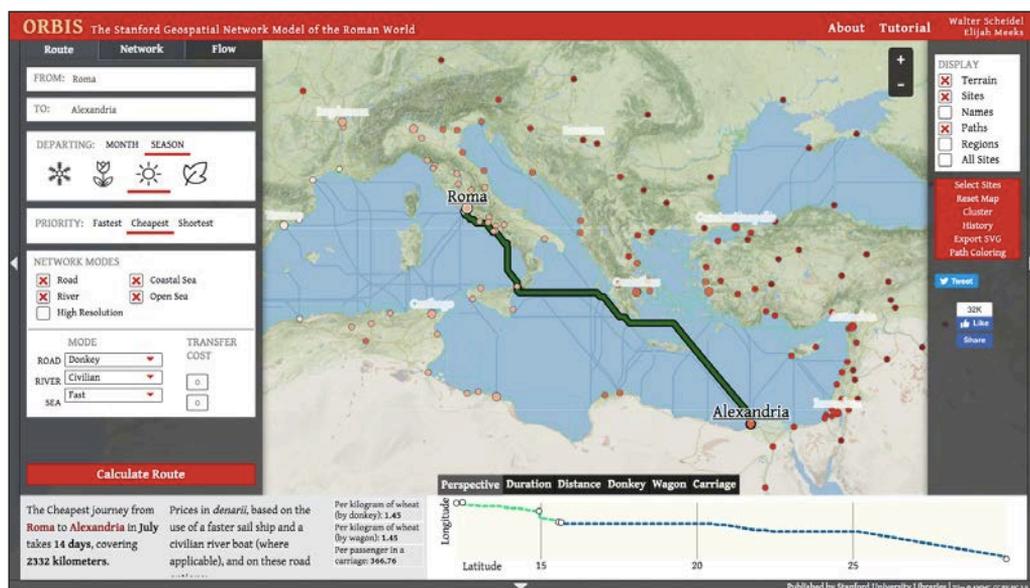
⁴³ <https://www.livius.org/>.

ORBIS: The Stanford Geospatial Network Model of the Roman World

Non poteva mancare in questo elenco l'Università di Stanford, in California, che ha creato, grazie al lavoro di W. Scheidel ed E. Meeks, il modello ORBIS dei trasporti via terra, fiume e mare del mondo romano intorno al 200 d.C., e che è in grado di calcolare la spesa di viaggio sia in termini di tempo che di costo⁴⁴. Il modello si basa su una versione molto semplificata della rete reale di trasporti e comunicazioni del mondo romano, fatta di centri abitati, vie e corsi d'acqua. Il *database* è composto da 632 siti, per lo più città, ma anche gole, passi e promontori, nodi importanti per le comunicazioni. I viaggi in mare sono arricchiti da variabili ambientali quali la direzione media dei venti e l'incidenza del moto ondoso e delle correnti per ogni mese dell'anno. A differenza dei progetti fin qui mostrati, il modello di ORBIS fa largo uso di calcoli statistici e probabilistici per definire i costi e stabilire, per ogni stagione dell'anno, qual è la via che meglio risponde ai parametri definiti dell'utente (ad esempio se la più veloce o la più economica).

Un esempio di calcolo del viaggio da Roma ad Alessandria d'Egitto, in estate, utilizzando l'opzione più economica: un viaggio di 2332 km della durata di 14 giorni.

Il mondo antico è spesso percepito come una realtà statica, e certamente il concetto di "distanza" ha subito un radicale cambiamento, anzi una vera ri-



⁴⁴ <http://orbis.stanford.edu/>, Stuart Dunn, «Review of ORBIS», *Journal of Digital Humanities* 1, n. 3 (2012), <http://journalofdigitalhumanities.org/1-3/review-of-orbis-project-by-stuart-dunn/>.

voluzione, negli ultimi cento anni. Per questo motivo è di estremo interesse – e molto istruttivo – visualizzare su una mappa le possibilità di movimento e di comunicazione nell’antichità e vedere rappresentata anche la grande differenza in termini di tempi e di costi che uno stesso viaggio potrà avere se intrapreso in diversi momenti dell’anno. Per propria scelta, il progetto ha un approccio dall’alto verso il basso, vale a dire che è più interessato alle macro-dinamiche – il sistema nella sua interezza, la connettività su larga scala – piuttosto che alle connessioni o variabili locali.

ORBIS è quindi la rappresentazione dell’Impero Romano nel momento della sua massima estensione, considerato dal punto di vista di uno dei suoi più importanti realizzazioni: l’efficiente e capillare sistema di comunicazioni. L’applicazione web è realizzata con diverse tecnologie, mentre la visualizzazione della mappa è realizzata con D3.js.

ToposText

ToposText⁴⁵ non è un atlante, e neppure un'applicazione webGIS vera e propria, ma fa uso di applicazioni di mappe web per collegare testi letterari e geografia. Inizialmente sviluppata per essere un'applicazione per dispositivi mobili, è ora disponibile anche come sito web. Si tratta di una raccolta di testi antichi accuratamente indicizzata dal punto di vista geografico,

in modo da fornire, per ciascun testo, una mappa complessiva dei luoghi citati; ciascun luogo, a propria volta, è legato a tutte le occorrenze testuali che lo menzionano. ToposText è uno strumento costruito con le API di Google Maps, molto utile per navigare geograficamente i testi antichi e, nell'altro verso, esplorare i luoghi attraverso i riferimenti che di questi si trovano nei testi.

Il progetto rende disponibili tutti i dati in formato aperto, anche se non viene esplicitata la licenza d'uso⁴⁶.

La mappa delle menzioni del toponimo Egitto in ToposText, <https://topostext.org/plate/271307REgy>.

The screenshot shows the ToposText interface for Egypt. At the top, there is a map of Egypt with a yellow location pin. Below the map, the page title is "Agyptos (Egypt) 5618 Egypt - Aiyuroc". On the left, there are metadata fields for the region, including its name, coordinates, and administrative details. The main part of the page is a table listing text entries with columns for Author, Title, Text, Date, and Full. The entries are sorted by date, showing various references to Egypt in ancient Greek literature.

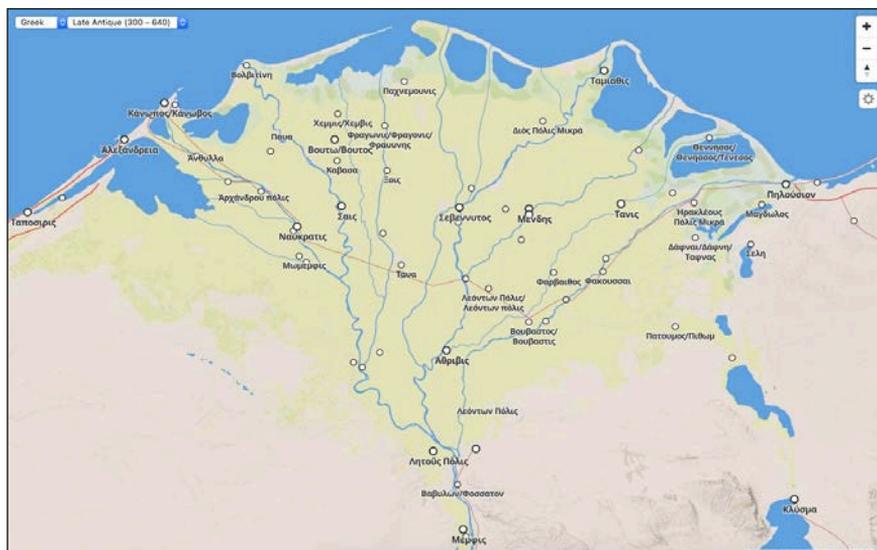
Author	Title	Text	Date	Full
Homer, the Iliad	[3.248	all the wealth that goeth in to Chryseïda, or to Thebes of Egypt, where resources in greater store are laid up in man's houses,	-1100	Read Text
Homer, the Odyssey	[3.253	the other five cities, they were taken by walls and assa to Egypt, where Menelaos gathered much gold and substance among people of an	-1100	Read Text
Homer, the Odyssey	[4.58	get home with my feet, I want to Cyprus, Phoenicia and the Egyptians; I want also to the Ethiopians, the Sabeans, and the Erimonians,	-1100	Read Text
Homer, the Odyssey	[6.113	worksheets which Alcandri wife of Polybus had given her; Polybus had in Egyptian Thebes, which is the richest city in the whole world; he gave	-1100	Read Text
Homer, the Odyssey	[4.223	been given to Helen by Polydamas wife of Theon, a woman of Egypt, where there grow all sorts of herbs, some good to eat	-1100	Read Text
Homer, the Odyssey	[4.357	went trying to come on her, but the gods detested me in Egypt; for my heartaches had not given them full satisfaction, and the [2 line]	-1100	Read Text
Homer, the Odyssey	[4.396	under the sea heropetide and whose name is Proteus. He is an Egyptian, and people say he is my father; he is Neleus's head	-1100	Read Text
Homer, the Odyssey	[4.464	lean house; so you have returned to the heaven-fed stream of Egypt, and offered holy hecatombs to the immortal gods that reign in [2 line]	-1100	Read Text
Homer, the Odyssey	[4.549	our seas; I again stationed my ships in the heaven-fed stream of Egypt, and offered hecatombs that were full and sufficient; when I had	-1100	Read Text
Homer, the Odyssey	[14.234	orders, took to their next devices, and reaped the land of the Egyptians, killing the men, and taking their wives and children captive; The [3 line]	-1100	Read Text
Homer, the Odyssey	[17.386	orders, took to their next devices, and reaped the land of the Egyptians, killing the men, and taking their wives and children captive; The [4 line]	-1100	Read Text
Homer's Hymn to Demeter 2	[25	see men will see to him; I reason he is bound for Egypt or for Cyprus or to the hyperboreans or further still; But in	-1100	Read Text
Aeschylus, Persians	[332	the other slaves, Artaban, had died near to the fountains of the Egyptian Nile, Abubas, and Pharnaces, and Pharnaces fell from one ship; Mardus, Ornyas's	-475	Read Text
Aeschylus, Suppliants	[1	[Enter a company of maidens, who have fled from Egypt and just landed on the shores of Argos; with them is their	-1100	Read Text
Aeschylus, Suppliants	[332	and without you what is accomplished for mankind? [The herald of the Egyptians is seen at a distance, with armed followers.]	-1100	Read Text
Aeschylus, Suppliants	[325	shout and call upon the gods; - you will not escape the Egyptian ship; Cry out, utter a strain of woe more bitter still. CHORUS:	-1100	Read Text
Pindar, Nemean Odes	[10.1	and the Gargen Melusis, [2] and many are the cities founded in Egypt by the devising of Euaphros; Her old Hippocleides goes sailing, when she	-1100	Read Text
Sophocles, Oedipus in Colonus	[330	the least? The thoughts and actions all Are fatal and ministered on Egyptian ways; For there the man of all the loom misdoers While the	-1100	Read Text
Herodotus, Histories	[1.1	themselves fertile to make long voyages by sea; And conveying merchandise of Egypt and of Assyria they arrived at other places and also at Argos; [2 line]	-1100	Read Text
Herodotus, Histories	[1.2	In this manner the Persians report that he came to Egypt; not agreeing therein with the historians, and thus they say was	-1100	Read Text

45 <https://topostext.org/>.

46 <https://topostext.org/TT-downloads>. È possibile scaricare i dati in formato KML (compreso in KMZ) e anche in triple RDF.

Pelagios Commons

Pelagios Commons⁴⁷ è un progetto per la costruzione di una comunità e un'infrastruttura per la diffusione dei Linked Open Data in ambito umanistico. La comunità è formata da enti, istituzioni e individui che, mediante gli strumenti che si passeranno in rassegna tra poco, raccolgono e rendono pubblici dati da loro creati e implementati. La rete scientifica è divisa in gruppi di lavoro tematici (*Working Groups* o WG) su argomenti specifici riguardanti l'applicazione di strumenti digitali alla ricerca antichistica. I temi alla base dei *Working Group* possono riferirsi a un particolare periodo cronologico o un ambito culturale specifico, oppure concentrarsi sulla creazione di un qualche nuovo strumento, *software*, servizio, mappa, *dataset* o altro.



La creazione della carta del Delta del Nilo in età tardoantica, con i toponimi nella versione greca è possibile grazie alla tecnologia di tile vettoriali. Fonte: <https://klokantech.github.io/roman-empire>

⁴⁷ <http://commons.pelagios.org/>. Rainer Simon, Elton Barker, e Leif Isaksen, «Exploring Pelagios: a visual browser for geo-tagged datasets», in *International Workshop on Supporting Users' Exploration of Digital Libraries, 23-27 September 2012, Paphos, Cyprus* (Paphos, 2012); Leif Isaksen, Rainer Simon, Elton T. E. Barker e Pau de Soto Canameres, «Pelagios and the Emerging Graph of Ancient World Data», in *ACM Web Science Conference, WebSci '14, Bloomington, IN, USA, June 23-26, 2014* (Bloomington, Indiana: ACM, 2014), 197–201; Rainer Simon, Elton Barker e Leif Isaksen, «The Pleiades Gazetteer and the Pelagios Project», in *Placing Names: Enriching and Integrating Gazetteers*, a c. di Merrick Lex Berman, Ruth Mostern, e Humphrey Southall (Bloomington: Indiana University Press, 2016), 97–109.

L'infrastruttura che Pelagios ha promosso in questi anni è composta principalmente di quattro strumenti principali: un servizio web di mappe XYZ, Recogito, Peripleo e una API per l'accesso automatizzato ai *dataset*.

Servizio web XYZ

Pelagios pubblica e mantiene un servizio web XYZ (*map tiles*) che costituisce una base aggiornata che ritrae l'Impero Romano al culmine della sua estensione⁴⁸. Si tratta della stessa base distribuita e resa disponibile anche dal progetto DARE. Oltre alla base *raster* è stata predisposta anche una base vettoriale che, attraverso un più recente protocollo che fa uso di *tile* vettoriali, permette non solo una migliore resa grafica (anche su schermi ad alta definizione, come quelli *retina*), ma anche la possibilità di una maggiore interazione con i dati, come ad esempio scegliere la lingua di visualizzazione dei toponimi (greco, latino, o quella attuale a seconda del paese di localizzazione), la possibilità di creare filtri dinamici sulla cronologia, tipologia o altri attributi dei dati, ecc.⁴⁹.

Recogito

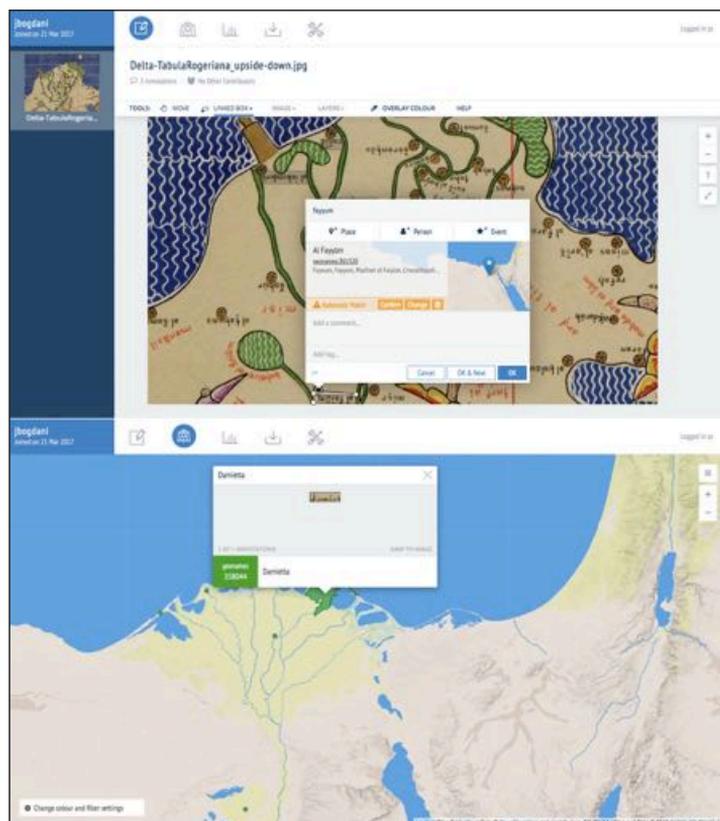
Recogito è un'applicazione web, liberamente disponibile *online*, che facilita la creazione di annotazioni semantiche e la produzione di Linked Open Data⁵⁰. In particolare, la piattaforma permette il caricamento di risorse digitali, in forma testuale o grafica (immagini di vario genere, come ad esempio una carta geografica, ma anche l'immagine di un papiro, di un manoscritto, di una iconografia, ecc.) e consente di evidenziare e segnalare su queste parole o porzioni di immagini che si riferiscono a luoghi, persone o eventi. Questa pratica è conosciuta con il nome di annotazione (*annotation*) e permette di aggiungere a una risorsa ulteriori livelli informativi, senza apportare cambiamenti alla risorsa stessa. Non è molto diversa dai sistemi manuali di sottolineatura e notazione marginale che spesso tutti noi adottiamo sui formati cartacei o digitali, ma la potenza della connessione Web e la possibilità di far riferimento a vocabolari condivisi ne fanno uno strumento di approfondimento notevole⁵¹.

⁴⁸ <http://commons.pelagios.org/2012/09/a-digital-map-of-the-roman-empire/>.

⁴⁹ <http://commons.pelagios.org/2017/11/building-the-roman-empire-vector-tile-map/>.

⁵⁰ <https://recogito.pelagios.org>.

⁵¹ La Web Annotation è dal 2017 uno standard W3C, parte della famiglia del



Un esempio di utilizzo della piattaforma Recogito. In alto si vede la fase di annotazione di uno stralcio relativo al Delta del Nilo di una carta storica (la cosiddetta *Tabula Rogeriana* redatta da al-Idrisi nel XII secolo); il toponimo al Faiūm viene collegato a Al Fayyūm, ovvero Geonames id 361320 (il sistema suggerisce le opzioni possibili in base a ricerche testuali semplici). In basso, i toponimi marcati possono essere visualizzati su una base cartografica moderna, anche se non si è provveduto all'inserimento manuale delle coordinate, che sono state estrapolate dal gazetteer di riferimento. Le occorrenze rimangono collegate alla porzione del documento di riferimento.

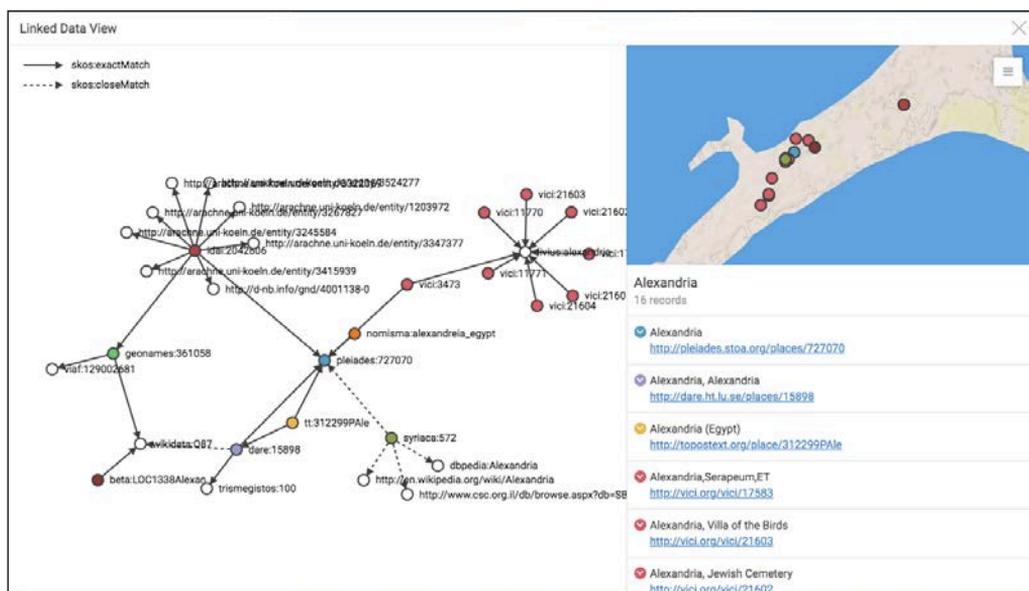
Queste occorrenze sono normalmente collegate a *gazetteer* esterni (per esempio al *gazetteer* di Pleiades nel caso della toponomastica antica o di Geonames per quella contemporanea), il che permette di collegare la risorsa digitale che viene annotata ad altri dati e informazioni disponibili in Rete. Nel caso dei luoghi, il collegamento con Pleiades permette la creazione automatica di mappe dinamiche a partire da testi o dalle immagini. Queste annotazioni possono successivamente essere esportate in una varietà di formati diversi, per un loro successivo utilizzo esterno in altre piattaforme che possono essere geografiche (tra i formati possibili ci sono i CSV e il GeoJSON, facilmente integrabili nei GIS), semantiche (è possibile esportare triplette RDF, in diversi formati), o orientate all'analisi dei testi (TEI). In forma sperimentale, tramite formati adatti per l'apprendimento automatico da parte delle macchine (*machine learning*), le anno-

Web semantico: sono stati definiti un *data-model* (<https://www.w3.org/TR/annotation-model/>), un protocollo (<https://www.w3.org/TR/annotation-protocol/>) e un vocabolario (<https://www.w3.org/TR/annotation-vocab/>). Per esempi d'utilizzo pratico generalizzato di questo tipo di tecnologia si veda il progetto Hypothesis.is, <https://web.hypothesis.is/>.

zioni possono essere anche utilizzate per ulteriori analisi automatizzate. Questa piattaforma rende quindi possibile l'estrazione in maniera estremamente semplice di alcuni concetti chiave – per ora luoghi, persone o eventi, espressi sotto forma di testo o immagine – da una risorsa digitale e la loro esportazione e utilizzo all'esterno, abbattendo la barriera della codifica e della validazione manuale delle triplette RDF.

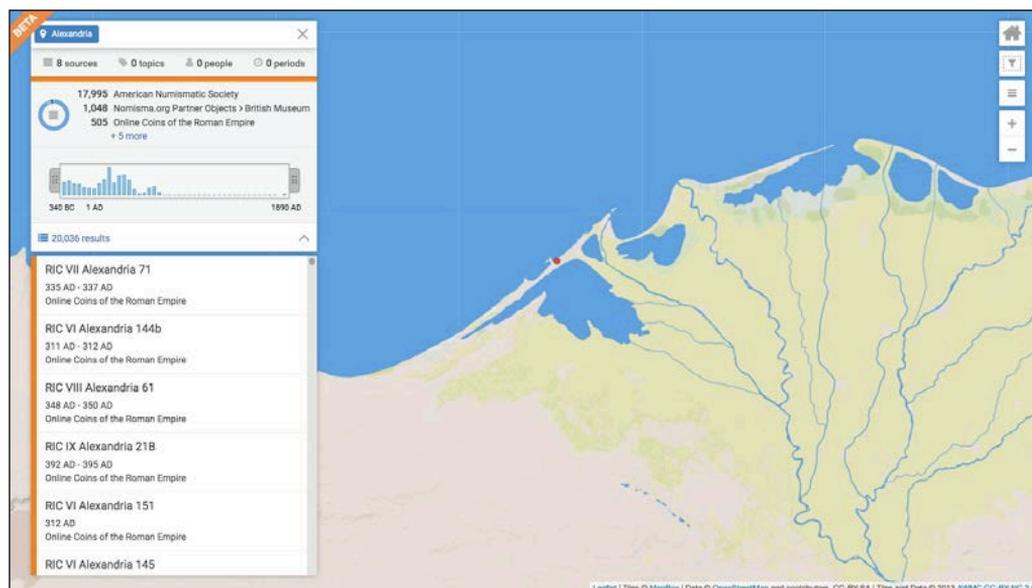
Peripleo

Peripleo⁵² è un servizio di ricerca su un insieme di contenuti e banche dati semantiche sviluppate dalla comunità: persone, progetti e istituzioni partner. Questi soggetti pubblicano e mantengono aggiornati sulla piattaforma comune i propri dati, debitamente strutturati. Il punto comune che le singole banche dati devono implementare è il collegamento ai luoghi, che costituiscono il vero perno dell'intera piattaforma. Il georiferimento – con vari gradi di precisione – diventa quindi un requisito indispensabile per la pubblicazione di dati principalmente archeologici (vi sono presenti a oggi



Linked Open Data in Peripleo, il caso di Alessandria d'Egitto: nel grafo è rappresentata la mappa completa delle forme possibili del toponimo nelle varie banche dati partecipanti, ciascuna accompagnata dai propri identificativi (URI). Questi identificativi sono collegati attraverso una puntualissima etichetta `skos:exactMatch` o una più generica `skos:closeMatch`.

⁵² <http://pelagios.org/peripleo>. Rainer Simon, Leif Isaksen, Elton Barker e pau Cañameres de Soto, «Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time», *Code4Lib* 31 (2016).



importanti collezioni di monete, epigrafi e oggetti di interesse archeologico), che è possibile ricercare e filtrare secondo la cronologia, tipologia o altri parametri. Ciascun fornitore di dati provvede in maniera autonoma e indipendente a mantenere *online* i propri *database*, i quali vengono elencati in una sorta di “indice”, compilato nell’ormai familiare formato RDF, secondo uno schema molto elastico noto con il nome di Pelagios Gazetteer Interconnection Format⁵³. L’“indice”, accessibile *online*, elenca e descrive – facendo uso della combinazione di varie ontologie – tutte le risorse presenti nel *database* che si vogliono rendere pubbliche attraverso Peripleo. Agli studiosi è lasciato ampio margine di libertà nella descrizione dei propri contenuti, ma è parimenti richiesta massima precisione e nella definizione degli elementi utili alla ricerca, come ad esempio la cronologia, e soprattutto nella disambiguazione dei luoghi, dato il ruolo decisivo che questi rivestono nell’esplorazione delle risorse. Gli indici vengono raccolti (ingl. *harvesting*) da Peripleo e registrati nella sua banca dati, resa accessibile dal portale geografico. Tramite quest’ultimo è possibile effettuare ricerche semantiche su un luogo di nostro interesse, accedendo a tutte le risorse che vi fanno riferimento, oltre che eseguire filtri cronologici e tipologici o limitare la ricerca a qualche *database* specifico. Per ogni oggetto estratto dalla ricerca – una moneta, una statua, un testo o un oggetto della vita quotidiana conservato in qualche museo o collezione – è poi possibile

Di Alessandria, Peripleo contiene il riferimento a 20.036 oggetti, sparsi in otto banche dati diverse, in gran parte monete. La ricerca può essere circoscritta a una cronologia più specifica o essere limitata a un unico database.

⁵³ Il formato è documentato al seguente indirizzo: <https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>.

consultare le informazioni aggiuntive che il *database* di riferimento contiene. Grazie poi al sistema degli URI (in questo caso URL) della specifica RDF, è sempre possibile in Peripleo risalire al *database* d'origine che ha generato il *record* e consultare la scheda completa.

Peripleo API

Come diversi altri progetti fin qui citati, anche Pelagios Commons abilita un accesso automatizzato ai propri dati attraverso un servizio dedicato, chiamato Peripleo API⁵⁴, che permette la libera interrogazione della banca dati principale anche da servizi web o *desktop* esterni. È possibile interrogare i dati mediante uno o più termini che si possono riferire alla tipologia del reperto (es. *coin* per le monete), al materiale (es. *gold* per oro) o a un luogo (es. *Athen**), ma è anche possibile eseguire *query* tipicamente geografiche e topologiche, come per esempio richiedere tutte le risorse comprese entro un rettangolo di coordinate date (*bounding box* o, più comunemente *bbox*) o che sono comprese entro un certo raggio a partire da una specifica coordinata. Per gli oggetti dei quali viene fornita una data, è possibile anche richiedere un arco cronologico ben definito. Il servizio restituisce risultati impaginati (divisi in gruppi, per evitare di sovraccaricare il *server* e ottenere dati troppo pesanti), in formato JSON.

Un esempio semplice di interrogazione potrebbe essere: “Trova tutte le monete di VI sec. a.C.”. Tale ricerca si può eseguire attraverso il seguente URL:

```
https://peripleo.pelagios.org/peripleo/  
search?query=coin&from=-600&to=-  
500&types=item&prettyprint=true
```

dove:

- <https://peripleo.pelagios.org/peripleo/search> costituisce l'URL di base
- `query=coin` è il termine ricercato, in questo caso “moneta”
- `from=-600` costituisce il primo termine cronologico, in questo caso il 600 a.C. (si noti il segno “-” prima del numero)
- `to=-500` costituisce il secondo termine cronologico, in questo caso il 500 a.C.
- `types=item` costituisce la tipologia del risultato richiesto; in questo

⁵⁴ <https://github.com/pelagios/peripleo/>.

caso si stanno cercando delle monete (oggetti, *item*), ma è anche possibile cercare banche dati (quale *database* contiene monete del VI sec. a.C.?)

- `prettyprint=true` è una direttiva che fa sì che il risultato sia impaginato in un formato leggibile anche da una persona (con margini, tabulazioni, interruzioni di linee); utile per un uso di prova, è superfluo per l'uso mediante applicazioni.

Una nota conclusiva riguarda i contenuti che questo progetto mette a disposizione, che attualmente vengono forniti da progetti e istituzioni di natura e scala molto differente, tra i quali anche *dataset* importanti come quelli dell'Istituto Archeologico Germanico⁵⁵, quelli del progetto EAGLE (Europeana Network of Ancient Greek and Latin Epigraphy)⁵⁶, dell'American Numismatic Society⁵⁷, di Pleiades stessa o del China Historical GIS⁵⁸. Al di là di questo, lo strumento è più una dimostrazione di fattibilità di quello che è possibile fare con dati strutturati, descritti, collegati e aperti che un vero e proprio strumento di ricerca. Peripleo traccia quindi una possibile linea metodologica orientata alla costruzione di motori di ricerca specializzati nel campo delle antichità (principalmente dell'archeologia), con strumenti derivati dalle tecnologie dell'informazione e comunicazione legate al mondo del Web e – fattore importante – mediante un approccio collaborativo e dal basso, derivato anch'esso dallo sviluppo di Internet. Non è chiaro quale sarà il futuro di questa iniziativa, ma certamente un atteggiamento che valorizza l'apporto dei singoli fornitori di dati (che in questo caso possono sviluppare la propria ricerca senza rigide limitazioni strutturali e curano in modo indipendente l'edizione dei propri dati) è sicuramente il più adatto per un mondo così fluido e dinamico come quello odierno rispetto a un approccio aggregante, centralizzato e omnicomprensivo, che tende quindi ad accentrare e strutturare in maniera rigida i dati. La discussione è quanto mai attuale e l'equilibrio tra una gestione centralizzata o decentralizzata degli archivi archeologici è difficile da trovare, in un momento nel quale le tecniche di analisi e trattamento dei dati derivate dagli studi sui cosiddetti *big data* sono sempre più applicate anche al dominio della storia e dell'archeologia⁵⁹.

⁵⁵ Si veda oltre la scheda dedicata a `iDAI.objects` `arachne`.

⁵⁶ <https://www.eagle-network.eu/>.

⁵⁷ <http://numismatics.org/>.

⁵⁸ <http://chgis.fas.harvard.edu/>

⁵⁹ Gabriele Gattiglia, «Think Big about Data: Archaeology and the Big Data Challenge», *Archäologische Informationen* 38 (2015): 12..

FastiOnline

FastiOnline è un sito web a cura dell'Associazione Internazionale di Archeologia Classica (AIAC), attivo dal 2003 e volto a fornire notizie e aggiornamenti su scavi e ricerche archeologiche, in particolare in Italia, ma più in generale aperto a tutto il bacino del Mediterraneo⁶⁰. Il sito porta avanti la missione di una precedente versione cartacea, "Fasti Archaeologici. Annual Bulletin of Classical Archaeology" (sempre a cura dell'AIAC), che, a causa dei costi elevati e della difficoltà a garantire uscite regolari, è stata interrotta. La banca dati, pur centralizzata, è gestita in maniera condivisa attraverso una serie di corrispondenti per ciascuno dei Paesi partecipanti, designati in base a specifici accordi con le istituzioni nazionali incaricate della ricerca archeologica. Si tratta di una struttura di stampo fortemente istituzionale, che intende garantire un buon grado di uniformità del dato. Il progetto non censisce i siti archeologici, bensì le singole ricerche effettuate all'interno di essi, ciascuna delle quali può essere articolata in più campagne: un sito archeologico grande e importante come Pompei, ad esempio, comprende in aprile 2019 circa 117 indagini schedate. Per ciascuna di queste vengono fornite brevi descrizioni, con una sintesi dei risultati e una serie di metadati sul progetto (enti e istituzioni coinvolte, responsabili scientifici e sul campo, ecc.).

FastiOnline è una risorsa molto utile per le tante informazioni su scavi e ricerche recenti, in gran parte ancora inedite, che è possibile reperirvi. Alla banca dati relativa agli scavi sono state poi aggiunte altre due sezioni parallele, una dedicata a progetti di ricognizione archeologica e l'altra a progetti di conservazione. Come quelle relative agli scavi, anche le nuove sezioni sono accompagnate da un visualizzatore web di dati GIS, che permette di individuare e ricercare i singoli interventi su una mappa. Per venire incontro a esigenze di edizione e comunicazione più approfondite, il portale ha sviluppato anche una propria rivista *online*, chiamata "FOLD&R", divisa, come il sito web, in tre serie: Italy, Archaeological Conservation e Survey, dove è possibile pubblicare rapporti preliminari o approfondimenti a partire dai temi del portale.

Una *repository* dalla forte impronta istituzionale, dunque, abbastanza aggiornata soprattutto per quanto riguarda il territorio italiano. La piattafor-

⁶⁰ <http://www.fastionline.org>, Helga Di Giuseppe e Elizabeth Fentress, «Open Archaeology: i Fasti e la pubblicazione online», in *Open Source Free Software e Open Format nei processi di ricerca archeologica*, a c. di Luca Bezzi, Stefano Costa, Denis Francisci, Piergiovanna Grossi e Damiano Lotto, (Roma: Edizioni Quasar, 2012), 85–93.

ma non prevede strumenti di interazione automatica da parte dei fornitori di dati (i singoli progetti di ricerca che curano le schede delle proprie campagne), i quali trasmettono le informazioni in formato Word ai referenti di area; sono questi che si occupano poi dell'immissione nel sistema. Non è possibile neppure un accesso automatico in lettura da parte di terzi (studiosi o applicazioni), anche se è stato predisposto un collegamento tra i siti presenti in FastiOnline e quelli di Pleiades. Inoltre, FastiOnline pubblica una parte del proprio *gazetteer* in Peripleo.



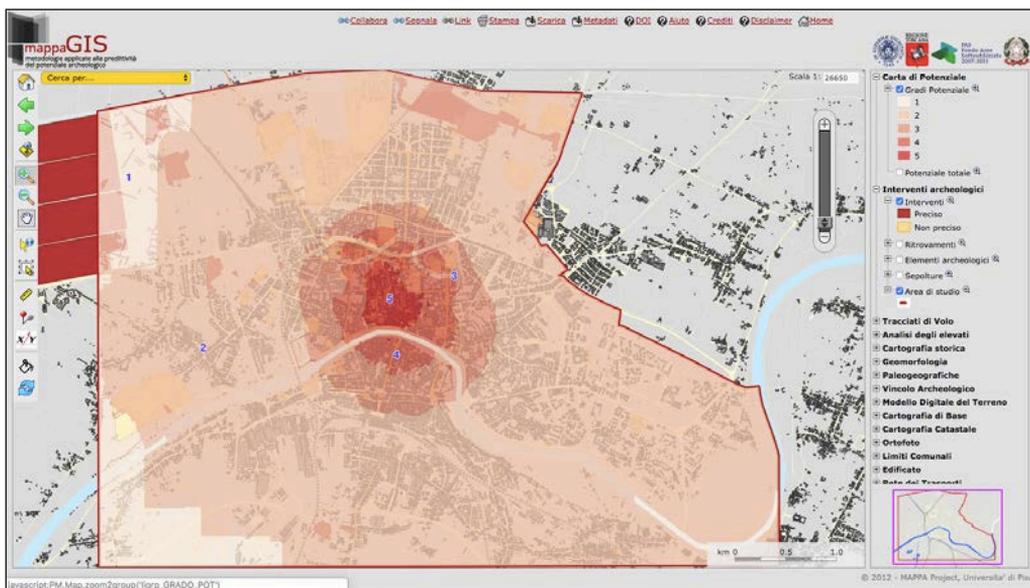
FastiOnline, indagini archeologiche in Italia e aree limitrofe, http://www.fastionline.org/excavation/map_view.php.

MAPPA Project

MAPPA Project⁶¹, finanziato dalla Regione Toscana e dall'Università di Pisa, ha come obiettivo quello di creare un webGIS dell'area urbana di Pisa, frutto dell'integrazione di informazioni archeologiche, geologiche, geomorfologiche. Il risultato finale è una carta del potenziale archeologico dell'area, realizzata mediante lo sviluppo di un algoritmo in grado di calcolarlo in senso predittivo. Il secondo risultato è costituito dal primo archivio italiano di Open Data in ambito archeologico⁶².

Schermata della mappa delle potenzialità archeologiche di Pisa, <http://mappagis.cs.dm.unipi.it:8081/mappa/mappa.phtml>.

Il progetto webGIS⁶³ raccoglie una grande quantità di informazioni geografiche di varia natura sul territorio urbano e periurbano di Pisa: trasporti, dati urbanistici, catastali, cartografia di base, DTM, geomorfologia, cartografia storica, vincoli archeologici, interventi e indagini archeologiche e infine la carta del potenziale archeologico della città, frutto di indagine multidisciplinare. Di particolare interesse sono le ricostruzioni della paleogeografia del centro urbano nel suo sviluppo cronologico (vengono forniti sette periodi che vanno dall'età protostorica all'età contemporanea). Il



⁶¹ <http://www.mappaproject.org>.

⁶² Francesca Anichini, MAPPA, *metodologie applicate alla predittività del potenziale archeologico. Vol. 1* (Roma: Ed. Nuova Cultura, 2012); Francesca Anichini, MAPPA, *metodologie applicate alla predittività del potenziale archeologico. Vol. 2* (Roma: Ed. Nuova Cultura, 2013).

⁶³ Tuttora consultabile all'indirizzo <http://mappagis.cs.dm.unipi.it:8081>.

webGIS integra anche strumenti di ricerca su diversi temi, archeologici o meno, e costituisce uno strumento fondamentale per la conoscenza diacronica dell'archeologia pisana. L'interfaccia pubblica è costruita utilizzando p.mapper⁶⁴, un *framework* PHP per il motore webGIS Mapserver, che si è già avuto modo di introdurre. La piattaforma è pubblicata in sola lettura, dal momento che il progetto non prevede l'implementazione del *dataset* da parte di utenti esterni.

L'algoritmo per il calcolo automatico del potenziale archeologico a partire dal *dataset* raccolto, sviluppato all'interno del progetto, si basa su uno dei più famosi algoritmi del Web, quello sviluppato da Larry Page e Sergey Brin per il loro motore di ricerca Google, noto con il nome di PageRank.

I motivi di tale scelta sono da ricercare nell'analogia tra i criteri usati per attribuire il potenziale archeologico e quelli utilizzati per assegnare l'importanza delle pagine web negli algoritmi dei motori di ricerca. Tale analogia si basa sul fatto che, da un punto di vista concettuale, l'elemento chiave per stimare il potenziale archeologico è l'identificazione delle relazioni che intercorrono tra i ritrovamenti archeologici: la presenza di un determinato ritrovamento nelle vicinanze di un altro può rafforzare o indebolire la probabilità che questi formino una struttura di livello organizzativo più complesso e, dunque, può incrementare o abbassare il potenziale archeologico dell'area in cui si trovano. Gli algoritmi di page ranking si basano su principi molto simili, secondo i quali ogni pagina web attribuisce importanza alle pagine che puntano a quella (mediante un link) e, viceversa, riceve importanza dalle pagine alle quali punta.⁶⁵

Altro risultato importante di questo progetto è la creazione del MOD⁶⁶, acronimo di Mappa Open Data, un archivio (*repository*) aperto per la pubblicazione di dati archeologici grezzi (elenchi e schede di unità stratigrafiche e reperti archeologici, analisi, immagini, fotografie e piante in vari formati, *matrix* stratigrafici, dati di rilievi, ecc.) e di quella che viene de-

⁶⁴ <https://www.pmapper.net/>.

⁶⁵ Nevio Dubbini, «L'algoritmo MAPPa», in *MAPPa, metodologie applicate alla predittività del potenziale archeologico*. Vol. 2, a c. di Francesca Anichini (Roma, 2013), 101–13.

⁶⁶ <http://mappaproject.arch.unipi.it/mod/>.

finita “letteratura grigia” (relazioni o report preliminari, diari di scavo e altre forme di annotazioni e appunti più o meno strutturati, molto comuni negli scavi archeologici). Questa raccolta e pubblicazione avviene nel pieno rispetto della legge sul diritto di autore e della difesa dei diritti morali di chi ha creato tali contenuti. L’attribuzione di identificativi stabili (DOI⁶⁷) e l’adozione di licenze aperte garantiscono la citabilità delle risorse, la tutela e la responsabilità scientifica degli autori. Il MOD era nato per contenere e pubblicare con libero accesso i dati raccolti all’interno del progetto per la creazione del webGIS, ma è stato poi aperto anche a contributi esterni riguardanti l’intero territorio nazionale. Anche in questo caso, lo strumento informatico è accompagnato da un più tradizionale strumento cartaceo, una collana chiamata Mappa Data Book, volta a offrire una sede di edizione tradizionale per pubblicazioni relative ai dati depositati nel MOD. Attualmente (aprile 2019) la collana è alla seconda uscita.

Purtroppo gli archivi del MOD (in totale 124) non sembrano essere stati aggiornati di recente, dal momento che gli ultimi depositi sono relativi all’anno 2015. Sicuramente la legislazione italiana in materia di beni archeologici e culturali, in generale, non è molto incoraggiante nei confronti della libera condivisione dei dati e agli studiosi è richiesto uno sforzo importante per cercare di orientarsi in questioni amministrative e legali riguardanti *copyright* e licenza.

⁶⁷ Il Digital Object Identifier o DOI è un sistema di identificazione univoca delle risorse digitali, ampiamente usato nel modo dell’accademia e della ricerca scientifica, ma non solo. Permette di assegnare un codice alfanumerico a una data risorsa e fa riferimento a una banca dati che permette il collegamento a un preciso indirizzo web (URL) dove è possibile trovare la risorsa o informazioni su di essa. È lo standard, oggi, dell’editoria scientifica digitale.

Archaeology Data Service (ADS)

Il MOD ha cercato di portare in Italia un servizio che l'Archaeology Data Service (ADS)⁶⁸ ha offerto e continua a offrire in Inghilterra da circa ventitré anni. Il servizio è reso disponibile da un'organizzazione senza scopo di lucro, con sede presso l'Università di York, che è anche la capofila di un progetto che coinvolge diversi altri centri universitari e di ricerca del Regno Unito. Ventitré anni, meno di una generazione, sono un brevissimo arco cronologico per l'archeologia, ma sono un periodo lunghissimo nell'era digitale, dove il tempo scorre più veloce e il cambiamento è repentino. Mantenere un dato digitale disponibile e consultabile per venti, trenta o più anni è una sfida che mette in campo non solo il sapere tecnico necessario per la costruzione e gestione di un simile archivio, ma anche la consapevolezza dei produttori di dati, e in questa direzione l'ADS ha fatto un lavoro straordinario di definizione di standard e buone pratiche per la corretta codifica e descrizione del dato archeologico. L'ADS, pur non avendo fini di lucro, richiede il pagamento di una tariffa per il deposito di ciascun *dataset*, in quanto è necessario un continuo rinnovo dei supporti di memoria e un costante aggiornamento dei formati dei *file* per garantirne l'accesso prolungato nel tempo. Pur essendo un servizio nato nel Regno Unito per venire incontro a esigenze di enti di ricerca inglesi, è possibile archiviare nell'ADS anche

Schermata dell'archivio dell'ADS e alcuni dei più recenti depositi.

The screenshot displays the ADS website interface. At the top, there is a navigation menu with links for HOME, SEARCH, DEPOSIT, RESEARCH, ADVICE, ABOUT, GALLERY, and HELP. Below the menu, a search bar is visible with a 'SEARCH' button. The main content area shows search filters categorized into WHAT, WHEN, WHERE, and RESOURCE. The 'WHAT' filter includes options like 'Event (1283)', 'Evidence (94)', 'Object (743)', 'Maritime Craft (16)', and 'Monument (1233)'. The 'WHEN' filter includes 'Prehistoric (333)', 'Roman (237)', 'Medieval (302)', 'Post Medieval (293)', 'Modern (280)', and 'Non UK Period Terms (2)'. The 'WHERE' filter includes 'British Isles (1418)', 'Continental Europe (81)', 'Middle East (16)', 'South America (4)', 'Africa (12)', and 'Asia (15)'. The 'RESOURCE' filter includes 'Funder (881)' and 'Programme (482)'. Below the filters, a list of search results is shown, each with a title, a brief description, and a year. The results include records such as 'Thames through Time. The Archaeology of the Gravel Terraces of the Upper and Middle Thames: The Medieval and Post-Medieval Periods AD 1000-2000', 'Isles of Solly Designated Wrecks Interpretation', 'Field Evaluation at Green Acres, Minsterworth, Gloucestershire', 'King Alfred Way, Newton Pookleford, Devon. Archaeological Excavation', 'Death, Burial and Identity: 3000 Years of Death in the Vale of Auldby', 'Cambridge Antiquarian Society Quarts Publications', 'Land at Cuckham Road, Nr Flabury, Worcestershire. Archaeological Evaluation', '97 London Road, Gloucester. Archaeological Evaluation', 'The Bull Hotel, Hockley, Essex: Historic Building Recording', 'Croome Court, High Green, Worcestershire Archaeological Observation', 'Land to Rear of Ebrington Arms May Lane Ebrington: Archaeological Evaluation', 'Barn West of Manor Farmhouse, Diddington, St Neots. Building Recording', 'Bewer Diversion, Gloucester Bus Station, Station Road, Archaeological Recording', 'Brooklands Farm, Chatterham Road, Evesham, Worcestershire', 'Inventory of Crafts and Trade in the Roman East (ICRATES): database of tileware', 'ARCHAIDE Portal for Publications and Outputs', 'Land adjacent to Edward Massey Gardens, Garnah's Road, Gloucester. Archaeological Field Evaluation', 'Cowlclose House, Barningham, County Durham: Building Recording', and 'Wingmoor Farm, Campden Road, Stratford-on-Avon, Warwickshire. Archaeological Evaluation'. The bottom of the page features the ADS logo, the University of York logo, and links for 'University of York Legal Statements | Website Terms and Conditions | Cookies | Privacy Policy' and 'Contact Us'.

⁶⁸ <https://archaeologydataservice.ac.uk/>.

dataset di istituzioni esterne relative a contesti stranieri.

Dal punto di vista dei contenuti, non ci sono limitazioni particolari sui tipi di *file* digitali che si possono archiviare: audio, *database*, dati di rilievi geofisici, di *laser scanning*, GIS, fotogrammetria, grafica *raster* e vettoriale, fogli di calcolo, dati per la realtà aumentata o più semplici testi. Esiste una lista dei formati consigliati, ma in generale l'utente è chiamato a rispettare pochi semplici consigli e molto buon senso. Un formato, secondo questi consigli, dovrebbe essere:

- adatto a rappresentare e registrare i dati, senza inutili forzature nella codifica;
- il più possibile stabile e non in fase di sviluppo o revisione costante (come avviene per molti formati proprietari che a cadenza regolare introducono incompatibilità con le versioni precedenti);
- supportato (letto e scritto) da tipi diversi di *hardware*;
- supportato (letto e scritto) da architetture e sistemi operativi diversi (Windows/Macintosh/Unix/Linux);
- descritto in maniera dettagliata da documentazione pubblicamente accessibile;
- idealmente, esente da restrizioni legali per il suo utilizzo, quali pagamento di diritti o obbligo di una specifica infrastruttura *hard/software*;
- molto diffuso, in quanto più un formato è diffuso più è probabile che venga supportato per lungo tempo⁶⁹.

L'ADS ha una politica molto efficace in questa direzione, dal momento che l'adozione di formati più particolari, che più si discostano dalle caratteristiche elencate poco sopra, è soggetta a tariffe di deposito maggiori, proprio perché richiede una lavorazione più lunga e complessa da parte degli archivisti.

I dati, una volta depositati, sono consultabili a libero accesso dal sito principale attraverso varie interfacce di ricerca, da quelle più semplici (ricerca libera di testo) fino a quelle guidate, attraverso mappe o percorsi ad albero, che, per scelte successive (geografiche, cronologiche, per soggetto, ecc.), guidano l'utente verso l'archivio richiesto.

I metadati dei singoli archivi sono disponibili anche in formato semantico (triple RDF) e il progetto ha implementato un'interfaccia di ricerca semantica per eseguire in maniera efficiente e veloce *query* in SPARQL nel

⁶⁹ <https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#File%20Management>.

database principale⁷⁰. È naturalmente possibile utilizzare questo servizio per la costruzione di applicazioni terze che facciano uso dei dati dell'ADS. Ai progetti vengono offerti anche altri servizi aggiuntivi, quali la creazione di interfacce di interrogazione personalizzate, perfettamente adatte alla struttura dei dati depositati, che vanno a creare in questo modo delle applicazioni web dedicate. Si tratta di servizi a pagamento, ma che, nel caso di progetti importanti, potrebbero risultare di grande utilità anche per gli utenti finali.

In questa *repository* archeologica non vengono archiviati solo dati, ma anche la sintesi delle ricerche. Un esempio è proprio la rivista dell'Università di York "Internet Archaeology"⁷¹, che dal 1996 pubblica sul Web articoli a libero accesso e il cui archivio è conservato a lungo termine presso l'ADS. La *repository* è tuttora molto attiva: in aprile 2019 vi erano già stati depositati quasi 40 archivi relativi all'anno in corso. Considerata la sua lunga storia, l'alta qualità dei servizi offerti e le garanzie sulla durevolezza dei dati è verosimile che il suo utilizzo venga incrementato nel prossimo futuro, anche in virtù dell'esplicita richiesta da parte dei grandi progetti di finanziamento europei e nazionali di una chiara politica di conservazione a lungo termine dei dati prodotti dalle ricerche finanziate. In Italia non esiste un'istituzione, un ente o un organismo statale, privato o a statuto misto, che si sia ancora prefisso uno scopo del genere. Il MOD di Pisa rimane un esperimento isolato, di successo per quanto riguarda la citabilità degli autori, il libero accesso all'informazione e la riusabilità dei dati, ma che offre scarse garanzie di conservazione a lungo termine, sia per la breve durata di vita del progetto, sia per la mancanza, sul sito, di un'ideale documentazione sulle politiche di conservazione dei dati, sul rinnovo dei supporti di memoria e sulle politiche di migrazione dei formati.

⁷⁰ <http://data.archaeologydataservice.ac.uk/query/>.

⁷¹ <http://intarch.ac.uk/>.

Open Context

Open Context⁷² è una risorsa ad accesso libero per la pubblicazione di dati derivanti da ricerche archeologiche e affini, nata nel 2006. È possibile depositarvi e rendere pubblici documenti, schede, elenchi, diari di scavo, immagini, mappe, vocabolari, tipologie e molto altro. Il progetto è fortemente orientato all'aspetto collaborativo, incentivando l'esplorazione e il riutilizzo dei dati presenti nei suoi archivi. In modo simile a quanto accade con ADS, ai singoli progetti che desiderano archiviare e pubblicare i dati viene richiesto un contributo economico⁷³. Il servizio prevede la pulizia e l'organizzazione dei dati da parte di *editor* esperti, l'assegnazione di identificativi stabili e anche l'attribuzione di DOI⁷⁴. Ai dati

Pagina di Open Context dedicata al progetto DINAA (Digital Index of North American Archaeology), uno dei più grandi con 508.719 oggetti schedati e distribuiti con Licenza CC0

The screenshot shows the Open Context website interface for the project 'Digital Index of North American Archaeology (DINAA)'. At the top, there is a navigation bar with 'About', 'Estimate', and 'Download' buttons. Below the navigation is a map of North America with a green overlay indicating the project's geographic focus. The main content area is divided into several sections: 'Project Abstract' which describes the project's goals and funding; 'Related Data Tables for Download' with a table listing various data types; 'Annotations (2)' which provides metadata for the project; and 'Copyright License' which states that the project is in the Public Domain. There is also a 'Suggested Citation' section and a 'Browse, Search Project' section with search and filter options.

presenti nei suoi archivi. In modo simile a quanto accade con ADS, ai singoli progetti che desiderano archiviare e pubblicare i dati viene richiesto un contributo economico⁷³. Il servizio prevede la pulizia e l'organizzazione dei dati da parte di *editor* esperti, l'assegnazione di identificativi stabili e anche l'attribuzione di DOI⁷⁴. Ai dati

⁷² <https://opencontext.org/>. Sarah Whitcher Kansa e Eric C. Kansa, «Data Beyond the Archive in Digital Archaeology. Advances» in *Archaeological Practice*, 6(02), 2018, 89-92. DOI <http://dx.doi.org/10.1017/aap.2018.7>

⁷³ All'indirizzo <https://opencontext.org/about/estimate> è disponibile un calcolatore automatico che permette di calcolare il costo di deposito in base alla complessità dei dati da depositare, quantificato a partire dal numero delle tabelle, dal numero di immagini, dal numero dei documenti statici (HTML, PDF, e simili), dal numero dei temi GIS e di altri documenti (modelli 3D, video, file relativi alla realtà virtuale, ecc.). I progetti relativi ad elaborati di tesi di vario grado, dottorato compreso, ricevono lo sconto del 25%.

⁷⁴ Per i DOI si veda sopra a pagina 178.

depositati vengono attribuite licenze Creative Commons per facilitarne il riutilizzo e, grazie al modello dei Linked Open Data, gli stessi dati possono essere connessi con altri *dataset* presenti in Open Context o anche fuori da questo contenitore. Per quanto riguarda la conservazione a lungo termine, Open Context si affida alle infrastrutture fornite dalla Biblioteca Digitale della California (California Digital Library)⁷⁵, che dovrebbero rappresentare una garanzia di stabilità.

Diversi sono i progetti importanti che hanno depositato i propri dati utilizzando questo servizio, come “Murlo. Poggio Civitate Excavation Project”⁷⁶, “Çatalhöyük Zooarchaeology” (127.714 oggetti)⁷⁷, “The Gabii Project”⁷⁸, “Petra Great Temple Excavations”⁷⁹, il “Digital Index of North American Archaeology (DINAA)” con circa 508.719 oggetti schedati⁸⁰ e molti altri.

Il deposito in Open Context prevede anche l’annotazione semantica dei dati, con l’adozione di vocabolari condivisi per molte chiavi di ricerca fondamentali, come ad esempio la cronologia o la definizione tipologica. Viene inoltre fornito un accesso programmatico ai dati contenuti mediante una apposita interfaccia API, in grado di restituire i dati in diversi formati, quali JSON (adatto per applicazioni web), GeoJSON (ottimale per applicazioni GIS) e (Geo)JSON-LD per i dati collegati. Sebbene alcune sezioni di questo portale siano ancora in fase di completamento, Open Context costituisce di certo un’iniziativa molto interessante sul tema della pubblicazione, conservazione a lungo termine e riutilizzo dei dati archeologici.

⁷⁵ https://merritt.cdlib.org/m/ucb_open_context

⁷⁶ Anthony Tuck, a c. di. «Murlo». Open Context. <http://opencontext.org/projects/DF043419-F23B-41DA-7E4D-EE52AF22F92F>. DOI: <https://doi.org/10.6078/M77P8W98>. ARK (Archive): <https://n2t.net/ark:/28722/k2222wm10>. 2012-07-06.

⁷⁷ David Orton, Nerissa Russell, Katheryn Twiss, Louise Martin e Sheelagh Frame, a c. di. «Çatalhöyük Zooarchaeology». Open Context. <http://opencontext.org/projects/1B426F7C-99EC-4322-4069-E8DBD927CCF1>. ARK (Archive): <https://n2t.net/ark:/28722/k2f47js2k>. 2013-02-13.

⁷⁸ Rachel Opitz, Marcello Mogetta e Nicola Terrenato, a c. di. «The Gabii Project». Open Context. <http://opencontext.org/projects/3585b372-8d2d-436c-9a4c-b5c10fce3ccd> DOI: <https://doi.org/10.6078/M7N014GK>. 2017-06-04.

⁷⁹ Il progetto dovrebbe essere attivo all’indirizzo <http://opencontext.org/projects/A5DDBEA2-B3C8-43F9-8151-33343CBDC857>, ma al momento della stesura di questo testo la pagina restituiva un errore. I 124.582 reperti relativi allo scavo sono comunque visibili e ricercabili all’indirizzo <https://opencontext.org/se-arch/?proj=10-petra-great-temple-excavations>.

⁸⁰ David G. Anderson, Joshua Wells, Stephen Yerka, Kelsey Noack Myers, Robert Carl DeMuth e Thaddeus Geoffrey Bissett, a c. di. «Digital Index of North American Archaeology (DINAA)». Open Context. <http://opencontext.org/projects/416A274C-CF88-4471-3E31-93DB825E9E4A> DOI: <https://doi.org/10.6078/M7N877QO>. 2015-10-21.

SITAR

Il Sistema Informativo Territoriale Archeologico di Roma (SITAR)⁸¹, a differenza delle iniziative fin qui esaminate, nasce da un'esigenza pratica di una più efficace gestione di un territorio complesso, quale quello della capitale d'Italia, geograficamente molto esteso, archeologicamente ricchissimo e urbanisticamente problematico. Si è trattato dunque della messa a sistema di tutte le informazioni a disposizione della Soprintendenza Speciale Archeologia Belle Arti e Paesaggio di Roma, messe *online* in lettura e consultazione, a seguito di un grosso sforzo di georiferimento e vettorializzazione, attraverso la pubblicazione di una piattaforma webGIS⁸² liberamente accessibile anche al grande pubblico. Il sistema gestisce una quantità notevole di informazioni di varia tipologia e provenienza: dati geologici, geomorfologici, fotografie aeree e ortofoto di varie date, dati LiDAR, temi dal Piano Territoriale Paesistico Regionale (PTPR) della Regione Lazio, dal Piano Regolatore di Roma Capitale, ecc., oltre a tutti i temi relativi alle tracce archeologiche, sia quelle monumentali visibili che quelle indagate nei decenni passati e presenti negli archivi della Soprintendenza. I temi archeologici sono disponibili anche attraverso filtri appositamente predisposti: cronologici⁸³, funzionali⁸⁴, ma anche relativi allo stato di conservazione⁸⁵. Ogni evidenza è interrogabile, in quanto collegata a un protocollo

⁸¹ <http://www.archeositarproject.it>. Il progetto è stato esposto in una serie di convegni organizzati a questo scopo e prontamente pubblicati: Mirella Serlorenzi, a c. di, *SITAR Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, 26 ottobre 2010)* (Roma: Iuno Edizioni, 2011); Mirella Serlorenzi e Ilaria Jovine, a c. di, *SITAR: Sistema Informativo Territoriale Archeologico di Roma: atti del II convegno, Roma 9 novembre 2011* (Roma: Iuno, 2011); Giorgia Leoni e Mirella Serlorenzi, a c. di, *Il SITAR nella rete della ricerca italiana: verso la conoscenza archeologica condivisa: atti del III convegno (Roma, Museo Nazionale Romano, 23-24 maggio 2013)* (Firenze: Edizioni all'Insegna del Giglio, 2015); Mirella Serlorenzi e Ilaria Jovine, a c. di, *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico. Atti del IV Convegno di Studi SITAR (Roma, 14 ottobre 2015)* (Firenze: All'Insegna del Giglio, 2017).

⁸² Oppure web Archaeological Information System (webAIS), come viene chiamato internamente al progetto. La piattaforma è raggiungibile all'indirizzo: <https://webais.archeositarproject.it/>.

⁸³ Le fasi pubblicate sono: preistoria, protostoria, età regia, età romana repubblicana, età romana imperiale, età romana, medioevo, età moderna, età contemporanea.

⁸⁴ Ad esempio, contesti funerari, infrastrutture idrauliche, viabilità e porti, strutture insediative, edifici pubblici, luoghi di culto, àmbiti produttivi, ecc.

⁸⁵ Elevati, strutture, rivestimenti parietali, rivestimenti pavimentali, fondazioni,

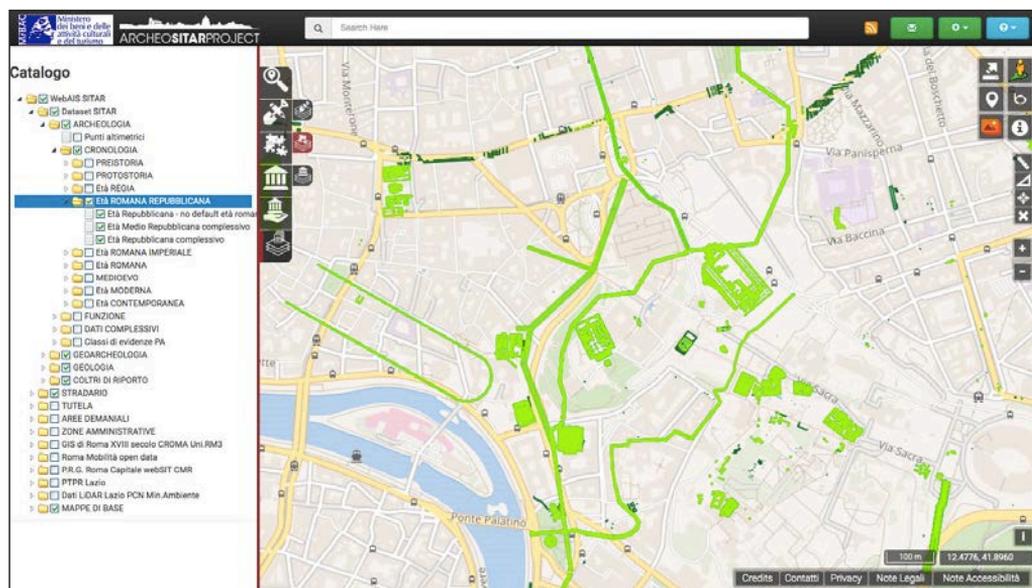
descrittivo che ne riassume le caratteristiche principali e fornisce una breve storia delle ricerche, una funzione di estrema utilità per aree e strutture ricostruite dalla documentazione d'archivio e oggi non più visibili. È difficile essere esaustivi nella descrizione di questo strumento così ricco e non è scopo di questo volume presentarlo nel dettaglio.

Di grande interesse sarà anche la pubblicazione dei dati ivi contenuti in altri formati che ne permettano l'utilizzo; è infatti annunciata⁸⁶ la pubblicazione dei vari *dataset* a libero accesso. È in corso la mappatura dei dati del SITAR nel modello CIDOC Conceptual Reference Model (CRM), un'ontologia per il Web semantico specifica per i Beni Culturali, sviluppata principalmente in ambito museale, che è stata già introdotta nel capitolo 2. Una volta terminata, questa operazione permetterà di accedere ai dati relativi all'archeologia di Roma, estrarli e riutilizzarli anche attraverso strumenti terzi.

Allo stato attuale, comunque, è possibile esportare in maniera manuale i vari dati che compongono la piattaforma, in 17 formati complessivi (*raster* e vettoriali).

Il progetto SITAR, insieme al MAPPA Project, ha il grande merito di aver grandemente contribuito alla discussione sui dati aperti nella ricerca archeologica⁸⁷.

Dettaglio delle evidenze archeologiche d'età repubblicana del centro di Roma, <https://webais.archeositarproject.it/>.



tagli, ecc.

⁸⁶ <http://www.archeositarproject.it/?p=4920>.

⁸⁷ Si vedano, ad esempio, gli atti dell'ultimo convegno SITAR, fortemente impor-

iDAI.objects arachne

Questa lista potrebbe diventare molto lunga senza tuttavia avvicinarsi alla completezza. Ma non si vuole qui creare un catalogo omnicomprensivo delle risorse disponibili *online*, quanto fornire una panoramica dei possibili usi delle tecnologie di Rete nel campo della ricerca archeologica e più in generale antichistica. Sono stati esclusi centinaia – o forse migliaia – di progetti che pubblicano i propri dati *online*, non perché non siano importanti, ma perché la tipologia e filosofia di pubblicazione è già stata esaurientemente trattata in altri casi. Sono inoltre state escluse – volutamente – tutte le grandi collezioni museali che pure pubblicano tantissimi dati di rilevanza archeologica in Rete⁸⁸: le loro specifiche esigenze di inventariazione e catalogazione hanno determinato la scelta di strutture informative particolari, che necessiterebbero di una trattazione dedicata.

Le categorie degli oggetti descritti e disponibili in Arachne, <https://arachne.dainst.org/categories>.

Non può, però, mancare una menzione al iDAI.objects arachne o semplicemente Arachne, il *database* centrale e in gran parte *open access* dell'Istituto Archeologico Germanico (DAI) e dell'Istituto di Archeologia dell'Università di Colonia. Si tratta di uno strumento di ricerca su centinaia di

tanti sulle istanze open, Serlorenzi e Iovine, *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico*. Atti del IV Congresso di Studi SITAR (Roma, 14 ottobre 2015).

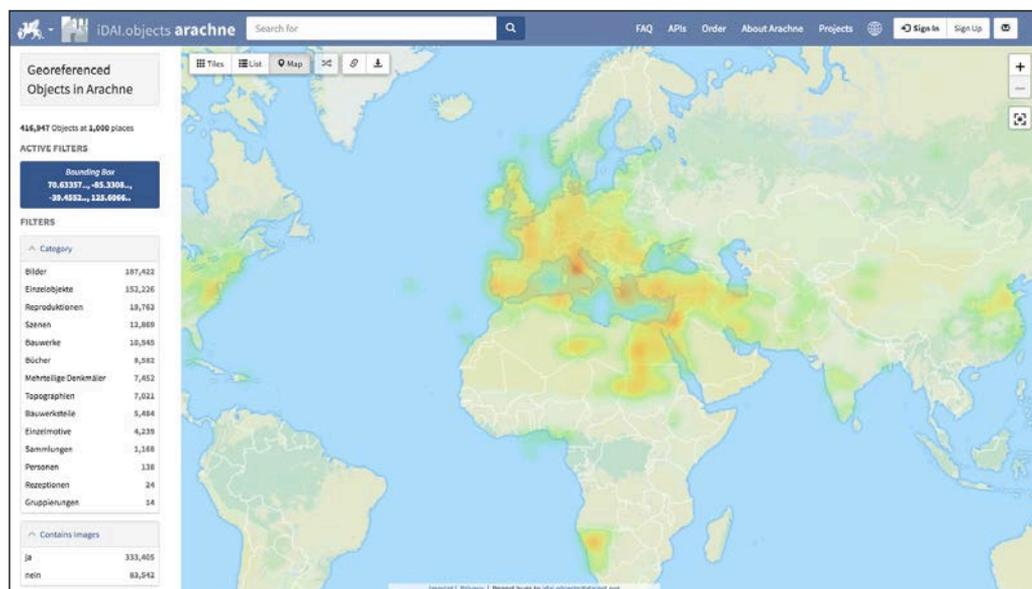
⁸⁸ Si è fatto cenno ad alcune collezioni museali e il possibile riutilizzo dei loro dati all'interno di progetti di ricerca quando si è trattato il progetto Kerameikos.

migliaia di *dataset* diversi, molti dei quali sono stati (e sono) in corso di digitalizzazione. I due concetti chiave che sostengono la struttura del sistema informativo sono:

1. tutti gli oggetti riguardanti il modo antico devono avere una minima base descrittiva e comparativa comune
2. ogni oggetto è legato a un determinato contesto.

Sono disponibili diverse interfacce di ricerca, geografiche e testuali, per classi di reperti, cronologia ecc. e il sistema informativo prevede varie categorie descrittive, concrete o astratte: oggetti, tipi, dati topografici (carte, mappe, rilievi), immagini, motivi, scene, edifici, riproduzioni, collezioni, libri, persone, iscrizioni, ecc.

Arachne prevede diversi strumenti anche per la condivisione automatizzata dei dati, tra i quali API ben documentati⁸⁹. Gran parte dei dati sono condivisi con licenze aperte e solo su alcuni *dataset* è necessario procedere all'autenticazione per poter accedere alle riproduzioni degli oggetti in alta risoluzione.



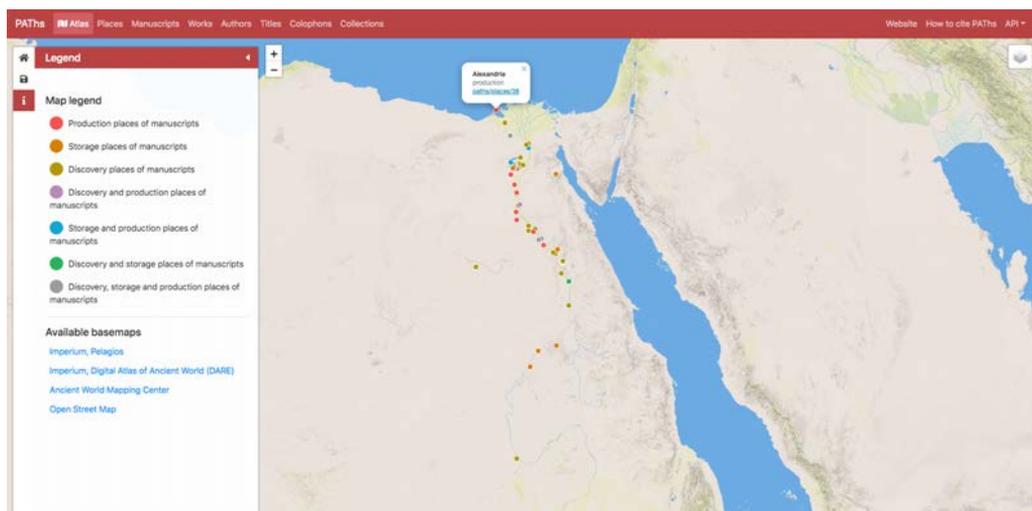
Interfaccia di ricerca geografica di Arachne, <https://arachne.dainst.org/map>.

⁸⁹ <https://github.com/dainst/arachnefrontend/blob/master/docs/rest-api-reference.md>.

PAThs. L'Atlante archeologico della letteratura Copta

“PAThs - Tracking Papyrus and Parchment Paths: An Archaeological Atlas of Coptic Literature. Literary Texts in their Geographical Context: Production, Copying, Usage, Dissemination and Storage” è un progetto finanziato dalla Commissione Europea per la Ricerca (ERC), diretto da Paola Buzi presso la Sapienza Università di Roma⁹⁰. Il progetto, tuttora in corso di svolgimento, è finalizzato alla creazione di una piattaforma digitale *online* dotata di interfacce geografiche, in grado di sintetizzare e rappresentare la geografia della produzione letteraria in lingua copta, nell'ambito dell'Egitto tardoantico e medievale. Si tratta di un progetto multidisciplinare, che unisce competenze filologiche, codicologiche, paleografiche, archeologiche, archeometriche e informatiche nel tentativo di esplorare un *corpus* di quasi 6200 manoscritti letterari e di ricostruire il loro contesto geografico e archeologico, andato perduto nella maggior parte dei casi.

Il progetto fa largo uso di molte delle risorse fin qui considerate. Il *ga-*



Schermata di una possibile visualizzazione geografica dell'Atlante archeologico della letteratura copta. Sulla base cartografica fornita dall' Ancient World Mapping center (le altre basi disponibili sono visualizzabili nell'elenco di destra) vengono visualizzati tutti i siti censiti dal progetto che hanno un rapporto con la produzione manoscritta. La tipologia del rapporto è indicata con diversi colori: i luoghi possono essere di produzione, di conservazione, di scoperta, oppure varie possibili combinazioni di queste.

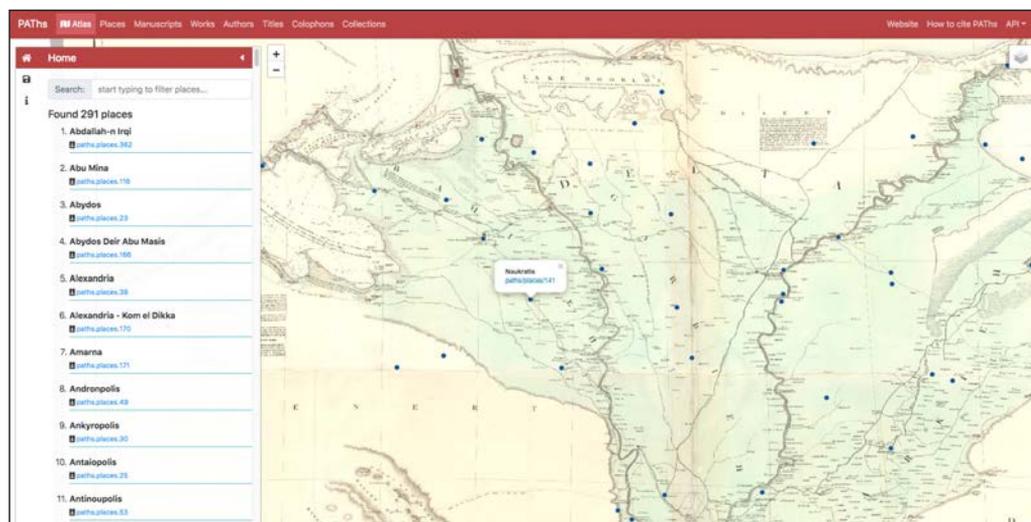
⁹⁰ Paola Buzi, Julian Bogdani, Nathan Carlig, Maria Chiara Giorda e Agostino Soldati, «“Tracking Papyrus and Parchment Paths”: A New International Project on Coptic Literature», *Rivista del Museo Egizio* 1 (2017), <https://doi.org/10.29353/rime.2017.656>.

zetteer dei luoghi antichi implementato da Pleiades⁹¹, ad esempio, è servito quale base di partenza per l'individuazione dei siti archeologici, la cui dettagliata descrizione è stata poi ulteriormente sviluppata e articolata seguendo le esigenze specifiche del progetto. Una metodologia analoga è stata seguita anche per la parte geografica, relativa al webGIS che funziona da interfaccia per l'accesso a gran parte dei dati.

Le basi cartografiche prodotte e liberamente distribuite da altri progetti, che si è già avuto modo di vedere in dettaglio nelle pagine precedenti (Ancient World Mapping Center, DARE, Pelagios, Google satellite, ecc.), sono state integrate nell'Atlante. In parallelo, sono stati creati nuovi temi cartografici, utili per l'inquadramento e l'analisi topografica, a partire da cartografia storica o da basi cartografiche più recenti edite negli ultimi ottant'anni. Questa preziosa documentazione grafica viene offerta come servizio WMS a libero accesso, quando la licenza dei dati di partenza ne permette la ri-distribuzione.

Alla pubblicazione della documentazione geografica (GIS) in vari standard e formati è dedicata una sezione importante di un portale parallelo, rivolto esclusivamente alla documentazione degli strumenti e delle metodologie utilizzate nel progetto, e grazie al quale è possibile anche accedere alle singole risorse⁹².

Particolare dell'Atlante PATHs. In blu i siti censiti dal progetto e sullo sfondo una carta del Delta del Nilo, redatta da A. Arrowsmith nel 1807 e frutto di vari rilievi degli anni precedenti, georiferito e reso disponibile attraverso il protocollo WMS



⁹¹ La banca dati di Pleiades è stata integrata con una seconda banca dati specialistica sul contesto egiziano, quella di Trismegistos, che comprende anche un'importante sezione dedicata ai luoghi, Herbert Verreth, *A survey of toponyms in Egypt in the Graeco-Roman period* (Trismegistos Online Publications, 2) (Köln, Leuven: Trismegistos Online Publications, 2013), <https://www.trismegistos.org/dl.php?id=15>.

⁹² Le risorse GIS sono disponibili all'indirizzo <https://docs.paths-erc.eu/gis/>.

Schermata della sezione dedicata ai luoghi (siti archeologici) del progetto PATHs.

L'archeologia, la geografia e, più in generale, le risorse GIS costituiscono solo una parte della struttura del sistema informativo digitale del progetto, incentrato su un'applicazione web basata su una banca dati relazionale⁹³. L'inserimento e la modifica dei dati sono limitati ai membri del progetto e ad altri studiosi che collaborano con PATHs, mentre la loro lettura non è soggetta a restrizioni. La banca dati principale, infine, è dotata di una interfaccia di comunicazione automatica (API) dedicata alle applicazioni dei progetti partner, per un accesso diretto alla totalità dei dati strutturati. In particolare, sono stati sviluppati strumenti dedicati per l'esportazione automatica dei dati contenuti nel *database* relazionale in formato geografico (GeoJSON), per l'utilizzo in programmi GIS tradizionali e in formato semantico (RDF), in particolare per la comunicazione con Pelagios.

PATHs Atlas Places Manuscripts Works Authors Titles Colophons Collections Website How to cite PATHs API

Introduction

Places

Show all Search Advanced search Saved queries

This section of the **PATHs** database contains a catalogue of Late Antique Egyptian sites active between the 3rd and 11th centuries.

A full description of the **Places** section is available at the [PATHs official documentation repository on GitHub](#). A [PDF version](#) is also available for offline access.

Places are firstly intended as physical spatial units that can be archaeologically observed and documented, i.e., they can be referred to a tangible area, with a specific topographical location and a distinctive configuration of material remains. Special attention is accorded to sites connected to the production, circulation, storage, preservation and discovery of manuscripts. Finally, major political, religious and cultural centres (for instance the episcopal sees), which are relevant for the reconstruction of the landscape of Byzantine Egypt, are also recorded, even though they no longer preserve archaeological traces. Likewise, toponyms attested in

- (*Das christlich-koptische Ägypten in arabischer Zeit*, Wiesbaden 1984-1992)
- Modern geographical location according to different scale degree (Region, Area)
- Ancient administrative affiliations (nomòs; province, according to the Diocletianic division of the country; episcopal see). For the Fayyum area, a specific reference to the *merides*-system is also recorded
- Site typology and its spatial and/or historical relationships with other places
- Dating of the Coptic phase within the chronological range of reference and the criteria upon which the attribution is based
- Brief description of the site summing up its main topographical and historical features as documented in the extant archaeological record
- Geo-localisation of the site as marked by its geographical coordinates
- Preceding and possibly subsequent phases to the Late Antique

⁹³ La banca dati è raggiungibile all'indirizzo: <https://db.bradypus.net/#/paths>. Prima della pubblicazione dell'atlante (<https://atlas.paths-erc.eu>), che offre accesso completo al dataset, l'applicazione relativa alla banca dati era aperta in sola lettura per gli studiosi che ne facevano richiesta. Sulla struttura della banca dati si veda Julian Bogdani, «The archaeological atlas of Coptic literature. A question of method», *Vicino Oriente* 21 (2017): 59–69.

PATHs rivolge una grande attenzione alla creazione e alla manutenzione di una rete molto articolata di connessioni con altri progetti e iniziative di ricerca che operano nell'ambito dell'Egitto romano, tardoantico e medievale. Questa rete viene materializzata nella creazione di elenchi molto puntuali di equivalenze tra identificativi di manoscritti, siti, opere, autori, collezioni, ecc. presenti nel *database* di PATHs e i loro corrispettivi in altri *dataset*. Attraverso queste connessioni è possibile ottenere dati aperti e collegati (LOD), i quali facilitano la navigazione tra *database* diversi e il recupero (anche automatizzato) di dati strutturati da ciascuno di questi. Una simile architettura costituisce un forte incentivo alla collaborazione tra le varie iniziative, che, grazie ai servizi del web semantico, possono integrare in maniera automatizzata i dati provenienti da ricerche precedenti e focalizzarsi maggiormente sugli aspetti specifici della propria ricerca.

Opere citate

Aliprandi, Simone. *Copyleft & opencontent: l'altra faccia del copyright*. Lodi: PrimaOra, 2005.

Aloia, Nicola, Ceri Binding, Sebastian Cuy, Martin Doerr, Bruno Fanini, Achille Felicetti, Johan Fihn, Dimitris Gavrilis, Guntram Geser, Hella Hollander, Carlo Meghini, Franco Niccolucci, Federico Nurra, Christos Papatheodorou, Julian Richards, Paola Ronzino, Roberto Scopiagno, Maria Theodoridou, Douglas Tudhope, Andreas Vlachidis e Holly Wright «Enabling European Archaeological Research: The ARIADNE E-Infrastructure». *Internet Archaeology*, n. 43 (6 marzo 2017). <https://doi.org/10.11141/ia.43.11>.

Alvoni, Giovanna, Ulrich Rausch, e Legoprint. *Scienze dell'antichità per via informatica: banche dati, Internet e risorse elettroniche nello studio dell'antichità classica*. Bologna: CLUEB, 2002.

Anderson, David G., Joshua Wells, Stephen Yerka, Kelsey Noack Myers, Robert Carl DeMuth e Thaddeus Geoffrey Bissett, a c.di. «Digital Index of North American Archaeology (DINAA)». *Open Context*. <http://opencontext.org/projects/416A274C-CF88-4471-3E31-93DB825E-9E4A> DOI: <https://doi.org/10.6078/M7N877Q0>. 2015-10-21.

Anichini, Francesca. *MAPPA, metodologie applicate alla predittività del potenziale archeologico. Vol. 1*. Roma: Ed. Nuova Cultura, 2012.

———. *MAPPA, metodologie applicate alla predittività del potenziale archeologico. Vol. 2*. Roma: Ed. Nuova Cultura, 2013.

Arizza, Marco, Valeria Boi, Alessandra Caravale, Augusto Palombini, e Alessandra Piergrossi, a c. di. *I dati archeologici. Accessibilità, proprietà, disseminazione (Roma, CNR, 23 maggio 2017)*. Firenze: All'Insegna del Giglio, 2018.

Bagnall, Roger S., e Sebastian Heath. «Roman Studies and Digital Resour-

- ces». *Journal of Roman Studies* 108 (2018): 171–89.
- Bagnara, Roberto, e Giancarlo Macchi Jánica, a c. di. *Open source, free software e open format nei processi di ricerca archeologici: atti del 1. Workshop (Grosseto, 8 maggio 2006)*. Firenze: Centro editoriale toscano, 2007.
- Basso, Patrizia, Alessandra Caravale, e Piergiovanna Grossi, a c. di. *AR-CHEOFOSS: free, libre and open source software e open format nei processi di ricerca archeologica: atti del IX Workshop, Verona, 19-20 giugno 2014*. Firenze: All'Insegna del Giglio, 2016.
- Beaulieu, Alan. *Learning SQL*. Sebastopol: O'Reilly Media, 2009.
- Benkler, Yochai, e Helen Nissenbaum. «Commons-based peer production and virtue». *Journal of political philosophy* 14, n. 4 (2006): 394–419. <https://doi.org/10.1111/j.1467-9760.2006.00235.x>.
- Berners-Lee, Tim. «Giant Global Graph». *DIG* (blog), 21 novembre 2007. <https://web.archive.org/web/20160713021037/http://dig.csail.mit.edu/breadcrumbs/node/215>.
- . «Information Management: A Proposal». Geneva: CERN, 1989. <http://cds.cern.ch/record/369245/files/dd-89-001.pdf>.
- Bezzi, Luca, Denis Francisci, Piergiovanna Grossi, e Damiano Lotto, a c. di. *Open source, free software e open format nei processi di ricerca archeologici: atti del 3. Workshop (Padova, 8-9 maggio 2008)*. Roma: Quasar, 2012.
- Bizer, Christian, Tom Heath, e Tim Berners-Lee. «Linked Data - The Story So Far». *International Journal on Semantic Web and Information Systems (IJSWIS)* 5, n. 3 (2009): 1–22. <https://doi.org/10.4018/jswis.2009081901>.
- Bogdani, Julian. «GIS per l'Archeologia». In *Groma 2. In profondità senza scavare*, a cura di Enrico Giorgi, 421–38. Bologna: BraDypUS, 2009.
- . «The archaeological atlas of Coptic literature. A question of method». *Vicino Oriente* 21 (2017): 59–69.
- Bogdani, Julian, e Eugenio De Mitri. «A Bibliography on the Application of GIS in Archaeology and Cultural Heritage». *Groma. Documenting Archaeology*, 21 dicembre 2017. <https://doi.org/10.12977/groma12>.
- Boi, Valeria, e Mirella Serlorenzi. «Archeologia preventiva, predittiva, po-

- tenziali archeologici. Una breve introduzione al panorama italiano». In *Proceedings of ArqueoFOSS: Free, Libre and Open Source Software e Open Format Nei Processi Di Ricerca Archeologica: VIII Edizione, Catania 2013*, a cura di Filippo Stanco e Giovanni Gallo. Oxford: Archaeopress, 2016.
- Bonacchi, Chiara. «The development of Public Archaeology in Italy: a review of recent efforts». *Public Archaeology* 12, n. 3 (2013): 211–16.
- Brogio, Gian Pietro. «Archeologia e istituzioni: statalismo o policentrismo?» *Archeologia medievale*, n. 24 (1997): 7–30.
- . «Archeologia pubblica in Italia: quale futuro». *Post-Classical Archaeologies* 2 (2012): 269–78.
- Burnard, Lou, C. M Sperberg-McQueen, e Fabio Ciotti. *Il manuale TEI Lite: introduzione alla codifica elettronica dei testi letterari*. Milano: Silvestre Bonnard, 2005.
- Buzi, Paola, Julian Bogdani, Nathan Carlig, Maria Chiara Giorda, e Agostino Soldati. «“Tracking Papyrus and Parchment Paths”: A New International Project on Coptic Literature». *Rivista del Museo Egizio* 1 (2017). <https://doi.org/10.29353/rime.2017.656>.
- Campana, Stefano, Gabriella Carpentiero, Marianna Cirillo, e Roberto Scopigno, a c. di. *CAA2015. Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. Oxford: Archaeopress, 2016.
- Campana, Stefano, e Maurizio Forte. *Remote sensing in archaeology: XI Ciclo di lezioni sulla ricerca applicata in archeologia, Certosa di Pontignano (Siena), 6-11 dicembre 1999*. Firenze: All’Insegna del Giglio, 2001.
- Campana, Stefano, e Maurizio Forte, a c. di. *From Space to Place: 2nd International Conference on Remote Sensing in Archaeology : Proceedings of the 2nd International Workshop, CNR, Rome, Italy, December 4-7, 2006*. Oxford: Archeopress, 2006.
- Cantone, Francesca, a c. di. *ArqueoFOSS: Open Source, Free Software e Open Format nei processi di ricerca archeologica. Atti del VI Workshop (Napoli, 9-10 giugno 2011)*. Pozzuoli: Naus Editoria, 2012.
- Carandini, Andrea. *Archeologia classica: vedere il tempo antico con gli occhi del 2000*. Torino: Einaudi, 2008.

- Carver, Geoff. «Archaeological Information Systems (AIS). Adapting GIS to Archaeological Contexts». *Archäologie Und Computer / 9*. 2004 9 (2004): 53.
- Cignoni, Paolo, Augusto Palombini, e Sofia Pescarin, a c. di. *Archeofoss: open source, free software e open format nei processi di ricerca archeologica. Atti del IV Workshop (Roma, 27-28 aprile 2009)*. Firenze: All'Insegna del Giglio, 2010.
- Codd, Edgar F. «A relational model of data for large shared data banks». *Communications of the ACM archive*. 13, n. 6 (1970): 377–87.
- . «Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks». IBM, 1969.
- Cohen, Daniel J, e Roy Rosenzweig. *Digital history: a guide to gathering, preserving, and presenting the past on the web*. Philadelphia: University of Pennsylvania Press, 2006.
- Conolly, James, e Mark W Lake. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press, 2006.
- Costa, Stefano, e Giovanni Pesce Luca, a c. di. *Open Source, Free Software e Open Format nei processi di ricerca archeologica Atti del II Workshop (Genova, 11 maggio 2007)*. London: Ubiquity Press]u[, 2013.
- Cristofani, Mario, e Riccardo Francovich. «Editoriale». *Archeologia e Calcolatori* 1 (1990): 1–7.
- Cristofori, Alessandro, Giovanni Geraci, e Carla Salvaterra. «Rassegna Strumenti informatici per l'Antichità Classica», 1998. <http://www.rassegna.unibo.it/>.
- Crockford, Douglas. «JavaScript: The World's Most Misunderstood Programming Language», 2011. <https://www.crockford.com/javascript/javascript.html>.
- Dal Maso, Cinzia, e Francesco Ripanti. *Archeostorie: manuale non convenzionale di archeologia vissuta*. Milano: Cisalpino, 2015.
- D'Andria, Francesco, a c. di. *Informatica e archeologia classica: atti del Convegno internazionale : Lecce, 12-13 maggio 1986*. Galatina (Lecce): Congedo editore, 1987.
- D'Andrea, Andrea. «Dati Digitali e Metodologia della Ricerca Archeologica». In *Atti del workshop L'integrazione dei dati archeologici digitali*

- *Esperienze e prospettive in Italia. Lecce, Italia, 1-2 Ottobre, 2015*, a cura di Paola Ronzino, 1634:10–17. Workshop Proceedings, 2016.
- . *Documentazione Archeologica, Standard e Trattamento Informatico*. Vol. 2. Budapest: Archaeolingua, 2006.
- D'Andrea, Andrea, Achille Felicetti, Sorin Hermon, Franco Niccolucci, Tommaso Zoppi e Roberto Bagnara, «I linguaggi standard del W3C e gli strumenti Open Source per la gestione dei dati archeologici», in *Open Source, Free Software e Open Format nei Processi di Ricerca Archeologica: Atti del I Workshop (Grosseto, 8 maggio 2006)*, a c. di Roberto Bagnara e Giancarlo Macchi Janica (Firenze: Centro editoriale toscano, 2007), 51–66.
- De Felice, Giuliano, e Maria Giuseppina Sibilano, a c. di. *Archeofoss: open source, free software e open format nei processi di ricerca archeologica. Atti del V Workshop (Foggia, 5-6 maggio 2010)*. Bari: Edipuglia, 2011.
- Di Giuseppe, Helga, e Elizabeth Fentress. «Open Archaeology: i Fasti e la pubblicazione online». In *Open Source Free Software e Open Format nei processi di ricerca archeologica*, a cura di Luca Bezzi, Stefano Costa, Denis Francisci, Piergiovanna Grossi, e Damiano Lotto, 85–93. Roma: Edizioni Quasar, 2012.
- Di Noia, Tommaso, Roberto De Virgilio, Eugenio Di Sciascio, e Francesco M. Donini. *Semantic web: tra ontologie e open data*. Santarcangelo di Romagna (RN): Maggioli editore, 2018.
- Dubbini, Nevio. «L'algoritmo MAPPA». In *MAPPA, metodologie applicate alla predittività del potenziale archeologico. Vol. 2*, a cura di Francesca Anichini, 101–13. Roma, 2013.
- Dunn, Stuart. «Review of ORBIS», *Journal of Digital Humanities* 1, n. 3 (2012), <http://journalofdigitalhumanities.org/1-3/review-of-orbis-project-by-stuart-dunn/>.
- Eide, Øyvind, Achille Felicetti, Christian-Emil Ore, Andrea D'Andrea, e Jon Holmen. «Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping». In *Open Digital Cultural Heritage System*, a cura di David B. Arnold, Franco Niccolucci, Daniel Pletinckx, e Luc Van Gool. Budapest: Archeolingua, 2011.
- Elliott, Tom, e Sean Gillies. «Digital Geography and Classics». *Digital*

- Humanities Quarterly* 003, n. 1 (26 febbraio 2009). <http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html>.
- Evans, Thomas L., e Patrick T. Daly. *Digital Archaeology Bridging Method and Theory*. Abingdon, Oxon; New York: Routledge, 2006.
- Ferrero, Marco. *SQL: la guida tascabile al linguaggio di interrogazione dei database*. Milano: Apogeo, 2004.
- Forte, Maurizio. *I Sistemi Informativi Geografici in archeologia*. Roma: MondoGis, 2002.
- Forte, Maurizio, e Stefano Campana. *Digital Methods and Remote Sensing in Archaeology Archaeology in the Age of Sensing*. Cham: Springer International Publishing, 2016.
- Forte, Maurizio, Stefano Campana, e Claudia Liuzza, a c. di. *Space, Time, Place Third International Conference on Remote Sensing in Archaeology: 17th-21st August 2009, Tiruchirappalli, Tamil Nadu, India*. Oxford: Archaeopress, 2010.
- Gabucci, Ada. *Informatica applicata all'archeologia*. Roma: Carocci, 2005.
- Gattiglia, Gabriele. «Think Big about Data: Archaeology and the Big Data Challenge». *Archäologische Informationen* 38 (2015): 12.
- Giorgi, Enrico, Federica Boschi, Michele Silani, e Julian Bogdani. «Documentare l'archeologia da Burnum (Sebenico, Croazia) a Suasa (Ancona): una tradizione rinnovata». *Archeologia e Calcolatori* 23 (2012): 261–82.
- Giorgi, Enrico, e Giuseppe Lepore, a c. di. *Archeologia nella valle del Cesano da Suasa a Santa Maria in Portuno: atti del Convegno per i venti anni di ricerche dell'Università di Bologna (Castelleone di Suasa, Corinaldo, San Lorenzo in Campo 18-19 dicembre 2008)*. Bologna: Ante Quem, 2012.
- Gottarelli, Antonio, a c. di. *Sistemi informativi e reti geografiche in archeologia: GIS-INTERNET. VII Ciclo di Lezioni sulla Ricerca applicata in Archeologia (Certosa di Pontignano 1995)*. Firenze: All'Insegna del Giglio, 1997.
- Gruber, Ethan, e Tyler Jo Smith. «Linked Open Greek Pottery». In *CAA2014: 21st Century Archaeology: Concepts, Methods and Tools: Proceedings of the 42nd Annual Conference on Computer Applications*

- and Quantitative Methods in Archaeology*, a cura di F Giligny. Oxford: Archaeopress, 2015.
- Harris, Edward C. *Principi di stratigrafia archeologica*. Roma: Carocci, 2017.
- Isaksen, Leif. «Archaeology and the Semantic Web». Doctoral Dissertation. University of Southampton. School of Electronics and Computer Science, 2011. <https://eprints.soton.ac.uk/206421/>.
- Isaksen, Leif, Rainer Simon, Elton T. E. Barker, e Pau Cañameres de Soto. «Pelagios and the Emerging Graph of Ancient World Data». In *ACM Web Science Conference, WebSci '14, Bloomington, IN, USA, June 23-26, 2014*, 197–201. Bloomington, Indiana: ACM, 2014. <https://doi.org/10.1145/2615569.2615693>.
- Kansa, Sarah Witcher e Eric C. Kansa, «Data Beyond the Archive in Digital Archaeology. Advances» in *Archaeological Practice*, 6(02), 2018, 89-92. DOI <http://dx.doi.org/10.1017/aap.2018.7>
- Kamermans, H., M. Leusen, e Ph. Verhagen. *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam, NETHERLANDS: Leiden University Press, 2009.
- Kleinrock, Leonard. *Communication Nets; Stochastic Message Flow and Delay*. New York: McGraw-Hill Book Company, 1964.
- Knappett, Carl. *Network Analysis in Archaeology. New Approaches to Regional Interaction*. Oxford: Oxford University Press, 2013.
- Law, Matt, e Morgan Colleen. «The archaeology of digital abandonment: online sustainability and archaeological aites». *Present Pasts* 6, n. 1 (2014): Art.2. <https://doi.org/10.5334/pp.58>.
- Lazzari, Marco. *Informatica umanistica*. Milano: Mc Graw-Hill, 2014.
- Leoni, Giorgia, e Mirella Serlorenzi, a c. di. *Il SITAR nella rete della ricerca italiana: verso la conoscenza archeologica condivisa: atti del III convegno (Roma, Museo Nazionale Romano, 23-24 maggio 2013)*. Firenze: Edizioni all'Insegna del Giglio, 2015.
- Licklider, Joseph Carl Robnett. «Man-Computer Symbiosis, IRE Transactions on Human Factors in Electronics». *IRE Transactions on Human Factors in Electronics*, HFE-1, n. 1 (marzo 1960).
- Lyon, Matthew, e Katie Hafner. *Where Wizards Stay Up Late: The Origins*

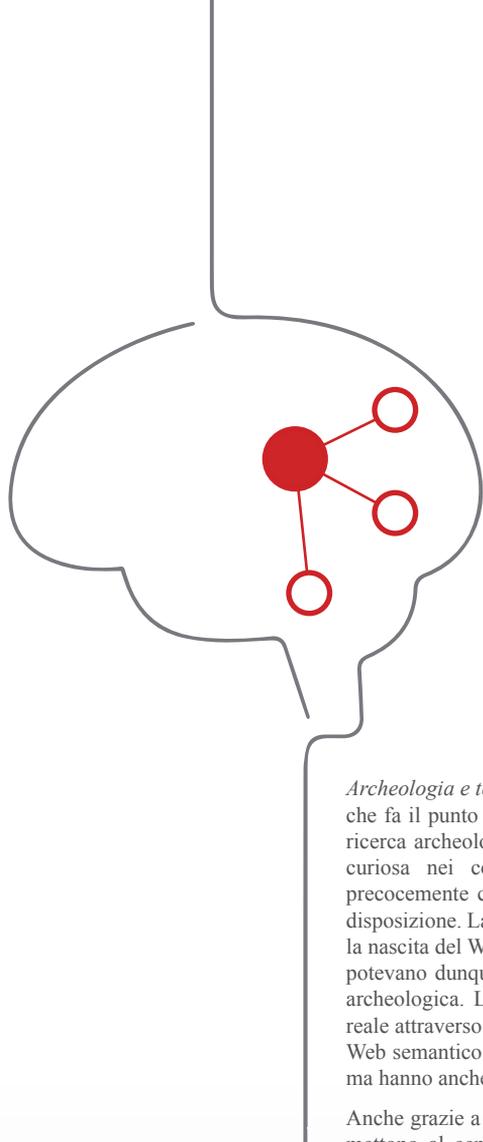
- Of The Internet*. Simon and Schuster, 1999.
- Mills, Barbara J. «Social Network Analysis in Archaeology». *Annual Review of Anthropology* 46, n. 1 (2017): 379–97. <https://doi.org/10.1146/annurev-anthro-102116-041423>.
- Moscato, Laura. «Lo Statuto di Anna e le origini del copyright nell’Inghilterra del Settecento». In *Fides humanitas ius : studi in onore di Luigi Labruna*, a cura di Luigi Labruna, Cosimo Cascione, e Carla Masi Doria. Napoli: Editoriale scientifica, 2007.
- Nield, Thomas. *Getting Started with SQL*. Sebastopol: O’Reilly Media, 2016.
- O’Reilly, Tim. *What Is Web 2.0*. Sebastopol: O’Reilly Media, 2009.
- Opitz, Rachel ,Marcello Mogetta e Nicola Terrenato, a c. di. «The Gabii Project». *Open Context*. <http://opencontext.org/projects/3585b372-8d2d-436c-9a4c-b5c10fce3ccd> DOI: <https://doi.org/10.6078/M7N-014GK>. 2017-06-04.
- Orlandi, Tito. *Informatica testuale. Teoria e prassi*. Roma-Bari: Laterza, 2010.
- Orton, David, Nerissa Russell, Katheryn Twiss, Louise Martin e Sheelagh Frame, a c. di. «Çatalhöyük Zooarchaeology». *Open Context*. <http://opencontext.org/projects/1B426F7C-99EC-4322-4069-E8DBD927C-CF1>. ARK (Archive): <https://n2t.net/ark:/28722/k2f47js2k>. 2013-02-13.
- Palazzolo, Maria Iolanda. «Per una storia del diritto d’autore in Italia: concetti, interessi, giurisprudenza». *Fabbrica del libro*, n. 2 (2011): 5–11.
- Parcak, Sarah H. *Satellite remote sensing for archaeology*. London; New York: Routledge, 2009.
- Pozzo, Riccardo. «Digital Humanities, Digital Cultural Heritage e l’istanza open». In *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico. Atti del IV Convegno di Studi SITAR (Roma, 14 ottobre 2015)*, a cura di Mirella Serlorenzi e Ilaria Iovine, 47–52. Firenze: All’Insegna del Giglio, 2017.
- Rabinowitz, Adam, Ryan Shaw, Sarah Buchanan e Patrick Golden, «Making sense of the ways we make sense of the past: The periodo project», *Bulletin of the Institute of Classical Studies* 59 (1 dicembre 2016): 42–55, <https://doi.org/10.1111/j.2041-5370.2016.12037.x>

- Rabinowitz, Adam, Ryan Shaw, e Patrick Golden, «Making up for Lost Time: Digital Epigraphy, Chronology, and the PeriodO Project», in *Crossing Experiences in Digital Epigraphy. From Practice to Discipline*, a c. di Annamaria De Santis e Irene Rossi (Berlin: De Gruyter, 2018), 202–15
- Radivojević, M., e J. Grujić. «Community structure of copper supply networks in the prehistoric Balkans: An independent evaluation of the archaeological record from the 7th to the 4th millennium BC». *Journal of Complex Networks* 6, n. 1 (24 luglio 2017): 106–24. <https://doi.org/10.1093/comnet/cnx013>.
- Renfrew, Colin, e Paul Bahn. *Archeologia. Teoria, metodi e pratica*. Tradotto da A. Vezzoli e F. Piccarreta. Bologna: Zanichelli, 2018.
- Richards, Julian D. «Digital Infrastructures for Archaeological Research: A European Perspective». *CSA Newsletter XXV*, n. 2 (2012). <http://csanet.org/newsletter/fall12/nlf1202.html>.
- Salus, Peter H. *Casting the Net: From ARPANET to Internet and Beyond*. Reading, Mass.: Addison-Wesley, 1995.
- Scianna, Andrea, e Benedetto Villa. «GIS Applications in Archaeology». *Archeologia e Calcolatori* 22 (2011): 337–63.
- Serlorenzi, Mirella, a c. di. *Archeofoss free, libre and open source software e open format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)*. Firenze: All'Insegna del Giglio, 2013.
- , a c. di. *SITAR Sistema Informativo Territoriale Archeologico di Roma. Atti del I Convegno (Roma, 26 ottobre 2010)*. Roma: Iuno Edizioni, 2011.
- Serlorenzi, Mirella, e Ilaria Iovine, a c. di. *Pensare in rete, pensare la rete per la ricerca, la tutela e la valorizzazione del patrimonio archeologico. Atti del IV Convegno di Studi SITAR (Roma, 14 ottobre 2015)*. Firenze: All'Insegna del Giglio, 2017.
- Serlorenzi, Mirella, e Ilaria Iovine, a c. di. *SITAR: Sistema Informativo Territoriale Archeologico di Roma: atti del II convegno, Roma 9 novembre 2011*. Roma: Iuno, 2011.
- Simon, Rainer, Elton Barker, e Leif Isaksen. «Exploring Pelagios: a visual browser for geo-tagged datasets». In *International Workshop on*

Supporting Users' Exploration of Digital Libraries, 23-27 September 2012, Paphos, Cyprus. Paphos, 2012.

- Simon, Rainer, Leif Isaksen, Elton Barker, e Pau Cañameres de Soto. «Pepi-pleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time». *Code4Lib* 31 (2016). <http://journal.code4lib.org/articles/11144>.
- Simon, Rainer, Leif Isaksen, Elton Barker, e Pau de Soto Cañameres. «The Pleiades Gazetteer and the Pelagios Project». In *Placing Names: Enriching and Integrating Gazetteers*, a cura di Merrick Lex Berman, Ruth Mostern, e Humphrey Southall, 97–109. Bloomington: Indiana University Press, 2016.
- Stanco, Filippo, e Giovanni Gallo, a c. di. *Proceedings of ArcheoFOSS: Free, Libre and Open Source Software e Open Format Nei Processi Di Ricerca Archeologica : VIII Edizione, Catania 2013*, 2016.
- Talbert, Richard J. A., e Roger S. Bagnall. *Barrington atlas of the Greek and Roman world*. Princeton, N.J. : Princeton University Press, 2000.
- Talbert, Richard John Alexander. *Rome's World: The Peutinger Map Reconsidered*. Cambridge: Cambridge University Press, 2014.
- Trabucco, Mario. «Pubblico ma non pubblico: prospettive normative sulla proprietà intellettuale dei dati archeologici». In *ArcheoFOSS: Open Source, Free Software e Open Format nei Processi di Ricerca Archeologica: Atti del IV Workshop (Roma, 27-28 aprile 2009)*, a cura di Paolo Cignoni, Augusto Palombini, e Sofia Pescarin, 65–70. Firenze: Edizioni all'Insegna del Giglio, 2010.
- Tuck, Anthony , a c. di. «Murlo». *Open Context*. <http://opencontext.org/projects/DF043419-F23B-41DA-7E4D-EE52AF22F92F>. DOI: <https://doi.org/10.6078/M77P8W98>. ARK (Archive): <https://n2t.net/ark:/28722/k2222wm10>. 2012-07-06
- University of Connecticut, Department of Anthropology, Arizona State University, e Archaeological Research Institute. *ArchNet: WWW virtual library of archaeology*. Tempe, Ariz.: Archaeological Research Institute, 1993. <https://web.archive.org/web/20150107082249/http://ari.asu.edu/archnet/>.
- Vannini, Guido. «Informatica per l'archeologia o archeologia per l'informatica?» In *Atti del I workshop Nazionale di Archeologia Computazionale, Napoli-Firenze 1999*, a cura di Andrea D'Andrea e Fabio Nico-

- lucci, *Archeologia e Calcolatori* 11 (2000): 311–15.
- Verreth, Herbert. *A survey of toponyms in Egypt in the Graeco-Roman period* (*Trismegistos Online Publications*, 2). Köln, Leuven: Trismegistos Online Publications, 2013. <https://www.trismegistos.org/dl.php?id=15>.
- Voorrips, Albertus. «Electronic Information Systems in archaeology. Some notes and comments». In *Methodological Trends and Future Perspectives in the Application of GIS in Archaeology*, a cura di Paola Moscati, 251–67. *Archeologia e Calcolatori* 9. Firenze: All’Insegna del Giglio, 1998.
- Wiseman, James, e Farouk El-Baz. *Remote Sensing in Archaeology*. New York: Springer, 2007.
- Wright, Holly, e Julian D. Richards. «Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures». *Journal of Field Archaeology* 43, n. S1 (2018): S60–S67. <https://doi.org/10.1080/00934690.2018.1511960>.



JULIAN
BOGDANI

ARCHEOLOGIA E TECNOLOGIE DI RETE

METODI STRUMENTI E RISORSE DIGITALI

Archeologia e tecnologie di rete. Metodi, strumenti e risorse digitali è un volume che fa il punto sulle applicazioni delle tecnologie di rete alla metodologia della ricerca archeologica. L'archeologia è sempre stata una disciplina estremamente curiosa nei confronti delle innovazioni tecnologiche, e ha sperimentato precocemente ciò che le scienze informatiche hanno messo di volta in volta a disposizione. La rapida e capillare diffusione delle tecnologie di rete, e soprattutto la nascita del World Wide Web alla fine degli anni Ottanta del secolo scorso, non potevano dunque non avere un importante impatto sui vari aspetti della ricerca archeologica. La pubblicazione dei dati, la possibilità di collaborare in tempo reale attraverso sistemi informativi condivisi e l'opportunità di ricerca offerta dal Web semantico e dai dati aperti e collegati (LOD) hanno aperto nuove frontiere, ma hanno anche riportato in luce vecchi e nuovi problemi.

Anche grazie a un'appendice che passa in rassegna diversi progetti e risorse che mettono al centro del proprio approccio proprio le tecnologie di rete, questo volume offre una panoramica aggiornata dello stato attuale della ricerca e mette in luce le criticità che verosimilmente ci si troverà ad affrontare nel prossimo futuro.

ISBN 978-88-98392-99-5



9 788898 392995

€ 20,00