

# Digitalia

Anno VI, Numero 2 - **2011**

Rivista del digitale nei beni culturali

ICCU-ROMA

# Seminario informativo DC NET. Nozioni di base sulle e-infrastructures e casi d'uso nell'ambito delle istituzioni culturali

**Maria Teresa Natale**

ICCU

Lo scorso 18 ottobre si è svolto presso la Biblioteca nazionale centrale di Roma il Seminario informativo DC NET, organizzato dall'Istituto centrale per il catalogo unico delle biblioteche italiane. Il workshop intendeva fornire nozioni di base sulle *e-infrastructures*, la loro tipologia e il loro utilizzo a *decision makers* ed esperti delle istituzioni culturali, che operano nei settori dell'ICT e della digitalizzazione del patrimonio culturale. Alcuni casi d'uso intendevano dimostrare come le *e-infrastructures* potrebbero servire efficacemente non solo il mondo della ricerca ma anche le istituzioni culturali<sup>1</sup>.

Nell'intervento d'apertura, il direttore dell'ICCU Rossella Caffo ha illustrato il senso del progetto europeo DC NET<sup>2</sup>, da lei coordinato. Il principale obiettivo di quest'iniziativa della durata di 24 mesi è favorire l'incontro tra le istituzioni culturali e le *e-infrastructures* (NREN, GRID e Cloud Computing).

La digitalizzazione massiva del patrimonio culturale, avviata da alcuni anni, ha aperto nuovi scenari che implicano una visione europea condivisa relativamente a strategie, politiche, standard, linee guida. Un percorso di dieci anni, avviato con il progetto MINERVA, che ha contribuito con i suoi risultati allo sviluppo di Europeana, l'iniziativa europea che ha dato maggior impulso alla digitalizzazione del patrimonio culturale europeo, garantendo visibilità in una prima fase soprattutto alle biblioteche e successivamente anche alle altre istituzioni culturali.

In Italia, le iniziative più importanti sono costituite da CulturalItalia, che applica gli stessi criteri di Europeana e ha la funzione di aggregatore nazionale *cross-domain* dei contenuti culturali digitali, e da Internet Culturale, l'aggregatore tematico delle biblioteche, interoperabile con CulturalItalia.

Tre sono le linee di azione a livello nazionale portate avanti dall'ICCU, nel settore

<sup>1</sup> Le presentazioni (slides, file audio e video) sono disponibili sul sito del progetto DC NET: <<http://www.dc-net.org/index.php?en/193/roma-seminario-informativo-dc-net-18-october-2011>>.

<sup>2</sup> DC NET, <<http://www.dc-net.org>>.

## Progetti

della digitalizzazione: 1) la realizzazione di progetti nazionali che mirano alla digitalizzazione di fondi rari e preziosi; 2) la partecipazione al progetto Google per la digitalizzazione di produzioni editoriali di pubblico dominio; 3) la partecipazione a progetti europei di digitalizzazione e buone pratiche.

Nel settore dei progetti europei, rientrano il progetto ATHENA<sup>3</sup>, appena concluso, che aveva l'obiettivo di favorire la partecipazione dei musei in Europea, e il progetto Linked Heritage<sup>4</sup>, avviato nel mese di aprile, che oltre a fornire a Europea contenuti digitali afferenti a istituzioni culturali intersettoriali approfondisce tematiche specifiche, come i *Linked Open Data*, le terminologie e la partnership pubblico-privato nel digitale.

Uno scenario che rende quanto mai necessaria l'apertura del mondo delle istituzioni culturali all'ambito delle *e-infrastructures* e della ricerca. Il progetto DC NET, quindi, oltre a favorire la conoscenza reciproca dei due settori, a livello strategico intende definire le priorità della ricerca sul patrimonio culturale digitale per delineare una *roadmap* per lo sviluppo di un'infrastruttura che fin dall'inizio vede tra le priorità lo *storage* e l'ampiezza di banda, ma anche grandi opportunità relativamente all'erogazione di servizi, per l'ottimizzazione delle risorse e la razionalizzazione della spesa.

Paolo Galluzzi (Museo Galileo), *keynote speaker* della giornata, ha esordito puntualizzando come il web abbia inciso in maniera radicale su un *modus operandi* radicato nella tradizione da secoli. Infatti, il modello di organizzazione delle conoscenze e dei contenuti per compartimentazioni distinte (musei, archivi, biblioteche) non ha più senso. Oggi gli oggetti vengono smaterializzati attraverso la digitalizzazione e potenziate le relazioni semantiche che li collegano, modificando la modalità di costruzione e sviluppo degli archivi culturali. Le comunità operano in uno spazio immateriale, con sistemi di pubblicazione innovativi, dove non è sempre facile distinguere l'autore e la data di pubblicazione/revisione del documento. La diffusione delle conoscenze diviene quindi "ubiqua, immediata, transeunte, autogestita, senza intermediazioni, dinamica".

Questo nuovo scenario però impone correzioni di rotta: l'accettazione di un nuovo concetto di autorialità, l'apertura a un modello di lavoro collaborativo che include anche l'apporto degli utenti, una nuova organizzazione del lavoro per produrre cultura. Un processo che sicuramente si presenta lungo e difficile, innanzitutto a causa di piante organiche obsolete e spesso inadeguate a sostenere questa rivoluzione. Va inoltre lamentata l'assenza di profili professionali specializzati (ad esempio, l'esperto della *preservation*), la scarsità di formazione e aggiornamento specifici, la mancanza di *turnover* che impedisce l'accesso nel settore dei beni culturali ai giovani, per loro natura più sensibili alle nuove tecnologie rispetto ai non nativi.

<sup>3</sup> ATHENA, <<http://www.athenaeurope.org/>>.

<sup>4</sup> Linked Heritage, <<http://www.linkedheritage.org/>>.

## Progetti

Antonella Fresa, coordinatore tecnico del progetto DC NET, è entrata nel merito delle attività del progetto ed il suo contributo è pubblicato in questo stesso fascicolo<sup>5</sup>.

Sul fronte delle *e-infrastructures*, Enzo Valente ha presentato il Consortium GARR<sup>6</sup> da lui diretto. Il GARR progetta e gestisce la rete telematica nazionale dell'università e della ricerca, offrendo a tali comunità un'infrastruttura digitale e servizi avanzati al passo con l'Europa per prestazioni e tecnologie. Consente inoltre di ottimizzare i costi di connettività a livello nazionale e internazionale, erogando servizi avanzati in base a specifiche comuni alle reti della ricerca in tutta Europa e nel mondo e garantisce l'interoperabilità di servizi e applicazioni, favorendo la partecipazione italiana a progetti europei e internazionali. Le stesse prestazioni vengono assicurate sia nelle città dell'Italia centrale, settentrionale e meridionale, sia nelle zone poco urbanizzate. GARR inoltre, oltre ad attività di formazione e divulgazione dedicate, supporta lo sviluppo di applicazioni e servizi nell'ambito dell'e-learning, dell'accesso multimediale a biblioteche e musei, della multi-videoconferenza e collaborazione remota, dell'HPC (High Performance Computing), GRID e *cloud computing*, SAN (Storage Area Network).

Valente ha anche illustrato GARR-X, un progetto di Next Generation Network che guiderà l'evoluzione della rete dell'università e della ricerca italiana per il prossimo quinquennio. L'obiettivo è quello di aumentare la flessibilità e l'efficienza del modello economico e tecnico per stare al passo con le richieste degli utenti e contribuire a ridurre il *digital divide*, offrendo gli stessi servizi su tutto il territorio nazionale e fornendo banda virtualmente illimitata e grande affidabilità anche per le attività caratterizzate da un elevato scambio di dati. Saranno così superate le differenze territoriali nella disponibilità di servizi e modalità di accesso, e arricchito il ventaglio di servizi e funzionalità di rete avanzate. Inoltre si punterà all'erogazione dei servizi ancora poco diffusi, come la banda su richiesta (*Bandwidth-on-Demand*) e la possibilità di effettuare collegamenti *end-to-end* anche a livello internazionale, realizzando reti ridisegnabili dinamicamente dagli utenti.

Federico Ruggieri (Istituto Nazionale di Fisica Nucleare) ha esordito affermando che le infrastrutture di Ricerca sono al centro del Triangolo della Conoscenza (Ricerca – Educazione – Innovazione) e che rappresentano un obiettivo chiave dell'Area Europea della Ricerca (ERA) e possono dare un contributo significativo alla Digital Agenda della Commissione Europea. In particolare, le infrastrutture GRID forniscono lo strato di servizi di calcolo ad alte prestazioni con risorse distribuite e condivise, basate su una rete ad alta velocità. "Una GRID di calcolo è un'infrastruttura hardware e software che fornisce un accesso a risorse di calcolo

<sup>5</sup> Antonella Fresa, DC NET, "Digitalia", 6 (2011), n. 2, p. 106.

<sup>6</sup> <<http://www.garr.it/>>.

## Progetti

ad alte prestazioni in maniera affidabile, coerente, pervasiva ed economica”.

Le GRID pertanto sono uno strumento estremamente potente per mettere in comune dati e risorse di calcolo. La difficoltà di apprendimento nella fase iniziale può essere superata con l'uso di portali orientati all'applicazione o servizio (*Science Gateways*). Il sistema di sicurezza di GRID si basa su un'infrastruttura a chiave elettronica pubblica/privata, fondata su algoritmi asimmetrici, in cui l'utente si autentica tramite un certificato digitale. Con gli strumenti standard vi è, pertanto, un alto livello di sicurezza, aumentabile per domini applicativi specifici. Allo stato attuale, l'Italia ha un'infrastruttura a livello europeo e mondiale (IGI) ed è trainante nelle attività di cooperazione internazionale. Non va dimenticato che alla base c'è ovviamente un'infrastruttura di rete per l'università e la ricerca d'avanguardia, sia a livello nazionale (GARR) che europeo (GEANT3).

L'intervento successivo ha visto Gabriella Contardi (Istituto centrale per il catalogo unico delle biblioteche italiane) illustrare il caso pratico del Servizio bibliotecario nazionale (SBN)<sup>7</sup>, le cui principali finalità sono la catalogazione partecipata e la disponibilità dei documenti, cui corrispondono in parallelo lo sviluppo dell'automazione delle biblioteche e la costituzione del catalogo collettivo nazionale. SBN, progetto avviato nel lontano 1984, si basa su un modello organizzativo incentrato sul decentramento e sull'articolazione territoriale in poli. Un sistema centrale configurato come “struttura di servizio alle biblioteche” gestisce l'infrastruttura per la catalogazione partecipata (Indice) e la base dati di sola lettura disponibile per la consultazione in Internet da parte degli utenti finali (OPAC), soddisfacendo circa 32 milioni di ricerche l'anno all'interno di un catalogo di 11,5 milioni di pubblicazioni con 58 milioni di localizzazioni di documenti. La relatrice ha puntualizzato che l'Indice - come infrastruttura funzionale alla catalogazione partecipata che le biblioteche effettuano online tramite un protocollo di colloquio - ha esigenze di alta affidabilità, velocità di connessione, rapida risoluzione di problemi in caso di interruzioni. L'infrastruttura di rete cui si appoggia l'Indice SBN è GARR, con un collegamento in fibra che assicura una connessione di capacità pari a 100 Mbps.

Rimanendo in ambito bibliotecario, Laura Tallandini (Sistema bibliotecario di ateneo dell'Università di Padova), ha illustrato l'esperienza del suo ateneo, presso il quale si sta implementando un sistema di autenticazione SSO per gli utenti, basato su software Shibboleth. L'opportunità di adottare questo sistema di autenticazione era nata dalla necessità di “superare la discontinuità degli accessi ai diversi servizi della biblioteca, in particolare la biblioteca digitale, e consentire una navigazione fluida, facile e completa” sia *on campus* sia *off campus*. Oggi, alcuni servizi sono ad accesso libero, mentre altri, tra i più importanti (come l'accesso a spa-

<sup>7</sup> <<http://www.sbn.it>>.

## Progetti

zi personalizzati), richiedono un'autenticazione da parte dell'utente. Si verificano spesso processi faticosi sia per l'utente stesso che per l'istituzione obbligata a mantenere sistemi di autenticazione diversi più o meno costosi. Ecco perché la decisione di utilizzare il software Shibboleth, ai fini di "facilitare la vita agli utenti", sia quelli istituzionali (docenti, ricercatori, assegnisti, *visiting professors*, tecnico-amministrativo, dottorandi, specializzandi, studenti), che hanno accesso libero a tutte le risorse elettroniche *on campus* e, via proxy, *off campus*; sia "Alumni" (cioè, neolaureati dell'Università di Padova), che hanno accesso limitato ad alcuni servizi con autenticazione, sia utenti non istituzionali che possono accedere a tutte le risorse ad accesso aperto (catalogo, archivi istituzionali ecc.) e ad alcuni servizi con autenticazione (es. aree riservate dell'OPAC), grazie a un sistema di autenticazione locale. Inoltre, visto il crescente uso dei dispositivi personali mobili l'ateneo sta studiando soluzioni per garantire l'accesso alle risorse online anche a coloro che si collegano via WiFi dalle biblioteche. Al termine dell'intervento Tallandini ha anche presentato il Portale Biblioteca Digitale che consente l'accesso integrato libero alle risorse elettroniche attraverso un'interfaccia di ricerca unica e intuitiva e che di fatto è il *gateway* verso tutti i servizi digitali di supporto al dominio della didattica e al dominio della ricerca (banche dati, riviste, cataloghi, collezioni digitali, siti web...). Attraverso il *gateway* Biblioteca Digitale si accede (al momento con autenticazione) anche al Portale per l'Accesso Integrato alle Risorse Elettroniche (AIRE), strumento che consente la ricerca simultanea su più risorse bibliografiche e l'accesso diretto ai testi pieni. Se oggi quindi per poter utilizzare tutte le funzionalità del portale sono richiesti due o tre step di autenticazione in sequenza, a seconda della modalità di accesso *on campus* od *off campus*, in un futuro molto vicino lo sviluppo Single sign-on (SSO) consentirà uno snellimento di queste procedure. Un attento impiego dello strumento tecnologico Shibboleth è dunque in grado di eliminare la necessità di autenticazioni in serie senza compromettere le regole dei contratti di accesso alle risorse elettroniche proprietarie; un problema può però essere rappresentato dalla necessità che gli editori configurino l'accesso alle loro risorse in modo Shibboleth-compliant.

La sessione mattutina si è conclusa con l'intervento di Iolanda Rolfo che in maniera organica ha illustrato il sito web del Museo Galileo di Firenze (già Museo di storia della scienza), istituzione culturale privata, che investe un terzo del bilancio dell'istituto per le attività e la gestione del laboratorio digitale interno.

Attivato con un primo collegamento a Internet a 2 Mbps nel 1994, rinnovato con maggiori funzionalità dieci anni dopo, dal 2009 il museo è entrato nella rete FiNet (1 Gbps in fibra ottica), che consente un'erogazione del servizio più affidabile, veloce e di qualità. Oggi dal sito web, rivolto a diverse fasce di utenza – *scholars*, *educational/edutainment*, turismo culturale – si accede a circa 1.000 schede di strumenti, più di 400 schede biografiche e altrettante schede di approfondimento,

## Progetti

1.800 immagini; alla ricostruzione tridimensionale esplorabile delle sale; a più di 150 video da vedere in streaming o da poter scaricare per una visione *offline* in ottima qualità; a un numero sempre crescente di mostre virtuali e apparati didattici interattivi.

Roberto Barbera (Università di Catania - Consorzio COMETA, partner di DC NET) ha spiegato come le biblioteche digitali su GRID possano contribuire a preservare e diffondere il patrimonio umanistico. Parallelamente all'evoluzione delle tecnologie dell'informazione e della comunicazione (mediante la potenza di un computer raddoppia ogni 18 mesi, la densità dei dischi ogni 12 mesi, la velocità delle reti ogni 9 mesi), il mondo della ricerca opera su piattaforme sempre più "distribuite" e "computazionali". L'e-Science implica l'esistenza di Organizzazioni virtuali di utenti che attraverso le *e-infrastructures* usufruiscono di applicazioni, dati, strumenti. Le *e-infrastructures* infatti offrono: sistemi di calcolo ad alte prestazioni e ad alto *throughput* (capacità di trasmissione effettivamente utilizzata), consentendo di affrontare e risolvere problemi computazionali di grandi dimensioni; sistemi di memorizzazione di massa ad alta densità; servizi di *storage* e *digital preservation*; sistemi di autenticazione diffusi e semplici da utilizzare; sistemi sofisticati di autorizzazione (chi può far cosa). Molte quindi sono le opportunità per le istituzioni culturali che potrebbero accrescere l'efficienza dei propri servizi – tanto per far qualche esempio – nei settori del restauro digitale, dello studio degli effetti ambientali, dell'analisi dei suoni, della replica geografica dei dati, dell'accesso federato ai dati, delle performance virtuali.

Il relatore ha poi illustrato un caso pratico: l'e-Culture Science Gateway<sup>8</sup> realizzato dal progetto europeo INDICATE, ovvero una piattaforma per la gestione di archivi digitali su GRID, di cui ha illustrato l'architettura, i meccanismi di autenticazione e autorizzazione, i casi d'uso. Questo pilot rappresenta un notevole passo avanti verso un uso più diffuso e più semplice delle *e-infrastructures* da parte di comunità di non-esperti in campo informatico. Esso intende dimostrare come il supporto alle Federazioni di Identità permetterà un'interazione più forte tra fornitori di *e-infrastructures* e operatori/istituzioni culturali e come esso potrà essere riutilizzato in tutti i contesti nei quali occorra creare ed accedere ad archivi digitali distribuiti in modo semplice, ancorché sicuro.

L'ultima parte del workshop è stata strutturata in forma di *tutorial* sulle *e-infrastructures*, molto utile per chi si avvicinava per la prima volta a questo ambiente. Gabriella Paolini (Consortium GARR) ha spiegato dal punto di vista tecnico il funzionamento di Internet: un insieme di reti interconnesse fra di loro su tutto il pianeta, con regole di funzionamento condivise e rispettate perché tutto funzioni,

<sup>8</sup> <<http://indicate-gw.conorzio-cometa.it/>>.

## Progetti

non governata da un gruppo, né da un ente, né da un'azienda, né gestito in modo centralizzato perché tutte le singole reti che formano Internet hanno una gestione autonoma. Dopo aver sottolineato che Internet non è sinonimo di World Wide Web, la relatrice si è soffermata diffusamente sui protocolli di comunicazione di dati che definiscono le regole per coordinare lo scambio di informazioni tra i diversi dispositivi e i processi che consentono questo scambio (pile, pacchetti ecc.).

Il workshop si è concluso con Sabrina Tomassini (Consortium GARR) che ha presentato un glossario commentato, utilissimo supporto per una comprensione basilica delle *e-infrastructures*. Ha chiarito tutta una serie di termini sempre più frequentemente usati, a partire da *Infrastruttura digitale*, utilizzato per indicare un insieme di risorse di calcolo, risorse di memorizzazione e applicazioni interconnesse da reti ad altissima capacità. Reti che mirano a dare un impulso alle attività della ricerca garantendo un facile e affidabile accesso alle strutture, alle risorse e ai mezzi di collaborazione, sfruttando la potenzialità della tecnologia. Sintetizzando, le infrastrutture digitali forniscono un semplice accesso su scala globale in modalità dedicata attraverso le quali il ricercatore può essere localizzato in qualsiasi parte del mondo. Possono essere raggiunti anche gruppi di ricerca molto piccoli e/o provenienti da diverse discipline, localizzati anche nei paesi in via di sviluppo.

L'adozione di queste nuove tecnologie è necessaria per gestire e memorizzare grandi moli di dati, garantire accesso trasparente a milioni di file da diversi domini amministrativi, mantenere basso il costo dell'accesso alle risorse. Solo attraverso le *e-infrastructures* è possibile avviare attività collaborative da gestire tramite piattaforme complesse e distribuite in grado di offrire nuove applicazioni e stabilire un nuovo approccio alla creazione di gruppi di ricerca, facilitando di molto l'accesso per le attività di formazione.