

ISSN (print): 2421-5783  
ISSN (on line): 2421-5562



Consiglio Nazionale delle Ricerche

**IRCES**

ISTITUTO DI RICERCA SULLA CRESCITA ECONOMICA SOSTENIBILE  
RESEARCH INSTITUTE ON SUSTAINABLE ECONOMIC GROWTH

# *Rapporto Tecnico*

*Numero 3, Agosto 2016*

FABB Repository dal progetto al prototipo.  
Nuove forme di conservazione, condivisione  
e valorizzazione di opere digitali

*Giancarlo Birello, Ivano Fucile, Valter Giovanetti, Anna Perin*



RAPPORTO TECNICO CNR-IRCRES

Anno 2, Numero 3, Agosto 2016

*Direttore Responsabile*

Secondo Rolfo

*Direzione e Redazione*

CNR-IRCRES

*Istituto di Ricerca sulla crescita economica sostenibile*

Via Real Collegio 30, 10024 Moncalieri (Torino), Italy

Tel. +39 011 6824.911

Fax +39 011 6824.966

segreteria@ircres.cnr.it

www.ircres.cnr.it

*Sede di Roma*

Via dei Taurini 19, 00185 Roma, Italy

Tel: 06 49937809

Fax: 06 49937808

*Sede di Milano*

Via Bassini 15, 20121 Milano, Italy

Tel: 02 23699501

Fax: 02 23699530

*Sede di Genova*

Università di Genova Via Balbi, 6 - 16126 Genova

Tel: 010-2465.459

Fax: 010-2099.826

*Segreteria di redazione*

Enrico Viarisio

enrico.viarisio@ircres.cnr.it



# FABB Repository dal progetto al prototipo. Nuove forme di conservazione, condivisione e valorizzazione di opere digitali

*Giancarlo Birello, Ivano Fucile, Valter Giovanetti*

*Ircres-CNR Ufficio IT*  
Strada delle Cacce, 73  
10135 Torino Italy

*Anna Perin\**

*Ircres-CNR Biblioteca*  
Via Real Collegio, 30  
10024 Moncalieri TO Italy

**ABSTRACT:** FABB project (Famine and Feast, Fame e Abbondanza) has been committed by Fondazione CRT. This technical report analyzes the strategies adopted and the main open-source software used. Ircres-CNR has deployed the software and server platforms of the repository, in a virtualized and redundant infrastructure, it also take care of the design, development and management of the web portal (front-end) for the presentation, research and consulting data of the digitalized items (lyrics, lyrics text, interviews, books, poems).

**KEY WORDS:** open-source, islandora, repository, digital archive, cultural heritage

**JEL CODES:** Z11

---

\*Corresponding author: [anna.perin@ircres.cnr.it](mailto:anna.perin@ircres.cnr.it)

## Indice

1	Introduzione.....	4
2	Architettura.....	5
2.1	Hardware.....	6
2.1.1	Storage.....	6
2.1.2	Hypervisor.....	7
2.2	Storage.....	8
2.2.1	Macchine virtuali e backup.....	8
2.2.2	Dati.....	9
2.3	Applicazioni.....	10
2.3.1	Repository: Fedora Repository.....	11
2.3.2	Front-end: Drupal.....	12
2.3.3	Framework: Islandora.....	12
2.3.4	Piattaforma di ricerca: Solr.....	13
3	Sito web di presentazione.....	15
4	Conclusioni.....	17
5	Sitografia.....	18
6	Bibliografia.....	19

## 1 Introduzione

L'Ufficio IT e la Biblioteca dell'Ircres-CNR da alcuni anni sono coinvolti in progetti di post-digitalizzazione in particolare per la gestione di tutta la fase di messa in rete di materiale con criteri di open-accessibility e utilizzando software open-source.

Il primo di questi progetti è DigiBESS, il repository di *Bess - Biblioteca Elettronica di Scienze Economiche e Sociali del Piemonte*, gruppo di 18 biblioteche e centri di documentazione di scienze economiche e sociali Piemontesi. La digitalizzazione riguarda materiale relativo all'economia e ai processi di trasformazione sociale ed economica della regione Piemonte con la finalità di conservazione e preservazione di materiale anche di difficile reperimento perché ormai fuori stampa o relativo a letteratura grigia, sono presenti materiali sia delle biblioteche di Bess che di altri enti Piemontesi che hanno dimostrato interesse per l'iniziativa. Il laboratorio DigiBESS ha sede presso l'Area della Ricerca CNR di Torino e, grazie al finanziamento della Compagnia di San Paolo che copre i costi del personale addetto alla digitalizzazione, ha una produzione costante che ha già superato ampiamente le 800.000 pagine.

Il progetto V2P2, riguarda invece la messa in rete di materiale dell'IPSP-CNR, quaderni di laboratorio, immagini, diapositive e dati della ricerca. Sono state in questo caso adottate nuove soluzioni rispetto DigiBESS utilizzando software più performanti e responsivi. Il progetto è attualmente in fase di stallo in attesa di finanziamenti per il suo sviluppo.

Infine il repository FABB di cui tratteremo le caratteristiche peculiari nelle pagine che seguono, *FABB significa Fame e abbondanza*, il progetto, realizzato con il contributo della [Fondazione Cassa di Risparmio di Torino](#), ha lo scopo di raccogliere e valorizzare documenti di diversa natura che toccano i temi della fame e dell'abbondanza così come questi sono raccontati in Piemonte, soprattutto in un contesto popolare. Il materiale raccolto comprende diverse forme della narrazione breve (fiabe, poesie, canzoni, racconti, memorie e testimonianze). Gli oggetti presentati sono parte di quel patrimonio culturale immateriale che focalizza le luci e le ombre dei percorsi di una storia minore che ha contribuito alla fondazione di una identità collettiva, intrecciata all'atmosfera che si respira in un determinato territorio, ai suoi colori, odori, suoni, alle parlate della gente..., insomma al *genius loci*, allo spirito del luogo.

Partendo quindi dalle caratteristiche disomogenee del materiale da caricare: estratti di libro, canzoni, testi di canzoni, interviste, video siamo già passati per il repository dalla fase progetto alla fase prototipo con alcune interessanti soluzioni adottate, ad esempio la trascrizione delle interviste, o l'arricchimento con un glossario.

Per tutti i repository la gestione della parte successiva alla digitalizzazione comincia con la disponibilità di spazio di memorizzazione per grossi volumi di dati per il deposito delle opere digitalizzate, con caratteristiche di stabilità, versatilità e dinamicità. Per proseguire con lo sviluppo e adattamento delle piattaforme server e dei software, inseriti su infrastrutture virtualizzate e ridondate che includono la progettazione e implementazione dei sistemi di repository, dedicati a mantenere, indicizzare e collegare i dati prodotti dalla digitalizzazione. Infine la progettazione, sviluppo e gestione dei portali web (front end) per la presentazione, ricerca e consultazione dei dati, in grado inoltre di gestire le policy di accesso ai dati e le relative autorizzazioni, la manutenzione ordinaria e consulenza continuativa sulle problematiche sistemiche e del software.

## 2 Architettura

L'architettura complessiva del repository FABB presenta una certa complessità, come si può rilevare dalla figura, riportandoci alla mente l'analogia con un puzzle.

Primo aspetto la varietà e la quantità di tecnologie coinvolte che hanno richiesto approfondita analisi per essere applicate e adattate alle nostre esigenze. Partiamo dalle nozioni base sistemistiche di multipath per la connessione ridondata allo storage e di virtualizzazione fino ad arrivare agli aspetti specifici di gestione di applicativi java e php, il tutto farcito da standard quali xml e xslt unito ad aspetti più prettamente bibliografici quali i metadati Dublin Core, applicati a oggetti digitali e inseriti in modelli legati da relazioni semantiche.

Altro aspetto interessante del puzzle è l'utilizzo quasi esclusivo di soluzioni open-source, che da un lato permettono un notevole risparmio economico e la possibilità di sfruttare quanto di più innovativo è disponibile nella comunità, ma per contro richiedono un grosso lavoro di ricerca e adattamento delle varie componenti per ottenere e soddisfare le nostre esigenze specifiche.

Procediamo ora alla descrizione dell'architettura suddividendola in tre strati: quello hardware, quello specifico dei file system e infine quello applicativo.

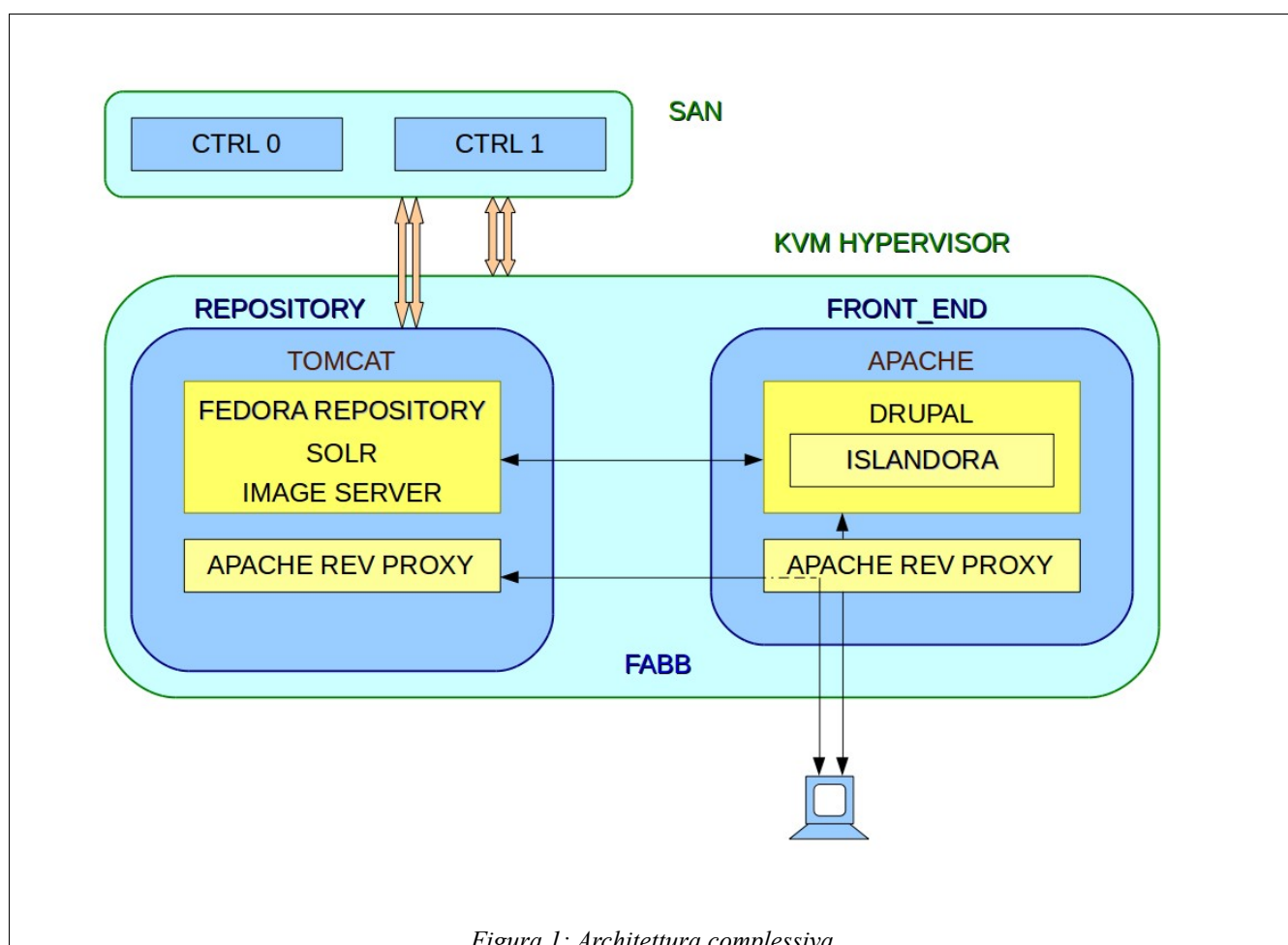


Figura 1: Architettura complessiva

## 2.1 Hardware

Le principali apparecchiature coinvolte nello sviluppo del progetto sono lo storage e il server impiegato come hypervisor per ospitare le macchine virtuali. Sono esclusi da questa trattazione i vari apparati di rete coinvolti nelle interconnessioni quali gli switch, i firewall ed ogni altro elemento relativo l'infrastruttura di rete.

### 2.1.1 Storage

Per lo storage è stata scelta una SAN Dell, il PowerVault MD3800i, con un buon rapporto qualità/prezzo, con doppio controller e supporto iSCSI.

Lo storage rende disponibili i vari volumi in cui è suddiviso lo spazio di memorizzazione tramite protocollo iSCSI. A differenza della soluzione a 2-nodi per il progetto DigiBESS, la ridondanza è a livello di controller quindi occorre che sul server connesso alla SAN sia attivata la funzione multipath, che in caso di fail-over provvede a commutare la connessione sul controller di backup. Nei test ha dato ottimi risultati di prestazioni ed affidabilità il modulo nativo Linux multipath senza dover utilizzare soluzioni proprietarie e non open-source.

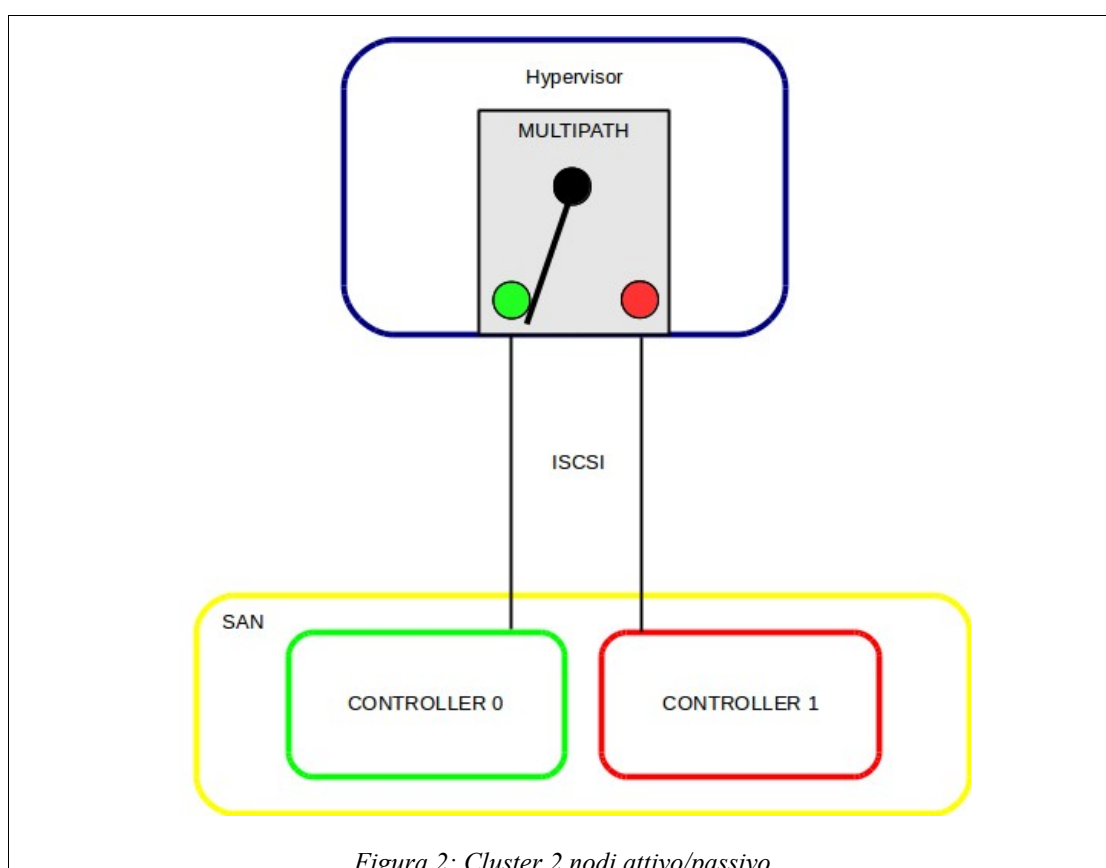


Figura 2: Cluster 2 nodi attivo/passivo

### 2.1.2 Hypervisor

All'inizio del progetto DigiBESS l'infrastruttura di virtualizzazione era basata su una soluzione proprietaria, dopo alcuni anni si è migrato a una soluzione open-source con KVM (Kernel-based Virtual Machine), ancora attuale e altamente affidabile, adottata da FABB, su cui vengono implementati tutti i server virtuali dell'Infrastruttura di rete.

Principali caratteristiche di KVM sono di essere una soluzione open-source supportata da una comunità molto attiva, la stabilità e l'affidabilità, la disponibilità di comodi strumenti di gestione ed il supporto multi-piattaforma dei sistemi operativi delle macchine virtuali. Inoltre l'installazione è molto pratica, nel caso della distribuzione Ubuntu server da noi utilizzata, può esser fatta contestualmente all'installazione del server stesso.

KVM si integra perfettamente con l'architettura LVM di gestione dello spazio disco e di cui parleremo nel capitolo seguente.

Per la gestione del networking delle macchine virtuali si è scelto OpenVSwitch, soluzione open-source perfettamente integrata con KVM, che permette di soddisfare tutte le esigenze di gestione delle VLANs, i protocolli IPv4 e IPv6 e le interfacce multiple delle macchine virtuali. Si è rivelato uno strumento altamente affidabile e potente e, dalla versione 14.04 di Ubuntu server, installabile direttamente dai repository ufficiali della distribuzione Ubuntu.



## 2.2 Storage

La diversità dei vari filesystem chiamati in causa dalle molteplici applicazioni e dai sistemi coinvolti nella realizzazione del progetto così come il loro posizionamento merita una nota particolare. Si va dalle implementazioni in grado di supportare lo storage di rete (SAN) alla memorizzazione annidata all'interno di altri file system per la gestione degli oggetti.

In particolare analizziamo la struttura di virtualizzazione e di memorizzazione dei dati del repository.

### 2.2.1 Macchine virtuali e backup

Gli hypervisor sono installati su server dotati di un proprio storage ad elevata affidabilità, dischi SAS (Serial-attached SCSI) in mirror che garantiscono anche un certo livello di prestazioni. Su questo storage locale sono di norma memorizzati i filesystem virtuali delle macchine ospitate sull'hypervisor, in particolare la partizione di sistema dei vari server. Gli hypervisor sono connessi inoltre ad alcune partizioni sulla SAN a loro dedicate, tramite protocollo iSCSI, potendo quindi usufruire, oltre dello storage locale, anche di questo ampio spazio di memorizzazione.

Data la versatilità ed affidabilità di LVM e la sua integrazione con KVM, i dischi delle macchine virtuali sono creati come Logical Volume all'interno dei Volume Group di LVM, che a loro volta sono basati su Physical Volume corrispondenti a dischi locali all'hypervisor.

I dischi virtuali così definiti possono sfruttare completamente tutte le caratteristiche di LVM quindi essere facilmente estesi o ridotti di dimensione, farne il backup a caldo grazie alla funzione di snapshot e poter ospitare un'ampia varietà di filesystem.

Per il backup delle immagini delle macchine virtuali si utilizza uno script open-source in perl (realizzato da Daniel Berteaud), molto pratico e veloce che sfrutta appieno gli snapshot di LVM e la possibilità di effettuare la copia del disco di sistema senza spegnere la macchina virtuale. Il backup viene memorizzato direttamente sulla SAN ed è quindi disponibile in caso di necessità per il ripristino sia sullo stesso hypervisor che su uno differente previa configurazione di una nuova macchina virtuale e la copia in un Logical Volume dell'immagine del disco. In quest'ultimo caso la funzione di backup può essere utilizzata anche per la migrazione delle macchine virtuali tra differenti hypervisor.

I server virtuali che necessitano di spazio particolarmente ampio, come ad esempio quello che ospita il repository, oltre alla partizione di sistema, sono connessi a partizioni direttamente sulla SAN, in questo caso il backup consiste sempre nella sola immagine del disco di sistema, per ovvie ragioni di spazio e di inutilità, essendo le partizioni sulla SAN già ridondate implicitamente.

### 2.2.2 Dati

Si è detto che per il repository si utilizza una connessione diretta tra la macchina virtuale e la partizione dati sulla SAN, tramite protocollo iSCSI, senza intermediazioni. Il filesystem con il quale viene formattato lo spazio remoto è a discrezione del sistema operativo del server e può quindi essere utilizzato uno qualsiasi o il più opportuno a seconda del caso, tenendo presente che lo spazio sarà acceduto solo e sempre da un'unica macchina alla volta.

Su questa base sono memorizzati gli oggetti del repository, assimilati a dei BLOB (Binary Large Objects) e in quanto tali è opportuna una forma efficiente di lettura e scrittura, ragione per cui, il repository utilizza un sistema di store specifico per questi oggetti. In particolare Fedora Repository utilizza Akubra data store, un progetto open-source di un'interfaccia per la memorizzazione di file adattabile ai principali sistemi di storage e tesa al più alto livello di interoperabilità tra i diversi sistemi. Per Akubra un Blob è un bitstream di lunghezza finita dotato di un identificativo e un Blob Store è concepito principalmente per fornire l'accesso in lettura e scrittura di un Blob.

Le opzioni di configurazione del repository permettono poter posizionare su una partizione qualsiasi del sistema lo store Akubra, nel nostro caso individuato su una partizione coincidente col volume iSCSI dello storage remoto. In questo modo abbiamo garantita l'affidabilità del sistema, la ridondanza dell'informazione e una veloce procedura di ripristino in caso di guasto del server, essendo sufficiente connettere al volume iSCSI un nuovo server che avrà a disposizione l'intero store del repository.

## 2.3 Applicazioni

Tutti gli applicativi utilizzati nel progetto richiederebbero una trattazione a parte per la loro importanza ed estensione, ci limiteremo quindi a segnalare delle principali applicazioni gli aspetti più rilevanti ai fini del progetto.

Innanzitutto abbiamo scelto di separare e distribuire su due server distinti le componenti relative al repository da quelle di presentazione del front-end. In questo modo si ottiene un miglioramento delle prestazioni generali del sistema ma soprattutto una semplificazione delle operazioni di mantenimento dei sistemi potendo agire in modo autonomo sui due server.

La suddivisione risponde anche alla logica di separare le applicazioni java da quelle php: le prime inserite in un unico contenitore Tomcat sul server del repository e le seconde dentro Apache sul secondo server previsto per la parte di front-end. Questa semplificazione permessa dalla natura propria delle applicazioni vede come unica eccezione l'installazione, sul server del repository, del web server Apache con la funzione di reverse-proxy per le applicazioni installate in Tomcat.

Riguardo il server Tomcat che ospita le numerose applicazioni java, si è scelto di non utilizzare il componente incluso nella distribuzione di Fedora Repository, ma di installarne uno esterno incluso nei pacchetti della distribuzione linux del server. Questa scelta ha come ragioni le maggiori prestazioni che si possono ottenere, come anche consigliato nei manuali d'installazione di Fedora Repository, oltre la semplicità di manutenzione del componente stesso essendo legato alla distribuzione del server e potendo così sfruttare le patch fornite con il sistema operativo.

All'interno di Tomcat sul server del repository troviamo varie applicazioni, che qui elenchiamo per completezza, delle quali solo alcune verranno trattate più nello specifico, in particolare abbiamo:

- adore-djatoka, trattamento immagini
- fedora, il repository server
- fedoragsearch, il motore di ricerca di fedora
- solr, motore di ricerca e indicizzazione

### 2.3.1 Repository: Fedora Repository

Il cuore dell'intero progetto è sicuramente il repository, cioè il sistema di memorizzazione e gestione degli oggetti digitali. Data l'esperienza positiva precedente con DigiBESS, Fedora Repository è stato il prodotto prescelto per FABB, sicuramente di spicco tra tutte le soluzioni possibili, sia per la diffusione che ha in questo settore sia per le numerose potenzialità che offre.

Il nome è l'acronimo di “Flexible Extensible Digital Object Repository Architecture”, il software è stato sviluppato in Java ed è frutto della comunità open-source, la cui fervida attività dimostra l'interesse diffuso esistente su questo prodotto, dagli aspetti più interni del sistema a tutte le varie applicazioni sviluppate a contorno.

Senza scendere nei particolari dell'architettura, cerchiamo di dare una panoramica su quello che Fedora Repository offre, in particolare la prima nota riguarda il fatto che il prodotto in sé è una base astratta che definisce un contesto e delle regole per la gestione di oggetti digitali. Queste assunzioni iniziali sono fatte pensando a sistemi di conservazione di opere digitali a lungo termine, in grado di gestire grossi volumi di dati e in modo flessibile per poter trattare i più svariati tipi di oggetti.

Nel repository gli elementi base sono gli oggetti che possiedono uno o più componenti chiamati “datastream”, i quali sono a loro volta dei contenuti, ad esempio un'immagine, oppure metadati che descrivono l'oggetto. I datastream possono essere memorizzati localmente sul server oppure referenziati tramite un url esterno. Gli oggetti posso dichiarare una o più relazioni con altri oggetti all'interno del repository, tramite asserzioni semantiche del tipo *soggetto-predicato-complemento*, che costituisce sicuramente una delle caratteristiche più evolute e con grandi potenzialità che offre questa architettura.

Quanto descritto finora abbinato alla possibilità di generare datastream virtuali, ad esempio la generazione di una miniatura da un'immagine ad alta definizione, in aggiunta all'identificazione univoca tramite namespace e Persistent Identifier (PID) degli oggetti, conferiscono al prodotto il carattere di repository durevole nel tempo di contenuti fruibili via web.

L'architettura di questo sistema è stata pensata per poter offrire la definizione personalizzata da parte dell'utente di modelli di oggetti, restando ferme le caratteristiche sopra elencate, potendo così sfruttare le potenzialità dell'ambiente adattato alle esigenze specifiche della soluzione. Un esempio tipo è proprio la nostra implementazione dove vengono utilizzati modelli quali collezioni, libri e pagine legati da semplici relazioni come “è membro della collezione” o “è parte di”, oggetti che prevedono datastream funzione di quanto il sistema di scansione produce, ossia immagini ad alta risoluzione, testo prodotto dal sistema di OCR e file pdf ricercabili.

Possiamo quindi concludere, tralasciando l'approfondimento di parecchie altre caratteristiche di cui è dotato Fedora Repository, con la disponibilità di funzioni richiamabili da script per la gestione dell'intero repository, che includono le operazioni di manutenzione e quelle di inserimento e rimozione di oggetti, semplificando così la fase di ingestione dei contenuti e rendendo l'operazione effettuabile in automatico.

### 2.3.2 Front-end: Drupal

Come interfaccia web verso gli utenti per presentare gli oggetti digitali, nel nostro caso brani audio, video, interviste con relative trascrizioni, testi di canzoni, libri e riviste, è stato scelto Drupal, un CMS (Content Management System) open-source, largamente diffuso e già utilizzato in rete per l'accesso a repository.

La scelta è stata dettata fondamentalmente dalla disponibilità del framework Islandora, che tratteremo nel seguito, che prevede proprio Drupal come base di sviluppo ed integrazione tra il repository ed il front-end verso gli utenti.

Su questo prodotto non c'è molto da dire, ovvero è largamente conosciuto e diffuso come piattaforma di presentazione web e non sarà oggetto del nostro lavoro il suo approfondimento, vogliamo solamente evidenziare come, rispetto ad altre soluzioni, sia più versatile ed aperto, soprattutto per chi voglia integrarlo e personalizzarlo con moduli aggiuntivi come fatto da Islandora.

Dal punto di vista della sicurezza Drupal prevede moduli per l'autenticazione ed autorizzazione in un'ottica di possibili integrazioni dell'applicazione nel caso siano richieste policy diverse di accesso ai vari contenuti. Queste funzionalità hanno le potenzialità di integrarsi con le policy del repository, che permettono un controllo granulare dell'accesso ai singoli datastream, così come con i modelli ed il framework Islandora, lasciando aperta anche questa possibilità per gli sviluppi futuri.

### 2.3.3 Framework: Islandora

Fedora Repository è un'ottima base per sviluppare applicazioni inerenti la conservazione e presentazione di oggetti digitali, infatti sono molte le applicazioni disponibili nella comunità che si poggiano su tale sistema, tra tutte ci è parsa più conveniente e vicina alle nostre esigenze Islandora, un framework open-source sviluppato dalla biblioteca Robertson della UPEI (University of Prince Edward Island, Canada).

Il framework, sviluppato dalla collaborazione di informatici e bibliotecari, nell'ottica di utilizzare al meglio le risorse disponibili nella comunità, costituisce un sistema completo perfettamente integrato di congiunzione e coordinamento tra il repository e il CMS Drupal, rendendo quest'ultimo l'interfaccia tramite la quale amministrare e presentare i contenuti del repository.

Per FABB non abbiamo utilizzato tutte le potenzialità che offre Islandora, ci siamo limitati a utilizzare solo alcune delle funzioni. In particolare quelle di presentazione, che per la loro impostazione "open", ci hanno permesso modifiche e integrazioni con una certa agilità, ottenendo alcune variazioni per noi importanti, come ad esempio l'utilizzo dei soli metadati Dublin Core, e quelle di batch ingesting molto utili per inserimento nel repository di numerosi oggetti come nel caso di libri o di oggetti voluminosi (in termini di byte) come nel caso dei video e delle interviste.

Senza voler anche in questo caso scendere in troppi particolari, il framework è costituito principalmente da due componenti: dei moduli lato Drupal che mettono in comunicazione il front-end ed il repository e dei modelli lato repository costituiti da oggetti e relativi datastream in grado di caratterizzare gli oggetti inseriti secondo uno schema specifico che rende semplice la presentazione al pubblico. Avere modelli direttamente nel repository presenta un'ulteriore particolarità, quella che nessuna informazione relativa agli oggetti è memorizzata nel database di Drupal o in nessun'altra parte del front-end, ma tutto risiede sotto forma di datastream ed oggetti all'interno del repository stesso.


La vivace comunità di Islandora ha prodotto, tra i vari moduli e componenti, oltre a quelli direttamente dedicati alla comunicazione con il repository, alcuni elementi particolarmente interessanti per il nostro progetto: il viewer dei libri, la visualizzazione di interviste con trascrizione ed il sistema di ricerca. Mentre dell'ultimo specificheremo meglio gli aspetti nel paragrafo seguente, il viewer dei libri è un componente derivato dall'iniziativa open-source Internet Archive BookReader che permette la lettura on-line senza dover far ricorso a soluzioni proprietarie esterne. Il viewer produce la visualizzazione delle pagine di un libro abbinate al relativo testo prodotto dall'OCR, potendo eventualmente effettuare lo zoom delle immagini e navigare tra le pagine del volume, con funzioni di ricerca del testo integrata. La visualizzazione di video di interviste con l'abbinamento del testo trascritto è frutto invece di un membro della comunità di Islandora, Ling Ling del Sheridan College di Toronto, il cui lavoro è stato fondamentale e utilizzato nel caso di quest'ultimo progetto. Il modulo visualizza in sincronia il testo con il video e permette la navigazione all'interno del video selezionando il brano desiderato.

#### **2.3.4 Piattaforma di ricerca: Solr**

Merita un cenno specifico la piattaforma di indicizzazione e ricerca, che costituisce per questo tipo di progetti sicuramente uno dei componenti più apprezzati ed utili per l'utente finale. La scelta è ricaduta su Solr, parte del progetto Apache Lucene, soprattutto per l'integrazione disponibile nei componenti prescelti come repository e front-end.

In Fedora Repository è disponibile un applicativo aggiuntivo “Fedora Generic Search Service”, integrato come servizio web all'interno di Tomcat, che offre preziose funzioni di indicizzazione e ricerca sfruttando come piattaforma proprio Solr ed integrandosi con il repository fino a rendere automatici gli aggiornamenti degli indici a seguito dell'ingesting di nuovi oggetti. Mentre tra i vari moduli sviluppati dalla comunità di Islandora, ne esiste uno specifico per integrare nel sito web di Drupal la ricerca tramite Solr all'interno del repository con tutte le principali funzioni richieste.

E' stato necessario personalizzare lo schema e la configurazione di Solr per adattarla alla nostra situazione, in particolare per definire quali datastream indicizzare e di quali oggetti, oltre al tipo di indicizzazione. Comunque è stato uno dei pochi interventi richiesti per ottenere l'indicizzazione full-text dei datastream relativi ai testi e quella a parole dei metadati Dublin Core. Il risultato ottenuto permette dal sito web la ricerca per parole chiave nei dati descrittivi con possibilità di filtri suggeriti dallo stesso sistema di indicizzazione (*facet*) ed eventualmente abbinata o in alternativa la ricerca full-text.



## Fame & Abbondanza Repository

Fame e abbondanza nelle forme brevi della narrativa popolare piemontese

[Home](#)   [Browse repository](#)   [Policies](#)

Home » polenta (31)

Field FullText

Search terms

polenta

+

**Search**

---


804 Items in 16 Collections

### Search results

(1 - 20 of 31)

1   2   next »   last »

---



**Title** Fame e Abbondanza in cinquantacinque poesie di Giovanni Rapetti - pag.41


**Author** Castelli, Franco, Emina, Antonella, Milanese, Piero

**Text** **Polenta** con crauti Fumo di cotichino e cavoli viene l'acquolina in bocca **polenta** sopra l'asse sotto a chi tocca cento cavoli nell'orto, l'inverno per una famiglia quando nevicava, ammazzato il maiale, la fame è riempita. Facevano il soffritto di cipolla tritata uno spicchio d'aglio schiacciato, conserva regolata tagliata una verza dura, bianca, lavata, giù nel padellino rosolato, dieci minuti

**Date** 2016

**Type** text

---



**Title** Fame e Abbondanza in cinquantacinque poesie di Giovanni Rapetti - pag.81


**Author** Castelli, Franco, Emina, Antonella, Milanese, Piero

**Text** Cose che hanno ottant'anni, te le ricordi ti sembrano fin belle quando le dici una volta è arrivato un uomo alla cascina a questuare, che ha fame, dagli un pezzo di pane della cesta. Quel pane è duro, ogni dieci giorni cuocivano dagli della **polenta** allora, qualcosa riempiva pane asciutto, la **polenta** dice che non gli piace capito che è un balordo, vuole venire dentro. Era seduto di fianco all'uscio

**Date** 2016

**Type** text

---



**Title** 900 - pag.5

**Author** Conte, Paolo

**Text** fermo sul legno che scricchiola per la sua vetustà... sì, venditori, vecchi incantatori, suggestionatori... la strada fatta là fuori è stata percorsa già... da-dam-da-dam... «Pesce veloce del Baltico» dice il menu, che contorno ha? «Torta di mais» e poi servono **Polenta** e baccalà..., cucina povera e umile fatta d'ingenuità, caduta nel gorgo

**Title**

- abbondanza (22) +-
- cinquantacinque (22) +-
- fame (22) +-
- giovanni (22) +-
- poesie (22) +-
- rapetti (22) +-
- estratto (5) +-
- novecento (5) +-
- piemontese (5) +-
- poeti (5) +-

Show more

**Author**

- antonella (22) +-
- castelli (22) +-
- emina (22) +-
- franco (22) +-
- milanese (22) +-
- piero (22) +-
- albina (5) +-
- giovanni (5) +-
- malarba (5) +-
- tesio (5) +-

Show more

**Words**

- polenta (31) +-
- mai (13) +-
- prima (13) +-
- volta (13) +-
- pane (12) +-
- paese (10) +-
- poeta (10) +-
- casa (9) +-
- cibo (9) +-
- dentro (9) +-

Show more

**Subject**

- fame (28) +-
- abbondanza (22) +-
- alimentazione (22) +-
- cibi (22) +-
- commestibili (22) +-
- erbe (22) +-
- giovanni (22) +-

Figura 3: Ricerca full-text e filtri



### 3 Sito web di presentazione

Oltre al front-end di Drupal per l'accesso al repository è stato realizzato un sito web responsive in Joomla all'indirizzo [fabb.irgres.cnr.it](http://fabb.irgres.cnr.it) utilizzando un template free di joomla freak. Scopo del sito è la presentazione del progetto, l'indicazione dei contatti, l'indirizzamento verso il repository e l'offerta di un glossario interattivo in grado di orientare l'utente nella ricerca delle voci all'interno del repository, in particolare il glossario risulta essere uno strumento fondamentale per la piena comprensione della terminologia in dialetto.

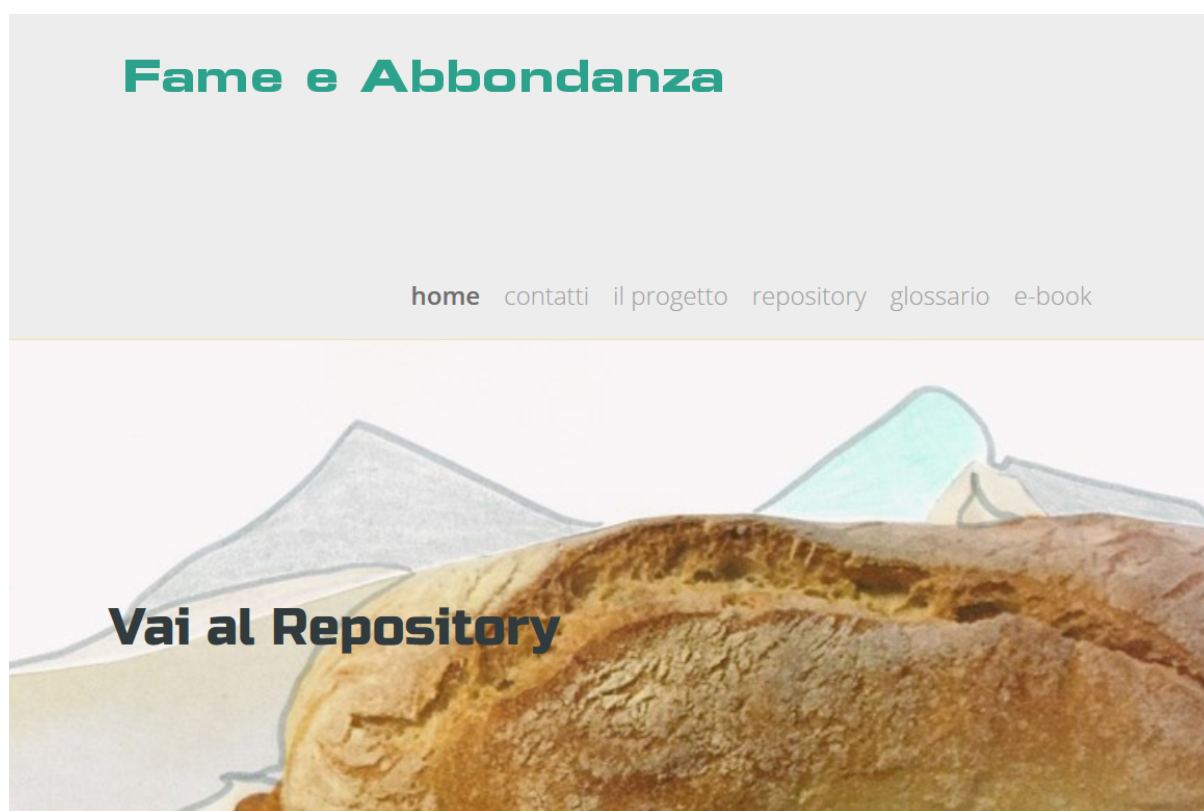


Figura 4: Sito web di presentazione

Dal sito web è inoltre accessibile l'e-book realizzato per il progetto dal titolo "*Fame e Abbondanza in cinquantacinque poesie di Giovanni Rapetti*". La peculiarità del volume è la sua piena fruizione per la lettura online anche dal sito web, realizzata con l'implementazione del modulo Joomla *FlippingBook*. Il visualizzatore consente lo sfogliamento del libro, l'ingrandimento delle singole pagine, la stampa delle pagine a video e il download del pdf.



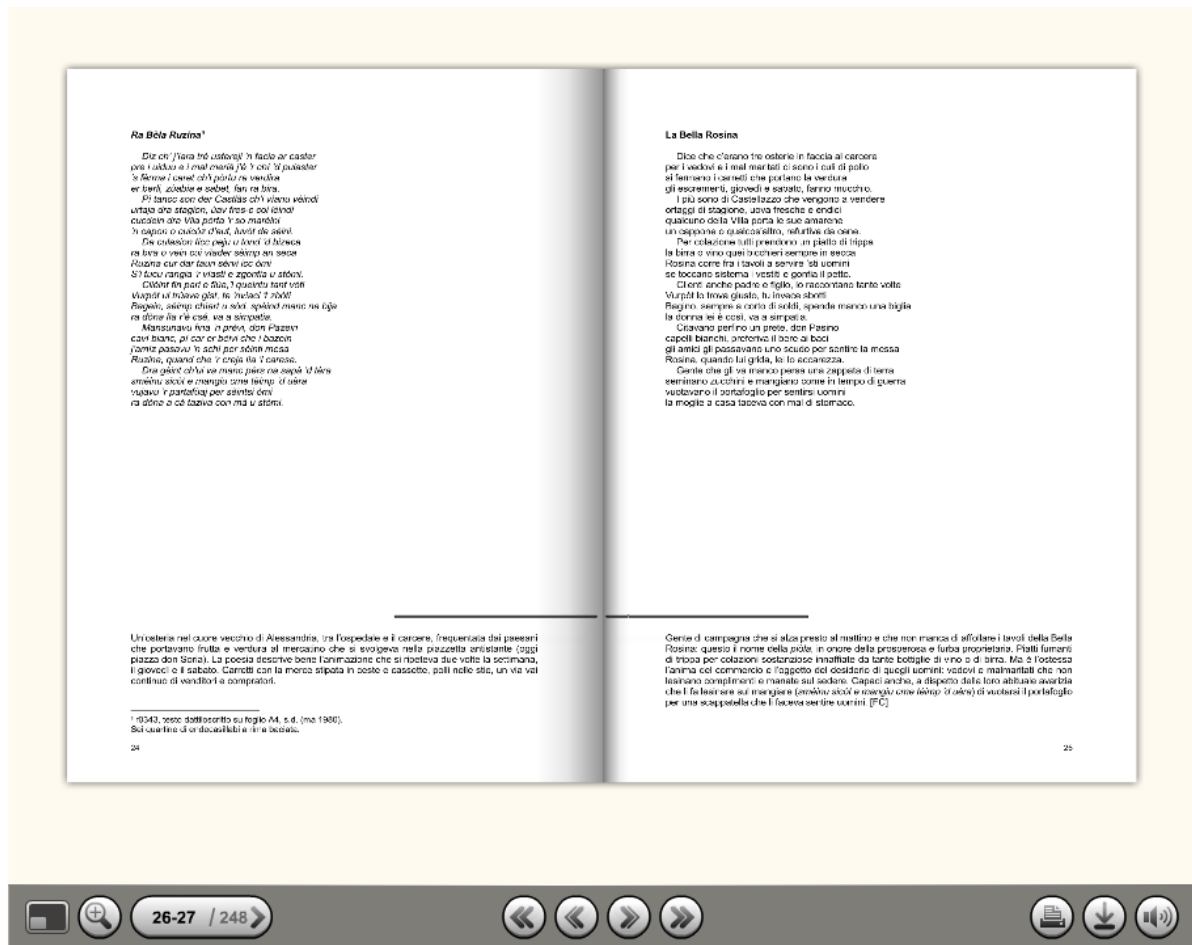


Figura 5: Visualizzatore libro in Joomla

## 4 Conclusioni

Il lavoro svolto è stato molto appassionante e costruttivo, nonostante ci fossimo già approcciati alla costruzione di repository è stato interessante affrontare nuove sfide, utilizzando le versioni più aggiornate dei programmi e trovando soluzioni “custom” per le specifiche tipologie di materiale da caricare.

Le potenzialità di un repository tematico sono molte, se partiamo da 3 parole chiave: conservare, valorizzare e rendere accessibile il materiale, questo già basterebbe per giustificare il lavoro svolto, ma il valore aggiunto in FABB è l’attenta scelta degli oggetti (testi, video, interviste) che, se pur molto diversi tra loro, tracciano un filo conduttore che guarda al passato e che riconduce all’identità culturale di un territorio.

Riassumendo, il progetto ha raggiunto un buon livello di sviluppo diventando di fatto un prototipo operativo, soprattutto è stata definita in modo chiaro e preciso l’architettura, nostro obiettivo è anche rendere disponibile l’esperienza fin qui sviluppata per eventualmente collaborare con chi interessato nello sviluppo di applicazioni analoghe e al riguardo si rimanda al sito di sviluppo accessibile all’indirizzo: <http://fabbdev.to.cnr.it/doku.php>.

Il repository FABB è stato presentato, a seguito della revisione e accettazione da parte di referee internazionali, al convegno mondiale annuale sugli archivi aperti. “OR 2016. 11th Annual Conference on Open Repositories” di Dublino 12-16/06/2016.



## 5 Sitografia

1. Apache Solr, <http://lucene.apache.org/solr/> visitato luglio 2016.
2. Apache Tomcat, <http://tomcat.apache.org/> visitato luglio 2016.
3. CNR Ismar, BSA digital Library, <http://biblio1.ve.ismar.cnr.it/> visitato luglio 2016.
4. DigiBESS repository, [www.digibess.it](http://www.digibess.it) visitato luglio 2016.
5. Docsplit <http://documentcloud.github.com/docsplit/> visitato luglio 2016.
6. Drupal Come for the software, stay for the community, <http://drupal.org/> visitato luglio 2016.
7. Dublin Core Metadata Initiative, <http://dublincore.org/> visitato luglio 2016.
8. FABB Repository, <http://fabb.to.cnr.it> visitato luglio 2016
9. FABB Sito presentazione, <http://fabb.ircres.cnr.it> visitato luglio 2016
10. FABB Sito sviluppo, <http://fabbdev.to.cnr.it/doku.php> visitato luglio 2016
11. Fedora Repository software repository, <http://fedora-commons.org/> visitato luglio 2016.
12. Joomla, <https://www.joomla.org/> visitato luglio 2016
13. KVM VM backup, <http://repo.firewall-services.com/misc/virt/virt-backup.pl> visitato luglio 2016.
14. Infrastruttura di rete del CNR in Piemonte, <http://www.to.cnr.it> visitato luglio 2016.
15. Islandora, building a rich digital repository ecosystem, <http://islandora.ca/> visitato luglio 2016.
16. Islandora developers forum, <https://github.com/islandora> visitato luglio 2016.
17. OR2016. 11th Annual Conference on Open Repositories, <http://or2016.net/> visitato luglio 2016.

## 6 Bibliografia

- Birello G., Perin A., Emina A., Zoppi I.M. (2016) FABB Repository. Restoration and enhancement of popular culture objects. Relazione presentata al OR 2016. 11<sup>th</sup> Annual Conference on Open Repositories, 12-16/06/2016, Dublin (IE)
- Birello G., Perin A., Fucile I., Giovanetti V. (2016) Fedora repository e Islandora: un framework innovativo open-source per la conservazione e presentazione di oggetti digitali, poster presentato al workshop Garr 2016 "Terabit generation, una comunità ad alte prestazioni", 18-21 Aprile 2016, Roma
- Birello G., Abbà S., Perin A. (2015) V2P2 and DigiBESS: case studies of open IT infrastructures for Islandora devel and production. Relation presented at Islandora Conference 2015, 3-7 August 2015, Charlottetown, Pei, Canada
- S. Abbà, G. Birello, M. Vallino, A. Perin, S. Ghignone and P. Caciagli (2015) [Shall we share? A repository for Open Research Data in agriculture and environmental sciences](#), Bulletin OEPP/EPPO Bulletin (2015) 45 (2), 311–316
- Birello Giancarlo, Giovanetti Valter, Fucile Ivano, Perin Anna (2014) [DigiBESS: an open-source repository for sharing digital preservation](#). Poster presented at “The reuse of digital cultural content in education, tourism and leisure” October 2, 2014, Biblioteca Nazionale Centrale, Rome Italy
- Birello G., Perin A. (2014) [Una biblioteca online accessibile da tutti. La digitalizzazione dell'archivio Bess: il caso Ceris-CNR in L'Italia che innova. Case History e protagonisti](#). Premio innovazione ICT Piemonte 2014
- Birello Giancarlo, Giovanetti Valter, Fucile Ivano, Perin Anna (2013) [Step by step installation guide of a digital preservation infrastructure part 2](#) Rapporto Tecnico Ceris-CNR N.45
- Birello Giancarlo, Giovanetti Valter, Fucile Ivano, Perin Anna (2013) [KVM: an open-source framework for virtualization](#) Rapporto Tecnico Ceris-CNR N.44
- Birello Giancarlo, Giovanetti Valter, Fucile Ivano, Perin Anna (2013) [OAI-PMH harvesting metadata and virtual datastream](#). Poster presented at Open Repository 2013 - Charlottetown, Prince Edward Island (Canada) July 8 - 12
- Birello Giancarlo, Giovanetti Valter, Fucile Ivano, Perin Anna (2013) [Digital preservation: an open-source architecture](#). Poster presented at 5th International Conference on Qualitative and Quantitative Methods in Libraries, 4 - 7 June 2013, "La Sapienza" University, Rome Italy
- Bertolla Giorgio, Birello Giancarlo, Perin Anna (2012) *Digibess, una biblioteca digitale open source*, Biblioteche Oggi, V.30 n.6 luglio-agosto 2012.
- Birello Giancarlo, Perin Anna (2012) *Conservazione delle opere digitali. Progetto Bess - Archivi Digitali*, Le tecnologie del CNR per la Pubblica Amministrazione, CNR, luglio 2012.
- Birello Giancarlo, Fucile Ivano, Giovanetti Valter, Perin Anna (2012) *Repository and Front End open-source for digital items*, poster presentato al convegno "[OR2012, the 7th International Conference on Open Repository](#)", Edimburgh GB 9-13 July 2012 ([PDF](#))
- Birello Giancarlo, Fucile Ivano, Giovanetti Valter, Perin Anna, Minuzzo Tiziano (2012)

Rapporto Tecnico Ceris-CNR N.42

[Step by step installation guide of a digital preservation infrastructure](#)

Birello Giancarlo, Fucile Ivano, Giovanetti Valter, Perin Anna (2012) Rapporto Tecnico Ceris-CNR N.41

[Storage in HA: manutenzione ordinaria e straordinaria](#)

Birello Giancarlo, Fucile Ivano, Giovanetti Valter, Perin Anna (2011) Rapporto Tecnico Ceris-CNR N.39

[Repository e Front-end open-source per la conservazione di opere digitali](#)

Birello Giancarlo, Fucile Ivano, Giovanetti Valter, Perin Anna (2011) Rapporto Tecnico Ceris-CNR N.37

[Storage in HA: cluster attivo/passivo open source.](#)