



**Vendredi 22 mars 2013 – 09h30 : Preuve et sécurité**

## **Cloud computing : Défis ou opportunités pour les archives ?**

**Jean-Daniel ZELLER**

Archiviste principal, Hôpitaux universitaires de Genève

Enseignant à la Haute école de gestion de Genève (département I&D)

### **Quelques définitions**

Le cloud computing repose sur une structure en couches où chaque couche technique repose sur la précédente. On distingue ainsi :

#### **Infrastructure as a Service**

*Infrastructure as a Service (IaaS)* permet aux utilisateurs d'accéder aux ressources informatiques sur lesquels ils peuvent déployer du logiciel, qui peut inclure des systèmes d'exploitation. Cependant, les utilisateurs n'ont pas accès à l'infrastructure sous-jacente du nuage.

L'exemple le plus frappant est sans doute la suite progressive d'*Amazon Web Services (AWS)*. L'*EC2* d'Amazon permet aux utilisateurs de créer et de gérer des instances de serveur virtuel, alors que leur *Simple Storage Service (S3)* fournit une infrastructure de stockage de données.

#### **Platform as a Service**

*Platform as a Service (PaaS)* donne aux utilisateurs la possibilité de déployer des applications sur l'infrastructure en nuage du fournisseur en utilisant des outils pris en charge par le fournisseur. L'infrastructure, y compris les systèmes d'exploitation, demeure hors du contrôle de l'utilisateur.

*Google App Engine* est un bon exemple de PaaS : les utilisateurs peuvent l'utiliser pour déployer des applications écrites en Python ou en Java, ce qui rend possible l'utilisation d'un certain nombre d'interfaces de programmation d'applications (API) intégré dans la plateforme. *Microsoft Windows Azure* permet de même aux utilisateurs de déployer des applications écrites en plusieurs langages

## Software as a Service

Le SaaS permet aux utilisateurs d'accéder aux applications exécutées sur l'infrastructure en nuage du fournisseur.

Généralement accessibles à l'aide d'un navigateur Web, ces applications offrent souvent des fonctionnalités qui étaient traditionnellement assurées par un logiciel dédié sur l'ordinateur de l'utilisateur. *Google Docs* mentionné précédemment est un exemple, un exemple, l'application de gestion de projet « 37 Signals » de Basecamp en est un autre. Ces applications comprennent souvent des fonctions de collaboration ou de partage qui seraient plus difficiles à mettre en œuvre dans le logiciel de bureautique.

A ces trois couches habituellement offertes par les prestataires s'ajoute actuellement des services de plus haut niveau.

## Business Process as a service (BPaaS)

Le *Business Process as a Service (BPaaS)* est la couche suivante de l'abstraction, après le SaaS. Il s'appuie sur la pile en fournissant un processus, une expertise **et des personnes qui permettent tout ou une partie d'un processus métier**, et le rend disponible via un modèle *pay-per-use*<sup>1</sup>.

## Archiving as a Service

Cas particulier de Business Process as a Service. Le terme a été introduit par Askhoj et al. dans leur tentative de concilier les fonctionnalités du Cloud computing avec le modèle de référence OAIS. (iConférence 2011)<sup>2</sup>

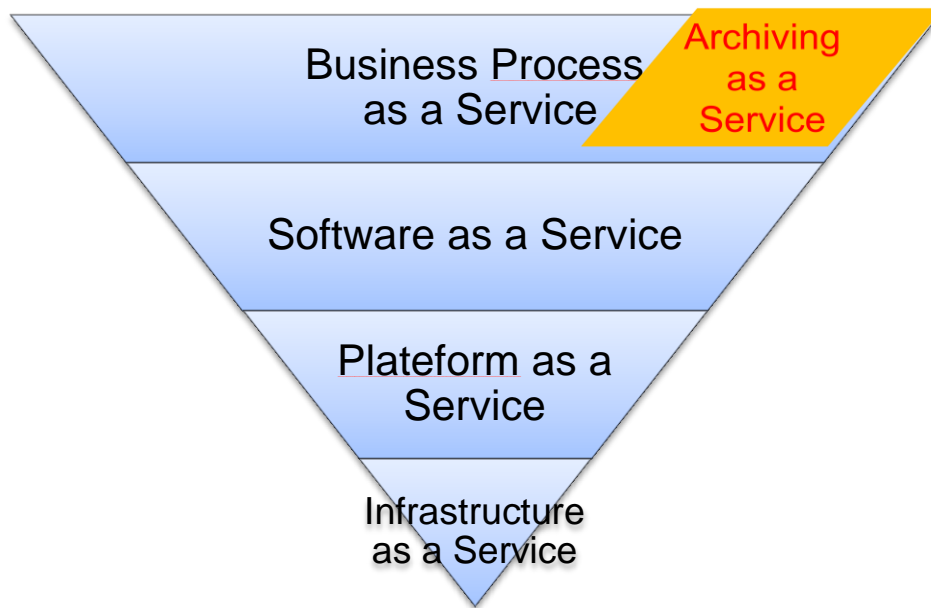
Le terme, parfois sous la dénomination de Archives as a Service, est utilisé par certains prestataires de services (NetApp, Symantec) pour vanter ce qui est en fait des services de stockage sécurisés.

---

<sup>1</sup> **How the Cloud is Transforming Business Process Management**, Aditya Mony & Kalyan Raman  
[http://www.finyear.com/How-the-Cloud-is-Transforming-Business-Process-Management\\_a23351.html](http://www.finyear.com/How-the-Cloud-is-Transforming-Business-Process-Management_a23351.html)

<sup>2</sup> **Archiving as a Service - A Model for the Provision of Shared Archiving Services Using Cloud Computing**, Jan Askhoj, Mitsuharu Nagamori, Shigeo Sugimoto, Proceeding Proceedings of the 2011 *iConference '11*, Pages 151-158 [ACM](#) New York, NY, USA ©2011 [table of contents](#) ISBN: 978-1-4503-0121-3 doi>[10.1145/1940761.1940782](https://doi.org/10.1145/1940761.1940782)

<http://fr.slideshare.net/janaskhoj/archiving-as-a-service-a-model-for-the-provision-of-shared-archiving-services-using-cloud-computing>



On peut représenter ces différentes couches comme s'appuyant les unes sur les autres. Le choix d'une pyramide inversée se justifie dans la mesure où l'utilisation des services de haut niveau est celle qui est la plus courante (voir plus loin le graphique sur le marché du cloud en France).

## Typologie des nuages

Les deux types d'offre les plus courantes sont le nuage public et le nuage privé. Ces appellations sont source de confusion terminologique puisque le nuage public désigne en fait les offres commerciales destinées à n'importe quel type d'utilisateur (individu, entreprise, administration), alors que le nuage privé désigne plutôt un nuage privatif (limité à un seul organisme utilisateur) quel que soit son statut (organisme privé ou public).

### Nuage public

Les services de nuage public **permettent à l'infrastructure d'être utilisé par le grand public** et ils offrent aux utilisateurs les avantages d'évolutivité rapide et d'une faible mise en place initiale des coûts. Google Apps est une suite d'applications SaaS populaire mise à disposition à travers un nuage public.

### Nuage privé

Un nuage privé est **possédé et exploité par un organisme ou un tiers pour le compte d'une organisation**. Pour cette raison, la plupart des avantages de l'infrastructure infonuagique tels que l'externalisation de l'infrastructure informatique et les économies d'échelle seront moins évidents, toutefois les risques de sécurité et les goulots d'étranglement dans le transfert des données peuvent être atténués.

On décrit aussi des offres composites qui sont :



## Nuage Communautaire

Avec les nuages communautaires l'infrastructure fournie est réservée à l'usage exclusif d'une communauté spécifique d'organisations qui ont partagé leurs préoccupations. Comme l'accès est restreint à des utilisateurs particuliers, les nuages communautaires peuvent présenter des risques de sécurité moindre que les nuages publics, mais avec des économies d'échelle plus limitées, ils ont tendance à être moins souple pour s'adapter aux besoins des utilisateurs et les coûts peuvent être plus élevés.

## Nuage hybride

Les nuages hybrides sont une composition de deux ou plusieurs infrastructures en nuage distinctes (privé, communautaire ou public) qui, tout en restant distincts, sont interconnectés pour permettre le transfert de données. Un nuage hybride privé-public peut être utilisé pour stocker des données sensibles en interne tandis que le stockage d'autres données est sous-traité.

## Quel avenir pour le cloud computing ?

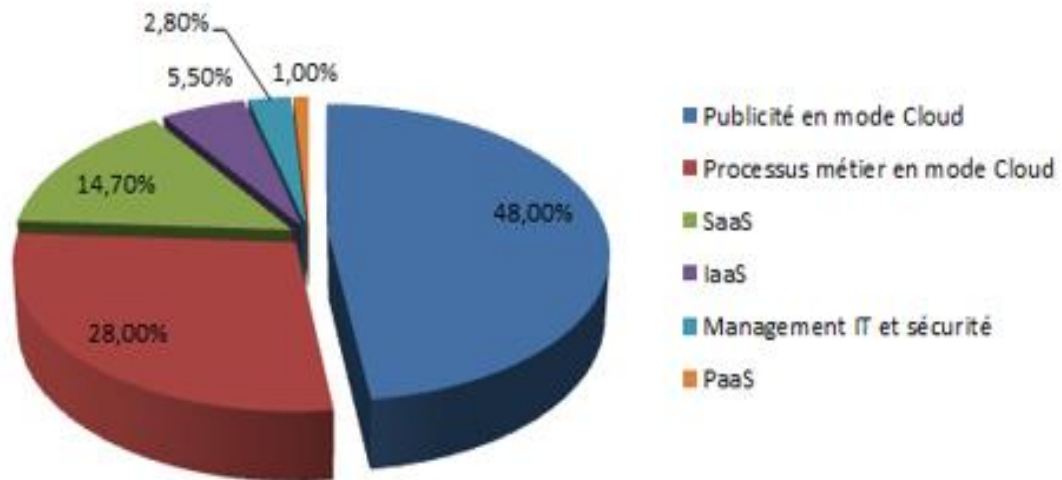


Source: [www.cloudtweaks.com](http://www.cloudtweaks.com)

*J'ai un problème pour vous dire quel est votre futur dans l'informatique,  
tout ce que je peux voir c'est des nuages !*

Tous les chroniqueurs spécialisés prédisent un usage croissant de l'informatique en nuage. Un retour sur le marché réel nuance quelque peu ces prédictions.

### Taux de pénétration des différents services de cloud (Gartner)



Selon les chiffres de Gartner sur l'usage du cloud en 2012, près de la moitié est utilisé pour des activités de publicité, par nature fugaces, et de peu d'intérêt dans le domaine de l'archivage.

Seul un quart des services cloud utilisés ressortent des fonctionnalités IaaS-PaaS-SaaS, qui sont les services de base.

**Environs 30% ressortent du *Business Process as a Service* qui serait susceptible de concerner des activités d'archivage.**

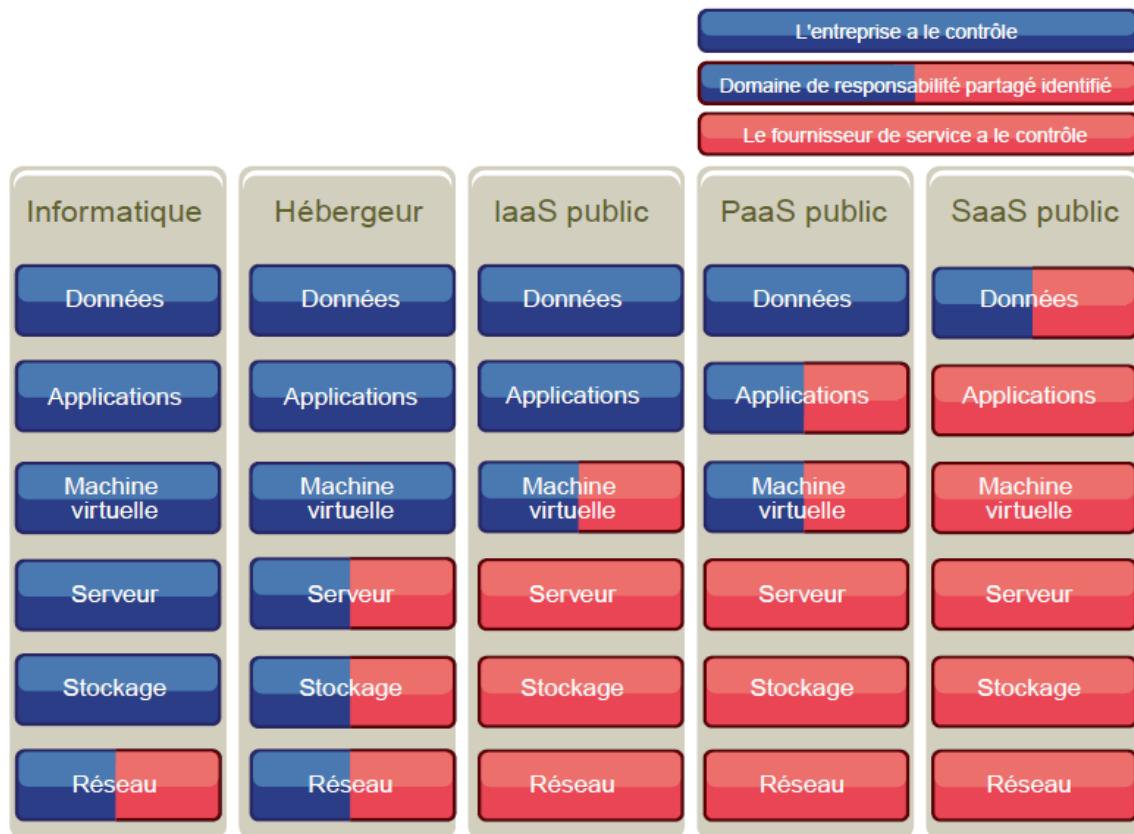
## Quelles garanties offre le cloud ?

En termes de :

1. Fiabilité - Sécurité
2. Continuité de services
3. Migration
4. Protection des données personnelles

### 1. Fiabilité – Sécurité

Les prestataires de services ont été initialement des hébergeurs de grandes quantités de données, qui ont trouvé rentable d'amortir leur énormes capacités de stockage en les louant à des tiers. Leur niveau de sécurité interne est assez élevé puisqu'ils avaient besoins de cette sécurité pour leurs propres données. On peut considérer qu'il est plus élevé que celui d'une PME qui n'a pas forcément l'équipe informatique nécessaire pour une sécurité informatique de haut niveau. Par contre par rapport à des organisations de plus grande envergure, la preuve reste à faire.



Source : Livre blanc sécurité du cloud computing<sup>3</sup>, page 7

Comme on le voit sur le tableau ci-dessus, dans le nuage (IaaS-PaaS, SaaS), c'est le prestataire de service qui a le contrôle majoritaire de la plupart des couches. Ceci est problématique dans la mesure où **9 sociétés sur 10 pensent que la sécurité des données relève de la responsabilité de leur fournisseur de services cloud** !<sup>4</sup> Alors que les prescriptions sur les données personnelles européennes attribuent légalement cette responsabilité au propriétaire des données.

## 2. Continuité de service

Il ne se passe pas une semaine sans que l'on notifie des ruptures de services chez le prestataire de services. Cela peut aller d'une simple indisponibilité de service de quelques heures, pas forcément trop grave, jusqu'à des pertes de données irrécupérables. La plupart du temps, les contrats de services n'offrent qu'une protection très limitée aux clients victimes.

## 3. Migration

La plupart des offres reposent sur des prestations non interopérables et ne garantissant pas la réversibilité (le fait de pouvoir transférer les données hébergées chez un autre prestataire ou sur les serveurs internes de l'organisme client).

<sup>3</sup> Syntec Numérique, **LIVRE BLANC SÉCURITÉ DU CLOUD COMPUTING, Analyse des risques, réponses et bonnes pratiques**, <http://web.archive.org/web/20130121141926/http://www.syntec-numerique.fr/Bibliotheque/Livres-Blancs-Cloud-Computing>

<sup>4</sup> Voir : <http://www.itrnews.com/articles/139084/9-societes-10-pensent-securite-donnees-releve-responsabilite-fournisseur-services-cloud.htm>

Consciente de cette problématique la communauté des télécoms, via l'Union Internationale des Télécommunication (UIT)<sup>5</sup> a entamé des réflexions pour normaliser les niveaux de service, mais ces travaux ne sont pas encore finalisés.

#### 4. Protection des données personnelles

Les prestataires américains (qui forment l'essentiel des prestataires de services en nuage) sont soumis au Patriot Act et au FISAA<sup>6</sup> même pour les données qu'ils gèrent en dehors du territoire américain, ce qui rendent illusoire la protection des données personnelles au sens européen. Il est par conséquent recommandable de ne pas mettre de données personnelles dans le nuage public<sup>7</sup>.

## Les réponses possibles à ce manque de garantie

### Des nuages privés nationaux (ou éventuellement européens) comme :

- Les offres Numergy et Cloudwatt, en France<sup>8</sup>
- Le projet de politique de nuage de la Confédération Suisse<sup>9</sup>
- Le projet de nuage sécurisé (sealed cloud) subventionné par le ministère de l'économie allemand (Sealed Cloud, selon le schéma ci-dessous)<sup>10</sup>

<sup>5</sup> Voir les travaux du Focus Group on Cloud Computing (FG Cloud) qui a produit à ce jour 7 Technical reports traitant différents aspects de la question, <http://www.itu.int/en/ITU-T/focusgroups/cloud/Pages/default.aspx>

<sup>6</sup> **Fighting cyber crime and protecting privacy in the Cloud, 2012**  
[http://www.bakchich.info/sites/bakchich.info/files/article\\_files/fisaa\\_0.pdf](http://www.bakchich.info/sites/bakchich.info/files/article_files/fisaa_0.pdf)  
<http://www.europarl.europa.eu/committees/en/studiesdownload.html?languageDocument=EN&file=79050>

<sup>7</sup> Voir recommandations de la CNIL, **Recommandations pour les entreprises qui envisagent de souscrire à des services de Cloud computing,**

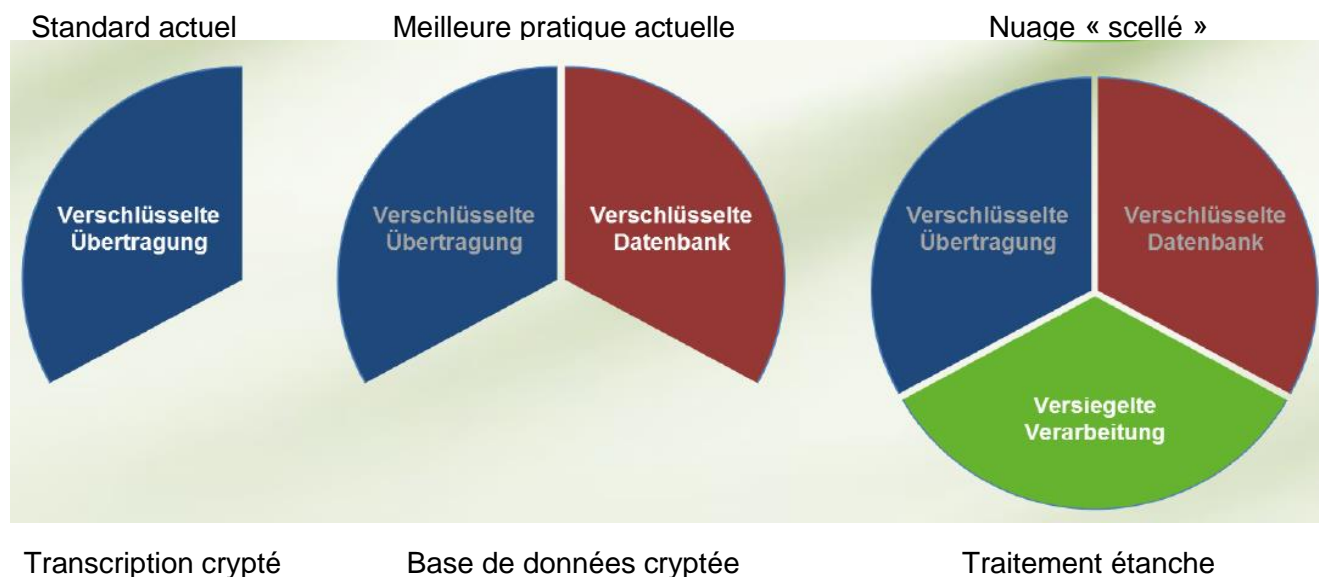
<sup>8</sup> **Numergy et Cloudwatt : Le cloud souverain made in France peut-il électrifier le marché ?** InfoDSI, mardi 8 janvier 2013, <http://www.infodsi.com/articles/137855/numergy-cloudwatt-cloud-souverain-made-in-france-peut-electrifier-marche.html>

**Cloudwatt et Numergy : forces et faiblesses des clouds souverains.** Journal du Net, 03.04.2013, <http://www.journaldunet.com/?id=1109660>

<sup>9</sup> <http://regarddejanus.wordpress.com/2012/01/03/projet-de-strategie-de-linformatique-en-nuage-des-autorites-suisse/>

<sup>10</sup> Voir : <http://www.usinenouvelle.com/article/au-cebit-uniscon-presente-le-cloud-sous-scelle.N192980> et [http://www.sealedcloud.de/?page\\_id=4](http://www.sealedcloud.de/?page_id=4)





## Faisabilité archivistique : Les propositions réglementaires et procédurales

Un certain nombre d'autorités archivistiques ont émis des recommandations quant à l'usage des services en nuages. Elles disent toutes que cet usage n'est pas exclu, mais les précautions contractuelles qu'elles proposent reviennent à exclure de fait l'utilisation de ces prestations<sup>11</sup>.

## Faisabilité archivistique : Les propositions organisationnelles

Le JISC anglais a effectué une évaluation des types de contenus hébergeables dans le nuage<sup>12</sup>. Bien que consacré aux données issues du monde académique, leur analyse peut être transposable aux autres organisations publiques. Après une analyse de l'offre disponible le rapport fourni un tableau des caractéristiques vantées du nuage (élasticité, service à la demande, accessibilité, intégrité des données, sécurité des données, optimisation des ressources, responsabilité, goulets d'étranglement) et de leur pertinence en terme de conservation. Voici quelques unes de leurs conclusions :

### Quelles sont les tâches de curation le plus viables dans le nuage ?

La capacité d'obtenir et de transférer rapidement des ressources informatiques, peut être appropriée pour les tâches peu fréquentes et intensives. Toutefois, si la tâche implique le transfert de gros volumes de données, alors le nuage devient moins approprié.

<sup>11</sup> Voir, Regard de Janus : <http://regarddejanus.wordpress.com/2011/07/17/archivistique-et-informatique-en-nuage-archiving-and-cloud-computing-premier-episode-%E2%80%93-recommandations-de-nouvelle-galles-du-sud-australie/>

<sup>12</sup> **Digital Curation and the Cloud Final Report**, Brian Aitken, Patrick McCann, Andrew McHugh, Kerry Miller , Produced by the Digital Curation Centre for JISC's Curation in the Cloud Workshop, Hallam Conference Centre 7th – 8th March 2012, [http://www.dcc.ac.uk/sites/default/files/documents/Curation-in-the-Cloud\\_master\\_final.pdf](http://www.dcc.ac.uk/sites/default/files/documents/Curation-in-the-Cloud_master_final.pdf)

**Quel modèle de service (privé-public) convient-il le mieux ?**

Le modèle public représente des risques qui font pencher vers le modèle privé, ce qui diminue les avantages économiques du cloud. Une solution infonuagique commerciale peut être utile pour les opérations par lots rares sur les données stockées ailleurs, mais elle peut être inappropriée pour le stockage de données sensibles ou personnelles.

**Quels sont les risques et les avantages ?**

Déplacer de grands volumes de données depuis et vers le nuage peut être difficile et si ce stockage est à l'extérieur de l'organisation propriétaire des données, et en particulier si elle est dans une juridiction différente, elles peuvent être exposées à d'importants risques juridiques.

**Est-ce que le nuage en-soi pose des problèmes de conservation ?**

L'adoption accrue des services en nuage à travers le paysage informatique est susceptible de conduire à la perte de données et à l'inaccessibilité dans certains cas, comme une conséquence des nombreux facteurs commerciaux et technologiques en jeu

**Faisabilité archivistique : Les propositions techniques**

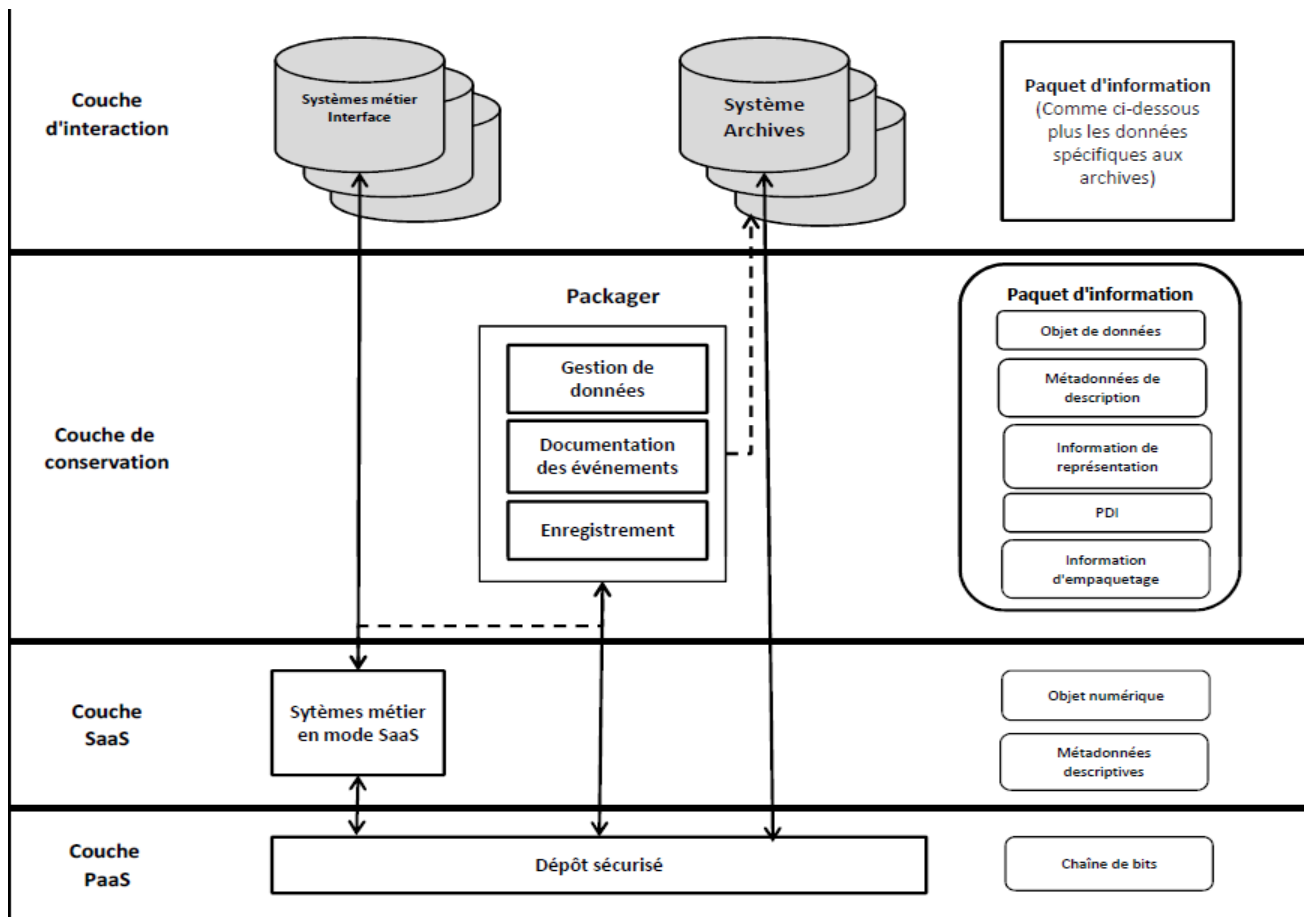
Une étude japonaise récente<sup>13</sup> sur les couches du nuage et le modèle OAIS tendrait à démontrer que les modèles sont compatibles, moyennant certaines adaptations.

Le concept en couche rend le nuage compatible avec OAIS moyennant certains aménagements. L'avantage avancé et l'existence dans un même dépôt virtuel des archives courantes et définitives, ce qui réduit le processus de versement à un simple transfert d'URI et l'ajout de métadonnées.

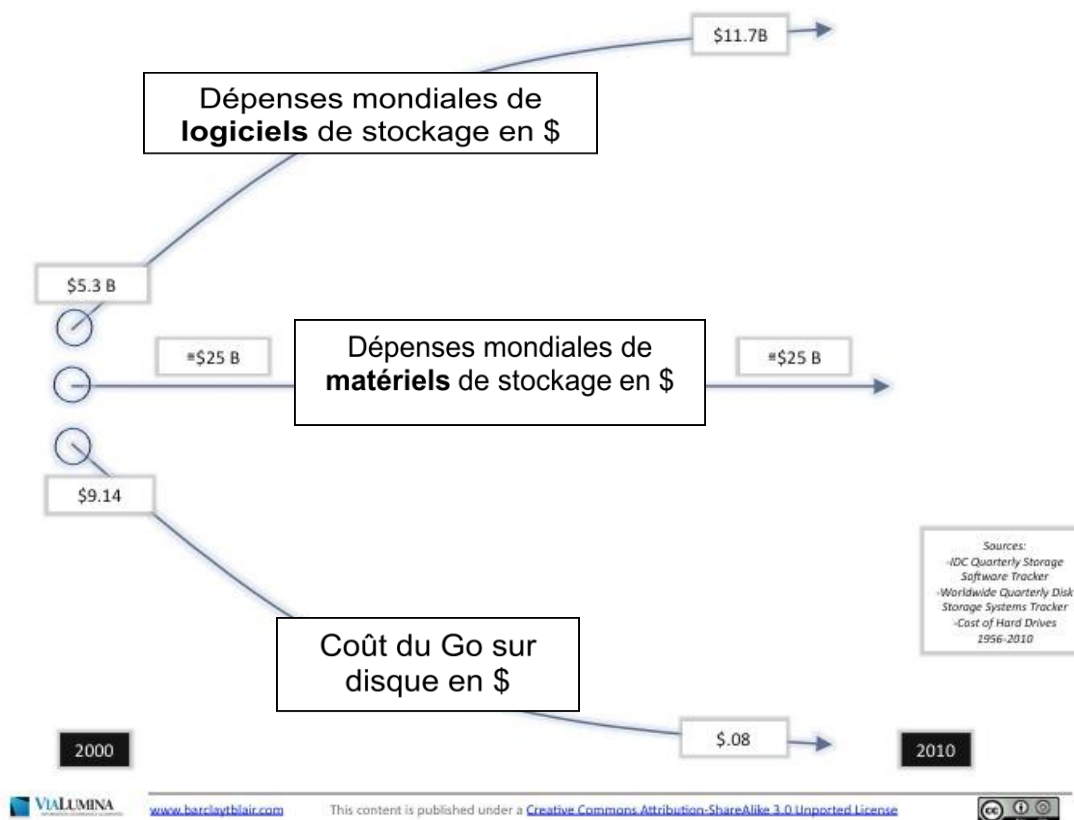
La correspondance entre les deux mondes pourrait être assurée d'une part par la création d'une couche « conservation » et d'une couche « d'interaction » en plus des couches PaaS et SaaS et par un « Packager » que l'équipe de recherche est en train de développer, selon le schéma suivant :

---

<sup>13</sup> **Preserving records in the cloud**, Jan Askhoj, Shigeo Sugimoto and Mitsuharu Nagamori, Graduate School of Library, Information and Media Studies, University of Tsukuba, Ibaraki, Japan, Records Management Journal, Vol. 21 Iss: 3 pp. 175 - 187: <http://dx.doi.org/10.1108/09565691111186858>



## La problématique des coûts



Source : IDC Quarterly Storage Software Tracker-Worldwide Quarterly Disk Storage Systems Tracker-Cost of Hard Drives 1956-2010

L'intérêt économique du cloud réside dans sa souplesse tarifaire, qui constitue un de ses principaux arguments de vente. C'est pertinent pour les données de faible durée de stockage (à cause du gain sur le non-investissement de matériel/logiciels qui pourraient être inutilisés avant la fin de l'amortissement) mais cela n'a pas de sens pour l'archivage à long terme (l'amortissement a de toute façon lieu). D'après David Rosenthal, c'est vrai dès la fin de la durée d'amortissement des disques dur, estimé environ à 4-5 ans<sup>14</sup>, c'est dire si l'archivage à long terme ne peut pas en profiter.

Ce gain est aussi pertinent pour de grande masse de données sur lesquelles on effectue des opérations conséquentes mais rarement. La puissance du nuage peut permettre de réaliser ces opérations sans avoir besoin d'investir un matériel coûteux, utilisé rarement. Par contre pour des données archivées, on accède plutôt régulièrement sur de petites parties des données stockées de manière quasi aléatoire. Ici aussi, le bénéfice de la structure tarifaire du cloud public n'est pas évident. Par contre, une mutualisation dans un cloud privé, comme celui du CINES<sup>15</sup> par exemple, constitue une piste tout à fait viable.

<sup>14</sup> Voir: David Rosenthal, **Talk at Fall 2012 CNI**, <http://blog.dshr.org/2012/12/talk-at-fall-2012-cni.html?m=1>

<sup>15</sup> La Gazette du CINES, Dossier archivage numérique pérenne, février 2013, <http://www.cines.fr/spip.php?article1057>

## Conclusions (provisoires)

En termes de **pérennité et d'intégrité**:

- En l'état, le cloud offre des fonctionnalités de **stockage**, **pas** des fonctionnalités **d'archivage**.
- L'absence de garantie de réversibilité rend quasiment impossible le transfert des données vers un autre prestataire ou le contractant (**vendor lock-in**<sup>16</sup>).

En termes de **sécurité et de traçabilité** :

- Bien que le niveau sécuritaire des prestataires soit parfois bien supérieur à celui des PME, les données restent exposées à des pertes ou des ruptures de service contre lesquelles les CGU actuelles n'offrent aucunes garanties sérieuses.
- La protection des données personnelles n'est actuellement pas garantie compte tenu de la soumission des prestataires au droit états-unien.

**Pour plus d'informations : Blog Le Regard de Janus**

[www.wordpress.regarddejanus.com](http://www.wordpress.regarddejanus.com)

Catégorie : *Informatique en nuage et cloud computing*

jean-daniel.zeller(at)hcuge.ch

---

<sup>16</sup> [http://fr.wikipedia.org/wiki/Enfermement\\_propri%C3%A9taire](http://fr.wikipedia.org/wiki/Enfermement_propri%C3%A9taire)