



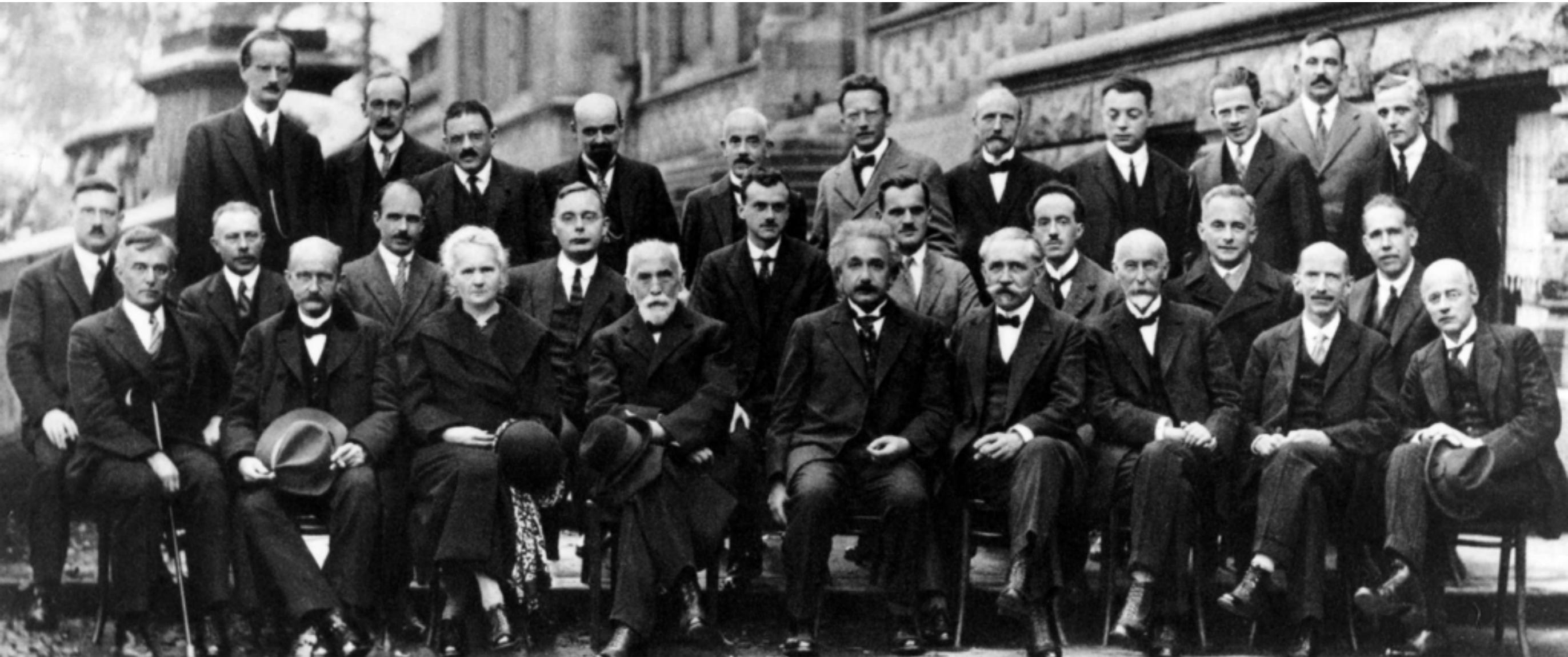
La préservation du patrimoine scientifique à l'heure du numérique

Journée FNRS - 31 janvier 2018

Anthony Leroy

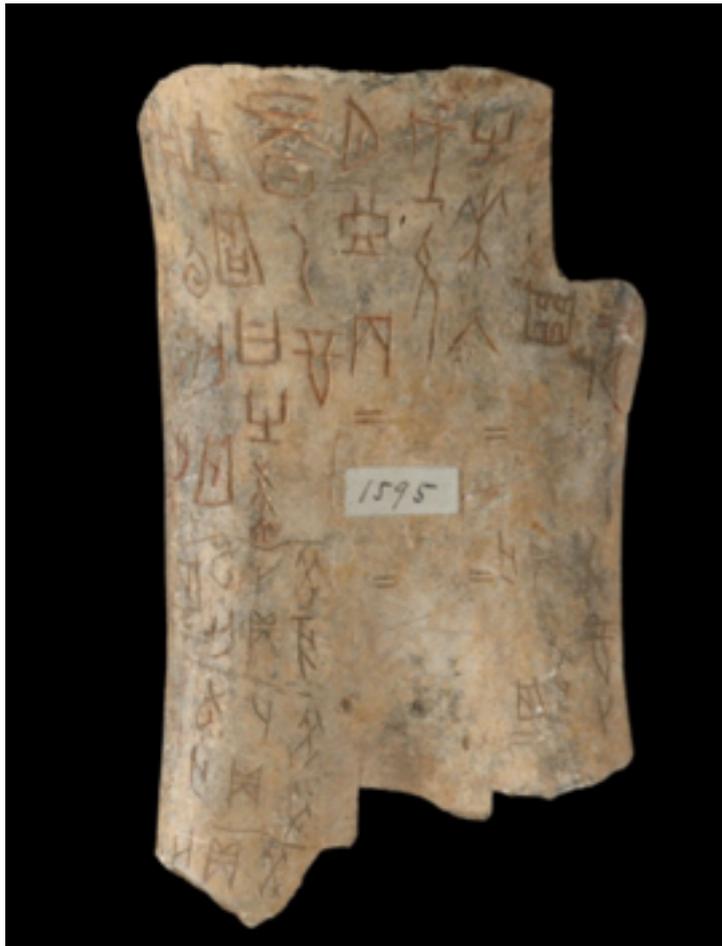
Département des Bibliothèques et de l'Information Scientifique (DBIS)
Université libre de Bruxelles

Qu'advientra-t-il des résultats de la recherche scientifique actuelle... dans 10 ans ? 100 ans ? 1000 ans ?

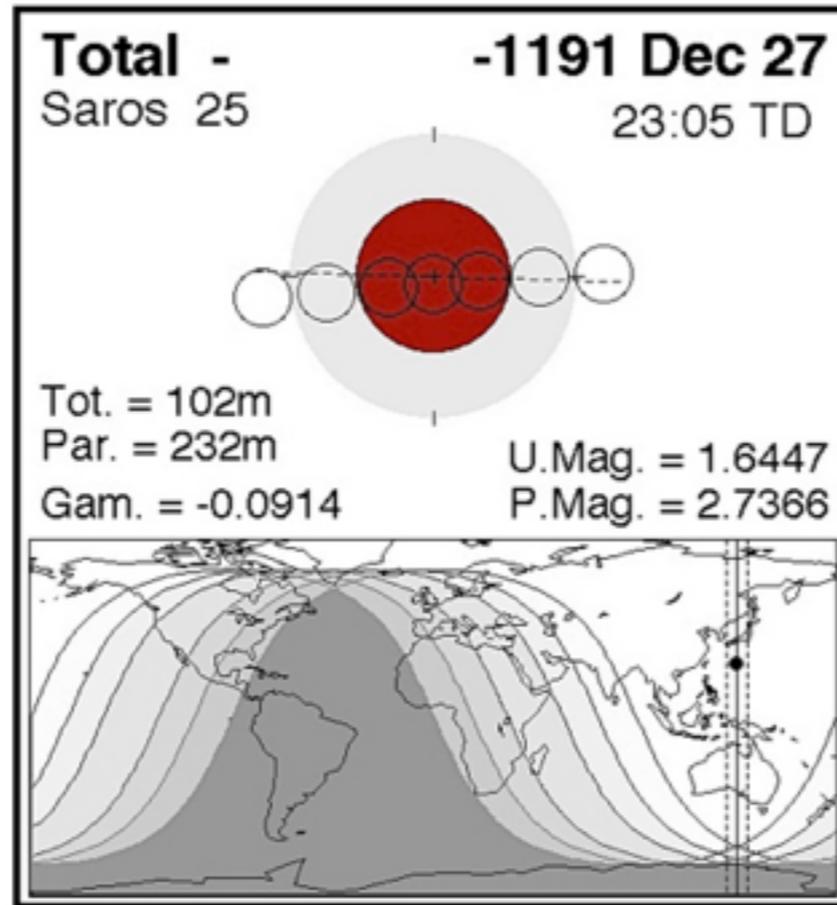


Participants au 5e Congrès Solvay, octobre 1927, à l'Institut international de physique Solvay dans le parc Léopold, Bruxelles

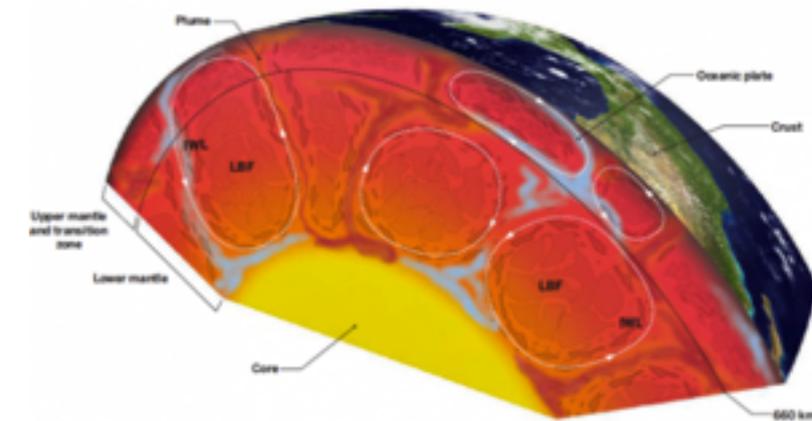
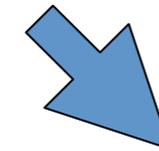
Sur les épaules des géants...



Inscriptions oraculaires, Dynastie Shang, 1191 BC

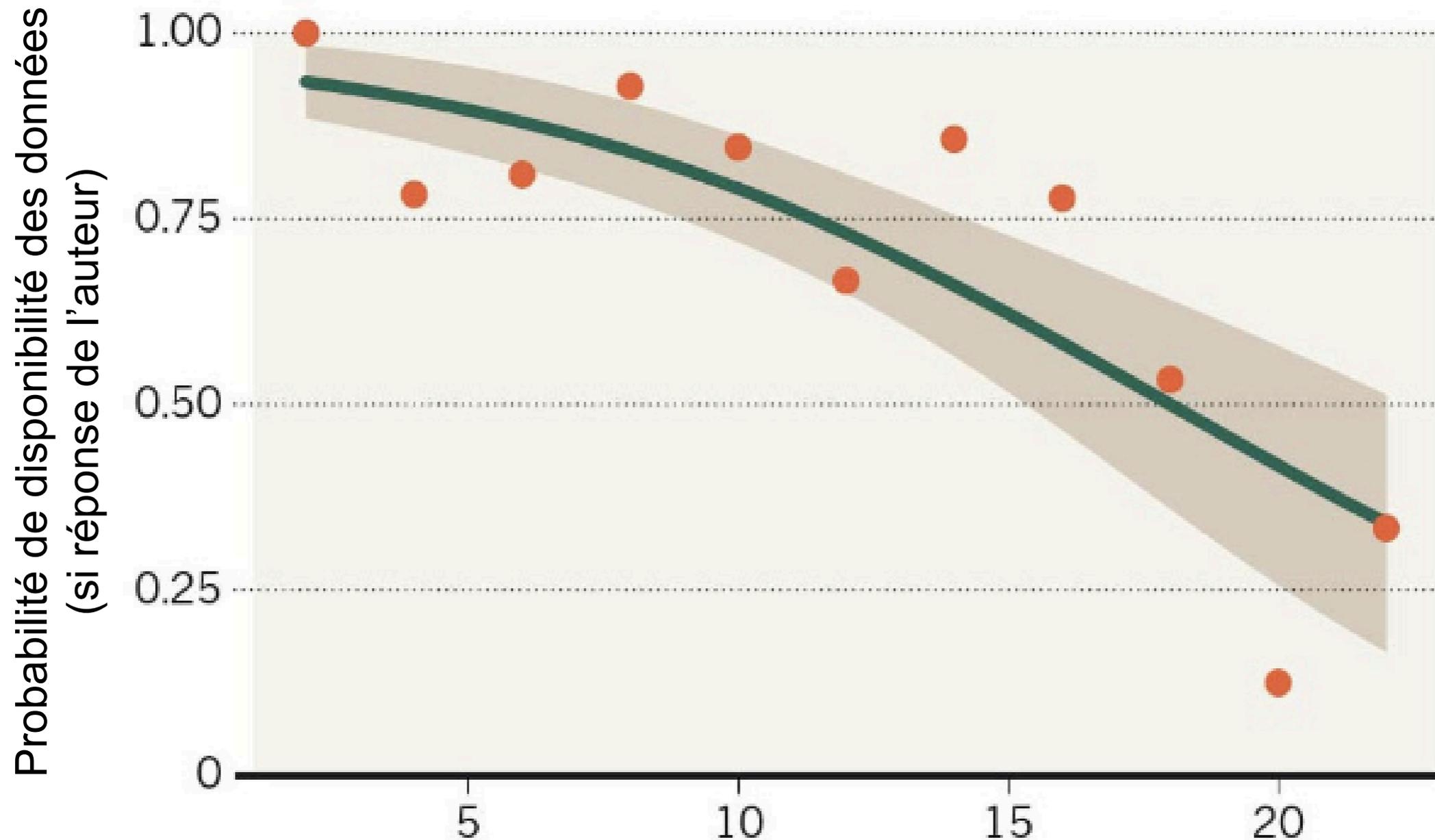


Modèle astronomique ($\Delta T = 7h / 3300$ ans)



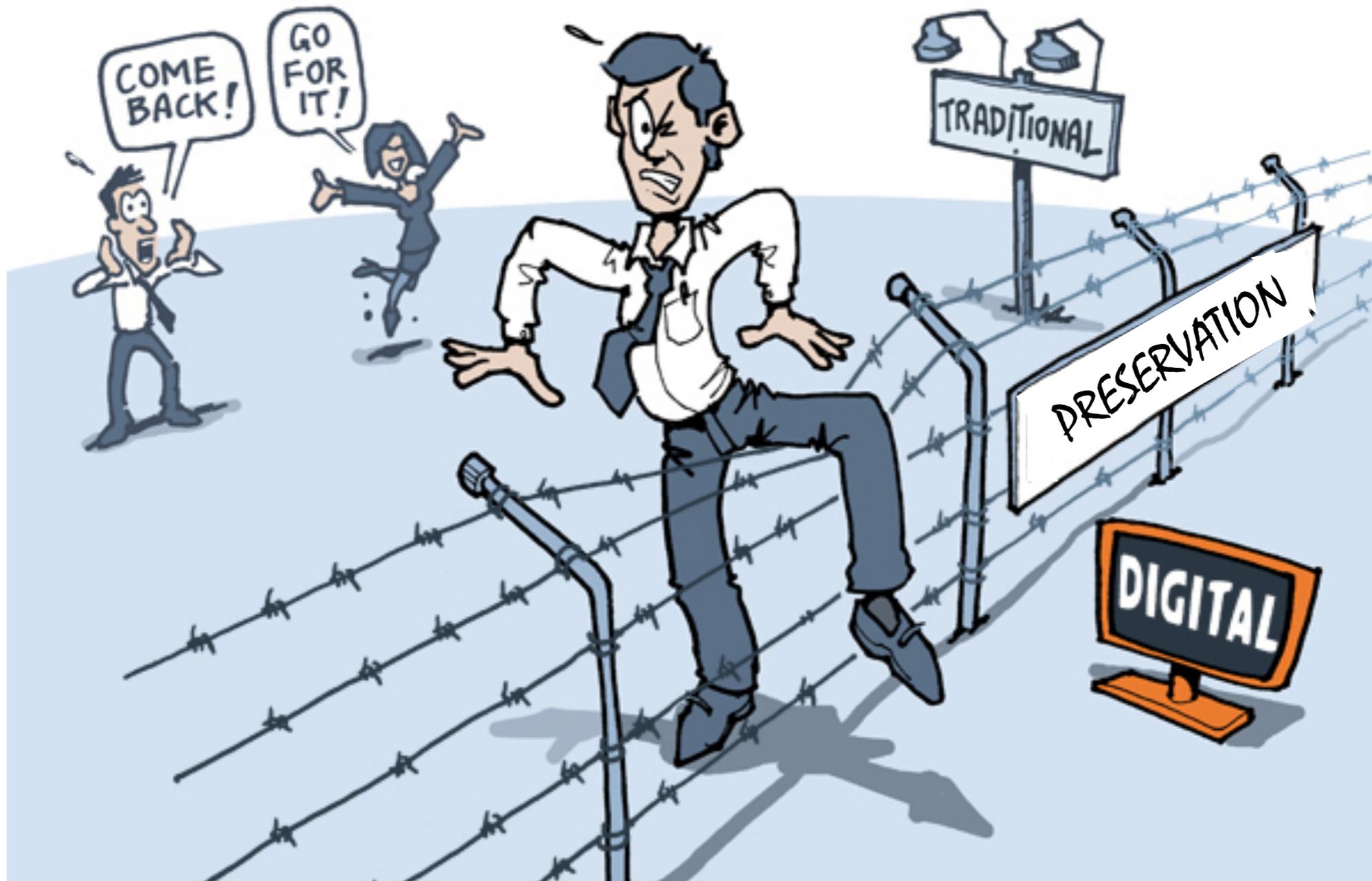
Coefficient de viscosité des manteaux inférieur et supérieur

Perte massive des données de recherche ces dernières 25 années !



80% des données en zoologie associées à des publications des années 90 sont définitivement **perdues**

Comment expliquer un tel paradoxe ?

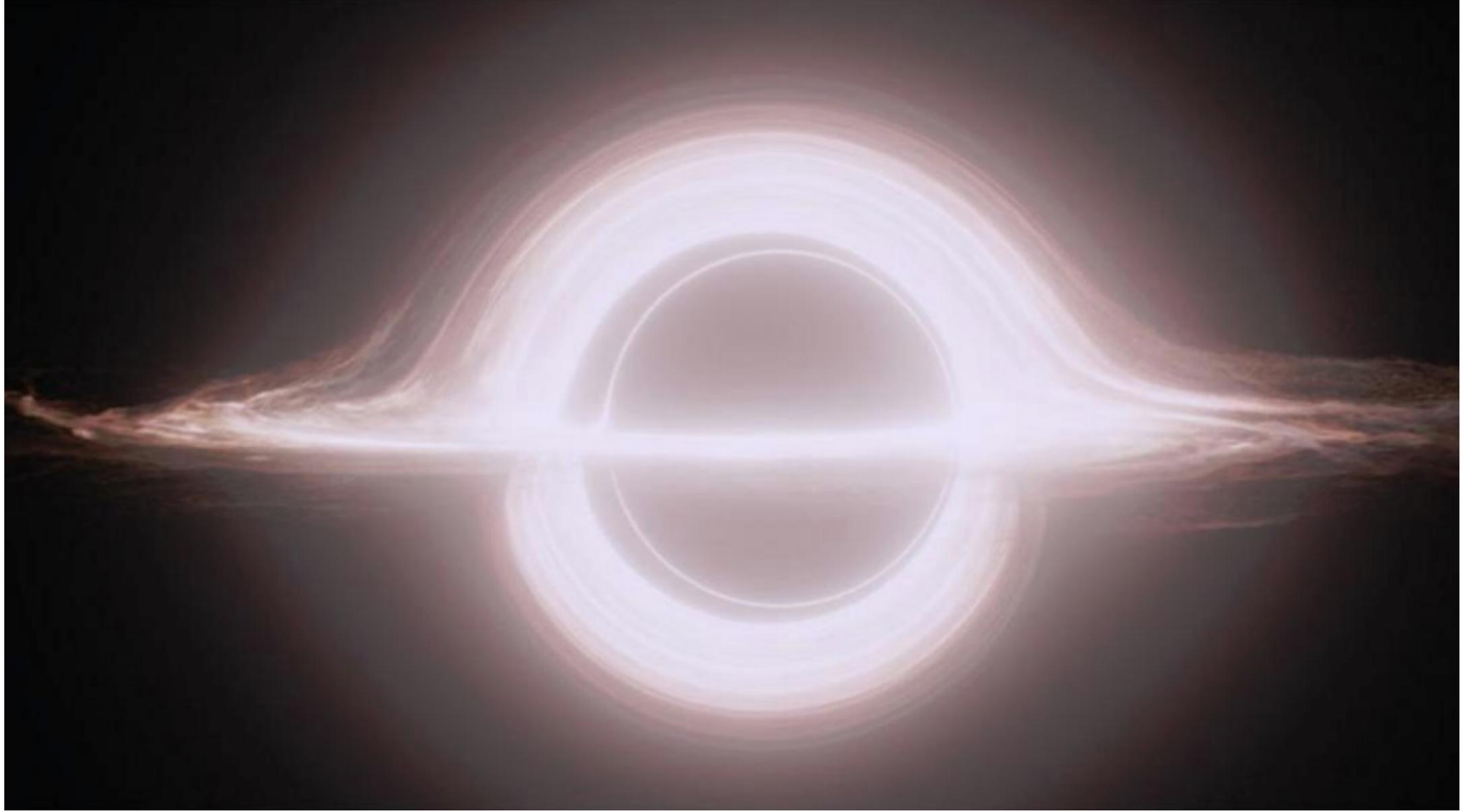


The transition from traditional to digital marketing – Roger Harvey 2014

Vers un “âge sombre” numérique?

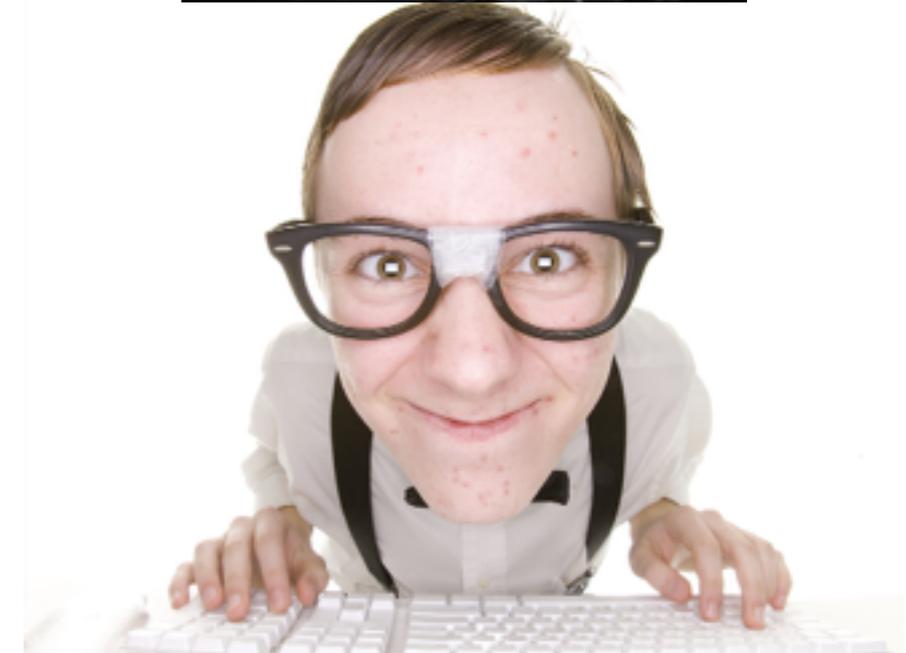
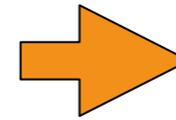


Vers un trou noir informationnel ?



A still image of Gargantua from the latest paper: Gravitational lensing by spinning black holes in astrophysics, and in the movie Interstellar (IOPScience). Image credit: Oliver James et al.

Qui est en charge de la préservation du patrimoine scientifique numérique de l'ULB?





L'équipe S&I compte 9 personnes



Benoît Pauwels
(directeur)



4 développeurs



1 agente d'aide à l'utilisateur



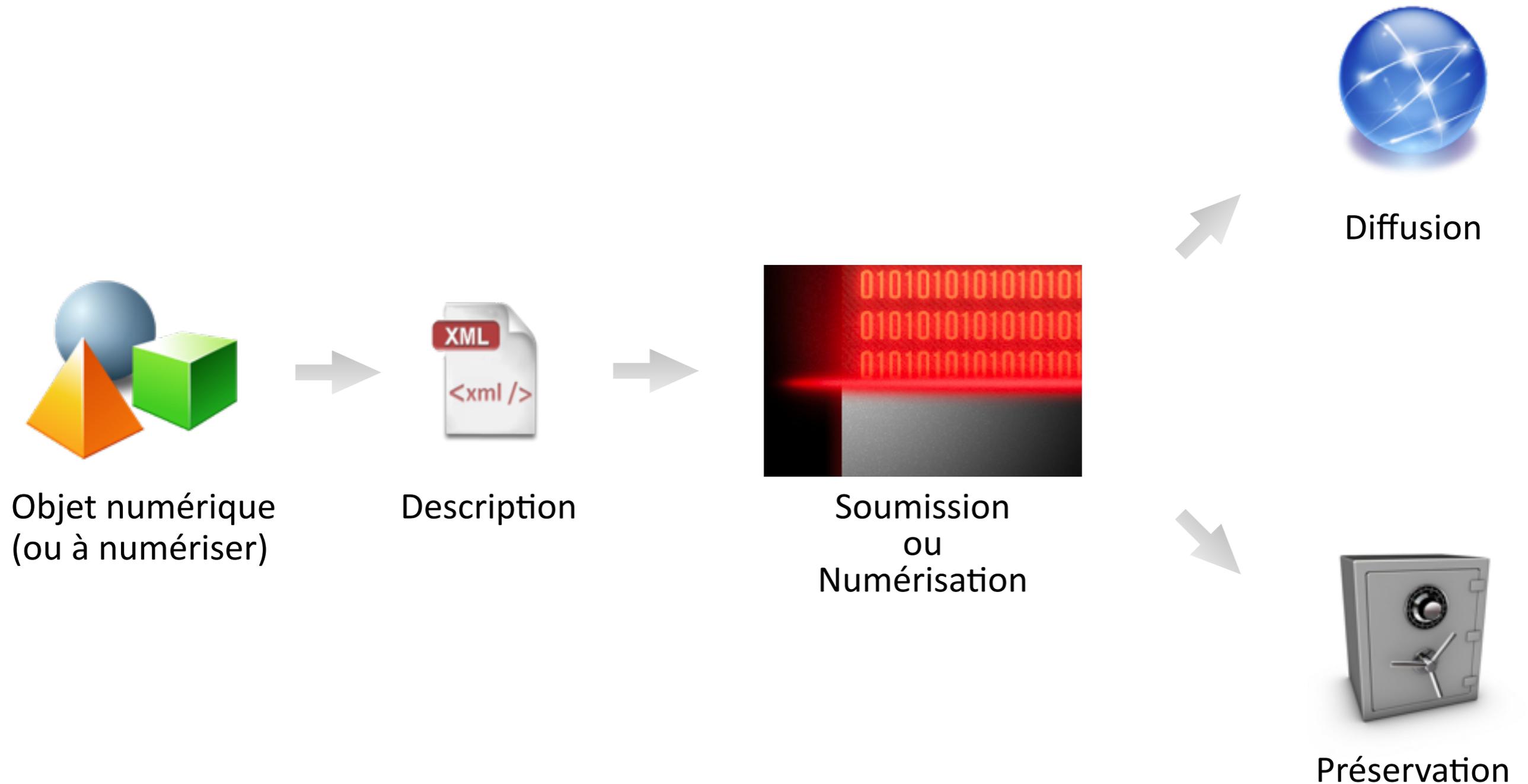
1 ingénieur système



2 agents de maintenance



Notre activité principale est de développer l'infrastructure logicielle permettant la description, la soumission, la diffusion et la préservation des objets numériques

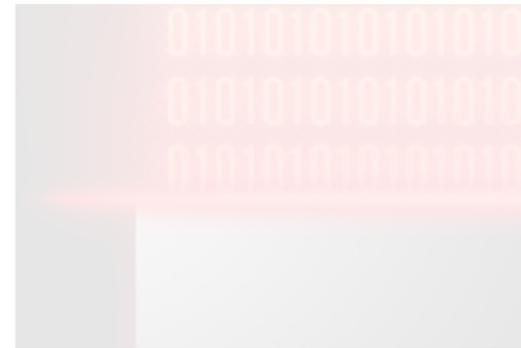




Notre activité principale est de développer l'infrastructure logicielle permettant la description, la soumission, la diffusion et la préservation **des objets numériques**



Description



Soumission
ou
Numérisation



Diffusion



Préservation



Quels types d'objets numériques?



**Natifs
numériques**



**Thèses de doctorat
numérisées**



**Tirés à part
numérisés**



**Données de la
recherche**



La production scientifique de l'ULB est stockée dans notre dépôt institutionnel

UNIVERSITÉ LIBRE DE BRUXELLES

Recherche avancée | Historique de recherche
Mon DI-fusion | À propos de DI-fusion | Contact

Passe-partout ▼
Recherche

51 documents ont été repérés pour la recherche "seth van hooland".
Voyez ci-dessous les numéros 1 à 20.

Tri Date de publication descendante

1 2 3 Suivant »

Going beyond Google Analytics: Extending the possibilities for the monitoring and reporting of usage data for cultural heritage collections
par Chardonens, Anne ; Coeckelbergs, Mathias ; Van Hooland, Seth ; [et al.]
Référence Proceedings of the 9th Qualitative and Quantitative Methods in Libraries International Conference (QQML2017)
Publication A Publier, 2017-03-12
Abstract de conférence [Accès en ligne](#)

Managing electronic records across organizational boundaries: the experience of the Belgian federal government in automating investigation processes
par Maroye, Laurence ; Van Hooland, Seth ; Aranguren Celorrio, Fiona ACF; Soyez, Sébastien; Losdyck, Bénédicte; Van Reck, Odile; De Terwangne, C.
Référence Records management journal, 27, 1
Publication Publié, 2017-02-07
Article révisé par les pairs [Détails](#)

Exploring archives with probabilistic models: Topic Modelling for the valorisation of digitised archives of the European Commission
par Hengchen, Simon ; Coeckelbergs, Mathias ; Van Hooland, Seth ; Verborgh, Ruben; Steiner, Thomas
Référence IEEE Proceedings of the First Workshop on Computational Archival Science
Publication Publié, 2016-12-02
Publication dans des actes [Accès en ligne](#)

Raffiner la recherche

Dépôt d'origine
ULB-DI (51)

Format
 Article révisé par les pairs (18)
 Publication dans des actes (9)
 Communication à un colloque (7)
 Abstract de conférence (4)
 Partie d'ouvrage collectif (3)
 plus...

Auteur
 Van Hooland, Seth (47)
 De Wilde, Max (11)
 Verborgh, Ruben (11)
 Coeckelbergs, Mathias (7)
 Hengchen, Simon (6)
 plus...

Date de publication
2016 (12)



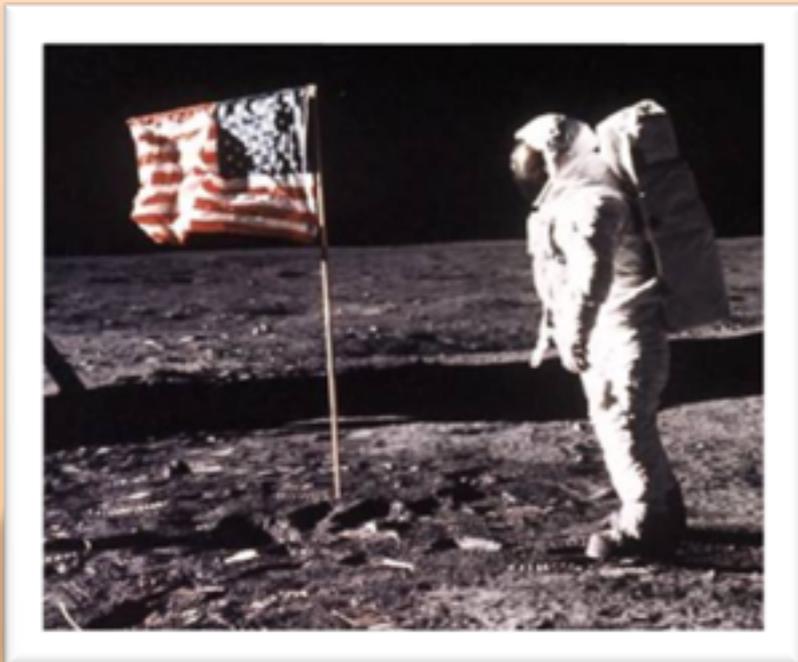
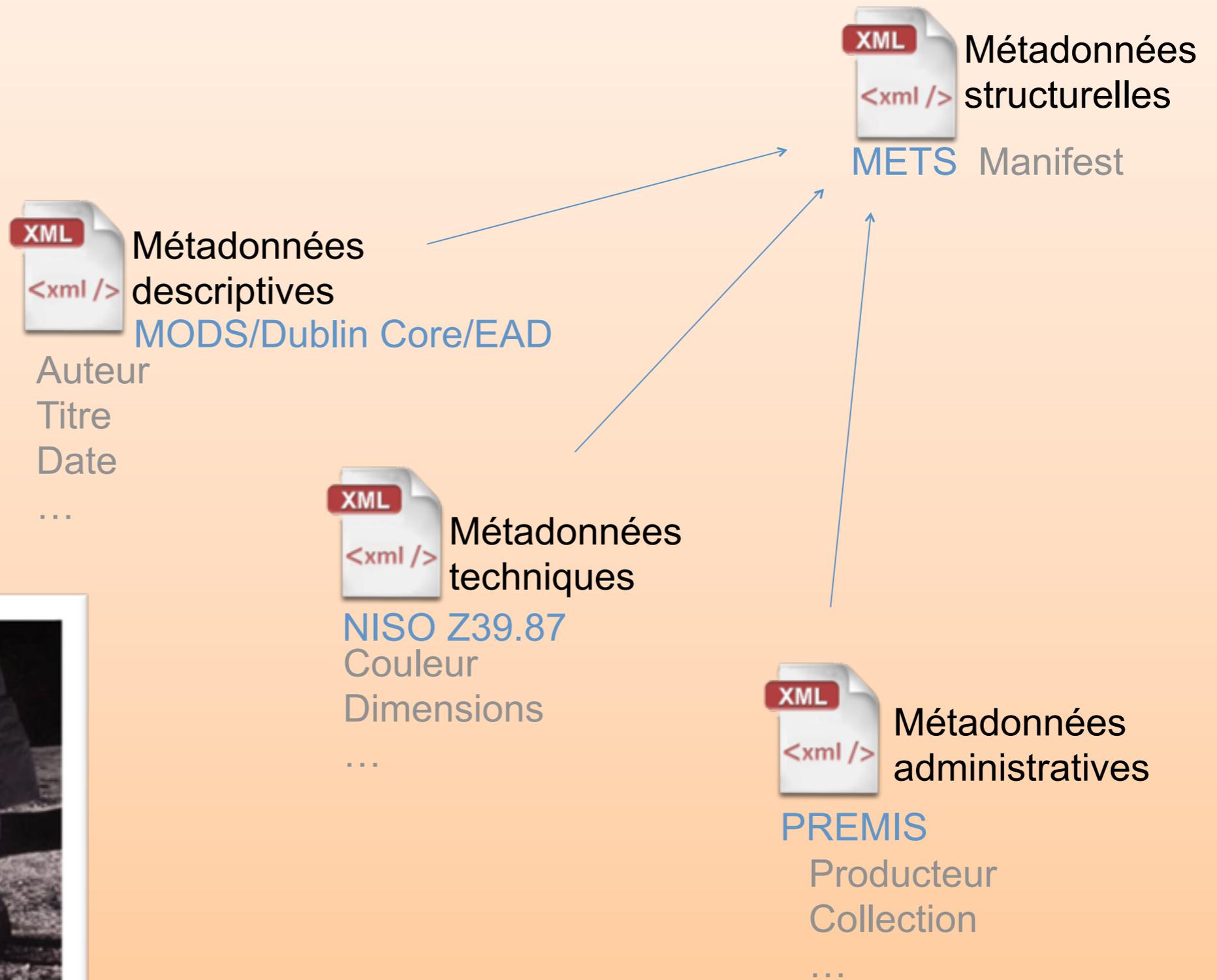
Notre activité principale est de développer l'infrastructure logicielle permettant la description, la soumission, la diffusion et la préservation des objets numériques



“Metadata is a love note for the future”, Jason Scott



Chaque objet numérique est décrit par un ensemble structuré de métadonnées





Ces métadonnées sont accessibles via DI-fusion

UNIVERSITÉ LIBRE DE BRUXELLES

[Recherche avancée](#) | [Historique de recherche](#) | [Mon DI-fusion](#) | [À propos de DI-fusion](#) | [Contact](#) |  

Passe-partout ▼ Recherche

[« Retourner aux résultats de recherche](#)  Citer

Contribution à l'étude du sérum chez les animaux vaccinés

par [Bordet, Jules](#) 
Publication Non publié, 1895
Thèse de doctorat

ACCÈS EN LIGNE DÉTAILS STATISTIQUES

Titre : Contribution à l'étude du sérum chez les animaux vaccinés
Auteur : Bordet, Jules
Pourvoyeur de diplôme : Université Libre de Bruxelles
Editeur : Université libre de Bruxelles, Faculté de Médecine - Médecine, Bruxelles
Date de défense : 1895
Sujet CREF : Médecine pathologie humaine
Volumes/pages : 1 v.
Intitulé du diplôme : Doctorat en sciences médicales
Note : info:eu-repo/semantics/nonPublished
Langue : Français
Identificateurs : ulbcat.ulb.ac.be:23317

Documents en relation

DI-fusion

The network collective: Rise and fall of a scientific paradigm
par Eichmann, Klaus K. , Bordet, Jules , Zinkernagel, Rolf Martin R.M.
Publication 2008

17



Notre activité principale est de développer l'infrastructure logicielle permettant la description, la **soumission**, la diffusion et la préservation des objets numériques



Objet numérique
(ou à numériser)



Description



Soumission
ou
Numérisation



Diffusion



Préservation



Le dépôt d'objets natifs numériques est réalisé à l'aide du portail de soumission de DI-fusion

DI-fusion
Le dépôt institutionnel

UNIVERSITÉ LIBRE DE BRUXELLES, UNIVERSITÉ D'EUROPE
ULB

My DI-fusion | Help

Book published as author : **blabla**

Type Describe Upload Check Finish

Help

Help

Uploading

File: No file chosen

URL:

Description:

Version:

Access rights:

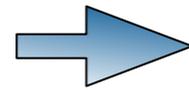
Embargo (YYYY-MM-DD):

Uploaded files

No file is currently attached to this item

©2009 DI-fusion | Archives & Bibliothèques de l'ULB | ULB | di-fusion@ulb.ac.be
Version: 3.1 (2013)

Pour les objets analogiques, on procède à leur numérisation à l'aide de scanners professionnels



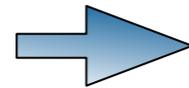
Scanner ADF

135 ppm RV A4 @ 300dpi

contrôle ultrasonique

A8 à A3

Bac: 500 pages



Scanner de livres

600 dpi

A2+

15cm épaisseur





Pourquoi numériser en interne et non faire appel à un prestataire externe?



- contrôle total de la production
- moins onéreux
- une fois en place, le workflow sera utilisé pour d'autres projets



- nécessite du personnel
- temps de démarrage plus long

Nous avons réalisé un vaste projet de numérisation rétrospective des thèses de l'ULB: +8.200 volumes, +1.8M pages





Notre workflow vise la tolérance zéro à l'erreur



Nommage de fichier automatique sur base des codes QR générés à partir du catalogue



scan recto-verso systématique



sauvegarde à la fois l'image brute et traitée



double scan avec détection d'avalage multiple : portrait et paysage



4 vérifications de comptage de pages



contrôle qualité assisté par ordinateur

Ce workflow nous permet de numériser 150 volumes de thèses par semaine en régime



~ 100 GB de données produites chaque jour



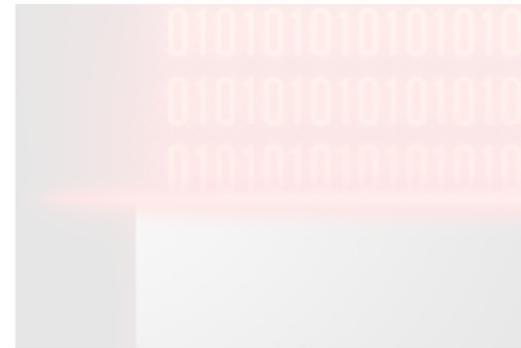
Notre activité principale est de développer l'infrastructure logicielle permettant la description, la soumission, la diffusion et la préservation des objets numériques



Objet numérique
(ou à numériser)



Description



Soumission
ou
Numérisation



Diffusion



Préservation

Les bibliothèques universitaires ont deux **missions** principales



Garantir l'accès aux objets
sélectionnés par les experts en
conservation



Préserver ces objets
notamment produits en interne

Garantir l'accès et préserver les objets physiques en bibliothèque... c'est assez simple.



Les bibliothèques universitaires ont deux **missions** principales



Garantir l'accès aux objets
sélectionnés par les experts en
conservation

Préserver ces objets
notamment produits en interne

A l'ère numérique, ces missions sont compromises en raison de:

- la perte de contrôle sur certains objets (accès par abonnement)
- la vulnérabilité des données numériques

Les bibliothèques universitaires ont deux **missions** principales



Garantir l'accès aux objets
sélectionnés par les experts en
conservation



Préserver ces objets
notamment produits en interne



Un oligopole et un marché juteux

5 éditeurs contrôlent plus de la moitié des publi. à facteur d'impact



**SPRINGER
NATURE**



Taylor & Francis
Taylor & Francis Group



6 éditeurs captent 65% des profits générés par l'édition scientifique



**SPRINGER
NATURE**



Taylor & Francis
Taylor & Francis Group

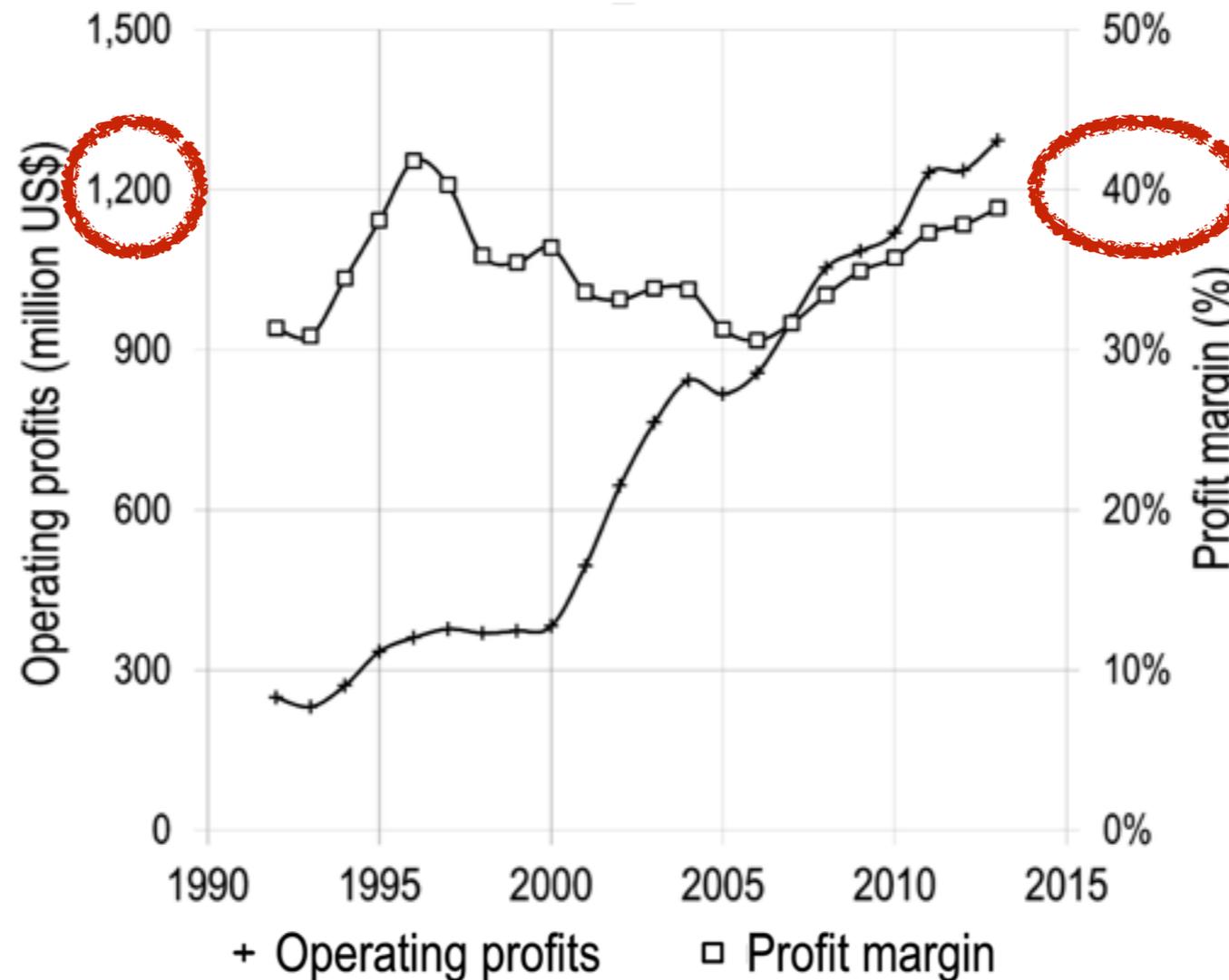




Des profits vertigineux en progression constante



Elsevier - Science, Technique et Médecine



Profit opérationnel (2016) : 2.58 Mrd €

Marge bénéficiaire sur STM supérieure à Apple, Google ou Amazon !



Comment de tels profits sont possibles ? Un business model de rêve !

L'auteur de l'article
n'est **pas rémunéré**
par l'éditeur et lui
cède ses droits



Argent public



Argent public



Les **réviseurs** des
articles ne sont
pas rémunérés
par l'éditeur

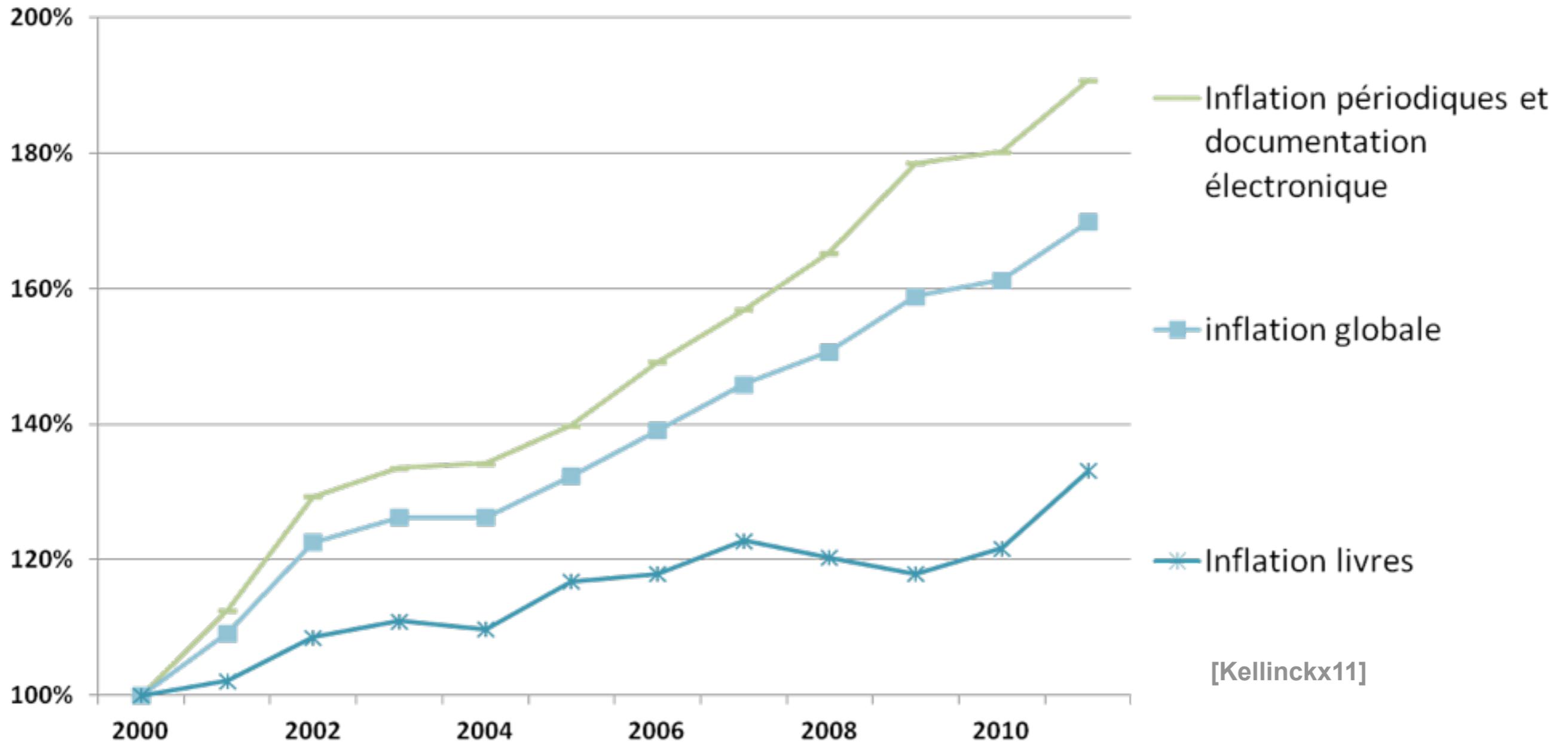


Argent public



L'**accès** aux articles
est **vendu** à prix d'or
aux bibliothèques

La décision de déléguer il y a 20 ans la publication des périodiques électroniques à des sociétés commerciales a d'importantes conséquences économiques aujourd'hui

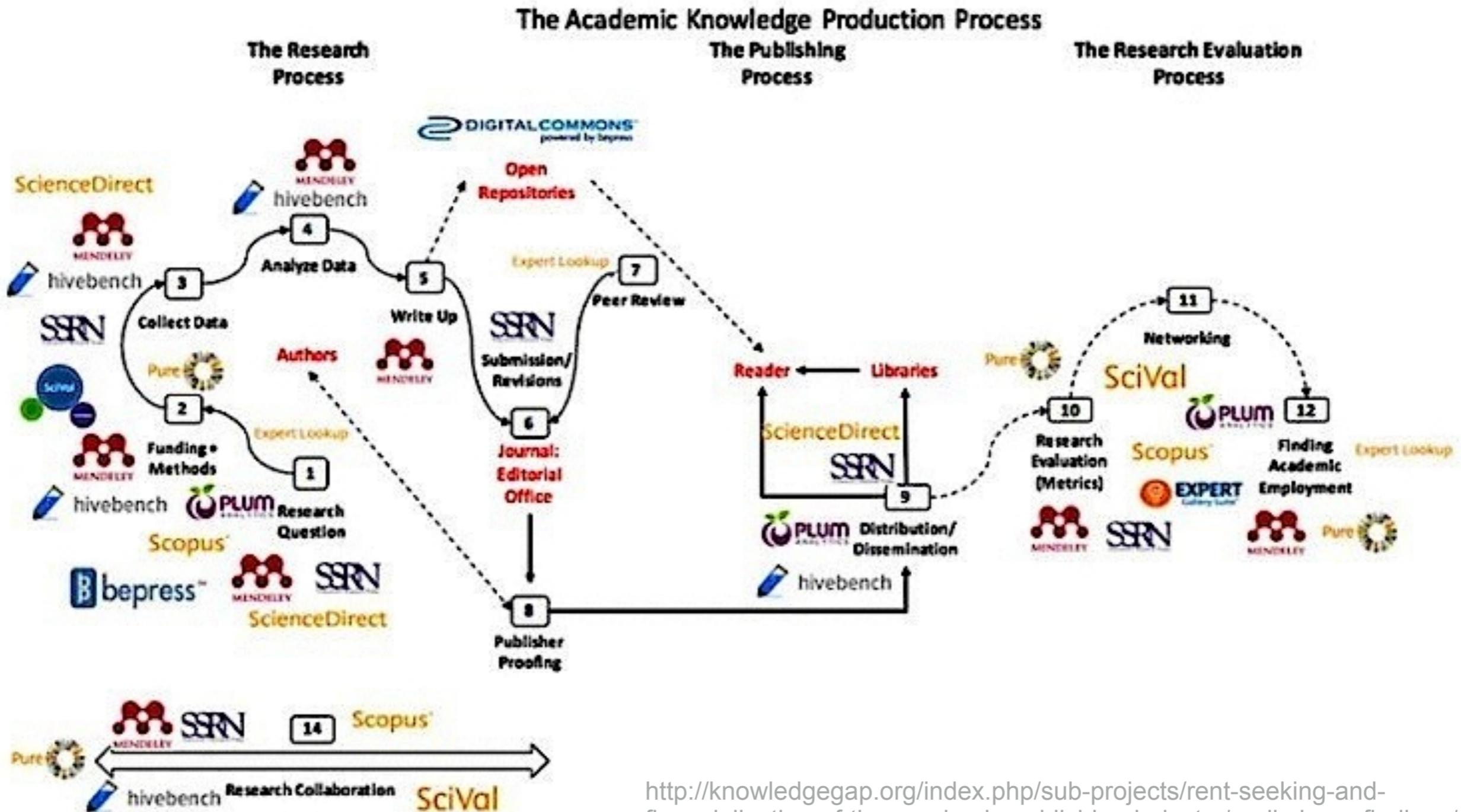


[Kellinckx11]

Inflation de 90% du budget documentaire de l'ULB en 11 ans



Vers un monopole ? Le modèle d'affaire d'Elsevier s'étend peu à peu à tout le processus de production de connaissance scientifique



<http://knowledgegap.org/index.php/sub-projects/rent-seeking-and-financialization-of-the-academic-publishing-industry/preliminary-findings/>



Au moins, la préservation des publications est assurée !?



ELSEVIER

Trois solutions d'accès post-abonnement existent



+70 éditeurs



+7400 titres

dark archive

\$450-\$15000

~\$2000

+140 éditeurs



+12500 titres

dark archive

\$1545-\$24720

~\$10000

+500 éditeurs



+9000 titres

light archive

\$2240-\$11180

~\$10000



Le problème de préservation des bibliothèques est résolu...



Les bibliothèques universitaires ont deux **missions** principales

Garantir l'accès aux objets sélectionnés par les experts en conservation

Préserver ces objets
notamment produits en interne

Reste donc un problème à résoudre:

Comment préserver nos ressources électroniques propres?





Qu'entend-on par préservation du patrimoine scientifique numérique ?

Stratégie et procédés de protection contre les menaces pesant sur les objets numériques d'intérêt dans le but de permettre leur **(ré)exploitation** dans le futur à **très long terme**.

A ne pas confondre avec :

- **le stockage** : enregistrement de données sur un support physique
- **la sauvegarde** : réplique des données dans le but de les restaurer en cas de perte
- **l'archivage au sens informatique** : stockage des données peu fréquemment utilisées sur des médias moins coûteux (typiquement bande magnétique)



A l'ère de l'externalisation... déléguer notre mission principale ?

Preservica[®]
Digital Preservation



SAAS

solutions
tiers-archivageur

IAAS

stockage en
cloud



Solutions clé-sur-porte des tiers-archiveurs ou de l'infrastructure de stockage: « pour N kEUR/an, nous nous chargeons de tout... »

Preservica[®]
Digital Preservation
(cloud edition)



Problèmes techniques



Problèmes juridiques
et économiques

Economique:

- Verouillage et dépendance envers le fournisseur de service ?

Juridiques:

- Quid en cas de faillite du fournisseur?
- Quid en cas de perte de données?

Techniques:

- Aucun contrôle sur les méthodes de préservation
- Migration vers un autre fournisseur?



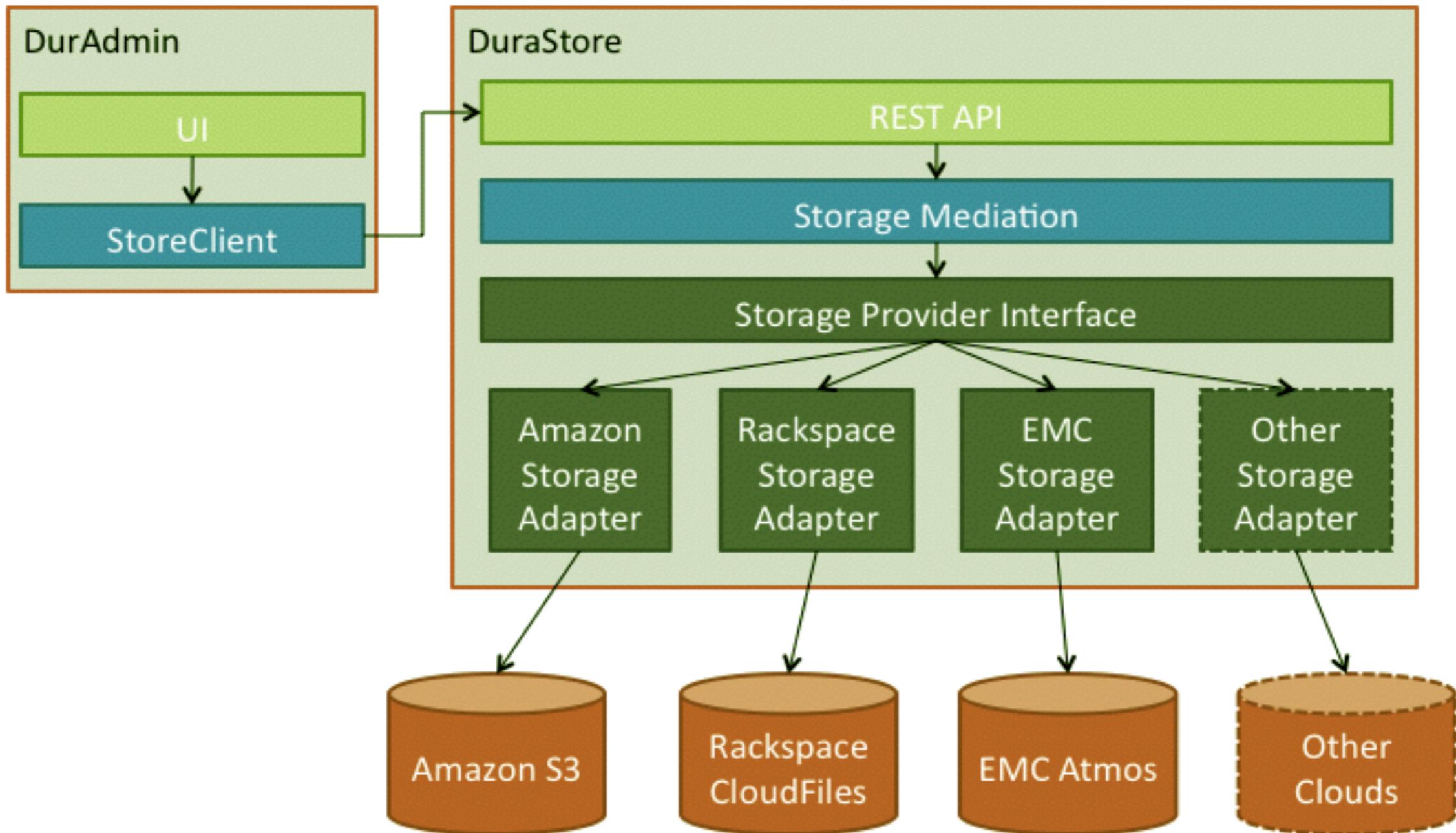
Stockage en Cloud: un autre oligopole ...



IaaS/PaaS market (Source: Jefferies)

Cloud vendor	Annualized revenue	% of market	Year-over-year growth
Amazon Web Services	\$18.34 billion	51%	42%
Microsoft Azure	\$6.17 billion	17%	89%
IBM Cloud	\$4.03 billion	11%	22%
Google Cloud Platform	\$2.05 billion	6%	125%
Alibaba Cloud	\$1.79 billion	5%	92%
Salesforce	\$1.78 billion	5%	31%
Oracle Cloud	\$1.59 billion	4%	20%
Subtotal	\$35.75 billion	86%	54%
Total Gartner estimate	\$41.79 billion	100%	33%

Duracloud, intermédiaire entre le client et le fournisseur de stockage virtualisé en ligne





La tendance en informatique est à l'externalisation (outsourcing)

En déléguant l'une de nos deux missions principales, nous deviendrons des **courtiers** en préservation et non plus des **conservateurs**.

Les universités comptent parmi les rares institutions multi-centenaires et sont donc à même de garantir la préservation à très long terme.

Comment rester **acteurs** de la préservation de nos objets numériques et non des **clients passifs** de services de préservation à durée de vie (probablement) limitée?



Trois piliers de la préservation numérique: technologie, organisation, ressources



technologie

organisation

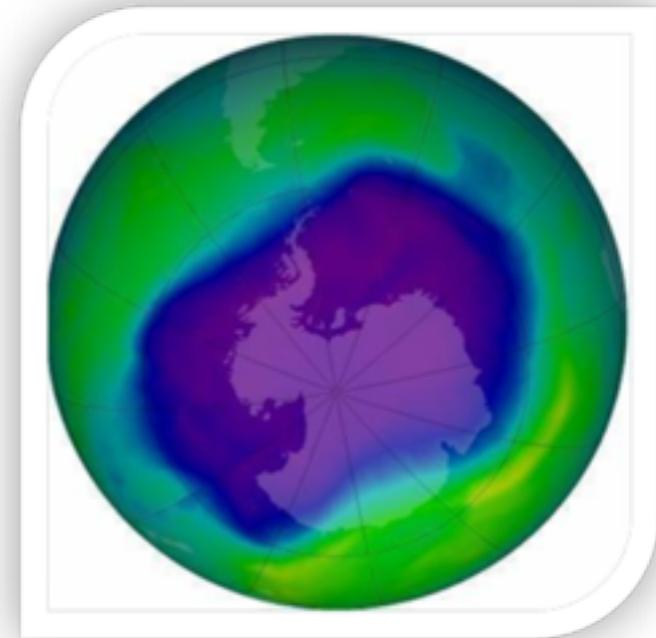
ressources
humaines et
financières

On distingue deux **niveaux** de préservation:



Préservation
des données **brutes**

Préservation
des données **logiques**



On distingue deux **niveaux** de préservation:



Préservation
des données **brutes**



Préservation
des données **logiques**





Quelles menaces mettent nos ressources numériques en péril?

7-5-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick



PROBLEME

SOLUTION



Dilbert, Scott Adams

1. Risque de catastrophes



<https://fic.kr/p/9MsNf9>

Risque difficile à évaluer... évènements à faible probabilité mais à impact majeur



It happens...



GOOGLE

Google suffers data loss as data center gets hit by lightning 4 times

BY NATALIE SHOEMAKER 08.21.2015 :: 9:30AM EST

Fire Causes Partial UniSuper Data Center Outage in Port Melbourne

As Irma Heads for Florida, One Miami Data Center is Especially Critical

While an outage is far from given, its effects would ripple far beyond cities in the hurricane's path.

Michigan County Offline After Data Center Fire

IT ser
after
count
west
have

Rich M

Toronto Flooding KOs Data Center Cooling Systems

A massive rainstorm caused widespread flooding and power outages Monday night in Toronto, which created challenges for some tenants at the city's largest data center hub. 151 Front Street maintained power, but experienced problems with cooling systems.

Kaseya, ConnectWise Pick Up Pieces After Irma Nails Florida HQs

Both companies supporting customers, but operations not expected to return to normal for days

Aldrin Brown | Sep 13, 2017

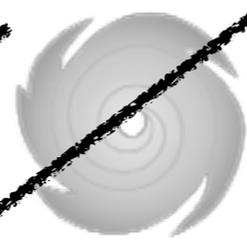
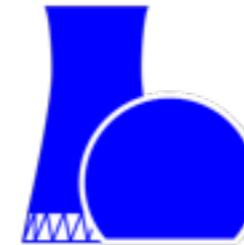
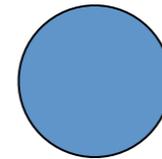
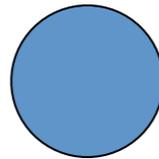
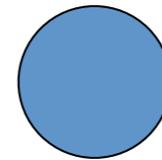
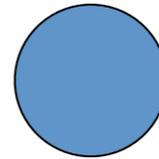
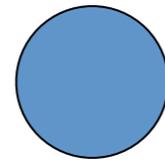


1. Risque de catastrophes

→ Redondance géographique

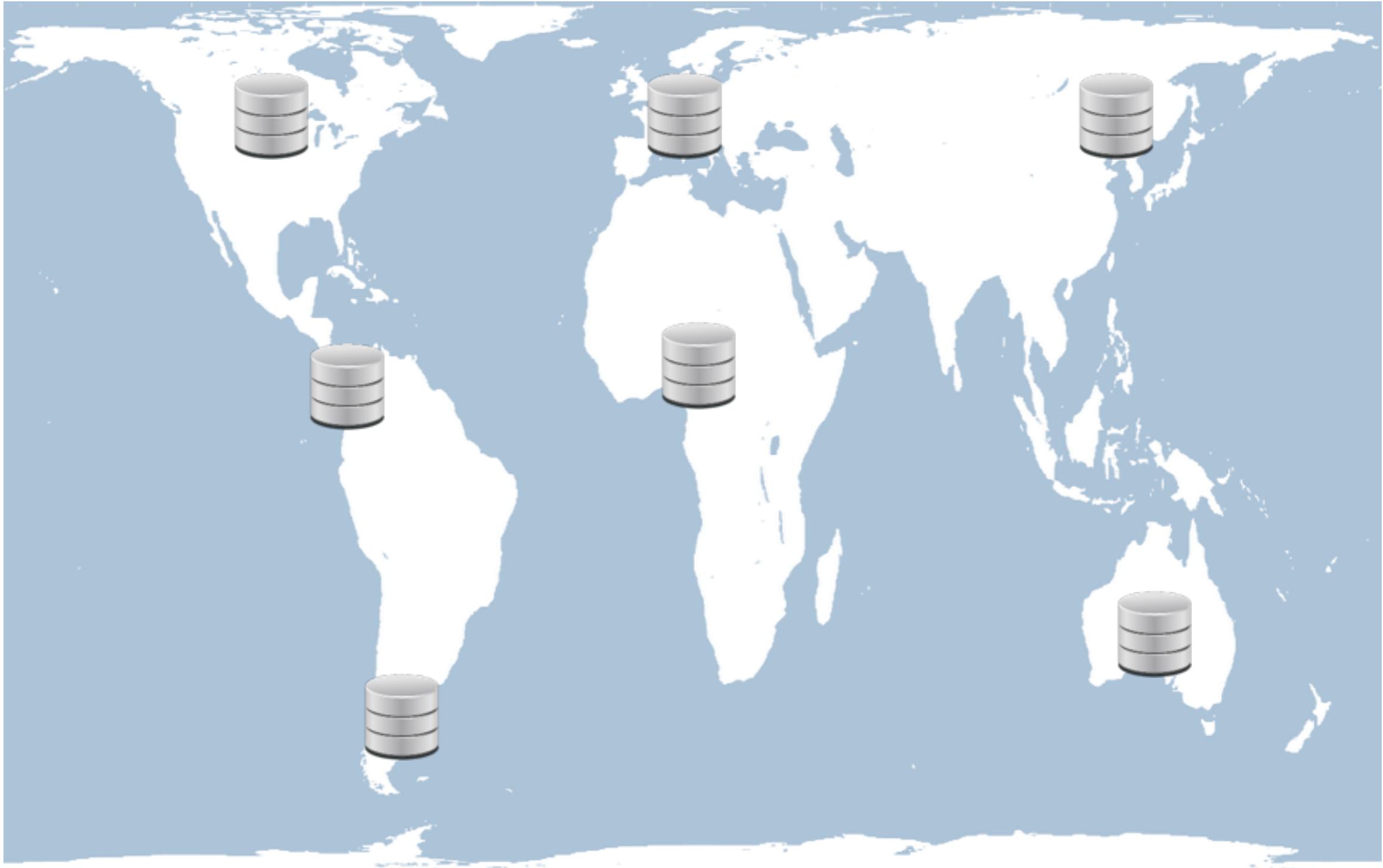


100km





Solution idéale : la géoréplication

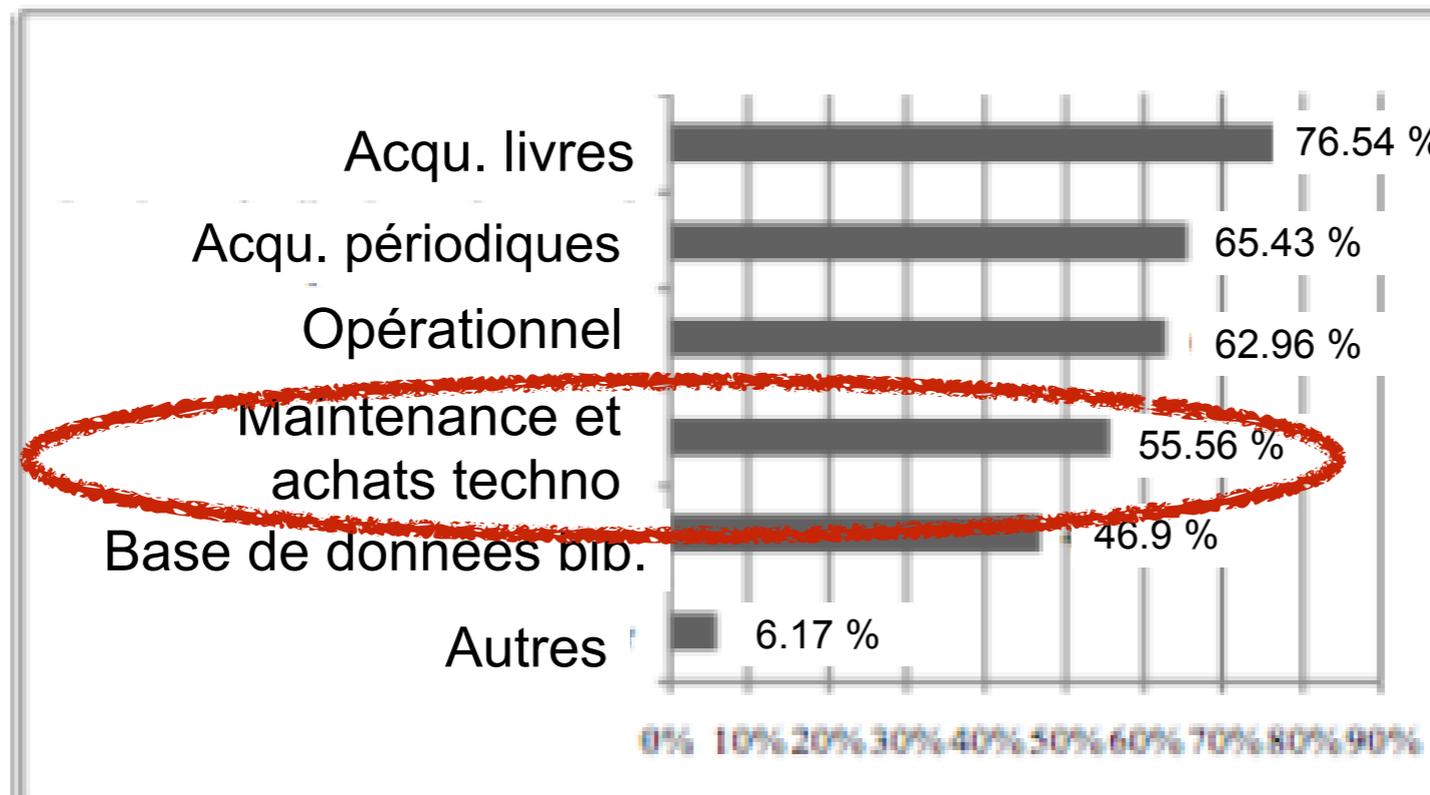


2. Risque économique



En cas de crise économique, les **restrictions budgétaires** concernent en priorité les activités ayant peu d'impact à très court terme

Restrictions budgétaires



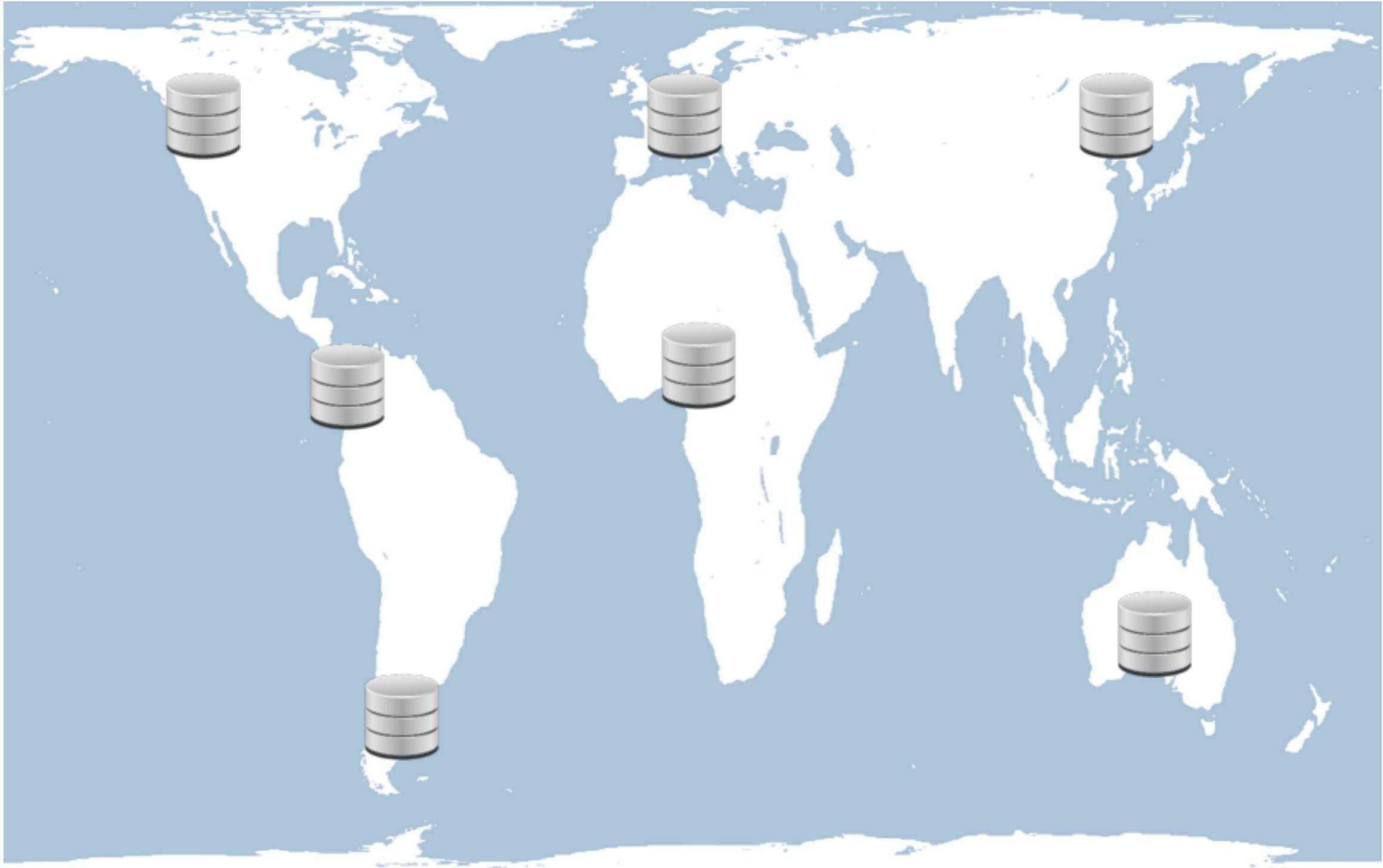
G. Giannakopoulos et al., Libraries in Crisis: A Glimpse over Greece and Cyprus, Procedia, Vol.147, 2014, <https://doi.org/10.1016/j.sbspro.2014.07.121>.

Si on n'investit plus dans la maintenance de l'infrastructure pendant 5 ans, le risque de perte croît considérablement.

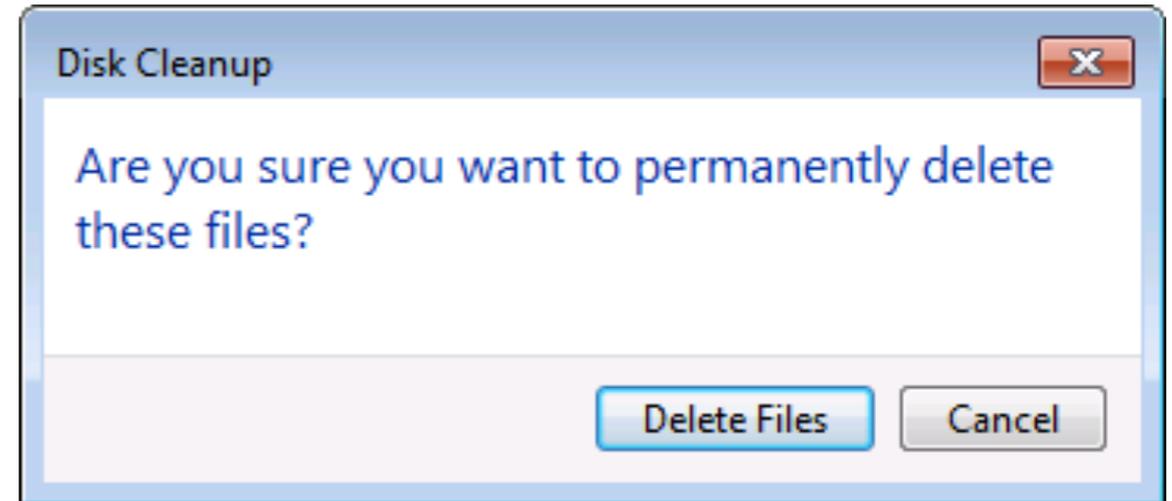


2. Risque économique

→ **Maîtrise des coûts**



3. Erreurs humaines





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

3. Erreurs humaines

→ Gestion technique individuelle





4. Risque organisationnel, politique, idéologique





Le patrimoine scientifique n'est pas épargné

Vanishing Canada: Why we're all losers in Ottawa's war on data

Records deleted, burned, tossed in Dumpsters. A Maclean's investigation on the crisis in government data

Anne Kingston

September 18, 2015

Sections

The Washington Post

Energy and Environment

Scientists are frantically copying U.S. climate data, fearing it might vanish under Trump

By Brady Dennis December 13, 2016

Most Read

WILDLIFE WATCH

U.S. Animal Abuse Records Deleted —What We Stand to Lose

By hiding online records of welfare violations, U.S. agency robs journalists, investigators, and the public of timely information—and takes pressure off abusers.



BUSINESS
INSIDER
UK

POLITICS

The Trump administration has told the EPA to remove its climate-change data from its website



Rafi Letzter

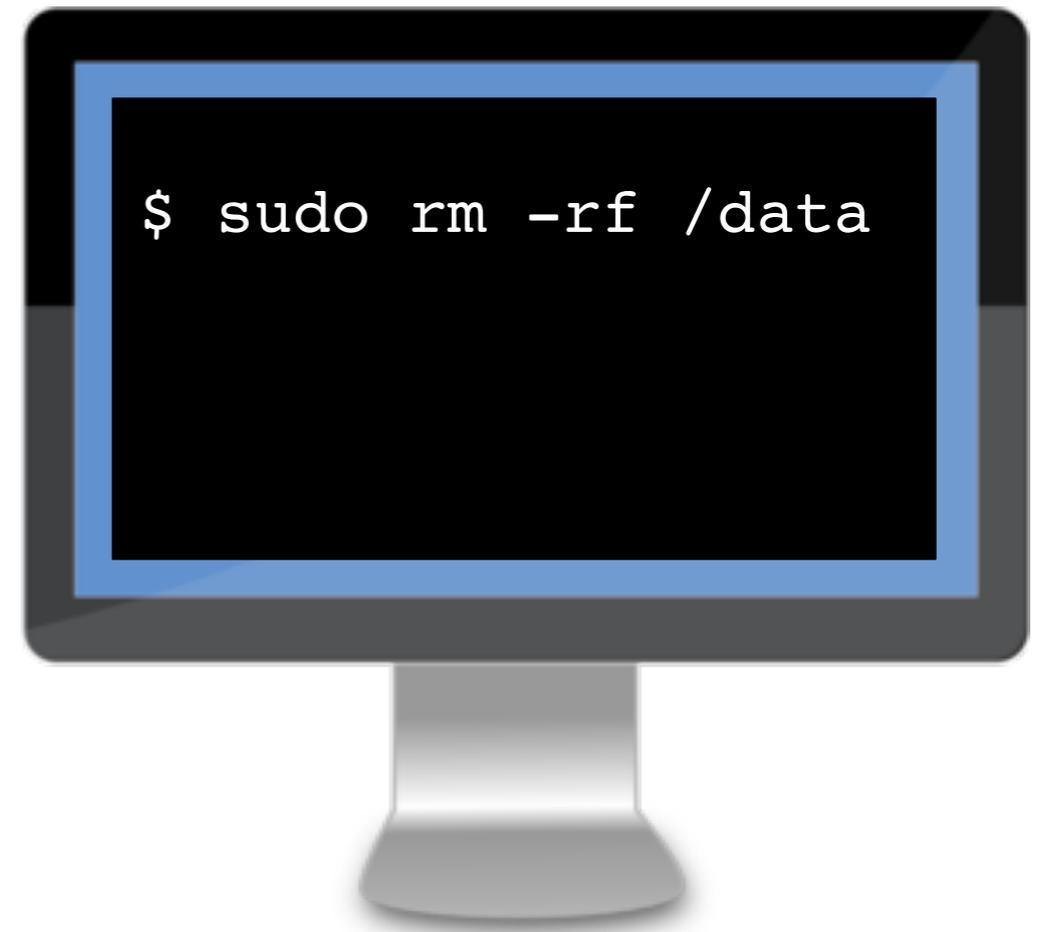
Jan. 25, 2017, 3:55 PM 1,781



Un autodafé numérique serait beaucoup plus efficace que son équivalent analogique



Au XV ème siècle



Aujourd'hui

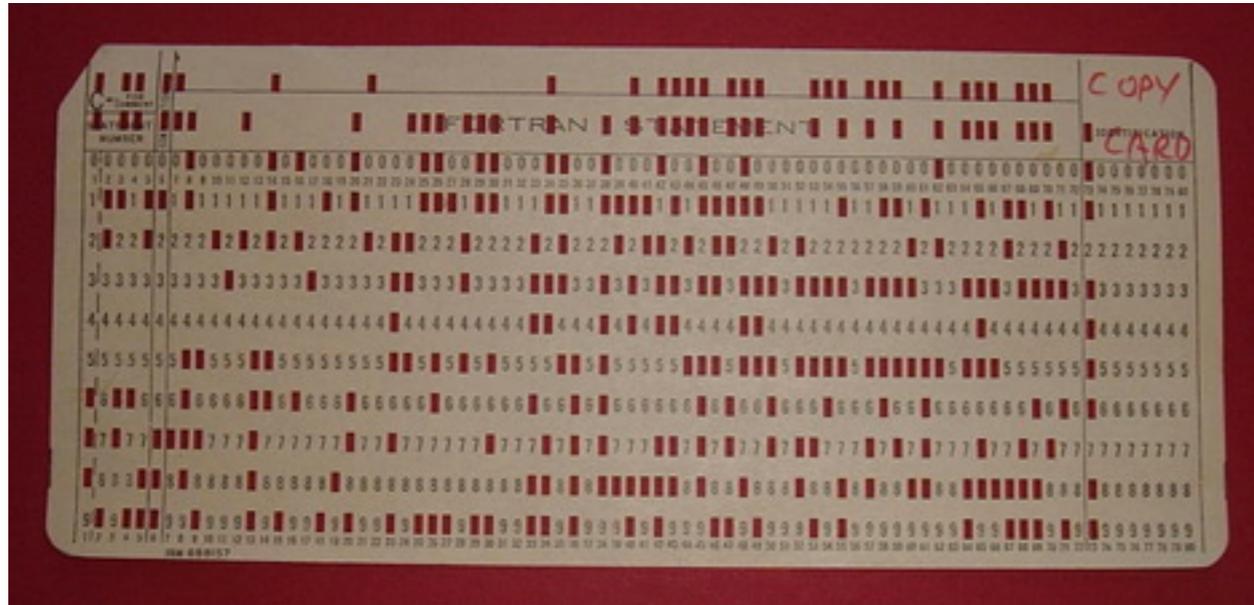


4. Risque organisationnel, politique...

→ **Indépendance administrative**



5. Obsolescence des supports



Rester bloqué dans un choix technologique coûte cher à moyen terme



Evolution des disques durs



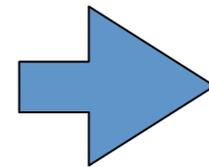
IBM 305 RAMAC (1956)

eq. 27.882 USD en 2017

5 Mo

~2000 h MTBF

910 kg



WD Gold (2017)

400 USD

12 To

2.5 M h MTBF

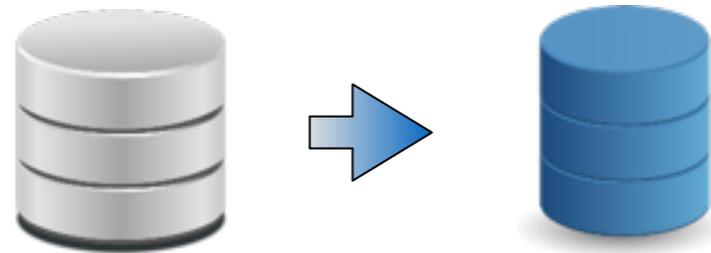
660 g



Nécessité de migrer vers le support le plus avantageux



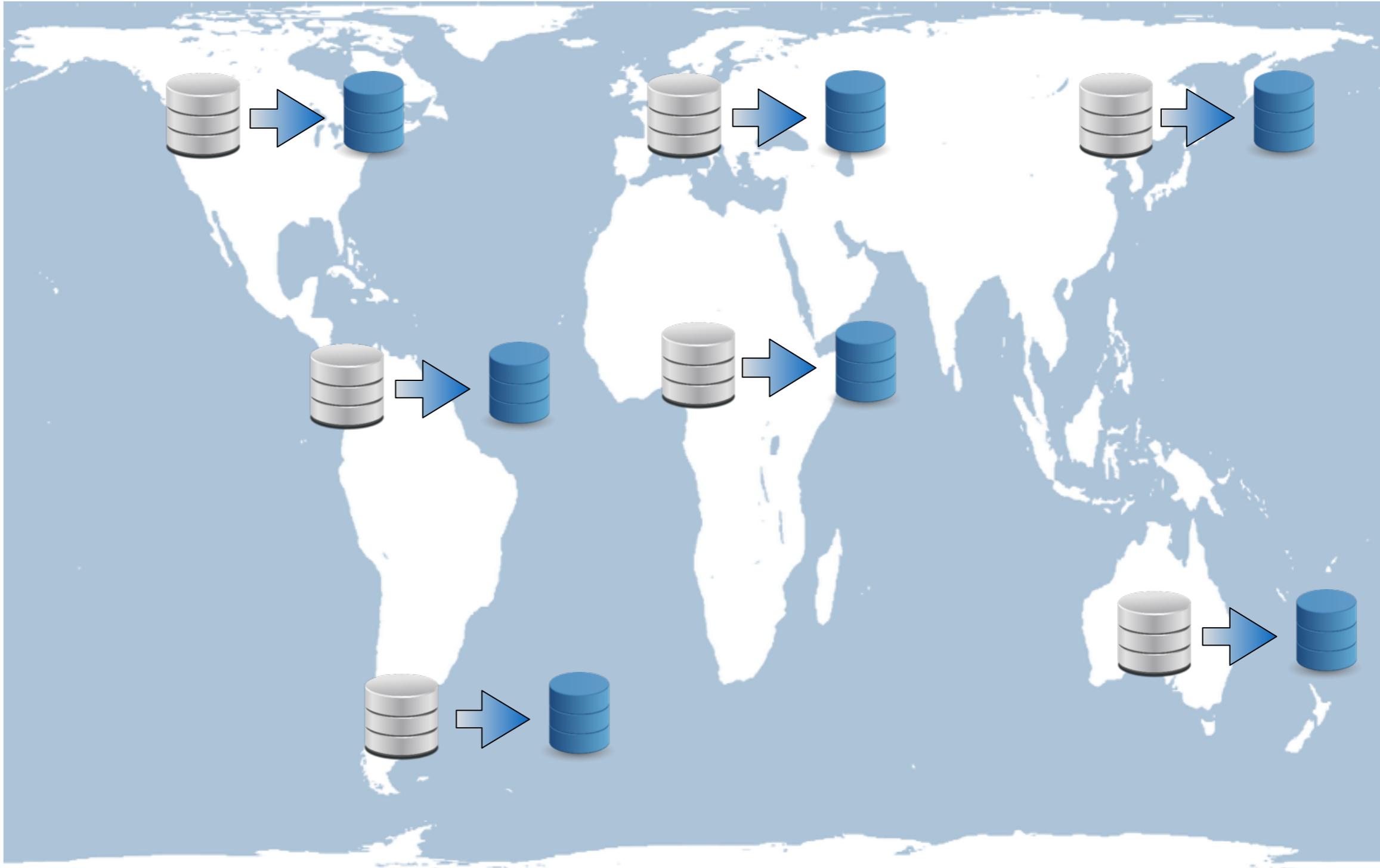
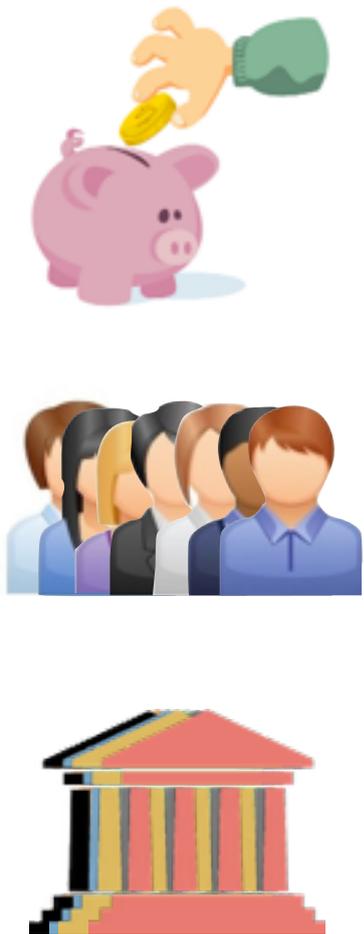
Migration régulière vers des supports de stockage récent de meilleur rapport capacité/fiabilité/prix



tous les 3 à 5 ans



5. Obsolescence des supports → **Migration des supports**

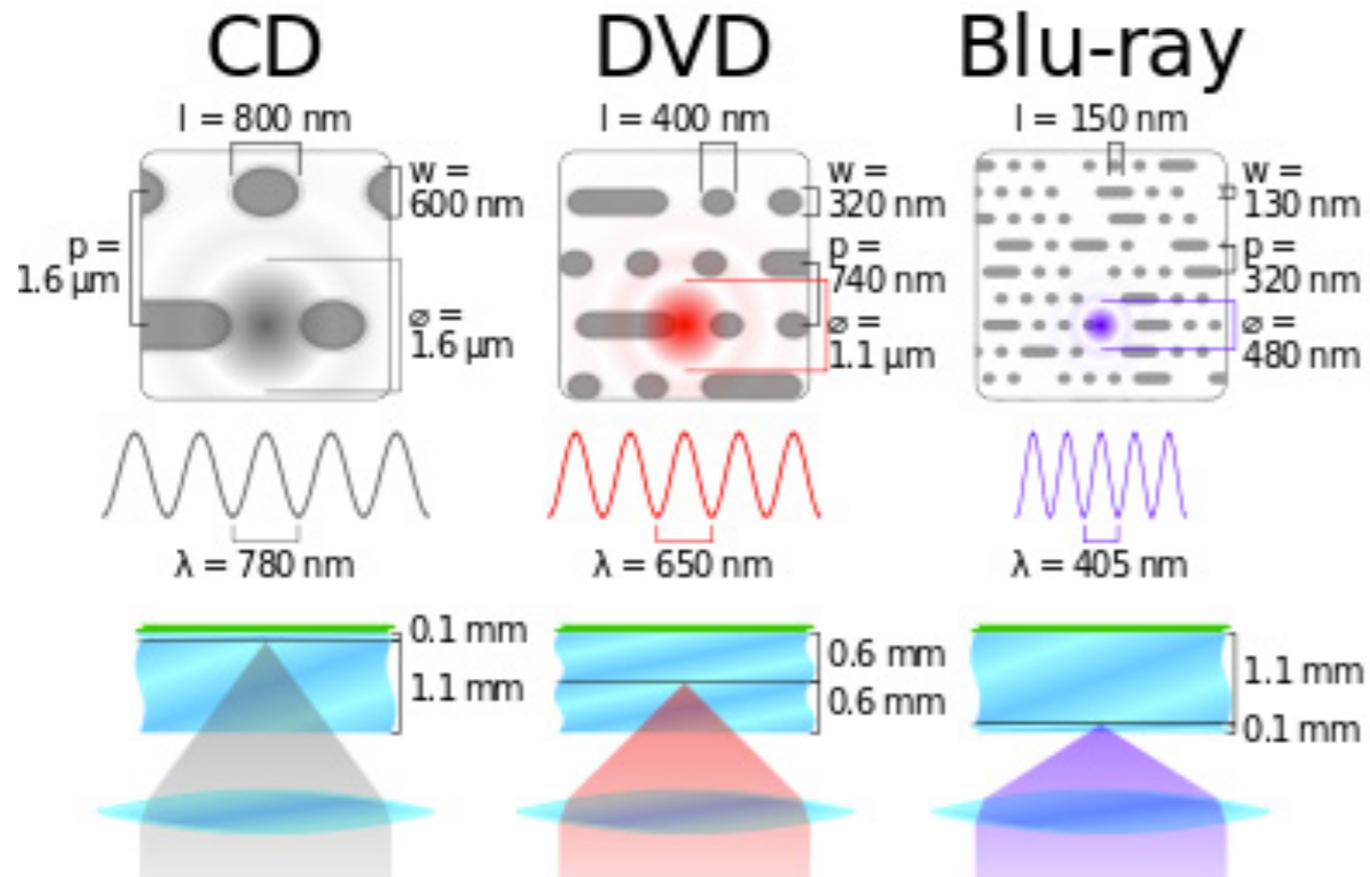




6. Défaillances matérielles



Durées de vie limitées des supports optiques de 5... à 200 ans

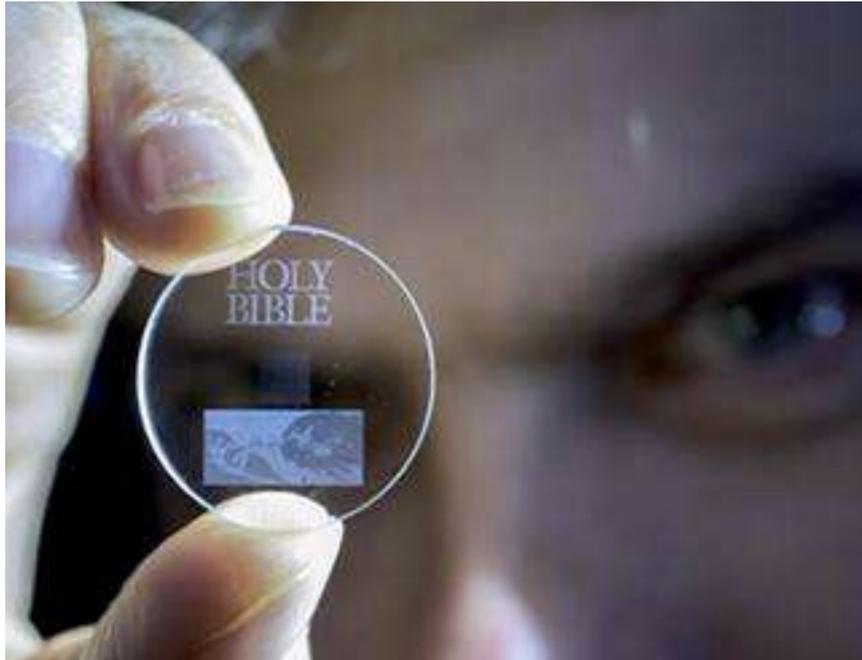


Problème :

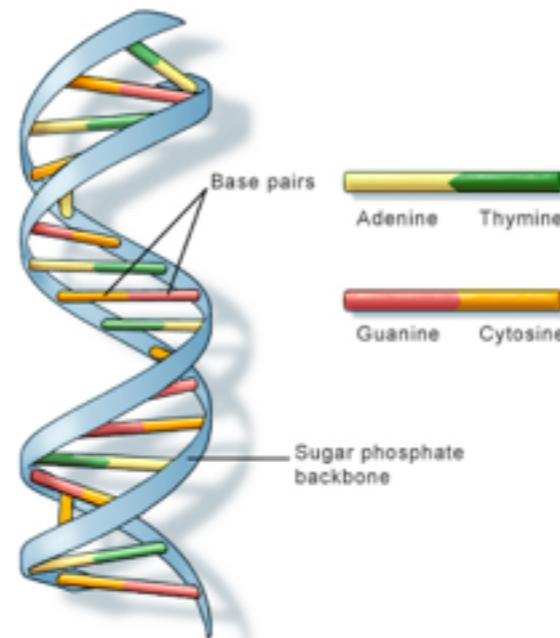
- faible capacité par rapport aux besoins
- manipulation fastidieuse pour des capacités de données importantes
- monitoring de l'intégrité nécessaire mais pas évident (consulter l'état à distance)



Les “super media” ...

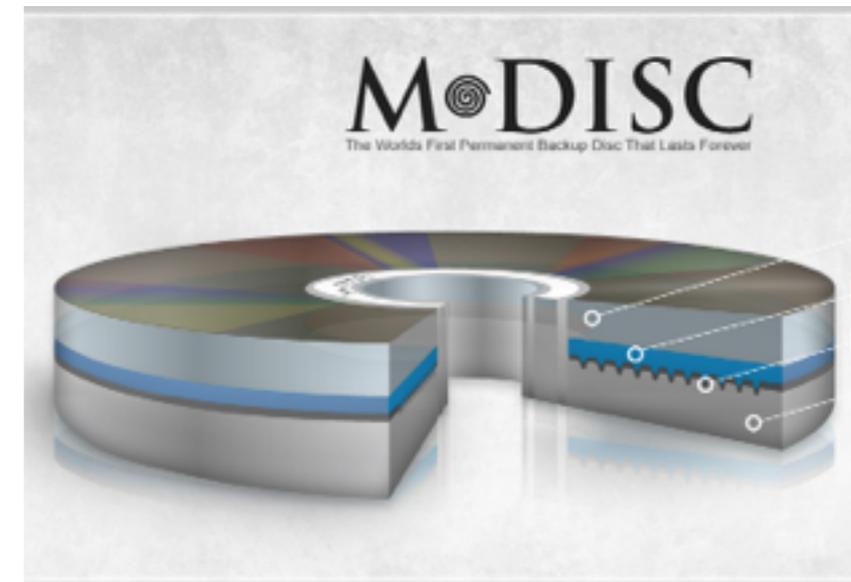


Stockage “Eternal 5D”
(>13 Mds années)



U.S. National Library of Medicine

Stockage ADN
7.5 M\$ / GB



M disc (1000 ans)

Problème (pour le moment) :

- viabilité économique à démontrer
- monitoring de l'intégrité pas évident

Pannes de disque dur

Hard Drive Annualized Failure Rates for Q4 2016

Reporting period 10/1/2016 - 12/31/2016 inclusive

MFG	Model	Drive Size	Drive Count	Avg. Age (months)	Drive Days	Drive Failures	Failure Rate
HGST	HDS723030ALA640	3 TB	978	61.21	90,415	9	3.63%
HGST	HDS5C3030ALA630	3 TB	4,476	55.87	412,752	13	1.15%
HGST	HDS5C4040ALE630	4 TB	2,625	45.35	241,665	4	0.60%
Toshiba	DT01ACA300	3 TB	46	44.12	4,232	-	0.00%
Seagate	ST4000DX000	4 TB	184	38.54	17,354	7	14.72%
WDC	WD30EFRX	3 TB	1,105	30.39	100,259	9	3.28%
HGST	HMS5C4040ALE640	4 TB	7,014	29.48	648,393	9	0.51%
WDC	WD60EFRX	6 TB	446	24.14	41,304	5	4.42%
HGST	HUH728080ALE600	8 TB	45	22.99	4,140	-	0.00%
Toshiba	MD04ABA500V	5 TB	45	22.15	4,140	-	0.00%
Seagate	ST4000DM000	4 TB	34,738	21.73	3,196,552	234	2.67%
Seagate	ST6000DX000	6 TB	1,889	21.48	173,720	8	1.68%
Toshiba	MD04ABA400V	4 TB	146	20.61	13,432	-	0.00%
WDC	WD40EFRX	4 TB	75	17.16	4,232	-	0.00%
HGST	HMS5C4040BLE640	4 TB	9,407	15.51	809,119	14	0.63%
Seagate	ST8000DM002	8 TB	8,660	4.72	663,697	30	1.65%
Seagate	ST8000NM0055	8 TB	60	1.44	1,560	-	0.00%
Totals			71,939		6,426,966	342	1.94%



Pour le moment, les supports magnétiques restent les plus avantageux



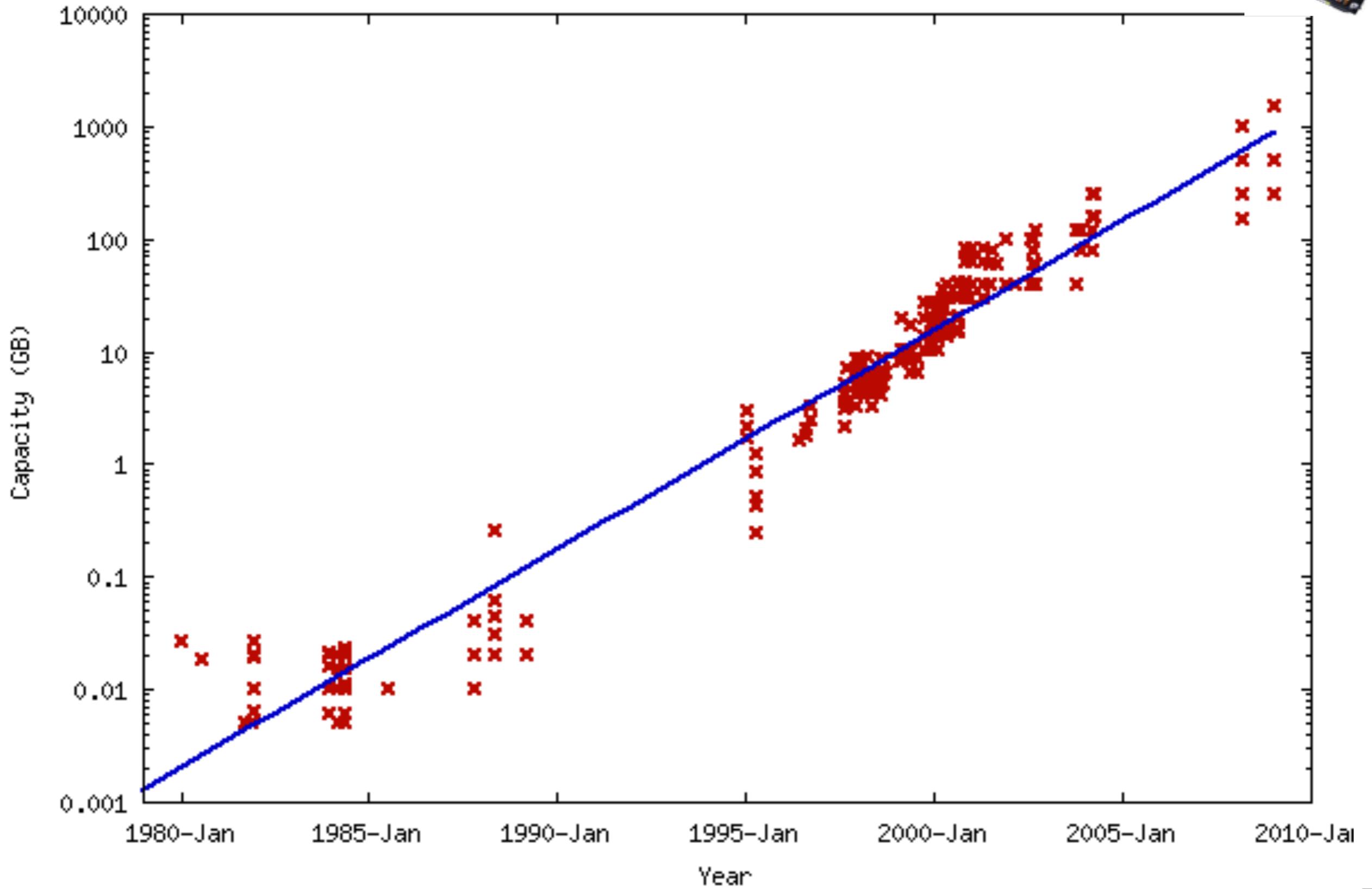
- faible coût à l'achat
- faible coût à maintenir
- fiable (MTBF en Mh)
- évolution de capacité régulière



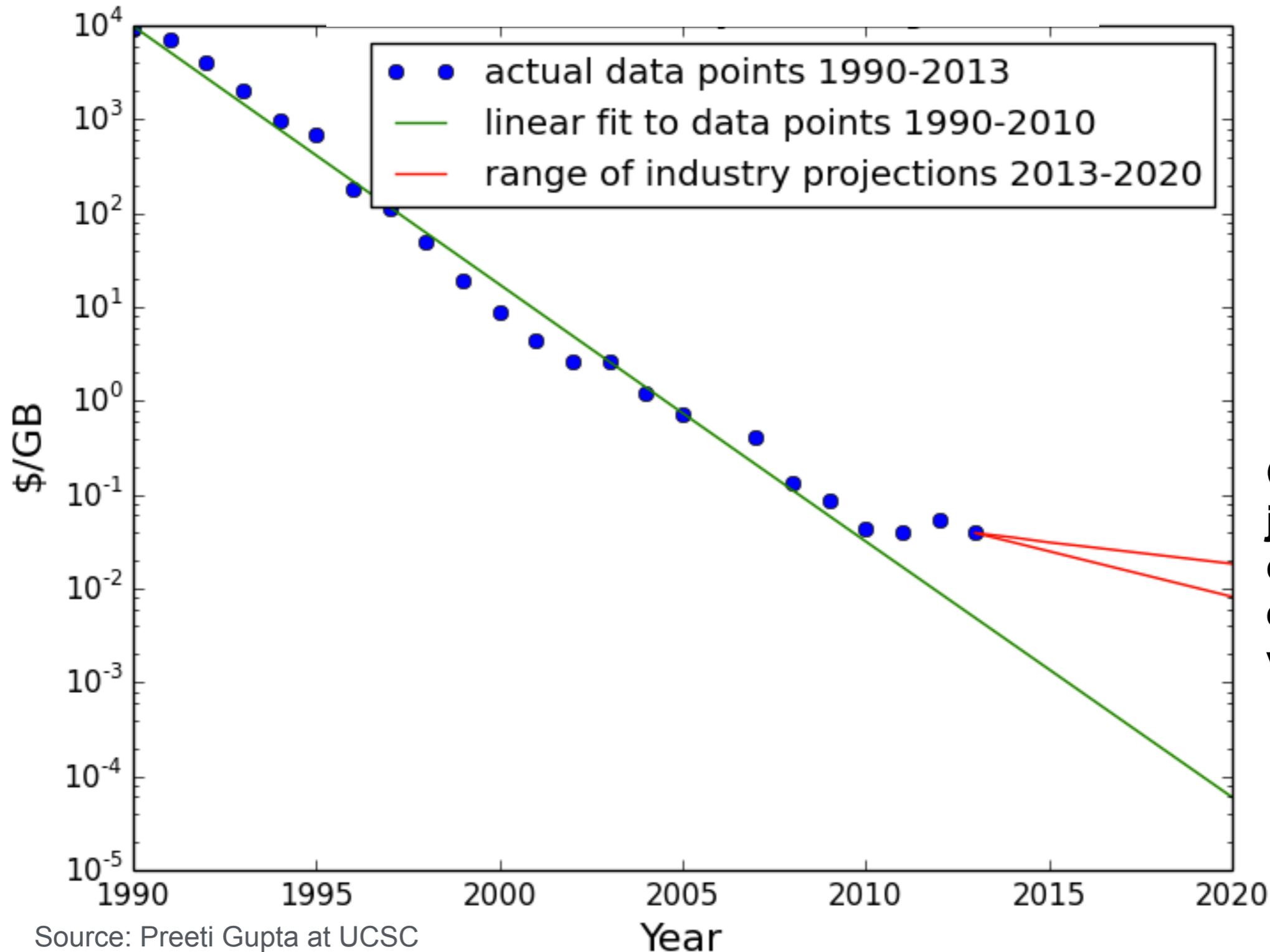
WD Gold
Enterprise-class

400 USD
12 To
2.5 M h MTBF
660 g

La loi de Kryder



Evolution du coût de stockage sur disque (coût par GB)



Coût du stockage jusqu'à 300x plus cher que si la loi était toujours valable

Le “bit-rot” ou dégradation de données



Stockage sous forme
de domaines magnétiques



Stockage sous forme
de charges électriques

Avec l'augmentation des capacités de stockage, le risque d'une corruption accidentelle de données augmente.

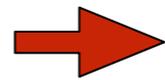
Solution: vérifier et réécrire les données régulièrement



Comment s'assurer de l'intégrité de ces copies soumises au risque de dégradation spontanée?

hash

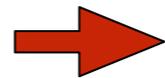
Université libre de Bruxelles



4e8b8da4a7315fd8b5e5dcc5ac313f6657d730c4

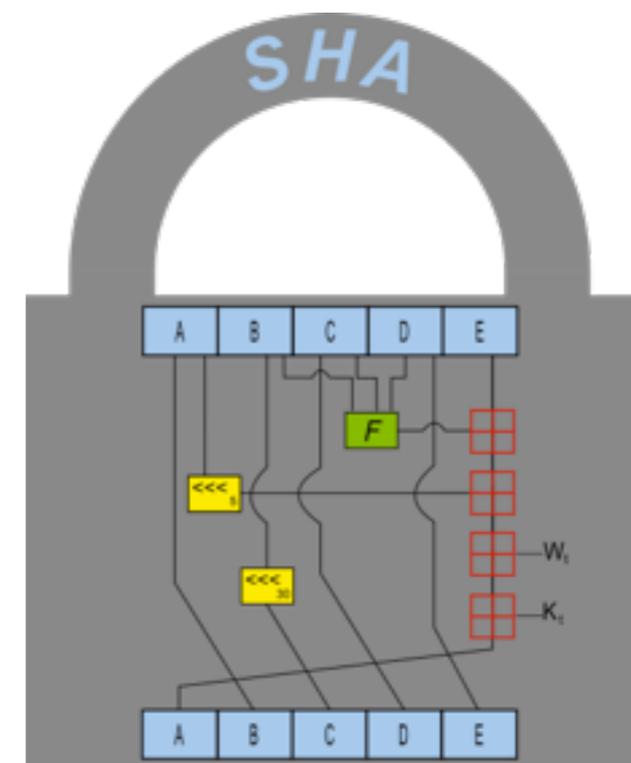


Université Libre de Bruxelles



bddef72a8beb5aab6b199fac36b1274ba77ecb06

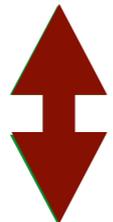
Contrôle d'intégrité efficace pour détecter les erreurs **involontaires**.





Il nous faut calculer les empreintes de chaque copie de l'objet et régulièrement recalculer et comparer à l'empreinte d'origine

Hash original



~~OK~~



Hash recalculé

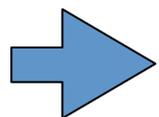
régulièrement

Problème :

Les bits qui constituent l'empreinte d'origine sont soumis exactement aux mêmes menaces que les bits qui constitue les données à préserver !

Si l'empreinte ne correspond pas à celle d'origine:

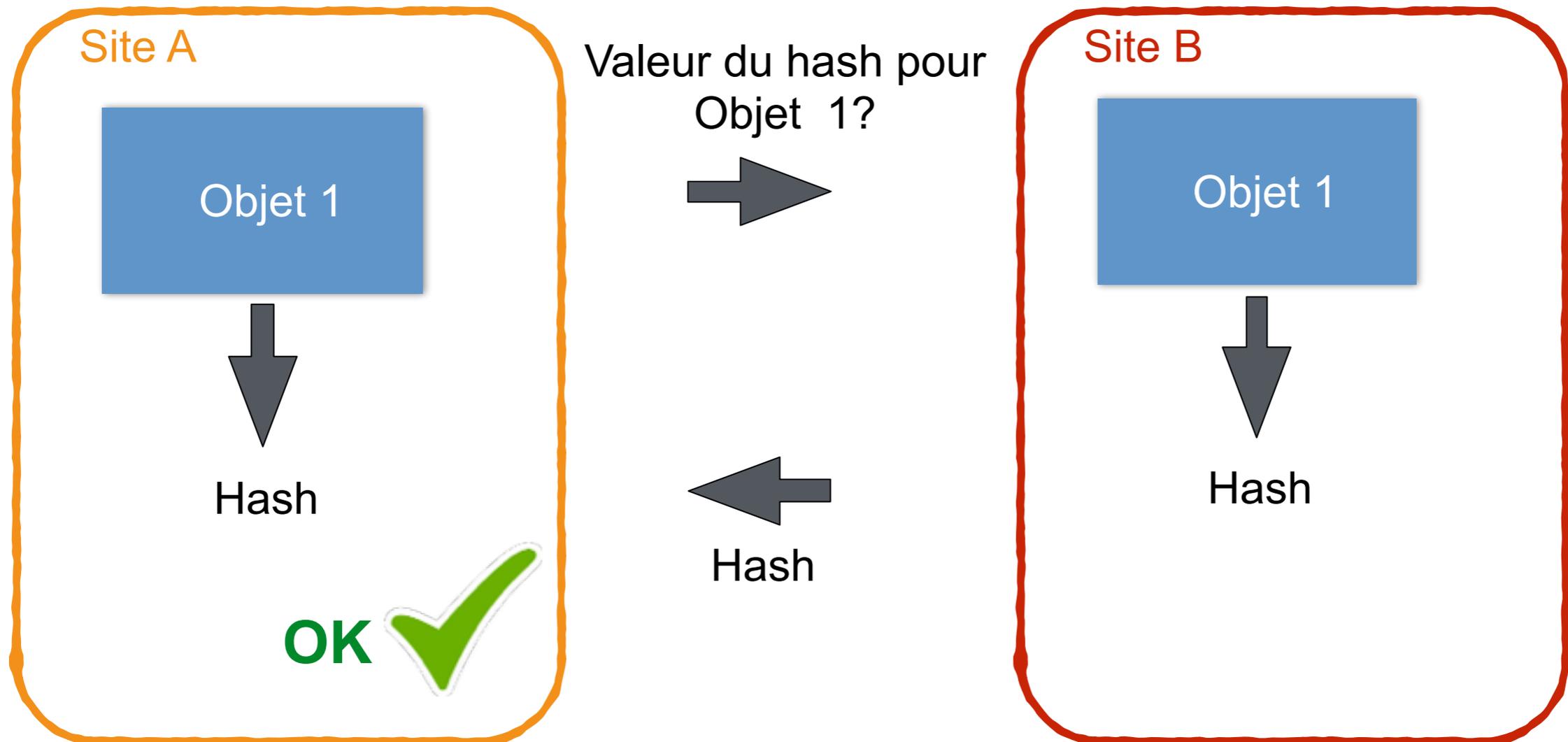
- soit l'objet est compromis
- soit l'empreinte d'origine est compromise
- soit les deux sont compromis



Nécessité de comparer aux autres copies existantes



Comparaison avec les autres copies du réseau



Si on multiplie ces vérifications avec tous les sites, on réduit sensiblement la probabilité de corruption globale.



6. Défaillances matérielles

→ **Surveillance continue des données**



Vérification régulière de l'intégrité des données

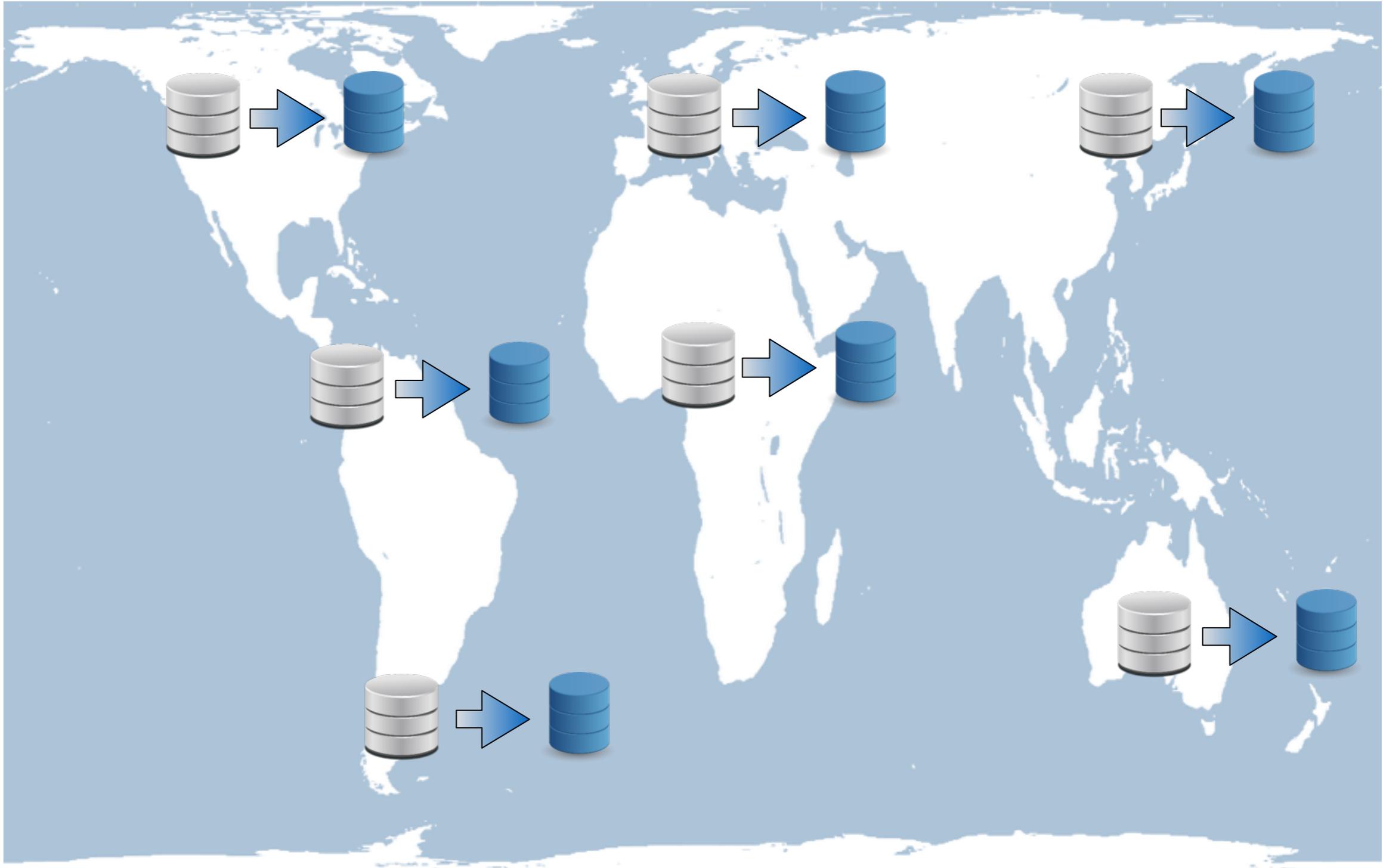
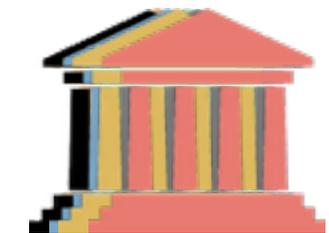
Remplacement des medias dès les premiers signes de défaillance

Hétérogénéité du matériel pour éviter les corrélations de pannes



6. Défaillances matérielles

→ **Surveillance continue des données**





7. Attaques internes ou externes



Mesures de sécurité :

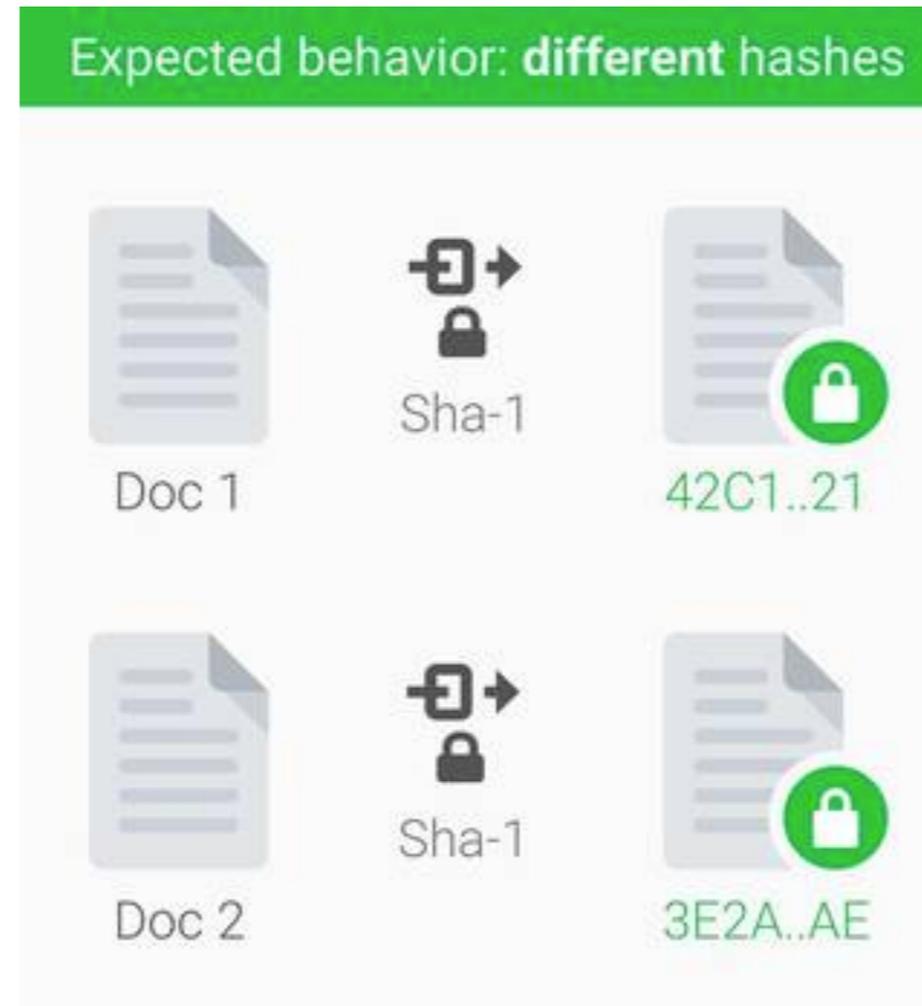
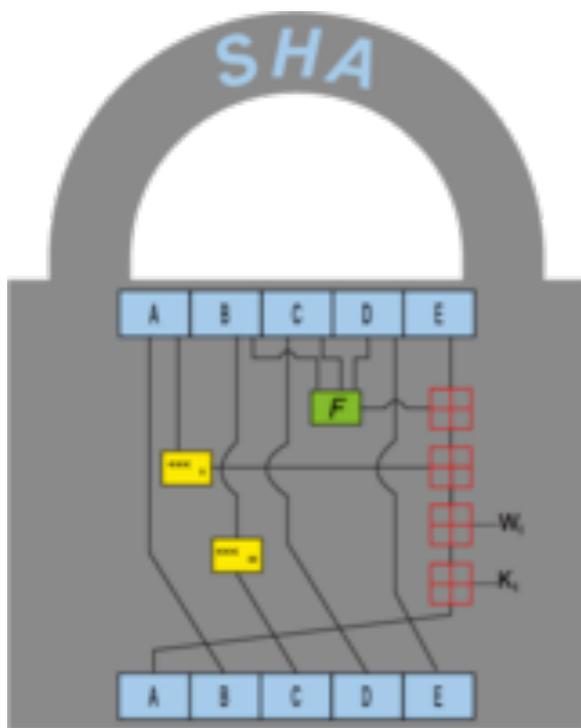
- Mise à jour logicielle régulière
- Authentification
- Pare-feu



Attaque externe : corruption volontaire

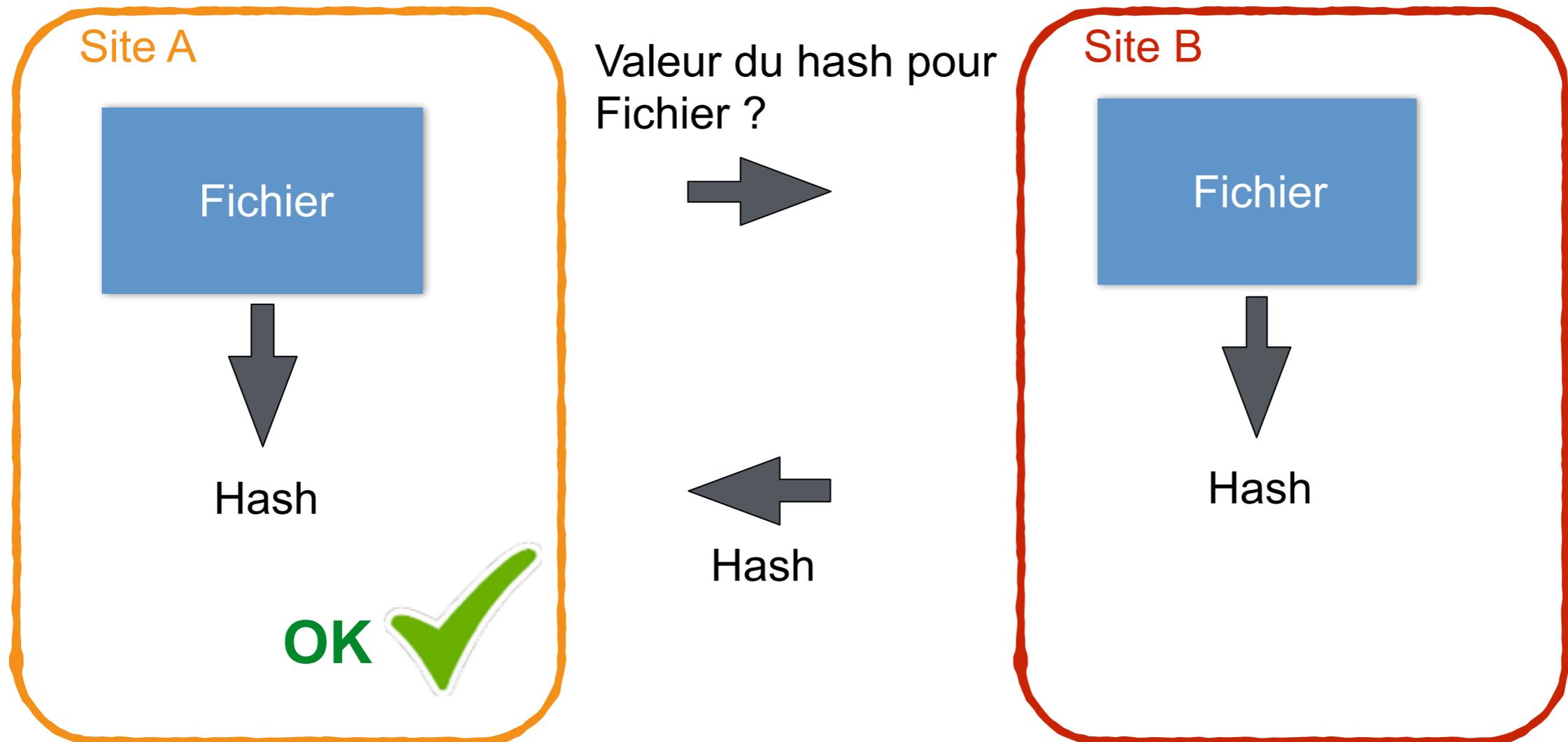
Si un attaquant veut **volontairement** modifier le contenu, il lui suffit d'adapter les empreintes au nouveau contenu.

Plus efficace, les collisions d'empreintes sont possibles :





Attaque interne : comment s'assurer qu'on a bien N copies dans le réseau ?

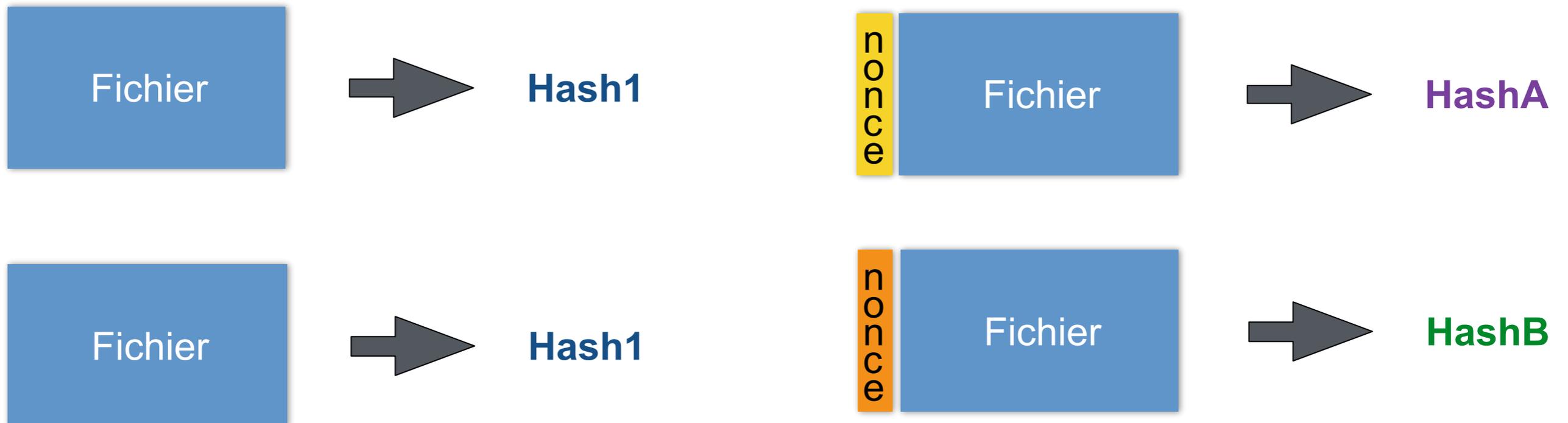


Il faudrait pouvoir s'assurer que les autres copies n'ont pas été détruites ou modifiées intentionnellement.



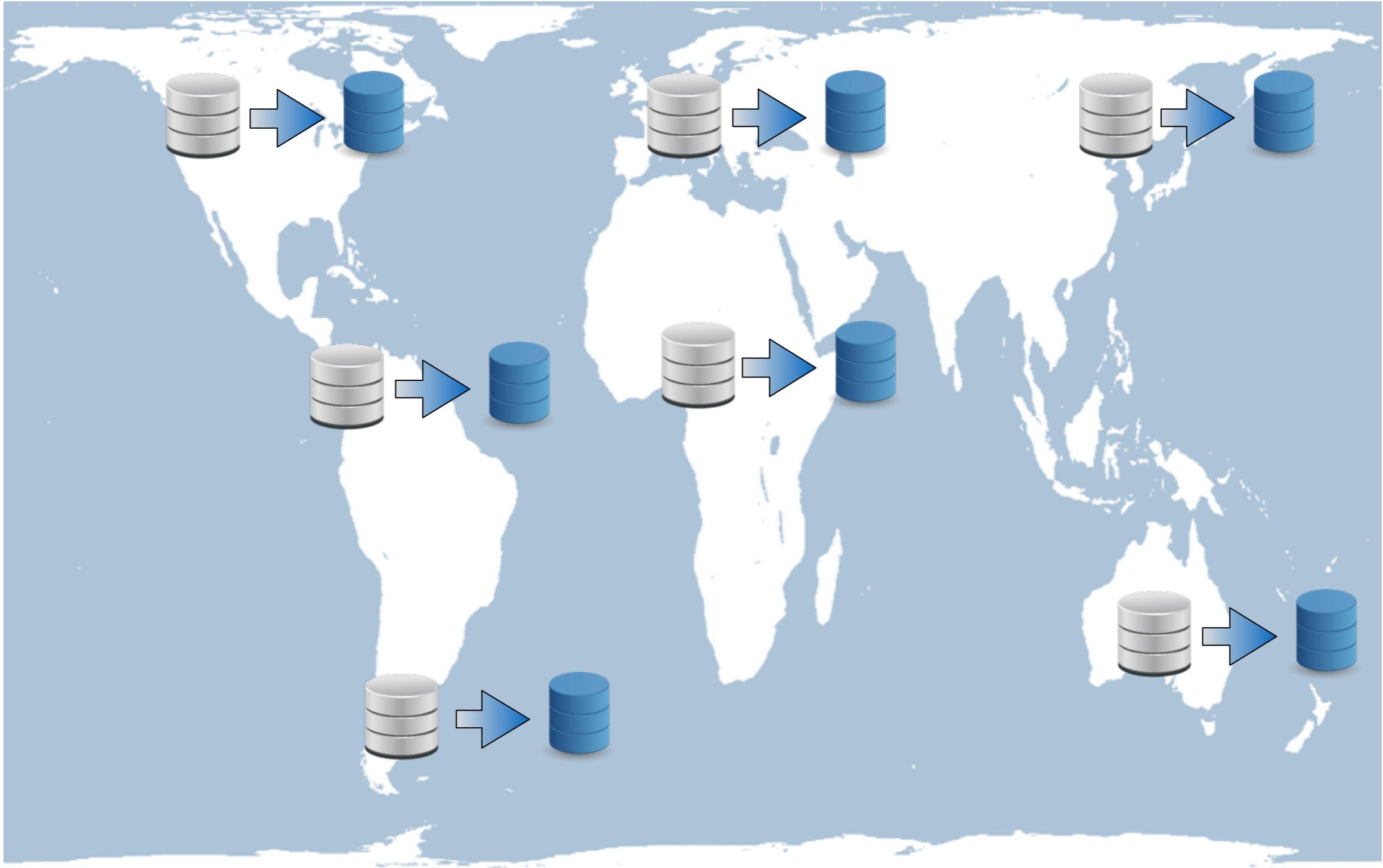
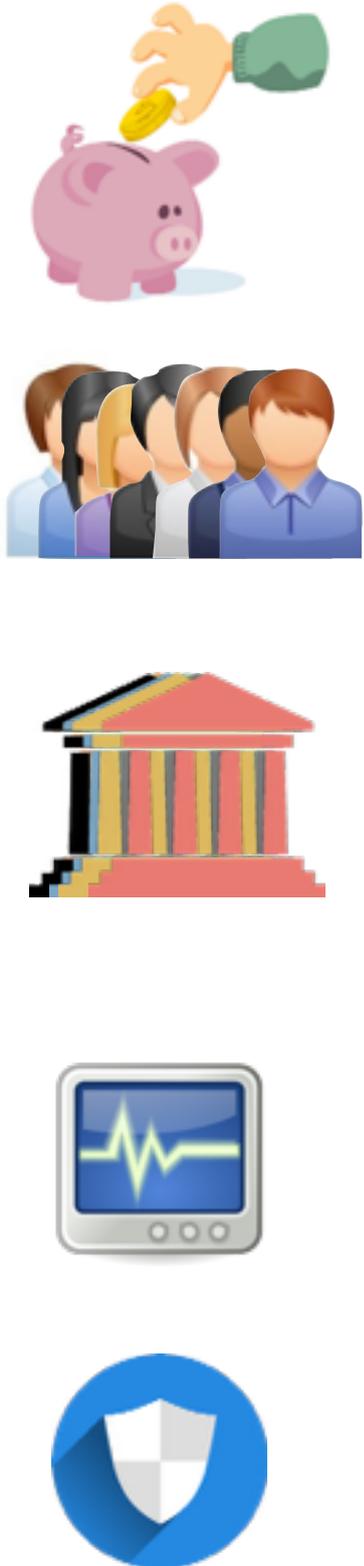
7. Protocole sécurisé de vérification des copies

Vérifier l'intégrité en temps réel en ajoutant un nombre aléatoire utilisé une seule fois au fichier (nonce)





7. Protocole sécurisé de vérification des copies





Une technologie permet de mettre en oeuvre un tel protocole sécurisé de vérification

Lots of Copies Keep Stuff Safe

logiciel open-source

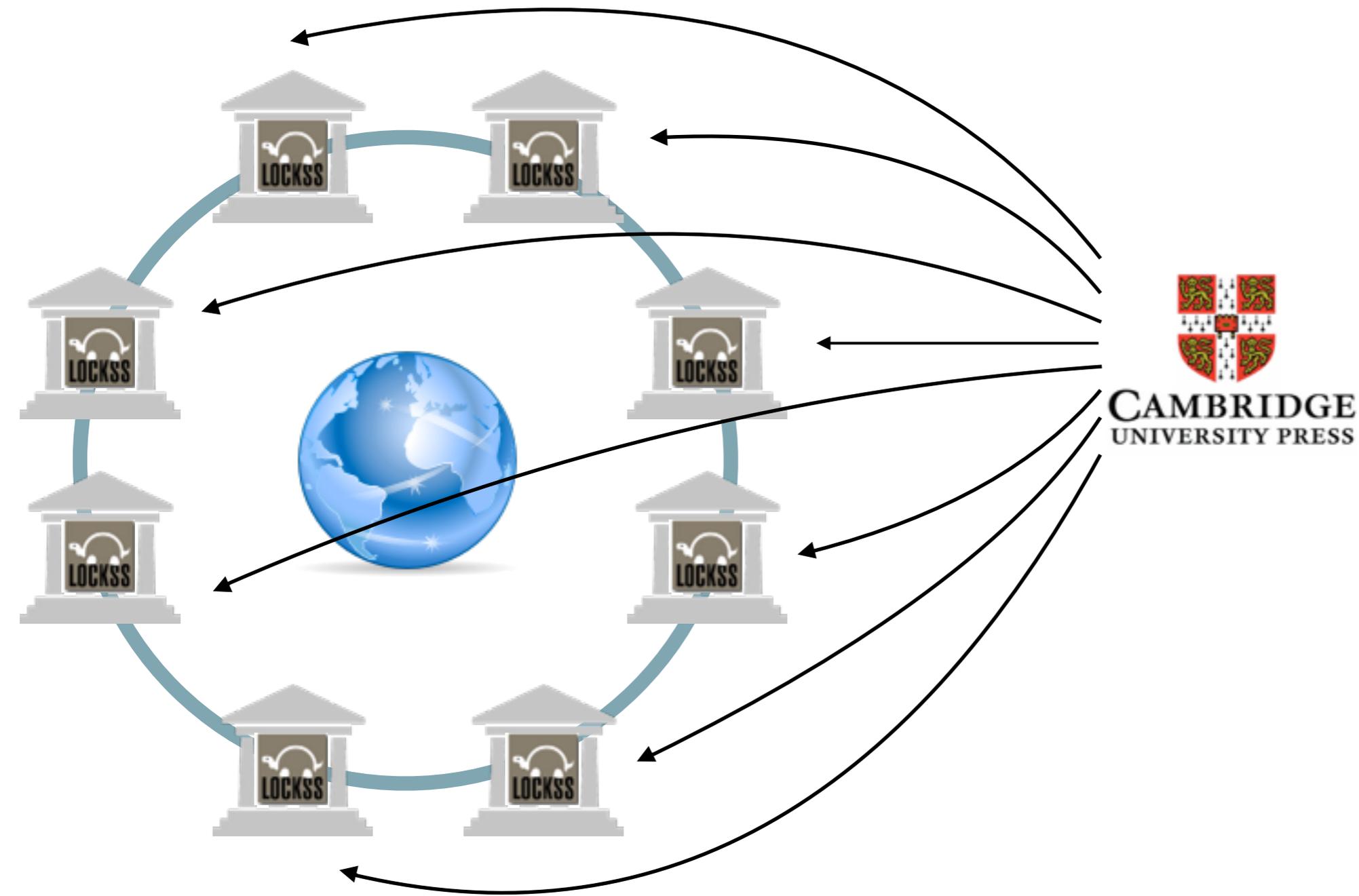
développé par David S. H. Rosenthal

pour le Global LOCKSS Network





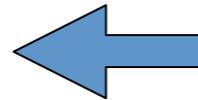
Le **Global LOCKSS Network** est un réseau mondial de préservation communautaire public



+200 noeuds au niveau mondial
coût modeste sur le long terme : simple PC
monitoring constant des données



La technologie LOCKSS permet de transformer un simple PC en nœud de préservation robuste



<http://difusion.academiewb.be/>

DI-fusion

Recherche avancée Historique de recherche Mon DI-fusion ULB | Mon DI UMONS | Aide

Passer-partout Recherche

Derniers dépôts

Platon et l'aporie du politique
par Legros, Robert Marie
Publication 1981

INFECTIONS A BACILLUS
CEREUS. A PROPOS DE 3 CAS
PERSONNELS
par Waks, Danielle, Serruys, Elisabeth
Publication 1981

Quelques tendances
fondamentales de la
philosophie du droit
par Legros, Robert Marie
Publication 1978-03

Les derniers dépôts
comme flux RSS

Afficher les derniers
dépôts

DI-fusion
Portail de consultation des dépôts institutionnels de l'Académie Wallonie-Bruxelles

DI-fusion est le portail de consultation des dépôts institutionnels numériques de l'Université libre de Bruxelles et de l'Université de Mons, membres de l'Académie Wallonie-Bruxelles. Chaque université dispose de son dépôt institutionnel, lequel constitue l'outil de référencement de sa production scientifique. L'interface de recherche DI-fusion permet de consulter les publications des professeurs et des chercheurs de l'ULB et de l'UMONS et les thèses qu'y ont été défendues.

Plus d'informations sur le Dépôt institutionnel de l'ULB.

Plus d'informations sur le Dépôt institutionnel de l'UMONS.

Astuces de recherche

Recherche d'expressions
Vous pouvez utiliser des guillemets pour combiner des mots entre eux:
ex. "Deuxième guerre mondiale"

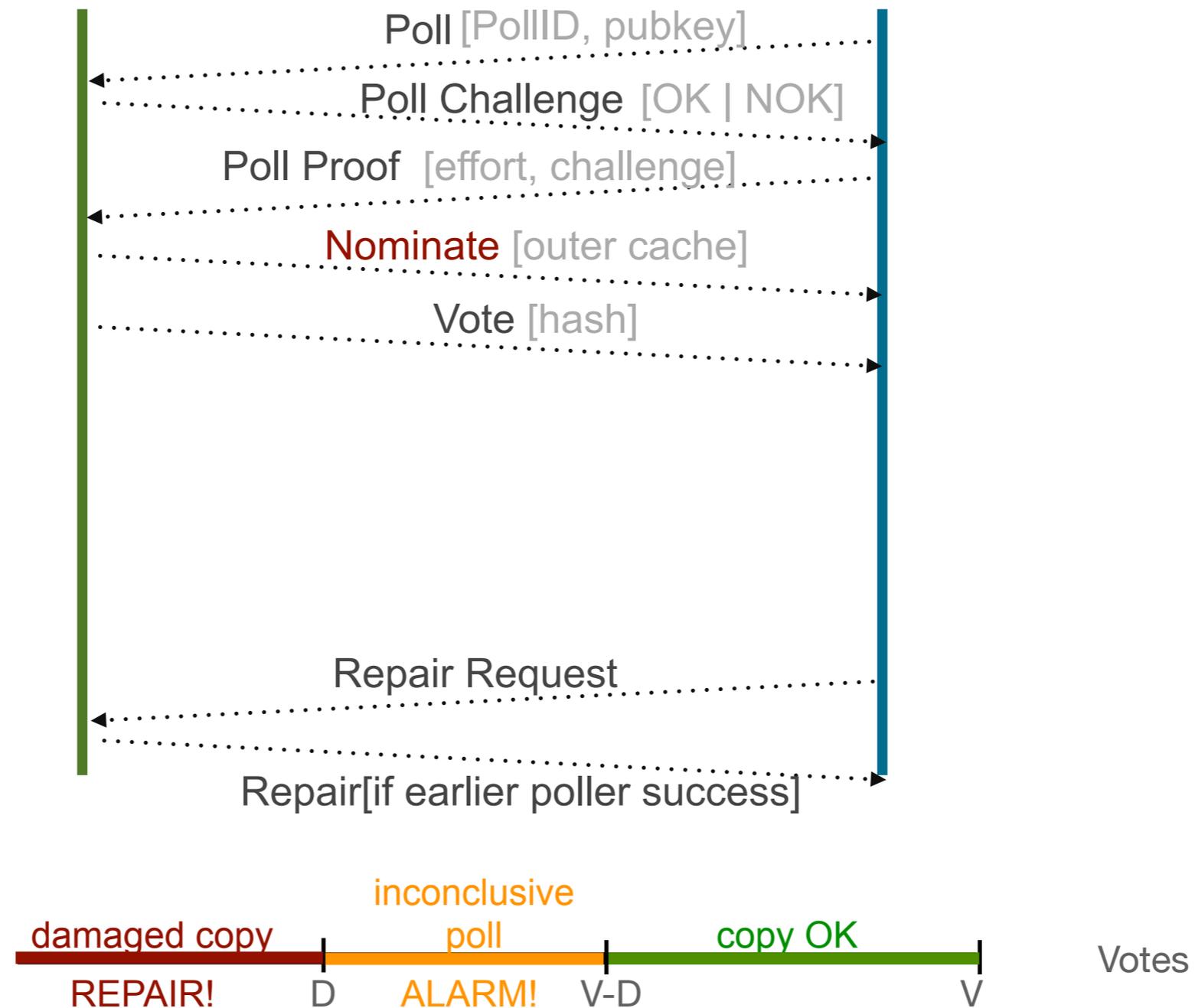
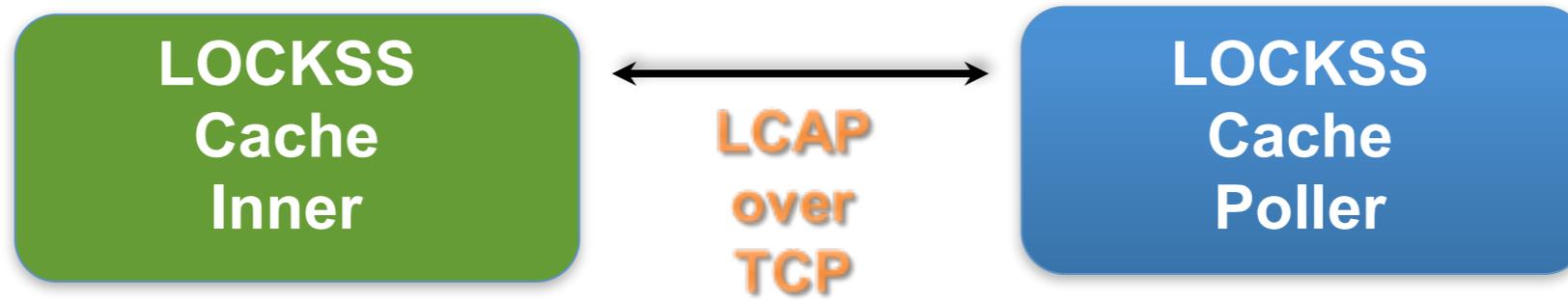
Troncatures et masques
Vous pouvez utiliser un * ou un ? pour représenter un caractère. Le * peut représenter 0 ou plusieurs caractères. Le ? peut représenter 1 seul caractère.
ex. histo* trouvera à la fois historique ainsi que histoires.

Recherche booléenne
Vous pouvez utiliser les opérateurs booléens AND, OR, NOT entre les mots ou les phrases pour combiner avec la logique booléenne.
ex. (chine OR inde) AND économie trouvera les documents qui traitent de l'économie de la chine ou de l'économie de l'inde.

A propos de DI-fusion | Bibliothèques de l'UMONS - Blogus Operandi - Helpdesk (ULB) | Helpdesk (UMONS) Conditions d'utilisation - Version : 1.2 (2010)



LOCKSS en pratique: monitoring de l'intégrité des caches





La solution de préservation idéale devrait intégrer tous les aspects évoqués précédemment

Redondance géographique

Monitoring des données

Migration des media

Authentification

Gestion indépendante des sites

Indépendance Administrative

Maîtrise des coûts

De nombreuses copies contrôlées périodiquement stockées sur des supports fiables, surveillés, et régulièrement mis à jour, dans un environnement logiciel sécurisé gérées par des personnes différentes dans des institutions indépendantes et à faible coût



Combien de copies ?

Idéalement en fonction de l'analyse de risque (difficile)

Le plus sera le mieux !

3 est le strict minimum

4 permet de survivre à une défaillance

7 permet de survivre à deux défaillances simultanées

Problème des généraux Byzantin (tolérance aux pannes)

$3f + 1 \leq n$ f : nb de défaillances tolérables simultanément



Trois piliers de la préservation numérique: technologie, organisation, ressources



Reste à constituer une association d'institutions partageant cette même approche de la préservation numérique.

SAFE FEderation PLN

SAFE Archive FEderation



SAFE PLN



international



structure organisationnelle légère



gestion technique distribuée



chaque partenaire peut surveiller son contenu au sein du réseau



les budgets restent complètement indépendants



SAFE

PLN : Où ?



Université
de Montréal

MEMORIAL
UNIVERSITY

UNIVERSITEIT
GENT

ULB



Universität Bielefeld

UCL
Université
catholique
de Louvain



UNIVERSITÉ
DE GENÈVE

A l'heure actuelle, le réseau compte 8 membres



SAFE

PLN : Qui ?



Martin Sévigny



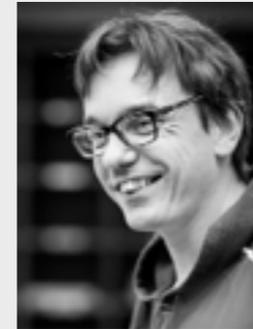
Christian Bélanger



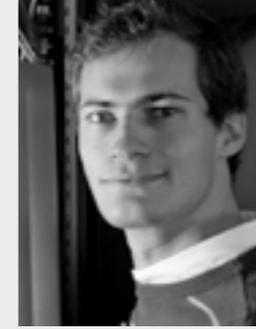
Friedrich Summann



Christian Pietsch



Patrick Hochstenbach



Omar Reygaert



Snorri Briem



Frank Hansen



Benoit Erken



Renaud Michotte



Krista Godfrey



Shawn Wagner



Jan Melichar



Hugues Cazeaux



Jean-Blaise Claivaz



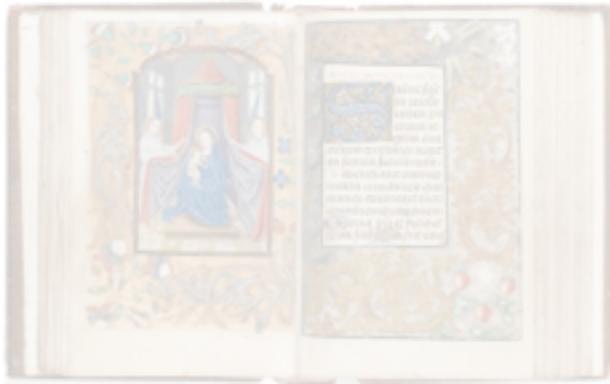
Benoit Pauwels



Anthony Leroy

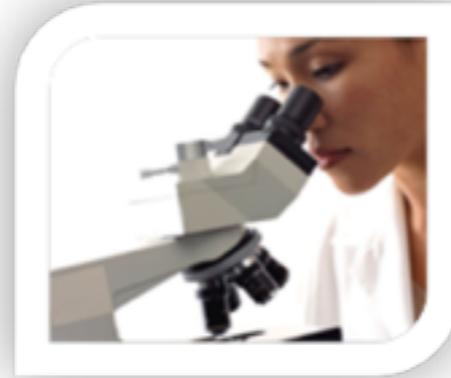
**SAFE**

PLN : Quoi ?



Livres numérisés
Cartes
Photos/Images

PDF/A
JPEG2000 lossless



**Thèses et
publications
(IR)**

PDF/A



Mémoires

PDF/A



SAFE

PLN : COMMENT ?



Monitoring physique du réseau:

Information globale de la disponibilité des services et des machines

Monitoring logique du réseau:

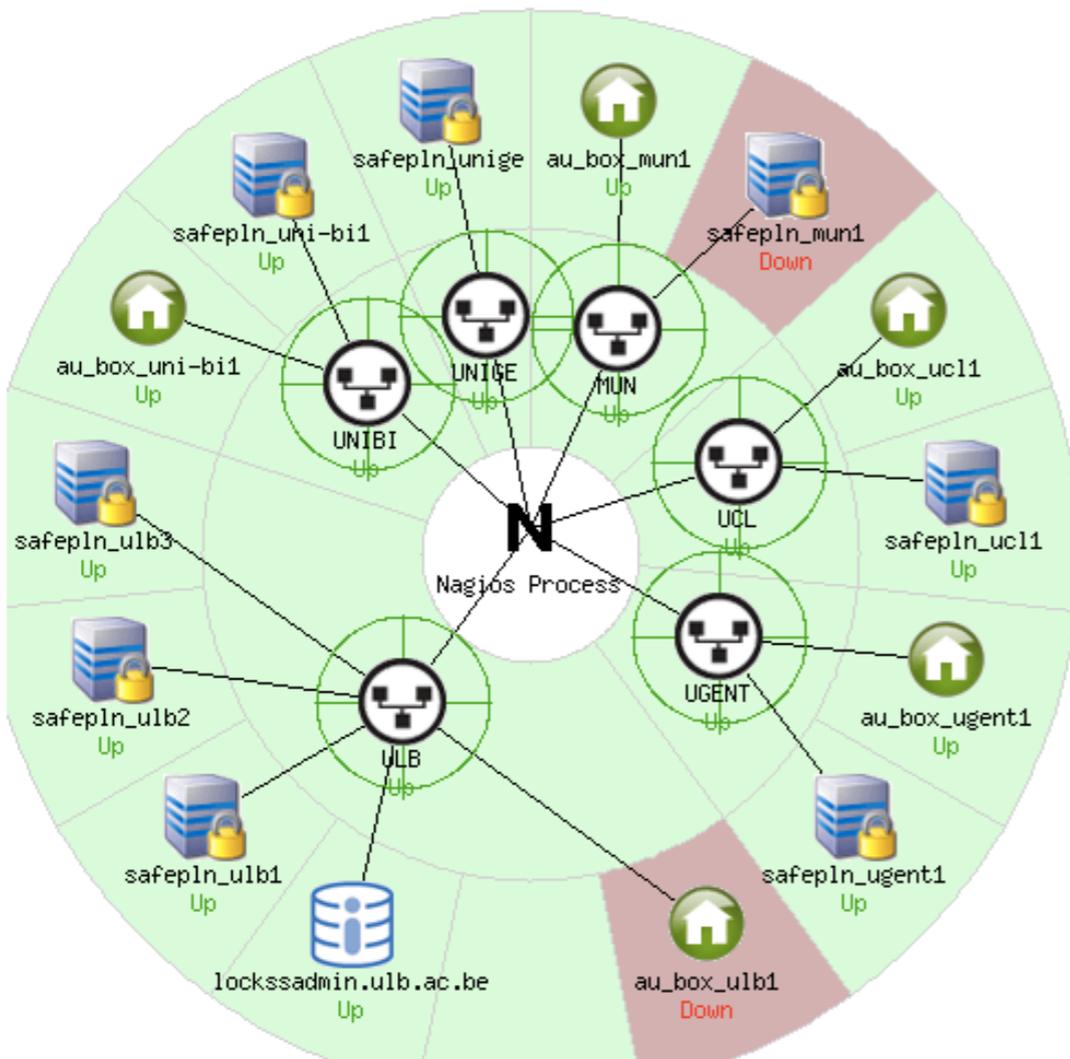
- Nos archives sont-elles **en sécurité**? (tous les votes indiquent que l'intégrité est préservée)
- **Où** les archives sont-elles préservées? (geo-replication efficace)



Quand un problème apparaît les outils envoient automatiquement un email aux administrateurs techniques.



Nous avons ainsi une vision globale de l'état du réseau qui nous permet de rapidement **détecter les problèmes**



Copy CSV Excel PDF Print Search:

AU Name	MUN	UGent	ULB1	ULB2	ULB3	UCL	Unibi	UNIGE
Academie Louvain Open Access Publications 2009	No Info	No Info	Warning	Success	Warning	No Info	Success	No Info
Academie Louvain Open Access Publications 2010	No Info	No Info	Success	Success	Success	No Info	Success	No Info
Academie Louvain Open Access Publications 2011	No Info	No Info	Success	Success	Success	No Info	Success	No Info
Academie Louvain Open Access Publications 2012	No Info	No Info	Success	Success	Success	No Info	Success	No Info
Academie Louvain Open Access Publications 2013	No Info	No Info	Success	Success	Success	No Info	Success	No Info
Bielefeld University Open Access Publications 2009	No Info	No Info	Warning	Success	Warning	No Info	Success	Danger
Bielefeld University Open Access Publications 2010	No Info	No Info	Success	Success	Success	No Info	Success	Danger
Bielefeld University Open Access Publications 2011	No Info	No Info	Warning	Success	Warning	No Info	Success	Danger
Bielefeld University Open Access Publications 2012	No Info	No Info	Success	Success	Success	No Info	Success	Danger
Bielefeld University Open Access Publications 2013	No Info	No Info	Success	Success	Success	No Info	Success	Danger

Showing 1 to 10 of 19 entries

Previous 1 2 Next

**Boîtes LOCKSS
&
Staging servers**

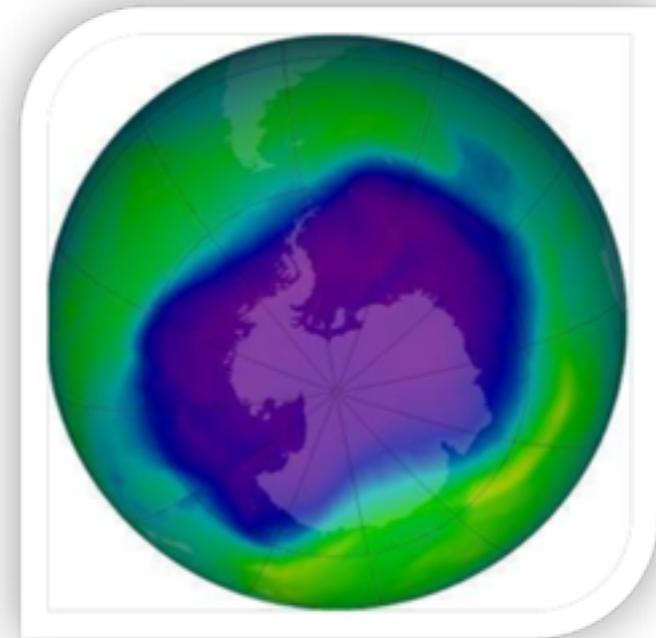


Archives

On distingue deux **niveaux** de préservation:

Préservation
des données **brutes**

Préservation
des données **logiques**

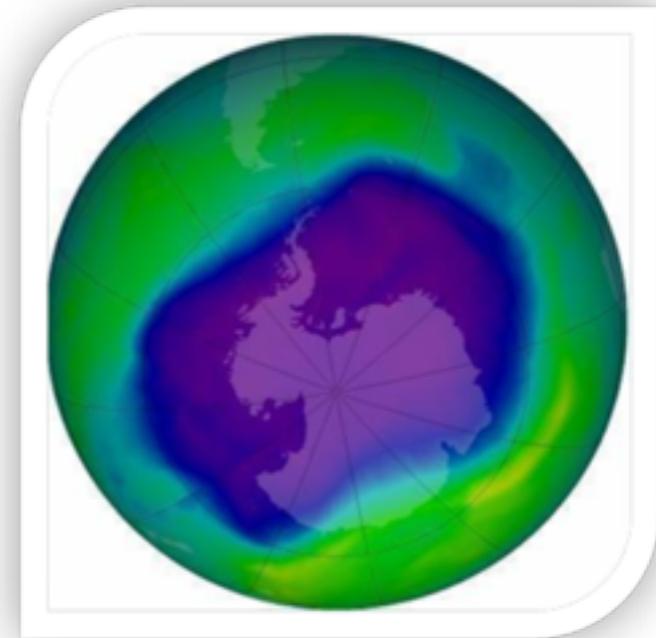


On distingue deux **niveaux** de préservation:



Préservation
des données **brutes**

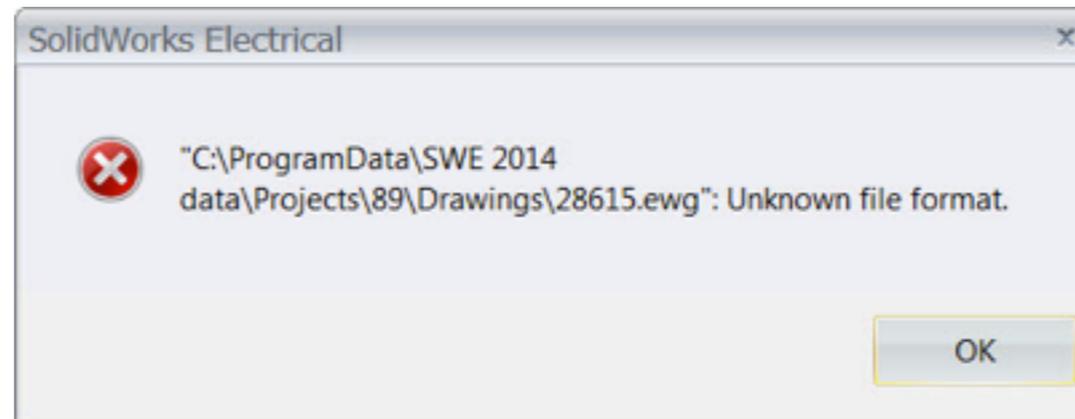
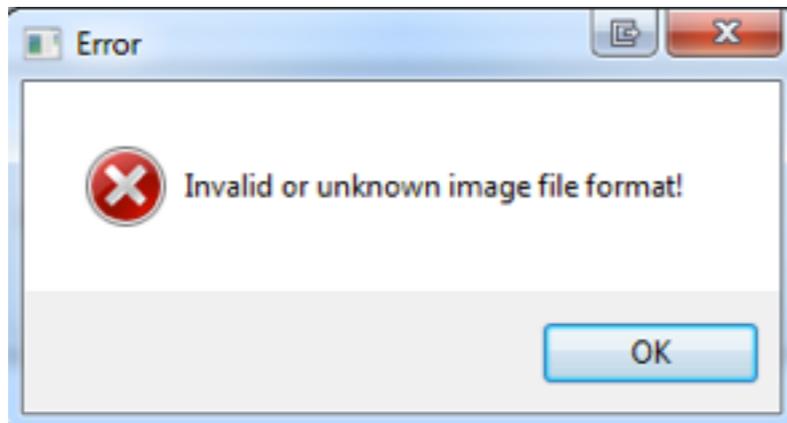
Préservation
des données **logiques**





8. Obsolescence des formats

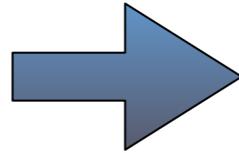
Même si l'intégrité des objets est préservée, encore faut-il que le format soit toujours interprétable dans le futur.



**Concerne surtout les formats propriétaires,
très peu les formats ouverts.**



Nous souhaitons préserver les objets numériques
pour la postérité



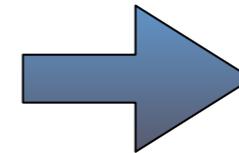
2013



Nous souhaitons que ces objets numériques demeurent identiques en termes de données brutes ET qu'ils soient toujours interprétables sur le long terme



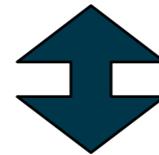
2300





Collaboration entre les dépôts numériques et service de préservation pour constituer une archive

Dépôts numériques



Service de préservation



Ce service de préservation aurait deux rôles principaux



Préserver l'interprétabilité et
l'exploitabilité des données

Assurer l'origine, l'authenticité
et l'intégrité des données



Migration de format (normalisation)

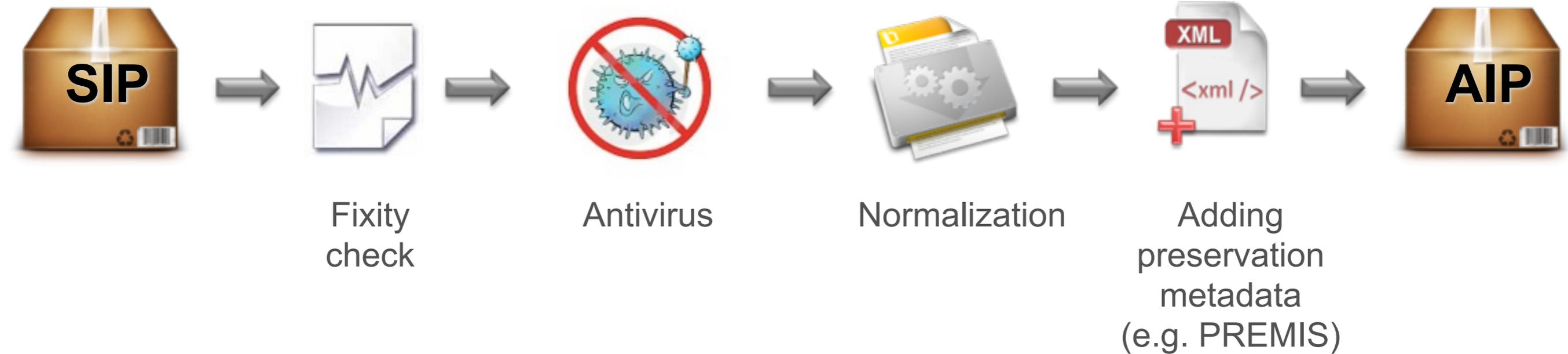
Préservation de l'environnement de
l'objet numérique



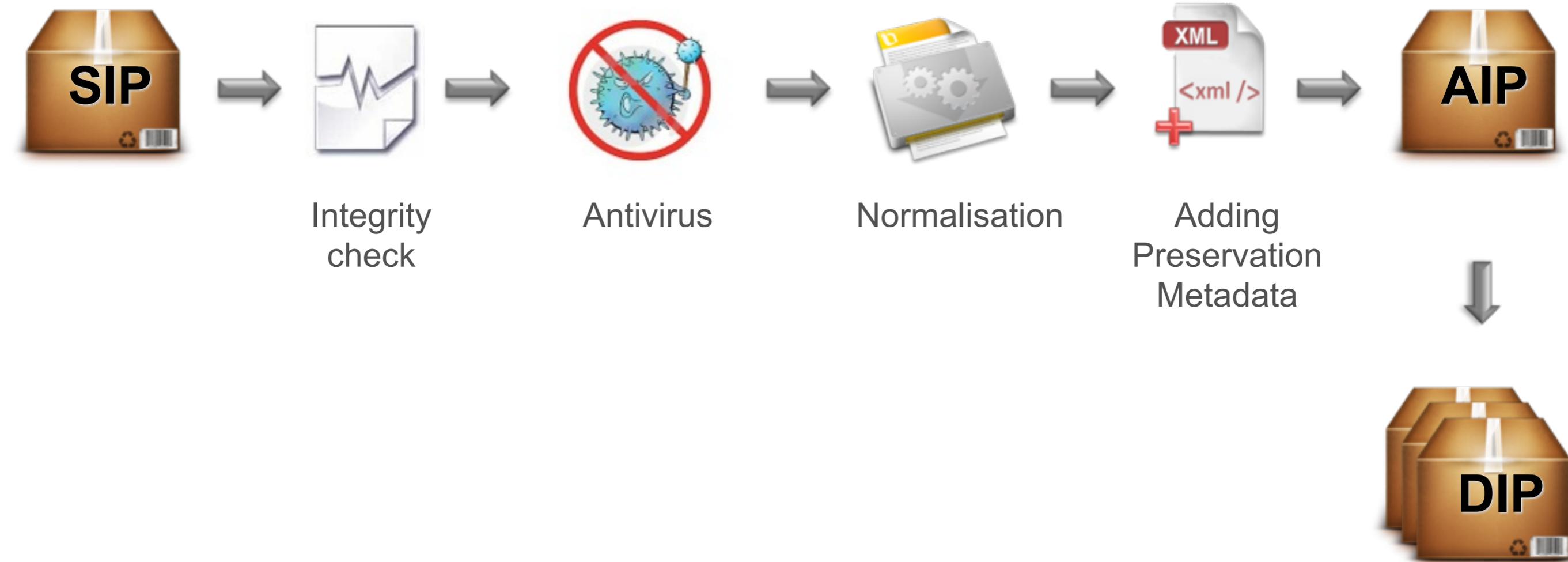
Ajout de méta-données de préservation

Log de toute modification

De l'objet d'origine à l'archive...

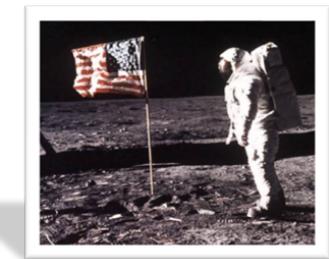


De l'objet d'origine à l'archive...



Les objets disséminés seraient dérivés directement des objets archivés:
Assurance que les objets disséminés sont représentatifs de ce qui est préservé

La normalisation



Vérification
intégrité

Vérification
antivirus

Normalisation

Ajout de
métadonnées
de préservation

Etapes de normalisation des fichiers préservés



Définir une politique de migration:

PLATO, DRAMBORA

Identification du format d'origine:

Jhove, DROID, FIDO

Evaluation du risque d'obsolescence de format:

AONS

Migration proprement dite à l'aide de convertisseurs open-source:

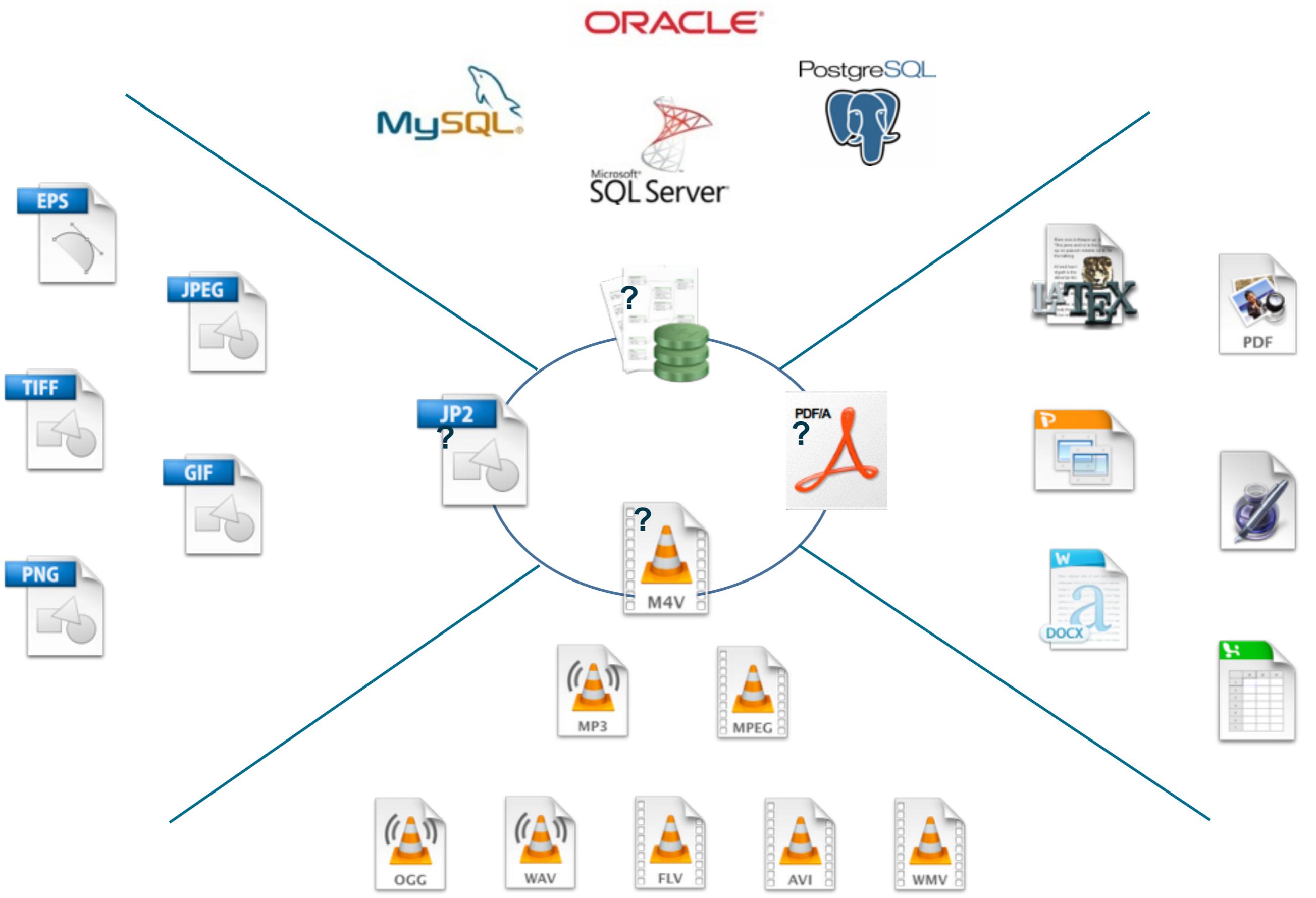
ffmpeg, unoconv(OpenOffice), imagemagik, inkscape

Ajout des méta-données de préservation concernant la migration

PREMIS

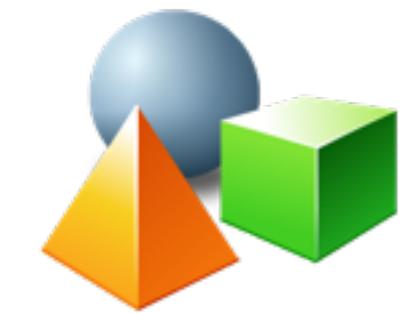


Normalisation des formats: définition d'un plan d'action





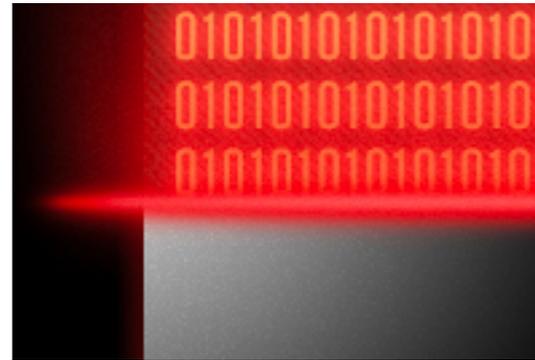
Notre activité principale est de développer l'infrastructure logicielle permettant la description, la soumission, la diffusion et la préservation des objets numériques



Objet numérique
(ou à numériser)



Description



Soumission
ou
Numérisation



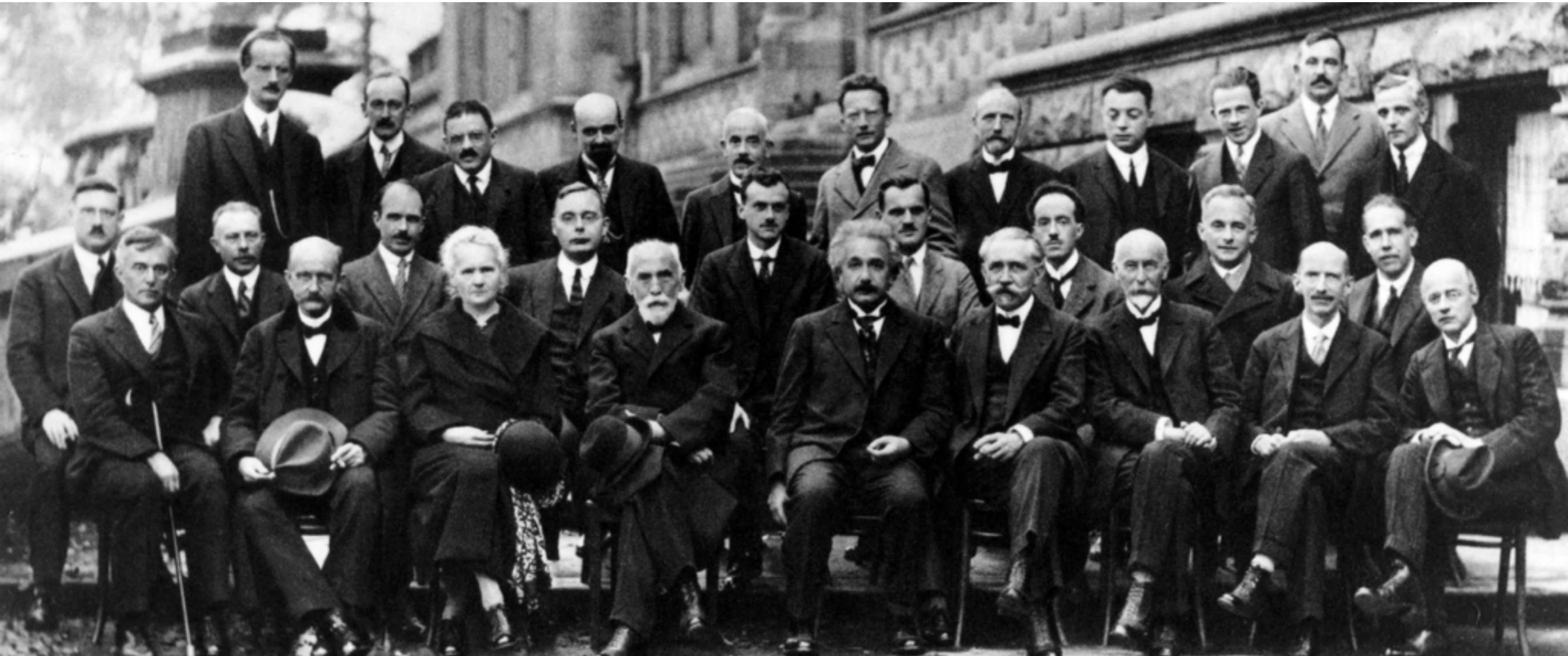
Diffusion



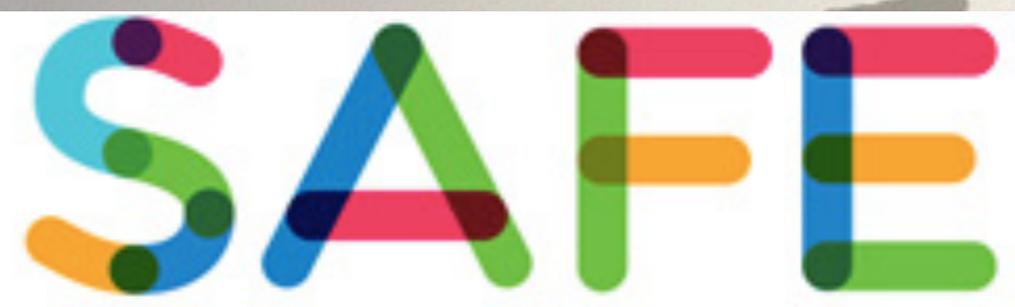
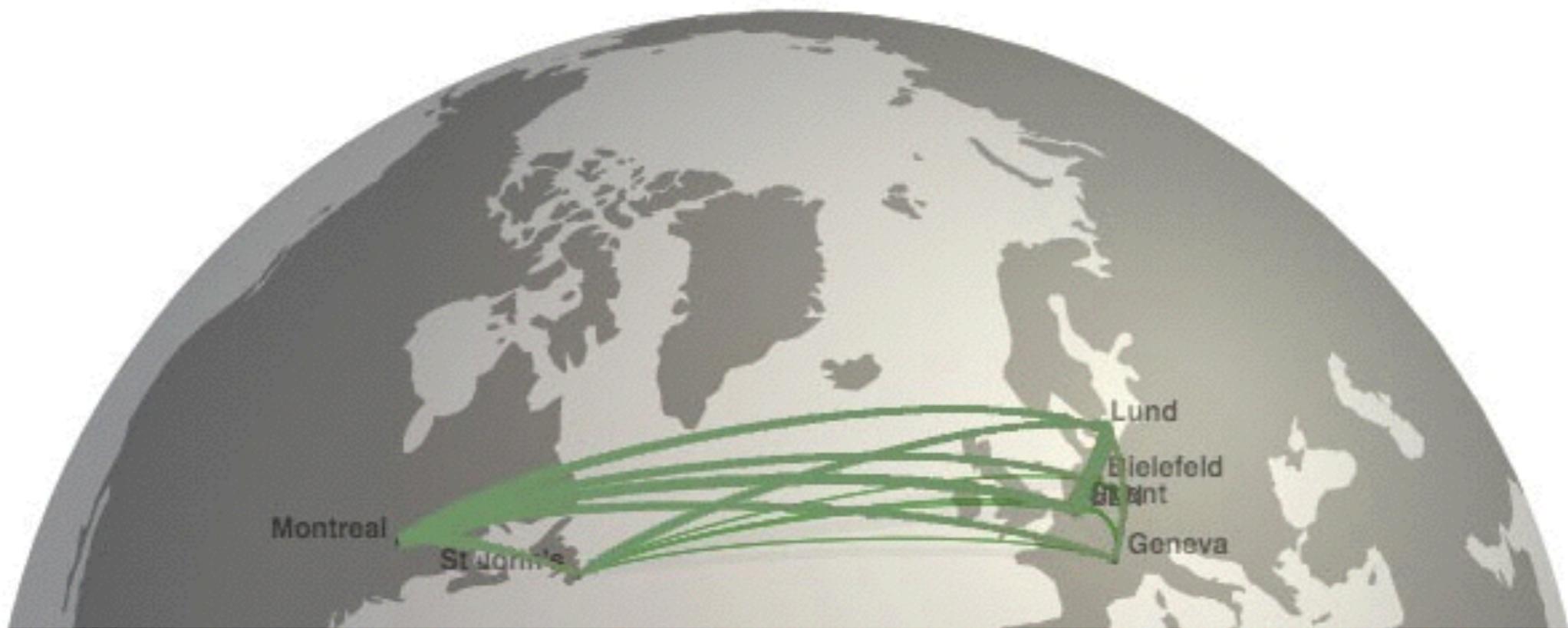
Préservation



Nous mettons tout en oeuvre afin de préserver le patrimoine scientifique numérique de l'ULB sur le long terme.



Participants au 5e Congrès Solvay, octobre 1927, à l'Institut international de physique Solvay dans le parc Léopold, Bruxelles



SAFE Archiving FEderation

