

Georg Büchler, Claire Röthlisberger-Jourdan, Martin Kaiser

Le CECO : une entreprise commune des Archives suisses pour faire face à l'archivage numérique

Contribution du CECO au Cahier No 13 de Documents (2012)

1 La genèse

La dématérialisation des administrations publiques influencera, et influence déjà, la matière première des Archives publiques : les dossiers et les documents.

Contrairement aux projets de numérisation qui sont le plus souvent le résultat d'une volonté concrète, la dématérialisation des processus administratifs est une contrainte ; il n'y a pas moyen d'y échapper. Tout service d'archives sera obligé, tôt ou tard, d'accepter des versements de documents numériques s'il veut rester fidèle à sa mission. Depuis la fin des années quatre-vingt-dix le monde archivistique a commencé à prendre conscience de cette nouvelle réalité. A l'époque, celle-ci ne faisait que se profiler à l'horizon, mais apparaissait déjà comme inévitable. Depuis, la science, les entreprises, mais aussi le monde archivistique ont commencé à y faire face.

En Suisse, un jalon important a été posé en 2002 conjointement par l'Association des Archivistes suisses (AAS) et par la Conférence des directrices et des directeurs des Archives cantonales et des Archives fédérales, ainsi que de la Principauté du Liechtenstein (CDA CH/FL) par la publication de la Stratégie globale pour la conservation à long terme des documents électroniques en Suisse. Elaborée par un collectif d'auteurs, épaulés par PriceWaterhouseCoopers, celle-ci avait pour but de dresser le point de la situation, de proposer des pistes de développement et d'établir une liste de recommandations. Parmi ces dernières se trouvait notamment la création d'un centre de coordination et de conseil. Il était incontestable que seule une collaboration étroite permettrait aux Archives (notamment petites et moyennes) d'aborder avec succès la tâche devant eux. Ainsi, les auteurs étaient formels : « L'institution d'un centre de coordination et de conseil est l'élément central des recommandations à court terme. »

La CDA n'a pas hésité à suivre cette recommandation et a engagé des négociations qui ont abouti en 2003 à la signature d'un accord administratif relatif à la création du Centre de coordination pour l'archivage de documents électroniques (CECO). En mai 2004 ce dernier a été ratifié par dix-sept cantons, la Confédération et la Principauté du Liechtenstein et est ainsi rentré en vigueur. Le CECO est entré en activité en novembre 2004 avec deux collaborateurs à temps partiel dans les bureaux des Archives fédérales à Berne qui continuent à l'héberger. Seule la moitié des cantons romands (les cantons de Vaud, Valais et Fribourg) en faisaient partie à ses débuts. Genève, Neuchâtel et le Jura ont rejoint le CECO dans la vague d'agrandissement de 2009 qui a porté le nombre de cantons membres à vingt-quatre et qui a également vu les premières villes (trois, puis cinq) adhérer à la structure. Tous les niveaux fédéraux sont ainsi représentés au sein du CECO. Il compte désormais trente-et-un membres et trois collaborateurs, ce qui correspond à deux temps complets.

2 La mission

La mission du CECO peut se résumer dans une phrase : soutenir les Archives membres dans toutes les questions que leur pose l'archivage de documents numériques. Plus concrètement cela comprend trois volets : fournir des instruments de base ; les mettre à l'œuvre dans des projets pilotes ; et collaborer directement avec les Archives membres.

2.1 Fournir des instruments de base

Avec le temps le CECO a ainsi développé une gamme de produits sur tous les niveaux de l'archivage numérique, s'étendant des standards et autres documents normatifs en passant par la réflexion théorique jusqu'aux simples outils permettant de résoudre une tâche précise. Loin d'être un guide exhaustif des prestations du CECO, cet article propose de les classer et de présenter quelques exemples particulièrement intéressants.

2.1.1 Fixer le cap : standards

Tout comme la collaboration entre Archives, la promotion et l'utilisation de standards ont depuis longtemps été considérés comme des ingrédients indispensables à l'archivage numérique. Il n'est donc guère étonnant que le CECO a été chargé, dès son origine, d'élaborer des standards. L'intérêt portait d'abord sur des questions de formats de fichiers, puis sur des exigences de base en matière d'archivage numérique et de gestion de documents. Cependant, n'étant pas un organisme de standardisation à proprement parler le CECO a dû se contenter de produire des standards et documents normatifs pour le seul cercle de ses Archives membres. Convaincus que la valeur d'un standard augmente avec sa diffusion, ces derniers ont mandaté le CECO de poursuivre l'œuvre de standardisation dans le cadre d'eCH, l'organisation suisse de normes de cyberadministration. Le groupe spécialisé eCH archivage numérique offre depuis juin 2011 le cadre pour une collaboration plus vaste entre Archives, entreprises et hautes écoles avec le but d'identifier des matières à standardiser et de développer des normes.

Exemple : le Cfa (Catalogue des formats de données d'archivage)

Le premier domaine dans lequel les Archives membres du CECO demandaient de l'assistance était celui des formats de données adaptés à l'archivage. Par conséquent, le CECO a publié, à la fin 2007, le Catalogue des formats de données d'archivage (Cfa), une œuvre réalisée en collaboration avec l'AAS. Ce catalogue dresse une liste de critères pour analyser l'aptitude à l'archivage d'un format de données et l'applique à une vingtaine de formats répandus pour en identifier les plus prometteurs. Citons, à titre d'exemple, le PDF/A, format de texte et notamment d'impression, et le SIARD, format d'archivage de bases de données développé par les Archives fédérales suisses. Le Cfa est ainsi devenu une base de décision pour les Archives, aussi bien pour l'acceptation (ou non) de formats dans des versements numériques que pour la formulation de recommandations à l'intention des administrations. La dernière mise à jour du Cfa date de la fin 2009.

2.1.2 Poser des bases théoriques : colloques et études

Tout comme l'archivage traditionnel, l'archivage numérique a besoin de solides fondations théoriques. Il n'est certes pas la tâche principale du CECO d'entreprendre

des réflexions scientifiques en matière d'archivage numérique, mais plutôt celle des universités et hautes écoles (suisses et internationales). Néanmoins le CECO est appelé à étudier certains sujets et à examiner des réflexions fondamentales autour de questions émergentes ou peu connues. Voilà pourquoi il organise des colloques pour discuter de problématiques nouvelles et qu'il en publie les résultats.

Exemple : un colloque sur l'évaluation archivistique de bases de données

L'archivage des données provenant de bases de données est un problème de l'archivage numérique fréquemment soulevé depuis quelques temps. Les applications fondées sur une base de données, appelées applications métiers, sont très répandues dans les administrations; elles sont en outre, d'un point de vue historique, relativement anciennes. Par conséquent, ces données font souvent partie des premiers versements numériques reçus par les Archives. Au-delà des problèmes techniques c'est l'évaluation archivistique de cette catégorie de documents qui interpelle les archivistes. En organisant un colloque sur l'évaluation archivistique de bases de données le CECO a tenté de rassembler plusieurs approches sur ce problème. Quelques points essentiels en ressortent. Puisque ce sont le contenu et la valeur de l'information qui intéressent en premier lieu, aucun processus ni logiciel n'est archivé. La base de l'évaluation est par conséquent une analyse détaillée du contenu de l'application spécialisée concernée, si possible en collaboration avec les offices soumis au versement. Il en résulte en règle générale un choix restreint des informations à archiver. La réduction du contenu est complémentaire à l'échantillonnage lorsqu'il s'agit de dossiers en série. Lors de l'archivage d'applications spécialisées, quelques informations concernant tous les enregistrements devraient être conservées. Pour accroître l'authenticité des informations archivées il convient de les compléter par une documentation supplémentaire du travail administratif et du processus de prise en charge archivistique, ainsi que d'utiliser le plus largement possible des interfaces et rapports de bases de données existants et employés.

Exemple : une étude des produits convertisseurs et de validation de PDF/A

Le rôle central que le format PDF/A peut jouer dans le contexte de l'archivage numérique a été souligné, entre autres, par le Cfa du CECO. Cependant la mise en œuvre s'avère difficile. Il n'est pas évident de trouver une issue du labyrinthe des différents outils qui proposent de produire ce format. C'est précisément ce que tente de faire une étude du CECO qui analyse la qualité de deux types d'outils proposés sur le marché : les convertisseurs, c'est-à-dire les logiciels qui transforment des fichiers dans des formats non-archivables comme Microsoft Word ou PowerPoint dans le PDF/A, et les validateurs, c'est-à-dire les logiciels qui vérifient la correspondance d'un fichier PDF/A avec la spécification de ce format. Cette dernière est volumineuse, et l'on peut constater qu'il est difficile d'y correspondre à 100%. Alors que la plupart des visionneuses PDF sont capables de composer avec des fichiers défectueux, l'archivage exige la plus grande correspondance possible d'un fichier avec sa spécification de format.

L'étude du CECO a démontré qu'il est important de bien définir ses exigences avant de choisir un convertisseur. S'il est correct que l'envergure et la qualité des résultats vont plus ou moins de pair avec le prix des logiciels, les exigences particulières des Archives peuvent les conduire à choisir un produit adapté à leurs besoins. En ce qui concerne les validateurs, la vitesse, la robustesse et la précision des produits testés laissent souvent à désirer. Il est néanmoins primordial de choisir un produit de qualité, à défaut de se laisser bernier par une fausse sécurité.

2.1.3 Résoudre des problèmes : outils et services

Loin des applications d'archivage numérique puissantes et coûteuses, nombre de petites Archives ont entamé l'archivage des premiers documents numériques qui sont arrivés sous leur tutelle. Tant qu'elles respectent certaines bases immuables les solutions intérimaires et ad hoc ne laissent rien craindre. Une priorisation des tâches s'impose et est raisonnable. Des outils informatiques restreints, mais spécialisés sont nécessaires pour manipuler ces documents. Là où de tels outils n'existent pas encore, le CECO s'emploie à les développer ou faire développer et à les publier sous une licence libre. Il met ainsi à la disposition de ses membres et des autres Archives un ensemble croissant d'instruments pour les assister dans la manipulation des documents numériques à archiver.

Exemple : une gamme d'outils pour manipuler le SIP

Le standard international OAIS, référence terminologique et conceptuelle centrale de l'archivage numérique, avait introduit il y a dix ans le concept du paquet d'informations à verser (SIP), l'interface entre le service versant et les Archives. Pour la Suisse, l'implémentation du SIP aux Archives fédérales était marquante. Cette spécification a été reprise et peaufinée par le CECO et fera l'objet d'une standardisation au sein du groupe spécialisé d'eCH. Autour de ce standard en devenir des outils foisonnent. Les AFS elles-mêmes ont publié le Package Handler, leur logiciel pour créer, valider, visionner et manipuler des SIP. Le CECO, lui, a développé pour ses membres SIP-Val, un outil autonome pour valider des SIP, c'est-à-dire vérifier que ceux-ci correspondent bien à la spécification. Un autre outil, nommé DI-Extractor, est en voie de développement et servira à extraire d'un SIP les métadonnées descriptives nécessaires pour peupler l'application métier des Archives. Le cas du SIP AFS/CECO/eCH montre ainsi de façon concluante aussi bien la valeur des standards (qui donnent naissance à un écosystème d'outils) que la complémentarité des composantes de l'archivage numérique (où il s'agit souvent de fermer des lacunes par un outil savamment désigné pour faire un grand pas vers une solution).

2.2 Mettre ces instruments à l'œuvre

Le succès de l'archivage se dispute dans la pratique. Il n'est pas suffisant de créer des bases théoriques, il faut les appliquer pour tester leur juste valeur. C'est la raison pour laquelle le CECO a toujours tenu à entreprendre des projets pilotes. Pour ce faire, il s'est à chaque fois joint à un petit groupe de ses Archives membres pour aborder un sujet jugé prioritaire.

Exemple : l'archivage de données de registre de l'assurance immobilière

Dans les cantons qui connaissent un établissement public d'assurance immobilière, les livres et dossiers de celle-ci constituent un fonds d'archives important. Ils renseignent sur l'histoire de bâtiments, l'urbanisation, le développement du paysage ainsi que sur des aspects économiques. Il n'est donc guère surprenant que ce sujet ait été choisi pour aborder l'archivage numérique.

Dans un premier projet pilote, suivi d'un projet plus approfondi, le CECO collabore avec les Archives de l'Etat de Lucerne, Zurich, Thurgovie, Schaffhouse et Appenzell Rhodes-Extérieures. C'est un véritable échange qui s'est développé entre les besoins du projet et les autres activités du CECO. En effet, ce projet a été l'occasion de développer et d'appliquer les instruments suivants :

- Des réflexions sur l'évaluation de bases de données ont été initialisées lors de la première phase du projet. Elles ont par la suite abouti au colloque susmentionné qui a contribué à clarifier la réflexion et à prendre des décisions fondées.
- Le Catalogue des formats de données d'archivage Cfa recommande l'utilisation du format SIARD pour l'archivage de bases de données. Cette décision a été mise en œuvre dans le projet.
- Deux outils du CECO facilitent l'utilisation de SIARD : le premier, csv2siard, transforme des fichiers depuis le format CSV (comma separated values), un format d'échange de données ancien et très répandu, au format SIARD. Cet outil est utilisé dans les cas où un export direct depuis la base de données dans le format SIARD n'est pas possible.
- Le deuxième outil est un validateur du format SIARD, SIARD-Val. Les validateurs sont cruciaux dans l'archivage numérique car ils permettent de vérifier qu'un fichier correspond le plus précisément possible à la spécification de son format.
- Si le fichier archivé est pris en charge par les Archives sous forme d'un SIP tous les outils de manipulation de SIP susmentionnés entrent en action.

Le processus adopté est esquissé dans la fig.1. Il commence par un export des contenus choisis au cours de l'évaluation archivistique depuis la base de données métier de l'assurance immobilière [1] dans plusieurs fichiers CSV [3] qui sont ensuite transformés par csv2siard [4] en un fichier d'archives au format SIARD [5].

Alternativement, un export direct depuis la base de données dans un fichier SIARD peut être implémenté. Après validation par SIARD-Val [6] le fichier SIARD est emballé dans un SIP (paquet d'informations à verser) [8] lequel à son tour doit être validé à l'aide de l'outil SIP-Val [9]. Le SIP valide peut ensuite être transféré au système de stockage [10]. Pour la consultation ultérieure plusieurs solutions sont envisageables. La plus simple consiste en une importation du fichier SIARD dans une base de données Access [11], créée ad hoc, qui permet de visualiser les contenus ou de générer d'autres fichiers de consultation, au format PDF/A par exemple [12].

Ainsi, un projet pilote comme celui qui s'occupe de l'assurance immobilière permet non seulement de voir les grandes lignes et le processus d'archivage dans son ensemble, mais offre également un contexte de test et d'évaluation des outils et autres instruments développés par le CECO. Il contribue ainsi à franchir le pas de la théorie vers la pratique.



2.3 Collaborer directement avec les Archives membres

Car c'est là le but ultime et la tâche centrale du CECO : assister les Archives membres dans l'archivage numérique. Les standards, les études et les outils qui constituent le portfolio du CECO n'ont d'utilité que quand ils sont mis en œuvre dans un service d'archives avec de réels documents d'archives. Il n'a jamais été question que le CECO enlève cette responsabilité aux Archives ; tout au plus il peut les soutenir. C'est précisément ce qu'il fait de plus en plus régulièrement. Une par une les Archives suisses se lancent dans l'archivage numérique à proprement parler. A différentes vitesses elles commencent à élaborer des stratégies et politiques, à rechercher des

solutions ad hoc pour leurs fonds numériques existants et à entrer en négociation avec leurs partenaires dans l'administration. Les outils du CECO leur apportent un soutien précieux pour certains aspects de leurs recherches. De plus les collaborateurs du CECO sont à la disposition des Archives membres pour les épauler plus concrètement. Ainsi, ils participent à des réunions entre Archives et services informatiques, relisent des rapports de stratégie, discutent du procédé à adopter dans le cadre d'un projet, servent de partenaire de réflexion ou répondent tout simplement à de petites questions. Quand cela est possible, ce soutien prend la forme d'un atelier ou d'une autre manifestation, dans le but de faire bénéficier le plus grand nombre de membres. Si cela n'est pas possible cette collaboration se fait en individuel. Ainsi, les Archives membres bénéficient de manière concrète de l'expérience acquise au cours des huit ans d'existence du CECO.

3 Conclusion

Il aura fallu au CECO divers essais pour trouver le bon équilibre dans ses prestations, tant du point de vue des méthodes de travail que des contenus. Il semble désormais que l'offre réponde de façon très précise aux attentes et besoins – ce qui n'empêchera pas le CECO de continuer à l'adapter, par exemple en intensifiant encore plus la collaboration directe avec les membres.

A travers son site web, <http://kost-ceco.ch/>, le CECO met à disposition la majeure partie de ses produits à tout service d'Archives.