

# Dig *Italia*

Anno V, Numero 2 - **2010**

Rivista del digitale nei beni culturali

ICCU-ROMA

# Imago et umbra

## Programma di digitalizzazione per l'Archivio storico della Pontificia Università Gregoriana: criteri, metodi e strumenti

**Andrea Bozzi – Martín Maria Morales – Marco Rufino**

*Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa – Archivio storico della Pontificia Università Gregoriana – Fondazione rinascimento digitale*

### Criteri<sup>1</sup>

Il titolo di questo progetto fa riferimento a un'espressione che si trova all'interno del *corpus* degli scritti di Nicolò Cusano («Sermo LXI, [...] Vita igitur istius mundi non est vita, sed imago et umbra vitae verae; ita de sapientia et prudentia et laetitia et singulis aliis») con la quale vengono indicate le realtà di questo mondo che, sebbene da una parte rappresentano (*imago*) nel contempo non lo fanno in modo totale e assoluto (*umbra*). In maniera analoga il progetto che qui si presenta, pur portando alla luce, in immagini, i documenti, non potrà mai esaurirne i contenuti.

L'Archivio della Pontificia Università Gregoriana (APUG)<sup>2</sup> possiede un patrimonio che testimonia l'attività intellettuale dei professori gesuiti del Collegio Romano, dalla sua fondazione nel 1551 alla soppressione della Compagnia di Gesù ed ancora dal 1824 all'Unità d'Italia. Questo insieme si presenta come esempio unico di raccolta documentaria attraverso la quale è possibile ricostruire la storia della didattica di una delle università tra le più celebri nell'Europa dei secoli XVI-XVII.

I fondi manoscritti oggi conservati presso l'APUG, provenienti dal patrimonio librario del Collegio Romano, sono il risultato di un complesso *iter*, non ancora del tutto concluso, di conservazione, nascondimento e restituzione. Nell'antico Collegio esistevano diverse "librerie", secondo i diversi corsi d'insegnamento: vi era una *bibliotheca* per la teologia, una per la filosofia, una terza per la retorica ed infine una quarta per gli studi inferiori. Oltre a queste quattro ve n'era però anche

<sup>1</sup> Martín Maria Morales (Archivio storico della Pontificia Università Gregoriana).

<sup>2</sup> Per la storia dell'archivio, descrizione del posseduto e i progetti in corso si veda il sito: <http://www.archiviopug.org>.

una quinta costituita da fondi provenienti da importanti lasciti testamentari: la *Bibliotheca maior*, o anche *secreta*<sup>3</sup>. In seguito alla soppressione della Compagnia di Gesù nel 1773 i fondi delle biblioteche paiono essere rimasti inalterati fino al ritorno dei Gesuiti nel 1824, mentre alcuni manoscritti insieme a codici, libri rari ed oggetti preziosi furono murati in un ripostiglio di cui, durante il secolo successivo, si conservò solo un vago ricordo. Fu in seguito all'emanazione della legge di soppressione delle Corporazioni religiose del 19 giugno 1873 che iniziò l'incameramento dei beni delle biblioteche dei diversi ordini. Nella denuncia del 1877 di Bartolomeo Podestà, primo bibliotecario della Biblioteca nazionale centrale di Roma (Bncr), si dà notizia della scoperta del ripostiglio. Tra i documenti rinvenuti, quelli considerati di scarso interesse, furono collocati in una soffitta dello stesso Collegio Romano sede della nascente Bncr. Nel 1948 la biblioteca, trovandosi nella necessità di liberare questa soffitta, decise di restituire il fondo al Generale della Compagnia.

Uno dei progetti intrapresi dall'APUG riguarda l'analisi dello stratificato processo, storico e culturale, che ha portato alla dispersione dei fondi gesuitici così come quelli di altri ordini religiosi. Questa ricerca implica una ricostruzione della composizione originaria dei fondi delle biblioteche del Collegio Romano a partire dai fondi gesuitici oggi conservati presso la Bncr, l'Archivio di Stato di Roma (ASR), la Biblioteca apostolica vaticana (BAV) e l'Archivum romanum societatis Iesu (ARSI). Ovviamente lo strumento principe per avviare questa e ogni altro tipo di ricerca è il catalogo. Dagli anni Quaranta fino al 2008 i ricercatori dell'APUG, per le intricate vicende che hanno interessato l'archivio, sono stati pochi. I cambiamenti di sede, via via più inaccessibile e disagiata, rappresentano, in parte, la caduta dell'interesse per un tipo di materiale che non alimenta più né l'insegnamento né la ricerca. Gli studiosi, qualora fossero riusciti ad accedere all'archivio, incontravano notevoli difficoltà, limitati da strumenti quali un inventario topografico incompleto e inesatto. Solo ad alcuni importanti manoscritti, grazie al contributo di singoli ricercatori, sono stati dedicati studi che, spesso, purtroppo non sono sfociati in pubblicazioni. Ma, come si vedrà, la vita di un archivio dipende anche dalla sua ricerca.

La limitatezza delle risorse economiche e umane per affrontare gli interventi di conservazione e valorizzazione dell'APUG ha condizionato fortemente le scelte operative degli ultimi due anni. Ad ogni modo, il cammino intrapreso si è manifestato vincente. L'impossibilità di acquistare un software per la catalogazione ha permesso:

- di avviare una collaborazione con l'Istituto centrale per il catalogo unico (ICCU)

<sup>3</sup> Cfr. Costanzo Bizzocchi, *La Biblioteca Segreta del Collegio Romano*, «Gesuiti della Provincia Romana», n. 3, 1971, p. 17.

- e l'acquisizione di MANUS integrando così l'APUG nel progetto di censimento nazionale dei manoscritti;
- di intraprendere un proficuo rapporto con la Fondazione rinascimento digitale di Firenze e con l'Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa (ILC) che consente l'elaborazione di un modulo specifico di Pinakes (d'ora in avanti citato con la sigla PK<sup>4</sup>) e Pinakes Text (PKT) per la realizzazione di un archivio digitale.

Inoltre, l'esiguità dei fondi investibili per la realizzazione di un laboratorio di digitalizzazione ha obbligato a pensare ad attrezzature meno costose, favorendo scelte che risultano, non solo più economiche, ma soprattutto capaci di rispondere alle sfide di un materiale che mal si presta ad operazioni di scansione massiva. Infine, la mancanza di personale specializzato ha aperto le porte dell'Archivio a tirocinanti provenienti principalmente dell'Università degli studi di Roma "La Sapienza". Grazie a ciò, l'APUG, inserito in una struttura universitaria, non è più solo un centro di documentazione ma anche un centro di formazione.



Nell'ambito degli archivi, in particolare di quelli ecclesiastici, si è passati da un sistema di "segretismo", in parte provocato dal confronto ideologico che dalla metà del XIX secolo si estese fino alla caduta del muro di Berlino, a un sistema sempre più aperto. Se prima la documentazione venne, in molti casi, celata, oggi si registra una disponibilità di documentazione che non sempre contribuisce ad accrescere le conoscenze. Il tipo di *segretismo* che si perpetua ancora oggi non è più legato alla volontà dell'archivista o dell'istituzione ma provocato da strumenti di ricerca inadatti che possono deviare lo studio, da un'assenza di dialogo tra archivisti e ricercatori e dalla carenza di politiche di conservazione che spesso impediscono la consultazione.

Figura 1. Sforza Pallavicino, *Metaphisica*, metà XVII sec. (APUG 390). Illustrazione del celebre calligrafo Galienus Conocius

<sup>4</sup> Per informazioni dettagliate su PK e PKT: <http://Pinakes.imss.fi.it>; <http://Pinakes.imss.fi.it/p3wiki>.

Per valorizzare queste testimonianze documentarie e consentire alla comunità scientifica di accedere a un patrimonio ancora in gran parte inedito l'APUG ha avviato diversi progetti: dalla basilare realizzazione di un OPAC a quella di un Archivio digitale. Aderendo al progetto di censimento nazionale MANUS online<sup>5</sup> dell'ICCU è iniziato nel 2008 il processo di catalogazione dei fondi storici dell'APUG. Questa costituisce la prima esperienza di realizzazione di un OPAC, relativo al materiale documentario manoscritto, all'interno della Compagnia di Gesù. I fondi attualmente in fase di inserimento a catalogo sono: il Fondo Curia, contenente circa 2.380 volumi (sec. XVI-XIX), restituito alla Compagnia di Gesù dalla Bncr nel 1948; il Fondo APUG, costituito da 3.271 volumi (sec. XVI-XIX), in parte miscelanei e provenienti dal Collegio Romano e da precedenti sedi dell'Università fino al 1946; il Fondo Collegio Romano, con testimonianze delle vicende istituzionali dell'Università nel XIX secolo. Grazie alla possibilità data dal *software* di gestire diversi livelli di descrizione (scheda breve/scheda estesa) si sta procedendo in due direzioni: da una parte vengono inseriti i dati ricavati da un'analisi codicologica e filologica approfondita, dall'altra vengono recuperati quelli derivanti dalle schedine dei cataloghi cartacei esistenti. Il progetto dell'ICCU, integrato a MANUS, di realizzazione di un Authority file all'interno di MANUS ha permesso la creazione, per ciascun tipo di responsabilità individuata sui manoscritti, di Authority record (corredati di notizie biografiche e bibliografiche dettagliate) relativi ai Gesuiti. L'APUG diventa così il principale punto di riferimento in materia gesuitica per gli archivi e le biblioteche partecipanti al progetto nazionale. Un catalogo unico permette, infatti, il ricongiungimento virtuale tra i documenti dell'archivio e quelli di altre sedi che conservano fondi gesuitici e la possibilità di confronto tra archivisti e bibliotecari di istituzioni diverse. Durante la catalogazione, inoltre, viene rilevato lo stato di conservazione dei manoscritti che, come si vedrà, è l'azione che deve precedere ogni intervento di digitalizzazione.

Accanto a quest'attività che consente all'archivio di essere presente in un catalogo unico con alcune tra le più importanti istituzioni culturali italiane (aderiscono attualmente al progetto oltre 150 enti), l'APUG ha avviato un progetto per la costituzione di un archivio digitale che valorizzi la complessità del suo patrimonio.

La parte più significativa del materiale documentario conservato presso l'archivio non rientra nelle normali casistiche cui si è abituati nell'ambito della produzione manoscritta. Questa documentazione si sta rivelando preziosa per una ricostruzione della storia dell'insegnamento nell'università che elaborò una consistente razionalizzazione degli studi<sup>6</sup>.

<sup>5</sup> Roberto Marcuccio, *Catalogare e fare ricerca con Manus Online*, «Biblioteche oggi», 28 (2010), n. 6, p. 33-49. Per accedere ai Fondi dell'APUG in MANUS: [http://manus.iccu.sbn.it//opac\\_SchedaBiblioteca.php?ID=168](http://manus.iccu.sbn.it//opac_SchedaBiblioteca.php?ID=168).

<sup>6</sup> Gesuiti, *Ratio atque institutio studiorum Societatis Iesu*, introduzione e traduzione di Angelo Bianchi, Milano: Rizzoli, 2002.

Da documenti<sup>7</sup> conservati in APUG emerge come alcuni manoscritti o stampati, fossero in possesso dei professori gesuiti sino alla loro morte o tramandati ai successori della cattedra per essere poi depositati nella “Camera dei manoscritti”<sup>8</sup>. All’interno della documentazione prodotta dai professori del Collegio Romano, spesso dettagliata con l’indicazione della data in cui si tennero le lezioni, è possibile individuare interventi successivi della stessa mano o di mani differenti che per decenni lavorarono sullo stesso materiale: in alcuni casi l’elemento di interesse si trova non tanto nel testo ma proprio in quelle piccole note a margine. L’indicizzazione di questi elementi, considerati normalmente secondari, è di grande rilevanza per lo studio non solo dell’attività del singolo gesuita, ma per analizzare la struttura e lo svolgimento dei singoli insegnamenti e, più in generale, le rotture e le continuità all’interno del sistema didattico del Collegio Romano. La storia dell’uso e del riutilizzo di materiali può diventare il fulcro di una ricerca che prenda in considerazione la rete di relazioni intervenute sia all’interno di un singolo manoscritto che in un *corpus* di materiale documentario più vasto.

Per offrire uno strumento che superi i limiti della ricerca tradizionale l’APUG ha firmato, con la Fondazione rinascimento digitale di Firenze e l’ILC di Pisa, un accordo di collaborazione che consentirà:

- la digitalizzazione integrale di alcune opere. I documenti saranno scelti in base a interventi di valorizzazione specifici: i due grandi progetti in corso riguardano il materiale documentario del Fondo Clavius e il celebre Epistolario di Athanasius Kircher, già digitalizzato in precedenza;
- la realizzazione di un archivio di conservazione dei materiali digitali;
- la fruizione dei dati attraverso l’impiego dell’applicazione Web open source PK, in cui confluiranno i documenti digitalizzati corredati di un apparato di metadati tradizionali, di trascrizioni del testo nelle sue diverse versioni e di descrittori semantici nelle diverse lingue.

L’intervento di digitalizzazione è volto a limitare l’accesso diretto al documento ai soli ricercatori che svolgono ricerche codicologiche, fornendo agli altri l’accesso al contenuto testuale in formato digitale. La digitalizzazione ha senso, infatti, solo in funzione di una conservazione materiale della documentazione originale. Pertanto sono stati affrontati i rischi, spesso sottovalutati, connessi alla movimentazione del materiale documentario, allo stress provocato dall’attività di riproduzione e alla scelta di apparecchiature non adatte. Per queste ragioni le diverse fasi del processo saranno gestite all’interno del laboratorio dell’APUG con attrezzature (*book*

<sup>7</sup> *Nota delle opere composte dal P. Nicolò M.a Pallavicino trovate dopo la sua morte alcune latine, alcune italiane*, carta inserita infine a F.C. 1088.

<sup>8</sup> *Antiche Consuetudini Della Comp.a; massime del Coll.o Rom.o raccolte da libri stampati o dalla tradizione di Padri che vissero prima della Soppressione etc.*, Fal. R, Archivio PUG Storia, doc. 8.

*cradle* corredato di un'apparecchiatura professionale per la fotografia digitale) realizzate secondo le specifiche esigenze dell'archivio: preservare materiale molto fragile, con legature che in molti casi non consentono l'utilizzo di scanner. Ogni fase del lavoro sarà eseguita da personale debitamente formato che si atterrà al protocollo elaborato per questo progetto. Attraverso un *software* di ritocco delle immagini, potranno essere realizzati interventi di restauro virtuale, producendo un oggetto digitale che consenta la miglior leggibilità di documenti che, spesso, presentano inchiostri sbiaditi o fortemente imbruniti. I materiali digitali prodotti, nel formato di conservazione TIFF, saranno a loro volta preservati in un *server* dedicato.

Per quanto riguarda gli strumenti di ricerca, essendo normalmente pensati da chi è abituato a incasellare staticamente dei dati in una griglia, non sempre corrispondono alle necessità pratiche di chi, talvolta, non ha le competenze necessarie per districarsi nel mondo creato da archivisti, bibliotecari o catalogatori, rischiando di conseguenza di non trovare dati fondamentali per il proprio lavoro.

Questo sistema aperto si propone di andare oltre i convenzionali cataloghi di collezioni digitali ed è per tale ragione che sarà necessario partire dall'assunto che esistono diversi livelli di osservazione.

Il concetto di osservazione, mutuato dalla teoria dei sistemi sociali di Niklas Luhmann, potrebbe fungere da cornice teorica nella quale pensare uno strumento di ricerca adatto a un'euristica che tenga conto delle complessità tanto del materiale che della ricerca. L'osservazione implica l'unità, in una stessa operazione, del distinguere e indicare: conoscere è marcare un lato e non un altro in quello che si osserva. Assumendo quindi come valida l'affermazione secondo la quale la realtà è sempre una realtà osservata, il documento, a sua volta testimonianza di un'osservazione, dovrebbe essere descritto lasciando aperta l'analisi alle nuove possibili differenziazioni.

Nell'individuazione dei casi d'uso non sarà sufficiente indicare gli attori che potenzialmente useranno quest'applicazione ma, nel panorama delle risorse online, è più che mai necessario chiedersi quali ricerche sono necessarie e in che modo possono essere svolte dall'utente remoto.

Attingere ad una conoscenza non significa che l'oggetto dell'analisi ci possa essere svelato nella sua "realtà", bensì che attraverso "l'invenzione" di dati esterni gli vengano attribuiti descrittori che lo rendano in qualche misura conoscibile.

I tempi della documentazione archivistica non sono i nostri tempi: la creazione di uno strumento di ricerca deve tenere conto della conflittualità tra il regime di storicità nel quale il manoscritto è stato prodotto e l'aspettativa di un risultato che si pretende, da quando la ricerca avviene attraverso la rete, praticamente immediato. Lo scotto da pagare per questa immediatezza può essere un esito della ricerca incongruente, insufficiente o troppo generico.

Si deve inoltre considerare come la descrizione di ciò che non è contemporaneo crei soltanto un'illusione di contemporaneità, attribuendo a un documento un valore che poteva originariamente non avere.

In un sistema aperto tutti i dati potranno essere integrati e costantemente aggiornati dagli utenti stessi, una volta abilitati, in modo da creare una piattaforma di lavoro che veda la collaborazione della comunità scientifica internazionale. Per garantire un uso appropriato dei dati relativi alla documentazione l'accesso alla banca dati online avverrà in modalità controllata definendo le attività dello specifico utente, monitorandone le ricerche in modo da evitare duplicazioni inutili e ridondanti di edizioni relative ai medesimi manoscritti.

Il *medium* del computer permette di differenziare, nella comunicazione, l'immissione di dati dalla richiesta di informazioni. Come nel caso della scrittura, non c'è unità tra immissione dei dati e comprensione. All'ampliamento della possibilità di gestire un numero sempre crescente di dati e relazioni dovrà accompagnarsi la necessità di una selezione che influirà sugli stessi contenuti.

PK consentirà la gestione di materiale eterogeneo a diversi livelli di descrizione: si pensi al caso di un testo a stampa glossato dall'autore in previsione di una ristampa dove compaiono anche note dei censori, glosse successive e si trovino inserite lettere relative proprio a quell'edizione. In questo caso i livelli di descrizione sono

almeno quattro: il testo edito, le glosse di commento, l'intervento dei censori e il carteggio.

Ovviamente questi diversi livelli prevedono l'utilizzo di standard diversi come, ad esempio, gli *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)*, ISBD(A), o la *Guida a una descrizione uniforme dei manoscritti e al loro censimento* che non entreranno in conflitto, ma dovranno integrarsi a seconda che la descrizione riguardi, nel caso specifico, materiale a stampa o manoscritto.

Attualmente nel caso di te-

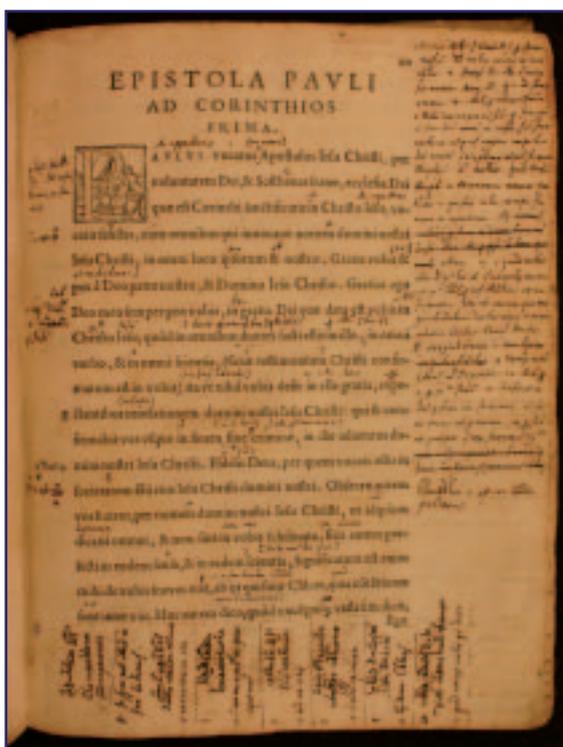


Figura 2. *Biblia. N.T. Ad Corinthios Epistula*, metà XVI sec. (APUG 429). Cinquecentina attribuibile al tipografo Bartholomaeus Gravius riccamente glossata da più mani e in più lingue

sti ibridi, stampa-manoscritto, si può procedere solo in due direzioni: o si descrive l'edizione inserendo come note di esemplare i riferimenti alla parte manoscritta, o si analizza il manoscritto indicando l'edizione come parti a stampa dove dati quali il titolo e i riferimenti alla pubblicazione non sono indicizzati. È evidente invece come in questo oggetto documentario non possano essere scisse le due descrizioni: soltanto analizzando i diversi livelli (stampa/manoscritto) e le relazioni intervenute (glosse/censura) su questo manufatto sarà possibile restituirne la complessità, parzialmente percepita attraverso l'immagine digitale.

Un altro esempio è quello di una legatura di riutilizzo costituita dal foglio di un manoscritto medievale: al singolo oggetto dovranno corrispondere sia i descrittori relativi alla legatura che quelli riguardanti la parte del manoscritto che, se identificata e collegata ad altre legature o a manoscritti mutili, consentirà di ricomporre virtualmente un documento oggi smembrato.

Come sottolineato all'inizio l'immissione di dati che descrivono una tale complessità, se non corredata da strumenti avanzati per la ricerca, potrebbe rivelarsi inutile. È per questo che gran parte del lavoro riguarderà l'elaborazione di una metodologia che consenta nuovi tipi di ricerca nel panorama del Web semantico. L'analisi di questi processi sarà approfondita nei due successivi interventi.

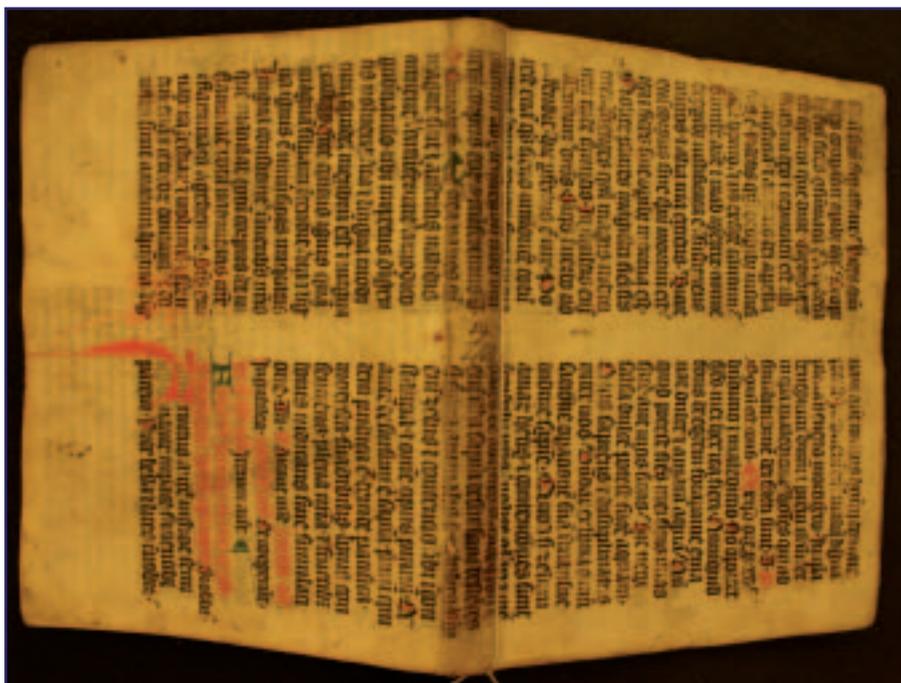


Figura 3. Pergamena medievale riutilizzata come legatura del codice F.C. 462

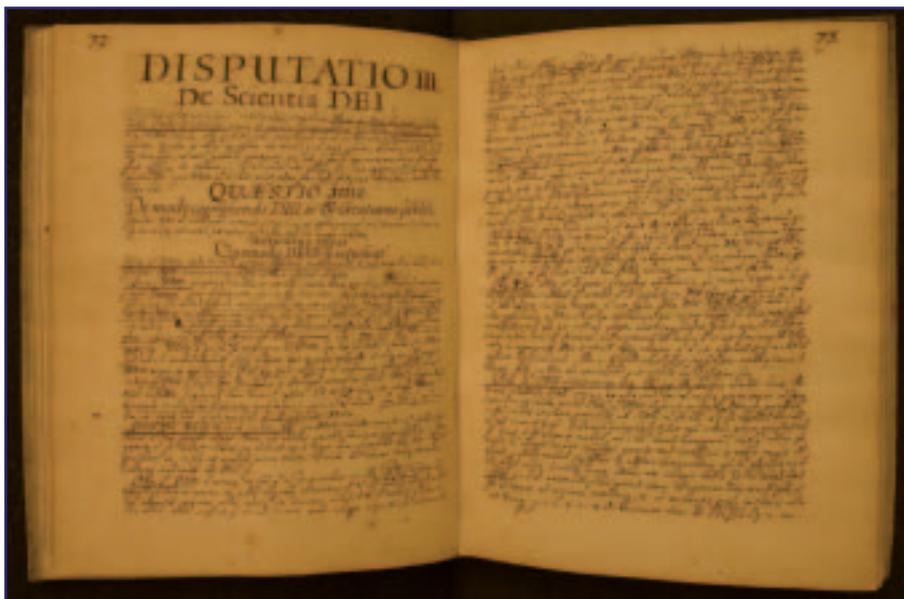


Figura 4. Martinus Juranich, *Tractatus de Deo uno et trino*, 1659 (F.C. 462, p. 72-73)

## Il metodo<sup>9</sup>

In questa sezione prenderò in esame gli aspetti di carattere metodologico che devono essere considerati nel momento in cui un'istituzione culturale (biblioteca, archivio, museo ecc.) diventa digitale e si pone l'obiettivo di rendere fruibili le fonti, in particolare quelle manoscritte inedite, ad un pubblico rappresentato principalmente da specialisti. Vale la pena di evidenziare come il processo di diffusione della cultura digitale (iniziato ormai da molti anni anche se proseguito, specialmente in Italia, con una maggiore lentezza e mancanza di coordinamento rispetto ad altri paesi europei, come, per esempio, il Regno Unito<sup>10</sup>) abbia conosciuto proprio di recente una forte accelerazione. Ciò è dovuto, tra l'altro, al fenomeno della diffusione e vendita di libri elettronici che si accompagna alla disponibilità sul mercato dei cosiddetti *ebook reader*. È molto recente la notizia che vede Amazon, leader indiscusso nella vendita di libri mediante ordinazioni online, aver distribuito per la prima volta un numero di testi elettronici superiore a quelli cartacei. La distribuzione di I-Pad da parte di Apple ha intensificato il fenomeno che ora rappresenta un momento di svolta anche per l'editoria italiana<sup>11</sup>.

<sup>9</sup> Andrea Bozzi (Istituto di linguistica computazionale "Antonio Zampolli" del CNR di Pisa).

<sup>10</sup> Si ricorda a tale proposito il pionieristico *Follett Report* che considerava la nuova tecnologia digitale per le biblioteche come strumento essenziale ai fini dell'educazione superiore. Il testo è leggibile all'indirizzo: <http://www.ukoln.ac.uk/services/papers/follett/report/intro.html>.

<sup>11</sup> Il catalogo dei libri elettronici in lingua italiana è ancora piuttosto limitato, ma le grandi case editrici, come, per esempio, il gruppo Mondadori e Feltrinelli, stanno investendo in questo senso.

Non possiamo certo sapere se e in quale misura tale rivoluzione nella diffusione del sapere (peraltro, non confinata al solo settore librario, ma relativa anche a quello musicale e filmico) condiziona anche la produzione di edizioni critiche, ovvero, il mondo della filologia e dell'editoria scientifica. Nel momento attuale, trovandoci di fronte ad un archivio di fonti con un valore culturale molto significativo e sottoposto a conversione digitale anche per aspetti di conservazione ibrida<sup>12</sup> e di modalità alternative di fruizione, dobbiamo stabilire quali modalità siano le più appropriate affinché l'Archivio storico della Pontificia Università Gregoriana (APUG)<sup>13</sup> possa mettere in valore il posseduto in fase di digitalizzazione indipendentemente dal fatto che tale documentazione sia poi eventualmente distribuita, secondo una non ancora precisabile modalità, su apparecchi portatili come i sopra menzionati *ebook reader* o su macchine collegate ad Internet e dotate di funzioni di annotazione e di ricerca sofisticate.

Partendo da queste premesse, dunque, conviene distinguere due momenti: un primo momento corrisponde a quello che deve collegare il progetto dell'APUG, all'ambiente delle *digital humanities* ed è quello sul quale mi soffermerò; un secondo momento, invece, è quello che mette in relazione i prodotti ottenuti nell'ambito di iniziative come questa con le nuove modalità di diffusione massiva dei dati digitali e che, almeno al momento attuale, non conviene affrontare. Il primo vede principalmente un utilizzo in ambito specialistico, mentre il secondo presuppone la presenza di un pubblico di lettori/utilizzatori molto vasto che viene invogliato ad accedere ai documenti anche per la presenza di elementi di assistenza aggiuntivi alla lettura ed alla comprensione (annotazioni, illustrazioni, note esplicative, accompagnamento sonoro, brevi filmati, ecc.). Potrei concludere questa parte dicendo che la prima fase riguarda la ricerca scientifica e la componente critico-editoriale della documentazione digitale, mentre la seconda riguarderebbe, gli aspetti didattici (strumenti di lettura e ricerca documentaria) che, ove opportunamente controllati, potrebbero valorizzare i percorsi formativi degli allievi delle scuole, specialmente a livello universitario.

Vediamo con maggiori dettagli gli aspetti di carattere metodologico sottesi al sistema informatico per la produzione della biblioteca digitale da parte dell'APUG

<sup>12</sup> La conservazione ibrida contempla diverse modalità di riproduzione di uno stesso documento in modo tale che di esso esistano più copie ottenute con tecnologie diverse (fotografiche e digitali), con lo scopo ambizioso di favorire la creazione di un sistema informativo integrato nel quale tradizione ed innovazione possano cooperare.

<sup>13</sup> Un accordo di collaborazione scientifica e tecnologica è stato firmato fra l'APUG, l'Istituto di linguistica computazionale "A. Zampolli" del CNR di Pisa (ILC) e la Fondazione rinascimento digitale di Firenze (FRD) proprio al fine di sviluppare metodi e strumenti adeguati alla valorizzazione del patrimonio librario storico posseduto dall'istituzione universitaria romana. Date le caratteristiche, descritte nella prima parte di questo contributo, di questa documentazione culturale si è resa necessaria un'attenta valutazione di quanto già realizzato a Pisa e Firenze affinché il sistema sia calibrato sulle specifiche esigenze di catalogazione, conservazione e fruizione dell'APUG.

con caratteristiche tali da garantirne lo studio, l'annotazione e, in una prospettiva ancora più ampia, la condivisione nell'ambito di un'infrastruttura di ricerca per le discipline storiche e filologiche che si ipotizza possa prendere l'avvio nei prossimi anni sulla base di accordi e progetti internazionali sostenuti dalla Commissione Europea e controllati dalla Fondazione Europea della Scienza di Strasburgo. Questo aspetto non va sottovalutato poiché appare sempre più necessaria una visione coordinata di tutte le attività che ruotano intorno alle *digital humanities* affinché si renda possibile o si semplifichi il processo di interoperabilità fra dati del *cultural heritage*, oggi difficilmente integrabili poiché solo in minima parte immessi sul Web con strumenti e metodi non dichiarati o al di fuori di standard condivisi.

Potremo sintetizzare questo approccio elencando gli elementi principali che sono stati presi in considerazione nella progettazione del sistema PKT, per una descrizione più dettagliata della quale si rinvia alla sezione successiva "Strumenti e funzionalità", utilizzato per la catalogazione, gestione ed interrogazione dei documenti digitalizzati a cura dell'APUG.

Un sistema che consenta di effettuare studi di carattere filologico e critico testuale su documenti digitali necessita di almeno tre componenti: la gestione delle immagini, dei testi e degli apparati (bibliografia, note critiche, annotazioni).

### **Le immagini**

Oltre ai consueti programmi che consentono di applicare varie tipologie di filtri per agevolare la lettura specialmente di quelle zone interessate da danni e da fenomeni di evanescenza degli inchiostri o acidificazione del supporto cartaceo, è importante disporre anche di strumenti in grado di trovare automaticamente o inquadrare manualmente le "zone parola" all'interno dell'immagine digitale. Questa operazione, non troppo complessa per i libri a stampa antichi con una netta spaziatura fra le parole, risulta utile soprattutto nel caso in cui il testo non debba essere interamente trascritto, ma solo alcuni termini significativi meritino di essere annotati ed indicizzati. Si pensi, per esempio, al caso di documenti nei quali sono presenti illustrazioni, grafici, figure geometriche, formule matematiche o chimiche, miniature che rappresentino il vero oggetto di analisi da parte di uno studioso. Il criterio seguito vuole consentire non solo le operazioni di selezione come quelle sopra citate, ma anche di tracciabilità di tutti coloro che siano in qualche modo intervenuti nelle operazioni di selezione. Viene in tal modo resa concreta una prospettiva di lavoro filologico collaborativo in rete. Abbiamo potuto constatare che una simile funzionalità risulta particolarmente utile nella digitalizzazione di libri di storia della scienza ove sono presenti numerose figure, spesso vero obiettivo di una ricerca e talvolta più importanti del testo che le accompagna poiché su esse si concentra l'attenzione del moderno commentatore ed editore.

## I testi

L'esperienza maturata in numerosi centri specializzati nello sviluppo di applicazioni di *text processing* prosegue ormai da parecchi decenni da quando, cioè, sono state realizzate anche dati testuali col fine di documentare stadi di sviluppo di una lingua o di produrre indagini lessicografiche specifiche. Il metodo da noi seguito per questa componente di PKT non si discosta da questi precedenti, ormai ben documentati, ma intende utilizzare in maniera particolarmente sviluppata il principio secondo il quale un testo sia effettivamente rappresentato da più livelli che devono essere marcati affinché non se ne disperda il contenuto informativo. I fenomeni di intertestualità come, per esempio, le citazioni che un autore riferisce altre sue opere o da quelle di autori diversi, appartengono ad un livello di testo specifico che si differenzia dal livello al quale appartiene il testo dell'autore che cita. Un elemento di granularità ancora maggiore potrebbe rendere necessaria la distinzione fra le citazioni da opere in prosa e quelle in poesia, oppure fra quelle che appartengono ad un periodo determinato rispetto ad altre. Questa serie di fenomeni, se opportunamente marcati, aumenta il livello informativo del testo e consente alla gestione automatica degli indici di tenerne conto. Come accennato in precedenza, un'applicazione orientata al lavoro filologico deve consentire, mediante un'interfaccia per l'utente finale semplice e capace di rispondere alle sue necessità di ricerca, di attivare tutti gli elementi di *mark-up* necessari. Essi, a loro volta, dovranno corrispondere direttamente (seguendo delle precise *guidelines*) o indirettamente (attivando un programma di conversione automatico) a sistemi di codifica del testo entrati nell'uso comune attraverso standard riconosciuti a livello internazionale. Sono proprio questi elementi che danno una garanzia che quanto viene svolto anche da un singolo studioso possa entrare a far parte, se lo si riterrà opportuno, di una comunità di studiosi che condividono programmi e risorse all'interno di una stessa infrastruttura di ricerca<sup>14</sup>.

Non è il caso di soffermarsi su un ulteriore ed importante aspetto correlato all'analisi del testo e, in particolare, alla produzione di indici: il sistema PKT presenta un'architettura che lo rende permeabile alle funzioni prodotte, per esempio, da un sistema di analisi morfologica automatica e lemmatizzazione. Un caso specifico appartenente a questo settore è dato dal sistema LemLat<sup>15</sup> per la produzione di in-

<sup>14</sup> Il tema delle *Research Infrastructures* (RI) si sta rilevando molto importante anche nel settore delle scienze umane dopo essere stato ampiamente discusso e analizzato per le cosiddette scienze dure. Le RI sono considerate un elemento imprescindibile per la costituzione di una *European Research Area* (ERA) nella quale ogni comunità di studiosi condivide standard, dati, sistemi, risultati al fine di competere, grazie a strumenti adeguati alla nuova società della conoscenza, con Paesi molto avanzati nel campo della ricerca scientifica e delle innovazioni tecnologiche. Per ERA si veda [http://ec.europa.eu/research/era/index\\_en.htm](http://ec.europa.eu/research/era/index_en.htm). Per quanto riguarda, invece, le RI si veda <http://cordis.europa.eu/infrastructures/>.

<sup>15</sup> Il sistema è stato originariamente disegnato e sviluppato per produrre l'indice lessicale dei grammatici latini antichi presso l'ILC di Pisa. Nel corso degli anni esso ha subito notevoli

dici lessicali latini, ma sono in fase di realizzazione simili strumenti per altre lingue, come, per esempio, l'arabo e l'occitano medievale.

### Le annotazioni

Le annotazioni sono costituite da più elementi diversi: annotazioni libere o informazioni di apparato critico. Nel primo caso contengono osservazioni personali dell'editore (ma anche di chiunque intenda intervenire, se autorizzato, a chiosare un passo, un'illustrazione, una singola espressione linguistica) e si presentano nella forma di testo non strutturato. Il metodo che abbiamo seguito nella progettazione del nostro sistema di filologia computazionale, anche in seguito a esplicite richieste di ricercatori appartenenti ad ambiti di studio molto diversi fra loro, prende in considerazione anche la possibilità di stabilire delle specifiche ontologie di dominio affinché sia il catalogo dei documenti che il testo da essi veicolato possano venire classificati secondo classi semantiche precostituite all'interno dell'ontologia. È inutile sottolineare che l'operazione di classificazione comporta un aggravio di tempo da parte dell'operatore che annota, anche se l'interfaccia venga predisposto in modo che le classi siano presenti, per esempio, in un menù a tendina dove la scelta è certamente rapida ed è garantita omogeneità di comportamenti. In ogni caso, il tempo e le energie spese nella fase di classificazione sono ben ripagate dalla maggiore quantità di informazioni che si recuperano nel momento in cui si accede al testo anche mediante le sottoclassi dello schema ontologico. Per fare un esempio, si potranno ritrovare tutti i passi dove è attestata la descrizione di parti anatomiche del corpo umano e il sistema restituirà i contesti che descrivono il cranio, le braccia, il tronco, ecc., risultato impossibile da ottenere grazie alla sola lettura di un indice alfabetico delle forme o dei lemmi. Un altro esempio molto interessante riguarda la ricerca linguistica su un *corpus* di opere che trattano, appunto, di fenomeni linguistici: un'ontologia specifica di questo dominio, che potrebbe anche essere relativa solo ad un periodo o ad un singolo grande linguista, consentirebbe di recuperare tutte le parti descrittive di fenomeni grafico-fonetici, presenti eventualmente in una determinata zona geografica. Le annotazioni che si presentano nella forma di un apparato critico servono principalmente per rappresentare almeno tre situazioni: la registrazione di varianti, nel caso in cui un testo sia trasmesso da più testimoni (nel progetto dell'APUG materiale documentario principalmente manoscritto) diversi. Fa parte di questo stesso ambito la registrazione di varianti presenti nelle diverse edizioni a stampa di una medesima opera, caso, talvolta, complicato dalla presenza di annotazioni manoscritte da parte dell'autore stesso che compaiono generalmente nei margini della pagina.

trasformazioni in concomitanza con gli sviluppi della tecnologia. Per informazioni, si veda la sezione *Documents* sul sito <http://www.ilc.cnr.it/lemlat/>, oppure Andrea Bozzi – Giuseppe Cappelli, *A Project for Latin Lexicography: 2. A Latin Morphological Analyzer*, «Computers and the Humanities», 24 (1990), n. 5-6, p. 421-426.

Un secondo caso è relativo ai manoscritti di autori moderni e contemporanei che sono intervenuti a più riprese nella modifica del proprio testo, molto raramente in quelli di autori antichi ove, piuttosto, si rinvengono chiose e commenti di altra mano. Un terzo fenomeno riguarda i testimoni unici, ovvero i documenti che riferiscono un testo non altrimenti noto: anche in questo caso, tuttavia, l'annotazione potrebbe assumere la veste di apparato critico qualora il filologo ravvisi errori banali o sostanziali e ritenga di intervenire restituendo il testo presunto originale con una lezione personale.

Il principio da noi seguito tiene presenti tutti questi aspetti e tende a risolverli applicando un principio unico.

Tali impostazioni metodologiche, in incontri con i referenti dell'APUG, sono state perfezionate *ad hoc* e su queste basi si è avviata la fase di progettazione di componenti software che avranno il compito di produrre i contenuti digitali sul Web e di mettere l'utenza, individuabile in una comunità di specialisti, in condizioni di navigare facilmente su dati complessi e di eseguire operazioni di ricerca intuitive e guidate.

## Strumenti e funzionalità<sup>16</sup>

### Gli strumenti

#### *Il progetto Pinakes*

Il progetto PK è il frutto dell'attività svolta all'interno del dipartimento di ricerca dell'Istituto e museo della storia della scienza di Firenze (ora Museo Galileo<sup>17</sup>). Dal 2006 PK è uno fra i principali progetti di ricerca della Fondazione rinascimento digitale<sup>18</sup>. Nell'anno successivo l'ILC si è associato al lavoro di ricerca, soprattutto per quel che riguarda la gestione del testo digitale.

PK è un'applicazione Web che consente la costruzione di una base di conoscenza distribuita e condivisa basata su ontologie<sup>19</sup> e dispone di uno specifico modulo, PKT per la realizzazione di edizioni digitali di testi critici. PK gestisce dati persistenti avvalendosi di strumenti quali Internet e le ontologie. L'applicazione è aperta, modulare, personalizzabile e gratuitamente scaricabile sotto il controllo della General Public License<sup>20</sup>. Si tratta di un'applicazione in grado di gestire molteplici

<sup>16</sup> Marco Rufino (Fondazione rinascimento digitale).

<sup>17</sup> Istituto e museo di storia della scienza, Firenze, <http://www.imss.fi.it>; Museo Galileo, Firenze, <http://www.museogalileo.it> (1 ottobre 2010).

<sup>18</sup> Fondazione rinascimento digitale, Firenze, <http://www.rinascimento-digitale.it> (1 ottobre 2010).

<sup>19</sup> Cfr. la voce "Ontologia (informatica)" in Wikipedia, l'enciclopedia libera: [http://it.wikipedia.org/wiki/Ontologia\\_\(informatica\)](http://it.wikipedia.org/wiki/Ontologia_(informatica)) (20 ottobre 2010).

Per altre informazioni sulla definizione di ontologia: Tom R. Gruber: *What is an Ontology?*, 1992, <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>, Stanford University: Knowledge System Laboratory, <http://www-ksl.stanford.edu> (1 ottobre 2010).

<sup>20</sup> General Public License, <http://www.gnu.org/licenses/gpl.html> (1 ottobre 2010).

tipologie di dati e oggetti o risorse digitali quali testi non strutturati e strutturati (HTML, XML, etc.), immagini, audio, audiovideo, oggetti tridimensionali, tabelle e grafici, espressioni matematiche, GIS, mappe e cartografie. Inoltre, essa consente una gestione multilingue: al momento attuale è possibile utilizzare l'italiano, l'inglese, il francese e il tedesco.

La versione Pinakes 3.0 è formata da tre componenti principali:

- Pinakes Admin: interfaccia Web di amministrazione per la definizione dei progetti, dello schema dei dati, degli utenti e per la configurazione di uno specifico *repository*;
- Pinakes Input: applicazione Web per il popolamento del *repository* e l'*editing* controllato dei dati in un modello PK;
- Pinakes Text: applicazione Web per l'edizione critica digitale di documenti testuali.

L'architettura software consiste di moduli Open Source per il cui sviluppo sono stati utilizzati:

- Linguaggio di sviluppo: JAVA (Jdk 1.5 o succ.);
- Servlet Engine: Tomcat 5.5.x + APACHE HTTP connectors;
- Web Server: APACHE HTTPD server 2.2.x;
- WebApplication Framework: Jakarta Struts; JBoss Seam;
- Web Service Framework: APACHE Axis 1.4;
- Database Engine: PostgreSQL 8.1 o successive;
- IDE: Netbeans 5.5.1.

### *Perché Pinakes*

PK nasce per dare una risposta alle problematiche e metodologie emerse dall'utilizzo dell'IT (Information Technology) per le ricerche e gli studi in ambito umanistico. Infatti, se da un lato l'introduzione dell'IT permette lo sviluppo di metodi formali all'interno delle discipline umanistiche, dall'altro richiede una significativa evoluzione delle metodologie di ricerca, fatto che tarda a realizzarsi anche a causa delle numerose ed ingiustificate resistenze. Tale ritardo influenza negativamente la possibilità di realizzare, comunicare e condividere il sapere attraverso le tecnologie del Web, cioè di individuare modalità ottimali di impiego delle innovazioni, sviluppando uno specifico *data modeling*.

Infatti, mentre sempre più spesso i risultati dell'attività di studio in ambito umanistico vengono codificati in formato digitale, l'utilizzo di differenti modelli per l'organizzazione di questi dati rende difficile, anche nell'ambito della stessa disciplina, lo scambio, l'interoperabilità, la condivisione ed il riuso dei risultati prodotti dall'attività della ricerca.

La caratteristica fondamentale di PK è quella di consentire la costruzione di una base di conoscenza distribuita e condivisa basata su ontologie. A questo PK affianca il fatto di essere un'applicazione Web, e quindi di permettere l'accesso per la modifica, l'implementazione e la consultazione dei dati senza limitazioni di tempo e di luogo.

Queste due caratteristiche garantiscono notevoli vantaggi:

- migliorano l'interscambio dei dati fra le persone e all'interno di istituzioni, enti e organizzazioni perché favorisce l'interoperabilità fra sistemi;
- permettono la creazione di un *data modeling* interoperabile per una gestione integrata ed evoluta dei dati;
- favoriscono, grazie ad una base di conoscenza comune, la condivisione e la ri-usabilità di metodi di modellizzazione, paradigmi, linguaggi e dati;
- permettono la pubblicazione sul Web dei dati provenienti da progetti differenti o da settori disciplinari diversi;
- facilitano il legame di dati contenuti in *repository* o basi dati diverse nonché di tipologie diverse di oggetti o risorse digitali;
- migliorano la ricercabilità dei dati poiché consentono interrogazioni semantiche;
- consentono il riutilizzo di dati digitali già presenti sia sul Web che all'interno di basi dati;
- consentono di riconoscere Proprietà Intellettuale e copyright dei risultati individuali;
- semplificano le collaborazioni interdisciplinari grazie ad un ambiente multilingue;
- offrono una nuova prospettiva e visione sulle relazioni fra i concetti, elemento indispensabile per aprire nuove strade ad ulteriori studi e, quindi, acquisire nuove conoscenze.

### *Ontologie e Web semantico*

Il termine Web semantico<sup>21</sup> indica un'evoluzione del Web attuale in cui le informazioni e i dati assumono una precisa caratterizzazione semantica che rappresenta la sostanziale novità rispetto al Web attuale. Esso renderà possibili elaborazioni au-

<sup>21</sup> Il World Wide Web, come si presenta oggi, non dispone di strumenti in grado di fornire informazioni *ad hoc*: ricerche effettuate sui motori di ricerca, Google *in primis*, solitamente forniscono una grande quantità di risultati, moltissimi dei quali con scarsa attinenza alla richiesta effettuata. Non esiste infatti uno strumento informatico capace di "comprendere" il contenuto di una pagina Internet: il Web, cioè, opera ancora in modo "letterale" (sulla base di ricerca di specifiche stringhe) e non dispone ancora di strumenti capaci di comprendere il valore "semantico" delle interrogazioni.

Questa limitazione è comune sia alla rete di Internet che alle Intranet di enti, istituzioni ed aziende, i cui documenti, dati e informazioni, che rappresentano la "conoscenza" specifica di ciascuna organizzazione, sono sì memorizzati in formato elettronico, ma in modo tale da restare quasi altrettanto inaccessibili di quando venivano archiviati su carta. Infatti, gran parte del contenuto di Internet è ancora progettato per essere "letto" piuttosto che per essere "elaborato" da programmi.

tomatiche molto più complesse ed efficaci. Per tale ragione assume un ruolo determinante la produzione di risorse semanticamente marcate e, soprattutto, l'attribuzione di una connotazione semantica a quelle già esistenti.

PK nasce proprio con questo intento: offrire gli strumenti per la costruzione di una base di conoscenza distribuita e condivisa basata su dati semantici e ontologie; esso dispone di meccanismi grazie ai quali i dati semantici possono essere contenuti da un modello ontologico esterno alla risorsa di riferimento.

La scelta di basare PK su ontologie è derivata dalla considerazione che l'uso di ontologie e le sottostanti tecnologie di modellazione contribuiscono all'organizzazione di architetture software con elevate capacità di interoperabilità, cooperazione, adattamento e capacità di evolvere.

In questa prospettiva, l'approccio ontologico comporta numerose opportunità:

- le ontologie permettono la rappresentazione esplicita di modelli semantici che combinano la non ambiguità necessaria alla specificità tecnica con la comprensibilità capace di colmare la distanza tra gli specialisti e i tecnici;
- il modello ontologico si adatta bene ad un contesto distribuito e permette la creazione di modelli per riuso, composizione e riconciliazione di frammenti sviluppati in modo concorrente e distribuito;
- le ontologie hanno la capacità di modellare domini che evolvono nel tempo, superando molte complessità che invece discendono dall'uso esclusivo di tecnologie di rappresentazione convenzionale. Questo, a sua volta, favorisce un approccio incrementale allo sviluppo capace di accompagnare processi di evoluzione verso modelli condivisi.

È quindi necessario un approccio che renda possibile in forma automatizzata un'interpretazione e una comprensione semantica dei dati. Qualcosa di profondamente diverso dalla nuova Babele di informazioni e linguaggi nella quale è facile perdersi: uno strumento, che può essere rappresentato dal Semantic Web (Web semantico) capace di far ottenere, ad ogni ricerca, risposte mirate e selettive. Il termine Semantic Web è stato proposto per la prima volta da Tim Berners-Lee (Cfr. Tim Berners-Lee – James Hendler – Ora Lassila, 2001: *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a devolution of new possibilities*, «Scientific American», may 2001. Tim Berners-Lee ha sottolineato che uno degli elementi fondamentali del web semantico sarà la compresenza di più ontologie. Se si vuole un sistema dinamico in grado di raffinarsi e funzionare su scala universale, bisognerà pagare il prezzo di una certa dose d'incoerenza. Gli obiettivi del Web Semantico sono di riportare chiarezza, formalità e organizzazione nei dati, collegando l'informazione presente nelle pagine Web a concetti astratti organizzati in una gerarchia (ontologia), a sua volta descritta in un meta-documento; permettendo a vari agenti intelligenti, di interpretare le informazioni e di addentrarsi nella rete navigando come esseri umani. Offrire, quindi, la possibilità di cogliere il contesto semantico di una fonte informativa interpretando le varie relazioni esistenti tra le risorse, formulare asserzioni sulle stesse, nonché controllare la loro attendibilità.

Per approfondire l'argomento: *W3C Semantic Web Activity*, <http://www.w3.org/2001/sw/> (1 ottobre 2010).

L'abituale organizzazione dei dati fatta ricorrendo al meccanismo della classificazione, può presentare diversi problemi derivanti essenzialmente dalle differenze di dominio, terminologia, scelta e significato delle varie caratteristiche ritenute importanti e dalle differenze fra le relazioni più significative. Anche sottili differenze fra le diverse terminologie risultano importanti e la definizione di vocabolari standard non risolve il problema.

All'opposto, un'ontologia, almeno nel senso in cui questo termine viene impiegato nel settore informatico, è un oggetto specifico, progettato per esprimere i concetti in maniera non ambigua e con un buon grado di copertura e di precisione rispetto al modello inteso. Una classificazione basata su metadati (ad esempio Dublin Core, Metadati amministrativi gestionali – MAG), pur essendo finalizzata all'integrazione dei dati, risente sempre della soggettività dei compilatori, mentre un'ontologia rappresenta un modello formale capace di accettare un livello di complessità maggiore, privilegiando la completezza e la correttezza logica rispetto alla comprensibilità umana. Tutto ciò offre una serie di evidenti vantaggi, ma richiede un lungo e profondo lavoro di riflessione per passare da una struttura lineare o relazionale (propria ad esempio delle schede per la catalogazione bibliografica, archivistica, artistica, ecc.) ad una struttura ontologica (e quindi gerarchica) per la costruzione di una base di conoscenza distribuita e condivisa.

### *L'Archivio storico della Pontificia Università Gregoriana*

Come ben illustrato nella prima parte di questo articolo, l'APUG costituisce una fonte archivistica di grande valore che rappresenta, per la sua complessità intrinseca, articolazione e varietà del materiale documentale posseduto, un'appassionante sfida per qualsiasi archivista.

Una sfida che si svolge su due piani: il primo, con caratteristiche più "materiali", sarà quello di rappresentare compiutamente la complessità della documentazione conservata presso l'archivio che, per la maggior parte, non rientra nelle normali casistiche della produzione manoscritta; il secondo, ancor più impegnativo, riuscire a "ricostruire", partendo, dalla utilizzazione continuativa della produzione manoscritta e dalle rielaborazioni successive, la storia dell'insegnamento e della sua evoluzione all'interno del Collegio Romano.

Questa prova assume una dimensione particolare per chi si propone, come PK, di rappresentare in modo logico e semanticamente definito uno specifico dominio di conoscenza, fornendo nel contempo strumenti automatizzati di accesso e consultazione, ma anche di analisi, annotazione e ricerca scientifica collaborativa.

Riteniamo che il progetto dell'APUG per la costituzione di un archivio digitale possa rappresentare uno dei banchi di prova più significativi per le ambizioni di PK nel proporsi come uno dei migliori e più efficienti strumenti di catalogazione archivistica e di rappresentazione semanticamente connotata della realtà contenuta nel materiale documentale.

### Le funzionalità

I tre componenti principali di PK forniscono all'utente l'insieme di funzionalità necessarie per:

- la definizione via Web dello Schema, dei progetti, degli utenti e la configurazione di un *repository* (Pinakes Admin);
- il popolamento via Web del *repository* e l'*editing* controllato dei dati in un modello PK (Pinakes Input);
- l'edizione critica digitale via Web di documenti testuali (Pinakes Text).

Pinakes Admin rappresenta l'area gestionale dove, sulla base di metodi già disponibili all'interno dell'applicazione (ontologia fondazionale<sup>22</sup>), l'utente può definire:

- quanto sia dettagliata la descrizione degli oggetti che vengono classificati;
- che tipo di relazione deve avere il suo oggetto con gli altri oggetti;
- cosa può o meno venire pubblicato sul Web;
- quali dati possono essere condivisi e quali no;
- chi fa che cosa e con quale livello di autorizzazione.

Pinakes Input rappresenta l'area della classificazione dove l'utente, nel rispetto dei vincoli di Proprietà intellettuale e copyright, può:

- inserire i dati utilizzando una propria definizione e, nel caso di modifiche che si rendano necessarie, come sia possibile effettuarle in base ai metodi adottati;
- raffinare la definizione del dominio di informazione e di conoscenza in base alle necessità incontrate nel corso della ricerca;
- consultare ed utilizzare i dati appartenenti ad altri progetti che potrebbero essere pertinenti alla propria ricerca;
- usufruire e/o offrire altre relazioni di progetto e indici di definizione;
- accedere ai depositi digitali di diversi progetti e utilizzare tutti gli oggetti digitali (immagini, video, musica, ecc.);
- definire una navigazione sui dati con lo scopo di modificarli;
- accedere tramite Web ad altri database allo scopo di catturare dati ed importarli in un progetto esistente;

<sup>22</sup> Con *Ontologia fondazionale* si vuole indicare il "modello concettuale" (*core ontology*) che PK mette a disposizione per aiutare gli utenti nella creazione di ontologie di dominio adatte all'analisi del proprio dominio di conoscenza.

L'ontologia fondazionale di PK fornisce un modello globale estensibile in cui possano essere messi in corrispondenza ed integrati dati provenienti da fonti eterogenee, grazie ad un insieme di classi e di predicati dotati di un significato determinato che possono venire estese dall'utente per poter rappresentare il proprio dominio di conoscenza.

- accedere a depositi digitali esterni al progetto Pinakes 3.0 per reperire nuove risorse o fare riferimento ad esse attraverso la creazione automatica di un indirizzo Web.

1. Pinakes Text rappresenta l'area di lavoro sui documenti digitali dove l'utente può:

Ricerca e consultare testi sulla base:

- del titolo;
- della responsabilità (autore, copista, legatore ecc.);
- di eventuali indici specifici (come quello per volumi – nel caso esista);
- del testo: le ricerche all'interno del testo consentiranno di individuare parole (interi o parziali), insiemi semplici o complessi di parole.

In particolare sarà possibile:

- individuare stringhe di caratteri appartenenti ad una parola intera o costituenti la parte iniziale, centrale o finale;
- indicare lo specifico ambito di ricerca, dal *full text* ad uno specifico livello in cui il testo è stato suddiviso (titolo, citazioni, note, etc.);
- attribuire valore distintivo agli accenti, agli spiriti o al maiuscolo;
- cercare una stringa di caratteri in concorrenza o in alternativa ad una seconda stringa, utilizzando operatori booleani di inclusione o esclusione;
- indicare quale intervallo di parole considerare soglia massima nella ricerca di due stringhe di caratteri con operatori booleani.

All'interno della pagina di consultazione sarà disponibile sia l'immagine digitale del documento che la trascrizione del testo in essa leggibile.

2. Fornire nuovi contenuti, che potranno aggiungere conoscenza, mediante l'inserimento di nuovi documenti con le relative immagini. A tale scopo si utilizza il linguaggio di *markup* della Text Encoding Initiative (TEI) come standard di riferimento. L'utilizzatore può:

- inserire (o collegare) nuove immagini di pagine di testo;
- inserire (o collegare) le trascrizioni del testo contenuto nelle immagini;
- collegare in forma dinamica il testo della trascrizione all'immagine corrispondente;
- inserire (o collegare) informazioni extra-testuali quali apparati e bibliografia.

3. Disporre di funzionalità necessarie alla ricerca filologica, quali:

- inserimento di annotazioni relative al testo;
- consultazione di annotazioni relative al testo;
- confronto e individuazione delle varianti del testo;
- indicizzazione del testo e delle varianti;
- selezione del testo caratterizzato da varianti;
- marcatura delle varianti d'apparato e la visione del testo marcato e della variante;
- lavoro collaborativo realizzato da più utenti in postazioni remote sul medesimo *corpus* di testi e di immagini che sono oggetto di studio, attribuendo a ciascuno studioso la propria specifica responsabilità su quanto elaborato.

Ulteriori funzionalità specialistiche sono inoltre in via di progettazione quali, per esempio:

- la gestione di versioni diverse degli stessi documenti (versionamento);
  - la possibilità di associare un ciclo di vita ai documenti condivisi (nello stesso modo di un articolo scientifico nel quale si riconoscono tre fasi, *draft*, *revision* e *published*) i quali, nelle varie fasi, sono soggetti a vedere modificati i vari livelli di autorizzazione che regolano le modalità e i tipi di intervento.
2. Effettuare annotazioni e, di conseguenza, ricerche semantiche sui testi, con la possibilità di definire un proprio modello concettuale, eventualmente derivandolo da una ontologia fondazionale, e utilizzarlo per descrivere la semantica delle entità presenti in un testo e le relazioni tra testi differenti.