

## **Considerazioni introduttive**

Nicola Palazzolo

(Università degli Studi di Perugia, Italia)

### **1.**

Sono trascorsi tre anni dal convegno di Firenze, *Diritto romano e scienze antichistiche nell'era digitale*,<sup>1</sup> nel quale si sono gettate le basi, se non di una collaborazione, quanto meno di una conoscenza tra ricercatori che – su sponde disciplinari diverse, conseguenza di una sciagurata divisione in rigidi settori scientifico-disciplinari – condividono però non solo lo stesso oggetto di studio, l'Antichità classica, e le stesse fonti, esaminate da punti di vista diversi, ma anche gli strumenti tecnologici utilizzati, le tecniche di reperimento e di trasmissione delle fonti, gli strumenti informatici per l'analisi dei testi, la ricostruzione delle parti mancanti, l'attribuzione di paternità.

Oggi siamo qui per vedere se è cambiato qualcosa, se si è fatto qualche passo in avanti, oltre quello, innegabile, di una maggiore consapevolezza della necessità di standard comuni, e quindi trasferibili ad altre ricerche e ad altre fonti, che può farci uscire finalmente da quella sorta di condanna che si portano dietro le scienze umanistiche, quella concezione proprietaria che impedisce qualsiasi cooperazione tra i vari sistemi.

Lo scopo del convegno è duplice: dar conto delle due ricerche PRIN, che hanno dato occasione all'organizzazione del convegno, quella che fa capo all'amico Paolo Mastandrea, e quella che moralmente fa capo ancora a me, ma che per ragioni anagrafiche è stata coordinata negli ultimi anni dal mio allievo Francesco Arcaria. Dar conto allora di queste due ricerche non tanto dal punto di vista dei risultati ottenuti nei due specifici ambiti disciplinari, ma piuttosto da quello delle metodologie digitali utilizzate per raggiungere questi risultati. Accanto a questo primo obiettivo il secondo è verificare se l'auspicio di tre anni fa, appunto di un inizio di collaborazione tra studiosi che spesso neppure si conoscevano, pur avendo notevoli punti di convergenza, ha avuto un seguito, nel senso che si è iniziato un percorso comune tra ricercatori provenienti da discipline apparentemente diverse.

---

<sup>1</sup> Gli Atti del convegno sono stati pubblicati in *Diritto romano e scienze antichistiche nell'era digitale = Convegno di studio* (Firenze 12-13 settembre 2011). Torino: Giappichelli, 2012.

Per il primo profilo, e limitandomi alla ricerca da me guidata, BIA-Net (Bibliotheca Iuris Antiqui in rete) è un risultato che va oltre l'utilità specifica per chi fa ricerca nel diritto romano. Non solo si è prodotto uno strumento di cui possono servirsi tutti gli studiosi di scienze antichistiche (testi filologicamente controllati e uniformati all'edizione critica comunemente accettata, bibliografia multidisciplinare e dotata di sofisticati strumenti semantici, un Thesaurus che si arricchisce continuamente - già oggi conta oltre 10.000 termini controllati, articolati in un albero classificatorio multilingue).

Tutto ciò sarebbe stato possibile anche con i CD-Rom, come lo avevamo fatto finora, ma lo si è fatto in linea, utilizzando tutte le codifiche e gli standard propri delle tecnologie digitali (XML come linguaggio di trasmissione, TEI per la struttura dei dati, Unicode per i caratteri), quindi standard aperti e dati esportabili e riutilizzabili in altri ambiti per ricerche diverse. È stata dura, sono occorsi parecchi anni di lavoro (rispetto all'ultima edizione su CD-ROM sono passati 12 anni, scontando le richieste pressanti e i malumori di chi in tutti i convegni ci chiedeva quando sarebbe uscita la nuova edizione), ma oggi possiamo dire con soddisfazione di aver resistito alle sirene, e di avercela fatta. Rimangono da risolvere alcuni problemi organizzativi relativi alla diffusione dei dati, ma penso che nei prossimi mesi potremo chiudere il lavoro. Daria Spampinato e Marco Giunti entreranno nei particolari relativamente a due degli aspetti della ricerca, che vanno proprio nella direzione dell'interoperabilità con altri archivi digitali.

Per il secondo profilo già l'organizzazione del convegno credo sia un notevole risultato. La collaborazione tra me, con i miei allievi sparsi per l'Italia, e Paolo Mastandrea con la sua scuola, con cui ci siamo conosciuti proprio a Firenze in quel primo convegno, è il frutto più tangibile di quello spirito che ci animò allora. E basta una scorsa al programma del convegno per accorgersi del passo avanti che è stato fatto: non più una vetrina delle cose fatte, allo scopo di farsi conoscere, ma una messa a fuoco del problema della collaborazione nelle sue varie sfaccettature.

Sul piano più strettamente dei progetti di ricerca un primo tentativo è stato fatto con il progetto europeo coordinato da Maurizio Lana di un corpus globale virtuale di testi latini digitalizzati, dalle origini al Rinascimento, in pratica una *digital library* del latino, nella quale si interrogano i diversi database come se fossero uno solo e si accede a tutti da un'interfaccia unica. Avevano dato la loro disponibilità al progetto diverse basi di dati: *DigilibLT* anzitutto, con i testi letterari in prosa latina dal II al VII sec. d.C.; *Musisque Deoque*, con la poesia latina dal III sec. a.C. al Rinascimento; la *Bibliotheca Iuris Antiqui*, con il corpus completo di testi giuridici latini e greci dalle origini a Giustiniano; e infine DIGIMED, con i testi medievali di argomento storico-geografico, insomma una collezione già così pressoché completa (circa 13 milioni di parole), e con prospettive di ulteriore allargamento, del patrimonio testuale della latinità.

Il progetto è stato presentato per due volte in occasione di bandi europei, e si è collocato in buona posizione, anche se alla fine non è stato finanziato. Io vorrei spingere a riprovare ancora, magari correggendo e migliorando il progetto nei punti che appaiono più deboli, ma a non rinunciare, perché la strada è questa. E comunque, al di là del finanziamento europeo, la strada della collaborazione e di integrazione tra i vari progetti di digitalizzazione si impone, e richiede da noi qui ed oggi, iniziative concrete che vadano in questa direzione. Proprio per questo abbiamo affidato la relazione di apertura al prof. Gregory Crane, che a livello mondiale dirige la più importante iniziativa di collaborazione per le lingue classiche, che è il Perseus Project, e che voglio ringraziare per la disponibilità, ma abbiamo pure studiosi come Franco Niccolucci o Silvia Orlandi, che stanno lavorando all'interno di gruppi di lavoro che fanno capo a Europeana, la grande biblioteca digitale europea.

## 2.

Quando parliamo di strumenti per la ricerca non possiamo ignorare che i principali strumenti di cui ci serviamo per la ricostruzione delle civiltà classiche sono i documenti scritti, sono quelli che chiamiamo 'testi'.

Abbiamo presente certamente che tutta una serie di altre ricerche si basa su documenti non testuali, della cosiddetta 'cultura materiale' (prendiamo ad esempio tipico l'archeologia), e che di questo tipo di documenti dobbiamo pur occuparci quando parliamo di applicazioni informatiche. Ma non c'è dubbio che la massa dei dati che noi trattiamo sono dati testuali, contenuti in supporti diversi (manoscritti, iscrizioni, papiri, ostraka, monete, ecc.), ma pur sempre testi.

E però occorre ripetere qui qualche distinzione, come ho già fatto nel precedente convegno, ma che voglio ribadire perché mi sembra non ancora chiara sino in fondo.

Cos'è un testo? È una successione di segni grafici, una successione di parole, è un oggetto fisico (libro, manoscritto, papiro) riprodotto in una serie di immagini digitali, o è un contenuto che l'autore ci vuole trasmettere, e che vogliamo analizzare a fini di studio e di ricerca? Sono tante le accezioni che possiamo dare all'espressione, e certo non voglio invadere il mestiere di altri facendo il tentativo di dare una risposta a questa domanda. Ma è chiaro che ciascuna di quelle accezioni cui prima accennavo dipende strettamente dal tipo di utilizzo che se ne intende fare.

Dico questo perché nell'ultimo decennio il tema della digitalizzazione del patrimonio testuale dell'umanità, attraverso la creazione di vere e proprie 'biblioteche digitali' si è progressivamente imposto come uno dei più significativi prodotti dell'applicazione delle nuove tecnologie digitali. C'è stato un grosso *battage* pubblicitario, anche commerciale, sulla velo-

cità di digitalizzazione dei grandi patrimoni bibliotecari. Ma spesso per i non addetti ai lavori una cosa non è chiara, che spesso invece è il limite principale di queste iniziative: il fatto che l'opera digitalizzata sia quasi sempre il supporto fisico, il libro inteso come bene culturale, da proteggere e conservare come tale, nella sua fisicità, non il suo contenuto, da mettere a disposizione dei ricercatori, e capace quindi di essere utilizzato all'interno di un'attività di ricerca. E ciò perché, nella gran parte dei casi, quando si parla di digitalizzazione di grandi patrimoni testuali, quasi sempre si tratta di immagini fotografate attraverso sistemi più o meno avanzati, magari con un sistema di indici che rimandano alle singole pagine in cui quell'argomento è trattato, ma senza che il testo in quanto tale possa essere trattato e analizzato sia dal punto di vista linguistico che da quello filologico.

Per rimanere dentro il mio settore disciplinare è stato con viva curiosità che sono andato a guardare nel sito dell'Università di Bologna un oggetto che, sotto il pomposo titolo di *Corpus Iuris online*<sup>2</sup> proponeva nient'altro che una collezione di immagini tratte da una prestigiosa edizione della Glossa accursiana, corredata da indici di riferimento: una bella iniziativa, non c'è che dire, sarà anch'essa una biblioteca digitale ma tutti comprendiamo che è cosa diversissima da un testo digitale, che può essere non solo ricercato, ma analizzato, manipolato a fini diversi.

Occorre che ci diciamo che ormai si tratta di strumenti certamente utili, ma utili a scopi diversi da quelli dell'analisi testuale. La biblioteca digitale di opere in formato immagine non può essere l'obiettivo principale da raggiungere, ma solo un primo passo, sul quale poi fondare un'edizione digitalizzata non di un'immagine, ma di un testo scritto, che sia agevolmente catturabile e manipolabile dal ricercatore stesso. In questa direzione sta appunto lavorando l'*Open Greek and Latin Project* dell'Università di Lipsia.

### 3.

Dicevo che occorre andare avanti verso l'interoperabilità dei nostri archivi e mettere in comune i nostri sforzi per raggiungere questo obiettivo. Abbiamo un patrimonio digitale notevolissimo, ma non tutto è allo stesso livello di elaborazione. Allora lo sforzo deve essere duplice:

- portare tutti i testi che abbiamo a uno standard comune, sia dal punto di vista formale (XML, TEI, ecc.), sia da quello sostanziale (correttezza filologica attestata dalla conformità a un'edizione critica accettata). So che è difficilissimo, che richiede tanta fatica e tempi non brevi.

---

2 URL <http://amshistorica.unibo.it> (2017-10-19).

Richiede il lavoro dell'informatico per la conversione del formato, ma richiede anche il lavoro dello specialista, che scelga l'edizione di riferimento e controlli il testo parola per parola dando il giusto valore ai segni diacritici utilizzati in quell'edizione per uniformarli a quelli dello standard prescelto. È il lavoro che abbiamo fatto per BIA-Net sui testi giuridici latini e greci: non solo il *Corpus Iuris giustiniano*, del quale ci sono in linea tantissime versioni anche testuali, ma delle quali si ignora quale sia l'edizione di riferimento, e quindi non attendibili, ma anche per tutte le fonti giuridiche pregiustiniane, nonché per le fonti greche, anch'esse uniformate agli standard TEI (e devo ringraziare per queste l'apporto che ci è stato dato dall'Istituto di Linguistica computazionale del CNR di Pisa, sia per il testo greco delle *Novellae*, sia per quello delle costituzioni greche del *Codex Iustinianus*, entrambi testi del tutto assenti online). Lo stesso discorso che facciamo per le fonti dobbiamo farlo per le basi di dati bibliografiche. È oggi ormai un risultato acquisito la possibilità di interrogazione contemporanea di molte basi di dati, e proprio nelle ricerche interdisciplinari, quali ormai sono e dovranno essere sempre più le nostre, il vantaggio è quanto mai evidente, ma anche qui i risultati possibili sono tanto più soddisfacenti in quanto tali basi di dati hanno conosciuto in fase di analisi concettuale e progettuale una normalizzazione del linguaggio di rappresentazione. Ciò significa per i produttori di banche dati bibliografiche la necessità di convertire in un linguaggio standardizzato (XML) tutti i metadati descrittivi delle singole unità bibliografiche presenti nella banca dati.

- Rendere esportabili e interoperabili le nostre banche dati, cioè garantire il riuso dei materiali digitalizzati senza vincoli particolari. Abbandonare quindi ogni residua concezione proprietaria, e affidarci interamente a tecnologie *open source* e *open access*, ovviamente con un giusto riconoscimento, e laddove possibile un'equa remunerazione, del lavoro svolto.

La prima sessione del convegno dovrà servire a farci capire a che punto siamo in questa strada.

#### 4.

Ma mettere in comune le risorse non è solo un problema tecnico e di standard. Il passaggio successivo è quello di avere strumenti più raffinati che ci consentano di reperire materiali che con i normali strumenti interni alle banche dati non si riuscirebbe a ottenere. Ecco che si apre tutto un campo di ricerca nel quale – almeno per restare al nostro campo – siamo molto indietro. Vi sono – è vero – diversi strumenti di cui ci serviamo ai fini della

ricerca nelle nostre banche dati, ma sono per lo più strumenti creati specificamente per un corpus spesso molto settoriale. Occorre chiedersi invece se non sia possibile andare alla ricerca di uno strumento che abbia da un lato sufficiente flessibilità per essere adattato alle esigenze di ricerca in uno specifico ambito scientifico, ma dall'altro che sia in grado di fare una ricerca ad ampio spettro entro corpora eterogenei, quali sono i nostri. Il web semantico e le ontologie dovrebbero andare in questa direzione. Cito solo l'esperienza di BIA-Net, nella quale – ed è stato uno dei punti di forza del progetto – si è creata un'apposita unità di ricerca (fatta non di giuristi né di informatici o di documentaristi, ma di filosofi, per la precisione di Logici) per studiare le possibilità che il vecchio *Thesaurus* di BIA (che pur si giovava di un patrimonio di oltre 10.000 termini strutturati entro una classificazione a 6 livelli di approfondimento) potesse non solo arricchirsi di nuovi termini ma incrementarsi con nuovi legami per essere utilizzabile nell'ambito di qualunque delle altre discipline antichistiche. Ma si parlerà anche di altri strumenti: Wordnets per le lingue classiche, ad esempio, è uno degli strumenti semantici che potrebbero essere utilizzati in tutti gli ambiti disciplinari delle scienze del mondo antico.

Ecco perché si è voluto dedicare un'apposita sessione del convegno agli strumenti semantici.

## **5.**

Peraltro la ricerca informatica in antichistica non è certo limitata alla creazione di corpora digitali più o meno ampi e agli strumenti avanzati per ricercare le occorrenze di un termine o di una frase nell'immenso corpus delle fonti antiche, o per ricercare la bibliografia collegata a un testo o anche solo a una frase. Dobbiamo andare oltre: e c'è qualcuno che già lo fa. Ad un certo punto sorge l'esigenza di poter, attraverso il computer, compiere non più una ricerca *dei* testi (per esempio, quelli che contengono un nome, una parola, o una sequenza di parole: le cosiddette concordanze), ma una ricerca *sui* testi, appunto l'analisi dei dati.

Allora vediamo fino a che punto le metodologie e anche gli strumenti tecnologici che alcuni di noi stanno sperimentando possono essere utilizzati anche da altri. Prendiamo il caso dell'attribuzione di paternità di un testo, prendiamo quello della ricostruzione di testi frammentari (generalmente papiri e iscrizioni), o infine quello della creazione di nuove edizioni critiche.

Quello dell'attribuzione di un testo disponibile nel mondo digitale è per esempio un campo nel quale molti studiosi potrebbero ritrovare più di un motivo di interesse. Nel precedente convegno a Firenze ha colpito la relazione di Maurizio Lana, che ha mostrato come, sulla base di metodi quantitativi, si possa tentare di individuare l'autore materiale di testi

legislativi imperiali in funzionari che abbiano ricoperto cariche di rilievo nella cancelleria imperiale, e di cui siano noti altri scritti.

La terza sessione del convegno sarà appunto dedicata a queste problematiche.

## 6.

E infine una parte che apparentemente è meno legata agli studi antichistici, ma che in realtà ci riguarda tutti. Quali le modalità organizzative e le forme giuridiche per un'impresa del genere, che non sia solo una mera attività di ricerca all'interno dei nostri studi e laboratori, ma si voglia proporre anche sul mercato? Quali i beni da tutelare e quali quelli da lasciare in uso libero da vincoli alla comunità scientifica? E da che cosa tutelarli? Quali i pericoli da cui difendersi? Gli attacchi degli hacker o la copiatura e l'esportazione abusiva dei dati? Tutela dal plagio o dalla contraffazione?

Da un altro punto di vista: si è parlato di tendenza all'open access; ma come fare a remunerare l'eventuale plus-valore che i metadati creati dall'attività dei ricercatori aggiungono al documento, che considerato nella sua essenzialità deve essere invece di pubblico dominio e ad accesso gratuito? Le due relazioni della quarta sessione cercheranno di rispondere a questi interrogativi.

## 7.

Chiuderà il convegno una Tavola Rotonda, nella quale vorremmo fare il punto della situazione su tutti questi problemi che, insieme, costituiscono la Biblioteca digitale, non un semplice corpus digitalizzato di documenti, un mero deposito nel quale andare a ricercare ciò che serve, ma piuttosto un complesso organizzato di servizi finalizzati all'utente-ricercatore. È possibile creare uno strumento simile per le scienze dell'Antichità, uno strumento che metta insieme corpora distinti e di differenti tipologie, anche se accomunati dagli stessi standard, ma anche strumenti adatti per le diverse tipologie di ricercatori, e organizzati in una struttura istituzionale, pubblica o privata, che ne consenta lo sfruttamento anche all'esterno delle comunità di ricerca che l'hanno creata? Io spero di sì, ed è questo il compito che vorrei affidare alle generazioni che ci seguono.

Il nostro, come dicevo, non vuole essere un convegno di presentazioni di cose fatte, una passerella di progetti realizzati o da realizzare. Ci saranno anche queste, ed è giusto farle conoscere, ma deve essere solo il punto di partenza. Il convegno deve essere anzitutto un confronto di idee, di proposte, di cose da fare insieme. Io mi auguro che da questo convegno

escano fuori i nuovi progetti per le scienze dell'Antichità, su cui cimentarci insieme nei prossimi anni.

Noi abbiamo fatto il primo passo, che forse era il più difficile: abbiamo fatto almeno il tentativo di dialogare tra noi, di scambiarci le esperienze. Adesso è il tempo di agire, e di agire insieme, perché le risorse pubbliche sono sempre di meno e quelle private per un campo di studio come il nostro richiedono menti illuminate che al momento non vediamo all'orizzonte.