

Nuove forme di fruizione del digitale. Il repository Byterfly: la conservazione e il riuso con l'open source, access e data

Giancarlo Birello, Anna Perin

CNR-IRCrES, Ist. di Ricerca sulla Crescita Economica Sostenibile, Moncalieri (TO)

Abstract. Questo paper descrive l'esperienza maturata dall'Ufficio IT e dalla Biblioteca CNR-IRCrES nello sviluppo di una biblioteca digitale, Byterfly, con requisiti di open source per architettura e software, open access per contenuti e open data per filosofia di conservazione, fruizione e riuso.

Keywords. Biblioteca digitale, repository, open source, open access, open data

Introduzione

Un repository è una biblioteca digitale con una raccolta mirata di oggetti che può includere testi, immagini, file audio e video, tutti archiviati in collezioni e associati a metadati che forniscono informazioni sulla risorsa. Tali oggetti devono essere ricercabili, accessibili e fruibili. L'ufficio IT e la biblioteca del CNR-IRCrES hanno sviluppato e gestiscono alcuni repository dal 2012, Byterfly è uno di questi (Birello et al., 2018) con requisiti di open source per architettura e software, open access per contenuti e open data per filosofia di conservazione, fruizione e riuso.

1. La biblioteca digitale

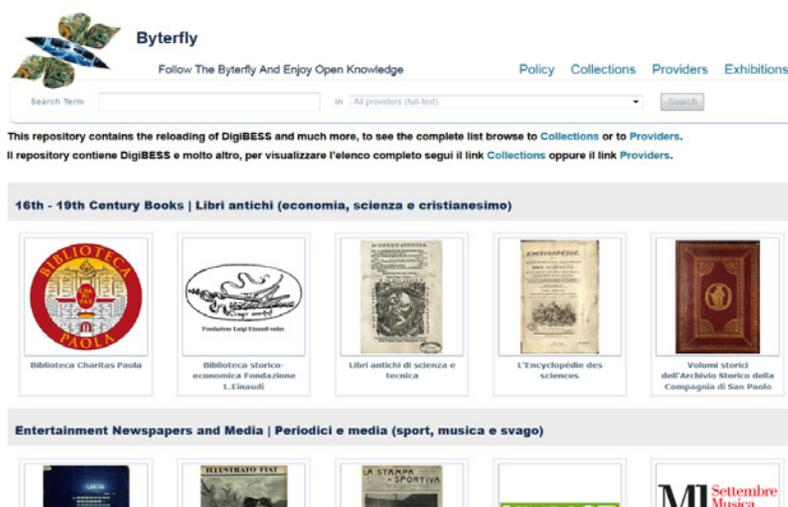
Per la sua realizzazione è stato utilizzato Fedora Repository, uno dei software più adottati a livello mondiale per la gestione delle risorse digitali, robusto, persistente e in grado di gestire big data. Per il front-end è stato utilizzato Dupal CMS e Islandora per l'accesso e la presentazione degli oggetti digitali, dotato di moduli espandibili e personalizzabili per tutti i tipi di media. Per le funzioni di ricerca è stato utilizzato Solr. I metadati adottati sono i Dublin Core. L'architettura è OAI-PMH e IIIF Image and Presentation API 2.1 compliant, il che consente la visibilità e la disseminazione sia attraverso meta-repository (Worldcat per esempio), ma anche semplicemente dai motori di ricerca tradizionali. Attualmente Byterfly ospita ventuno fornitori di contenuti (provider). Ogni provider può avere una o più collezioni. Le collezioni presenti (65) sono molto diversificate, ma consideriamo questo un punto di forza e non un limite.

Nella biblioteca digitale è depositato molto materiale sulle scienze sociali ed economiche del Piemonte inclusi volumi dell'Università di Torino (appartenenti alla biblioteca Bobbio e alla biblioteca di Economia e Management), della Fondazione Einaudi,

periodici del Gruppo Fiat (FGA Automotive), della Camera di Commercio di Torino, del Museo dell'Automobile di Torino, libri d'arte (provider Fondazione 1563 per Arte e Cultura di Torino). Inoltre, ospitiamo libri antichi (sec. XVI-XVIII) di teologia e religione (provider: Ordine dei Minimi di San Francesco di Paola), volumi antichi di scienza e tecnica (provider: CNR-IRCrES) e materiali, libri e immagini del festival musicale internazionale MITO Settembre Musica organizzato ogni anno a Torino e Milano. I documenti sono in italiano, latino (libri antichi), inglese, francese, spagnolo e tedesco.

Per meglio orientare l'utente alcune collezioni sono state visivamente raggruppate nella homepage in 5 macro aree: Libri antichi, periodici e media, riviste di economia e impresa, riviste di scienze sociali, libri moderni di economia e scienze sociali. Attualmente sono depositate ed accessibili più di 1.300.000 pagine, in continua espansione.

Fig. 1
Home page Byterfly
www.byterfly.eu



2. L'architettura

L'architettura si basa su una piattaforma virtualizzata che utilizza KVM come hypervisor in combinazione con OpenVSwitch per il networking e iSCSI con multipath per gestire la SAN (Storage Area Network).

L'architettura attuale richiede tre macchine virtuali:

- il back-end, che include il repository per gestire gli oggetti digitali;
- il server IIIF (International Image Interoperability Framework) per il trattamento delle immagini;
- il front-end, che ospita l'interfaccia per l'accesso pubblico agli oggetti e ai dati del repository.

Uno storage altamente affidabile (SAN) completa l'architettura fornendo al repository un ampio spazio per la conservazione degli oggetti e all'hypervisor lo storage per i backup delle macchine virtuali.



Fig. 2.
Architettura Byterfly
a tre macchine
virtuali

Stessa architettura è utilizzata anche per il repository ASA – Archivio Studi Adriatici (Armeli Minicante et al., 2017).

La macchina virtuale di back-end ospita Fedora Repository, BlazeGraph e Solr, ognuno dei quali viene eseguito in un contenitore servlet Java separato: i primi due, in un paio di istanze di Tomcat, mentre Solr è in esecuzione in un contenitore servlet Jetty incluso nella distribuzione di Solr. Fedora Repository gestisce l'attività di conservazione dei contenuti della Byterfly e include le API per l'ingesting programmabile, la descrizione semantica delle relazioni tra gli oggetti e la loro gestione basata su modelli. BlazeGraph è il gestore del triple-store delle relazioni tra gli oggetti del repository. Infine, Solr è il motore di ricerca e indicizzazione di tutti gli oggetti del repository. La configurazione è personalizzata per indicizzare anche i full-text dei volumi oltre a tutti gli elementi Dublin Core dei metadati al fine di fornire i campi da utilizzare nel front-end Islandora, nelle pagine di ricerca e visualizzazione.

Il server IIIF Cantaloupe fornisce l'accesso alle immagini conservate nel repository tramite le Image API 2.1 dello standard. Può accedere sia alla versione JPEG2000 che TIFF delle immagini per poter fornire al front-end i riquadri opportunamente scalati richiesti dai visualizzatori OpenSeaDragon e Internet Archive Bookreader. Cantaloupe è un generatore di derivative ad alte prestazioni con potenti funzionalità di cache ed estremamente versatile, in grado di fornire immagini a più coppie repository/front-end contemporaneamente.

Server front-end "Open-access" significa poter accedere a tutti gli oggetti digitali e ai dati conservati nel repository in modo semplice. Questo è uno dei motivi principali per cui abbiamo scelto Islandora molti anni fa e abbiamo confermato la scelta anche per Byterfly. Il framework Islandora è integrato in Drupal CMS, supportato da una community dinamica ed estesa. Reso inoltre interessante per il supporto IIIF e l'uso intensivo di Views personalizzabili per visualizzare nel front-end collezioni, oggetti e metadati in base alle più svariate esigenze. Non ultima la recente implementazione delle IIIF Presentation API 2.1 tramite un modulo di Islandora che rendono i contenuti della Byterfly compatibili con gli standard richiesti da Europeana.

3. Le policy

Le policy adottate per il repository sono disponibili in una pagina web sul portale e si basano sulle indicazioni fornite da OpenDOAR (The Directory of Open Access Repositories) che, con un semplice strumento online, aiuta a formulare e presentare le policy del proprio repository con opzioni raccomandate minime per la conformità di adesione al movimento Open Access. Le policy coprono tutti gli aspetti del repository, in particolare: Policy dei metadati, dei dati, dei contenuti, di sottomissione e di conservazione.

4. Mostre digitali

È stato inoltre realizzato un sito di mostre digitali (exhibitions) che consente una fruizione trasversale del materiale contenuto nella biblioteca digitale. Attualmente è presente la mostra “1919. Cronache dalla Storia: immagini e documenti d’archivio. Dal quotidiano alla voglia di cambiamento”. Realizzata in occasione del centenario della fine della Grande Guerra (1915-1918), offre un percorso organizzato attorno a quattro grandi temi: “La ripresa delle attività”, “Economia e politica”, “La donna e la società”, “Lo sport e lo svago”.

5. Sviluppi futuri

Da diversi anni collaboriamo allo sviluppo del software per la realizzazione di repository con alcune comunità dedite all’open source, negli ultimi mesi è nata una stretta collaborazione con il Metropolitan New York Library Council (metro.org) che porterà nei prossimi mesi al rilascio di una nuova architettura, semplificata e flessibile che prevede metadati dinamici, altamente integrata con Drupal e che sfrutta a pieno le potenzialità del IIIF.

6. Conclusioni

Obiettivo della Byterfly è mettere al centro ogni collezione ospitata e, anche se piccola, sfruttare i vantaggi dell’architettura generale. Massima attenzione alla fruibilità degli oggetti, semplicità di utilizzo della piattaforma da parte dell’utente e preservazione a lungo termine dei materiali contenuti.

Riferimenti bibliografici

Birello G., Perin A. (2018) Follow the Byterfly and enjoy open knowledge, Rapporto Tecnico CNR-IRCrES, n. 6/2018. <http://dx.doi.org/10.23760/2421-5562.2018.006>

Armeli Minicante S., Birello G., Sigovini M., Minuzzo T., Perin A., Ceregato A. (2017): Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algarium Veneticum and the Archivio di Studi Adriatici Case Study, *Journal of Library Metadata*, Volume 17, Issue 2, Pages: 111-125, DOI:10.1080/19386389.2017.1355165

Autori



Giancarlo Birello - giancarlo.birello@ircres.cnr.it

È System e Network Manager presso l'ufficio IT di CNR-IRCrES, referente regionale per l'infrastruttura di rete CNR Piemonte e Access Point Manager per il GARR. Ha iniziato ad usare programmi open source per la conservazione e presentazione di oggetti digitali dieci anni fa, ha recentemente implementato la biblioteca digitale Byterfly. È coinvolto in alcune comunità internazionali per lo sviluppo dei software utilizzati.

Anna Perin - anna.perin@ircres.cnr.it

È responsabile della Biblioteca CNR-IRCrES dove svolge tutte le attività di back office e front office per i propri utenti, servizi di reference e di orientamento per l'utilizzo delle risorse della biblioteca e document delivery. Si occupa della realizzazione e gestione dei siti web di Istituto. Ha sviluppato competenze sui repository di ultima generazione occupandosi in particolare di metadati, policy, usabilità e fruibilità per l'utente.

