ISSN online: 2282-5665

Consiglio Nazionale delle Ricerche

CERIS ISTITUTO DI RICERCA SULL'IMPRESA E LO SVILUPPO

Ottobre

2013

Rapporto tecnico N.45

STEP BY STEP INSTALLATION GUIDE OF A DIGITAL PRESERVATION INFRASTRUCTURE PART 2

Giancarlo Birello, Ivano Fucile,
Valter Giovanetti, Anna Perin

Consiglio Nazionale delle Ricerche

Istituto di Ricerche sull'Impresa e Lo Sviluppo

RAPPORTO TECNICO CNR-CERIS
Anno 8, N° 45; Ottobre 2013

Direttore Responsabile
Secondo Rolfo

Direzione e Redazione
CNR-Ceris
Istituto di Ricerca sull'Impresa e lo Sviluppo
Via Real Collegio, 30
10024 Moncalieri (Torino), Italy
Tel. +39 011 6824.911
Fax +39 011 6824.966
segreteria@ceris.cnr.it
http://www.ceris.cnr.it

Sede di Roma
Via dei Taurini, 19
00185 Roma, Italy
Tel. 06 49937810
Fax 06 49937884

Sede di Milano
Via Bassini, 15
20121 Milano, Italy
tel. 02 23699501
Fax 02 23699530

Segreteria di redazione
Enrico Viarisio
e.viarisio@ceris.cnr.it

STEP BY STEP INSTALLATION GUIDE OF A DIGITAL PRESERVATION INFRASTRUCTURE PART 2

Giancarlo Birello, Ivano Fucile Valter Giovanetti*
(CNR-Ceris, IT Office)

Anna Perin
(CNR-Ceris, Library)

CNR-Ceris
IT Office
Strada delle Cacce, 73
10135 Torino – Italy
Phone: +39 011 3977533/4/5

CNR-Ceris
Library
Via Real Collegio, 30
10024 Moncalieri (Torino) – Italy
Phone: +39 011 6824928

* Corresponding author: g.birello@ceris.cnr.it

ABSTRACT: The Ceris-CNR project of digital preservation infrastructure has been committed by Bess (Social Science Electronic Library of Piemonte) for years 2011-2012 and confirmed for year 2013 sponsored by Compagnia di San Paolo of Turin.
Ceris-CNR role is to handle all the post-scan of the digitalization, for this purpose it has deployed the software and server platforms of the repository and also the web portal for the presentation, research and consulting.
This report is the second part of step by step guide to build the digital archive infrastructure.

KEY WORDS: open-source, islandora, repository, digital archive, cms

Table of Contents

1 Introduction

The CNR Ceris project of digital preservation infrastructure has been committed by Bess (Social Science Electronic Library of Piemonte) for years 2011-2012 and confirmed for year 2013.

Bess is a group of eighteen socioeconomic libraries in Piemonte (Italy) included Ceris library, they share a common specialization even if they are private foundations, research institutes, and university libraries that means different libraries in terms of size, parent institution, purpose, financial endowment, as well as collections.

One of the initiative promoted by Bess and sponsored by Compagnia di San Paolo of Turin, is the creation of a digital repository of sources of Piedmontese society and economy.

Bess has set up a digitalization laboratory, to be directly used by the members, for the conservation and preservation of collections included out of print and gray literature materials;

External partners are welcome to digitalize and share their archives, Istituto Gramsci for example has licensed “Sisifo” review as well as Centro Storico FIAT with “Illustrato Fiat” and “Blu Lancia” and Archivio Storico Lavazza with “Notizie Lavazza” that are now accessible through Bess archive. Other institutions have agree to share some collections yet.

The resulting repository is serving as a source of regional and economic information to the whole community.

CNR Ceris role is to handle all the post-scan of the digitalization and 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. Currently (September, 2013) 416.751 pages are available in the repository.

CNR Ceris 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, research and consulting data of the digitalized items. Moreover the repository is OAI-PMH compliant and is well harvested by the main National and International meta-repository .

Here listed some evidence of numbers, hardware and software used, most of the aspects will be analyzed in this paper or are available in the *"Step by step Installation Guide of a Digital Preservation Infrastructure."*, Rapporto Tecnico N.42, maggio 2012, CNR Ceris:

- files: pdf/a, high resolution tiff, txt file
- Metadata
- 2-node active/passive open-source cluster
- Kernel-based Virtual Machine (KVM) hypervisor
- repository: Fedora Commons
- harvesting OAI-PMH
- scripting for ingesting
- Custom models and datastreams
- front-end server: Drupal and Islandora
- Solr - search platform from the Apache Lucene project

The open-source community helped us to build on our project and this paper is at disposal of the open-source community, in particular we have to thank you Islandora Team with which we have exchanged many code pages.

2 Front-end server

The following customizations regard code modifications of Islandora 6.x modules in drupal front-end CMS as developed in the first part of this guide.

2.1 Collapsed breadcrumb

To limit title length in breadcrumb to avoid display overflow.

Edit Main breadcrumb (not Search breadcrumb) builder code:

```
nano -w  sites/all/modules/islandora-github/ObjectHelper.inc                    [6.1.1]

...
function getBreadcrumbs($pid, &$breadcrumbs, $level=10) {
    module_load_include('inc', 'fedora_repository', 'api/fedora_utils');
    // Before executing the query, we hve a base case of accessing the top-level collection
    global $base_url;
    if ($pid == variable_get('fedora_repository_pid', 'islandora:root')) {
      $breadcrumbs[] = l(t('Digital repository'), 'fedora/repository');
      $breadcrumbs[] = l(t('Home'), $base_url);
    }
    else {
      $query_string = 'select $parentObject $title $content from <#ri>
                       where (<info:fedora/' . $pid . '> <dc:title> $title
                       and $parentObject <fedora-model:hasModel> $content
                       and (<info:fedora/' . $pid . '> <fedora-rels-ext:isMemberOfCollection> $parentObject
                       or <info:fedora/' . $pid . '> <fedora-rels-ext:isMemberOf> $parentObject
                       or <info:fedora/' . $pid . '> <fedora-rels-ext:isPartOf> $parentObject)
                       and $parentObject <fedora-model:state> <info:fedora/fedora-system:def/model#Active>)
                       minus $content <mulgara:is> <info:fedora/fedora-system:FedoraObject-3.0>
                       order by $parentObject';
      $query_string = htmlentities(urlencode($query_string));
      $url = variable_get('fedora_repository_url', 'http://localhost:8080/fedora/risearch');
      $url .= "?type=tuples&flush=TRUE&format=CSV&limit=1&offset=0&lang=itql&stream=on&query=" . $query_string;

      $result = preg_split('/[\r\n]+/', do_curl($url));
      array_shift($result); // throw away first line
      $matches = str_getcsv(join("\n", $result));
      if ($matches !== FALSE) {
        $parent = preg_replace('/^info:fedora\//', '', $matches[0]);
        if ((strlen($parent) >0) && ($level <10)) {
          if (substr($matches[1],-11,6) != '- page') {
            if (strlen($matches[1]) > 25){
              $breadcrumbs[] = l(htmlentities(substr($matches[1],0,25)) . '...', 'fedora/repository/' . $pid);
            }
            else {
              $breadcrumbs[] = l($matches[1], 'fedora/repository/' . $pid);
            }
          }
        }
        if ($parent == variable_get('fedora_repository_pid', 'islandora:root')) {
          $breadcrumbs[] = l(t('Digital repository'), 'fedora/repository');
          $breadcrumbs[] = l(t('Home'), $base_url);
        }
        elseif ($level > 0) {
          $this->getBreadcrumbs($parent, $breadcrumbs, $level - 1);
        }
      }
    }
  }
...
```

The following picture shows code result:

Picture 1: Collapsed breadcrumb

2.2 Custom Theme

Create a custom theme for drupal containing style definitions for standard and custom html tags.

- SSH login on drupal server
- Clone Garland theme

```
cd /usr/share/rootsitedir
mkdir sites/all/themes/bessparent
cp -R themes/garland/* sites/all/themes/bessparent/
rm -R sites/all/themes/bessparent/minnelli
```

- Configure parent theme

```
mv sites/all/themes/bessparent/garland.info
  sites/all/themes/bessparent/bessparent.info
nano -w sites/all/themes/bessparent/bessparent.info
name = DigiBessParent
description = Garland clone.
version = VERSION
core = 6.x
engine = phptemplate
stylesheets[all][] = style.css
stylesheets[print][] = print.css

; Information added by drupal.org packaging script on 2012-02-29
version = "6.25"
project = "drupal"
datestamp = "1330534547"
```

- Create child theme

```
mkdir sites/all/themes/digibess
nano -w sites/all/themes/digibess/digibess.info

name = DigiBess
description = DigiBess theme.
version = VERSION
core = 6.x
base theme = bessparent
stylesheets[all][] = digibess.css

; Information added by drupal.org packaging script on 2012-02-29
;version = "6.25"
;project = "drupal"
;datestamp = "1330534547"
```

- Create child template to override functions

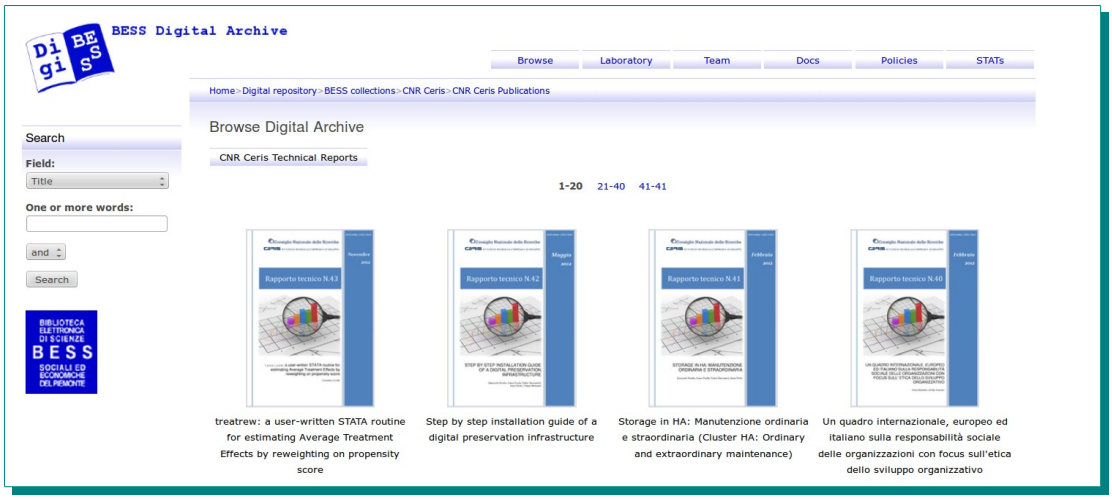
```
nano -w sites/all/themes/digibess/template.php

<?php
function digibess_breadcrumb($breadcrumb) {
  if (!empty($breadcrumb)) {
    return '<div class="breadcrumb">'. implode('>', $breadcrumb) .'</div>';
  }
}
```

- Create child CSS to override formatting

```
nano -w  sites/all/themes/digibess/digibess.css                    [6.1.2]
```

- Copy modified *bg-content.png* in *sites/all/themes/digibess/images*
- Copy new *bg-tabs.png* in *sites/all/themes/digibess/images*
- Enable theme browsing to *siteURL/admin/build/themes*



Picture 2: DigiBESS theme example

2.3 Custom Collection tabs

Collection view tabs are defined in *CollectionClass.inc* file.

We modified this file to show custom title and a new tab for collection information page contained in INFO datastream.

```
nano -w  sites/all/modules/islandora-github/CollectionClass.inc                [6.1.3]

...
function showFieldSets($page_number) {
    module_load_include('inc', 'fedora_repository', 'api/fedora_item');
    global $base_url;
    $tabset = array();
    global $user;
    $objectHelper = new ObjectHelper();
    $item = new Fedora_Item($this->pid);
    $query = NULL;
    if ($item->exists() && array_key_exists('QUERY', $item->datastreams)) {
      $query = $item->get_datastream_dissemination('QUERY');
    }
    $results = $this->getRelatedItems($this->pid, $query);
    $colleinfo = NULL;
    if ($item->exists() && array_key_exists('INFO', $item->datastreams)) {
      $colleinfo = $item->get_datastream_dissemination('INFO');
    }
    $collection_items = $this->renderCollection($results, $this->pid, NULL, NULL, $page_number);
    $collection_item = new Fedora_Item($this->pid);
    // Check the form post to see if we are in the middle of an ingest operation.
    $show_ingest_tab = (!empty($_POST['form_id']) && $_POST['form_id'] == 'fedora_repository_ingest_form');
    $add_to_collection = $this->getIngestInterface();
    $view_selected =true;
    $coll_selected =false;
    $qstring = $_GET['q'];
    $qparts = explode('/', $qstring);
    $tail = end($qparts);
    if ($tail == 'info') {
      $view_selected = false;
      $coll_selected = true;
    }
    drupal_set_message();
    $tabset['1'] = array(
      '#type' => 'tabpage',
      '#title' => $collection_item->objectProfile->objLabel,
      '#selected' => $view_selected,
      '#content' => $collection_items,
      '#weight' => '1'
    );
    if ($colleinfo != NULL) {
      $tabset['2'] = array(
        '#type' => 'tabpage',
        '#title' => 'Collection info',
        '#selected' => $coll_selected,
        '#content' => $colleinfo,
        '#weight' => '2'
      );
    }
    return $tabset;
  }
...
```

At repository level you have to add a datastream INFO to Collection object in text/html format with information about collection. Available header tags are h2 (without background) and h4 (with theme background), for example:

```
<html>
<head>
<meta content="text/html; charset=ISO-8859-1"
http-equiv="content-type">
<title>Lettera da Tecnocity</title>
</head>
<body>
<h2>Lettera da Tecnocity (1984 ? 1992)</h2>
<span style="font-weight: bold;">Direttore: Marcello Pacini</span><br
style="font-weight: bold;">
<span style="font-weight: bold;">Un progetto della Fondazione Agnelli
per il rilancio dell?area metropolitana torinese</span><br>
<br>
<h4>LO SFONDO</h4>
<br>
Nel decennio tra gli anni ottanta e novanta la Fondazione Giovanni
Agnelli è impegnata in un ampio programma volto a studiare le
condizioni che rendono possibile e governano l?innovazione tecnologica
...
```