

DALL'OPEN DATA ALLA PREDITTIVITÀ. NUOVI MODI DI FAR CRESCERE L'ARCHEOLOGIA ITALIANA

1. INTRODUZIONE

Il progetto MAPPA (metodologie applicate alla predittività del potenziale archeologico; ANICHINI *et al.* 2012; ANICHINI *et al.* 2013) è stato un progetto di ricerca sviluppato nel biennio 2011/2013 dall'Università di Pisa con un finanziamento della Regione Toscana e la collaborazione del Comune di Pisa, della Direzione Regionale per i Beni Culturali e Paesaggistici della Toscana e delle Soprintendenze per i Beni Archeologici della Toscana e per i Beni Architettonici, Paesaggistici, Artistici, Storici ed Etnoantropologici delle province di Pisa e Livorno. Ponendosi l'obiettivo di realizzare la carta del potenziale archeologico dell'area urbana di Pisa, il progetto puntava a creare uno strumento che offrisse un modo nuovo di interagire con il patrimonio archeologico ancora sepolto e non conosciuto. Se, infatti, le disposizioni di legge (D.Lgs. 263/2006 artt. 95, 96 e conseguente C.M. 10/2012) normano, nel caso delle opere pubbliche, le procedure per la valutazione dell'impatto archeologico dalla fase di progettazione preliminare fino alla fase esecutiva, la carta del potenziale archeologico offre un supporto alle scelte di pianificazione urbanistica e territoriale in una fase ancora precedente, agevolando le procedure di valutazione di impatto archeologico e consentendo di passare da un "generico" grado di potenziale archeologico ad uno "specifico" grado di rischio connesso alla natura del progetto da realizzare. In previsione che la normativa vigente sia estesa anche all'ambito delle opere private – com'è già nel resto d'Europa – la carta di potenziale archeologico è uno strumento in grado di favorire la cooperazione tra Soprintendenze e Amministrazioni locali, innescando un circolo virtuoso che rafforza l'azione di tutela e consente lo sviluppo dell'insediamento moderno, rendendo fin dall'inizio tutti consapevoli del potenziale e dunque del probabile rischio archeologico, nonché dell'entità – in termini di costi e tempi – delle indagini archeologiche che potranno essere richieste. In questo modo il Ministero, gli Enti locali, i committenti pubblici e privati e i professionisti incaricati dei progetti sono messi in condizione di operare con maggiore efficacia.

Da un punto di vista pratico, la carta del potenziale archeologico totale emerge dalla somma pesata delle singole carte di potenziale archeologico di periodo, prodotte dall'elaborazione algoritmica di una molteplicità di variabili, che vanno dalla tipologia dei ritrovamenti archeologici, al loro potenziale informativo, alle relazioni funzionali e topografiche che ciascuno di essi è in grado di sviluppare in senso sincronico e diacronico e in relazione

al contesto urbanistico e ambientale in cui si collocano. Le variabili inserite nel calcolo altro non sono che i dati archeologici, storici e geomorfologici che il progetto ha provveduto a disseminare secondo i paradigmi dell'open access e dell'open data.

L'open access è stato declinato col duplice binario della diffusione aperta di tutti i risultati scientifici prodotti nel corso del progetto sia in italiano, sia in inglese, scegliendo principalmente la cosiddetta gold road to open access, sostenuti in questo dalla casa editrice Nuova Cultura di Roma; e della diffusione di tutti i prodotti georeferenziati della ricerca tramite webGIS accessibile a tutti, senza accessi differenziati (ANICHINI, GATTIGLIA 2012; NOTI 2012).

L'open data è stato, invece, esplicitato attraverso la realizzazione del repository open data dell'archeologia italiana MOD Mappa Open Data (ANICHINI *et al.* 2013a; ANICHINI *et al.* 2013b), implementato fino ad oggi su base volontaria; il parere legale elaborato da Marco Ciurcina in proposito ha finalmente consentito di fare chiarezza sulla paternità intellettuale della documentazione archeologica e di conseguenza su a chi spettano legittimamente i diritti di pubblicazione di quella documentazione (CIURCINA 2013).

M.L.G.

2. OPEN DATA

L'apertura dei dati non è fine a se stessa, ma gli archeologi, pur esprimendo un generico atteggiamento positivo verso gli open data, non sembrano avere ancora una reale consapevolezza delle potenzialità dei dati aperti¹. Chi riutilizza i dati sembra farlo con la stessa modalità con la quale utilizza una qualsiasi altra fonte bibliografica, non sfruttando la possibilità di riproccarli, mentre forte viene avvertito il rischio di inattendibilità dei dati. È probabile che la comunità archeologica diffidi della propria capacità di valutare l'affidabilità di una fonte con la quale non ha ancora dimestichezza. In realtà qualunque dato può essere più o meno affidabile, indipendentemente dalle modalità e dalla rapidità di pubblicazione. Solitamente un dato archeologico pubblicato a stampa assume un certo grado di affidabilità a seconda della credibilità del suo autore e dell'eventuale processo di peer review a cui è stato sottoposto; con gli open data, invece, la credibilità è determinata dalla possibilità di verificarne l'origine, ripercorrendo a ritroso il processo interpretativo. La validazione passa, dunque, alla responsabilità di ogni singolo

¹ Cfr. in particolare ANICHINI 2013, con l'analisi dei dati del sondaggio promosso dal progetto MAPPA sugli open data e un recente post su OpenPompei (<http://www.openpompei.it/2014/01/29/gli-archeologi-e-la-disabitudine-agli-open-data/>; ultimo accesso 13 febbraio 2014).

archeologo, rafforzando la ricerca e la revisione tra pari in un'ottica aperta e condivisa (open peer review).

Questo origina una certa preoccupazione tra gli archeologi ancora poco avvezzi a una discussione aperta sui propri dati (e quindi sul proprio lavoro in termini di metodologia e risultati) e genera una certa ansia da prestazione per pubblicare dataset inappuntabili, senza tener conto che «un dataset perfetto è un dataset sospetto»².

Gli open data servono, ma sono ancora un mondo tutto da scoprire. Perché ciò avvenga, è necessario che gli archeologi vedano nella condivisione aperta delle informazioni il vero volano della conoscenza e dello sviluppo della ricerca, della tutela, di un nuovo e più ampio concetto di cultura del patrimonio archeologico e di un ritorno possibile in termini economici. In breve possiamo asserire che gli open data possono:

- portare all'innalzamento degli standard qualitativi di documentazione;
- migliorare il lavoro degli archeologi professionisti consentendo loro di reperire con facilità i dati necessari all'archeologia preventiva (VIArch);
- rendere più efficace l'azione di tutela, attraverso strategie che facciano sentire il cittadino partecipe della protezione del suo patrimonio storico/archeologico, e non il contrario, perché l'ampia condivisione, in una comunità proiettata verso l'idea di patrimonio storico archeologico come bene comune, diventa progressivamente garanzia di controllo, di qualità, di conservazione.

La condivisione dei dati è una delle poche strade oggi percorribili per far progredire le conoscenze, senza dover sostenere costi proibitivi, ottimizzando – in un'ottica di archeologia sostenibile – l'utilizzo di quella mole di dati che quotidianamente vengono prodotti e rimangono sottoimpiegati, anche grazie allo sviluppo in direzione dei Linked Open Data verso cui le ultime posizioni assunte dal MiBACT³ fanno sperare si stia andando. Infine gli open data archeologici sono un investimento ad alta redditività: come ha dimostrato una ricerca inglese (BEAGRIE, HOUGHTON 2013), negli ultimi 30 anni ogni sterlina investita in open data ne ha restituite 8,3!

F.A.

3. DAGLI OPEN DATA AL MODELLO MATEMATICO

Il progetto MAPPA ha proposto un modo per “riciclare” i dati attraverso l'applicazione di modelli matematici all'archeologia, modelli, cioè, che

² Carlo Citter nel corso di *Opening the Past* 2013.

³ Si veda *CulturaItalia* (<http://www.culturaItalia.it/>) o la stessa pagina Open Data del Ministero (<http://www.beniculturali.it/mibac/export/MiBAC/sito-MiBAC/MenuPrincipale/Trasparenza/Open-Data/index.html>).

permettono di ricreare le regole di un determinato fenomeno, piuttosto che di misurarlo in maniera statistica. Nello specifico, si è deciso di rielaborare l'algoritmo di PageRank (BINI *et al.* 2011; LANGVILLE, MEYER 2006; DUBBINI, GATTIGLIA 2013; DUBBINI 2013), partendo dall'assunto che i criteri utilizzati per l'attribuzione del potenziale archeologico siano molto simili a quelli adoperati dai motori di ricerca per l'assegnazione di importanza alle pagine web, in quanto entrambi fondati sulle relazioni reciproche. Punto di partenza è stata la definizione del potenziale archeologico, che in sintesi abbiamo inteso come la possibilità che nel sottosuolo sia preservata una stratificazione archeologica di maggiore o minore importanza, e dei parametri necessari a stimarlo (ANICHINI *et al.* 2011). Per consentire all'algoritmo di funzionare, i dati archeologici sono stati categorizzati (FABIANI, GATTIGLIA 2012) e parametrati in base al loro potenziale informativo sommando tre diverse componenti:

- la tipologia di informazione archeologica;
- la mobilità/inamovibilità del ritrovamento archeologico;
- la frequenza/rarità di ogni ritrovamento all'interno di un singolo periodo.

Per calcolare la tipologia di informazione è stata stilata una lista di 19 ambiti informativi per i quali ogni ritrovamento può fornire indicazioni: organizzazione del territorio, attività e tecniche produttive, tecniche costruttive, strategie e tecniche militari, igienico-sanitarie, sistemi idrici, viabilità/trasporti, commerci, coltivazione/allevamento, alimentazione, gestione dei rifiuti, antropologia, culto/pratiche funerarie, aspetti politico-istituzionali, indicatori sociali e di genere, tempo libero, costume, fauna/flora, ambiente (geomorfologia). Definiti questi ambiti, ad ogni categoria di ritrovamenti è stato assegnato un valore, ottenuto sommando il valore 1 ogni volta che la categoria forniva informazioni su quell'ambito informativo, altrimenti 0. A questo è stata aggiunta la mobilità/inamovibilità dei ritrovamenti, un parametro connesso alla consistenza/labilità delle diverse tracce archeologiche (1 per ritrovamenti archeologici realizzati in muratura; 0,5 per ritrovamenti archeologici che prevedono una muratura, ma che possono essere anche meno strutturati; 0 per ritrovamenti labili). L'assunto è che a resti archeologici più consistenti corrispondano valori di potenziale più alti, perché esiste una maggiore possibilità di identificare i resti stessi. Infine, sulla base della maggiore o minore frequenza con cui può essere rinvenuto in ogni singolo periodo storico, un medesimo tipo di ritrovamento assume un valore basato sul parametro frequenza/rarità (0 assente, 1 frequente, 3 raro, 6 unico), consentendo così di modulare il potenziale di ogni tipo di ritrovamento nelle diverse epoche storiche.

Quindi si è proceduto alla creazione algoritmica delle aree funzionali, vale a dire dei livelli di base dell'organizzazione spaziale e funzionale cittadina,

in base alla tipologia dei dati archeologici: area urbana, suburbana, rurale. Ogni dato archeologico, in questo modo, risulta correlato all'area funzionale in cui si trova. Nella complessa rete delle relazioni funzionali, sono state inoltre considerate le associazioni tra le tipologie di ritrovamenti che si instaurano sia sul piano sincronico all'interno delle aree funzionali, sia su quello diacronico (ANICHINI *et al.* 2013c).

Infine abbiamo distinto (DUBBINI, GATTIGLIA 2013) tra potenziale archeologico assoluto e potenziale archeologico stimato: il primo rappresenta il valore di potenziale delle aree per le quali vi erano dati disponibili; il secondo la valutazione del valore informativo della stratificazione archeologica espressa dall'algoritmo.

F.F.

4. L'ALGORITMO MAPPA

Per fornire una stima del potenziale archeologico "totale", l'algoritmo MAPPA lavora a partire dalla stima del potenziale di ogni singolo periodo archeologico. Per questo motivo il sottosuolo dell'area di studio è stato diviso in celle tridimensionali che formano dei layer corrispondenti ai sette periodi archeologici considerati: Protostoria, Periodo Etrusco (Fig. 1), Periodo Romano e Tardoantico (Fig. 2), Alto Medioevo (Fig. 3), Basso Medioevo (Fig. 4), Età Moderna, Età Contemporanea.

Tecnicamente è stata seguita la seguente procedura. A seconda della natura dei dati disponibili, i valori del potenziale assoluto sono stati suddivisi in diversi file di input per poter essere processati dall'algoritmo:

- dati con geolocalizzazione certa: dati con coordinate spaziali conosciute, compresi quelli di incerta datazione e attribuzione, che sono stati organizzati in una matrice $n1 \times n2$, nella quale l'elemento i, j rappresenta il potenziale assoluto della cella dovuto ai dati di geolocalizzazione certa collocati dentro alla cella stessa;
- dati di incerta geolocalizzazione: dati con coordinate spaziali non note con precisione, riferiti ad una localizzazione generica, che sono stati organizzati in una matrice $n1 \times n2$, nella quale l'elemento i, j rappresenta il potenziale assoluto della cella dovuto ai dati di geolocalizzazione incerta collocati dentro alla cella stessa;
- forme: nel caso dei dati con geolocalizzazione certa, è stato definito, dove possibile, un profilo ai ritrovamenti, ad esempio ricostruendo il contorno di una casa a partire da una serie di lacerti murari. Ad ogni forma è stato assegnato un valore numerico, compreso tra 1 e 6 (1 è il valore di precisione massima), che esprime il grado di precisione con il quale le dimensioni e l'orientamento della forma possono essere dedotti. Il valore del potenziale

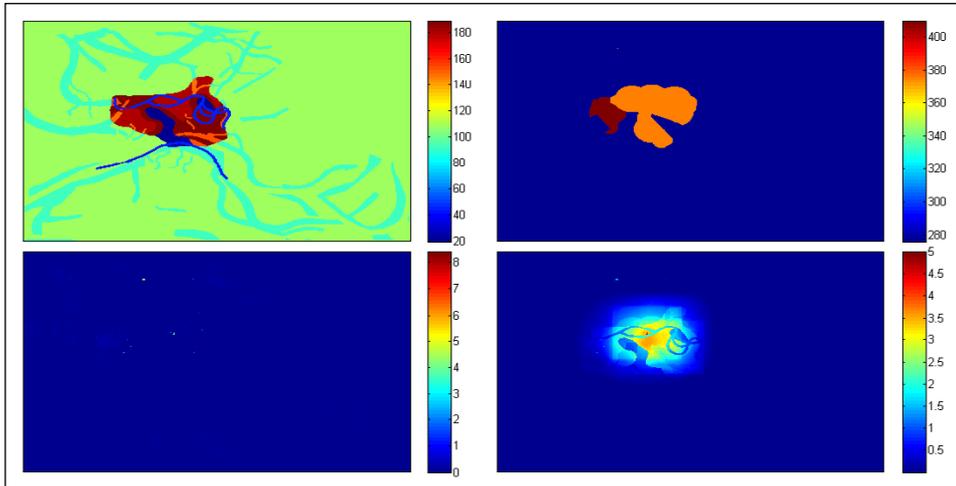


Fig. 1 – Da sinistra in alto, in senso orario: a) il dato geomorfologico: fiumi e area umida sono in blu, pianura e alto morfologico in rosso, il dato geomorfologico medio (dove i dati non sono disponibili) in verde, e i paleovalvi incerti in celeste dentro al verde, in arancione dentro al rosso; b) le aree funzionali: in arancione l'area urbana, in rosso l'area suburbana, in blu l'area rurale; c) il potenziale archeologico stimato; d) il potenziale assoluto. I valori dei parametri sono: speed_up = 0.8, rel = 0.5, yield = 0.5.

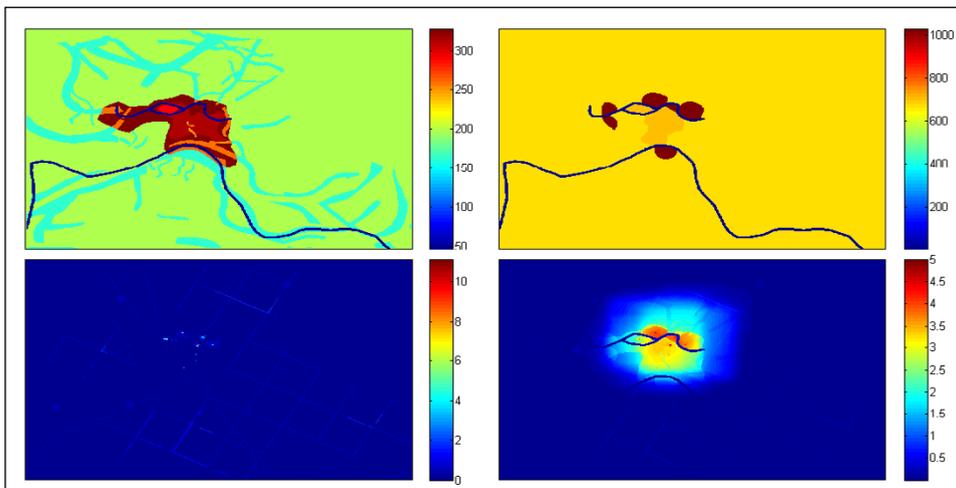


Fig. 2 – Da sinistra in alto, in senso orario: a) il dato geomorfologico: fiumi in blu, pianura e alto morfologico in rosso, il dato geomorfologico medio (dove i dati non sono disponibili) in verde, e i paleovalvi incerti in celeste dentro al verde, in arancione dentro al rosso; b) le aree funzionali: in giallo scuro l'area urbana, in rosso l'area suburbana, in giallo l'area rurale; c) il potenziale archeologico stimato; d) il potenziale assoluto. I valori dei parametri sono: speed_up = 0.75, rel = 0.5, yield = 0.5.

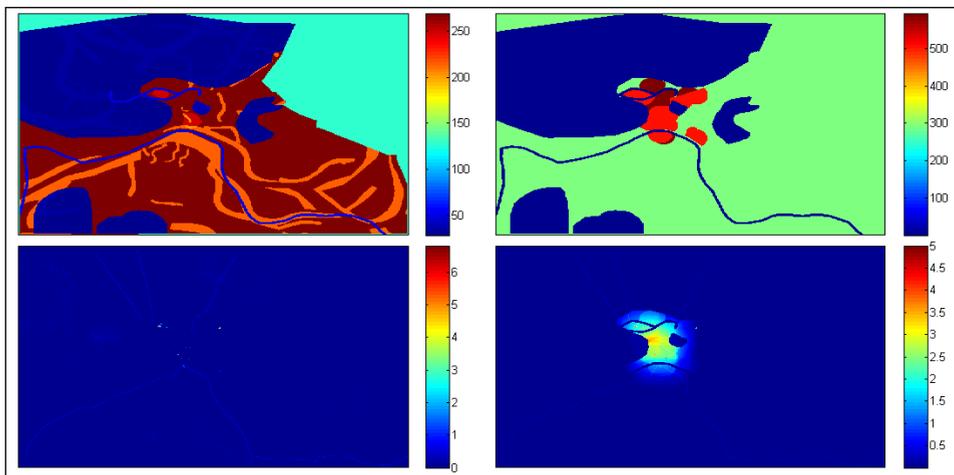


Fig. 3 – Da sinistra in alto, in senso orario: a) il dato geomorfologico: fiumi e area umida sono in blu, pianura e alto morfologico in rosso, il dato geomorfologico medio (dove i dati non sono disponibili) in verde, e i paleovalvi incerti in celeste dentro al verde, in arancione dentro al rosso; b) le aree funzionali: in arancione l'area urbana, in rosso l'area suburbana, in verde l'area rurale; c) il potenziale archeologico stimato; d) il potenziale assoluto. I valori dei parametri sono: speed_up = 0.7, rel = 0.5, yield = 0.5.

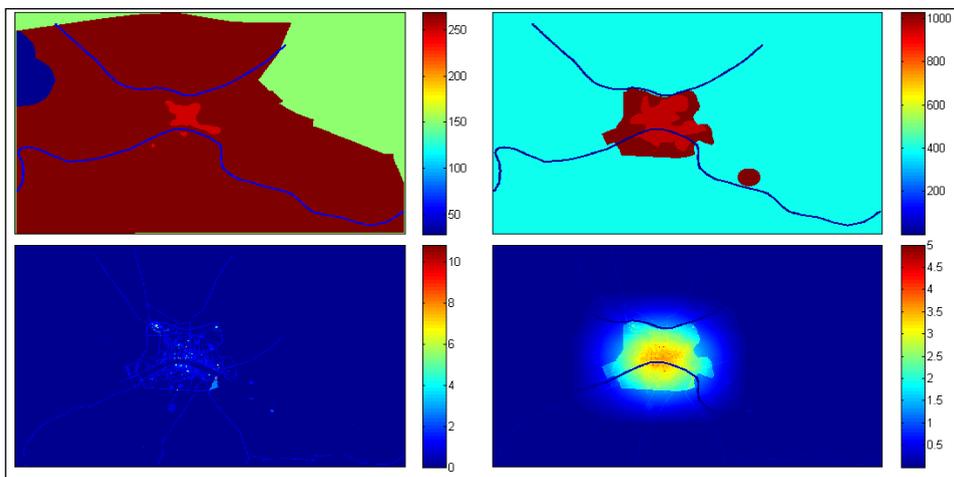


Fig. 4 – Da sinistra in alto, in senso orario: a) il dato geomorfologico: fiumi e area umida sono in blu, pianura e alto morfologico in rosso, il dato geomorfologico medio (dove i dati non sono disponibili) in verde, e le celle al di fuori dell'area di lavoro sono in verde; b) le aree funzionali: in rosso chiaro l'area urbana, in rosso scuro l'area suburbana, in celeste l'area rurale; c) il potenziale archeologico stimato; d) il potenziale assoluto. I valori dei parametri sono: speed_up = 0.7, rel = 0.5, yield = 0.5.

assoluto delle forme è stato calcolato dividendo il potenziale assoluto della categoria corrispondente alla forma per il valore di precisione della forma stessa. I dati delle forme sono stati organizzati in una matrice $n1 \times n2$, nella quale l'elemento i,j rappresenta il potenziale assoluto della cella dato dalle forme;

– dati geomorfologici: questi dati sono relativi alle diverse *facies* geomorfologiche individuate. Al dato geomorfologico è stato assegnato un valore di potenziale assoluto, sommando i valori dei potenziali assoluti delle categorie di ritrovamenti che possono essere presenti nelle diverse *facies*. Quando questo dato non era disponibile, è stato sostituito dalla media dei valori delle *facies* geomorfologiche. Nel caso di paleotracce riferibili a paleoalvei di datazione incerta, il valore geomorfologico è stato pari a $1/n$ rispetto al corrispondente valore del fiume, visto che l'incertezza sulla collocazione cronologica è da dividere tra gli n diversi periodi archeologici nei quali poteva rientrare. I dati geomorfologici sono stati organizzati in una matrice $n1 \times n2$, nella quale l'elemento i,j rappresenta il potenziale assoluto della cella determinato dai dati geomorfologici relativi alla cella stessa.

Oltre a questi dati di input, l'algoritmo MAPPA definisce e utilizza le aree funzionali (cfr. § 3) e altri due file di input che quantificano le relazioni sincroniche e diacroniche tra i ritrovamenti e che permettono di aumentare il potere diffusivo del potenziale archeologico, sulla base della probabile presenza di ulteriori ritrovamenti nelle vicinanze di alcune determinate (ma ancora circoscritte) categorie di ritrovamenti:

– relazioni sincroniche: ad ogni categoria di ritrovamenti, tenendo in considerazione l'area funzionale nella quale si trova, è stato associato l'insieme delle categorie di ritrovamenti che più probabilmente⁴ si trovano a una distanza minore di 50 m da questa. Questo file di input contiene, per ogni cella, la somma dei potenziali assoluti delle categorie associate ai ritrovamenti in quella cella;

– relazioni diacroniche: ad ogni categoria di ritrovamenti, tenendo conto dell'area funzionale nella quale si trova, è stato associato l'insieme delle categorie di ritrovamenti che più probabilmente si collocano ad una distanza minore di 50 m, nel periodo cronologicamente precedente o successivo. Questo file di input contiene, per ogni cella, la somma dei valori di potenziale assoluto delle categorie associate ai ritrovamenti localizzati nella cella stessa.

La scelta delle dimensioni delle celle deriva da alcune considerazioni: i dati archeologici sono localizzati con precisione (ovviamente quando la loro geolocalizzazione è certa), mentre i dati geomorfologici non possono essere forniti con una precisione maggiore di 10×10 m, a causa del numero di punti

⁴ Con l'espressione "più probabilmente" intendiamo relative almeno al 75% delle attestazioni.

quotati disponibili per la realizzazione di ciascun DEM di periodo; esiste un *trade-off* tra la dimensione delle celle e il numero totale delle celle che coprono l'area di lavoro: più piccola è la dimensione delle celle, più grande è il numero totale delle celle che coprono l'area e dunque più grande il numero delle celle per le quali il potenziale archeologico deve essere stimato⁵. Per risolvere questo problema abbiamo cercato (numericamente) il massimo di una funzione che rappresenta la differenza tra la quantità di informazione che fornisce la presenza, non la natura, dei dati, e l'area delle celle divisa per il numero totale delle celle, attraverso il calcolo dell'entropia di Shannon (DUBBINI, GATTIGLIA 2013). Questo ha indicato come dimensione "ottimale" quella delle celle con lato tra 10 e 14 m: abbiamo quindi scelto il valore 10×10 m.

Matematicamente parlando, il modello ha bisogno di una matrice S dei pesi e di un vettore D dei valori di potenziale assoluto, che rappresenta l'informazione disponibile pregressa. L'algoritmo è stato applicato a ogni periodo separatamente e i risultati sono stati poi "sommati", per ottenere una visione unica. L'implementazione pratica del modello è stata eseguita mediante il software Matlab. Il vettore D , che rappresenta i dati disponibili, è una matrice di dimensioni $n1 \times n2$, messa nella forma (*reshape* in Matlab) di un vettore di lunghezza n , ed è ottenuta dalla somma delle matrici dei potenziali assoluti relativi ai dati di localizzazione certa, incerta e delle forme. La matrice dei pesi S è una matrice $n \times n$ di cui l'elemento i,j rappresenta il peso (il valore) del link tra la cella j e la cella i . Più precisamente, la matrice S è calcolata nella seguente maniera:

– ogni cella recante almeno un ritrovamento distribuisce la sua importanza entro una maschera quadrata di celle, centrata nella cella stessa. Il lato di questo quadrato è proporzionale (di costante 1/5) al valore del potenziale assoluto dell'area funzionale alla quale la cella appartiene. Abbiamo usato il valore dell'area funzionale in questa maniera, perché il potenziale assoluto delle aree funzionali è stato espresso proprio in funzione dei possibili ritrovamenti che possono verificarsi in quella determinata area: quindi, in aree funzionali di maggior valore è giustificato aumentare la forza di diffusione del potenziale e viceversa;

– una volta costruita la maschera, il peso totale (ovvero la somma dei pesi distribuiti da ogni cella) dentro la maschera è dato dal valore dell'area funzionale più un valore proporzionale alla somma delle relazioni sincroniche e diacroniche. Abbiamo usato questo valore per definire il peso totale, perché la "quantità di potenziale" che una cella diffonde è influenzata dalla probabilità di avere ritrovamenti di alto o basso valore potenziale nelle vicinanze: i

⁵ Nel caso limite in cui volessimo stimare il potenziale archeologico di ogni punto dell'area di lavoro con precisione arbitraria, avremmo un numero finito di dati di input, ma dovremmo stimare il potenziale di un numero infinito di punti.

valori di potenziale assoluto delle aree funzionali e delle relazioni sincroniche e diacroniche esprimono proprio questa possibilità;

– la distribuzione dei pesi dentro alla maschera attorno a ogni cella è data dalla distribuzione uniforme, pesata (in parti eguali) per i valori geomorfologici e per i valori delle aree funzionali nella maschera. Questo perché il dato geomorfologico costituisce una sorta di influenza “di base” sulla diffusione del potenziale archeologico e perché i valori delle aree funzionali – in quanto proporzionali al potenziale dei ritrovamenti possibili in quell’area – possono essere usati per pesare la diffusione del potenziale.

L’algoritmo consiste nell’applicazione di una procedura base, applicata ripetutamente:

– il vettore D che rappresenta i dati disponibili e la matrice S dei pesi sono generati come descritto sopra;

– sono eseguite le seguenti iterazioni:

for $i = 1, \dots, 1000$

$$A = S \cdot x + \begin{bmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{bmatrix} \cdot x;$$

$$A = (1 - \text{yield}) \cdot A + \text{yield} \cdot x;$$

$$u = [1 \ 1 \ \dots \ 1];$$

$$y = \text{rel} \cdot (A) + (1 - \text{rel}) \cdot [D \cdot (u^T \cdot x)];$$

$$y = \frac{y}{\sum_{i=1}^n y^i}$$

$$x = y;$$

end

$$D = \text{speed_up} \cdot x + (1 - \text{speed_up}) \cdot D$$

In queste formule x è un vettore colonna stocastico (cioè la somma delle sue componenti è uguale a 1) di dimensione n , usato come condizione iniziale per l’applicazione delle iterazioni descritte dal ciclo *for* (il risultato di queste iterazioni è indipendente alla condizione iniziale). I parametri utilizzati nell’algoritmo sono:

– $\text{maxit} \in \mathbb{N}$ è il numero di volte in cui i passi 1. e 2. sono eseguiti. Ogni volta che i passi 1. e 2. sono eseguiti, l’algoritmo fa uno step di predizione del potenziale archeologico. Dopo ogni step il risultato è preso come il nuovo punto di partenza per lo step successivo. Più maxit è grande, più la stima del potenziale si distacca dai dati originali;

- $speed_up \in [0,1]$ è il peso che esprime la parte del nuovo potenziale dato dall'applicazione dei passi 1. e 2. e la parte data dal potenziale del precedente step. Più $speed_up$ si avvicina a 1, meno il risultato ad ogni passo è legato ai dati del precedente step;
- $rel \in [0,1]$ è il parametro che regola quanto teniamo in considerazione il potenziale dato dalle relazioni (ovvero dalla matrice dei pesi S), rispetto a quanto teniamo in considerazione il potenziale dato dai valori del potenziale assoluto. Più rel si avvicina a 1, meno il valore del potenziale assoluto è tenuto in considerazione, e più la matrice dei pesi S è preminente nella stima del potenziale;
- $yield \in [0,1]$ è la quantità di potenziale che ogni cella tiene per sé, rispetto al resto, che viene distribuito alle altre celle sulla base della matrice dei pesi S . Più $yield$ è vicino a 1, più ogni cella mantiene il potenziale per se stessa non diffondendolo.

La validazione dell'algoritmo (DUBBINI, GATTIGLIA 2013) è stata effettuata sul potenziale archeologico totale, in base a 14 nuovi carotaggi (CAPITANI *et al.* 2013), e ha evidenziato come la proporzione dei livelli di potenziale esattamente stimati sia $9/13 = 69.2\%$, mentre il massimo errore sia 1, con un errore medio di 0.3077. Se non facciamo riferimento ai livelli di potenziale, il massimo errore è 1.5737, mentre l'errore medio è 0.5495, e la varianza dell'errore è 0.2421. È utile notare, inoltre, che quando la stima è in disaccordo con il potenziale assoluto dei nuovi carotaggi, questa è sempre più grande. Questo è in accordo con il fatto che l'area del carotaggio è di circa 35 cm², mentre la stima del potenziale riguarda una cella di 100 m²: un carotaggio è meno informativo, perché non occupa l'intera cella.

N.D.

5. LA CARTA DI POTENZIALE ARCHEOLOGICO

La carta del potenziale totale è un prodotto digitale pubblicato su MappaGIS che si compone:

- della carta di potenziale archeologico totale;
- delle carte di potenziale archeologico di periodo;
- della carta dei gradi di potenziale archeologico, ovvero della visualizzazione sintetica della carta del potenziale totale suddivisa in 5 livelli.

Nella carta di potenziale archeologico totale (Fig. 5), ognuna delle celle di 10×10 m di lato, in cui è suddivisa l'area possiede un proprio specifico valore compreso tra 1 e 5 espresso come valori continui, mentre la carta dei gradi di potenziale archeologico propone una lettura semplificata, statica, evidenziando in maniera immediata le aree a maggiore o minore potenziale, risultando più idonea alla pianificazione urbanistica e territoriale su ampia scala (ANICHINI *et al.* 2013d).



Fig. 5 – Carta di potenziale archeologico totale dell'area urbana e suburbana di Pisa.

6. CONCLUSIONI

Facendo un bilancio complessivo (DUBBINI, GATTIGLIA 2013), il PageRank si è rivelato un modello matematico funzionale i cui punti di forza risultano:

- la capacità di assegnare il potenziale archeologico alle celle, classificandole sulla base delle loro interazioni;
- l'utilizzo congiunto di dataset di diversa natura (archeologica, geomorfologica, storica, etc.) e differente cronologia per produrre un unico output: la stima del potenziale archeologico;
- la formalizzazione algoritmica dell'intero processo. Il processo interpretativo archeologico entra in gioco, ma in un modo tale che, una volta cambiati i parametri in accordo con le diverse interpretazioni, la procedura algoritmica funzioni esattamente nella stessa maniera;
- il trattamento dei dati come se fossero una rete complessa;
- la potenza dell'elaborazione matematica come strumento di ricerca, oltre che di tutela, soprattutto per la sua capacità di predire l'organizzazione degli spazi urbani.

Facciamo tuttavia notare alcuni possibili miglioramenti per gli studi futuri:

- il modello PageRank sembra lavorare meglio con dati che si concentrano in un'area piuttosto che con dati di tipo policentrico. Questo è dovuto anche al modo in cui sono definite le aree funzionali: nel caso monocentrico bisogna tracciare i confini delle aree funzionali, mentre nel caso policentrico bisogna predirne la collocazione in più punti dello spazio. Non abbiamo provato a fare tale inferenza e non è detto che il modello PageRank sia il metodo adatto

per individuare la posizione di nuove aree che siano indipendenti dalle altre e che, quindi, debba essere integrato con soluzioni differenti, anche per poter esportare il modello su situazioni territoriali di maggiore ampiezza;

– è necessario valutare la bontà della stima fornita dal modello PageRank. Nel corso del progetto abbiamo basato il test sui risultati di 14 nuovi carotaggi, ma non abbiamo studiato un metodo generale che permetta di decidere quando e se un modello PageRank possa essere appropriato;

– abbiamo lavorato soltanto su dati positivi, mentre sarebbe importante avere anche i dati di tipo negativo: conoscere l'assenza di una determinata stratificazione archeologica sarebbe utile tanto quanto la conoscenza della presenza, perché entrambe queste componenti contribuiscono a fissare i valori iniziali di potenziale archeologico alla stessa maniera. Si potrebbe, così, sviluppare un modello per diffondere gli zeri del potenziale, complementare al modello che diffonde il potenziale.

Infine, l'utilizzo di un modello matematico ha reso evidente il fatto che, intrapresa la strada della digitalizzazione, è ormai giunto il tempo per l'archeologia di avviarsi in quella della "datificazione", per potersi finalmente confrontare con i Big Data. Sebbene questo richieda un cambio di paradigma da *hypothesis-driven research* a *data-driven research*, dalla ricerca dell'accuratezza all'accettazione dell'imprecisione dei dati, dal concetto di causalità a quello di correlazione, si tratta di un salto non impossibile per una disciplina che ha già a che fare con la complessità, la frammentarietà, la varietà (per fonti, scala e formato) dei dati. Proiettare l'archeologia nel campo dei Big Data la renderà maggiormente sostenibile e aprirà prospettive di ricerca inimmaginabili come ad esempio l'analisi immediata delle distribuzioni dei reperti e dei flussi commerciali del passato.

G.G.

FRANCESCA ANICHINI, NEVIO DUBBINI, FABIO FABIANI, GABRIELE GATTIGLIA,
MARIA LETIZIA GUALANDI
Progetto MAPPA – Università di Pisa

BIBLIOGRAFIA

- ANICHINI F. 2013, *MAPPA survey: gli Open Data nell'archeologia italiana*, in ANICHINI et al. 2013, 121-132 (doi: 10.4458/0917-09).
- ANICHINI F., BINI M., FABIANI F., GATTIGLIA G., GIACOMELLI S., GUALANDI M.L., PAPPALARDO M., SARTI G. 2011, *Definition of the parameters of the Archaeological Potential of an urban area*, «MapPapers», 1, 47-49 (doi:10.4456/MAPPA.2011.02).
- ANICHINI F., FABIANI F., GATTIGLIA G., GUALANDI M.L. 2012, *MAPPA. Metodologie Applicate alla Predittività del Potenziale Archeologico*, vol. I, Roma, Edizioni Nuova Cultura.
- ANICHINI F., GATTIGLIA G. 2012, *A qualcuno piace 'webgis'*. Indicazioni pratiche per un utilizzo archeologico consapevole, in ANICHINI et al. 2012, 73-86 (doi: 10.4458/8219-08).

- ANICHINI F., DUBBINI N., FABIANI F., GATTIGLIA G., GUALANDI M.L. 2013, *MAPPA. Metodologie Applicate alla Predittività del Potenziale Archeologico*, vol. II, Roma, Edizioni Nuova Cultura.
- ANICHINI F., CIURCINA M., NOTI V. 2013a, *Il MOD: l'archivio Open Data dell'archeologia italiana*, in ANICHINI et al. 2013, 121-132 (doi: 10.4458/0917-09).
- ANICHINI F., GATTIGLIA G., GUALANDI M.L., NOTI V. 2013b, *MOD (MAPPA Open Data). Conservare, disseminare, collaborare: un archivio open data per l'archeologia italiana*, «Archeologia e Calcolatori», Suppl. 4, 2013, 45-52.
- ANICHINI F., FABIANI F., GATTIGLIA G., GHIZZANI MARCÌA F. 2013c, *Questioni di pesi: valori, parametri, relazioni per il calcolo del potenziale archeologico*, in ANICHINI et al. 2013, 89-100 (doi: 10.4458/0917-05).
- ANICHINI F., FABIANI F., DUBBINI N., GATTIGLIA G., GHIZZANI MARCÌA F. GUALANDI M.L. 2013d, *La carta di potenziale archeologico*, in ANICHINI et al. 2013, 115-120 (doi: 10.4458/0917-09).
- BEAGRIE N., HOUGHTON J. 2013, *The Value and Impact of the Archaeology Data Service. A Study and Methods for Enhancing Sustainability*. Final Report (http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf).
- BINI D., DUBBINI N., STEFFÈ S. 2011, *Mathematical models for the determination of archaeological potential*, «MapPapers», 1, 77-85 (doi:10.4456/MAPPA.2011.08).
- CAPITANI M., FABIANI F., SCIUTO C., TARANTINO G. 2013, *(MAPPAProject) Pisa. Carotaggi di verifica del potenziale archeologico (Relazione)*, Pisa, MOD (doi: 10.4456/MAPPA.2013.21).
- CIURCINA 2013, *Parere legale sul portale Mappa Open Data*, «MapPapers», 4, 87-106 (doi: 10.4456/MAPPA.2013.76).
- DUBBINI N. 2013, *L'algoritmo MAPPA*, in ANICHINI et al. 2013, 101-114 (doi: 10.4458/0917-06).
- DUBBINI N., GATTIGLIA G. 2013, *A PageRank based predictive model for the estimation of the archaeological potential of an urban area*, in A.C. ADDISON, G. GUIDI, L. DE LUCA, S. PESCARIN (eds.), *Proceedings of the 2013 Digital Heritage International Congress*, New York, IEEE, 571-578.
- FABIANI F., GATTIGLIA G. 2012, *La struttura informatica di archiviazione*, in ANICHINI et al. 2012, 41-71 (doi: 10.4458/8219-07).
- LANGVILLE A.N., MEYER C.D. 2006, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton, Princeton University Press.
- NOTI V. 2012, *Il webGIS del progetto MAPPA: architettura del sistema e scenari futuri*, in ANICHINI F et al. 2012, 87-96 (doi: 10.4458/8219-09).

ABSTRACT

Data and the possibility to circulate and spread them are the key infrastructure of archaeology. Part of the (interpretative) information underlying data is connected to the know-how of each single researcher. Data, instead, are firm points. In this paper, we describe how open data can be used for determining the archaeological potential. We used a sizable number of datasets, in order to consider the problem of estimation of archaeological potential in all of its aspects: archaeological, historical, and geomorphological data. As the identification of relations among finds is a key issue in archaeological interpretation process, we applied a modified version of the PageRank algorithm. The procedure included a categorization of archaeological data, the assignment of initial values of potential to the available data, the creation of geomorphological maps, the definition of functional areas (i.e. the levels of spatial and functional organization: urban, suburban and rural areas), and the application of the PageRank based algorithm. The model has been applied on the urban area of Pisa, and tested through the data of 14 new cores. The map of archaeological potential consists of the composition of the 7 layers, one for each archaeological period under consideration. The results, including the archaeological potential map, are to be considered as the first steps towards an automatic, formally definable, and repeatable approach to the computation of archaeological potential.