Build your own digital preservation network to safeguard your collections

ILIDE conference 2022

Anthony LEROY - PhD Eng.
Université libre de Bruxelles

As University Libraries, we have two main **missions**

Providing **access** to objects selected by curators

Preserving those objects

For analog objects, guaranteeing access and preservation is relatively simple

Books Images Public Domain Free Clipart

But in the digital world, those missions are compromised

we lost control on some digital objects (subscription)

digital objects are vulnerable, especially our local **unique** scientific production and cultural heritage

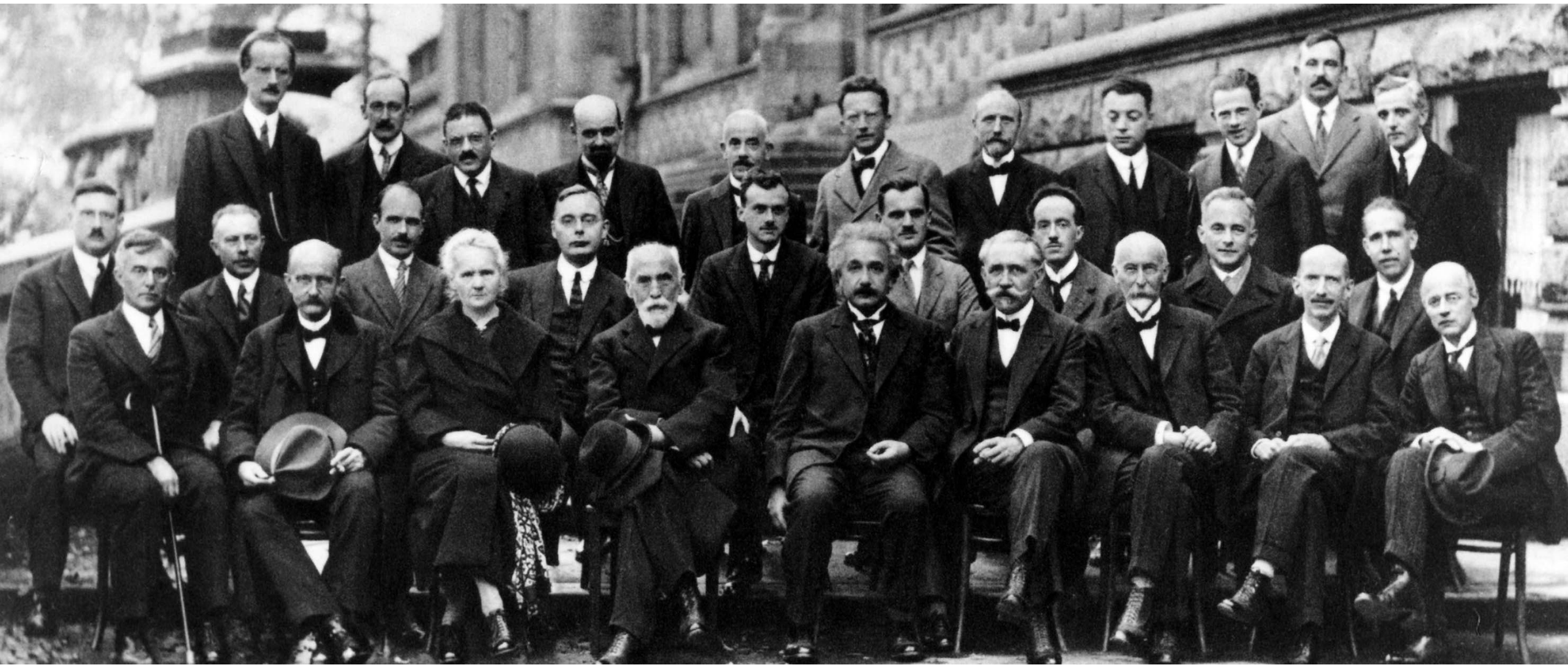
theses

laboratory notebooks

research data

cultural heritage

What will happen to the current research outcome in 100 years?



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

Most institutions maintain an IR to archive and disseminate their digital collections

It supports the mission of guaranteeing **access** to objects.

But what about digital **preservation** ?

Efficient digital preservation is complex

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

An effective digital preservation solution needs to be based on a **threat-model**.

Multiple threats endanger our archives

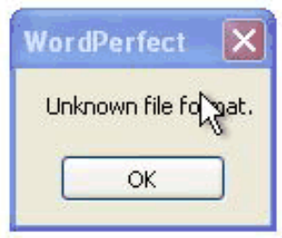
Natural disasters

Storage media failure

Internal or external attacks

Media obsolescence

Human failure



Format obsolescence

Economic breakdown

Organizational issues

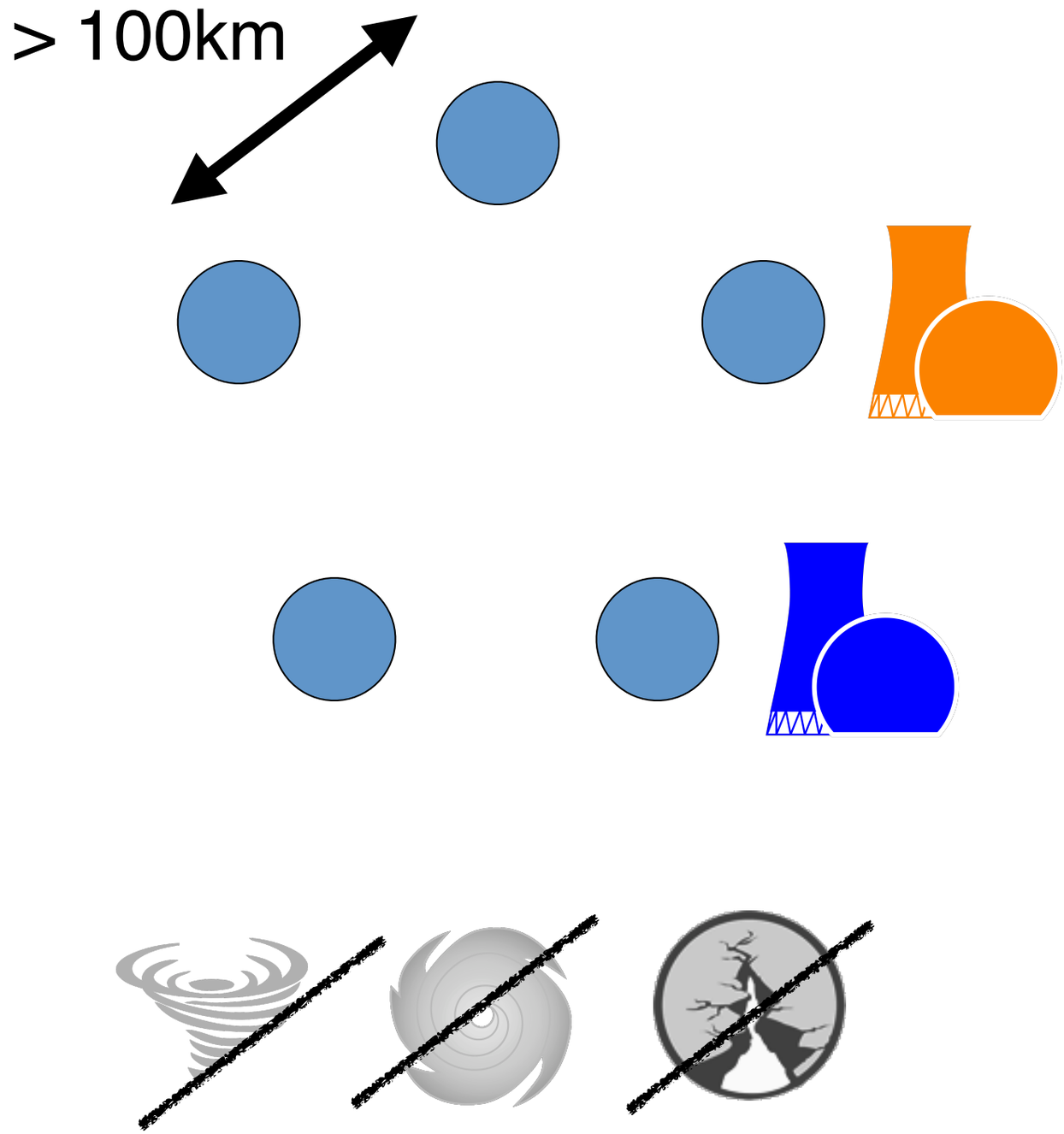
1. Natural and man-made disasters

https://flic.kr/p/9MsNf9

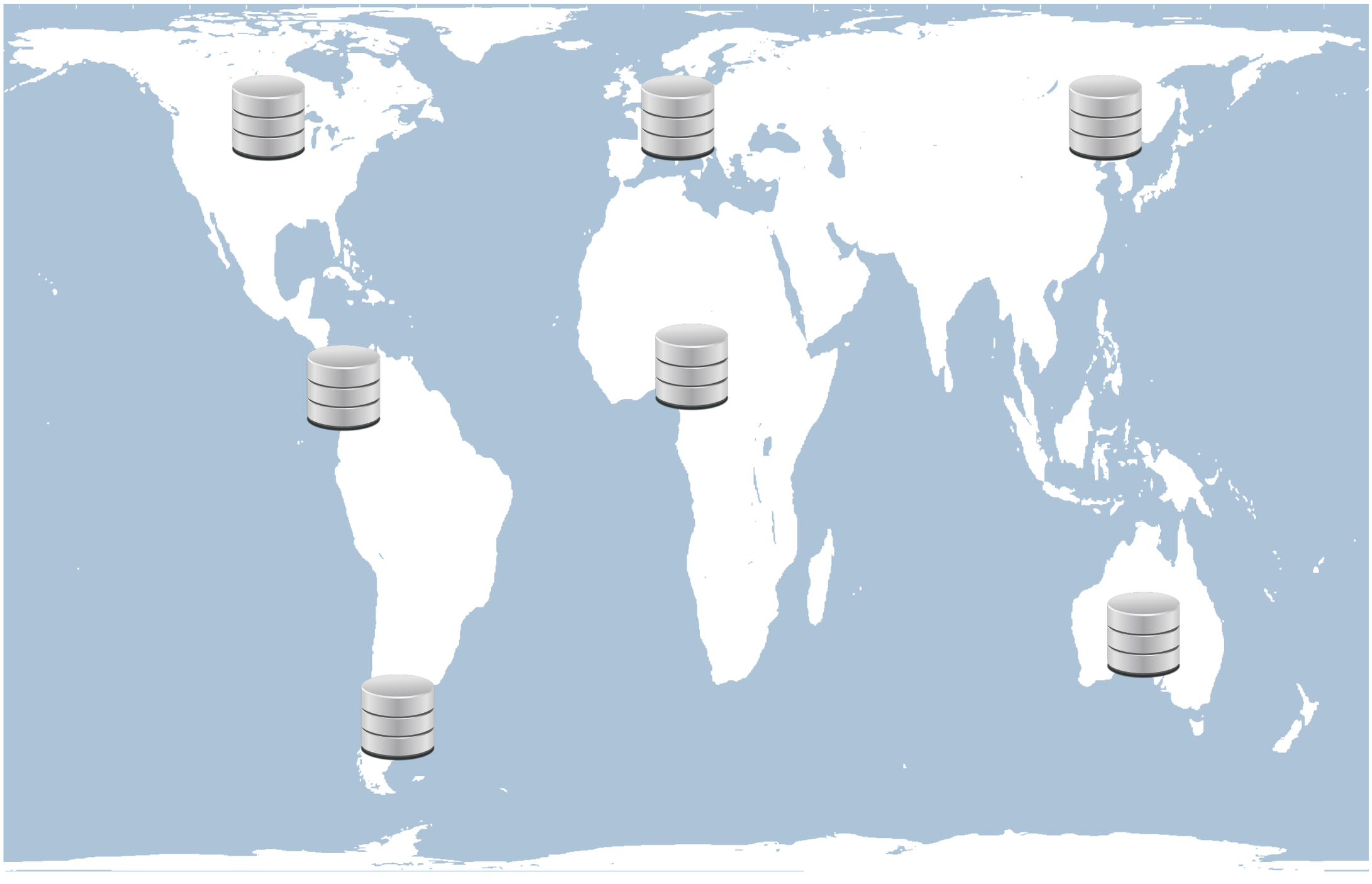
Very difficult risk to evaluate… low-probability with major impact events

1. Natural and man-made disasters

→ Geo-replication

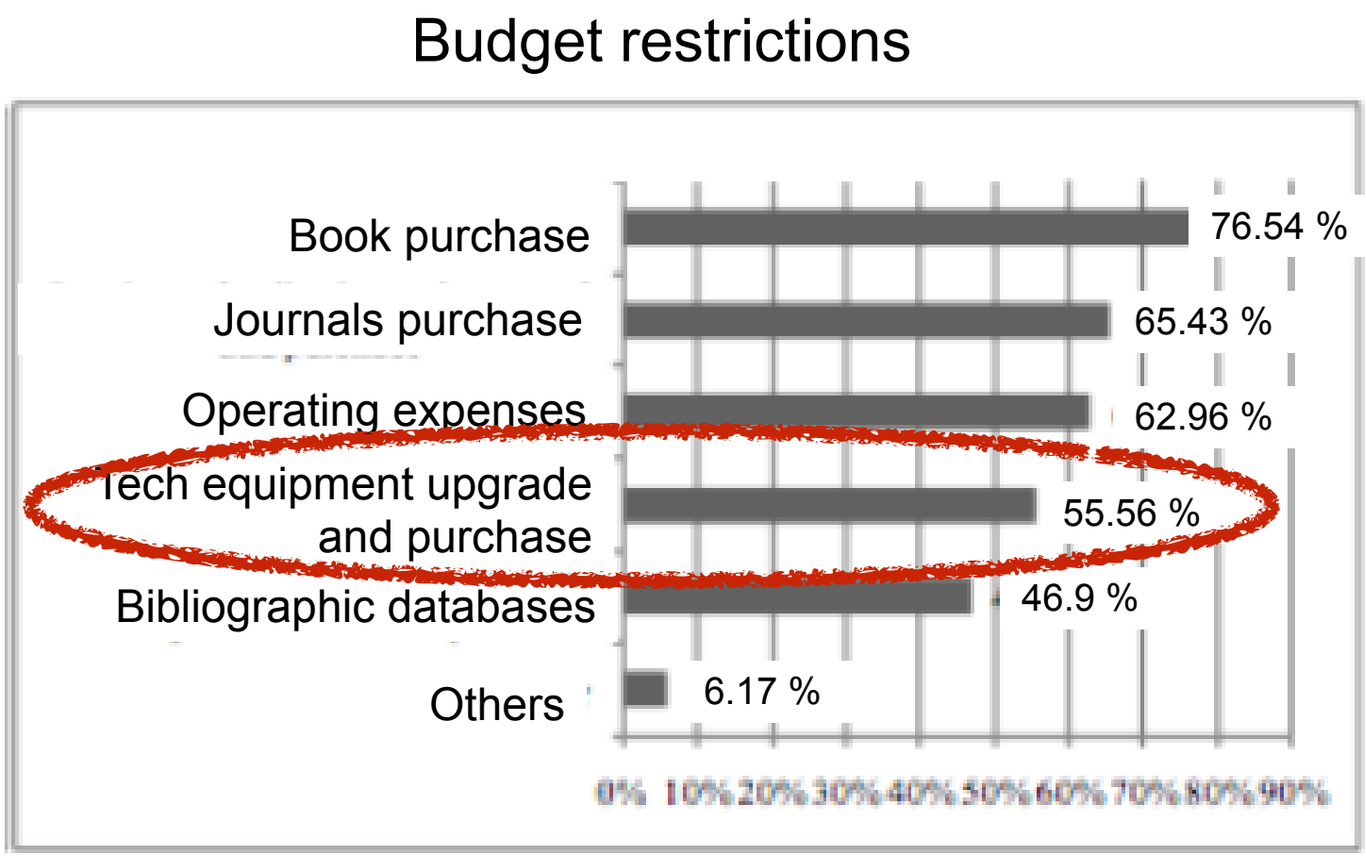


The ideal solution : geo-replication of copies



2. Economic risk

In times of economic crisis, budget restrictions first and foremost concern activities that do have low impact on the short term future

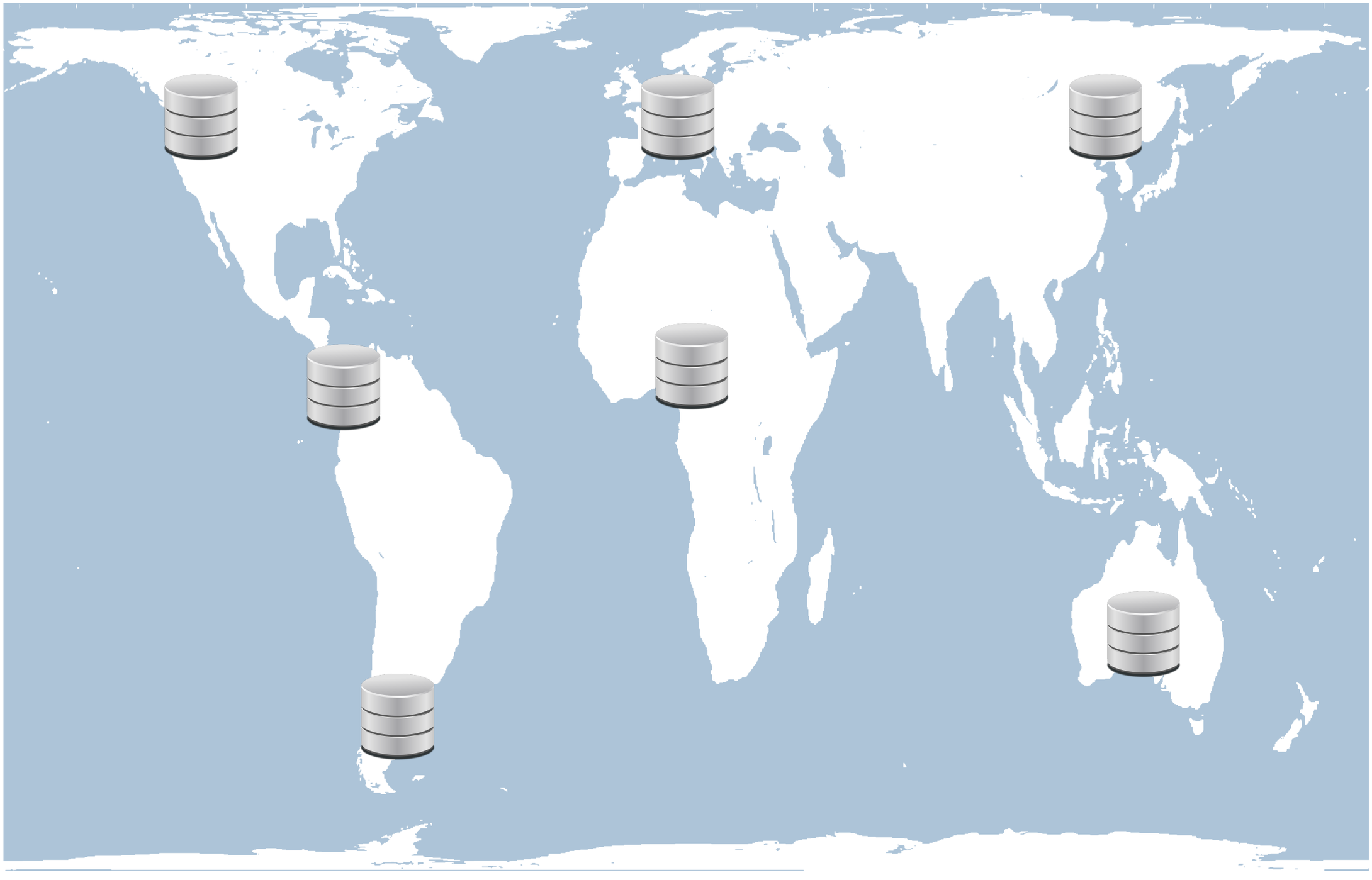


G. Giannakopoulos et al., Libraries in Crisis: A Glimpse over Greece and Cyprus, Procedia, Vol.147, 2014, https://doi.org/10.1016/j.sbspro.2014.07.121.

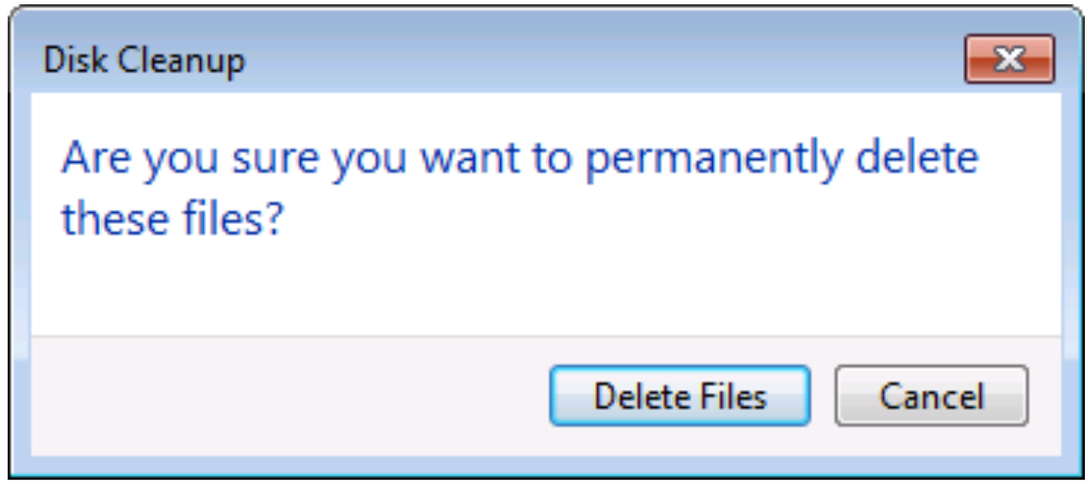
If investments in the infrastructure maintenance cannot be done for just 5 years, the risk of data loss considerably increases.

2. Economic risk

→ Cost control



3. Human errors



It happens…

Amazon's Cloud Crash Disaster Permanently Destroyed Many Customers' Data

Henry Blodget
Apr. 28, 2011, 7:10 AM 112,400

AWS Outage that Broke the Internet Caused by Mistyped Command

Amazon says Tuesday's mayhem resulted from mistake during a routine debugging exercise.

Yevgeniy Sverdlik | Mar 02, 2017

British Airways: Engineer Wrongly Disconnected Data Center Power Supply

How a single tech could cause so much damage, and why the backup system failed remain open questions.

Bloomberg | Jun 06, 2017

Human Error Cited in Hosting.com Outage

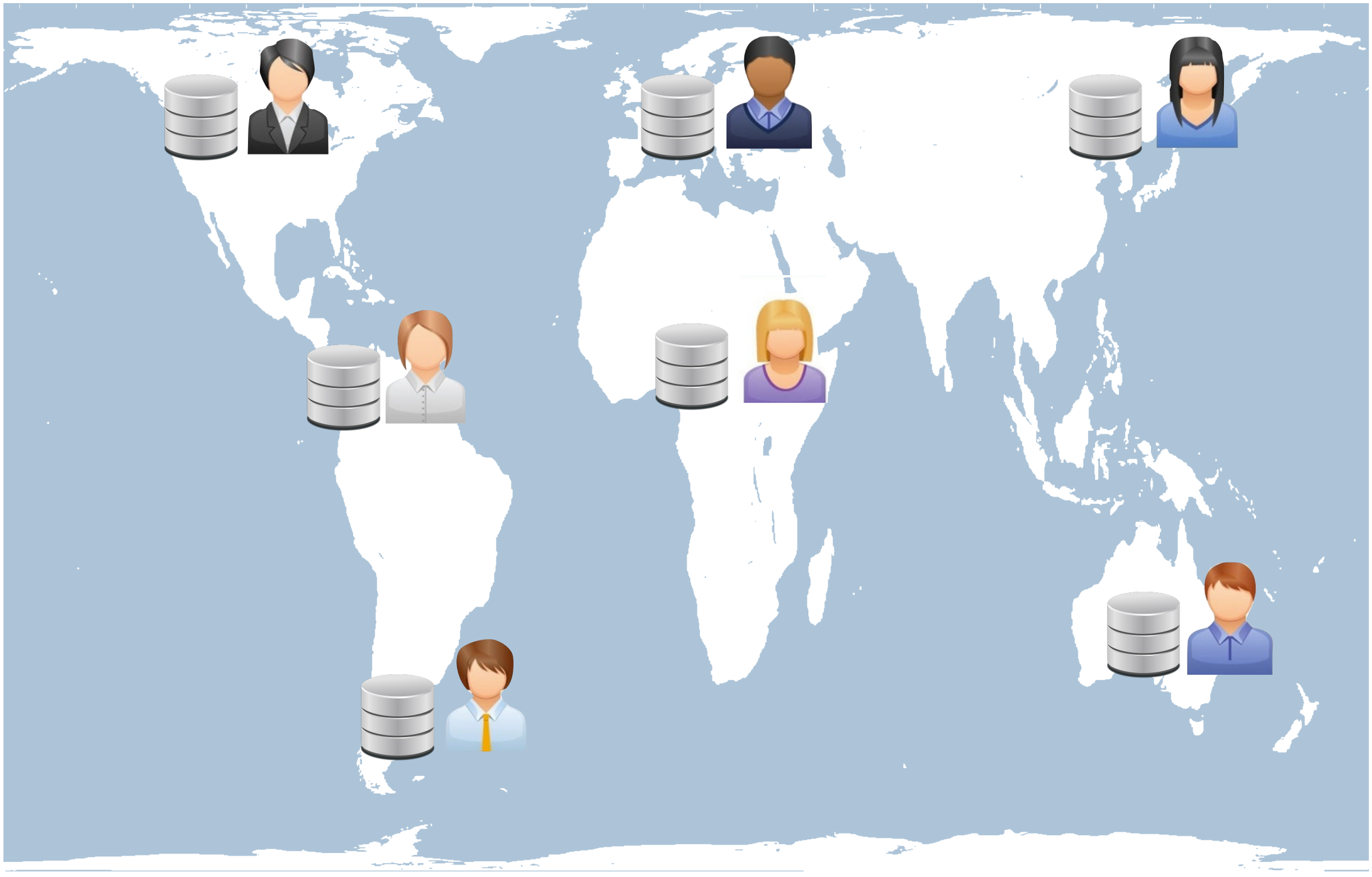
Hosting.com said human error was responsible for a data center power outage that left more than 1,100 customers without service. The downtime occurred as the company was conducting preventive maintenance on a UPS system in the company's data center in Newark, Del.

Rich Miller | Jul 28, 2012

http://www.datacenterknowledge.com/

3. Human errors

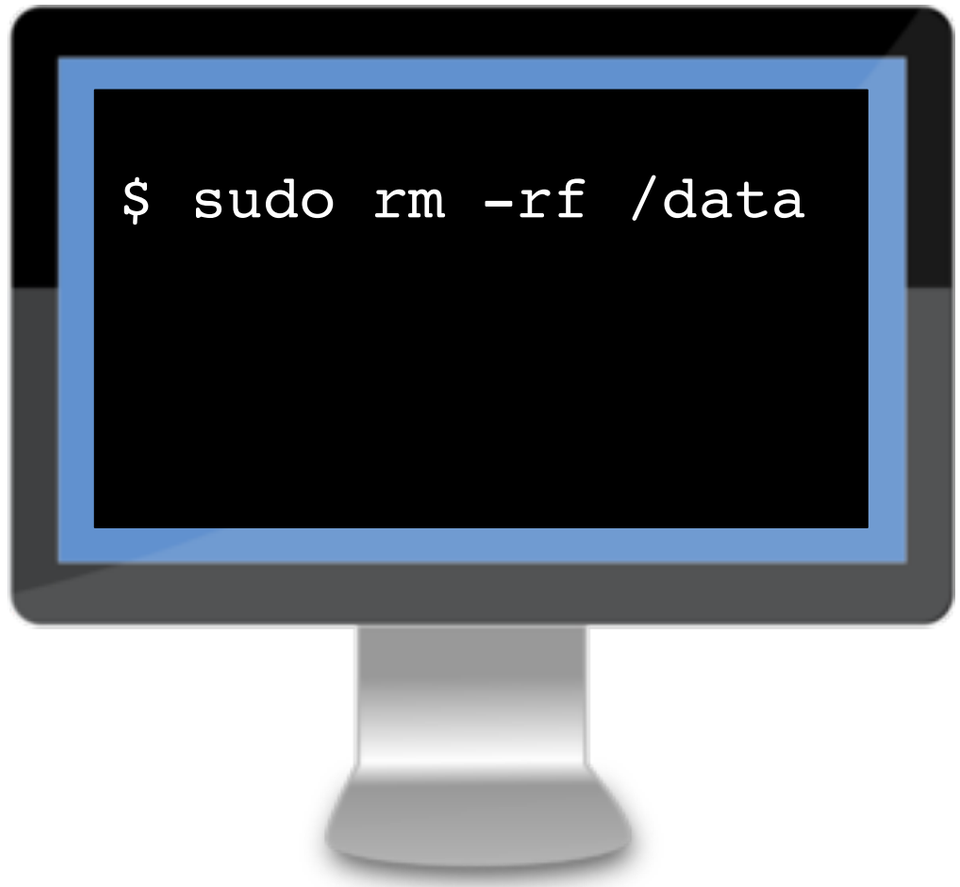
→ Independent technical management



4. Organizational, political and ideological risks

A digital autodafé is much more efficient than its analog equivalent

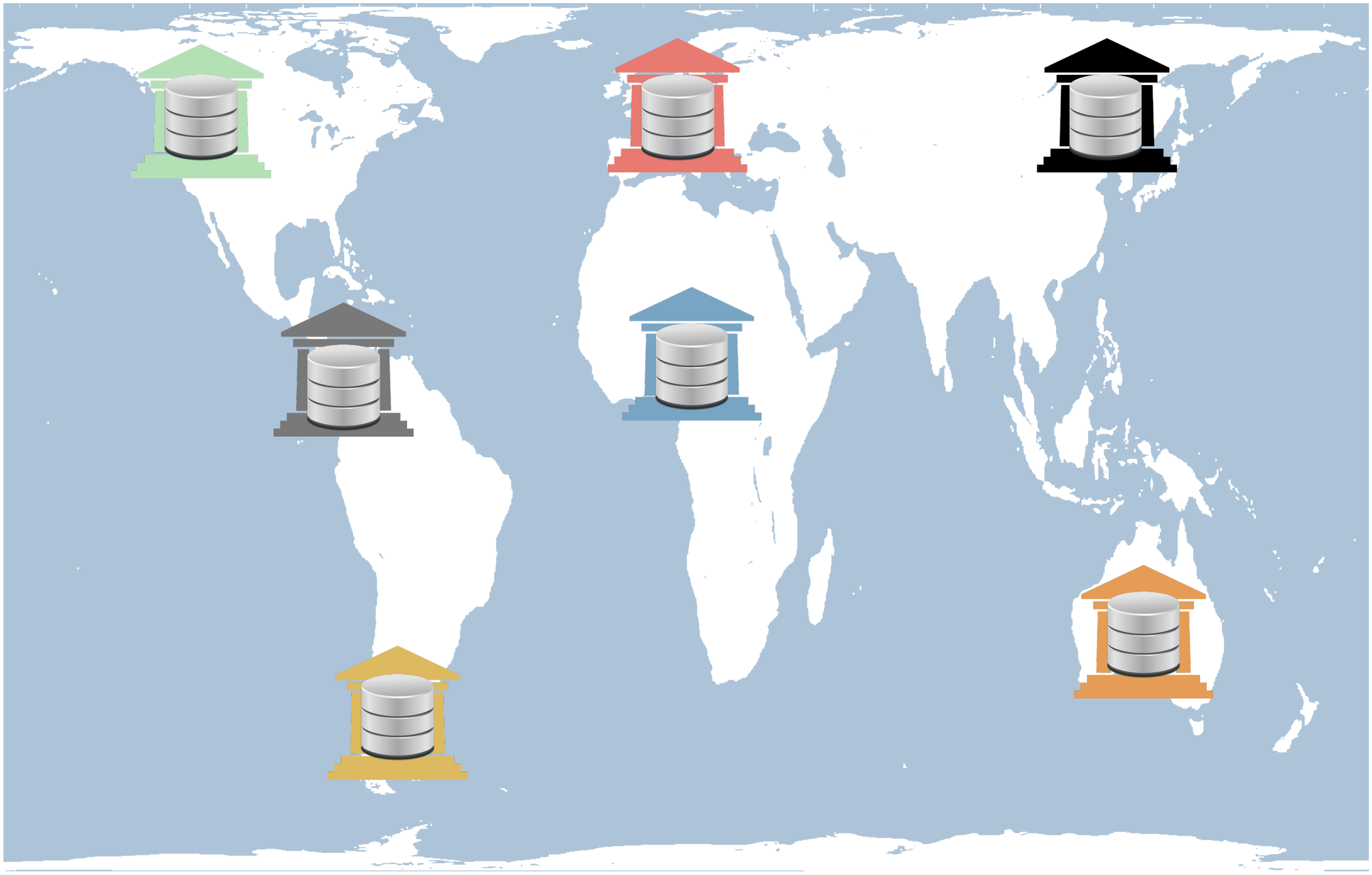
In the 15th century



Today

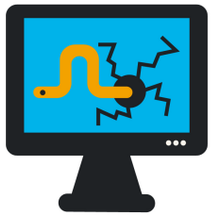
4. Organizational risk

→ Administrative independence



5. External or internal attacks

access-based attacks



viruses
worms

ransomware

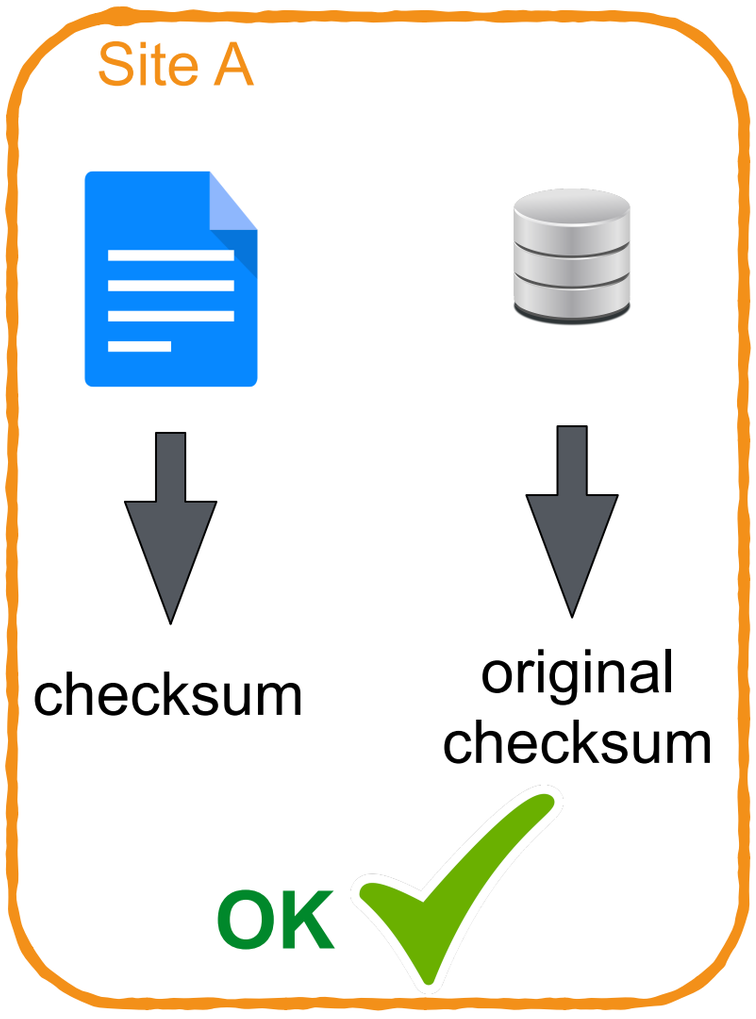
cyber espionage

Security measures :

- Regular software updates
- Authentication
- Firewall

Image from macrovector_official on Freepik

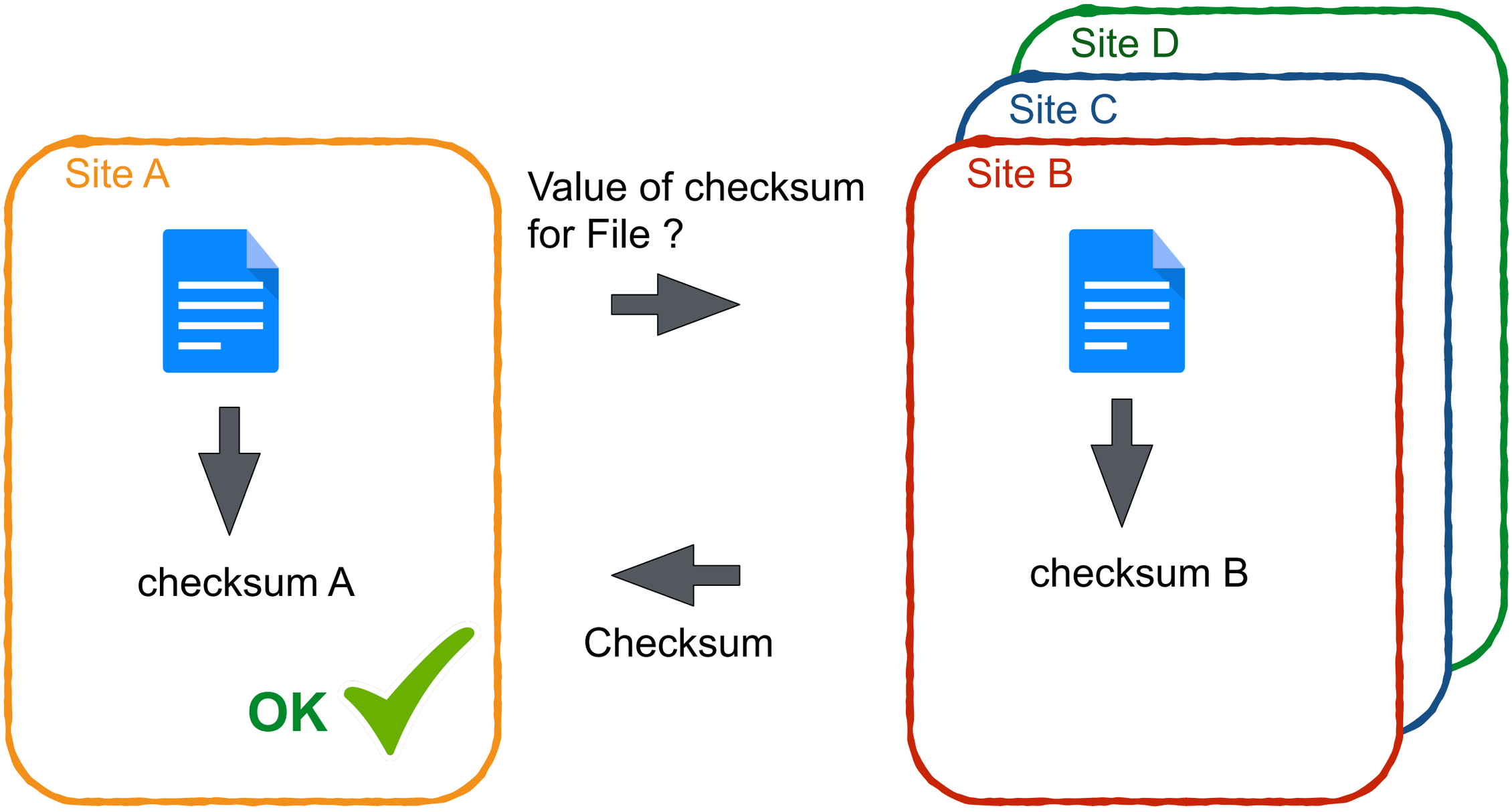
Integrity monitoring solution: checksums



If checksums do not match:

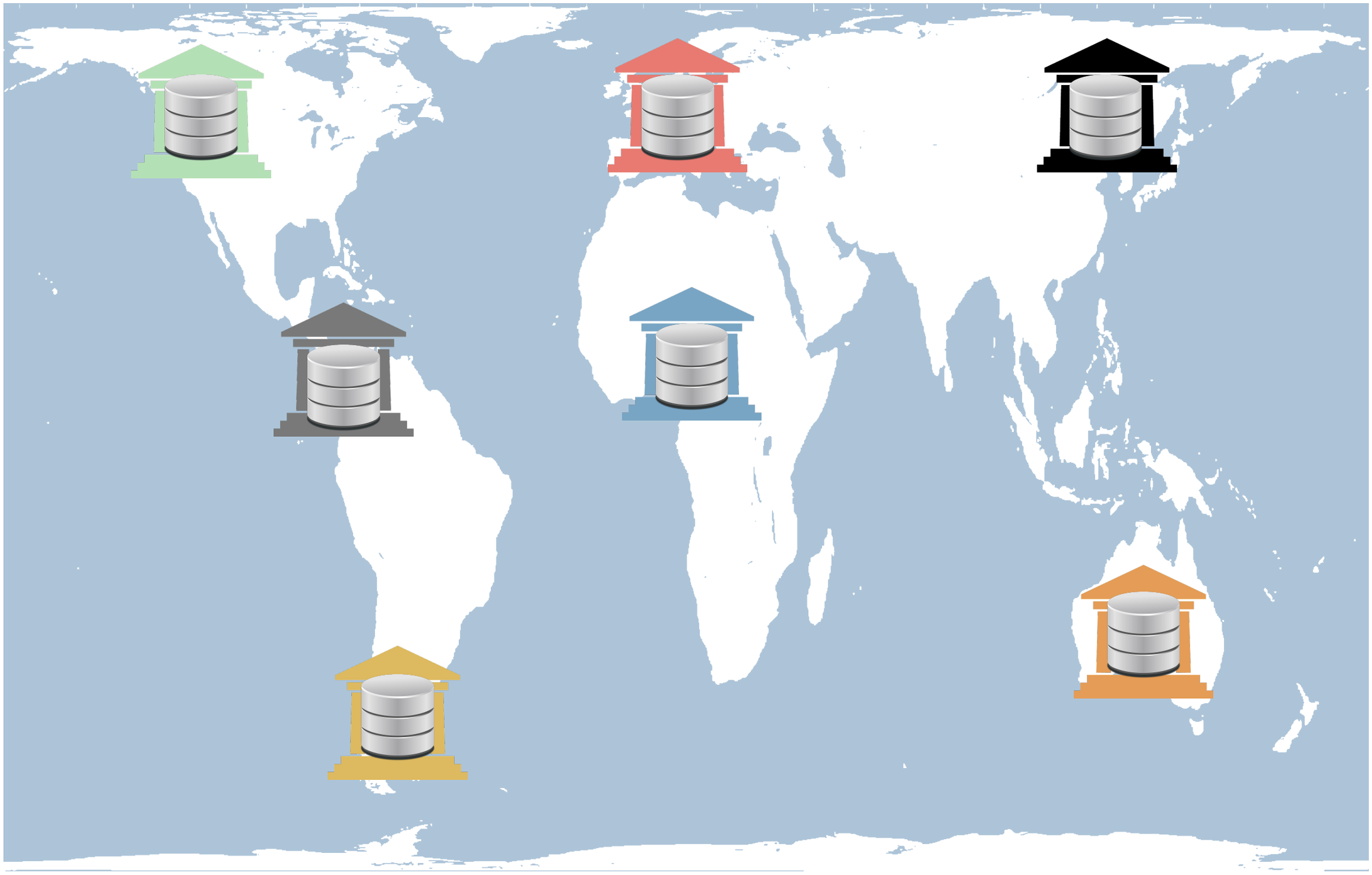
- Is the copy compromised?
- Is the checksum compromised?
- Are they both compromised?

Integrity monitoring solution: in a network



We need a **better** integrity check method based on **real-time computation** that can only be performed when in **possession** of an intact copy.

5. Secure protocol for monitoring the integrity of all copies

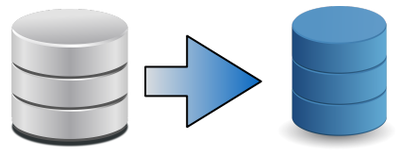


“We need to be **active** players of preservation, not passive clients of third-party preservation services”

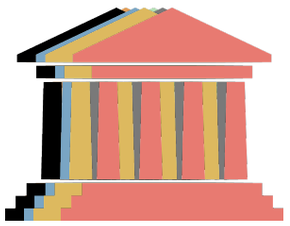
(Katherine Skinner, 2011)

The ideal preservation solution should combine all the aspects previously discussed

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



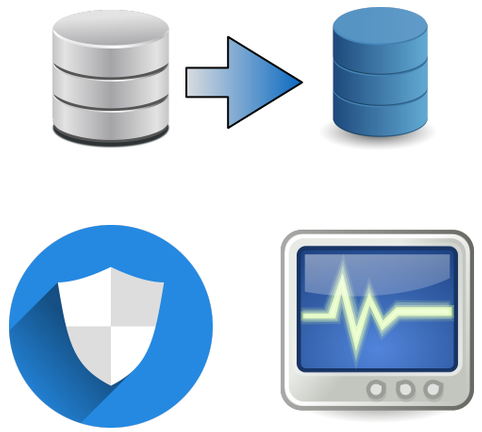
managed by different people around the globe in independent institutions



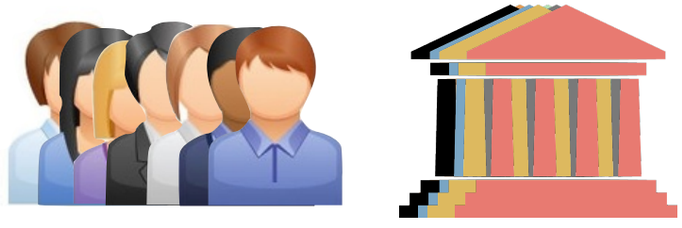
and at low and controlled cost.

A preservation solution is based on three pillars: technology, organization and resources

Technology



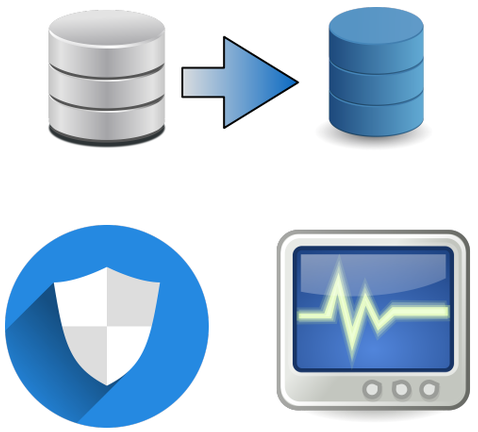
Organization



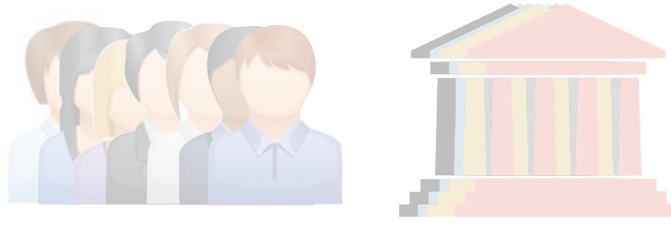
Resources

A preservation solution is based on three pillars: technology, organization and resources

Technology



Organization



Resources

LOCKSS: a well-proven technology

open source software

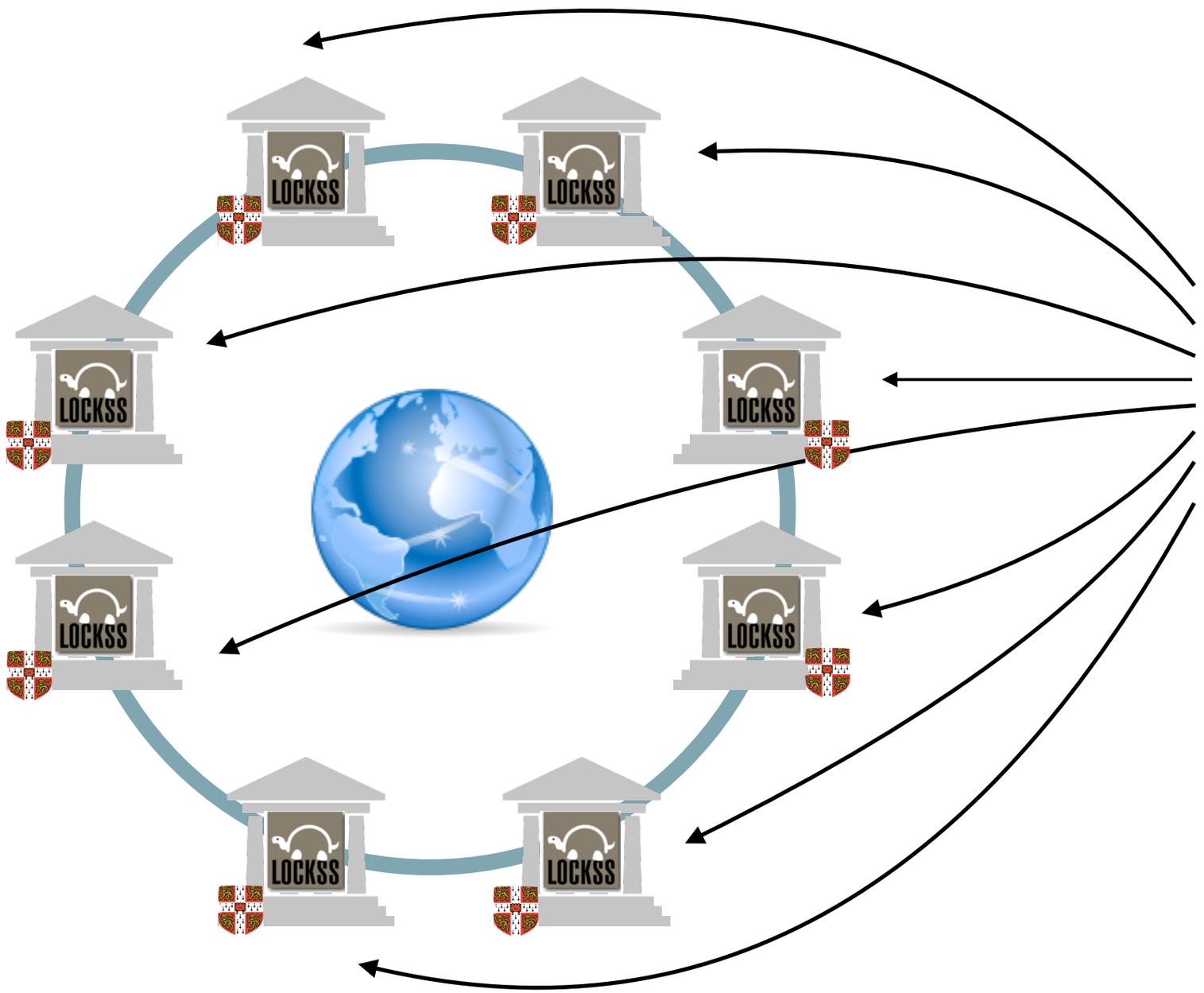
originally for PCA (Global LOCKSS Network)

awarded technology (perfect score - TRAC certif.)

a robust integrity check and repair protocol

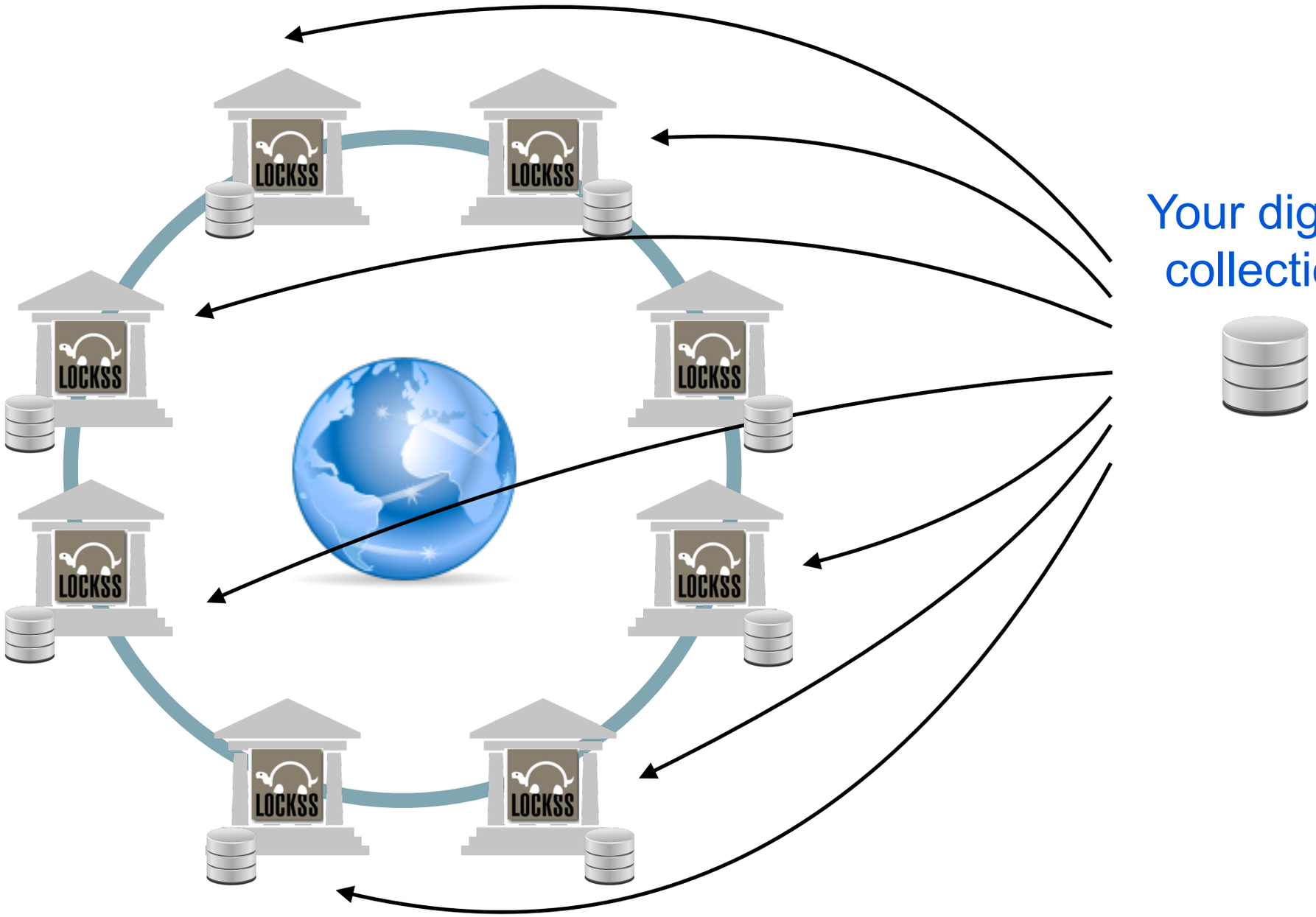
modernized software architecture (LOCKSS 2.0)

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



CAMBRIDGE
UNIVERSITY PRESS

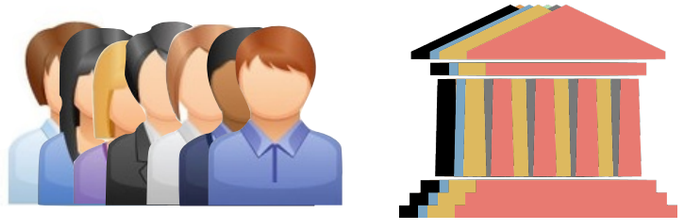
LOCKSS preservation capabilities are equally well-suited to **any digital collection**



A preservation solution is based on three pillars: technology, organization and resources

Technology

Organization



Resources

A diversity of LOCKSS networks preserve a wide range of digital content

General interest scientific journals

LOCKSS

CLOCKSS

PKP
PUBLIC
KNOWLEDGE
PROJECT

Scientific journals of regional interest

Cariniana

Government information

CGI

FDSYS
FEDERAL DIGITAL SYSTEM

Content of common interest (incl. grey literature)

THE ALABAMA DIGITAL PRESERVATION NETWORK
PRESERVING ALABAMA'S DIGITAL RESOURCES

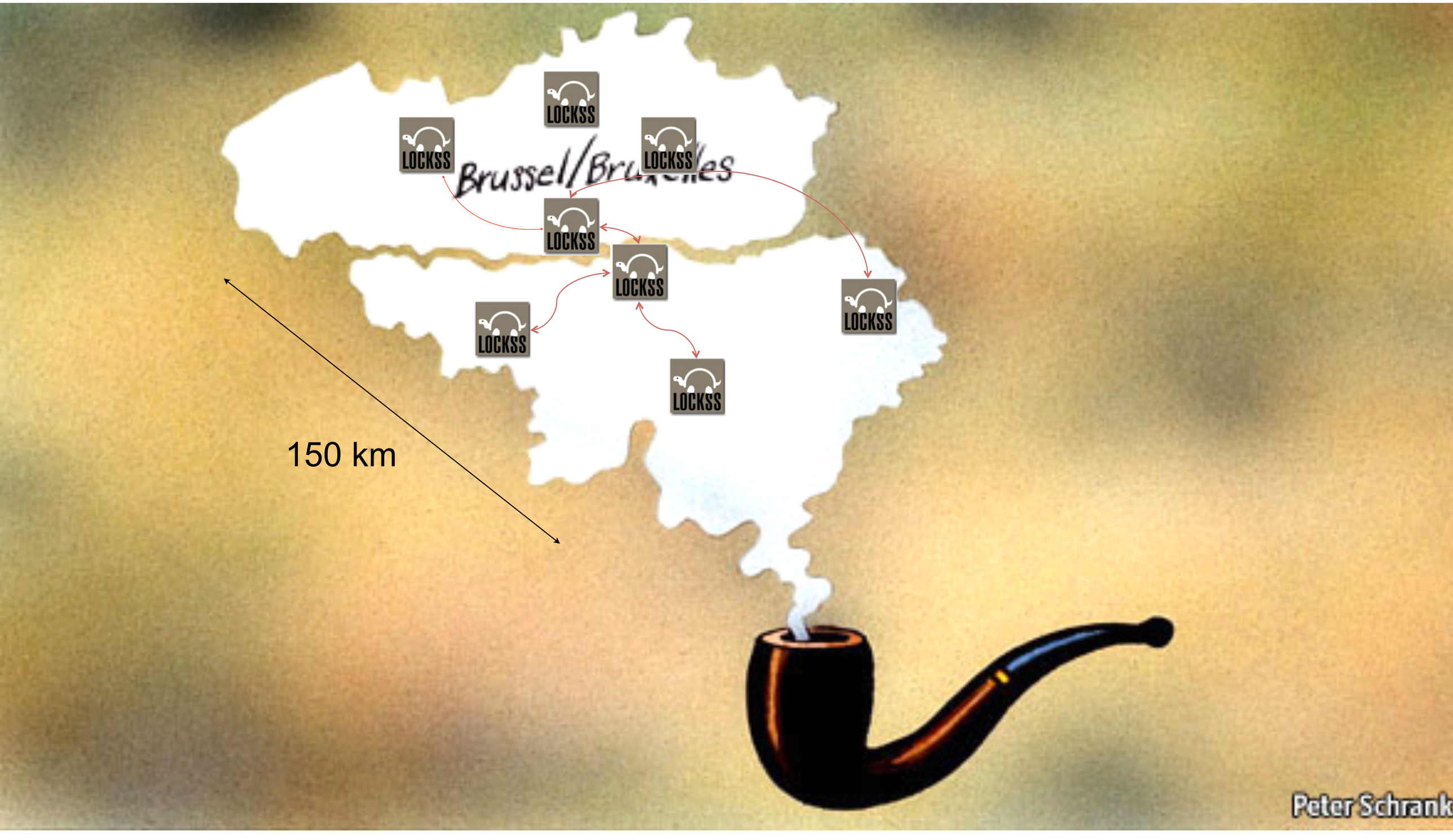
WestVault

Meta Archive

DIGITALCOMMONS™
powered by bepress

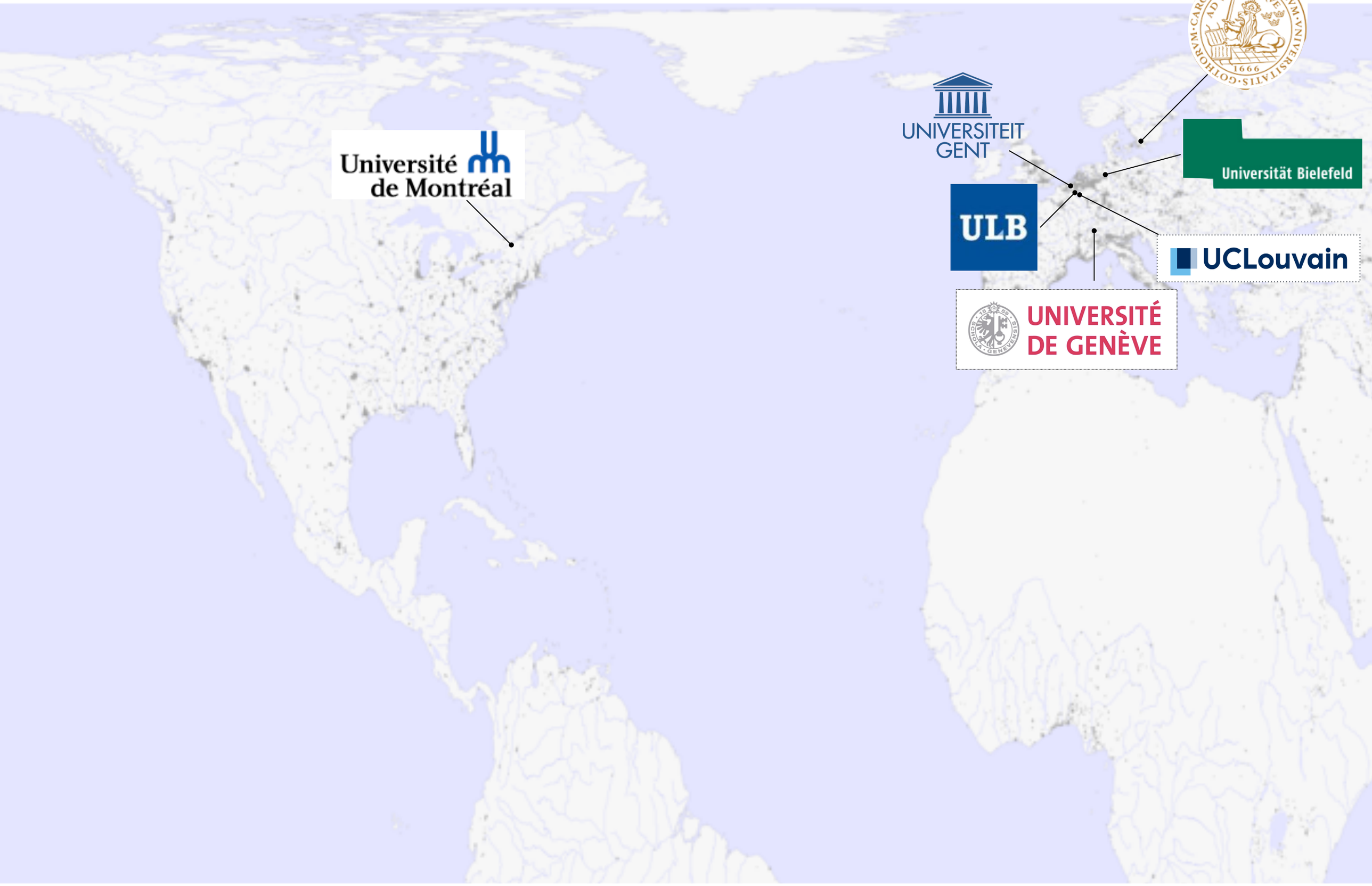
SAFE

Ceci n’est pas un PLN !



Oh, come on! Tell me you have a backup of Belgium!

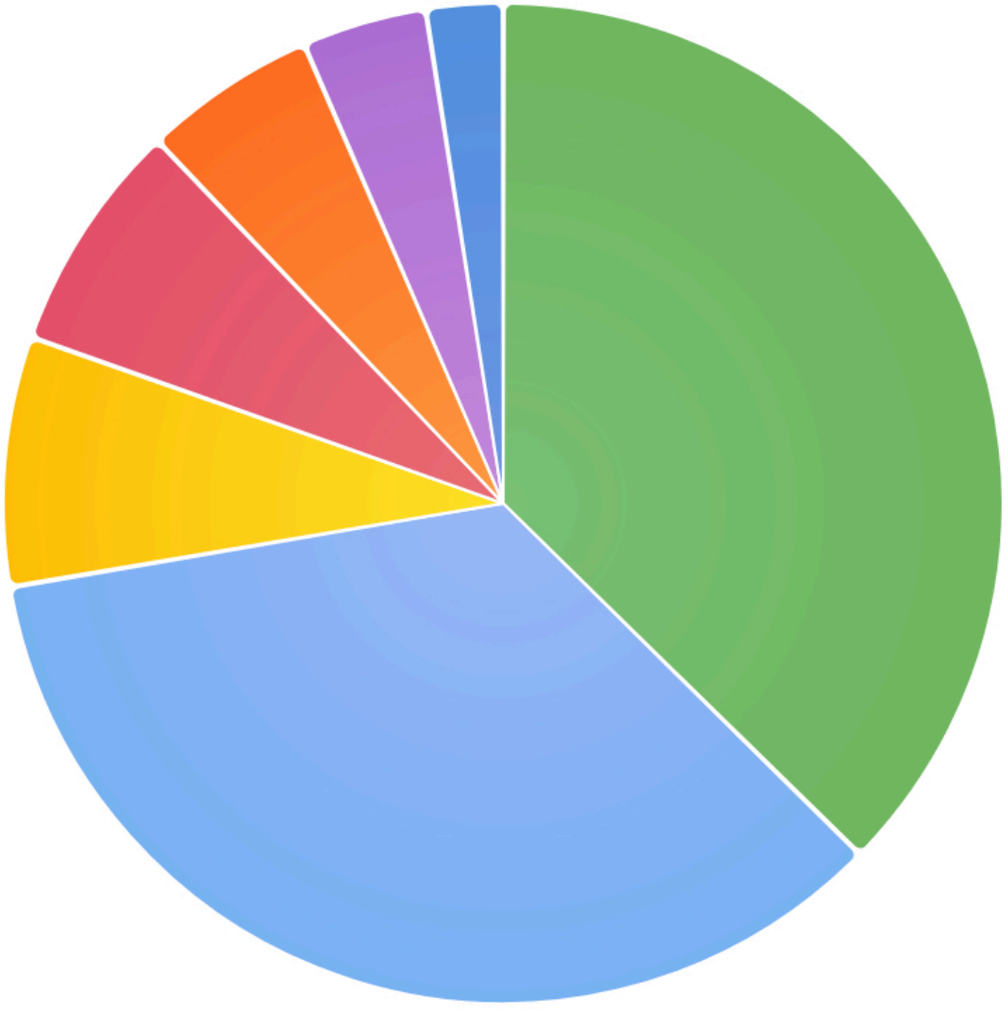
SAFE has 7 members



What does SAFE preserve?

Open access content collected from our Institutional Repositories:

- theses
- books, reports
- publications



Université de Montréal	332 GB
Universiteit Gent	311 GB
Lund University	71 GB
Universite de Geneve	66 GB
Uni. Bielefeld	49 GB
Université Libre de Bruxelles	35 GB
UCLouvain	21 GB

1.2 TB
53 collections - 125k objects

SAFE

SAFE Archive FEderation

international federation

geo-replication in completely independent sites

light organizational structure

7+ nodes

budgets remain fully independent

economic risk mitigation

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

LOCKSS Network dashboard

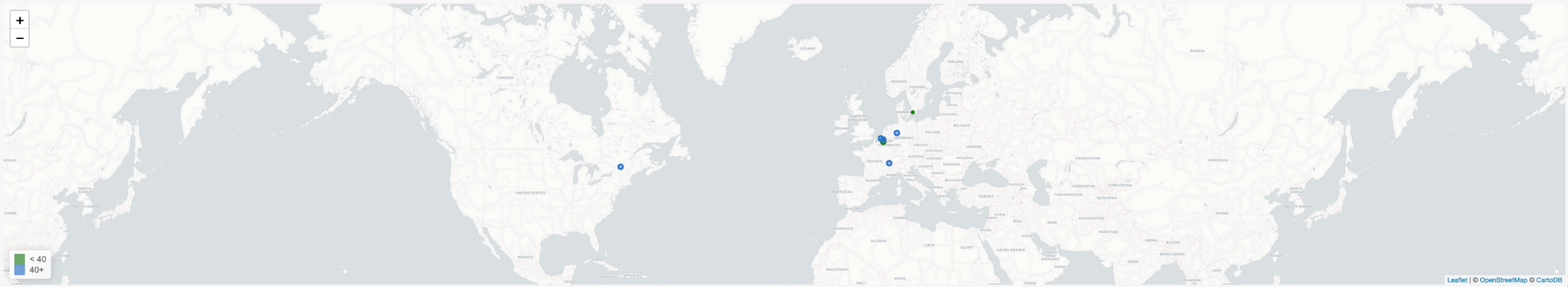
General / LOCKSS network monitor

Last 1 year

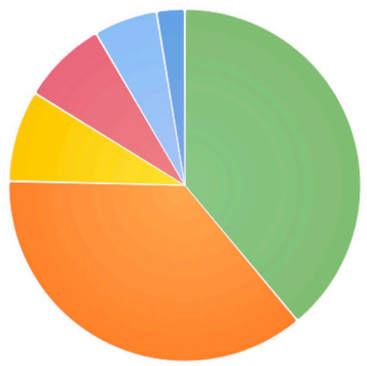
LOCKSS network dashboard

Total archive size	Active Boxes	Active AUs	Avg copies/AU	Average poll agreement	Last completed poll	Min. free capacity
1.16 TB	7	51	6.8	99.3%	2022-06-19 14:11:06	702 MB

LOCKSS Network map

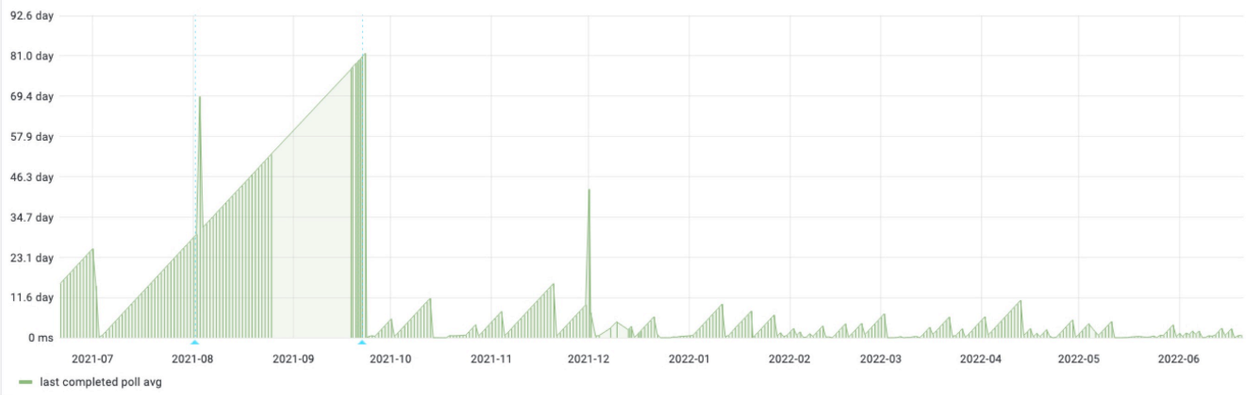


Archival Units per publisher



	Value
Université de Montréal	332 GB
Universiteit Gent	311 GB
Lund University	71 GB
Universite de Geneve	66 GB
Uni. Bielefeld	49 GB
UCLouvain	21 GB

Maximum time since last completed poll



LOCKSS boxes

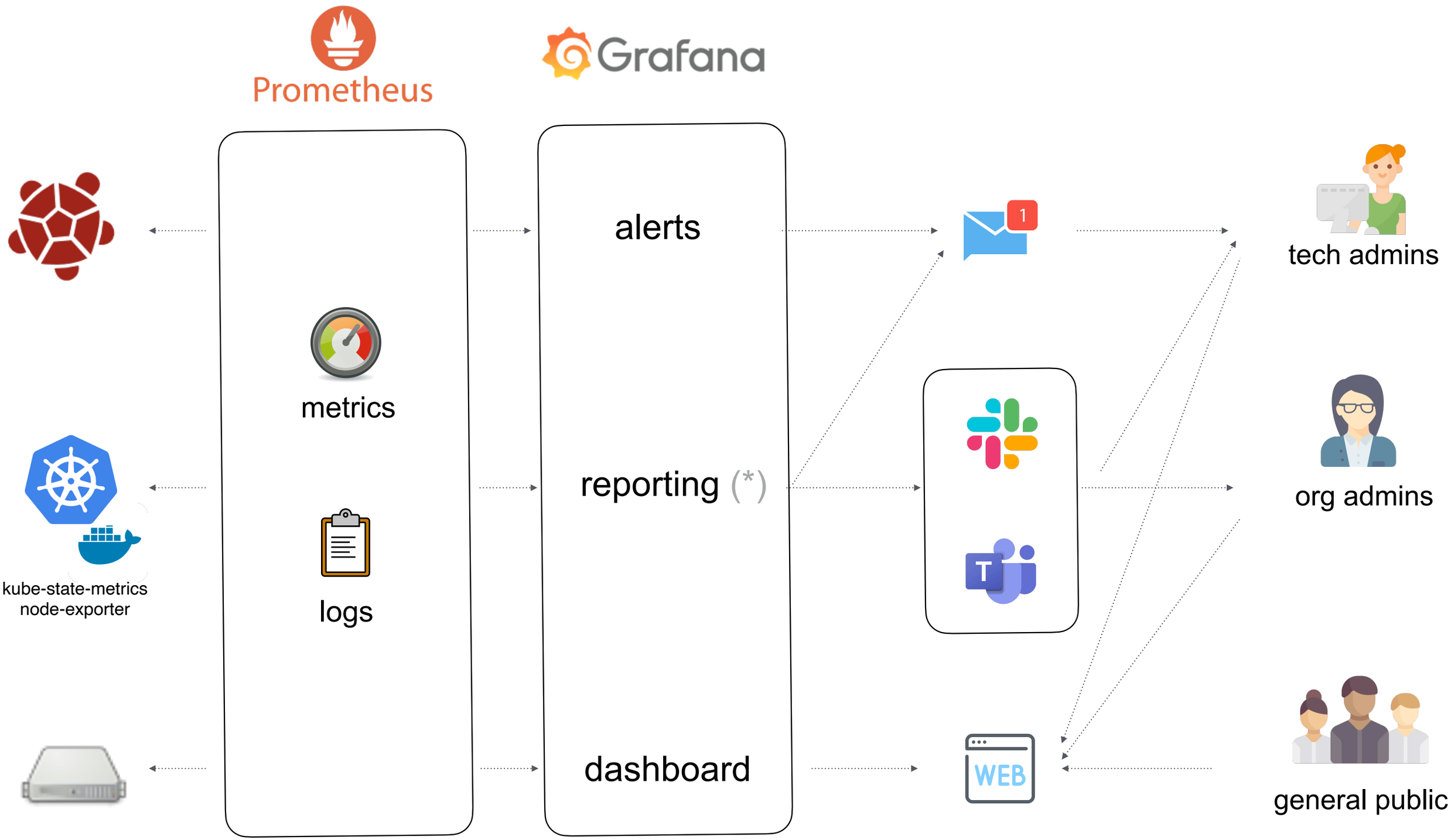
LOCKSS box name ↑	IP address	Country	Daemon version	Active AUs	Active since	Platform	Java version
Lund	130.235.140.200	Sweden	1.75.9	52	2017-07-05 20:23:24	Linux RPM	1.8.0_332
UCL	130.104.5.94	Belgium	1.75.9	50	2017-07-05 20:23:16	Linux RPM	1.8.0_312
UdeM	132.204.9.246	Canada	1.75.7	48	2017-07-05 20:23:28	Linux RPM	1.8.0_275
UGent	157.193.230.142	Belgium	1.75.9	46	2020-02-05 08:33:54	Linux RPM	1.8.0_322
ULB1	164.15.1.85	Belgium	1.75.9	46	2017-07-05 20:23:15	Linux RPM	1.8.0_272

Archive status across the network

AU status across the network

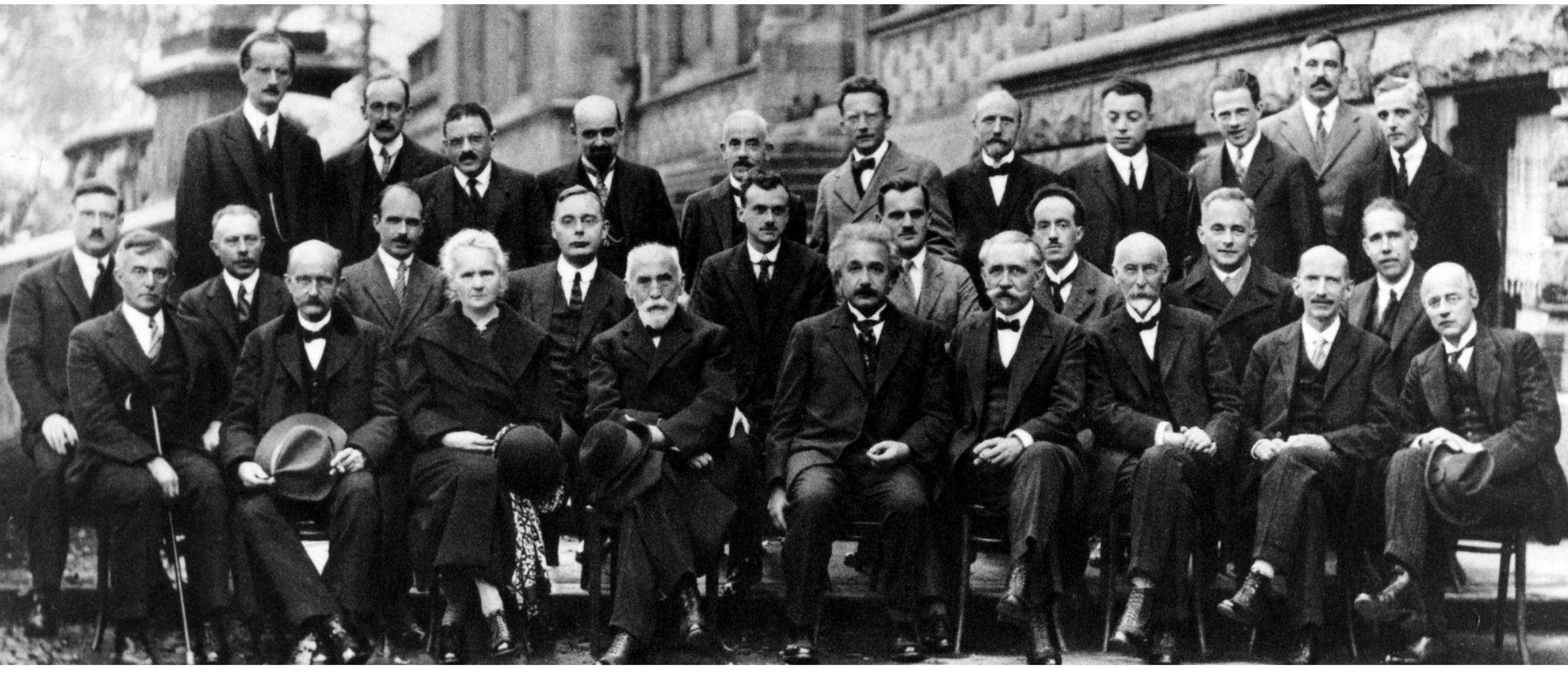
name	ulb	ucl	unige	udem	lund	ugent	unibi
Université de Montréal - Open access collection 2017	100%	100%	100%	100%	100%	100%	100%
Université de Montréal - Open access collection 2016	100%	100%	100%	100%	100%	100%	100%
Université de Montréal - Open access collection 2014-2015	100%	100%	100%	100%	100%	100%	100%
Université de Montréal - Open access collection 2010-2013	100%	100%	100%	100%	100%	100%	100%
Université de Montréal - Open access collection 2005-2009	100%	100%	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2020	100%	No Quorum	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2018	100%	100%	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2015-2017	100%	100%	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2011-2014	100%	100%	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2005-2010	100%	100%	100%	100%	100%	100%	100%
Uni. Bielefeld - Open access Pub collection 2004	100%	100%	100%	100%	100%	100%	100%
UCLouvain Open Access ThesisCM 2017	99%	100%	99%	100%	99%	72%	100%
Open Access Theses from Lund University 2017	100%	100%	100%	100%	100%	100%	100%
Universite de Geneve Publications 2019	100%	No Quorum	100%	100%	100%	100%	100%
Universite de Geneve Publications 2018	100%	100%	100%	100%	100%	100%	100%
Universite de Geneve Publications 2017	100%	100%	100%	100%	100%	100%	100%
Universite de Geneve Publications 2016	100%	100%	100%	100%	100%	100%	100%
Universite de Geneve Publications 2015	100%	100%	100%	100%	100%	100%	100%
Universite de Geneve Publications 2014	100%	100%	100%	100%	100%	100%	100%

Alerts and reporting



(* requires Grafana Enterprise)

Thanks to our robust preservation solution, we are confident that our objects will be preserved for future generation



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

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