



Stockage numérique pas cher : mythe ou réalité ?

Réunion plénière du CR2PA du 11 octobre 2012
SANOFI, 9 rue du Président Allende – Gentilly (94)– 14h-17h

Compte rendu

Contenu

1. <i>Accueil du président</i>	1
2. <i>Présentation du projet de records management et archivage de Sanofi</i>	2
3. <i>Le stockage numérique pas cher : mythe ou réalité ?</i>	3
4. <i>Actualités du CR2PA</i>	7

1. Accueil du président

Le Président Richard Cazeneuve remercie Sanofi pour son accueil ainsi que l'assistance, nombreuse : les membres du Club, de nouvelles entreprises qui, si elles sont intéressées, nous rejoindrons sans doute, et un panel d'éditeurs qui prouvent, par leur présence, l'intérêt des échanges entre les promoteurs de l'archivage managérial et ceux qui les utilisent.

Il rappelle brièvement les objectifs du Club pour les participants qui nous rejoignent pour la première fois.

2. Présentation du projet de records management et archivage de Sanofi

par Olivier de BOISBOISSEL et Marie BESSARD

Le projet

Le projet se situe au niveau du Groupe Sanofi, constitué principalement par acquisitions successives (Rhône Poulenc, Hoechst AG, Synthélabo...).

Il existe une « Records Management Policy », des « Retention Schedules » (« tableaux de gestion »), une gouvernance *records management* (rencontre 2 fois par an).

Pour le domaine pharmaceutique, les obligations règlementaires sont très présentes avec des exigences importantes en matière de traçabilité des « records ».

Les objectifs 2012-2013 du projet :

- faire vivre la politique records management et archives;
- développer des indicateurs pour l'analyse de risque, analyser la rentabilité de l'investissement en archivage électronique;
- Renforcer la politique de destruction : il y a encore aujourd'hui trop de conservation papier ou électronique sans intérêt pour les métiers ou pour la preuve (à l'exception des archives historiques);
- développer un partenariat fort avec l'informatique : accompagner les informaticiens car le domaine de la preuve n'est pas le métier de l'IT; à noter le retard pris par les archivistes en général, absents des enjeux du « mass archiving » par exemple.
- apporter un support méthodologique aux filiales du Groupe.

Un nouveau modèle économique se met en place au sein du Groupe Sanofi ; l'orientation au plus près des clients et les marchés émergents représentent les opportunités de croissance pour demain.

Tous les ans, il est fait un effort de consolidation des coûts : opérations d'archivage électronique, opérations de destruction; analyse des problèmes et de leurs conséquences. Le Records Management reste centralisé, mais les Records Managers doivent apprendre à travailler en mode fonctionnel et transverse et dans une logique de réduction de coûts.

La politique archives est sponsorisée depuis 2008 par la Direction générale.

Notre constat en 2012 est que notre politique doit être appliquée et contrôlée beaucoup plus régulièrement, avec des indicateurs (KPI).

Il faut noter pour auditer; l'audit est indispensable au fonctionnement du système, sinon cela reste un vœu pieu, un dialogue d'experts...

Pour l'environnement électronique, on part du postulat qu'il ne doit pas être plus difficile, plus compliqué de gérer les supports numériques que le papier; c'est assurément plus complexe mais pas plus compliqué.

Questionnaire R.M. annuel auprès des entités du Groupe qui sera opérationnel en 2013 :

- avez-vous une organisation ?
- avez-vous une politique ?
- avez-vous des *retentions schedules* ?
- si oui, sont-ils à jour ?
- avez-vous un cleaning day ?
- combien de documents détruits ?
- combien de documents archivés?
- avez-vous un outil de RM?
- avez-vous réalisé un audit?
- avez-vous des contentieux?

On constate en général que plus de 50% des documents sont des « transitory records » et non des « records ». Par ailleurs, certaines durées de conservation appliquées en France sont souvent trop

longues au regard des problématiques électroniques et devront être re-questionnées (ex : 120 ans pour les dossiers du personnel).

Archivage électronique

On peut se demander comment il se fait-il qu'il n'existe pas d'outil d'archivage électronique simple, peu coûteux et visible?

L'Outil PReSERVE pour l'archivage électronique (**P**latform for **R**etention and **S**afety of **E**lectronic **R**ecords leveraging their **V**alue for the **E**nterprise) est une composante phare de la politique de records management. Le but est d'offrir une solution d'archivage unique et commune (archivage technique basé sur l'OAIS) à toutes les entités du Groupe.

PReServe doit accepter tous types de records (« document engageant et informations qui s'y rapportent » mais « record » est plus rapide), y compris les courriers et des données plus structurées; Sanofi veut centraliser l'information.

Les archives physiques sont gérées avec le logiciel Spark Archives; le module archivage électronique a été évalué au regard de Documentum.

La question qui se pose est : quel est le coût du maintien des solutions informatiques qui ne sont plus utilisés *versus* leur versement dans un outil d'archivage?

Il y a deux écoles : un outil monde ou une solution locale, avec des problèmes différents, la solution monde étant coûteuse et longue à mettre en œuvre ou la solution locale qui permet une mise en place rapide mais sur un périmètre restreint.

3. Le stockage numérique pas cher : mythe ou réalité ?

Table ronde avec

Emmanuelle OLIVIÉ-PAUL, directrice associée du cabinet d'études Markess international
Éric SOUBEYRAN, responsable solutions et applications au département informatique de Sanofi

Daniel STEVENIN, inspecteur général de RTE, président de l'AFOPE

Dominique SAHLER, expert Architecture stockage numérique Groupe PSA Peugeot Citroën

animée par Philippe GIORGI, directeur de la gestion de la connaissance et de la conservation des données de Bouygues Telecom

Le volume de données peut-il croître indéfiniment?

Dominique SAHLER : on a observé dans le groupe que les volumes ont été multipliés par 14 en dix ans; il s'agit du volume des données que l'on maîtrise, soit 15 Po (celui qu'on ne maîtrise pas est estimé à 4 fois ce volume, dans les espaces non gérés par la direction informatique). Dans le même temps, on observe une réduction du coût d'achat d'un Go de stockage : divisé par 12. Mais il faut y ajouter les coûts récurrents de fonctionnement...

Éric SOUBEYRAN : pour la pharmacie, le volume de données ne peut que croître (*row data*, molécules...) et on veut en général les conserver sur des durées longues (100 ans). Le coût du Go baisse mais le coût gestion ne baisse pas ! Toutefois, le coût dans l'outil Preserve est moins élevé que le coût de stockage dans l'application d'origine.

Emmanuelle OLIVIÉ-PAUL : les entreprises n'ont pas encore atteint la maturité face à ces questions. Tous les employés ont accès chez eux, et utilisent des services en ligne où le robinet est ouvert, de façon quasiment illimitée. La question de bien archiver pour faire des économies est une nouvelle approche.

Éric SOUBEYRAN : nous avons engagé un programme de rationalisation des *data centers* au niveau du groupe, avec un objectif de créer 3 *data centers* au lieu de 46 aujourd'hui.

Dominique SAHLER : dans l'effort de réduction des achats de Go, nous sommes aidés par les technologies, la mise en concurrence, et le *down sizing*. Nous arrivons à utiliser les supports de stockage au juste nécessaire; la technologie offre aujourd'hui des baies intelligentes qui couplent les disques bon marché peu performants avec autres plus efficaces mais également plus chers ; un système intelligent va placer les données au bon endroit en fonction de l'usage qui en est fait.

Daniel STEVENIN : les coûts de fonctionnement sont importants. Souvent, les directions générales regardent les investissements et oublient les coûts de fonctionnement qui sont pourtant prépondérants dans l'informatique. On sait réduire les coûts mais la nature a horreur du vide et, s'il n'y a pas un effort managérial constant pour bénéficier des investissements, on a des systèmes qui continuent ou recommencent à déborder. Oui, on peut faire plus avec moins, mais plutôt que d'avoir de belles politiques, il faut s'obliger aux efforts nécessaires. Les petites contraintes (blocage des boîtes mail, etc.) ne suffisent pas; il faut des actes managériaux.

Les serveurs sont-ils responsables du réchauffement climatique?

Daniel STEVENIN : le Gartner estimait il y a quelques années que les serveurs étaient responsables de 1% du CO2 sur la planète. Aujourd'hui, les *data centers* s'efforcent à des économies d'énergie, et leur impact sur le réchauffement climatique devrait diminuer.

Éric SOUBEYRAN : je ne poserais pas la question comme cela. L'archivage électronique est une nécessité, incontournable. Il y a un lien entre le stockage et la consommation : les *data centers* consomment beaucoup d'électricité. Ceci dit, on a prévu d'éloigner des *data centers* en zone froide (Chamonix); il y a certainement un impact mais il faudrait le mesurer réellement.

Y a-t-il un risque à conserver des données périmées?

Emmanuelle OLIVIÉ-PAUL : comme il a été dit, les projets d'archivage sont incontournables; on ne peut faire autrement que d'archiver car on y est obligé par la législation. En revanche, il faut estimer les données, les identifier de manière naturelle et intuitive dès le départ, car ce ne sera pas fait, sauf si on bloque les boîtes mail pour contraindre les utilisateurs. Il faut indexer, classer, organiser, ce qui doit faire gagner du temps dans le travail au quotidien. En fait, la question n'est plus aujourd'hui « Combien ça coûte d'archiver? » mais « Quelles sont les données à archiver? ». L'autre question est la lisibilité des données dans 20 ou 50 ans. Je remarque que l'on continue à utiliser les microfilms. Par ailleurs, tout ne relève pas de la conformité réglementaire, il y a aussi la recherche des données conservées, avec la question du classement et des moteurs de recherche. Il existe des solutions pour mutualiser l'archivage à durée limitée (1 ou 3 ans) et diminuer certains coûts. Enfin, on peut organiser la destruction, automatique ou non. Mais le travail d'indexation est un préalable; on ne peut mettre tous les employés dans l'obligation de trier et détruire, même s'il existe un « cleaning day ».

Y a-t-il compatibilité entre droit à l'oubli et archivage?

Daniel STEVENIN : je ne dirais pas les choses tout à fait comme ça : l'archivage est la phase finale de gestion de cycle de vie. Le droit à l'oubli est plus général car l'utilisation des technologies fait croître les risques de diffusion, surtout pour les personnes (cf la CNIL). Il faudrait qu'un certain nombre de données puissent se perdre (mais c'est plus difficile avec le numérique donc cela exige plus de rigueur). L'archivage étant la phase ultime, il faut penser la gestion des données en amont. Une des difficultés est que cette sensibilité peut se ne révéler que plus tard.

De l'homme et de la machine, qui est responsable?

Éric SOUBEYRAN : l'informaticien garde tout, même après le départ du salarié de l'entreprise, mais si on réfléchit, pourquoi garder ce qui est inutile s'il y a risque ? On revient à la question de l'interdiction des mails.

Dominique SAHLER : il existe un schéma en forme de pyramide qui montre les différentes valeurs des données : brut (au plus large), utile, alloué, affecté, utilisé, dormant (figé). La question est : qui peut prendre la responsabilité de la purge? La réponse est que c'est ingérable sauf si c'est pris en compte dans le cycle de vie de la donnée en fonction des dates définies lors de la création.

Alors, finalement, le stockage numérique est cher ou pas cher?

Emmanuelle OLIVIÉ-PAUL : jusqu'ici on stockait sans s'intéresser aux données; là, on inverse la problématique. Il y a un gros travail à faire en amont et les projets d'archivage sont une opportunité pour poser la question, jusqu'à la donnée personnelle. Il existe des solutions automatiques pour signaler que le contenu est personnel et faciliter sa destruction. Il faut réfléchir également au moyen d'optimiser les infrastructures. Les ressources des serveurs ne sont souvent utilisées qu'à 20%. Il faut repenser la manière dont les données sont stockées, en tenant compte de l'inflation des données (vidéo...). Enfin, les pratiques des individus chez eux ont un impact non négligeable car on ne pourra pas imposer un autre modèle dans l'entreprise; ou alors, si on impose quelque chose, il faudra compenser en offrant un service au quotidien.

Daniel STEVENIN : le stockage n'est pas cher si on ne stocke que des données qui ont du prix. Si les données sont stratégiques, elles doivent être archivées et stockées, on ne peut pas dire que c'est cher ou pas. Je parle pour mon entreprise qui est française.... c'est peut-être différent au plan international.

Dominique SAHLER : le stockage est cher, oui ; il faut faire évoluer les choses mais il est difficile d'expliquer aux nouvelles générations, utilisatrices d'un Youtube gratuit la problématique du coût de stockage...

Éric SOUBEYRAN : on doit lier cette question au rôle du records manager; c'est cher parce qu'il n'y a pas de gouvernance! Il faut contrôler les données à la source mais il faut reconnaître qu'on a le plus grand mal à le faire. Les technologies nouvelles permettront sans doute des gains. Aujourd'hui les serveurs virtuels sont utilisés à 100%.

Débat avec les participants

Olivier de BOISBOISSEL : je voudrais des précisions sur les courbes de PSA.

Dominique SAHLER : en données brutes, les données sont multipliées par 14 et le coût d'achat du Go est divisé par 12. Mais à chaque fois que l'on stocke, il faut ajouter les coûts de sauvegarde, d'énergie, etc. tout un ensemble de coûts de fonctionnement. Ne considérer que le coût d'achat n'est pas représentatif du coût réel des données stockées.

Olivier de BOISBOISSEL : les réflexes ne vont pas changer; on aura toujours plus à stocker car la demande de stockage va augmenter.

Dominique SAHLER : oui, avec la vidéo notamment; on a donc besoin d'une politique pour conserver au juste nécessaire. La courbe est inexorable, et elle va encore augmenter dans les 5 ans à venir.

Jean-Daniel ZELLER (Hôpitaux universitaires de Genève) : nous avons fait une étude qui montre un accroissement des archives intermédiaires de + 5% par an ces 20 dernières années, ce qui se répercute sur les archives définitives (5-10% du tout). C'est le même problème partout : + 5 % par an pour le papier; et pour l'informatique (selon les statistiques des Archives fédérales suisses), le rythme est de 15 à 20% par an. Avec l'effet de rattrapage, cela signifie que les données archivées (haute valeur) sont passées de 5 à 20%. Il faut aussi prendre en compte dans les coûts le temps d'acquisition et de traitement des données (temps humain incompressible).

Emmanuelle OLIVIÉ-PAUL : nous sommes dans une phase de transition entre le papier et le numérique. Si on prend l'exemple de la dématérialisation des factures : l'entreprise garde à la fois le papier et le numérique (entre autres, parce que la réglementation n'est pas homogène au plan européen); on mesurera les bénéfices quand il n'y aura plus du tout de papier. Aujourd'hui on cumule les inconvénients. Je reviens sur la nécessité de travailler en amont sur la donnée : il y a des données à détruire, même au bout d'une heure ! Il faut travailler sur les référentiels. D'abord le projet de gouvernance, puis l'infrastructure.

Richard CAZENEUVE : dans ces données dormantes évoquées, quel est le pourcentage de données inutiles? Pour l'élimination du papier, on vide son bureau. Mais avec le numérique aujourd'hui, les serveurs sont un déversoir virtuel et sans contrôle! Ma question : pourquoi continuer à déverser sans contrôle, alors que pour d'autres activités de l'entreprise on a institué des contrôles? L'information est partout et personne ne s'en soucie. C'est une boucle infernale! Pourquoi ne pas qualifier l'information? Tout ce qui ne couvre pas un risque et qui n'est pas doté de métadonnées qui permettent de le rechercher n'a pas d'intérêt à être conservé. C'est une fuite en avant. Olivier de Boisboissel a parlé tout à l'heure des indicateurs; c'est un problème que les records managers devront régler.

Éric SOUBEYRAN : j'approuve le propos! L'informatique a un côté séduisant; facile mais les choses évoluent; on va entrer dans un monde de restriction avec des effets sur les utilisateurs.

Emmanuelle OLIVIÉ-PAUL : pourcentage de données dormantes inutiles : 80 %...

Dominique SAHLER : dormantes signifient sans accès, ce qui justifie l'éloignement mais le fait de ne pas être utilisées ne veut pas dire pour autant qu'on peut les détruire, en tout cas l'informaticien ne peut porter cette réponse.

Emmanuelle OLIVIÉ-PAUL : il nous faut tous se mettre autour de la table; le sujet concerne tout le monde. Il n'y a pas réellement de lien aujourd'hui entre les acteurs, personne ne s'écoute...

Dominique SAINTVILLE (INA) : je ne m'occupe ni d'archivage institutionnel ni d'informatique mais des archives audiovisuelles de la radio et de la télévision, qui posent des problèmes de stockage, de sécurité, de pérennité. Nous avons en commun une problématique commune, le stockage numérique. Or, je suis frappée de ne pas avoir entendu au cours du débat les mots de « redondance » et de « migration ».

A l'Ina nous organisons la « redondance » : pour un même document, plusieurs fichiers numériques sont créés (formats de conservation, d'exploitation, de consultation), ces fichiers sont fixés sur des supports différents (serveurs, librairie robotisée, et bandes informatiques stockées sur étagère) ; il y a aussi un site miroir distant en cas de panne grave.

Par ailleurs, nous sommes contraints d'organiser des migrations périodiques pour faire face à l'obsolescence des formats numériques et aux défis de la haute définition.

Dominique SAHLER : la redondance (les répliques pour le PCA, le PRA...), oui, cela fait partie des coefficients entre « vue client » et « volume réel de données ». Sur les formats, nous n'avons pas tout à fait les mêmes questions car nous n'avons pas beaucoup de sujets vidéo chez PSA, mais en revanche, nous avons beaucoup de CAO, avec une compatibilité ascendante qui ne va plus loin que N-1. Pour réaliser les archivages long terme, il faut régulièrement migrer les données dans les dernières versions et on regarde également les possibilités de format neutre qui existent pour les données 3D et qui permettent de ne plus être lié directement au format de la version du logiciel utilisé pour les créer. C'est un peu comme lorsque l'on utilise un format PDF pour un document texte le rendant ainsi indépendant du logiciel source et visualisable sur tout type de poste.

Dominique DELATTE (Crédit Agricole) : on parle de la baisse du coût de stockage et des coûts de fonctionnement qui baissent moins, mais quel est le coût d'un archivage piloté, avec une démarche de records management ? Connaît-on le TCO [*total cost of ownership* / coût total de possession] des documents archivés correctement ? Le coût de stockage proprement dit n'est pas le coût majeur...

Olivier de BOISBOISSEL : ces indicateurs n'existent nulle part! Le coût du document archivé, le coût du repérage, les coûts de stockage papier sont noyés. Les coûts d'archivage numérique sont relatifs, selon la volumétrie. Ce qui est frappant, c'est que tout le monde tourne autour de la question mais il n'y a pas d'incitation de la DG. Je pense que le stockage va être externalisé; cela semble la bonne solution pour sortir du problème. Le stockage numérique est cher dans l'absolu mais relatif par rapport aux données, aux politiques, aux volumétries. Il faut rationaliser; maintenir les applications au juste besoin; on a identifié X application inutiles. Je ne crois pas à l'internalisation.

Marie-Anne CHABIN : ce sont les prestataires en gestion externalisée d'archives papier qui, il y a 40 ans, ont commencé à fournir les premiers indicateurs d'évaluation de l'archivage, notamment des coûts de gestion. Cela reste à faire pour le numérique.

Emmanuelle OLIVIÉ-PAUL : une entreprise sur deux à un contrat de service car il est trop lourd de gérer en interne. On note deux tendances de fond : la gouvernance de l'information, surtout dans les entreprises anglo-saxonnes (on y trouve souvent un référent pour cette gouvernance) et l'approche par l'infrastructure, surtout en France jusqu'à il y a 2-3 ans. On voit le décalage. Il faut mettre l'information avant l'infrastructure.

Françoise PHILIPPE-LAROCHE (Oséo) : l'été, on jette tout (la preuve que ça ne sert à rien...). On fait face à une tendance à la convergence d'outils du fait de la convergence d'entreprises. Les changements d'outils fonctionnels remettent en cause les questions d'archivage.

4. Actualités du CR2PA

Référentiel du CR2PA sur « l'archivage managérial »

Richard Cazeneuve présente la publication prochaine d'un **référentiel du CR2PA intitulé « L'archivage managérial »** qui expose le concept qui est le slogan du Club et qui est la version française du records management.

Le message est bien de placer l'humain en amont de la machine, en tenant compte de deux éléments qui fragilisent la démarche d'archivage : le facteur humain et le temps différé.

Le schéma présenté compare le système de gouvernance de l'information à risque dans l'entreprise à un système végétal : sans bonnes racines, il ne peut se développer; et sans branches solides, il aura du mal à résister au tsunami numérique.

Groupes de travail et divers

Marie-Anne Chabin donne des nouvelles des **groupes de travail**, notamment le GT « Benchmark » qui regroupe les entreprises: ASP, Banque de France, Cofidis, Renault, Vilogia. Ce groupe de travail a préparé un questionnaire d'enquête sur les pratiques d'archivage qui sera diffusé prochainement.

Le CR2PA, avec Clotilde Cucchi-Vignier (Total) a rencontré une délégation d'archivistes chinois de collectivités territoriales de la Région du Jiangsu, intéressée par les spécificités de la conservation des documents en entreprise: évaluation de la valeur de l'information pour une entreprise, couverture des risques liés à la non présentation des documents.

Le site web www.cr2pa.fr est en cours de rénovation.

Richard Cazeneuve invite les participants à lire, s'ils ne l'ont pas déjà fait, le recueil de billets de blog de Marie-Anne Chabin, intitulé « Sérénité et autres curiosités » qui, sous forme souvent humoristique, s'efforce de réhabiliter l'archivage.

L'archivage des mails engageants

Daniel Colas présente l'actualité de **l'archivage des mails engageants** : la doctrine du CR2PA sur les mails, y compris sur les pratiques de quotas qui ne font que déplacer le problème, a trouvé un écho ; le CR2PA est heureux de saluer la sortie de la solution Cimail-Temis-Atos qui est en ligne avec ce qu'il préconise ; on espère des développements à cette actualité.

À signaler sur le sujet le récent article de 01 informatique « Trier et archiver son courrier électronique pour être prêt en cas de litige » qui fédère les témoignages concordants d'un juriste (Jean-Emmanuel Ray), d'un éditeur (Cimail) et d'un expert en archivage (Marie-Anne Chabin).

La **prochaine rencontre du CR2PA** inaugurera une nouvelle formule avec une discussion éditeurs-utilisateurs, un débat entre maîtrise d'ouvrage et éditeurs autour des bonnes tendances pour la mise en œuvre de l'archivage managérial : c'est une innovation!

Ce sera le 4 décembre 2012, le même jour que l'Assemblée générale statutaire du CR2PA.

Compte rendu rédigé par Marie-Anne Chabin (relu et corrigé par les intervenants, diffusé le 5.11.2012).