Preserving digital records and databases

Preservation and Archiving Special Interest Group 2015

Luis Faria lfaria@keep.pt
KEEP SOLUTIONS

March 13, 2015
Advanced services for Information Management and Digital Preservation

Research
30%

Open-source services
30%

Commercial products
40%
<table>
<thead>
<tr>
<th>Products and services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional repositories</td>
</tr>
<tr>
<td>Archival management</td>
</tr>
<tr>
<td>Library management</td>
</tr>
<tr>
<td>Museum collection management</td>
</tr>
<tr>
<td>Federated search portals</td>
</tr>
<tr>
<td>Digital preservation</td>
</tr>
</tbody>
</table>
E-ARK

Create and pilot an international methodology for archiving digital records and databases

Partners:

National archives: Denmark, Hungary and Estonia

Ministries: Eslovenia, Norway, Portugal and Spain

Universities: Portsmouth, Cologne and Lisbon

Companies: KEEP SOLUTIONS, Magenta, ES Solutions
RODA
REPOSITORY OF AUTHENTIC DIGITAL OBJECTS

www.roda-community.org

Open source
Vendor freedom, no licensing costs

Ready to use
By archivists, producers and consumers

Authenticity
Preservation metadata, ISO 16363 compliant
Search in the repository:

demo

<p>| 75 | DEMO_IMAGE_DEMO_VIDEO_MP4_01 | 2013-01-01 | 2014-01-01 |
| 68 | DEMO_TEXT_TESTE_IMAGEM_JPEG  | 2013-01-01 | 2014-01-01 |
| DEMO | Demonstration fonds         | 2013-12-31 | 2013-12-31 |
| 36 | Big Buck Bunny              | 2013-01-01 | 2013-01-01 |
| 6  | Nature photos               | 2013-07-02 | 2013-07-02 |
| asd | balikk                       | 0001-01-01 | 0002-01-01 |
| TEXT | Text samples...            | 2013-01-01 | 2013-12-31 |
| IMAGE | Image samples            | 2013-01-03 | 2013-12-31 |
| AUDIO | Audio samples              | 2013-01-01 | 2013-12-31 |
| VIDEO | Video samples              | 2013-01-01 | 2013-12-31 |
| 445 | Test                        | 2012-06-24 | 2012-06-24 |
| 64  | TESTE_DOC_PDF_01            | 2013-01-01 | 2013-09-01 |
| 420 | Teste_roda                  | 2010-01-01 | 2015-01-01 |
| 395 | moje                        | 0001-01-01 | 0002-01-01 |</p>
<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
<th>View</th>
<th>Preservation</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEDIDOS-2012</td>
<td>Identification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIARQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST/TEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>231</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUDIO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATABASES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMAGES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;indefinida&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIDEO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZV_01_01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEST_SZ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Identification**
- Reference: 20
- Complete reference: PT/DGARQ/TEST/TEST/IMAGES/20
- Handle: http://hdl.handle.net/10384/20
- Title: National Geographic
- Level: item
- Initial date: 2009-06-01
- Final date: 2009-06-07
- Country code: PT
- Repository code: DGARQ
- Origination: Ifaria
- Acquisition number: NG
- Acquisition date: 2009-06-04

**Content and structure**
- Scope and content: demonstration

**Notes**
- Note: Pacote de submissão validado semanticamente por admin
**Ingestion**

**Description:** The objects inside the SIP were added to the preservation repository.

**Result:** success

**Ingest details:**

- O SIP foi ingerido com sucesso como filho do objeto roda:21.
- Foram criados 1.
The Database Preservation Toolkit allows conversion between Database formats, including connection to live systems, for purposes of digitally preserving databases. The toolkit allows conversion of live or backed-up databases into preservation formats such as DBML or SIARD, XML-based formats created for the purpose of database preservation. The toolkit also allows conversion of the preservation formats back into live systems to allow the full functionality of databases. For example, it supports a specialized export into MySQL, optimized for PhpMyAdmin, so the database can be fully experimented using a web interface.

This toolkit was part of the RODA project and now has been released as a project by its own due to the increasing interest on this particular feature.

The toolkit is created as a platform that uses input and output modules. Each module supports read and/or write to a particular database format or live system. New modules can easily be added by implementation of a new interface and adding of new drivers.

www.database-preservation.com
Preservation format: SIARD

1.0
Developed by the Swiss Federal Archives
Swiss standard since 2013 (eCH-0165)
Based on standards: XML, SQL:1999, Unicode

2.0 / SIARD-E (under development)
Under development by E-ARK and Swiss Federal Archives
Support for multiple medium storage
Stricter data types
Support for arrays and user-defined types
db-preservation-toolkit

Import modules
- MySQL
- Oracle'12
- SQL Server
- PostgreSQL
- DB2
- MS Access
- ODBC
- DBML
- SIARD

Streaming data model

Export modules
- MySQL
- SQL Server
- PostgreSQL
- DB2
- MS Access
- ODBC
- DBML
- SIARD
- Viewer
In the Archive
In the Archive

MySQL
Oracle'12
SQL Server
PostgreSQL
DB2
MS Access
ODBC

DB Preservation Toolkit
SIARD
DB Preservation Toolkit

Viewer
Live DBMS
OLAP
Throughput

Row-intensive databases

10,000 rows/s

LOB-intensive databases

30 MB/s
Other preservation services

- ISO 16363 consultancy
- Format characterisation and risk analysis
- QA and validation of digitisation projects
- Large-scale format migration
