

#### LOCKSS Distributed Digital Preservation Networks

Leroy, Anthony 2018 Dostupný z http://www.nusl.cz/ntk/nusl-387442

Dílo je chráněno podle autorského zákona č. 121/2000 Sb. Licence Creative Commons Uveďte původ-Zachovejte licenci 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL). Datum stažení: 27.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .



#### LOCKSS DISTRIBUTED DIGITAL PRESERVATION NETWORKS

#### Conference on Grey Literature and Repositories 2018

This presentation is licensed under the Creative Commons: <u>CC-BY-SA-4.0</u>, via <u>http://repozitar.techlib.cz/record/1296</u>

Anthony Leroy Université libre de Bruxelles



1



### University Libraries have two main **missions**

**To guarantee access to objects** selected by curators

To preserve those objects especially our own production

# For analog objects, guaranteeing access and preservation is relatively simple







### University Libraries have two main **missions**

To guarantee access to objects selected by curators

To preserve those objects especially our own production

In the digital era, those missions are compromised:

we lost control on some digital objects (access via subscription)
the vulnerability of digital objects

# What will happen to the current research outcome in 10 yrs? 100 yrs?



Attendees of the 5th Solvay Congrès, October 1927, Institut international de physique Solvay, Brussels

### Massive loss of research data in the last 25 years



80% of data in zoology linked to publications in the nineties are definitely lost

As a digital library, our core business is to develop the software infrastructure to describe, submit, disseminate and preserve digital objects





Strategies and processes to protect against the threats endangering digital objects of interest with the aim of (re)using them in the future on the very long term.

#### Not to be confused with :

- **storage:** recording data on a physical medium
- **backup** : replicating data in order to restore them quickly in case of loss
- **archiving in the ordinary IT meaning**: moving less frequently used data on cheaper storage media (typically magnetic tapes)

An effective digital preservation solution should thus be based on a threat-model.



#### Multiple threats endanger our archives



Natural disasters

► Geo-replication



Storage media failure

**Data monitoring** 





Authentication



Media obsolescenceMedia migration



Human failure

Independent site technical admin



Format obsolescenceFormat migration



Economic breakdown





Organizational issues

Independent site administrations

### A robust preservation solution should provide mitigation measures for all these risks

EARNING



"We need to be active players of preservation, not passive clients of third-party preservation services" (Skinner11)

### As clients, our only guarantee of preservation would be the **service level agreement.**



#### No more technical problems...

#### Legal issues

- What if the service provider goes bankrupt?
- What if data gets lost? Can we claim for damages?

#### **Technical issues**

- No control on the archiving technical policy
- Is migration to another provider possible?

#### **Control is key in digital preservation**





at the cost of many legal and economical issues

# Universities are amongst the oldest institutions around

Average company lifespan on S&P 500 Index (in years)



What organization is better qualified to preserve data on the long term?

### managed by different people

and at low cost.

in independent institutions

Lots of copies regularly verified stored on reliable monitored media, periodically updated, in a secure software environment,

#### The ideal preservation solution should involve all the aspects previously discussed







## The ideal preservation solution should involve all the aspects previously discussed



The 3 pillars of digital preservation: organisation, ressources and technology





#### When is a digital preservation solution good enough ?



It all depends on your threat-model (which includes your limited resources).



### Ideally, it should be evaluated based on a detailed risk and threat assessment.



But in practice, the reliability is very difficult to evaluate.

#### Hence, the more copies, the



#### More copies

Less correlated copies

More reliable copies

Faster failures detection and repair

Less aggressive compression



#### open source software

LOCKSS

originally for PCA (Global LOCKSS Network)

awarded technology (perfect score - TRAC certif.)

a robust integrity check and repair protocol

#### How does LOCKSS work in practice ? Illustration with the **Global LOCKSS Network**



### LOCKSS core : a unique robust integrity monitoring and repairing protocol







state-of-the-art community OSS to cope with fast evolving web



independent modules interacting through REST-API



de-silo components: make poll/repair an independent module



support for large-scale distributed storage (HDFS)



aligning with web archiving standards (WARC-native)



one box = multiple containers possibly on different machines

EARNING LOCKSS preservation capabilities are equally well-suited to any other type of digital content



UNIVERSITAL WALLONIE-











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





#### "poll/repair" is the key component for us



share resources, not money



distributed technical administration, no automation and centralized administration of boxes

# LOCKSS empowers university libraries to fulfill their essential mission of preserving digital knowledge

