

Le projet de préservation numérique à long terme de l'Etat luxembourgeois?

Introduction

« La préservation numérique doit être considérée comme une mission continue, qui ne se limite pas à des activités projet. Elle représente une responsabilité publique permanente. »

Cette citation est extraite de l'Appel de Berlin², une publication datant de 2013 d'une initiative d'experts du terrain et de chercheurs en matière de préservation numérique. Elle met l'accent sur un bouleversement fondamental du métier d'archiviste du 21^e siècle : il est illusoire de croire que la problématique du numérique pourrait se résoudre avec l'acte d'achat et de mise en place d'une solution informatique.

Je vais partager avec vous dans la suite de mon exposé les principaux défis soulevés par l'analyse, la conception et la mise en œuvre de la solution de préservation numérique des Archives nationales de Luxembourg ; tout en sachant que nous, archivistes, signons des deux mains l'affirmation de l'Appel de Berlin que le plus dur reste à venir par la suite.

Définitions

Je voudrais commencer par une ou plusieurs tentatives de définition de ce qu'est ou plutôt de ce que la « préservation numérique à long terme » n'est pas :

- La préservation numérique n'est pas à confondre avec l'acte de numérisation de documents papier dans le but de pouvoir les gérer de manière numérique.
- La préservation numérique n'est pas non plus à confondre avec l'acte de réaliser des copies de sauvegarde – des backups - dans le but d'éviter la perte d'informations en cas de dysfonctionnement du support.

Les désignations d'origine anglo-saxonne « *long-term digital preservation* » ou « *digital preservation* » tout court, n'ont pas de traduction française pertinente, si ce n'est « archivage électronique pérenne » - traduction notamment utilisée par les standards et normes en la matière, dont le « *Modèle de référence pour un Système ouvert d'archivage d'information (OAIS)* »³ qu'on va revoir plus en détail par la suite.

Qu'entend-on par le « long terme » ?

Selon la norme ISO 14721⁴, le « *long terme* » est défini comme étant suffisamment long pour être soumis :

- à l'impact des changements technologiques,

¹ Archives nationales de Luxembourg

² Berliner Appell zum Erhalt des digitalen Kulturerbes, Initiative Nachhaltigkeit in der Digitalen Welt, September 2013 (<http://www.berliner-appell.org>)

³ « Modèle de référence pour un Système ouvert d'archivage d'information (OAIS) » (Livre Bleu), Comité Consultatif pour les Systèmes de Données Spatiales, CCSDS 650.0B-1 (F), Mars 2005

⁴ http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

- y compris à la prise en compte de nouveaux supports et nouveaux formats de données
- ou à des changements de la communauté d'utilisateurs.

Le « long terme » peut se poursuivre indéfiniment.

Cette définition a le mérite d'être claire mais ce n'est pas forcément rassurant comme défi, en particulier pour le métier informatique pour lequel le « long terme » commence déjà au-delà de la barrière magique des 10 ans – 10 ans qui constituent une exigence légale usuelle et juste encore maîtrisable sans la mise en place d'un système de préservation.

Quels sont donc les objets qu'un tel système se devra préserver ?

Nous parlons dans ce contexte d'objets numériques – un objet numérique étant une représentation d'une unité de description – le plus souvent un document ou un dossier, composée d'une part des données elles-mêmes (le ou les fichiers) ainsi que des métadonnées y relatives.

Les fichiers sont en général des documents informatiques comme nous les connaissons tous (p.ex. des fichiers bureautiques relativement simple comme le Word, le PDF), mais également des objets plus complexes (comme p.ex. des fonds complets, des dossiers numérisés structurés, des bases de données ou des données cadastrales).

Les métadonnées sont habituellement des informations de provenance, de contexte, de référence et d'intégrité, mais peuvent contenir d'autres propriétés significatives.

Le modèle OAIS

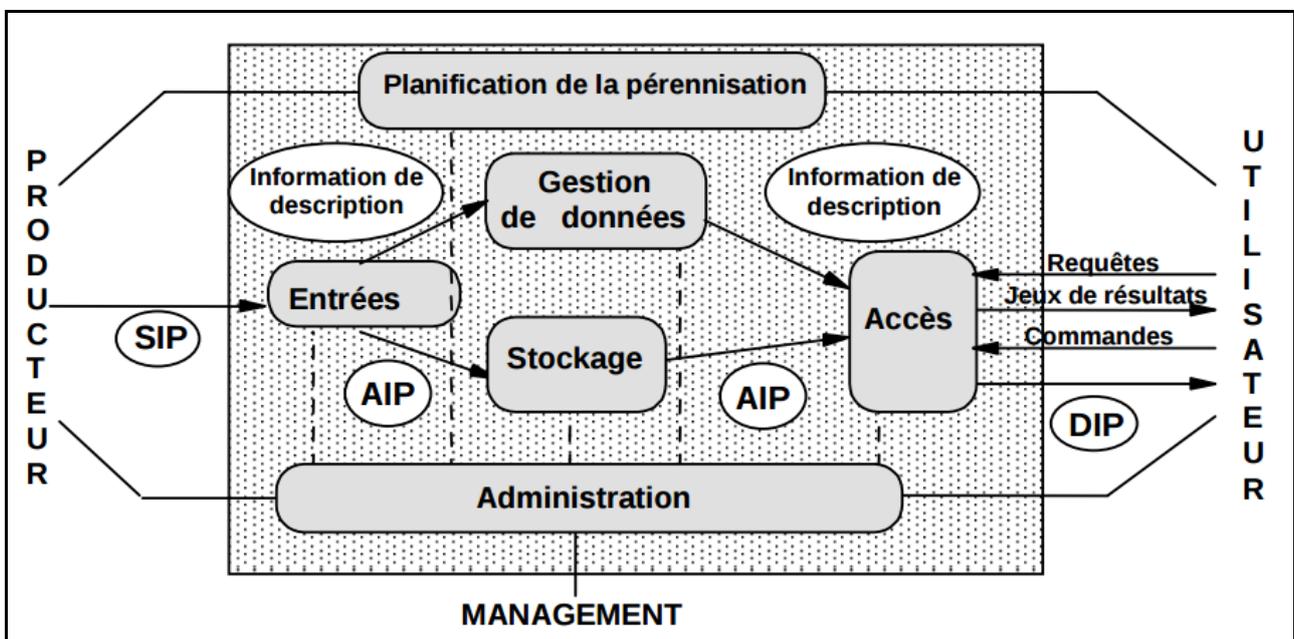


Schéma du principe de fonctionnement d'une archive électronique selon la norme OAIS

Le document « *Modèle de référence pour un Système ouvert d'archivage d'information (OAIS)* » de 2005 dont ce schéma est extrait visait à favoriser un large consensus sur les principes de fonctionnement d'une archive électronique chargée de conserver définitivement. Ce fut un succès incontestable, car à ce jour, aucun fournisseur ni aucun

consultant n'osent vous approcher sans connaître au moins les premières pages de l'OAIS. Ce qui rend cette norme également très intéressante, est le fait que le modèle peut s'appliquer à tout type de supports, même s'il a initialement été conçu pour le numérique.

L'OAIS conçoit l'archivage comme un service rendu par un prestataire à deux types de parties prenantes :

1. Le **producteur** des documents ou des fonds versés, qui peut le cas échéant se faire représenter dans cette fonction par un intermédiaire appelé service versant.
2. L'**utilisateur** qui accède les informations préservées. L'OAIS ne distingue ici pas particulièrement entre différents types de consommateurs – que ce soit le grand public ou le producteur lui-même.

Les blocs fonctionnels sont supposés opérer sous la responsabilité du service d'archives (« **management** ») :

3. Les **entrées** de documents sous forme de versement de fonds, appelé dans l'OAIS « *ingest* » comprennent le processus structuré de transfert des fonds du producteur au service d'archives.

Le versement est décomposé ensuite en deux parties : les fonds transférés d'une part et l'inventaire de ces fonds d'autre part.

- Les informations d'inventaire sont généralement stockées à part afin de permettre une gestion et des recherches plus aisées.
 - Parallèlement, les données elles-mêmes – les fichiers documents donc - sont stockées avec leurs principales métadonnées sous forme de paquets autosuffisants.
4. Un quatrième bloc fonctionnel couvre l'**accès** aux fonds, et se charge de sujets aussi divers que la recherche, la consultation ou encore la restitution, ainsi que la vérification des droits et le traçage des accès.
 5. Le cinquième pavé traite de la préservation ou **pérennisation** - concept clé de l'OAIS que nous pouvons assimiler dans un premier temps au pendant numérique du contrôle régulier de l'état de conservation et le cas échéant de la restauration des fonds. La réalité est un peu plus complexe – nous y reviendrons plus tard.
 6. Et enfin, nous avons encore un bloc de gestion **administrative**, qui n'est pas vital pour cet exposé, mais qui a sa raison d'être, car elle est supposée garantir le bon fonctionnement de l'entité, le respect des procédures et des standards en matière de sécurité des informations.

Si nous ne regardons ce schéma que du point de vue de fonds analogiques - papier ou parchemin par exemple - nous aurions tendance à dire que ceci est bien connu et maîtrisé:

- Nous recevons des versements,
- que nous ré-encodons dans nos inventaires,
- et que nous stockons ensuite dans nos dépôts,
- pour les sortir en cas de demande d'un lecteur,
- ou le cas échéant pour nous en occuper dans nos ateliers de restauration.

Pourquoi donc le numérique viendrait soudainement perturber cette mécanique bien rôdée ?

Les défis

L'information numérique est fragile et nous sommes forcés à nous en occuper en continu.

Je n'ai probablement pas à vous en convaincre, vous l'aurez tous déjà vécu à plus petite échelle :

- une disquette restée trop longtemps au soleil sur le banc de fenêtre,
- un stick USB passé au lave-linge,
- un disque dur qui commence à faire des bruits bizarres.

Tous les supports informatiques ont une durée de vie limitée due à l'usure normale ou au dommage accidentel de leurs composants. La durée de vie moyenne d'un disque dur dans un environnement professionnel est de l'ordre de 5 ans⁵. Imaginez ce que le cumul de ces risques individuels de pannes peut représenter pour une archive électronique à l'échelle de centaines de téraoctets ou de pétaoctets.

S'ajoute à ceci le phénomène de l'obsolescence⁶, à la fois en matière de supports et de formats.

Les équipements de stockage ne tombent pas uniquement en panne, mais la technologie sous-jacente a également la fâcheuse tendance à devenir obsolète. Les cycles de vie des produits techniques sont courts, surtout à l'échelle d'un archivage perpétuel. Qui d'entre nous est encore en mesure de lire des disquettes des 5.25" ou des bandes magnétiques IBM 3480⁷?

Par obsolescence des formats de fichiers, on entend la disparition d'un format au profit d'une version plus évoluée ou d'un format alternatif. Citons à titre d'exemple :

- le format WordPerfect, très populaire dans les années 90 que vous arrivez encore à lire avec un peu de chance,
- le format .DOC qui a considérablement évolué sous le même nom les dernières décennies et dont les premières versions sont quasiment illisibles avec les logiciels actuels, ou encore
- le fameux format de traitement de texte Epistole relativement populaire dans le secteur public dans les années 80, qui reste une énigme à casser.

En règle générale, des passerelles ou convertisseurs sont offerts par les éditeurs logiciels ou la communauté utilisateurs. Or, ces conversions impliquent souvent une perte d'information plus ou moins substantielle, pertes qui risquent de s'accumuler au fil des migrations.

⁵ Ensuring the Longevity of Digital Information, Jeff Rothenberg, February 22, 1999

⁶ Digital Longevity: the lifespan of digital files , Julian Jackson (<http://www.dpconline.org/events/previous-events/306-digital-longevity>)

⁷ https://en.wikipedia.org/wiki/IBM_3480_Family

N'oublions pas non plus que ces outils de conversion de format ont également une durée de vie limitée, car ils sont souvent liés à des logiciels et des systèmes d'exploitation particuliers, pour lesquels les éditeurs de logiciels aiment bien vendre de nouvelles versions.

Tout ceci, couplé à une croissance exponentielle des volumes d'informations de plus en plus exclusivement numériques, nous force à trouver des outils et méthodes pour attaquer ce défi de manière efficace.

L'approche

La meilleure stratégie à ce jour est la préservation active, c'est-à-dire une combinaison de redondance et de migration continue des supports et des formats. Cette stratégie accepte la fragilité et l'obsolescence comme des risques inhérents au numérique, et elle réplique avec un effort préventif continu, le plus automatisé et transparent possible, pour l'utilisateur.

La redondance matérielle des supports implique une duplication des informations sur de multiples systèmes locaux voire sur des sites géographiquement distants pour se prémunir contre les incidents mineurs et majeurs. Il est recommandable dans ce contexte de recourir également à des technologies et des constructeurs différents pour ne pas créer des points de défaillance couplés.

La redondance des informations implique, au-delà de la simple duplication, la gestion de multiples versions ainsi que la transformation de formats à risque dans de nouveaux formats supportés.

J'évite le terme formats pérennes. Nous considérons en effet que tous les formats de fichiers, qu'ils soient ouverts et documentés ou propriétaires, ont une espérance de vie plus ou moins longue, mais aucun d'entre eux ne peut être une solution à terme. Toute notre approche se base sur cette appréciation.

Vous auriez entretemps compris que cette approche est complexe et couteuse⁸, et nous croyons qu'il est seulement raisonnable de s'y attaquer dans une constellation d'un partenariat projet solide associée à une perspective de mutualisation conséquente.

Le partenariat projet

En 2011, trois institutions – les Archives nationales (ANLux), la Bibliothèque nationale (BnL) et le Centre des technologies de l'information de l'État (CTIE) ont décidé de mettre en commun leurs compétences respectives afin de réaliser une plateforme mutualisée de préservation numérique, afin de dégager ainsi un maximum de synergies, et de développer une solution aussi économique que possible au profit de l'ensemble des administrations de l'Etat.

⁸ Cost Model for Digital Preservation, Slats, Verdegem, Nationaal Archief of the Netherlands, 2005

En complément à une base de besoins communs qui sont la collecte, la conservation et la préservation ainsi que l'accès aux documents en question, les trois institutions ont chacune des objectifs spécifiques en fonction de leur métier.

Les ANLux souhaitent gérer

- les résultats de leurs campagnes de numérisation de fonds d'archives, et
- les versements numériques des producteurs d'archives historiques, que ce soient des documents nés numériques ou éventuellement déjà dématérialisés par le producteur.

La BnL souhaite gérer

- les résultats de ses campagnes de numérisation des contenus tombant sous leur responsabilité,
- les résultats du moissonnage du web luxembourgeois, et
- les rentrées du dépôt légal de publications numériques luxembourgeoises.

Le CTIE souhaite

- mettre en place et exploiter l'infrastructure technique mutualisée pour la solution, et
- proposer cette plateforme de préservation numérique en tant que service à l'ensemble des acteurs étatiques.

Conformément à ses besoins et contraintes spécifiques, chaque partenaire met l'accent sur un certain nombre de points d'attention particuliers qu'il couvre au bénéfice de tous dans le projet :

Les ANLux

- veillent à la compatibilité de la solution avec le projet de loi sur l'archivage,
- s'investissent dans la modélisation des structures informationnelles et des processus de versement, notamment en ce qui concerne les plans de classement, les tableaux de tri et les jeux de métadonnées communes, et
- sont évidemment particulièrement intéressés à mettre en place une gestion bout-à-bout du cycle de vie des futurs documents d'archives.

La BnL

- veille à la mise en œuvre du règlement grand-ducal sur le dépôt légal numérique,
- se focalise particulièrement sur la modélisation d'objets complexes et le traitement de formats exotiques (eBooks, pages web), ainsi que sur
- l'accès grand public aux informations.

Le CTIE

- souhaite se conformer aux exigences sécuritaires de la loi sur l'archivage légal afin de profiter des avantages en matière de valeur légale des documents archivés,

- veillera tout particulièrement à la standardisation, l'automatisation et l'économicité de la solution.

Le périmètre fonctionnel côté Archives

Quel est le périmètre de ce que les Archives nationales souhaitent mettre en place dans le projet ?

Même si l'applicabilité de la théorie des trois âges des archives dans le numérique est discutable, voire contestée par certains, nous y trouvons toujours nos repères, surtout quand il s'agit de faire des parallèles entre le système que nous sommes en train de construire et les archives papier.

- Un dossier est créé, puis clôturé au bout d'un certain temps sous la responsabilité et dans l'environnement du producteur. Nous travaillons typiquement à une granularité du dossier.
- Il devient ensuite semi-actif et passe aux archives intermédiaires, toujours sous la responsabilité du producteur. Si le dossier doit être modifié, il peut être restitué et réactivé.
- A un certain moment, la durée d'utilité du dossier vient à échéance, et une décision quant à son sort final est prise – décision qui, en principe, a été convenue d'avance au niveau du tableau de gestion. Pour simplifier la suite : le dossier est alors soit détruit soit versé aux archives historiques pour conservation indéterminée.
- Le dossier reste là-aussi consultable pour le producteur, et sous des conditions prédéfinies le cas échéant également pour le grand public. Dans des contextes spécifiques, le dossier peut même être restitué au producteur. Le séjour du dossier aux archives historiques est sous la responsabilité des Archives nationales.

Ce cycle de vie contient plusieurs événements particulièrement intéressants, sur lesquels je voudrais insister :

- La création du dossier se passe dans l'environnement du producteur. Ceci est une description volontairement vague pour une large panoplie de possibilités allant du simple serveur de fichiers jusqu'à la gestion électronique de documents, en passant par des applications métier génératrices de documents. Nous touchons ici à des environnements déjà en place pour lesquels il sera crucial, de convenir d'un jeu de métadonnées minimales communes à alimenter dès la création des dossiers ou des documents entre le producteur et l'archive.
- Pour toute la suite du cycle de vie, il est vital de fixer également d'entrée un plan de classement et un tableau de gestion.
- Le versement du dossier aux archives intermédiaires s'accompagne d'un changement d'environnement. En effet, le concept de mutualisation de la plateforme prévoit la mise à disposition d'un silo d'archives intermédiaires propre à chaque producteur sur l'infrastructure du CTIE. Il faudra donc concevoir des moyens conviviaux de transferts manuels ou automatiques, selon un certain

protocole, des dossiers électroniques par l'archiviste respectivement ses applications métier.

- L'application du tri en fin de durée d'utilité dans les archives intermédiaires débouche sur une décision relativement binaire : soit le dossier peut être détruit, soit il est à verser aux archives historiques. Le tri ainsi que les deux options de traitement qui s'en suivent sont pris en charge par le système, avec les contrôles et notifications nécessaires, sous les conditions déjà annoncées (plan de classement, tableau de tri, jeu de métadonnées minimales).
- Ce scénario de service bout-à-bout complet n'empêche pas un producteur de gérer p.ex. ses archives électroniques intermédiaires en autonomie sur sa propre plateforme hors CTIE, mais le défi de préservation se pose déjà à ce moment et est d'envergure, d'autant plus si le producteur veut potentiellement rentrer sous la couverture de la loi sur l'archivage électronique. Il y a donc à notre avis un certain intérêt à profiter du service proposé.
- La consultation du dossier versé aux archives historiques doit évidemment être possible. Afin de permettre une transparence par rapport aux futures transformations, relocalisations et autres traitements informatiques, l'équipe projet envisage entre autres l'utilisation d'identifiants et de liens persistants qui permettra un accès perpétuel aux documents, dans le respect des délais de communicabilité.

Le projet

Où en sommes-nous concrètement dans le projet?

Après une longue phase

- de sensibilisation des décideurs,
- d'étude de marché et
- de visites de référence,

le projet proprement dit a débuté en 2014 avec

- la mise en place progressive des équipes,
- la préparation d'un cahier des charges et
- la publication d'un appel à manifestation d'intérêt, suivi
- d'un appel d'offre européen début 2015 pour aboutir sur
- le choix d'un produit fin 2015.

En parallèle,

- des analyses des contextes légaux ont été menées,
- la conception des futurs processus métier et
- la définition des jeux de métadonnées standardisées communes

ont été abordées.

Nous avons également élaboré

- des documents, comme la politique d'archivage ou
- un modèle de la future convention d'archivage.

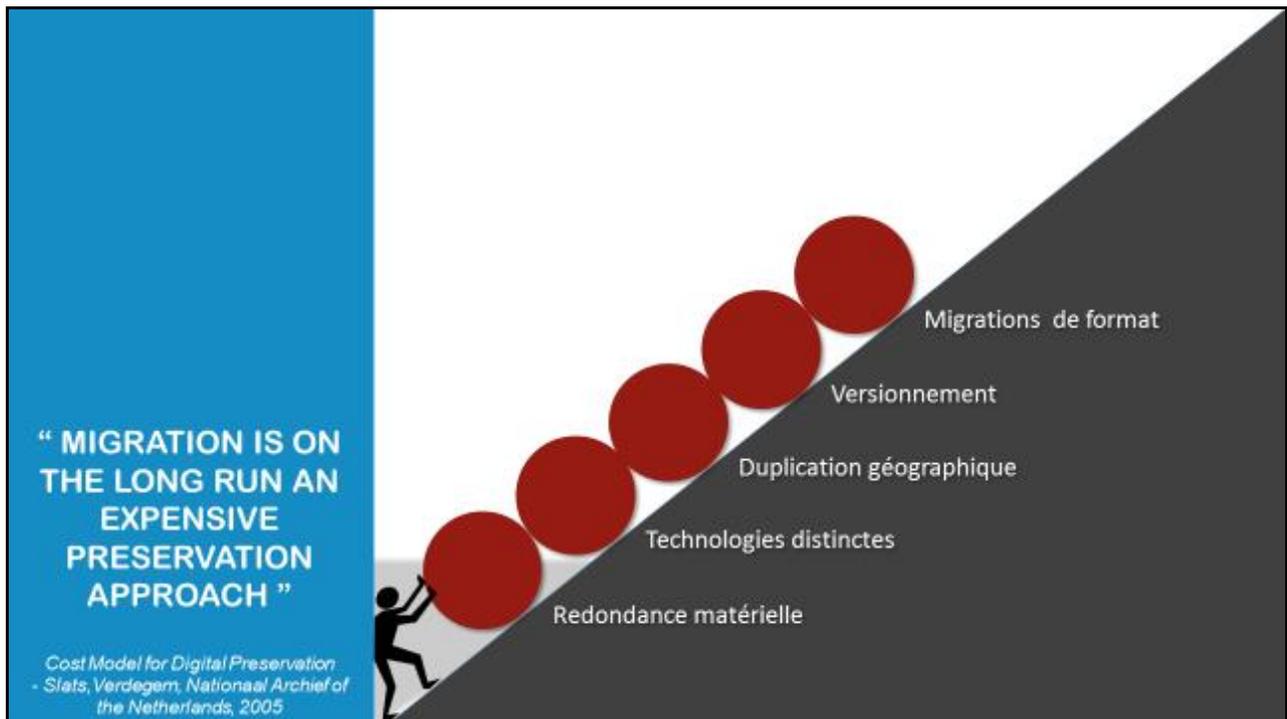
Pour l'ensemble des travaux, nous essayons de nous inspirer voire nous nous alignons largement sur les exigences normatives telles que

- l'ISO 15489 (qui traite du Records management),
- la 14721 (le fameux OAIS),
- la 30300 (la gestion de documents d'activité),
- la famille des 27000 (définissant les exigences en matière de sécurité de l'information),
- la norme AFNOR MEDONA (la modélisation des échanges de données pour l'archivage),
- et last but not least la règle technique PSDC de l'ILNAS

En tout, une équipe de 8 personnes des trois partenaires travaillent à temps plein ou partiel sur le projet, et rassemble des compétences métiers archivistique, bibliothéconomique et informatique.

Nous sommes actuellement en train de mettre en place l'infrastructure matérielle serveurs et stockage pour la mise en place du logiciel en vue de commencer mi-2016 à les adapter à nos besoins respectifs.

En parallèle, les équipes au CTIE ont commencé à aligner leurs modes de fonctionnement aux exigences sécuritaires strictes en vue d'une future certification PSDC sous la loi du 25 juillet 2015 sur l'archivage électronique.



Synthèse

Nous n'en sommes qu'aux premières phases de notre projet, et encore sans expérience opérationnelle pertinente. Cependant, déjà aujourd'hui, il est déjà clair pour toutes les personnes impliquées qu'une transformation profonde du métier de l'archiviste est en cours. Elle ne s'annonce pas, elle est bel et bien en cours.

Au cours des analyses préalables, nous avons été en contact avec un grand nombre d'institutions publiques menant ou ayant déjà finalisé de tels projets. Certains étaient à l'époque au même stade que nous, d'autres s'y sont déjà mis plus tôt et sont déjà en exploitation depuis quelques années, mais rares sont ceux qui ont des expériences pertinentes de plus de cinq ans. Nous profitons clairement du fait de ne pas être parmi les adeptes précoces.

Les modèles conceptuels, standard et normes en la matière existent. Le marché des fournisseurs et prestataires de services est en train de se structurer. Des solutions techniques commencent à prendre un niveau de maturité intéressant, et des offres à la portée des plus petites institutions dans la « *cloud* » commencent à apparaître.

Attention : les efforts d'implémentation et d'exploitation sont considérables, tant au niveau ressources humaines que financiers. Non seulement les systèmes informatiques sont complexes, mais le métier d'archiviste est dans une phase de transition, dans laquelle l'introduction du nouveau support et la volonté d'homogénéiser le traitement des dossiers papier, numériques et hybrides nécessite une mise en question des processus et des habitudes de travail en interne, mais également avec les producteurs et lecteurs.

L'archivage électronique est un travail d'équipe et nécessite une collaboration étroite de multiples profils, archivistes, analystes, informaticiens et même juristes – avec le besoin d'un vocabulaire commun, d'un chevauchement des compétences et la volonté d'une collaboration rapprochée sur plusieurs années dans des domaines où chacun, seul, ne se sent pas forcément à l'aise.

Nous sommes en train de construire une plateforme mutualisée, proposant une solution d'archivage intermédiaire et historique pour l'ensemble des acteurs du secteur public luxembourgeois. Notre montage de projet est complexe, multi-métier et ambitieux au niveau fonctionnel, et nous travaillons dans des contextes légaux non encore stabilisés.

Mais c'est justement cela qui rend le projet tellement intéressant !