

L'archivage numérique et le rôle de l'informaticien

Franklin Boumda Mouafo

Citer ce document / Cite this document :

Boumda Mouafo Franklin. L'archivage numérique et le rôle de l'informaticien. In: La Gazette des archives, n°240, 2015-4. Voyages extraordinairement numériques : 10 ans d'archivage électronique, et demain? pp. 323-334;

doi : 10.3406/gazar.2015.5312

http://www.persee.fr/doc/gazar_0016-5522_2015_num_240_4_5312

Document généré le 01/02/2018

L'archivage numérique et le rôle de l'informaticien

Franklin BOUMDA MOUAFO

Il serait complètement anachronique d'être du XXI^e siècle et de rester insensible à la révolution numérique. Dans tous les domaines, le numérique s'est fait une place, aussi minime soit-elle. L'archivage par exemple, existe depuis bien longtemps, et témoigne désormais de la force des nouvelles technologies. Cette activité a permis de construire l'histoire des peuples et les générations successives s'y appuient pour léguer un héritage à leurs successeurs. La production papier s'amenuise peu à peu, pendant que les données numériques se multiplient. Au fil du temps, la société prend conscience des facilités qu'offrent l'informatisation et l'automatisation de certains procédés, et s'ouvre progressivement à l'archivage des données numériques. Les principes et les fondamentaux qui régissent l'archivage de ce type de données sont les mêmes que ceux de l'archivage au sens classique (ou papier) que l'on connaît, mais les tâches sont spécifiques. Plusieurs entreprises, et institutions publiques s'y consacrent entièrement ou partiellement, grâce à leurs services ou leurs produits. Leurs équipes sont constituées d'archivistes et d'informaticiens qui collaborent afin de faciliter les processus, allant de la production à l'archivage des données.

Dans cet article, il sera question dans un premier temps de présenter l'archivage tel qu'il est perçu par un informaticien non-averti. Ensuite, de montrer combien son rôle est déterminant et incontournable, et de dire pourquoi un informaticien plutôt qu'un archiviste est pertinent pour réaliser certaines activités bien particulières. La dernière partie sera consacrée à la collaboration entre les deux métiers dans le cadre de l'archivage numérique.

L'archivage vu par l'informaticien

Conception erronée ou incomplète

Le nombre de services dédiés à la gestion d'un système d'archivage électronique (SAE) a considérablement augmenté au cours de ces dix dernières années. Il est difficile de dire si c'est le numérique qui s'est orienté vers l'archivage, ou plutôt si c'est l'archivage qui s'est ouvert au numérique. Dans les deux cas, on trouve une forme de franchissement que l'on pourrait assimiler à l'évolution. Prenons par exemple le cas du format PDF (*portable document format*) qui a été créé dans le but de faciliter les échanges d'informations : ce format standardisé est désormais très utilisé pour l'archivage de documents électroniques ; on parle actuellement de PDF/A et de ses dérivés.

Archivistes et informaticiens n'ont pas la même compréhension du terme archivage. Pour les premiers, c'est leur métier. Pour les seconds, sans généraliser, ils en ont une idée plus ou moins claire. On ne peut pas reprocher à l'informaticien, fraîchement sorti de sa formation, d'associer archivage à « *tar.gz* »¹ ou « *zip* ». On ne peut pas non plus lui en vouloir s'il confond sauvegarde ou stockage et archivage. Par défaut, c'est ce qu'il sait. C'est une extrapolation certes, mais c'est une réalité. Afin d'illustrer ces écrits, une enquête interne au sein d'une structure a été réalisée auprès d'informaticiens de spécialités différentes². Ils ont été répartis en deux groupes : le premier est constitué de ceux n'ayant jamais exercé dans un service d'archivage et le second, de ceux y ayant travaillé ou y travaillant actuellement.

Deux questions leur ont été posées. Premièrement : « que vous évoque le terme "archivage" ? ». À cette question, voici quelques réponses obtenues auprès du premier groupe :

- Matthieu (ingénieur support utilisateur) : « l'archivage, c'est le fait de garder/conservé des documents avec une classification (titre, date, personne, forme, informations sur le contenu, quantité) en vue d'une consultation potentielle ultérieure » ;

¹ Extension d'une archive compressée sous les systèmes Unix et dérivés (voir : doc.ubuntu-fr.org/tar).

² Au total 20 réponses ont été recueillies : dont 14 du premier groupe et 6 du deuxième groupe.

▪ Nicolas (administrateur système) : « c'est le stockage de documents dont la consultation est ponctuelle et rare (données anciennes, logs, etc.) dans un emplacement qui lui est réservé » ;

▪ Gérard (technicien réseaux informatiques) : « c'est stocker des données/fichiers à long terme, récupérables et retrouvables. Boîte archive avec classement et lexique de recherche ».

Dans ces trois réponses, on retrouve les mots : classification, tri, stocker ou son synonyme conserver. Ces réponses ne sont pas complètes mais ne s'éloignent pas de la définition. On remarque qu'ils ignorent ou du moins n'ont pas mentionné qu'il existe des règles et des normes qui régissent la pratique de l'archivage. Toutes les réponses recueillies du premier groupe sont, à quelques mots près, similaires. Il ne faudrait pas en faire une généralité. L'informaticien s'intéresse à bien d'autres domaines que le sien. Il s'agit tout simplement ici de dire comment l'archivage est perçu de manière générale par l'informaticien non-averti. L'exploitation de données numériques recommande de faire une sauvegarde de son système, de ses données et de les conserver le plus longtemps possible et de manière sécurisée. L'informaticien n'est donc pas étranger aux pratiques de l'archivage, il n'a simplement pas la définition exacte du mot.

La première question portait sur l'archivage au sens large, la seconde a été d'avantage axée sur le numérique.

Que vous évoque l'archivage numérique ?

Statistiquement, sur les informaticiens interrogés, 25 % travaillent actuellement dans un service dédié à la gestion d'un SAE (SSAE), 5 % y ont travaillé et les 70 % restant n'y ont jamais travaillé. Seuls les 30 % y ayant déjà travaillé ou y travaillant encore ont été capables de donner une réponse détaillée, complète et correcte à la question de ce que leur évoque le terme archivage. En rajoutant le mot numérique à la question, les réponses du groupe des non-avertis n'ont pas été plus détaillées que celle-ci : « c'est l'archivage des fichiers ou données numériques ». Cela s'explique par le fait que, de nos jours, il n'existe pas une formation spécialisée pour l'archivage dans le domaine de l'informatique. Pourtant les formations d'archivistes sont sensibilisées à l'existence de la spécialité archiviste numérique. L'avantage du diplômé en informatique par rapport à l'archiviste est que, hormis les tâches spécifiques à l'archivage, il a des

bases informatiques lui permettant de s'adapter. Il apprend au fur et à mesure de ses missions.

Suite à la question « comment vous êtes-vous retrouvés dans ce domaine ? », posée aux informaticiens du deuxième groupe, donc ceux ayant travaillé ou travaillant actuellement dans un service dédié à l'archivage numérique, voici quelques réponses :

- Pierre (ingénieur d'études ayant travaillé pendant dix ans au sein d'un SSAE) : « j'étais déjà dans l'entreprise au moment où le ministère a assigné cette mission comme partie de nos activités. Le document des spécifications fonctionnelles faisait quatre pages au total, c'était nouveau. J'ai découvert en y travaillant » ;

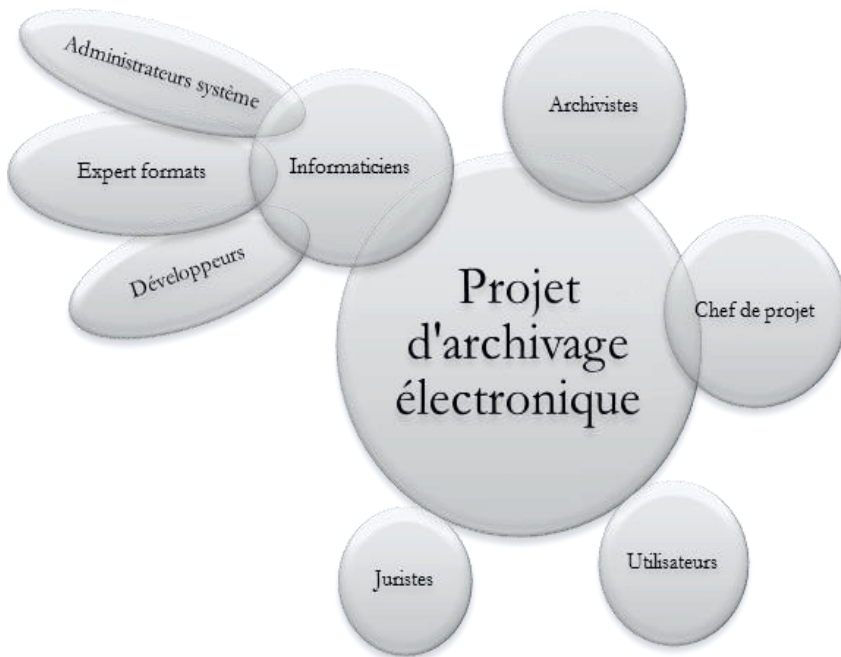
- André (ingénieur d'études travaillant depuis un an et demi au sein d'un SSAE) : « j'ai intégré ce service grâce à un concours. En lisant la fiche de poste, j'ai été attiré par l'expertise format. Le développement n'a rien de nouveau pour moi, parce que c'est ce que je fais depuis très longtemps » ;

- Benoît (développeur travaillant depuis huit ans dans un SSAE) : « à l'époque j'étais stagiaire et j'ai été recruté par la suite. J'ai trouvé l'opportunité intéressante parce qu'il s'agissait d'un sujet innovant, un défi pour la société. C'est d'ailleurs pour cela que j'ai postulé ».

Parmi les 30 % d'informaticiens, aucun n'a affirmé avoir toujours voulu exercer dans un SSAE. Ils sont issus de parcours différents et se sont retrouvés dans ce domaine par une offre d'emploi ou par concours. En d'autres termes, l'archivage numérique était nouveau pour eux, mais le fait que l'offre soit inscrite dans le domaine de l'informatique a attisé leur curiosité. Ce serait tout de même important de noter que les informaticiens ne sont pas les seuls à être peu informés sur l'archivage numérique. En fait, il existe également de nombreux archivistes pas du tout informés ou ayant peu de connaissances sur cette problématique. Une situation qui repose en partie sur le fait que ce soit une activité récente. Les originaux papiers étant encore indispensables pour de nombreuses entreprises, la démarche de dématérialisation est progressive. Ainsi, le marché de l'archivage numérique est moins important, en termes d'offres d'emploi, que le nombre d'informaticiens et d'archivistes issus des formations qualifiantes. Une situation qui s'explique par le fait que, bien qu'ils soient de plus en plus nombreux à s'y intéresser, les établissements aspirants butent sur la complexité de développement d'un projet d'archivage.

Le projet d'archivage numérique

Même si l'archivage numérique n'est pas très connu, c'est une activité qui se pratique depuis plusieurs d'années. Au CINES¹ par exemple, un département dédié – le département Archivage et diffusion (DAD) – a pour mission principale de développer, maintenir et faire évoluer les infrastructures d'archivage dont la plateforme d'archivage du CINES (PAC) lancée en 2006. Le département est constitué d'une douzaine de personnes, parmi lesquels des informaticiens et des archivistes. Les tâches sont attribuées par corps de compétences et spécialités. Plus concrètement, parmi les informaticiens on dénombre : des chefs de projets, des experts en formats de fichiers numériques, des développeurs, des administrateurs systèmes et des administrateurs de bases de données.



Profils et domaines de compétences impliqués dans un projet d'archivage électronique
©Franklin BOUMDA MOUAFO

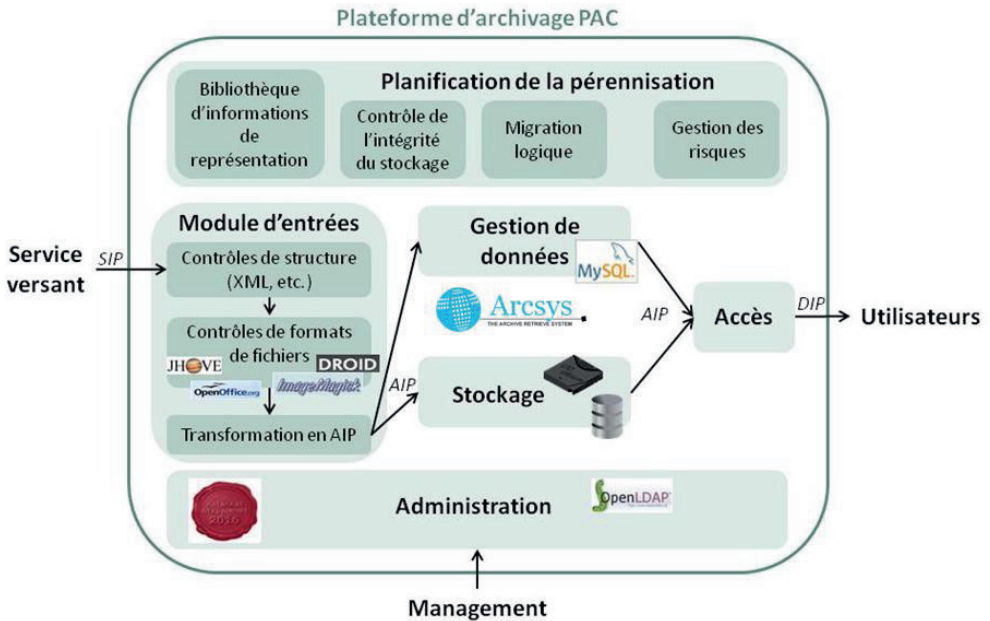
¹ Centre informatique national de l'enseignement supérieur (www.cines.fr).

Le service d'archivage numérique a débuté au CINES en 2004, suite à une décision du ministère de l'Enseignement supérieur et de la Recherche, avec pour objectif d'assurer la pérennité des thèses électroniques. Depuis, la plateforme a bien évolué et s'est étendue pour accueillir d'autres types de données que des thèses. Pour en arriver là, il a fallu comprendre l'état de l'art d'un domaine encore très récent il y a une dizaine d'années. Sur le plan national et international, public et privé, il n'existait pas de solution pouvant être adaptée à ce besoin. Rapidement, une équipe a été constituée pour d'une part définir et mettre en place les méthodes et processus métiers, et d'autre part développer la collaboration entre les archivistes, les bibliothécaires et les informaticiens. Aujourd'hui, le projet est bien consolidé, et chaque acteur a un rôle et des missions bien définis.

Le rôle de l'informaticien

L'importance du rôle de l'informaticien dans la mise en œuvre d'un projet de ce type réside dans la spécificité de ses tâches pour ledit projet. Pour cela, nous ne parlerons pas de l'informaticien en tant que chef de projet d'un SAE par exemple car tout projet implique la gestion. La nouveauté réside dans ce que fait l'informaticien, et qui ne peut être fait que par lui dans un projet d'archivage électronique.

Le schéma ci-après, à travers l'architecture qu'il présente, expose la majeure partie des activités de l'informaticien dans un projet d'archivage comme PAC. Qu'est-ce qui diffère de l'activité classique de l'informaticien ? Étant donné la criticité des données manipulées et du système à déployer, personne d'autre n'est plus approprié que l'informaticien pour mettre en œuvre une solution capable d'interagir avec un utilisateur lambda (qui sait, ou non, se servir d'une interface logicielle). C'est du développement. Il est donc clair, comme pour la plupart des logiciels de service ou de production, qu'une plateforme d'archivage nécessite deux types d'interfaces : une interface d'administration réservée à la gestion, et une interface utilisateur, plus connues sous les noms respectifs de *back-office* et *front-office*.



Architecture de la plateforme PAC ©Cines

L'expertise formats

À titre d'exemple de *front-office*, le CINES offre aux utilisateurs *via* le site Web « Facile »¹, l'opportunité de vérifier la conformité de leurs fichiers. On distingue un grand nombre de formats de fichiers numériques. Parmi ces formats, certains sont propriétaires, d'autres sont libres ; certains sont basés sur des normes et d'autres non ; certains existent depuis peu, d'autres depuis très longtemps. Certains de ces formats sont obsolètes et d'autres ne le sont pas, ou du moins, évoluent en versions. La pérennité des documents archivés dépend principalement de la capacité à les exploiter dans les années à venir. Pour cela, les normes et les règles de conformité sont implémentées sous forme d'outils numériques permettant de contrôler les fichiers. Il revient à l'expert formats d'étudier les différents formats par catégories (audio, image, vidéo, texte, scientifique, etc.) et de sélectionner ceux qu'il lui semble pertinent d'archiver au sein de son service. Cette sélection ne se fait pas par conviction. L'expert

¹ facile.cines.fr

parcourt les spécifications des formats, s'appuie également sur l'expérience de professionnels, des communautés et de scientifiques qui peuvent lui apporter suffisamment ou peu d'arguments sur un format. Il doit pouvoir répondre aux questions telles que : ce format est-il ouvert ? Quel avenir pour ce format ? Est-il suffisamment spécifié ? Etc.

Derrière le site « Facile » par exemple, tourne un logiciel (rassemblant plusieurs outils) qui contrôle les formats médias (audio, image et vidéo), les formats textes, les formats connus uniquement dans les communautés scientifiques et d'autres types de formats comme le PDF. Le développeur est capable d'associer plusieurs de ces outils entre eux, mais il faut une compétence en plus pour savoir distinguer les messages d'erreurs et les retours de ces outils. C'est le rôle de l'expertise formats. Savoir « comprendre » le format afin de répondre aux attentes, aux inquiétudes, aux questions et incompréhensions des utilisateurs du système d'archivage. C'est à lui qu'il revient d'étaler les différents choix de formats d'archivage possibles, et de présenter les avantages et inconvénients de chacun d'eux. Ce qui permettra à l'utilisateur de faire son choix.

Parfois, il peut arriver qu'il y ait des dysfonctionnements dans le code de l'application pour cause d'un cas exceptionnel de fichier. Les logiciels ou outils utilisés nécessitent d'être mis à jour ou d'être remplacés par des nouveaux qui ont des résultats plus intéressants. L'expert formats doit donc régulièrement réaliser des tests d'outils, veiller à ce que l'outil utilisé corresponde au plus performant et l'adapter si possible à son environnement de validation de fichiers. Hormis cette veille technologique fréquente, le développement de validateurs s'impose parfois. Tous les formats n'ont pas de validateur, ou en ont qui ne correspondent aux besoins. Il faut nécessairement en développer un dans ce cas. Ces situations expliquent pourquoi l'expert formats doit nécessairement développer, ou du moins savoir manipuler du code. Il y a des exceptions, mais dans la plupart des cas, les experts formats sont des informaticiens.

L'administration du système d'archivage électronique

Développer un système d'archivage électronique exige des ressources matérielles et logicielles. Il faut définir l'architecture de manière rigoureuse et confier l'administration à une personne expérimentée. Entre le service versant et le service pilote d'un SAE, il faut définir des protocoles d'échanges, prenant en compte l'aspect privé des données et un niveau de sécurité important

permettant de protéger les fichiers. C'est un aspect très important dans le fonctionnement du SAE. C'est à ce niveau qu'est choisi le logiciel d'archivage, qui réceptionne les données et qui écrit les références dans la base de données. L'administrateur assure la supervision de cette base et crée des processus de sauvegarde avec des niveaux d'alertes. L'administration d'un système d'archivage électronique se démarque de l'administration d'un système de fichiers ou d'un serveur par ce genre de tâches spécifiques, propres à l'archivage. La personne qui endosse ce rôle doit :

- gérer les migrations des archives ;
- décider des types de supports pour chaque profil d'archivage ;
- définir les protocoles de reprises des services (en cas de panne, arrêt ou de plantage) ;
- effectuer une veille technologique pour les logiciels utilisés et les mettre à jour ;
- contrôler les journaux (ou logs) ;
- définir les protocoles de connexion ;
- contrôler régulièrement les supports et l'espace de stockage disponible ;
- rédiger la documentation technique sur tous les protocoles utilisés ;
- intervenir en cas de panne et suivre les évolutions.

De manière générale, l'administration du système d'archivage électronique englobe les tâches d'administration d'un système informatique classique et des tâches plus spécifiques énumérées précédemment. Il doit comprendre le projet et bien saisir les enjeux afin de savoir comment gérer ses ressources et les données.

La gestion des risques

Dans le cadre de l'archivage classique, la prévention des risques est très importante. Ce n'est pas empêcher un risque que de mettre en place des méthodes ou un ensemble de mesures et de plans d'intervention rapides en cas de sinistre ou d'évènement nécessitant une limitation des dégradations. La gestion des risques de l'archivage numérique s'inspire du même principe. Il s'agit en quelques mots d'identifier les risques et de les évaluer ; on pourrait se permettre de terminer cette définition en rajoutant « afin d'éviter leur occurrence ». Ce ne sont pas les différences qui existent entre les données matérielles et les données numériques reconnues pour être très sensibles qui

nous importent ici. Mais, pour la fourniture adéquate d'un SAE et garantir sa fiabilité, il faut assurer sa capacité à restituer les données telles que stockées par le propriétaire. Les données numériques sont confrontées à des menaces qui engagent la responsabilité du service d'archives. Chaque profil évoqué ci-avant¹ dans un SSAE est associé à une catégorie de risques. À titre d'exemple, les formats de fichiers évoluent à la même cadence que le numérique de nos jours, et peuvent vite devenir illisibles et incompréhensibles. L'expert formats se doit donc d'assurer la veille, afin de mettre à jour, en temps et heure, ses outils, leurs versions, et intégrer cela dans son code si nécessaire. S'il y a besoin de migrer les fichiers, ce qui arrive parfois en cas d'obsolescence d'une version, il est chargé d'assurer cette tâche après accord entre le service versant et le SSAE. Hormis l'obsolescence du format, la fiabilité du système peut ou non garantir l'intégrité du fichier. L'administrateur système joue un rôle clé dans ce sens. Il doit planifier les sauvegardes, sécuriser les accès physiques au matériel et les accès logiciels au système pour que l'authenticité du fichier soit maintenue. Il se doit également de veiller à un espace disque adapté à la taille des données à archiver. La liste des risques ne se résume pas à ceux évoqués. Certains sont plus reliés à l'aspect informatique qu'à l'aspect archivistique, mais la gestion des risques interpelle la conscience de chaque membre du service. La cohésion dans les activités du SSAE est maintenue grâce à la bonne coopération de chaque corps de métier présent en son sein.

La collaboration archiviste-informaticien

Il n'est pas rare de voir l'archiviste travailler avec l'informaticien. L'archivage électronique reste d'ailleurs l'un des rares domaines, si ce n'est le seul, où ils sont obligés d'échanger leurs expertises. Sans omettre que l'informaticien a toujours besoin de comprendre les problématiques de son environnement de travail afin de s'y adapter. Cette partie n'est pas réservée au dévoilement d'une technique ou d'une méthode de travail entre l'informaticien et l'e-archiviste (archiviste numérique), mais il s'agit plutôt d'expliquer pourquoi ils sont amenés à travailler ensemble. L'informatique et l'archivage sont impliqués à parts égales. Pour créer l'architecture logicielle de la plateforme d'archivage, la fondation doit nécessairement être archivistique. L'e-archiviste prescrit les règles et les normes à respecter. Il intervient également dans la constitution des

¹ Voir schéma p. 323.

paquets à archiver et établit la liste des métadonnées qui seront à conserver avec ces paquets. Puis l'informaticien entre en action pour dire au système quels sont les éléments à vérifier dans un paquet émis par un utilisateur, s'assurer que le paquet est complet et répondre aux critères préalablement définis. Ce n'est pas intuitif car l'informaticien se doit de comprendre, un minimum, les bases de l'archivage. Il doit être informé et renseigné afin de pouvoir communiquer avec l'archiviste. De son côté, ce dernier a besoin de quelques notions informatiques, notamment sur la formalisation des métadonnées. Généralement, les schémas des bordereaux de versements sont écrits en XML. Certains archivistes ont l'opportunité de découvrir certaines de ces notions au cours de leur formation : les bases de données, les langages de manipulation de métadonnées comme RDF¹, etc. La collaboration informaticien-archiviste est même plus qu'indispensable dans un SSAE. De la mise en place aux développements, l'e-archiviste aide au respect des normes et bonnes pratiques archivistiques, l'informaticien les implémente, et ils réfléchissent ensemble aux évolutions.

Retours d'expériences

À la suite de notre enquête réalisée auprès des deux groupes d'informaticiens, nous avons recueilli des témoignages d'informaticiens et d'archivistes² travaillant ensemble sur un projet d'archivage électronique.

Alexia et Camille (e-archivistes) : « travailler sur ce projet est une première expérience de collaboration directe avec des informaticiens. La difficulté ? C'est le vocabulaire. Quand nous leur parlons de pratiques purement archivistiques, au début, ils sont contrariés, puis parviennent à apprendre avec le temps. Quand eux commencent à nous parler de développement informatique ou d'aspects techniques, nous sommes perdues. Des deux côtés, des exigences existent qu'il faut absolument respecter, ce qui n'est pas toujours évident. La solution ? Faire des glossaires pour expliquer les termes afin de faciliter la communication. Dans tous les cas, on trouve toujours une solution qui convienne à tous, et finalement, on arrive à se comprendre ».

Alexandre et Philippe (informaticiens) : « la manipulation d'un document physique et d'un document électronique n'est pas pareille. Dans tout type de

¹ *Resource Description Framework* (www.w3.org/RDF/).

² Les informaticiens qui témoignent font partie du groupe des six expérimentés auprès desquels l'enquête de la page 2 a été menée. Les deux e-archivistes sont issues de deux structures différentes.

secteur, on a besoin d'un archiviste. Dans un projet d'archivage électronique, l'archiviste doit saisir la différence entre l'accès à un document électronique et la consultation d'un document dans une bibliothèque. C'est pour cela que l'e-archiviste (ou archiviste électronique) doit nécessairement avoir quelques bases en informatique. Il doit comprendre des concepts comme le modèle relationnel par exemple, parce que la classification et l'indexation dans une boîte ne sont pas similaires au stockage sur une base de données. Pour l'informaticien, il est nécessaire de s'intéresser au vocabulaire archivistique pour progresser, pour la suite du projet. Aujourd'hui nous sommes capables de répondre aux questions des services versants, tant sur le plan archivistique qu'informatique ».

Conclusion

L'archivage électronique est une discipline récente qui permet de mettre en avant deux domaines complémentaires. L'informatisation contribue à la dématérialisation de l'archivage, qui permet à son tour d'appliquer les normes et les règles archivistiques aux fichiers numériques. Ce n'est pas une nouveauté que de voir les deux compétences impliquées travailler ensemble. Dans un autre cadre, on connaît le pouvoir envahissant du numérique. Aider à retrouver l'emplacement d'un document dans une bibliothèque ou dans un musée, est un exemple de collaboration des deux domaines. Sauf que dans le contexte présent, les deux domaines s'emboîtent : bien plus qu'une coopération, c'est un jumelage. Certaines formations d'archivistes offrent désormais des bases informatiques à leurs étudiants, une action à promouvoir. L'informaticien occupe un rôle bien plus déterminant dans un projet d'archivage électronique au même titre que l'archiviste. Dans l'avenir, l'intérêt pour l'archivage numérique sera plus important et il faudra pouvoir répondre à cette forte demande. Un prototype de système d'archivage pourrait dès lors faire son apparition au profit d'un ou plusieurs postes d'archivistes ou d'informaticiens. En attendant, il reste encore difficile d'envisager un SSAE sans ces deux compétences.

Franklin BOUMDA MOUAFO
Ingénieur d'études en informatique - Expert formats
boumson_franky@yahoo.fr