

Da BIA su CD-Rom a BIANet: il lungo cammino di una banca dati fortunata

La *Bibliotheca Iuris Antiqui* (BIA), diffusa nella sua prima edizione su CD-rom nel 1994, e poi nella seconda (BIA2000) nel 2002¹, costituisce ancora oggi l'unico esempio di sistema informativo integrato dedicato allo studio dei diritti dell'antichità. Si tratta di un prodotto ormai talmente noto e apprezzato da parte dell'intera comunità scientifica internazionale, che non può non costituire la base di partenza per ogni ulteriore sviluppo di strumenti informatici ausiliari alla ricerca romanistica.

Il successo di BIA è intimamente connesso ad una decisione fondamentale assunta già nel 1994²: la ricerca romanistica porta con sé l'esigenza di individuare rapidamente, per ciascun tema di studio, tanto le fonti antiche pertinenti, quanto le elaborazioni dottrinali succedutesi nel tempo che a quelle fonti fanno riferimento. Per soddisfare tale esigenza, piuttosto che trasferire su distinti supporti informatici i tradizionali repertori bibliografici e i maggiori indici delle fonti, è risultato preferibile ricorrere ad un sistema integrato, che, gestendo insieme l'archivio delle fonti e quello della bibliografia di riferimento, consente di 'navigare' da un archivio all'altro, moltiplicando così – grazie all'adozione di strategie ipertestuali – le possibilità di reperimento di informazioni pertinenti rispetto al proprio tema di ricerca.

Un'ulteriore indicazione che emerge già dalle edizioni di BIA su CD-rom riguarda l'ideazione e la realizzazione dell'archivio *Thesaurus* (l'archivio dei codici di classificazione numerici e dei descrittori in linguaggio naturale utilizzati al fine di operare la descrizione semantica dei documenti presenti nell'archivio bibliografico): si tratta di uno strumento di ricerca concettuale inizialmente sottovalutato dagli utenti di BIA, ma che nell'uso avanzato è risultato utile sia allo specialista – che può avvalersi dell'indicizzazione al fine di ridurre il 'rumore' tipico delle ricerche soltanto testuali (si pensi, ad esempio, a chi, volendo reperire soltanto i riferimenti bibliografici in materia di diritto di proprietà, debba cercare le occorrenze della parola 'proprietà' nel titolo di ciascuna scheda bibliografica), nonché allo scopo di semplificare la ricerca tra documenti redatti in lingue, moderne ed antiche, diverse – sia,

1 *Bibliotheca Iuris Antiqui (BIA2000). Sistema informativo integrato sui diritti dell'antichità*. Direzione scientifica di Nicola Palazzolo (Catania, Torre, 2002) p. 102 + CD-Rom.

2 Il gruppo di ricerca originario, che ha dato luogo all'ideazione e poi alla realizzazione delle prime versioni della banca dati, era costituito dal suo direttore Nicola Palazzolo, nonché da Francesco Arcaria, Orazio Licandro, Lucio Maggio, Patrizia Sciuto, Antonella Mazzola. Ad essi si sono poi affiancati, nel corso degli anni, diversi altri collaboratori, nonché varie unità di ricerca periferiche presso Università o centri di ricerca (Perugia, Genova, Parma, Catanzaro, ITTIG/CNR). Il gruppo aveva sede presso il C.I.R. (Centro per l'Informatica romanistica) un centro di ricerca dell'Università di Catania, poi soppresso nel 2015.

soprattutto, al non esperto della materia (un laureando, ma anche un cultore del diritto vigente) che ha poca confidenza con i termini propri del linguaggio giuridico romano.

Le considerazioni di cui sopra avevano indotto l'originario gruppo di ricerca, già negli anni '80, a progettare un nuovo sistema informativo in linea, anch'esso di tipo integrato, composto da tre archivi: *Opera* (archivio bibliografico), *Fontes* (archivio delle fonti), *Thesaurus* (archivio dei termini controllati). Ciascun archivio, oltre a potere essere consultato separatamente dagli altri, è stato organizzato in modo tale da consentire una ricerca integrata, che permette cioè di passare da un archivio all'altro. Il successo del CD-Rom di BIA, sia in Italia che in tantissimi paesi, anche quelli in teoria molto distanti dal diritto romano, è stata la conferma migliore della bontà di quell'intuizione. E la riprova di questo successo è stata il fatto che ininterrottamente dal 1997 al 2009 i progetti biennali di ricerca collegati all'evoluzione della banca dati sono stati sempre approvati con giudizi estremamente lusinghieri, e finanziati dal MURST (poi MIUR) nell'ambito del Piano PRIN (Prodotti di rilevante interesse nazionale).

La gestazione di BIA-Net è stata al contrario lunghissima e per molti versi impreveduta. Non si trattava infatti solo di trasferire la tecnologia già utilizzata per il supporto ottico su un supporto a distanza, ma di realizzare un prodotto interamente nuovo, che rispondesse agli standards più avanzati della tecnologia online. E' peraltro noto che la trascrizione su supporto remoto dei testi delle fonti o della bibliografia ai fini della pubblicazione on-line comporta sempre un trattamento dei documenti da memorizzare notevolmente più complesso di quello necessario per la pubblicazione su CD-ROM. Comporta infatti la necessità di strutturare gli archivi secondo formati standard riconosciuti a livello internazionale per la digitalizzazione degli archivi testuali e bibliografici, tra i quali i più accreditati sono oggi quelli in linguaggio XML. La caratteristica del linguaggio XML è la separazione concettuale tra la rappresentazione logica dei dati e la loro visualizzazione tramite un media. Quindi lo scopo dell'utilizzazione dello XML è principalmente quello di rappresentare la struttura logica delle informazioni, per poi demandare a un software di presentazione le funzionalità proprie della visualizzazione.

Per quanto riguarda, ad esempio, l'archivio OPERA, ai fini dell'edizione on line, è stato necessario convertire in un linguaggio aperto i records bibliografici dell'archivio OPERA. A questo scopo è stato coinvolto un gruppo di ricerca operante presso il Dipartimento di Matematica e Informatica dell'Università di Catania³, che ha utilizzato lo standard XML Dublin Core, nella versione *qualified*, seguendo le specifiche della *Dublin Core Metadata Initiative*. Ciò ha comportato la predisposizione di svariati DTD (*Document Type Description*) che rappresentassero in maniera completa ed esaustiva i

³ Il gruppo di ricerca che ha progettato e sviluppato la prima versione on line della banca dati era formato, oltre che da vari dottorandi in Informatica dell'Università di Catania, da contrattisti CNR, che hanno lavorato in équipe, sotto la direzione di Daria Spampinato, tecnologo di ruolo del CNR.

vari tipi di documenti (monografie, raccolte di scritti, atti di convegni, articoli su riviste, ecc.).

Relativamente poi all'altro archivio, quello delle fonti giuridiche romane (FONTES) il sistema prescelto è stato la codifica XML/TEI (*Text Encoding Initiative*), che è divenuto uno standard internazionale *de facto* per la digitalizzazione dei testi in ambito umanistico (più precisamente lo standard XML/TEI P5). All'interno dello standard XML è stato poi necessario elaborare la struttura di ogni singolo documento attraverso una sua specifica DTD (*Document Type Definition*) che includesse una serie di elementi aggiuntivi concernenti la struttura fisica (volume, pagina, riga) e quella concettuale (libri, capitoli, paragrafi, frammenti) del testo.

Infine, nella versione on line del *Thesaurus* si è scelto, rispetto all'edizione su CD-Rom, di utilizzare gli standards correnti per la realizzazione dei Thesauri in linea, e cioè indicare le principali relazioni, sia quelle gerarchiche verticali (Broader Term, Narrower Term), rappresentate dalle classi dello schema di classificazione (es.: diritti di obbligazione > contratti verbali > *stipulatio*), sia quelle di affinità orizzontali (Related Term) (es.: alla classe 'Epoca repubblicana. Singoli reati' fanno capo i concetti: *sacrilegium*, *calumnia*, *conspiratio*, ecc.), ed infine quelle di sinonimia, antinomia, variante ortografica o di scrittura, preferenza linguistica (UF/USE) (es.: *XII Tabulae* = *lex XII tabularum*; *matrimonium* = *nuptiae*; *adstipulator* = *adpromissor*). Il formato prescelto (SKOS/RDF) consente il riuso dei suoi dati aperti e agevola l'interoperabilità con altri strumenti di indicizzazione semantica, con enciclopedie, dizionari, etc.

Come si vede, quindi, un lavoro di grande impegno, che ha coinvolto professionalità le più diverse, e che soltanto dopo molti anni è riuscito a prendere forma, dopo che la banca dati è stata incorporata nel prestigioso sistema informatico dell'Istituto di Teoria e tecniche dell'Informazione Giuridica del CNR, con sede a Firenze⁴.

Profondamente rinnovati sono stati, peraltro, anche i contenuti della Banca dati BIA-Net – nella forma e nella sostanza – rispetto a quelli, pur già innovativi, del CDRom BIA2000.

L'archivio bibliografico (*Opera*) di BIA, che nell'edizione su CD-Rom del 2002, andava dal 1940 al 1998, nell'edizione on line BIA-Net presenta ad oggi una copertura temporale che si estende fino al 2005, per un complesso di circa 50.000 opere tra monografie, raccolte di scritti, testate di riviste romanistiche, articoli, che si occupano di diritto romano e antico, in parte tratti da banche dati in linea o da repertori bibliografici specializzati (in particolare

4 Sono estremamente grato al Direttore pro-tempore dell'Istituto, dott. Sebastiano Faro, oltre che per la lungimiranza dimostrata nell'accogliere BIA-Net tra le prestigiose banche dati prodotte dall'ITTIG (ora IGSG), per aver autorizzato i tecnici dell'istituto a compiere le verifiche e operare le correzioni necessarie per la messa in linea della banca dati. Notevole è stato il lavoro compiuto dai tecnici dell'IGSG/CNR per sistemare vari malfunzionamenti o parti mancanti del programma. Non si sarebbe giunti a questo risultato senza la preziosa collaborazione degli informatici dell'Istituto, tra i quali mi piace ricordare il dott. Fabrizio Turchi e il dott. Gerardo Giardiello.

IVRA), ed in una parte maggiore memorizzati manualmente. L'aggiornamento della base bibliografica procede regolarmente, sia pure con una qualche lentezza, dovuta al fatto che l'indicizzazione dei singoli documenti viene fatta ancora manualmente da esperti della materia, anche se è allo studio presso il CNR un sistema che consentirebbe di automatizzare parzialmente, e quindi di velocizzare, il lavoro di raccolta e di indicizzazione. Si è ritenuto tuttavia, allo scopo di non privare gli utenti di un importante supporto alla ricerca, di aggiornare comunque la base bibliografica con i dati essenziali delle singole pubblicazioni, integrandoli via via nel tempo con i metadati aggiunti dal gruppo di ricerca.

Per quanto concerne l'archivio FONTES, in previsione dell'edizione on line già il vecchio gruppo di ricerca, convinto della necessità che un'edizione digitale delle fonti dovesse garantire anzitutto la correttezza filologica del testo digitale, e specialmente la conformità ad un'edizione critica di riferimento, aveva – sia pure per una parte della base testuale di BIA - abbandonato i vecchi modelli di cui si era avvalso sin allora. Già nella seconda edizione di BIA su CD-ROM, BIA2000 (2002) era stato perciò avviato un profondo lavoro di 'restauro filologico', che ha portato alla revisione integrale del testo elettronico del Digesto giustiniano, che è stato reso conforme all'*editio minor* di Mommsen-Krüger (più precisamente l'*editio stereotypa* XII, Berolini 1911). Il lavoro di restauro è continuato negli anni successivi (dopo il 2002) da parte di diverse unità di ricerca, con la revisione integrale (con l'eccezione delle costituzioni in lingua greca), effettuata da una delle unità del gruppo (presso l'Università di Perugia), del testo del Codex Iustinianus, che è stato reso conforme a quello dell'*editio stereotypa* X, Berolini 1929, a cura di Paul Krüger, di quello delle Istituzioni giustiniane, reso conforme all'edizione XII di Mommsen-Krüger, e di quello del Codex Theodosianus, reso conforme all'edizione di Mommsen-Meyer; in tempi più recenti è stato acquisito il testo delle Novelle giustiniane, reperito presso l'ILC/CNR di Pisa in un vecchio formato elettronico degli anni '70, e convertito, previo restauro filologico (effettuato, grazie ad una convenzione, presso l'Università di Cagliari), in un formato compatibile con i nuovi programmi di *information retrieval on line*. Presso il gruppo di ricerca originario alcuni collaboratori esterni hanno provveduto poi al restauro filologico delle Istituzioni di Gaio, reso conforme all'edizione oggi più accreditata (Nelson), nonché delle altre fonti giuridiche pregiustiniane, secondo l'edizione contenuta in *Fontes Iuris Romani* vol. II. Infine un importante lavoro di integrazione è stato realizzato a Pisa presso l'ILC/CNR mediante la lettura ottica con vari strumenti software delle costituzioni greche del Codex Iustinianus⁵, che fino ad allora erano assenti nell'edizione contenuta in BIA2000, e che oggi sono visibili solo su BIA-Net, perché non esistono in nessun'altra raccolta digitale, né in linea né su CD-ROM.

5 Sono grato in particolare al dott. Federico Boschetti, ricercatore presso l'ILC di Pisa, per il suo prezioso contributo, tanto riguardo alla conversione del vecchio testo delle Novellae, quanto per la lettura ottica delle costituzioni greche del Codex giustiniano.

Un sistema di navigazione interno all'archivio *Fontes*, appositamente realizzato presso l'Istituto ISTC/CNR di Catania, consente poi di collegare il singolo documento recuperato (generalmente frammento o paragrafo) a quelli che lo precedono e lo seguono; cosicché è possibile 'scorrere' in avanti o all'indietro lungo tutto l'archivio al fine di potere leggere il testo recuperato all'interno del contesto di riferimento. Per realizzare ciò, si è dovuto disporre i documenti dell'archivio *Fontes* in una sequenza che non fosse rigidamente alfanumerica, bensì alfanumerico-logica (D.1.2.2.10 segue, com'è logico, D.1.2.2.9 e non D.1.2.2.1, come sarebbe avvenuto in una sequenza alfanumerica).

Infine, il *Thesaurus Iuris Romani*, uno strumento unico nel panorama, assai ampio, dei Thesauri specialistici, e la cui utilità si era già rilevata decisiva nel collegamento funzionale con l'archivio bibliografico di BIA, è stato in questi ultimi anni arricchito e migliorato nella sua struttura con l'apporto delle competenze specialistiche del Gruppo "Soggettario" della Biblioteca Nazionale di Firenze, a seguito di specifica convenzione con l'ITTIG/CNR⁶. Ad oggi (ma esso si arricchisce ogni giorno) il Thesaurus consta di oltre 10.000 termini tra descrittori e non descrittori, mentre attende di essere presto messa in linea una nuova classe, la n. 9, relativa alla Geografia del mondo antico.

⁶ Sento il dovere di ringraziare la dott.ssa Anna Lucarelli, responsabile del Nuovo Soggettario della Biblioteca Nazionale di Firenze, e la sua validissima collaboratrice dott.ssa Elisabetta Francioni per l'intelligente apporto di consigli e suggerimenti.

