arbido

2018/2 Automatisation: opportunité ou menace?

Morisod Pascal, historien et informaticien, spécialiste I+D

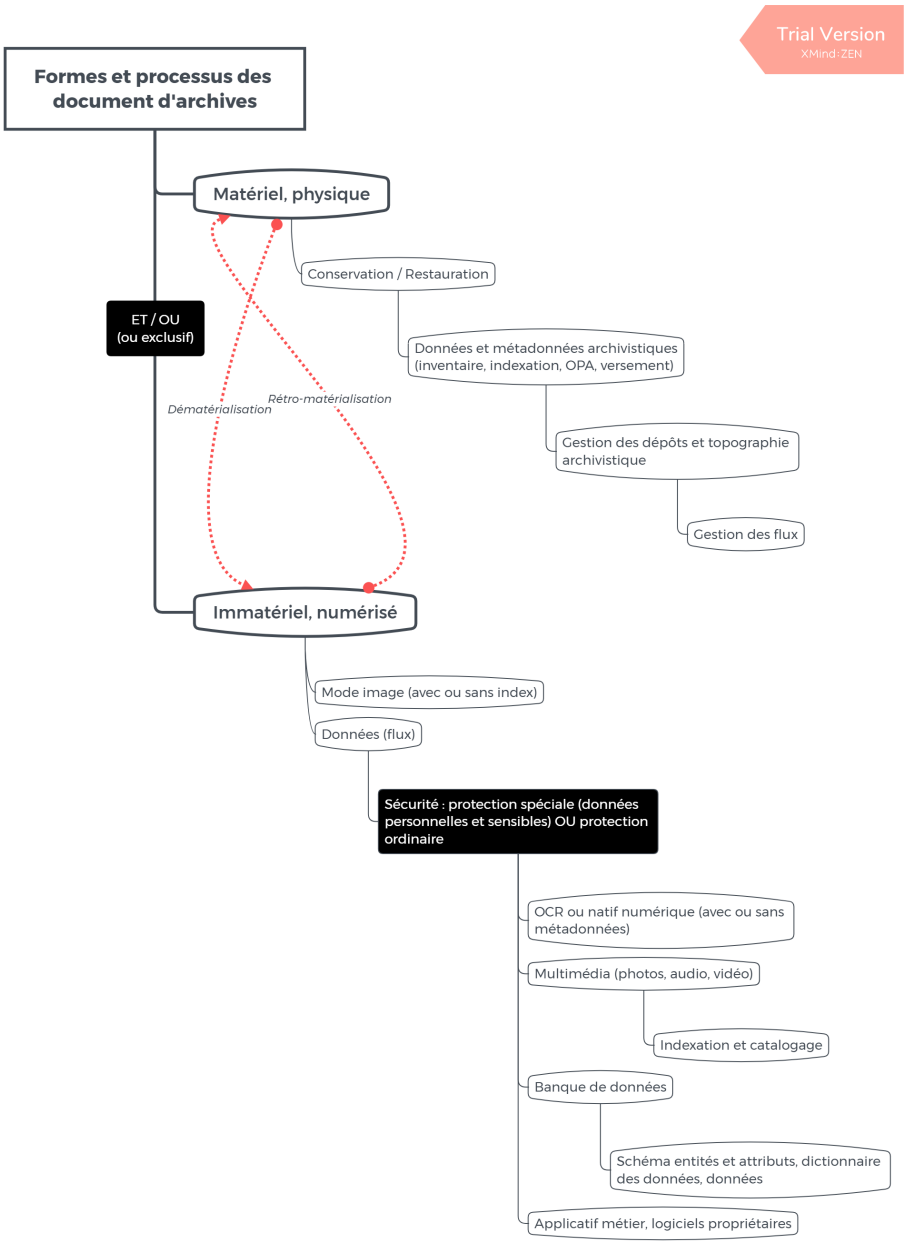
Des archives, des machines et des hommes, un heureux ménage à trois?

Avec les TIC, une nouvelle ère s'est ouverte pour les archivistes, qui font de la gestion électronique documentaire à large portée tout en valorisant leurs anciens stocks. Le document d'archives évolue et les principes inscrits dans la loi nécessitent d'être affinés en se penchant sur les différentes formes d'archives. Elles impliquent des processus de traitement variés, soutenus par des moyens technologiques à haut potentiel, qui se renouvellent sans cesse.

Des archives, mais quelles archives?

Aujourd'hui, le patrimoine archivistique est à discuter et évaluer, saisir et numériser, stocker et sécuriser, diffuser et valoriser. L'automatisation paraît incontournable pour traiter des volumes importants de documents et données, qui se mesurent en téraoctets (To), voire en pétaoctets (Po). En affirmant l'irréductibilité du support et en associant au document toute donnée informatique connexe, les archivistes se sont judicieusement laissé toutes les portes ouvertes 1. Encore faut-il les franchir...

Pour nourrir la réflexion, une ébauche de modèle pourra aider à caractériser l'aspect formel d'une archive dans une approche intégrée, afin de coller au plus près des processus de gestion et de déterminer les compétences que requière le traitement d'archives. Il permet également de se poser de nombreuses questions, au hasard: comment archiver une base de données? que faire des données propriétaires? que faire des originaux papiers face à la dématérialisation? les documents numérisés sont-ils véritablement conformes aux originaux? les données personnelles et sensibles peuvent-elle être traitées comme les autres?



La forme des documents et des données à archiver implique des processus techniques et réglementaires spécifiques faisant appel à de nombreuses compétences.

Des professionnels, mais quels professionnels?

De nos jours, l'écosystème archivistique ressemble plus à un fond marin habité de nombreux organismes vivants qu'à un abysse peuplé d'antiques créatures. Il y a davantage de diversité et de régénération, parfois moins de profondeur et de lisibilité dans la portée des actions. Voici trois catégories de compétences nécessaires à la gestion des archives:

- Les **compétences classiques**: cette catégorie regroupe les savoirs de l'historien et des spécialistes des sciences auxiliaires (héraldique, généalogie, onomastique, paléographie, sigillographie, etc.) ou connexes (latin, grec), avec des spécialisations par période, une expertise «locale» en références, que ce soit au niveau métier, ou dans les pratiques institutionnelles.
- Les **compétences modernes**: dans cette catégorie se trouvent les savoirs enseignés sous le triplet archivistique, bibliothéconomie, information documentaire. Il s'agit de compétences orientées «sciences de l'information», avec une part d'informatique appliquée.
- Les **compétences d'ingénierie**: elles regroupent les savoirs, qui peuvent être réduits à des produits, des développeurs d'applications, des éditeurs de logiciels, des services informatiques. Ce sont souvent des ressources externalisées qui nécessitent un travail collaboratif, au gré des mandats. Les applications propriétaires en font également partie.

Les compétences issues de ces trois groupes doivent pousser les acteurs à former des chaînes de compréhension verticales (entre elles) et horizontales (transversalité à l'échelle d'une administration, par exemple) afin d'assurer la qualité d'un service d'archives. La communication devient également essentielle. En ce sens, des méthodes de projets ont vu le jour, chargées de régler les échanges entre les différents partenaires et de consigner de manière structurée les étapes d'un projet. La communication joue également un rôle important au niveau pédagogique, envers le public et les partenaires, ainsi qu'au niveau médiatique, afin d'assurer une présence qui valorise les démarches.

Des technologies, mais quelles technologies?

L'année 2018 consacre une bonne part de son actualité à l'intelligence artificielle (IA), en particulier à l'apprentissage automatisé des machines (*learning machine*), rendu célèbre par des robots capables de monter un meuble Ikea[2] ou de peler un concombre[3]. L'IA s'occupe des données non structurées, discontinues, discrètes, qui sont évidemment présentes dans un service d'archives, ne serait-ce qu'en commençant par les documents manuscrits. Si l'IA offre un océan de possibilités[4], nous n'évoquons ici que quelques exemples de technologies qui peuvent potentiellement s'appliquer aux archives.

Automatisation des flux d'informations

La gestion électronique documentaire prend en charge l'entier de la production documentaire et informationnelle d'un organisme. Devant la masse des données produites, l'automatisation devient inévitable, afin de faire transiter les informations tout au long du cycle de vie des documents et des données. Les règles de conservation et de consultation fixées en amont par la doctrine archivistique doivent s'accompagner d'une gestion automatisée des flux vers l'archivage historique, ainsi que d'une gestion des accès aux archives, en plus des moyens de contrôle fournis par l'algorithmique (redondances, par exemple).

Indexation automatisée des images

Dans le cadre d'archives photographiques, la reconnaissance d'images a considérablement évolué et l'indexation classique, faite par un documentaliste qui passe ses journées à taguer des images, semble avoir atteint ses limites. Avec les développements de l'algorithmique et de l'intelligence artificielle, l'indexation automatisée ouvre de nouvelles perspectives, tout comme l'indexation induite par le web des données, capable de capter le savoir humain à distance et de classer des images à l'aide de robots-indexeur.

Un programme basé sur l'intelligence artificielle et développé par Google en 2016 est capable de générer une légende d'image de manière assez convaincante, bien que générant encore des erreurs[5]. L'algorithme est capable d'indiquer avec précision ce qu'il voit, mais également de comprendre les relations entre les différents éléments. Du côté du web des données (*linked data*), le principe consiste à décrire les données de façon structurée afin qu'elles interagissent avec d'autres services orientés web. Un des avantages pour les archives est de provoquer la rencontre de ses contenus avec le public, qui en retour peut également enrichir le service d'archives. Le retour sur investissement est réel, tout en favorisant l'interaction avec le public.

Dans le cadre de téléversement d'images sur Wikimedia Commons, les images peuvent être vues par des humains comme scannées par des robots (les bots) capables de repérer certaines parties. La diffusion de telles images dans Wikipédia est ensuite très simple, et permet de capter de l'information en provenance des internautes Wikipédiens. Dans l' exemple ci-dessous, l'image a été mise en ligne par les Archives cantonales jurassiennes, puis placée dans l'article Wikipédia «Fête des Vignerons», des informations sur le contexte de la photographie ont ensuite été éditées par un Wikipédien, qui peut être contacté au besoin.

Char de Palès, personnifiée par Hermance Taverney, lors du cortège de la fête de 1905. Photo d'Eugène Cattin.

ArCJ, 137 J 3077 a

Grâce au web des données et à l'intelligence artificielle, Wikimedia Commons et Wikipédia permettent la rencontre entre les documents d'archives et un public érudit capable de les valoriser.

Technologie blockchain et service d'archives

La technologie *blockchain*, encore lui faudra-t-elle confirmer ses premières promesses, a pour but la création d'un internet de la valeur (*internet of value*)6., qui vise à réaliser des transactions financières de n'importe quelle taille, de pair à pair (sans passer par un intermédiaire financier), sécurisées, extrêmement rapides et à moindre coût. Les réseaux blockchain peuvent se développer de manière dédiée pour des applications ou domaines bien précis, tout en offrant des caractéristiques communes. Ethereum.org, dont le siège se situe dans la crypto-vallée de Zoug, a été développé pour permettre d'adosser à des transactions financières des contrats intelligents (*smart contracts*). Cette technologie pourrait permettre de lever d'importants blocages au sein d'un service d'archives. Résumons le cas en trois points:

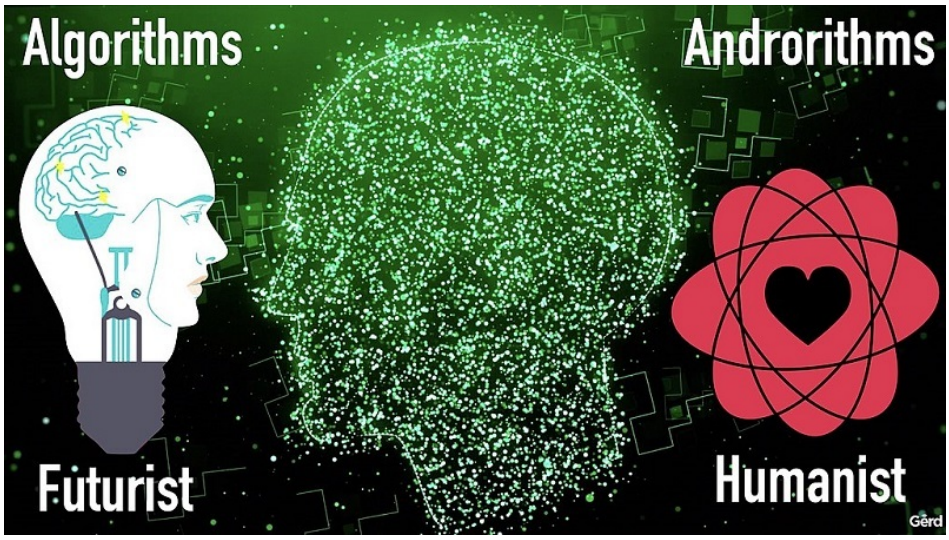
- Il est fréquent que des demandes du public proviennent d'autres continents, ou d'autres régions du pays, ou encore d'autres unités administratives;
- Les émoluments, souvent moindres, sont difficiles à percevoir, du moins ils rallongent les temps et délais de traitement d'une demande et en augmentent les coûts;
- Les archives génèrent de nombreux formulaires (*smart contracts*) et doivent se transmettre de manière sécurisée, a fortiori quand elles sont certifiées conformes.

Cette technologie pourrait dès lors grandement faciliter les échanges d'archives au niveau international comme local, de manière sécurisée, en offrant une simplification dans la perception des émoluments, ainsi qu'un gain de temps considérable dans la chaîne de traitement d'une demande.

La technologie blockchain pourrait faciliter les échanges d'archives ainsi que la perception des émoluments (détail de la page d'accueil du réseau Ethereum, une plateforme d'applications fondée sur la blockchain @ ethereum.org.).

Menace ou promesse: une question humaine

«L'intelligence artificielle est comme la force dans Star Wars: elle doit éviter le côté obscure» [7]. Il en va de même avec l'automatisation. Si elle ne sert qu'à faire des économies, si les archives sont uniquement évaluées par de monolithiques règles automatisées, si les systèmes ne sont utilisables que par une partie du public faute d'une pédagogie, si les acteurs se sentent perdus dans ce nouvel ordre, nous sommes alors assurément face à une menace. Le vrai est que tout est à conquérir: les parts de marché dans la gestion des documents suivies d'une implémentation heureuse, la qualité des projets soutenus par les bonnes technologies, une communication efficace entre les différents partenaires, une pédagogie qui permette aux utilisateurs d'intégrer les nouveaux outils, la satisfaction d'un public transgénérationnel, le renouvellement des perspectives scientifiques à l'aune du *big data*. Il revient d'abord aux acteurs de développer une vision et une stratégie, d'affirmer leurs convictions, de convaincre leurs pairs. Les conditions d'une promesse sont accomplies si les machines sont au service des humains, et non l'inverse.



Intelligence humaine simulée versus intelligence humaine sensible. Les ordinateurs raisonnent, les hommes pensent. Auteur: Gerd Leonhard.

Les éventuelles opinions et analyses exprimées dans cet article n'engagent pas l'institution dans laquelle l'auteur travaille; il s'exprime et rédige ici à titre privé. Néanmoins, l'auteur remercie Gilbert Coutaz, directeur des Archives cantonales vaudoises, pour ses commentaires avisés.

1 Voir par exemple l'article 3 de la loi fédérale sur l'archivage, dont de nombreux cantons s'inspirent.
2 Lire par exemple: http://www.lefigaro.fr/societes/2018/04/21/20005-20180421ARTFIG00051-un-robot-assemble-une-chaise-ikea-en-moins-de-dix-minutes.php.
3 Voir «La Suisse, leader en intelligence artificielle», quotidien *Le Temps*, Lausanne, 31 janvier 2018.
4 Pour un tour d'horizon, voir «ce qu'il faut attendre de l'IA» in *Archimag guide pratique* no 61, Paris: Serda, pp. 3-5.
5 Voir par exemple la section consacrée à ce programme dans Delestre, N. et Malandain, N., *Du web des documents au web sémantique*, Bois-Guillaume: Klog, pp. 61-63. Un article consacré à ce sujet, Google impressionne avec son outil de légende de photos basé sur une IA, permet de jauger l'étendue des progrès en la matière.
6 Voir par exemple l'article suivant pour plus d'informations: http://fintechnews.ch/blockcha...
7 Tiré de «J.-C. Heudin: avec l'intelligence artificielle, nombre d'applications sont devenues possibles», in *Archimag Guide pratique* 61, Paris: Serda, 2018, p. 6.

Pascal Morisod

Après une formation en Histoire et Informatique et méthodes mathématiques à l'Université de Lausanne, Pascal Morisod rejoint la jurassique et jurassienne Paléontologie A16 et l'équipe du Système d'information documentaire, en charge de la gestion du patrimoine informationnel et physique issu des creusements de la Transjurane (2009-2015). Il occupe ensuite le poste d'Archiviste cantonal adjoint du canton du Jura, en charge des projets «nouvelles technologies» (2016-2017). Depuis 2018, il est archiviste et référent TIC aux Archives cantonales vaudoises.

Résumé

Français
Avec les TIC, une nouvelle ère s'est ouverte pour les archivistes, qui font de la gestion électronique documentaire à large portée tout en valorisant leurs anciens stocks. Le document d'archives évolue et les principes inscrits dans la loi nécessitent d'être affinés en se penchant sur les différentes formes d'archives. Elles impliquent des processus de traitement variés, soutenus par des moyens technologiques à haut potentiel, qui se renouvellent sans cesse.