

Digital preservation: an open-source architecture



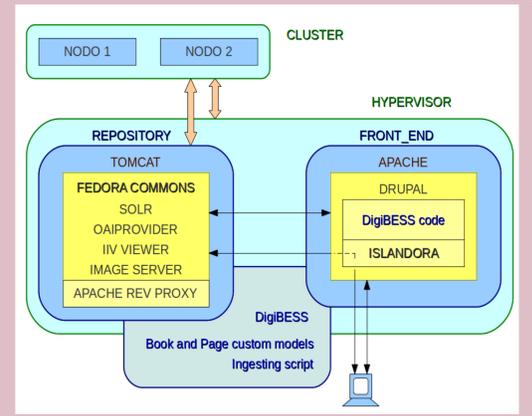
G.Birello, I.Fucile, V.Giovanetti, A.Perin

5th International Conference on Qualitative and Quantitative Methods in Libraries 4 - 7 June 2013, "La Sapienza" University, Rome Italy

Project

Ceris-CNR IT Office and Ceris-CNR Library are involved in a **digitization project** promoted by Bess (Electronic Library of Economic and Social Sciences in Piedmont Area, Italy) and commissioned to handle all the post-scan of the digitization, **DIGIBESS** repository (economic and social science books of Piedmont Area, Italy). The sponsorship of the project is Compagnia di San Paolo of Turin (Italy). This Poster analyzes the **strategies** adopted and the main **open-source** software used. Ceris-CNR had to provide for the **management of large volumes of data** with the availability of space storage for the digitized works with characteristics of stability, versatility and dynamism. Ceris-CNR has deployed the **software and server platforms** of the repository, in a virtualized and redundant infrastructure and also take care of the design, development and management of the web portal (**front-end**) for the presentation, search and consulting data of digitalized items

Architecture of the Repository



Numbers, Hardware and Software

- ✓ 395.662 **digitalized pages**, **2.216 books** in two years, still in progress
- ✓ **files**: high resolution tiff, OCR txt file, pdf /a, metadata
- ✓ **14 TB** disk base space at disposal
- ✓ **2-node active/passive** open-source cluster
- ✓ High Availability **Hypervisor** using cluster storage
- ✓ **Fedora Commons** repository
- ✓ harvesting **OAI-PMH**
- ✓ **scripting** for ingesting
- ✓ **Islandora** model and modules
- ✓ **Drupal** front-end
- ✓ **Solr** - search platform from the Apache Lucene project
- ✓ **Full-text** search

Front-end

The repository allows you to search in the fields **Title, Author, Full-text, Keywords** entering one or more words. It is possible to refine your search with **facet**.

The book selected can be **downloaded** in pdf or high resolution TIFF. You can also **read online** the book and turn pages on the screen.

The pages can be resized and an additional function enables the text of the page selected in txt.

DIGIBESS exposes **OAI-PMH** interface for external metadata harvesting, is indexed in international portals OpenDOAR, BASE, IESR, ROAR. The books are indexed in **WorldCat Catalogue, Pleiadi, Cultura Italia** and soon in **Europeana**

References

Repository site: <http://www.digibess.it> **Developer site:** <http://dev.digibess.it>
Digibess, una biblioteca digitale open source, Biblioteche Oggi, V.30 n.6 luglio-agosto 2012.

Technical Reports on Ceris-CNR Website (www.ceris.cnr.it):

- N.42** Step by step installation guide of a digital preservation infrastructure
- N.41** Storage in HA: manutenzione ordinaria e straordinaria
- N.39** Repository e Front-end open-source per la conservazione di opere digitali
- N.37** Storage in HA: cluster attivo/passivo open source

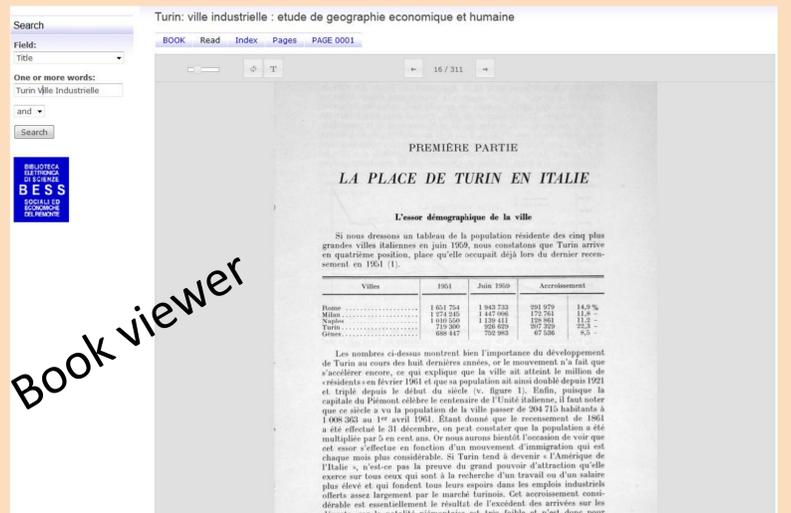
Our thanks to open-source community



Dublin Core



Book viewer



OAI-PMH

