SAFE Archive FEderation

LOCKSS Webinar
Nov 14, 2017

Anthony Leroy
Université libre de Bruxelles, Belgium
international federation
geo-replication in completely independent sites

light organizational structure
7+ nodes

distributed technical administration
local admin only, no automation

each partner monitors the status of his content in the network
global verification that the preservation is performed correctly

budgets remain fully independent
economic risk mitigation
SAFE is based on a simple memorandum of understanding.

The agreement has been signed by the heads of libraries of the participating institutions.

The parties acknowledge that an equal amount of resources will be made available to the execution of the SAFE PLN project. All parties will make a LOCKSS cache node available for the SAFE PLN grid with a capacity of maximum 4 TB of storage.

All parties agree that the maximum raw storage capacity per institution in the SAFE PLN is 400GB.

The storage capacity requirements will be adapted every year by mutual agreement.
SAFE currently has 8 members
SAFE: WHO?
: WHAT?

Digitized books
Maps
Pictures

Theses and publications (IR)

Academic content
Research data
SAFE : WHAT ?

- Université libre de Bruxelles: 34 GB
- Université catholique de Louvain: 8 GB
- Bielefeld University: 37 GB
- Universiteit Gent: 261 GB
- Université de Genève: 25 GB
- Lund University: 26 GB

Note: Some institutions are still in the process of making their AUs available for production
SAFE : WHICH ?

2012-ETD-ugent: Archival Unit

manifest.html
metadata.xml

archive-ugent-be-[UUID] : BagIt

- bagit.txt
- bag-info.txt
- manifest-md5.txt
- marc.xml
- mets.xml
- data/
  - file AC.pdf
  - file LS.txt
  - file MA.tar
The content that is ready to be archived is frozen in a staging server called an **AU Publisher**.
Physical network monitoring:
Nagios provides global information about host and services availability

Logical network monitoring:
- Are our AUs safe? (all votes agree on data integrity)
- Where are our AUs preserved? (efficient geo-replication)

When a problem occurs, the tool sends an email alert to technical admins.
Logical network monitoring tool

plnmonitor daemon

plnmonitor wicket app

read only access (debug)
- java daemon daily collecting debug information from our boxes
- uses LOCKSS WebService API (available since v. 1.65.5)
- updates a local database with daily boxes status
- keeps history

for eachBox in allSAFEBoxes:
    getPlatformConfiguration()
    queryRepositorySpaces()
    queryAus();
    queryPeers();
    queryRepositories()

:8081/ws/DaemonStatusService?wsdl
Network monitoring tool

### SAFE PLN Monitor

<table>
<thead>
<tr>
<th>Active Boxes</th>
<th>Total AU</th>
<th>© Last Poll</th>
<th>MB Total Size</th>
<th>MB Min Free Capacity</th>
<th>© Last Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>16</td>
<td>196h ago</td>
<td>312 GB</td>
<td>205 GB</td>
<td>3h ago</td>
</tr>
</tbody>
</table>

#### Detailed Status - SAFE PLN Network

**SAFE Boxes Storage**

- **ULB1**: 3 TB
- **UCL**: 9 TB
- **Unibi**: 3 TB
- **UdeM**: 5 TB
- **UNIGE**: 2 TB
- **UGent**: 2 TB
- **MUN**: 4 TB
- **Lund**: 4 TB

**AU repartition**

- **Universiteit Gent**: 261 GB
- **Universite de Genove**: 25 GB
- **Lund University**: 26 GB

#### SAFE Boxes Info

<table>
<thead>
<tr>
<th>Box Id</th>
<th>IP address</th>
<th>UI Port</th>
<th>Country</th>
<th>Active AU</th>
<th>Tech Admin</th>
<th>Org Admin</th>
<th>Daemon Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>129.70.43.35</td>
<td>8081</td>
<td>Germany</td>
<td>1</td>
<td>Christian Pietsch</td>
<td>Friedrich Summann</td>
<td>1.72.3</td>
</tr>
<tr>
<td>2</td>
<td>130.104.5.34</td>
<td>8081</td>
<td>Belgium</td>
<td>9</td>
<td>Renaud Michotte</td>
<td>Benoît Erken</td>
<td>1.67.5</td>
</tr>
</tbody>
</table>
Archival Units status: global view

<table>
<thead>
<tr>
<th>AU Name</th>
<th>UNIGE</th>
<th>UdeM</th>
<th>ULB1</th>
<th>UCL</th>
<th>Unibi</th>
<th>UGent</th>
<th>MUN</th>
<th>Lund</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universite de Geneve Publications 2009</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2010</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2011</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2012</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2013</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2014</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2015</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universite de Geneve Publications 2016</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
<tr>
<td>Universiteit Gent Open Access Publications 2011</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
</tr>
<tr>
<td>Universiteit Gent Open Access Publications 2012</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
<td>No Info</td>
<td>No Info</td>
<td>Success</td>
<td>No Info</td>
<td>Success</td>
</tr>
</tbody>
</table>
Key LOCKSS aspects for SAFE

“poll/repair” is the key component for us

share resources, not money

distributed technical administration, no automation and centralized administration of boxes

the current ingest mechanism based on web crawling is not particularly suited for our academic digital assets collections
Needs for SAFE

- More in-depth documentation
- (documented) APIs

Needs for CLNs in general

- Configuration templates for various case studies
SAFE Archiving FEderation