

LIBER: UN PROGETTO DI DIGITALIZZAZIONE DEI TESTI IN SCRITTURA LINEARE B

1. INTRODUZIONE

In questa nota sono presentati i risultati preliminari di un progetto di digitalizzazione dei testi in scrittura lineare B in corso di realizzazione presso l'Istituto di Studi sul Mediterraneo Antico del Consiglio Nazionale delle Ricerche (ISMA-CNR). Tale progetto, che porta il nome di LiBER, acronimo di *Linear B Electronic Resources*, è stato avviato nel 2011 presso l'ex Istituto di Studi sulle Civiltà dell'Egeo e del Vicino Oriente (ICEVO-CNR) ed è stato successivamente sviluppato grazie al supporto logistico e alle infrastrutture informatiche dell'ISMA e dell'Area della Ricerca di Roma 1 del CNR. Dal novembre 2013 è online un sito web appositamente dedicato (<http://liber.isma.cnr.it/>), che permette di accedere alle risorse disponibili.

2. SCOPO DEL PROGETTO

Scopo generale di LiBER è quello di produrre un'edizione elettronica aggiornata dei testi in lineare B, scrittura utilizzata a Creta e nella Grecia micenea tra il XV e il XIII sec. a.C. per redigere documenti d'archivio di natura contabile. Più precisamente, lo scopo è quello di realizzare un database di testi interrogabile online attraverso un motore di ricerca appositamente concepito per il trattamento delle scritture logo-sillabiche – quale è la lineare B – e capace di filtrare e combinare i documenti sulla base di criteri testuali, archeologici, paleografici e topografici, con la possibilità di presentare i risultati come elenchi di testi, come indici di parole e logogrammi o come mappe dei “findspot” dei testi.

Come si può vedere dalle pagine del progetto MNAMON della Scuola Normale Superiore di Pisa dedicate alla lineare B¹, le risorse online dedicate a questa scrittura sono relativamente scarse e pochissimi sono al momento gli approcci di tipo informatico ai testi. Uniche eccezioni degne di nota sono gli *Índices Generales de la Lineal B*² di F. Aura Jorro e il progetto DAMOS: *Database of Mycenaean at Oslo*³ di F. Aurora. Il primo contiene indici diretti e inversi delle parole e un indice dei logogrammi della lineare B, di tipo non

¹ <http://lila.sns.it/mnamon/index.php?page=Scrittura&id=20> (accesso del 26/06/2014).

² <http://www.cervantesvirtual.com/bib/porta/diccionariomicenico/index.html> (accesso del 26/06/2014).

³ <https://www2.hf.uio.no/damos/> (accesso del 26/06/2014).

interattivo, compilati sulla base delle edizioni cartacee correnti. Il secondo contiene tutti i testi in lineare B in traslitterazione, riprodotti in forma semplificata e interrogabili attraverso un apposito motore di ricerca.

LiBER intende ampliare e migliorare questo genere di approccio ai documenti in lineare B, mettendo a disposizione dell'utente finale non solo una trascrizione epigrafica corretta dei testi e strumenti per la produzione di indici di tipo interattivo, ma anche risorse utili per lo studio degli aspetti topografici, paleografici e cronologici sia dei documenti in quanto tali che dei dati ricavabili dall'interrogazione del database.

3. ORIGINE E SVILUPPO DELL'IDEA

L'idea di realizzare un *Document Management System* di questo tipo è nata a seguito di una serie di riflessioni, che si possono così sintetizzare.

I documenti in lineare B, che contengono per lo più censimenti di beni o transazioni economiche su tavolette d'argilla, offrono una testimonianza imprescindibile per le ricerche sulla società e l'economia palaziale micenea. Tuttavia, per essere correttamente utilizzati come fonti storiche, devono prima essere ricomposti nella loro unità originaria. È raro infatti che una tavoletta in lineare B costituisca da sola un documento contabile. Al contrario, un tipico documento contabile è costituito da decine di tavolette, scritte in genere in uno stesso luogo e da uno stesso individuo per uno stesso scopo e caratterizzate dalle stesse formule e dagli stessi logogrammi.

Purtroppo, spesso lo stato frammentario delle tavolette ostacola l'operazione di ricostruzione dei documenti, ragione per cui in molti casi l'attribuzione di uno o più frammenti a un certo documento – come del resto l'attribuzione di uno o più frammenti a una stessa tavoletta non più integra – si può fare solo combinando le informazioni testuali superstiti con quelle archeologiche (contesto e cronologia), topografiche (findspot) e paleografiche (identificazione dello scriba autore della tavoletta frammentaria). Tale esigenza è ben nota da tempo a tutti gli specialisti del settore ed è stata recepita a vari livelli anche nelle edizioni dei testi. Questi, infatti, sono in genere raggruppati per serie coerenti dal punto di vista del contenuto e, nel caso dell'archivio di Cnosso, sono ulteriormente suddivisi in "set", sottogruppi corrispondenti ad altrettanti documenti contabili nel senso sopra indicato⁴.

I singoli documenti d'archivio, d'altro canto, dischiudono il loro potenziale informativo solo se correttamente posti in relazione tra loro, un obiettivo che si può raggiungere solo combinando, anche in questo caso, le informazioni testuali con quelle extratestuali, di natura archeologica, topografica

⁴ Si fa riferimento qui non ai *corpora*, ma alle edizioni in traslitterazione (*FDC IV, KT5, PTT I, TITHEMY*): nei *corpora*, infatti, i testi non sono ordinati per serie, ma per numero di inventario.

e paleografica (documenti redatti da uno stesso scriba in uno o più uffici; documenti redatti da più scribi in uno stesso ufficio; documenti redatti da più scribi in più uffici, ma legati tra loro da uno stesso scopo amministrativo, etc.). La stessa combinazione di criteri, infine, si rivela fondamentale anche per la comprensione degli assetti amministrativi dei vari centri palaziali, sia in termini di relazioni fra scribi, uffici e dipartimenti (gerarchie scribali), sia per la ricostruzione dei flussi delle informazioni (all'interno dei centri amministrativi o fra questi e i rispettivi territori) e dei sistemi di gestione e di archiviazione dei dati presso le singole amministrazioni (esistenza o meno di archivi centrali, loro articolazione e consistenza).

La combinazione dei criteri testuali ed extratestuali che presiedono alla ricostruzione di una tavoletta, di un intero documento o delle relazioni esistenti tra più documenti, tuttavia, si rivela spesso un'operazione complicata e dai risultati incerti. Come dimostra la storia degli studi, inoltre, le ricostruzioni possono mutare facilmente o per la modifica di una lettura o per l'attribuzione di una tavoletta a un diverso scriba o per la revisione di una cronologia o di un findspot o per vari altri motivi. Spesso, inoltre, a causa del numero elevato delle variabili in gioco, le soluzioni corrette tendono ad essere trovate solo dopo vari tentativi.

A fronte delle peculiarità della documentazione e delle difficoltà oggettive nella ricostruzione e nell'analisi della struttura degli archivi in lineare B, si è ritenuto utile, dunque, sviluppare uno strumento di gestione dei documenti capace di applicare e combinare tra loro in modo semplice ed efficace i criteri sopra menzionati, in modo da favorire, da un lato, la verifica del livello di attendibilità dei raggruppamenti testuali correnti (quelli pubblicati nelle edizioni cartacee di riferimento) e, dall'altro, lo sviluppo di nuove ipotesi, soprattutto in relazione ai testi che ancora sfuggono a una classificazione e a una interpretazione convincenti.

Si è così deciso di inserire in un database i testi in traslitterazione pubblicati nelle edizioni cartacee più recenti, aggiornandole con le modifiche e le integrazioni pubblicate in data successiva e aggiungendo per i singoli testi, laddove disponibili, le informazioni su scribi, findspot, cronologie e luoghi di conservazione.

Si è ritenuto opportuno, inoltre, presentare i dati sia in forma tabulare, favorendone così la visione sinottica e la selezione in base ai criteri di aggregazione sopra menzionati, sia in forma più "tradizionale", impaginando i singoli testi, i rispettivi apparati critici e le informazioni extratestuali come nelle comuni edizioni cartacee, di cui peraltro si è deciso di rispettare fedelmente le regole di trascrizione epigrafica. Alle pagine così concepite si è d'altra parte ritenuto utile aggiungere funzionalità di tipo interattivo, fra cui in particolare la possibilità di selezionare gli elementi testuali come chiavi di ricerca all'interno del database.

Si è ritenuto infine fondamentale dotare il sistema di un motore di ricerca testuale, di un apparato fotografico e di funzionalità webGIS. Il motore di ricerca, calibrato sulle specificità delle scritture logo-sillabiche e realizzato sulla base di un modello già testato sulle scritture cuneiformi (cfr. *infra*), è compatibile con la sintassi delle *Espressioni Regolari* ed è stato concepito per produrre indici di parole ed elenchi di testi in base a chiavi di ricerca anche complesse, eventualmente integrate da notazioni epigrafiche (caratteri puntati, corsivi, segni di lacuna, di rasura, etc.). Quanto alle fotografie, che accompagnano i testi e gli apparati critici come nelle edizioni cartacee da cui sono tratte, si è ritenuto utile presentarle sotto forma di miniature zoomabili, per consentire sia di esaminare il layout dei testi sia di studiare il tracciato dei segni (due elementi fondamentali per l'attribuzione dei testi alle mani scribali). Infine, le funzionalità webGIS sono state integrate in modo tale da generare mappe interattive dei findspot sia dei testi selezionati dalla tabella del database sia di quelli filtrati attraverso le query inviate al motore di ricerca.

Qui di seguito sono illustrati e discussi alcuni aspetti tecnici legati all'implementazione delle funzionalità sopra descritte.

4. ASPETTI TECNICI DI LIBER

4.1 *La codifica dei testi*

La realizzazione di un *Document Management System* per il trattamento di scritture logo-sillabiche presenta un'ampia serie di incognite e variabili, non sempre prevedibili in fase di progettazione. Sebbene nel corso delle ultime due decadi enormi sforzi siano stati riservati alla realizzazione di strumenti condivisi e soluzioni di ampio respiro per l'annotazione e la marcatura dei documenti, sia antichi sia moderni, questa fervente attività di ricerca si è concentrata per ovvie ragioni quasi esclusivamente su scritture alfabetiche ed ha individuato nel metalinguaggio di marcatura XML la migliore tra le opzioni disponibili. In questo panorama ha acquisito sempre maggior credito uno standard denominato TEI (Text Encoding Initiative)⁵, uno schema generalistico adatto alla codifica e rappresentazione digitale di un'ampia gamma di tipologie testuali, mentre una sua variante nota con il nome di EpiDoc è stata appositamente progettata per l'edizione di documenti antichi⁶.

Per quanto riguarda le scritture logo-sillabiche, l'esempio più significativo di impiego di un metalinguaggio di marcatura è stato impiegato dal progetto ORACC, realizzato e mantenuto da Steve Tinney⁷ per la codifica

⁵ <http://www.tei-c.org/index.xml> (accesso del 03/01/2014).

⁶ <http://sourceforge.net/p/epidoc/wiki/Home/> (accesso del 20/01/2014).

⁷ Si vedano le osservazioni in "TEI: TEI for Oracc", <http://oracc.museum.upenn.edu/doc/developer/tei> (accesso del 20/01/2014) e relativi rimandi.

dei testi in cuneiforme provenienti dall'area Mesopotamica e databili tra il III e il I millennio a.C. Il sostanziale beneficio di un sistema di marcatura gerarchico quale l'XML, impiegando, come nel caso di ORACC, una variante della codifica TEI-EpiDoc, risiede nell'adesione ad uno standard di marcatura condiviso, con un'ampia comunità e sviluppato senza soluzione di continuità a partire dai primi anni '90.

La prima e più ovvia conseguenza di un simile approccio è determinata dalla "portabilità" del documento codificato. Questo aspetto, tuttavia, risulta secondario rispetto al processo di marcatura in sé: l'utilità di un testo strutturato, infatti, trova fondamento nella possibilità di assegnare uno o più attributi ai singoli elementi logici che compongono il documento, siano essi i singoli lemmi o porzioni più o meno ampie di testo (per es. da interi periodi a singoli grafemi). L'impiego di un modello XML di codifica, seppur appositamente progettato per questo scopo, tuttavia, presenta una serie di limiti non secondari, specialmente se rapportati al presente ambito di ricerca: nella maggior parte dei casi, il risultato acquisisce considerevoli "ridondanze", non tutti i documenti possono essere ridotti a una struttura strettamente gerarchica⁸, l'impiego di un sistema di interrogazione generalistico come XQuery rende molto complicata l'organizzazione di complesse analisi testuali e, infine, possono sorgere problemi di gestione legati alla "leggibilità" del risultato prodotto (MAURIZIO, ORSINI 2010).

In ragione di questa serie di inconvenienti, per LiBER si è optato per una soluzione mista, basata su una marcatura "leggera" – una sorta di "trascrizione diplomatica" delle singole tavolette – e sull'adozione di un database relazionale per l'immagazzinamento dei documenti trattati e dei rispettivi attributi.

La trascrizione utilizzata per la codifica dei testi in lineare B è stata mutuata da quella impiegata per il progetto SINLEQIUNNINI, un framework per la gestione dei testi cuneiformi in traslitterazione, risultato di una pluriennale collaborazione tra ISMA-CNR, Università Ca' Foscari di Venezia e Università degli Studi di Napoli l'Orientale⁹. Tale approccio trae a sua volta origine da un metodo di marcatura ideato da G. Buccellati alla fine degli anni '60 per la codifica delle tavolette cuneiformi antico-babilonesi e in seguito impiegato per quelle provenienti dagli archivi palatini di Ebla (Siria)

⁸ Questo aspetto non fa altro che aumentare la "ridondanza" del testo codificato. Si veda ad esempio <http://www.tei-c.org/release/doc/tei-p5-doc/it/html/NH.html> (accesso del 10/01/2014).

⁹ La collaborazione vede coinvolti Francesco Di Filippo (ISMA-CNR), Carlo Zaccagnini (Università degli Studi di Napoli l'Orientale) e Lucio Milano (Università Ca' Foscari, Venezia). A partire dal 2006 il framework SINLEQIUNNINI è stato impiegato con successo per la realizzazione degli archivi digitali dei testi cuneiformi provenienti da Ebla (*Ebla Digital Archives* [EbDA], <http://virgo.unive.it/eblaonline/>) e da Emar (*Middle Euphrates Digital Archive* [MEDA], <http://virgo.unive.it/emaronline/cgi-bin/index.cgi>). Per maggiori dettagli riguardo al framework e ai singoli progetti, si rimanda alla pagina web <http://www.pankus.com/>.

(BUCCELLATI 1977, 2011)¹⁰. Si tratta di un sistema di marcatura in linea che, basandosi sulle peculiarità delle sequenze grafemiche del cuneiforme, permette di ridurre al minimo l'utilizzo di marcatori (tag), senza per questo ridurre l'efficacia della codifica. Nella sua versione originaria, tuttavia, questo sistema è quasi esclusivamente concepito per l'annotazione epigrafica dei testi: la trascrizione consente di implementare la codifica con alcune categorie lessicali, ma non è possibile nessuna notazione morfo-sintattica. In termini pratici, la sua originaria applicazione trova concreta efficacia come strumento per l'estrazione di alcune tipologie di informazioni dai testi, ma soprattutto per rendere al meglio le complessità grafiche del sistema di scrittura cuneiforme, in particolare quello del III millennio a.C., pur ovviamente non riducendosi a quello che in gergo informatico potrebbe definirsi come un semplice "foglio di stile". Questo approccio originario è stato in seguito implementato nel progetto SINLEQIUNNINI attraverso l'adozione di un database relazionale per la gestione della multiforme serie di attributi necessari alla marcatura dei testi e alla loro interrogazione.

Volendo riassumere in breve il processo di codifica delle tavolette in lineare B utilizzato per LiBER, lo si può descrivere come segue. Una prima fase prevede la codifica manuale dei documenti secondo un sistema di marcatura diplomatica appositamente progettato. Un secondo stadio consiste nella segmentazione dei documenti in unità minime di analisi ("tokenizzazione"), che, in base alle diverse necessità, possono essere le singole parole così come i singoli grafemi. I diversi elementi logici, in questa fase, preservano sia gli attributi necessari alla loro corretta formattazione sia alcuni aspetti prettamente epigrafici (per es. determinativi, numerazione degli omofoni, etc.). Il processo termina con l'inserimento degli elementi logici (token) all'interno delle diverse tabelle del database secondo stringenti regole di normalizzazione. Quest'ultimo stadio permette la rimozione dal dataset di duplicati e elementi ridondanti secondo l'impiego di *Normal Forms* e la loro gestione attraverso indici relazionali (per un'esauriente panoramica sulle "forme normali" cfr. POWELL 2006, 73 ss). È solo in questa fase che ogni singolo token viene arricchito con le pertinenti informazioni tipologiche ed eventualmente morfo-sintattiche. Non sempre è possibile estrarre questo genere di informazioni: nel caso di LiBER, ad esempio, non è stato progettato alcun modulo per la lemmatizzazione. Software di questo tipo sono stati realizzati solo per i *corpora* accadici codificati attraverso il framework SINLEQIUNNINI.

A differenza di un processo di marcatura gerarchico, che predilige la creazione di un unico file di testo comprensivo di tutte le informazioni (da quelle testuali a quelle archeologiche, da quelle topografiche a quelle paleografiche), LiBER prevede che il documento venga segmentato in unità minime,

¹⁰ http://128.97.6.202/gb-info/digital/gb_digital.html.

calibrate in base alle necessità, e che ogni frammento venga immagazzinato nel pertinente contesto logico, comprensivo sia degli attributi specifici sia dei necessari indici per la ricostruzione del documento nella sua interezza.

Per completezza, infine, è utile soffermarsi su quello che potrebbe essere considerato come il principale limite di un simile processo di marcatura: la terzietà rispetto ad uno standard d'uso corrente e, dunque, l'attuale impossibilità di esportare i testi strutturati in un formato condiviso capace di favorire un livello di interoperabilità. Tuttavia, va altresì constatato come lo stato attuale degli studi sulla trascrizione delle fonti primarie in scritture logo-sillabiche si trovi ancora in una fase fortemente pre-paradigmatica, seppur in precipitosa accelerazione. Sebbene l'ultima decade abbia visto il fiorire di numerose iniziative tese alla formalizzazione di documenti risalenti alla più remota antichità, in modo particolare nell'ambito della ricerca Vicino Orientale, non si è ancora affermato uno standard di codifica di tipo generalistico, come è invece accaduto in altri settori della filologia computazionale quali quelli legati agli studi classici¹¹. A questo, infine, va aggiunta l'attuale impossibilità di usufruire degli strumenti che si vanno sempre più affermando nel campo della linguistica computazionale. Basti qui riportare lo schema presentato in CALZOLARI, LENCI (2004) che riassume procedure e potenzialità del *Natural Language Processing* (NLP) (Fig. 1).

Appare evidente come non ci sia nulla di più distante dall'attuale portata dei diversi progetti di digitalizzazione dei *corpora* di testi storici logo-sillabici. La prospettiva, semmai, andrebbe ribaltata: ogni possibile strumento di analisi per questa tipologia di fonti dovrebbe tenere in conto, in primo luogo, la straordinaria molteplicità di variabili caratteristiche dei diversi sistemi grafici (cfr. al riguardo le diverse ipotesi di lavoro relative agli studi cuneiformi presentate in BARTHÉLEMY 1998, 2009; MACKS 2002; SMITH 2007).

Lasciando a margine le future possibilità offerte da più raffinati strumenti di analisi linguistica, per LiBER è parso opportuno concentrarsi sulla vasta gamma di problemi filologici ed epigrafici legati ad un progetto di digitalizzazione di fonti primarie in traslitterazione. Seppur a discapito della "portabilità" dei testi codificati, dunque, si sono potuti sperimentare alcuni significativi benefici rispetto ad un approccio basato su una marcatura gerarchica, specialmente per quanto riguarda i problemi sopra descritti della "ridondanza" e della semplicità di gestione. Il procedimento utilizzato, prevedendo una marcatura "leggera", ha permesso da un lato di preservare un documento "leggibile" ancorché codificato, dall'altro di ridurre al minimo la sua dimensione e le ripetizioni interne al documento stesso. Se la cosiddetta leggibilità è direttamente collegata alla velocità di creazione, gestione e ag-

¹¹ Si veda ad esempio il progetto EPIDOC e gli esempi in http://wiki.tei-c.org/index.php/Samples_of_TEI_texts (accesso del 21/01/2014).

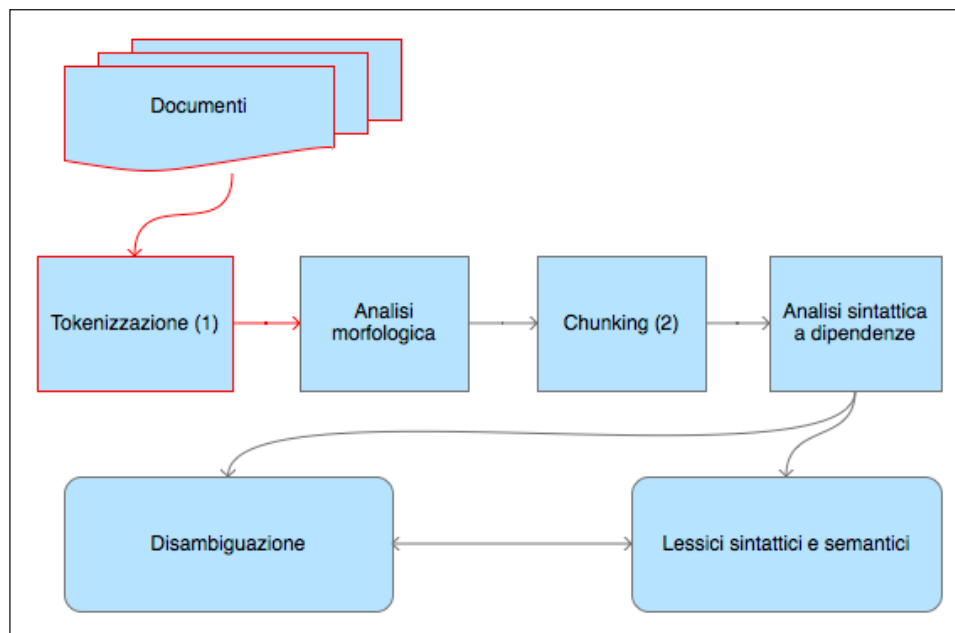


Fig. 1 – Processo di analisi in NLP (da CALZOLARI, LENCI 2004). Note al grafico: 1) tokenizzazione = segmentazione del testo in unità minime di analisi; 2) chunking = disambiguazione morfosintattica e segmentazione del testo in sequenze di gruppi sintattici.

giornamento del documento strutturato – caratteristiche non secondarie per un progetto che può avvalersi solo di limitate risorse – il vantaggio sostanziale di un simile approccio metodologico, comunque, risiede nella possibilità di eliminare l'impressionante quantità di marcatori, tag di apertura e di chiusura per ogni singolo elemento logico o gruppi di elementi, necessari per il processo di serializzazione di una codifica XML.

La marcatura del token attraverso un sistema di “record ↔ attributi”, suddiviso in tabelle secondo un rigido criterio di normalizzazione, ha permesso di ridurre considerevolmente le dimensioni della banca dati. La prima e più significativa conseguenza di quest'impostazione riguarda i tempi di latenza necessari per le operazioni di deserializzazione dell'intera banca dati, tempi che si sono dimostrati sostanzialmente trascurabili, sia per quanto riguarda la stampa a schermo dei singoli documenti sia, soprattutto, per l'esecuzione delle interrogazioni.

È proprio quest'ultimo aspetto che si è voluto privilegiare in LiBER a discapito di ogni altra considerazione d'ordine metodologico. Si è voluto creare, infatti, uno strumento flessibile e veloce per il trattamento del potenziale informativo dei documenti, sia esso inteso a livello testuale, archeolo-

gico, topografico e paleografico. Contrariamente a quanto avviene con un approccio basato su una marcatura XML, dunque, il progetto è in primo luogo un efficiente motore di ricerca appositamente progettato per combinare la multiforme varietà dei dati di un archivio, mentre l'edizione digitale del testo in sé risulta essere complementare.

4.2 Sviluppo di strumenti per una gestione integrata dei documenti e dei loro contesti

La più significativa innovazione introdotta in LiBER, almeno quella direttamente connessa agli aspetti tecnico-progettuali, riguarda le potenzialità offerte dall'integrazione di un webGIS in un progetto di natura epigrafico-filologica. Non è certo questa la sede per addentrarsi in profondità in una descrizione di un simile strumento e delle innovazioni introdotte nel campo delle *Digital Humanities* (cfr. in particolare DJINDJIAN 2008 e HARRIS 2010). Viceversa, sarà opportuno soffermarsi brevemente su alcuni aspetti progettuali, sulle opzioni scelte e le metodologie impiegate per la realizzazione della componente geografica di LiBER.

Nel corso delle ultime tre decadi si sta assistendo ad una serie di repentini sviluppi tecnologici che in una misura insospettabile hanno alterato i modi e il ruolo della rappresentazione cartografica, avendo accresciuto la portata e la disponibilità di dati in formato digitale, così come gli strumenti per la loro gestione (Fig. 2). In questo panorama di repentini mutamenti non è stato semplice selezionare gli strumenti adeguati per la realizzazione di un'infrastruttura capace di arricchire un già denso patrimonio storico-filologico con informazioni di natura spaziale. Tuttavia, è proprio questo processo che si intende qui descrivere brevemente.

L'infrastruttura geografica di LiBER, interamente basata su software e protocolli rilasciati con licenza libera (Open Source), può essere sinotticamente rappresentata dallo schema presente in Fig. 3.

Le carte archeologiche sono state elaborate a partire dagli originali delle edizioni di riferimento¹² e georeferenziate attraverso punti di controllo recuperati dalle foto satellitari distribuite da Google. Questo processo ha introdotto nella rappresentazione cartografica una distorsione grafica e alcuni "errori" rispetto alle originali piante archeologiche. La suddetta distorsione è dovuta alla proiezione geografica scelta, che è a sua volta determinata da esigenze progettuali. Le differenze sono sostanziali, come è possibile osservare nella Fig. 4.

La georeferenziazione, inoltre, ha introdotto anche due tipologie di "errore". Il primo è determinato dal processo stesso di georeferenziazione,

¹² Per Micene, TREUIL *et al.* 2008, 390, fig. 50; per Midea, DEMAKOPOULOU *et al.* 2002, 42, fig. 41; per Tirinto, RUDOLPH 1973, tav. 9 (necropoli di Profitis Ilias: rilievo di H. Sulze, 1932) e RIEGER, BÖSER 1990, carta 8 "Tiryns" (palazzo e città bassa: rilievo di W. Böser e H. Daske, 1980).

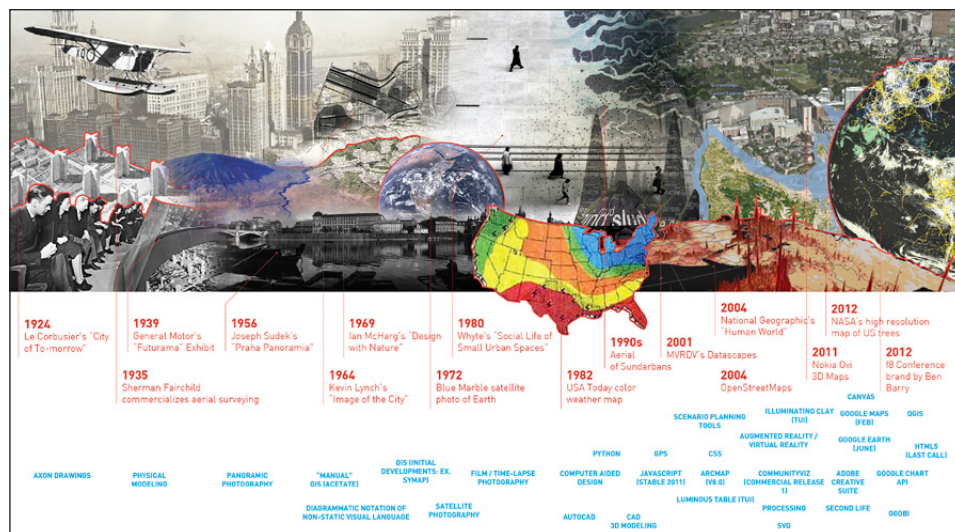


Fig. 2 – Sviluppi paralleli della rappresentazione cartografica “urbana” e delle tecnologie informatiche (da KENNEDY 2012).

che determina una inevitabile distorsione di alcune geometrie nell’operare la riproiezione della griglia del raster: sebbene il procedimento impiegato (la trasformazione polinomiale di primo grado, BROVELLI, MINGHINI 2012) permetta solo il ridimensionamento, la traslazione e la rotazione della carta, un generale errore orizzontale da 1 a 5 m circa è riscontrabile in quelle porzioni dell’output che sono più distanti dai punti di controllo. Diverso è l’errore orizzontale determinato dall’impiego di punti di controllo ottenuti per mezzo della cartografia digitale di Google. Queste immagini satellitari hanno un errore orizzontale che può variare da 1 m in contesti urbani a 30 m in contesti extra-urbani, un errore che per giunta non è omogeneamente distribuito (POTERE 2008; BENKER, LANGFORD, PAVLIG 2011).

Ne deriva che, nonostante un rigoroso processo di acquisizione, il risultato non può che risentire di questo limite e, nella sostanza, la cartografia digitale, così come i findspot delle tavolette individuati proprio a partire da quella cartografia, non possono considerarsi validi in termini assoluti: ad esempio, non possono essere paragonati per precisione ad una mappatura operata sul territorio con una Stazione Totale. Ciò che, al contrario, è assolutamente corretto nella rappresentazione geografica di LiBER sono i rapporti relativi tra le strutture e i findspot: se è ravvisabile una qualche forma di traslazione delle evidenze rispetto alla loro posizione reale – ed è probabile che ci sia – questa è comunque omogenea e non inficia il risultato complessivo.

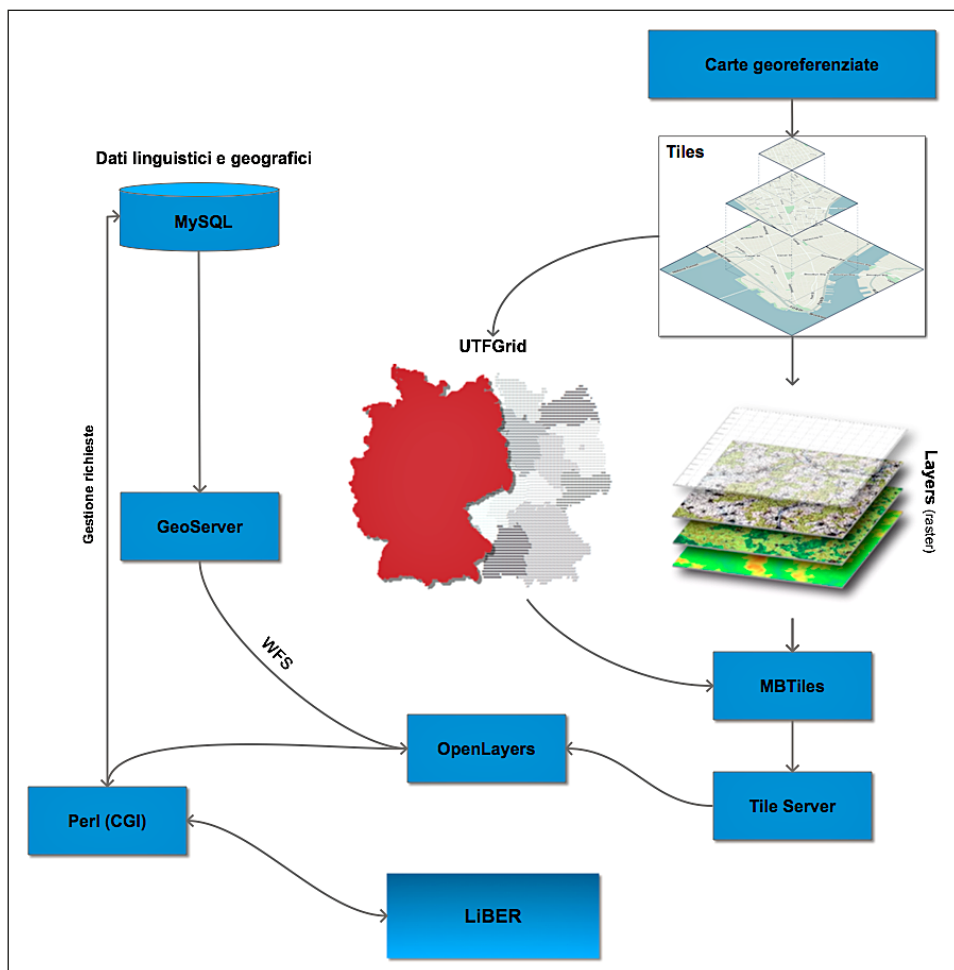


Fig. 3 – Schema dell'infrastruttura spaziale di LiBER.

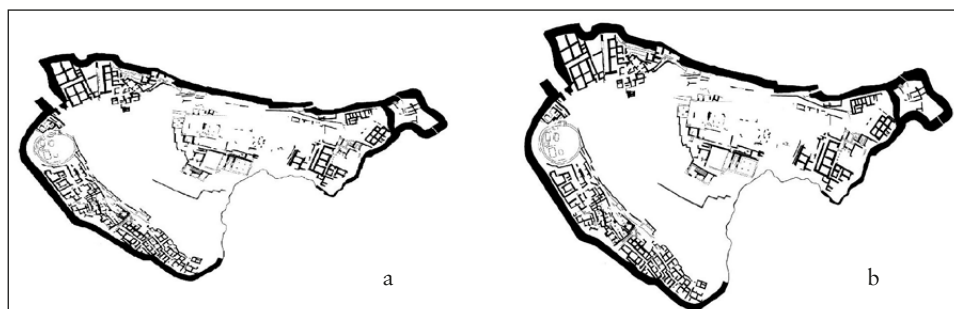


Fig. 4 – Distorsioni geometriche determinate da differenti sistemi geografici di proiezione: A) Micene (epsg:4326), B) Micene (epsg:3857).

Le carte archeologiche così prodotte sono state, infine, elaborate attraverso un software per il *tiling*. Si tratta di un procedimento per la costituzione di piccole “piastrelle” di 256x256 pixel suddivise per i necessari livelli di zoom della *slippy map* e che possono essere riunite per creare il ben noto effetto ottico di una mappa continua. Per il procedimento si è scelto il formato di output *.mbtiles*, un file database SQLite, capace non solo di eliminare le “tegole” duplicate (per es. celle vuote o uguali tra loro), con un considerevole risparmio in termini di spazio di archiviazione, ma anche di aggiungere al layer raster un secondo livello non visibile con un ulteriore livello di dati.

In tutti quei casi in cui l’informazione vada considerata come un necessario complemento della cartografia storica – informazione con la quale l’utente non può direttamente interagire se non attraverso le operazioni di zoom – la trasmissione dei dati non avviene attraverso l’inserimento di un layer vettoriale, come di consuetudine, ma tramite un layer UTFGrid. Questa soluzione si basa su un processo di rasterizzazione delle informazioni contenute nel layer vettoriale in una matrice di caratteri ASCII che vengono fusi, per così dire, con la cartografia raster: ad ogni carattere compreso nella matrice corrisponde un solo attributo del layer vettoriale originario, così che la pertinente porzione di informazione è già presente nella mappa e va solo deserializzata in locale.

In LiBER tutte le tipologie di dati che costituiscono un essenziale complemento dell’informazione geografica, e con i quali l’utente non può interagire attraverso filtri, sono trattati attraverso il protocollo UTFGrid. Questo layer si colloca esattamente al di sotto del *base layer* visibile, costituito da una carta dei rilievi e da quella delle strutture antiche, e rimane in attesa fino a quando l’utente non muove il mouse al di sopra di una delle strutture censite: è solo a questo punto che le informazioni pertinenti vengono visualizzate, riducendo al massimo i tempi di latenza. Per la mappa di base si è scelto di utilizzare una versione delle *slippy maps* distribuite da Google, dalla quale sono state eliminate tutte le feature geografiche ad eccezione dei rilievi e delle etichette dei mari.

Le altre tipologie di layer vettoriali necessarie all’arricchimento delle carte archeologiche (per es. i *findspot* delle tavolette) sono immagazzinate nel database principale. Come database server si è scelto di impiegare MySQL in associazione alle estensioni per la gestione dei dati spaziali tramite GeoServer. Si è optato per questa soluzione, a discapito di una più complessa infrastruttura basata sulla coppia PostgreSQL-PostGIS, vista la relativa semplicità delle operazioni richieste: i *findspot* delle singole tavolette sono stati immagazzinati come elementi puntuali caratterizzati da coppie di valori decimali (latitudine - longitudine) e tradotti in un formato nativo per MySQL (WKB). Attualmente, non avendo necessità di complesse query spaziali (se non alcune misurazioni euclidee delle distanze tra *findspot*), l’infrastruttura messa in opera si è dimostrata più che affidabile.

GeoServer, invece, ha il compito di gestire la trasmissione delle informazioni in formato vettoriale (GeoJSON): i dati immagazzinati in MySQL e le richieste a GeoServer riguardano esclusivamente le tipologie di informazioni variabili, ovvero i gruppi di dati che l'utente può filtrare attraverso una serie di selettori e/o campi di testo impostati nel front-end web del progetto. Queste richieste sono risolte per mezzo del protocollo WFS (*Web Feature Service*), un protocollo di straordinaria efficacia per la selezione di porzioni del dataset, ma non particolarmente efficiente per transazioni di grandi quantità di dati, che invece sono gestiti dal protocollo UTFGrid.

Infine, l'intero sistema è gestito da Open Layers, un framework javascript per la visualizzazione e l'interazione di mappe e dati cartografici tramite browser.

5. ESEMPI DI INTERROGAZIONE DEL DATABASE

Alcuni esempi di interrogazione del database possono dare un'idea più concreta delle potenzialità del sistema.

Se dalla tabella del database si selezionano le iscrizioni di Micene (inserendo la sigla "MY" nel search box "Site") e poi si filtrano i record così ottenuti inserendo la sigla "LH III B2" nel search box "Chronology", si ottiene l'elenco completo e aggiornato di tutte le iscrizioni del sito risalenti alla fase finale dell'età palaziale (Fig. 5); se poi si clicca sul pulsante "Plot data", si può avere la distribuzione topografica dei testi selezionati con l'indicazione contestuale del numero di iscrizioni rinvenute in ogni singolo findspot (Fig. 6). Se, viceversa, dopo aver selezionato le iscrizioni di Micene, si seleziona lo scriba 62 dal menu a tendina "Scribe", si ottiene l'elenco completo dei testi vergati da tale funzionario. I documenti dello scriba 62, poi, possono essere o rappresentati su mappa (cliccando sul pulsante "Plot data") o riprodotti in sequenza (cliccando sul pulsante "Print data"), secondo un principio di aggregazione testuale totalmente nuovo rispetto a quelli in uso nelle edizioni cartacee e utile, ad esempio, a facilitare i confronti di tipo lessicale o formulare all'interno della produzione di uno o più scribi (Fig. 7).

Se nel search box del motore di ricerca si digita "NI" e si spunta il check box "Case sensitive", si ottiene l'elenco completo delle attestazioni della sigla NI (che nei testi in lineare B indica i fichi), indipendentemente dal fatto che questa sia di lettura certa o incerta (NI o NI) o che sia attestata da sola o in legatura con un altro segno (per es. *155^{VAS}+NI: vaso di fichi). L'elenco generato dal motore di ricerca indica anche i findspot, gli scribi e le cronologie dei singoli documenti, così che è possibile avere informazioni precise sulla distribuzione dei dati ricercati (per siti, findspot, scribi e fasi cronologiche) (Fig. 8). I pulsanti "Print resultset" e "Plot resultset on map", infine, permettono

Site	Document	Findspot	Scribe	Chronology	Inv. No.
MY	Fu 711	N.E. Acropolis	—	LH III B2	Nauplio M. 26703
MY	L 710	House of Columns (Magazine of the pithoi)	—	LH III B2	Nauplio M. 26701
MY	Oi 701	Citadel House (Room 1)	63	LH III B2	Nauplio M. 13855
MY	Oi 702	Citadel House (Room 4)	64	LH III B2	Nauplio M. 13856
MY	Oi 703	Citadel House (Room 4)	64	LH III B2	Nauplio M. 13857
MY	Oi 704	Citadel House (Room 4)	64	LH III B2	Nauplio M. 13858
MY	Oi 705	Citadel House (Room 4)	—	LH III B2	Nauplio M. 13859
MY	Oi 706	Citadel House (Room 4)	—	LH III B2	Nauplio M. 13860
MY	Oi 708	Citadel House (Room 4)	—	LH III B2	Nauplio M. 13862
MY	Wt 700	Citadel House (not stratified)	—	LH III B2	Nauplio M. 59-107
MY	Wt 713	Acropolis (West of Building K)	—	LH III B2	?
MY	X 707	Citadel House (Room 4)	—	LH III B2	Nauplio M. 13861

Fig. 5 – Tabella del database di LiBER con i testi di Micene risalenti alla fase LH III B2.

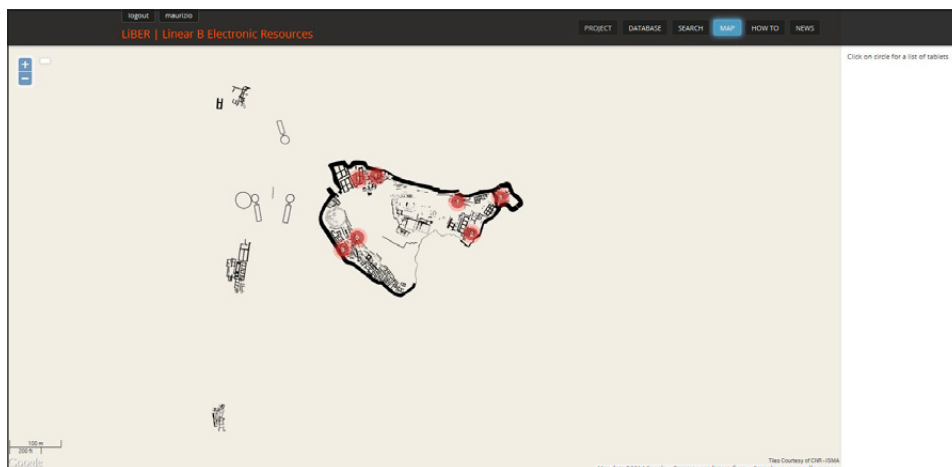


Fig. 6 – Mappa dei findspot dei testi di Micene risalenti alla fase LH III B2.

rispettivamente di esaminare i testi contenenti i dati ricercati e di ottenere la distribuzione geografica (siti) e topografica (findspot) degli stessi.

Un ultimo esempio può infine illustrare la flessibilità e le ulteriori potenzialità del motore di ricerca, che, come detto, è compatibile con la sintassi

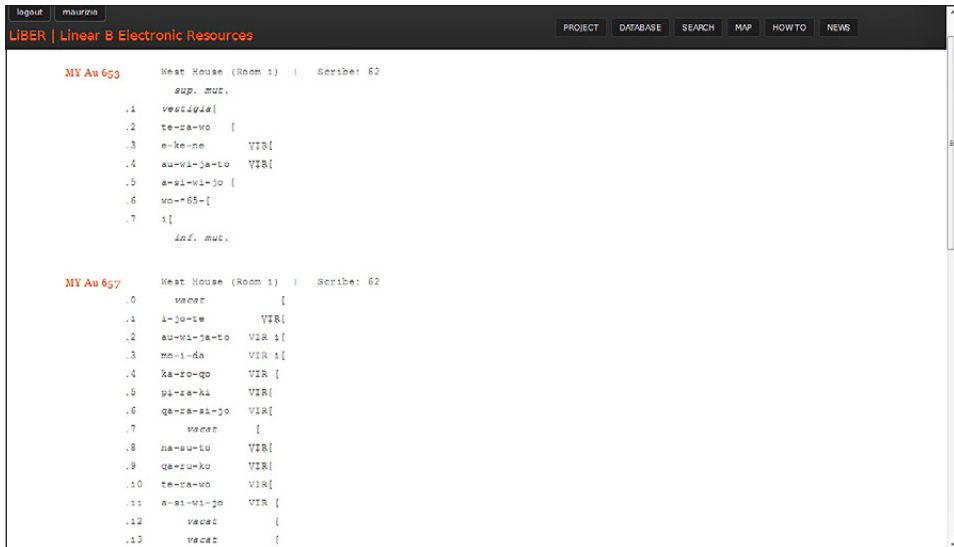


Fig. 7 – Elenco dei testi dello scriba 62 di Micene.

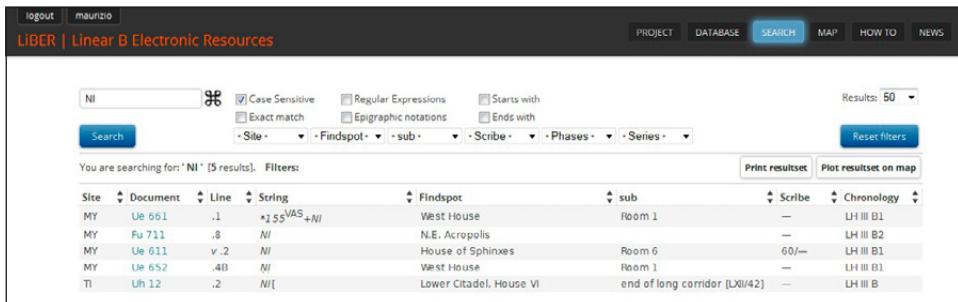


Fig. 8 – Indice delle attestazioni della sigla NI “fichi”.

delle *Espressioni Regolari*. Se nel search box del motore di ricerca si digita la formula $(\backslash *[0-9]\{3\})\{([A-Z]\{3,4\})\}$, si spuntano i check box “Case sensitive” e “Regular Expression” e si seleziona “Tiryns” dal menu a tendina “Site”, si ottiene l’elenco completo dei logogrammi attestati nei documenti di Tirinto, ordinato alfabeticamente e completo delle eventuali letture alternative menzionate negli apparati critici dei singoli testi (Fig. 9). Anche in questo caso, come nel precedente, i testi dell’elenco generato dal motore di ricerca possono essere rappresentati su mappa (mediante il pulsante “Plot resultset on map”) o riprodotti in sequenza (mediante il pulsante “Print resultset”).

LiBER | Linear B Electronic Resources

SEARCH

Search: Case Sensitive Regular Expressions Starts with Exact match Epigraphic notations Ends with

Results: 50

You are searching for: '(*[0-9]{3})|(*[A-Z]{3,4})' (18 results). Filters: TI |

Site	Document	Line	String	Findspot	sub	Scribe	Chronology
TI	SI 5	.1	ARM	Lower Town, Sector W		—	undated
TI	SI 5	.2	ARM	Lower Town, Sector W		—	undated
TI	X 17	.2	JAJR + JMA + JLANA	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	Cb 4	.2	BOIS	Lower town, Sector H, House O		—	LH III B-C
TI	EF 2		GRA	Lower Town, Sector H, House M	under South wall	—	LH III B
TI	SI 8	.3]ROTA + [MO]	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	SI 9	.2]ROTA	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	SI 9	.3]ROTA	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	SI 10]ROTA	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.2	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.3]VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.4	VIR[Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.5	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.6	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.7	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.8	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.10	VIR	Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B
TI	AI 7	.13	VIR[Lower Citadel, House VI	end of long corridor [LXII/42]	—	LH III B

Fig. 9 – Indice dei logogrammi attestati nei documenti in lineare B di Tirinto.

Per la descrizione dettagliata delle funzionalità del sistema si può consultare la pagina “How To” del sito web di LiBER.

6. STATO DI AVANZAMENTO DEL PROGETTO E PROSPETTIVE

Al momento, il database di LiBER contiene i testi in lineare B provenienti dai siti palaziali di Micene, Midea e Tirinto. Si tratta di un campione di testi limitato (179 in tutto), ma sufficientemente rappresentativo delle caratteristiche dei materiali epigrafici in lineare B. Il fatto che i testi, pur provenendo da più siti, appartengano a una regione geograficamente coerente (l’Argolide micenea) e che i depositi d’archivio risalgano, anche in uno stesso sito, a epoche diverse, permette inoltre di effettuare test e ricerche che vanno al di là dei singoli documenti o dei rapporti fra i documenti di uno stesso archivio. È possibile, cioè, analizzare la diffusione areale e cronologica di singoli elementi testuali (vocaboli, logogrammi, etc.) o, come dimostrano i primi utilizzi di LiBER da parte della comunità degli egeisti, testare ipotesi complesse basate sulla distribuzione dei testi per fasi cronologiche distinte. Ad esempio, in un contributo attualmente in corso di stampa, finalizzato all’individuazione dei criteri epigrafici ed archeologici utili a definire edifici e siti micenei come sedi del potere palaziale, P. Darcque e F. Rougemont citano LiBER come un nuovo strumento utile all’analisi della distribuzione geografica e cronologica delle evidenze epigrafiche (DARCQUE, ROUGEMONT c.s.).

Le ricerche “multiarchivio” di questo tipo sono per il momento limitate a causa del numero ridotto di testi disponibili, ma con il progressivo ampliarsi del database verranno a costituire una funzionalità importante del sistema, in grado di fornire risposte analitiche a interrogativi complessi di varia natura (epigrafica, archivistica, prosopografica, paleografica, linguistica, etc.).

La prossima fase del progetto prevede che al database siano progressivamente aggiunti i documenti dell’archivio in lineare B di Cnosso. Date le peculiarità di tale archivio, tale aggiunta costituirà un’ulteriore fase sperimentale, durante la quale saranno implementate soluzioni non ancora integrate nel sistema, come la ricerca delle co-occorrenze testuali (selezione delle epigrafi in funzione di due o più chiavi di ricerca testuali) e la visualizzazione dei microcontesti costituiti dagli elementi testuali immediatamente adiacenti a quelli ricercati.

MAURIZIO DEL FREO, FRANCESCO DI FILIPPO
Istituto di Studi sul Mediterraneo Antico
CNR-Roma

BIBLIOGRAFIA

- BARTHÉLEMY F. 1998, *A morphological analyzer for Akkadian verbal forms with a model of phonetic transformations*, in M. ROSNER (ed.), *Proceedings of the Workshop on Computational Approaches to Semitic Languages (Montreal 1998)*, Montreal, Université de Montréal, 73-81.
- BARTHÉLEMY F. 2009, *Une description morphologique structurée en arbre du verbe akkadien qui utilise des structures de traits et des transducteurs multirubans*, «*Traitement Automatique des Langues*», 50, 2, 173-99.
- BENKER S.C., LANGFORD R.P., PAVLIS T.L. 2011, *Positional accuracy of the Google Earth terrain model derived from stratigraphic unconformities in the Big Bend region, Texas, USA*, «*Geocarto International*», 26, 4, 291-303.
- BROVELLI M.A., MINGHINI M. 2012, *Georeferencing old maps: A polynomial-based approach for Como historical cadastres*, «*e-Perimtron*», 7, 3, 97-110.
- BUCCELLATI G. 1977, *The Old Babylonian Linguistic Analysis Project: Goals, procedures and first results*, in A. ZAMPOLLI, N. CALZOLARI (eds.), *Computational and Mathematical Linguistics. Proceedings of the International Conference on Computational Linguistics (Pisa 1973)*, Firenze, Olschki, 385-404.
- BUCCELLATI G. 2011, *Digital edition and graphemic analysis of the Ebla texts*, in L. MILANO (ed.), *Archivi Reali di Ebla, Edizione Digitale, 1*, *Cybernetica Mesopotamica*, CD 4, Malibu, Undena Publications.
- CALZOLARI N., LENCI A. 2004, *Linguistica computazionale: strumenti e risorse per il trattamento automatico della lingua*, «*Mondo Digitale*», 2, 56-69.
- DARCQUE P., ROUGEMONT F. c.s., *Palaces and ‘Palaces’: Mycenaean texts and contexts in the Argolid and neighbouring regions*, in *Proceedings of the International Conference Mycenaean up to Date: The Archaeology of the North-Eastern Peloponnese. Current Concepts and New Directions*, Athens, Swedish Institute at Athens.
- DEMAKOPOULOU K., DIVARI-VALAKOU N., SCHALLIN A.-L., EKROTH G., LINDBLOM A., NILSSON M., SJÖGREN L. 2002, *Excavations in Midea 2000 and 2001*, «*Opuscula Atheniensia*», 27, 27-58.

- DJINDJIAN F. 2008, *Webmapping in the Historical and Archaeological Sciences. An Introduction*, in F. DJINDJIAN, H. NOIZET, L. COSTA, F. POUGET, *Webmapping dans les sciences historiques et archéologiques, Actes du Colloque international (Paris 2008)*, «Archeologia e Calcolatori», 19, 9-16.
- FDC IV = ARAVANTINOS V.L., DEL FREO M., GODART L., SACCONI A. 2005, *Thèbes. Fouilles de la Cadmée. IV. Les textes de Thèbes (1-433). Translittération et tableaux des scribes*, Pisa-Roma, IEPI.
- HARRIS T.M. 2010, *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, Bloomington, Indiana University Press.
- KENNEDY S.J. 2012, *Transforming Big Data into Knowledge: Experimental Techniques in Dynamic Visualization*, Ph.D. Diss., MIT, Cambridge, MA.
- KT5 = KILLEN J.T., OLIVIER J.-P. 1989, *The Knossos Tablets. Fifth Edition*, Salamanca, Ediciones Universidad de Salamanca.
- MACKS A. 2002, *Parsing Akkadian verbs with Prolog*, in M. ROSNER, S. WINTNER (eds.), *Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic Languages*, Philadelphia, Association for Computational Linguistics, 3-8.
- MAURIZIO M., ORSINI R. 2010, *A model and a language for large textual databases*, in S. BERGAMASCHI, S. LODI, R. MARTOGLIA, C. SARTORI (eds.), *Proceedings of the Eighteenth Italian Symposium on Advanced Database Systems, SEBD 2010*, Bologna, Esculapio Editore, 254-265.
- POTERE D. 2008, *Horizontal positional accuracy of Google Earth's high-resolution imagery archive*, «Sensors», 8, 7973-7981.
- POWELL G. 2006, *Beginning Database Design*, New York, John Wiley & Sons.
- PTT I = BENNETT E.L. Jr., OLIVIER J.-P. 1973, *The Pylos Tablets Transcribed. Part I: Texts and Notes*, Roma, Edizioni dell'Ateneo.
- RIEGER A., BÖSER W. 1990, *Tiryms. Forschungen und Berichte, Band XI*, Mainz am Rhein, Philipp von Zabern.
- RUDOLPH W. 1973, *Die Nekropole am Probitis Elias bei Tiryms*, in *Tiryms. Forschungen und Berichte, Band VI*, Mainz am Rhein, Philipp von Zabern.
- SMITH S.P. 2007, *Hurrian Orthographic Interference in Nuzi Akkadian: A Computational Comparative Graphemic Analysis*, Ph.D. Diss., Harvard.
- TITHEMY = MELENA J.L., OLIVIER J.-P. 1991, *The Tablets and Nodules in Linear B from Tiryms, Thebes and Mycenaenae*, Salamanca, Ediciones Universidad de Salamanca.
- TREUIL R., DARQUE P., POURSAT J.-CL., TOUCHAIS G. 2008, *Les civilisations égéennes du néolithique et de l'âge du bronze, Deuxième édition refondue*, Paris, Presses Universitaires de France.

ABSTRACT

This article focuses on the preliminary results of a CNR-ISMA ongoing project for the digital edition of Linear B texts, having the ultimate goal of providing scholars, and all those who are interested in the Mycenaean world, with an updated edition of these documents. LiBER (Linear B Electronic Resources) is a document management system which is able to process a variety of materials, such as the logo-syllabic script preserved by these ancient records and their physical supports, as well as to project all relevant data into a dynamic archaeological map. In particular, LiBER has been designed to manage structured texts and all the information available about their chronology, paleography and spatial distribution. The purpose of this article is to illustrate the general philosophy which lies behind the conception of this kind of enterprise and the solutions adopted for the encoding of this specific logo-syllabic script – by exploring drawbacks and potentials of descriptive markup languages and a database driven approach – and for the representation of data through dynamic maps.