



HAL
open science

Document numérique : L'informatique en quête d'un corps

Aurélien Bénel

► **To cite this version:**

Aurélien Bénel. Document numérique : L'informatique en quête d'un corps. La Gazette des Archives , 2021, Du matériel à l'immatériel, 262, pp.45-60. hal-03259294

HAL Id: hal-03259294

<https://utt.hal.science/hal-03259294v1>

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Document numérique : L'informatique en quête d'un corps

Aurélien Bénel

Introduction

L'influence de l'informatique sur la pratique des archivistes est indéniable. On connaît peut-être moins celle de l'archivistique sur l'ingénierie informatique (notamment pour les systèmes d'information destinés aux professions intellectuelles). Cette influence s'explique par la quête insatiable par les informaticiens d'objets et de gestes métaphoriques, supposés plus intuitifs, à substituer à la matérialité technique de leur propre domaine, réputée absconse. Loin des discours marketing sur la dématérialisation, cette informatique, en quête d'un "corps", s'est tournée vers l'archivistique à l'aube de ce siècle, à l'occasion d'un regain d'intérêt pour la notion de "document numérique"^{1 2}.

Le 6 décembre 1999 est créé l'*Institut scientifique du document numérique* destiné à "fédérer les efforts de recherche dans ce domaine à Lyon et en Rhône-Alpes"³. Laboratoire hors-les-murs, cet institut regroupe à l'époque des chercheurs issus non seulement de plusieurs domaines de l'informatique (notamment en systèmes d'information et en reconnaissance de formes) mais également en Sciences de l'information ainsi qu'en ingénierie documentaire appliquée aux Lettres. Le tout est alors coordonné par l'École nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB). Les doctorants de cet institut, chargés de rendre compte de leurs travaux de manière synthétique, découvrent qu'ils sont obligés d'aborder le document à travers son cycle de vie⁴, pour contourner l'impossibilité de produire une définition qui fasse consensus entre toutes ces disciplines.

Dans mon domaine, celui des systèmes d'information, l'acception du terme "document" est à l'époque celle de "documentation", réactivée par la publication

¹ Roger T. Pédaque (Ed.). *La redocumentarisation du monde*. Cépaduès, 2007.

² Aurélien Bénel, Chérifa Boukacem-Zeghmouri. Le retour du document. *Archimag*, 177. Serda, 2004. p. 33–34. <https://hal-utt.archives-ouvertes.fr/hal-02955125hal-02955125>

³ Institut scientifique du document numérique, *ISDN : Présentation*, 2001. https://web.archive.org/web/20051012105512/http://isdn.enssib.fr/archives/axe3/isdn_presentation.pdf

⁴ Chérifa Boukacem, Antoine Gagneux, Aurélien Bénel, *Manifeste des doctorants de l'Institut scientifique du document numérique*, mars 2001.

de la spécification de XML (en février 1998). Cette acception remonte au moins à 1978, date à laquelle IBM proposait une “*Document composition facility*” (basée sur GML, le lointain ancêtre de XML). Les *success stories* de ce champ de l’informatique parlent d’elles-mêmes : IBM, Airbus, les communautés européennes d’alors, le CERN... Autrement dit, il s’agit de rationaliser la documentation des organisations de très grande taille, en considérant en particulier :

- le document comme instance d’un type de documents, autrement dit comme instance d’une structure générique (*Format*),
- le document comme source de documents dérivés, produits automatiquement pour un usage donné (*Transformations*)⁵.

De plus, la rencontre entre informaticiens spécialistes du traitement des textes structurés et ceux spécialistes du traitement des images nous amènent peu à peu à envisager :

- le document comme réservoir de fragments potentiels que l’on peut “tamiser” à l’aide de cribles sur mesure (*Motifs*)⁶.

Par ailleurs, la fréquentation des spécialistes de lexicométrie (comptage des mots, notamment dans le domaine des Lettres) me fait découvrir :

- le document comme répétitions de formes dont émergent des singularités statistiques (*Fouille de textes*).

Ces quatre approches du document accentuent sa *virtualité* : plus que ce qu’il est, ces approches s’intéressent à ce qu’il pourrait engendrer à la suite de traitements automatiques. Or l’automatisation qui fait partie intégrante de ce que l’on pourrait appeler “le mythe de l’informatique”, ne correspond cependant qu’à une petite partie de ce que l’informatique nous apporte dans notre travail, nos loisirs et notre vie quotidienne. Cette virtualité aurait tendance à nous faire oublier la quête de *matérialité* qui est celle de l’informaticien, de l’informaticien quand il analyse les

⁵ Serge Garlatti, Michel Crampes, Actes de la conférence “Documents Virtuels Personnalisables”, Brest, 10-11 juillet 2002.

⁶ Aurélien Bénel, Sylvie Calabretto, Véronique Eglin, Jérôme Gensel, Elisabeth Murisasco, *et al.*. Vers un CTRL +F amélioré pour tout type de document numérique ? Techniques et enjeux de la recherche de motifs. Sèdes F.; Ogier J.-M.; Marquis P. *Information, Interaction, Intelligence – Le point sur le i3*, Cepadues, 2012. p. 215–238. <https://hal.archives-ouvertes.fr/hal-01353187>

pratiques existantes des usagers avec leurs objets et leurs gestes, mais aussi de l'informaticien quand il conçoit des objets et des gestes métaphoriques à l'intérieur du logiciel chargé d'assister ces pratiques.

À l'occasion des groupes de travail interdisciplinaires de l'ISDN, les informaticiens que nous sommes prenons connaissance d'un texte célèbre pour les sciences de l'information mais qui nous était complètement inconnu : un texte de 1951, écrit par Suzanne Briet⁷, alors bibliothécaire à la Bibliothèque nationale de France et directrice de l'Institut national des techniques de la documentation. Ce texte développe un exemple audacieux : celui d'une antilope qui, dans son milieu naturel, n'avait rien d'un document mais qui va le devenir en étant sélectionnée par un explorateur comme preuve de l'existence d'une nouvelle espèce, puis ramenée au Jardin des plantes, conservée avec d'autres animaux, et qui va se voir attribuer un identifiant unique, et noter les circonstances de sa capture. On va également essayer de la classer, c'est-à-dire de la positionner par rapport aux autres documents du même corpus. En complément de ce premier processus de documentation, un second est à l'œuvre qui va générer d'autres documents à partir du premier : des photographies vont être prises de cette antilope, des articles dans la presse vont en faire la publicité, des articles scientifiques vont faire référence à cette antilope, visible au Jardin des plantes en tant que preuve. Cette preuve pose cependant le problème de sa conservation : puisqu'il s'agit d'un animal vivant, elle peut mourir. On va donc anticiper sa mort et préparer sa conservation *post-mortem* en enregistrant son cri. Par ailleurs, au moment de sa mort, on procédera à son empaillage et ainsi on substituera à l'ancien document, un nouveau. Dans cette petite histoire, outre le corps de l'antilope, apparaissent celui de l'explorateur, des employés du zoo, du photographe, des journalistes, des scientifiques, du technicien du son, du taxidermiste : ils apparaissent à travers les gestes concrets qu'ils exécutent autour de la matière que constitue l'antilope.

Plusieurs auteurs ont témoigné de la manière dont la perception brietienne du document avait déplacé leur perception du document numérique^{8 9}. Pour notre

⁷ Suzanne Briet, *Qu'est-ce que la documentation ?*, Collection de documentologie. Éditions documentaires, industrielles et techniques, Paris, 1951.

<http://martinetl.free.fr/suzannebriet/questcequeladocumentation/>

⁸ Michael K. Buckland. 1997. What Is a 'Document'? *JASIS* 48 (9), 1997. p. 804–809.

<https://people.ischool.berkeley.edu/~buckland/whatdoc.html>

⁹ Manuel Zacklad. Processus de documentation dans les Documents pour l'Action (DopA).

Actes du colloque "Le numérique : Impact sur le cycle de vie du document pour une analyse

interdisciplinaire", EBSI & ENSSIB, 2004. <https://www.enssib.fr/bibliotheque->

[numerique/documents/1209-processus-de-documentation-dans-les-documents-pour-l-action-dopa.pdf](https://www.enssib.fr/bibliotheque-numerique/documents/1209-processus-de-documentation-dans-les-documents-pour-l-action-dopa.pdf)

part, c'est à l'occasion de la création d'une unité d'enseignement sur les systèmes documentaires (en 2007) que nous avons pleinement pris conscience du fait qu'elle ouvrirait six nouvelles perspectives, bien plus incarnées que les quatre précédentes :

- le document comme preuve volontairement conservée pour être mobilisable (*Identification*),
- le document comme item, rangé et décrit notamment en fonction de sa situation de production (*Description*),
- le document comme porteur d'un contenu interprétable qui le situe par rapport aux autres documents du corpus (*Indexation matière*),
- le document comme support d'une multiplicité d'interprétations par ses lecteurs (*Annotation/Citation*),
- le document comme témoignage attesté par un tiers (*Authenticité*),
- le document comme superposition de couches d'écriture ou comme comparaison de ses variantes (*Révision*).

Plus que de simples sections dans un état de l'art, plus que différentes “écoles” traitant du document, ces dix aspects (4+6) du document sont pour nous autant de points d'attention à porter à un système documentaire, qu'il s'agisse d'analyser ce qui existe ou de concevoir un futur possible. Nous proposons dans cet article d'explorer les six points d'attention relatifs au corps du document et aux corps de ceux qui contribuent à sa conservation et son accès. La méthode qui en résulte sera mise en œuvre sur deux projets très différents : un autour des cahiers de laboratoire (destinés aux chercheurs académiques ou aux innovateurs en entreprise), l'autre autour des vitraux du département de l'Aube. Pour les besoins de l'article, ces deux projets, pourtant déjà implémentés^{10 11}, seront présentés comme “en devenir” et définis au fur et à mesure, notre but étant de placer le lecteur en situation d'écriture d'un cahier des charges.

¹⁰ Lorraine Tosi, Aurélien Bénel. L'authenticité à l'épreuve du numérique – Le cas des cahiers de laboratoire. *19ème Colloque International sur le Document Électronique*, Nov. 2016, Athènes, Grèce. <https://hal.archives-ouvertes.fr/hal-02954835>

¹¹ Aurélien Bénel. Archives numériques et construction du sens ou « Comment échapper au Web sémantique ? ». *Meta/morphoses. Les archives, bouillons de culture numérique*, *La Gazette des Archives* 245 (1), 2017. p. 163–177. <https://hal.archives-ouvertes.fr/hal-02372470>

1. Identification

Première phase de notre méthode d'analyse et de conception de systèmes documentaires, l'identification des documents est analogue au tableau que nous dresse Suzanne Briet. Il s'agit de repérer les corps en présence :

- les corps humains d'abord, contribuant, chacun dans son rôle et sa visée, à la constitution de “documents comme preuves”,
- les “corps documentaires”, ensuite, en précisant les frontières de chaque type de document.

Étude de cas n°1

La méthode doit nous éviter des approches trop hâtives et simplistes. Dans le cas des cahiers de laboratoire, comme nous allons le voir, il serait contre-productif de considérer les cahiers eux-mêmes comme les documents. Considérons d'abord les acteurs en présence : innovateurs, collaborateurs, employeur, concurrents, tribunal.

L'innovateur, sur sa paillasse ou à son bureau, note ses idées ainsi que les résultats de ses expérimentations d'abord pour lui-même. Comme sur un journal de bord, il s'agit de rendre compte, de témoigner jour après jour d'une démarche d'exploration, avec ses multiples pistes, ses impasses, ses retours en arrière, ses petites victoires. Tout doit être documenté pour ne rien oublier, au cas où il faudrait apprendre de ses erreurs, réitérer les réussites, ou reprendre des pistes délaissées.

Son employeur rassemble ses cahiers, une fois terminés. Preuves potentielles d'une invention, les cahiers sont gardés au coffre-fort. Preuves de l'antériorité d'une idée pour une personne parmi une équipe, ils permettent aussi de récompenser chacun, en fonction de sa “part d'inventivité”.

Dans le cas d'un litige entre l'employeur et un concurrent, les cahiers de laboratoire peuvent rejoindre les pièces du procès. Ils deviennent preuve d'antériorité de la possession d'une invention par l'une des parties. Ces éléments d'antériorité portent, comme nous le verrons plus tard, non sur le cahier lui-même, mais sur chacune de ses pages. En effet, chacune d'elles porte une date différente, attestée par un témoin.

Notons aussi que les corpus se superposent : les pages du cahier appartiennent, au moins séquentiellement¹², au corpus de l'innovateur, à celui de l'employeur et parfois à celui du tribunal.

Étude de cas n°2

Dans le domaine du vitrail, le département de l'Aube est le plus riche de France, avec pas moins de 9000 mètres carrés de verrières. Du point de vue de la conservation du patrimoine, le vitrail lui-même pourrait être considéré comme un document. Cependant la visée de notre projet de système documentaire n'est pas la conservation des vitraux mais la production et la diffusion des connaissances scientifiques relatives à ces vitraux (notamment en Histoire de l'art et en Histoire des religions). Quels sont les corps en présence autour de ces vitraux ?

Le premier est le bien nommé “*Corpus vitrearum*”, un organisme international rassemblant, pour la France, des personnes du ministère de la Culture et des enseignants-chercheurs, avec un double objectif de conservation et d'étude. Cet organisme s'est doté d'une méthode de numérotation des baies : du chevet (0) à la façade, nombres pairs pour le côté droit, impairs pour le côté gauche ; à partir de 100 pour le niveau supérieur et ainsi de suite. Ses membres chargés du recensement des vitraux dans l'Aube ont publié¹³ pour chacune des baies, une notice descriptive, témoignage attestant de l'état du vitrail, de sa datation, de ses techniques, de ce qu'il représente.

Viennent ensuite le corps des visiteurs déambulant dans les édifices, habitants, touristes, amateurs plus ou moins éclairés, souvent en quête de “clefs” pour mieux lire les vitraux et les apprécier, mais donnant corps à des photographies, des photographies un peu fragiles dans les appareils ou les mobiles multifonctions, mais qui prennent chair au gré des synchronisations avec leur photothèque (sur leur disque dur personnel ou celui des géants du numérique).

Enfin, le corps des historiens, dans leurs universités à plusieurs centaines voire milliers de kilomètres des vitraux qu'ils étudient ou qu'ils pourraient étudier.

¹² La question du devenir du cahier et de ses copies lorsque l'innovateur termine son contrat reste une question ouverte, même si le droit tranche en faveur de l'employeur. Dans la recherche publique où la tenue d'un cahier est souvent laissée au bon vouloir de l'innovateur, il semble difficile par exemple de lui interdire de garder une copie de son travail, sans anéantir les efforts de promotion de ces dispositifs.

¹³ Comité français du *Corpus vitrearum*, *Les vitraux de Champagne-Ardenne*, Recensement des vitraux anciens de la France, volume 4. Ed. du CNRS, 1992.

Le système documentaire à concevoir permettrait, à défaut de rapprocher les corps des visiteurs et ceux des historiens, de leur rendre accessibles, aux uns et aux autres, les photographies et leurs clefs d'interprétation comme des biens communs.

2. Description

Deuxième facette indispensable de la constitution d'un document : la description de sa situation de production. Contrairement à la “donnée” dont le terme indique que l'on se pose trop rarement la question de sa source (“Donnée par qui ?”), le document atteste : il est un témoignage (*testimonium*, comme l'appellent les philologues). Et en cela, son statut, sa portée, dépendent notamment de son auteur, de sa date, de ses destinataires. Il s'agit donc dans un premier temps, d'identifier pour chaque type de documents les indices de leur situation de production, d'imaginer d'une certaine manière la scène et les corps en présence.

À ce sujet, Manuel Zacklad parle de la nécessité, dans le processus de “documentarisation”, de munir le futur document d’“attributs standards”¹⁴. De fait, la standardisation de ces descriptions est souvent un point litigieux : les “schémas de description” sont nombreux (Dublin Core, EAD, MODS, RIC, etc.) et chacun a ses défenseurs. Il ne s'agit pas tant pour nous de prendre parti que de nous interroger sur la prétention souvent hégémonique de ces schémas et leur visée d'une compatibilité des fonds *a priori*, alors que ceux-ci n'ont pas nécessairement vocation à être combinés. Peut-être faut-il y voir une résurgence du vieux mythe de la bibliothèque universelle : ayant pris acte, avec les incendies de la bibliothèque d'Alexandrie, du caractère vain de regrouper tous les documents en un même lieu, il s'agirait, comme l'avait imaginé Paul Otlet¹⁵, de recréer une bibliothèque universelle de manière virtuelle en standardisant la description de documents pourtant conservés aux quatre coins de la planète.

Adoptons une approche pragmatique : demandons-nous pour chaque type de document, avec quels autres systèmes et donc quels autres fonds, nous souhaiterions une compatibilité et quel serait l'ensemble minimal d'attributs nécessaire pour cela.

¹⁴ Zacklad, M. Documentarisation processes in Documents for Action (DofA): The status of annotations and associated cooperation technologies. *Computer Supported Cooperative Work* 15, 2006. p. 205–228. <https://doi.org/10.1007/s10606-006-9019-y>

¹⁵ Raphaële Cornille, Stéphanie Manfroid et Manuela Valentino, *Le Mundaneum : les archives de la connaissance*, Les Impressions nouvelles, Bruxelles, 2008. p. 40-41.

Ne nous empêchons pas non plus, si c'est nécessaire, d'ajouter aux attributs standards des attributs spécifiques à notre fonds documentaire. En effet, beaucoup d'infrastructures logicielles (CMIS, OAI-PMH...) ne sont plus limitées à un schéma de description imposé, mais sont capables de combiner dans la description d'un document, des attributs issus de plusieurs schémas de description voire des attributs sur mesure.

Pour finir, nous devons concevoir les fonctionnalités associées à cette description des documents. Outre les habituelles fonctionnalités de recherche par attributs et d'affichage de ces attributs en regard du document, il nous semble indispensable de les utiliser comme critères de tri et comme liens entre documents analogues. Autrement dit, l'idée est de matérialiser la proximité qui pouvait exister entre documents dans un fonds physique. On s'attachera en particulier à trier l'affichage de nos fonds numérisés selon l'organisation des documents physiques (comme une sorte de transposition du principe archivistique de *respect des fonds* visant à la préservation du contexte original des documents).

Étude de cas n°1

Dans le cas de la réinvention du cahier de laboratoire, les attributs de description à apporter aux documents numériques versés au dossier sont assez classiques : auteurs, date, destinataires. Il est à noter cependant que dans une perspective de propriété intellectuelle :

- est considéré comme auteur celui qui dépose le document le premier dans l'archive,
- dans le cas d'une co-écriture, il faudra être capable de déterminer les contributions de chacun (comme on le verra plus loin) ,
- la date sera la plus précoce possible, idéalement celle d'un dépôt préalable à la diffusion restreinte,
- dans le cas d'une diffusion restreinte concomitante au dépôt, l'horodatage devra être suffisamment précis pour éviter toute ambiguïté vis-à-vis d'un dépôt postérieur par un destinataire,
- indiquer comme destinataires tous ceux qui ont eu accès au document permettra par exemple, avant de s'engager dans un dépôt de brevet, de s'assurer que chacun a signé un accord de non-divulgateion.

Notons qu'en filtrant les documents de l'archive en fonction d'une personne (apparaissant en tant qu'auteur ou destinataire), il est possible de reconstituer le

cahier de laboratoire de chacun. Par ailleurs, le tri par date permet de retrouver l'apparence d'un journal.

Étude de cas n°2

Le système documentaire relatif aux vitraux de l'Aube permet aux visiteurs de partager des photographies prises *in situ* (de baies entières ou de panneaux de ces baies). Ce partage, en tant que bien commun, ne doit évidemment pas se faire au détriment du respect des droits moraux du photographe. Pour cela, il est donc capital de connaître l'auteur ainsi que la date de prise de vue. L'auteur sera en général le déposant (sauf par exemple lors de l'intégration de fonds pré-existants). La date de prise de vue peut être extraite dans la plupart des cas du fichier de la photographie (sauf dans le cas de fonds anciens numérisés, par exemple les photographies des inventaires qui ont suivi la loi de 1905).

Pour que ces photographies puissent être mises en relation avec les notices du recensement, il est nécessaire par ailleurs de décrire pour chaque photographie le nom de l'édifice ainsi que le numéro de la baie.

Notons par ailleurs que pour une compatibilité avec des services de cartographie, de vue par satellite ou de vues panoramiques, il est nécessaire d'harmoniser le nom de l'édifice avec les données présentes sur ces plateformes. Les données de ces plateformes étant établies dans un premier temps à partir de l'annuaire, elles présentent souvent des imprécisions historiques (“église Saint-Rémy” au lieu d’“église Saint-Remi”) ou de localisation (l'adresse administrative d'une “église” étant celle du presbytère et non de l'édifice). Si ces données sont souvent déjà corrigées pour les édifices les plus touristiques, nous devons, pour les autres, suggérer une correction, donner les raisons de cette demande et espérer une validation de la demande.

Enfin, dans le cas spécifique d'une photographie issue d'un fonds ancien, il faudra pouvoir indiquer sa cote et ainsi garder le lien avec la matérialité du document conservé aux archives départementales.

3. Indexation matière

Le terme d'*indexation matière* est paradoxal : quoi de plus intellectuel en effet que l'exercice visant à situer un document par rapport aux autres vis-à-vis de son sujet ?

Une matérialité possible pourrait être celle des bibliothèques dans lesquelles¹⁶ on peut “explorer des rayons où se trouvent par exemple rassemblés tous les livres sur un sujet donné” et “trouver à côté du livre qu'on était allé chercher un autre livre qu'on ne cherchait pas”^{17 18}. Mais, comme l'archivistique nous l'enseigne, cette matérialité du sujet est incompatible avec celle du *respect des fonds* : elle est une transplantation dans un monde étranger (*alien world*¹⁹), celui du lecteur. En tant que telle, il est illusoire d'en faire une place unique, pérenne et consensuelle^{20 21}.

S'il y avait une matérialité à trouver à l'indexation, ce serait peut-être justement celle de l'*index*, le doigt pointé de celui qui a lu à destination de celui qui n'a pas encore lu²². Cette “indirection” est par exemple celle des fiches cartonnées qui permettent à Paul Otlet de placer les références au document dans plusieurs fichiers triés de manière différente²³, en particulier par “reversion des composants” d'un “indice composé” tel que “Les ARTS dans leurs rapports avec la LITTÉRATURE en FRANCE au XVIIe s.”²⁴ (1895-1905). En faisant cela, il ouvre la voie à ce que Ranganathan théoriserait sous le terme de “classification à facettes” (1933), dans lequel chaque document vient se placer dans un espace conceptuel à plusieurs dimensions.

Mais pourquoi se limiter à 4 ou 5 dimensions, quand on peut considérer que chaque mot-clé en est une et que l'on peut donc les combiner à l'envi ? C'est pour donner corps à cet étonnant modèle que les bibliothèques américaines se munissent dans les années 1960 de poinçonneuses et de longues aiguilles. Le principe est d'utiliser des fiches cartonnées (*edge-notched cards*) présentant au centre la référence du document (écrite à la machine ou à la main) et sur les bords une zone perforée pour chaque mot-clé possible, entaillée au poinçon quand le mot-clé est affecté au document. Il devient alors possible en introduisant manuellement

¹⁶ Grâce aux systèmes classificatoires, tels que celui de Melvil Dewey (1876), censés offrir à chaque document une place et une seule dans une arborescence de sujets.

¹⁷ Umberto Eco, *De Bibliotheca*, Trad. Éliane Deschamps-Pria, L'Échoppe, Caen, 1986. p.27.

¹⁸ Eva Sandri, La sérendipité sur internet : égarement documentaire ou recherche créatrice ?, *Cygne noir*, 1, 2013. <http://www.revuecygnoir.org/numero/article/la-serendipite-sur-internet>

¹⁹ Jeanette White Ford, *Archival principles and practice : a guide for archives management*. McFarland, 1990.

²⁰ Georges Perec, Notes brèves sur l'art et la manière de ranger ses livres, *Penser/Classer*, Hachette, 1985.

²¹ Clay Shirky, Ontology is Overrated: Categories, Links, and Tags, *O'Reilly Emerging Technology Conference [ETech'2005]*, San Diego, March 14-17, 2005.

²² Laure Amélie Guitard, Indexation par sujet en archivistique et en bibliothéconomie : du pareil au même ? *Documentation et bibliothèques* 59 (4), 2013.

²³ Raphaële Cornille *et al.*, *Op. cit.* fig. p. 24.

²⁴ *Ibid.*, fig. p. 36.

une aiguille dans l'une des perforations de retirer d'un ensemble de fiches toutes celles qui ne sont pas affectées au mot-clé correspondant. En jouant sur le paquet choisi ainsi que sur la répétition de l'opération, il devient même possible de matérialiser les opérateurs dits “booléens” (ET, OU, NON). Non sans humour, certains surnomment ce dispositif un “ordinateur – aiguille à tricoter” (*knitting needle computer*).

Dans ces trois premiers modèles d'indexation (classificatoire, à facettes et combinatoire), qu'ils soient “pré-coordonnés” (en imposant une hiérarchie de classification) ou “post-coordonnés” (en imposant une liste d'autorité et des règles de combinaison), il s'agissait de contrecarrer le caractère supposé arbitraire de toute lecture. Une autre approche revient à prendre acte de cette diversité de lectures et d'interprétations : si l'indexation matière est un “point d'accès”²⁵, chaque lecture en tant qu'accès vient ouvrir un nouveau sentier ou rend un chemin existant un peu plus visible (ce qui n'est pas sans rappeler la métaphore d'Origène à propos de l'interprétation des Écritures²⁶). De cette approche, le Web 2.0 (appelé aussi “Web participatif”) nous propose une version “pop” : les “folksonomies” (mot-valise associant *taxonomie* et *folks*, les gens, le peuple). Chacun peut laisser sur un document, en son nom propre, un ensemble d'étiquettes (*tags*), c'est-à-dire des mots librement choisis mais réutilisables d'une personne à une autre afin de “faire nombre” (y compris de manière visuelle par la taille des *tags* dans des “nuages de *tags*”). Dans la plupart des systèmes documentaires, le nombre de lecteurs est trop faible pour espérer cet effet de masse, qualifié parfois de “sagesse des foules”, que l'on a pu observer sur les plateformes de partage de photographies ou de signets. Par contre, pour des fonds beaucoup plus spécialisés, ce type de dispositifs a pour intérêt de mettre en visibilité la figure de l’“amateur éclairé” et le produit de son travail. Du côté des archives, par exemple, on peut penser aux plateformes contributives qui permettent aux généalogistes de repérer et de partager les noms propres dans les pièces d'archive qu'ils ont sous les yeux, afin de faciliter les recherches entreprises par d'autres généalogistes²⁷.

Nous l'avons compris, l'indexation matière est comme un doigt pointé par un lecteur à destination d'un futur lecteur. Le concepteur de système documentaire

²⁵ Laure Amélie Guitard, *Op. cit.*

²⁶ Origène, *Traité des principes*, IIIe s. (trad. M. Harl, G. Dorival, A. Le Boulluec, Études augustiniennes, 1976).

²⁷ Archives de l'Aube, *MyArchives, Présentation du projet*, Services numériques culturels innovants. Ministère de la culture, 2010. <https://www.culture.gouv.fr/Media/Thematiques/Enseignement-superieur-et-recherche/Archives-anciens-programmes/Fiches-projets-CultureLabs/Projets-CultureLabs-2010/MyArchives-Archives-departementales-de-l-Aube>

doit donc d'abord trouver des exemples de description de contenu de documents susceptibles d'intéresser de futurs lecteurs. Il s'agit ensuite d'identifier qui pourrait créer une telle indexation sachant que dès lors que le fonds rassemble des documents singuliers, contrairement aux livres des bibliothèques, leur indexation n'est jamais réutilisable et ses coûts sont donc extrêmement élevés²⁸. On tâchera par conséquent de repérer des personnes qui assurent déjà une forme de lecture de ces documents et de cartographie du fonds, des personnes qui pourraient accueillir la fonctionnalité technique comme introduisant un nouveau support à leurs pratiques plutôt que des tâches supplémentaires. Une fois identifiés quelques exemples d'indexation matière et le profil des indexeurs, il devient possible de choisir un des quatre modèles²⁹ (classificatoire, à facettes, combinatoire, folksonomique), de montrer comment sa structure générique subsume les exemples et comment les indexeurs pourront la faire évoluer.

Étude de cas n°1

Si les cahiers de laboratoire sont tenus page après page, ils sont consultés de manière thématique notamment à l'occasion de la déclaration d'une invention à l'employeur, lors de l'écriture d'un brevet ou préalablement à l'écriture d'un contrat de prestation ou de collaboration. Dans ces différents cas, le point d'accès est l'*invention*. Associer une invention à un document n'est généralement pas possible au moment de l'écriture du document qui pourtant la décrit : reconnaître et nommer une invention nécessite que les différentes idées qui la composent fassent "système". C'est à ce moment seulement que l'inventeur peut nommer l'invention et tâcher d'identifier les différents documents qui en exposent ses principes. Cette relecture fastidieuse visant à indexer les documents peut être simplifiée si une première indexation a été mise en place préalablement : une indexation relative au(x) projet(s), autrement dit, relative à la visée initiale des recherches. Le modèle d'indexation est donc à facettes : une facette des projets et une des inventions. Les valeurs doivent pouvoir être librement définies par l'inventeur puisque, par définition, les inventions ne peuvent être préexistantes dans un vocabulaire contrôlé et que, par ailleurs, la liste des projets actifs se renouvelle très régulièrement.

²⁸ Laure Amélie Guitard, *Op. cit.*

²⁹ Voir de les combiner ou de les adapter.

Étude de cas n°2

Alors que la phase de description concernait les photographie des vitraux, celle de l'indexation matière porte sur leur sujet c'est-à-dire sur les vitraux eux-mêmes. Leurs notices descriptives, précédemment évoquées, nous donnent une idée assez précise de ce qui peut intéresser l'expert ou l'amateur éclairé : la scène et les personnages représentés, la datation de l'œuvre, le maître verrier auquel on l'attribue, des techniques du verre remarquables... Pour un tel public, un modèle à facettes pourrait reprendre ces différentes catégories³⁰. Les valeurs, facilement amendables en raison de la grande richesse du domaine, devraient également pouvoir être structurées hiérarchiquement, afin de rendre compte de niveaux de détails différents (des dates précises attestées par des sources historiques, comme des portions de siècle dans le cas de datation par le style ; des attributions attestées ou hypothétiques à un maître verrier ou à son atelier) ou pour rendre compte par exemple de la structure de l'écrit duquel est tiré le récit. Ces notices descriptives, on l'a compris, fournissent une base essentielle pour l'indexation matière, cependant le passage d'un texte à des termes réutilisables affectés à des facettes est une opération intellectuelle complexe et coûteuse. À terme, on peut espérer par exemple que des doctorants en Histoire de l'art, utilisent la plateforme pour établir leur “catalogue”³¹ et contribuent ainsi à l'indexation, cependant, il est indispensable, comme dans la plupart des espaces participatifs, de commencer par une “amorçage” (*bootstrap*) afin d'attirer les contributeurs futurs.

4. Annotation/Citation

“Nous avons peut-être les mêmes livres mais nous surlignons des passages totalement différents” exprime, dans un dessin de Barbara Smaller pour le *New Yorker*, un couple devant les bibliothèques de son salon. Dans nos archives où le stylo est proscrit ou dans nos bibliothèques où le surligneur est honni (pour des raisons évidentes de conservation), on aurait peut-être tendance à l'oublier mais la lecture de travail – lecture active à des fins de mémorisation et d'interprétation – laisse des traces.

L'annotation est tellement indissociable du travail intellectuel que parfois elle s'institutionnalise et devient partie intégrante du document : dans les manuscrits du Moyen-Âge, sous forme de *scholies* (reprises dans la marge d'un terme pour en

³⁰ Pour le grand public, par contre, des travaux de thèse démarrent sur ce que pourrait être le premier point d'accès de celui/celle qui découvre le domaine et surtout sur la manière de lui transmettre petit à petit le goût et la possibilité de développer son expertise.

³¹ Annexe d'une thèse recensant les œuvres étudiées.

expliciter ou en discuter la signification), de gloses marginales ou interlinéaires, de rubriques (étymologiquement “ce qui est en rouge”, afin de faire ressortir une phrase ou de donner un titre à un passage), mais aussi dans les premiers imprimés, avec notamment la structure concentrique unique des commentaires du *Talmud de Babylone*. Ces exemples, par leur matérialité, illustrent le double mouvement permis par la lecture des annotations³² :

- un premier allant de l’intérieur du texte vers l’extérieur, avec les commentaires et les références à d’autres ouvrages,
- un second allant de l’extérieur vers l’intérieur du texte avec les rubriques, prolongeant d’une certaine manière l’indexation matière à l’intérieur du document.

Mais comme l’illustre l’exemple introductif et les interdits des archives et des bibliothèques, la matérialité de la page apporte aussi ses limites que pourraient lever des annotations numériques partagées³³ : sur le papier, les traces d’interprétation des uns font obstacle à celles des autres, alors que le numérique permettrait de choisir les annotations à superposer au document comme autant de feuilles de calque^{34 35 36}.

Ce mouvement vers un support numérique destiné à la lecture active est initié en 1965 par Theodor Nelson. Alors étudiant en sociologie, il imagine de détourner les “calculateurs” (*computers*) de l’époque pour en faire des “machines littéraires”³⁷ : des machines sur lesquelles chaque lecteur pourrait librement annoter, comparer côte-à-côte, commenter, citer dans de nouveaux documents. C’est ce qu’il appelle

³² Aurélien Bénéel, Jean-Pierre Cahier, Matthieu Tixier, LaSuli : un outil pour le travail intellectuel. *Actes du 14e colloque international sur le document électronique (CIDE)*. Europa, 2011.

³³ Jacques Virbel, Annotation dynamique et lecture expérimentale : vers une nouvelle glose ?, *Littérature 96*, 1995. p. 91–105.

³⁴ Martin Röscheisen, Christian Morgensen, Terry Winograd, Beyond browsing: Shared comments, SOAPs, trails, and on-line communities, *Computer Networks and ISDN Systems*, 27 (6), 1995. 739-749.

³⁵ Manuel Zacklad, Aurélien Bénéel, Sandra Bringay, Catherine Barry-Gréboval, Jean Charlet, Michel Treins, Gabriella Salzano, Olivier Curé, Françoise Détienne, Flore Barcellini, Jean-Marie Burkhardt, Warren Sack, Sylvie Guibert, Françoise Darses, Jean-Françoise Boujut, Gaëlle Lortal, Myriam Lewkowicz, Amalia Todirascu-Courtier, William Turner, Marie-Josèphe Pierrat, Mathilde de Saint Leger, Processus d’annotation dans les documents pour l’action : textualité et médiation de la coopération. Roger T. Pédaque (Eds.). *La redocumentarisation du monde*, Cepaduès, 2007. p. 149–166.

³⁶ Aurélien Bénéel *et al.*, *Op. cit.*

³⁷ Theodor Holm Nelson, *Literary Machines*, Mindful Press, 1993.

un “hypertexte”, une notion dont nous ne connaissons aujourd'hui, selon lui, que des formes “triviales”³⁸, puisqu'elle suppose la possibilité pour tous de créer des liens bidirectionnels entre n'importe quels fragments (assurant conjointement la liberté de citation et la reconnaissance du travail des créateurs) dans une architecture totalement décentralisée (donc sans aucune mainmise de la part des propriétaires de serveurs ou des nations)³⁹.

Pour prendre en compte l'interprétation par les lecteurs, le concepteur de système documentaire doit d'abord déterminer le type de trace le plus approprié (commentaires, surlignement et éventuellement typage de fragments, florilèges, comparaisons, parcours de lecture, références bibliographiques, citations, etc.). Mais parce que les interprétations seront multiples, le concepteur doit ensuite imaginer les formes d'affichage ainsi que l'architecture technique qui permettront une libre expression des points de vue sans pour autant submerger le lecteur.

Étude de cas n°1

Par la diversité des éléments qui le constituent, par le nombre restreint de ses lecteurs, par sa forme de journal, par l'obligation de garder les pages en l'état, le cahier de laboratoire se prête peu aux annotations. Par contre, les éléments du cahier sont cités : parfois dans les contrats de collaboration (pour matérialiser les savoir-faire existant chez chacune des parties avant le début de la collaboration) mais surtout dans les déclarations d'invention.

Le système documentaire fournira un support de rédaction de ces déclarations d'invention, rédaction se traduisant par la citation, la reformulation et le réagencement d'éléments épars, textuels et graphiques, tirés des cahiers de laboratoire pour une nouvelle visée. Pour le rédacteur du cahier de laboratoire – salarié auquel une “mission inventive” a été confiée – la déclaration d'invention à l'employeur lui permet d'obtenir une “rémunération supplémentaire” pour toute invention brevetable (dans le cas des fonctionnaires, la prime est partagée entre les co-inventeurs en fonction de la contribution de chacun). Pour l'employeur, l'enjeu de la déclaration d'invention est de déterminer si l'invention est brevetable et, si elle l'est, de voir si l'utilité industrielle et commerciale supposée de l'invention

³⁸ Theodor Holm Nelson, Xanalogical Structure, Needed Now More Than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-use, *ACM computing surveys* 31 (4es), ACM, 1999.

³⁹ Aurélien Béné. 15 ans de Web 2.0 : Imitation pop ou extension du domaine de l'Hypermédia ?, *Actes de la 15e édition de la conférence Hypertextes et hypermédiias, produits, outils et méthodes (H2PTM 2019)*, ISTE, 2019. p. 111-125.

justifie les coûts de dépôt et d'exploitation d'un brevet. Pour répondre à ces objectifs, la déclaration d'invention (comme c'est souvent le cas dans les établissements de recherche) sera structurée de manière à mettre en évidence les critères de brevetabilité (solution technique à un problème technique, jamais divulguée, allant au-delà d'une simple juxtaposition de procédés connus de "l'homme du métier"), mais aussi les perspectives commerciales, ainsi que la part de chacun dans l'invention.

Étude de cas n°2

Le système construit autour des photographies de vitraux accueillera les contributions de tous : tout autant les doctorants souhaitant rendre visible au plus grand nombre leur travail quotidien, que des historiens d'art cherchant à rendre visibles leurs compétences à l'occasion par exemple d'une recherche d'emploi, que les amateurs éclairés et tous ceux qui se découvriront une passion pour les vitraux grâce au projet. Pour autant le statut de ces contributions devra être différencié :

- certaines seront destinées au plus grand nombre, d'autres seront hyper-spécifiques,
- certaines seront "formalisées" (à l'aide d'une sorte de thésaurus propre à une personne ou un collectif, ou encore de liens bidirectionnels "voir aussi", par exemple, entre les vitraux basés sur le même modèle préparatoire, appelé "carton"), d'autres pourront être faites en texte libre sous forme de commentaires ou de liens vers des ressources externes (inventaires, articles de recherche...).

Chaque type d'espace (points de vue "officiels", points de vue de chercheurs, commentaires des visiteurs) aura ses propres règles de contribution et de visibilité. Une contribution, reconnue comme de qualité par un collectif, pourra être publiée dans son espace, et ainsi acquérir une nouvelle visibilité.

Les contributions personnelles devront si possible pouvoir être exportées (pour être intégrées par exemple à un mémoire ou un ouvrage). Par ailleurs, leur mise en visibilité à l'extérieur du système, auprès de l'entourage des contributeurs (via les plateformes dites de "réseaux sociaux") devra être facilitée à des fins tout à la fois de motivation des contributeurs actuels que de recrutement de nouveaux.

5. Authenticité

Suzanne Briet nous le rappelait, le document est une *preuve*. Le terme de “preuve”, n'est pas à prendre au sens mathématique (“cohérence d'une démonstration”), ni au sens scientifique que pouvait lui donner les positivistes logiques (“certitude de la correspondance d'un discours avec le réel”), mais au sens plus juridique, administratif ou même philologique de “témoignage attesté par un tiers”⁴⁰.

Cette portée administrative du document, permettant de faire valoir des droits (notamment de propriété), explique non seulement l'histoire des archives départementales (créées dès l'An V de la toute jeune république française), mais aussi la composition de leurs fonds (dont une grande part est dévolue aux actes notariaux et aux documents produits par les services publics), sans oublier le contrôle et la garantie de l'État incarnés par le conservateur. Dans les bâtiments, la place relative des corps des usagers et des personnels empêche très largement les falsifications : l'accès réservé aux magasins, la consultation sous le regard de tous dans la salle de lecture, la médiation par le président de salle notamment au moment de rendre les pièces d'archive. Mais l'histoire et l'authenticité du document d'archive commence avant le magasin, dans son versement par les services producteurs et son acceptation, mais aussi, chez les producteurs, par les mains qui l'ont signé, tamponné ou marqué d'un sceau.

L'impact du numérique sur l'authenticité est ambivalent : il n'a jamais été aussi facile de modifier un document, mais jamais aussi facile non plus de “tracer” l'ensemble des actions autour de ce document. À l'intérieur d'une organisation, la confiance attribuée à l'authenticité des documents présents dans le système documentaire est très largement celle que l'on accorde au Responsable de la sécurité des systèmes d'information (RSSI) pour assurer la fiabilité de la génération automatique et de la conservation des métadonnées associées à un contenu.

Quand la personne à qui est destinée la preuve ne dépend pas de la même organisation que celle qui gère le système documentaire (et n'a donc pas de raison de lui faire confiance), les métadonnées du document ne suffisent plus. Il faut faire alors appel à des techniques numériques non centralisées garantissant l'authenticité ou, pour être plus précis, suivant les cas, la *non-répudiation*, l'*autorité* ou l'*antériorité*. Traditionnellement :

⁴⁰ Aurélien Bénel, Christophe Lejeune, Partager des corpus et leurs analyses à l'heure du Web 2.0, *Degrés 136-137*, section “m”, 2009.

- sont engagés par un document ses signataires : ils ne peuvent le répudier ou l'amender sans l'accord de toutes les parties ;
- l'autorité d'un document vient de celui qui signe ce document (souvent accompagnée du tampon relatif à son institution et à sa fonction) ;
- est considéré comme auteur d'un document celui qui peut fournir l'exemplaire le plus ancien de ce document (la date devant être attestée par un tiers assermenté ou réputé neutre).

En informatique, cette garantie de l'authenticité, sans centralité, repose sur des techniques cryptographiques :

- la *signature électronique* (par cryptographie asymétrique⁴¹), qui telle un cachet de cire sur un pli, ne peut être apposée que par le possesseur du sceau (*clé privée*) et qui est cassé par toute modification du contenu ;
- l'*horodatage certifié*, qui consiste, de manière entièrement automatisée et selon des procédures organisationnelles et techniques précises, à faire signer par un tiers (neutre) l'association d'un contenu et d'un horodatage⁴² ;
- l'*arbre de Merkle* (ex : *Git*, *Blockchain*) qui à condition d'être largement répliqué chez des acteurs neutres ou d'intérêts opposés (ouvrant ainsi à une confrontation des sources), permet de chaîner de manière pérenne des éléments et donc de dater de manière relative les documents référencés par un même entrepôt (technique éventuellement combinable avec la précédente⁴³).

Étude de cas n°1

La forme traditionnelle du cahier de laboratoire est celle de pages cousues entre elles, numérotées, remplies des notes quotidiennes d'une personne (sans laisser d'espaces), tamponnées régulièrement par un notaire ou relues et signées par un "témoin" le plus indépendant possible des intérêts du projet⁴⁴. Cette forme met en

⁴¹ Whitfield Diffie, Martin Hellman, New Directions in Cryptography. *IEEE Transactions on Information Theory* archive 22 (6), 1976. p. 644–654.
<https://ee.stanford.edu/~hellman/publications/24.pdf>

⁴² Stuart Haber, W. Scott Stornetta, How to Time-Stamp a Digital Document. *Journal of cryptology* 3 (2), 1991. p. 99–111. <http://dx.doi.org/10.1007/BF00196791>

⁴³ *Ibid.*

⁴⁴ *Ibid.*

évidence l'importance de l'antériorité en matière de propriété intellectuelle. Pour autant, la jurisprudence nous invite à prendre des distances avec cette forme :

- les cahiers de laboratoire peuvent être récusés que ce soit pour fraude avérée ou parfois seulement en raison de négligences (difficilement évitables vu la lourdeur des règles à appliquer) ;
- inversement, sont déclarés comme recevables des documents de formes diverses (lettres, e-mails, déclarations d'invention, brouillons, fichiers informatiques) voire des témoignages devant la cour, du moment qu'ils émanent d'une personne indépendante des parties et constituent pour l'invention un *terminus ante quem* (date avant laquelle l'évènement a eu lieu).

Tout d'abord, notre système documentaire aura recours massivement à l'horodatage certifié des documents versés au “cahier de laboratoire réinventé”. En plus de l'horodatage lors du dépôt volontaire d'un document dans un état officiel, le système encouragera un horodatage précoce :

- par scrutation⁴⁵ des documents longs en cours d'écriture (afin de gagner quelques jours voire quelques semaines entre la dernière sauvegarde et le dépôt volontaire),
- par interception⁴⁶ des documents partagés avec des collaborateurs.

Ces horodatages automatiques, si précieux dans notre cas, sont rendus possibles par le fait que le tiers horodateur délègue son autorité à un automate, toujours disponible, pour tous, et de manière gratuite. Cette simplicité de la procédure d'horodatage certifié contraste avec la complexité de la signature électronique quand celle-ci est effectuée par les usagers (dans l'usage quotidien mais surtout dans sa mise en place organisationnelle et technique). Cette signature ne sera utilisée (conjointement à de l'horodatage certifié) que pour la déclaration d'invention qui reste beaucoup plus rare, en fréquence et en nombre de personnes concernées.

Étude de cas n°2

⁴⁵ À la demande de l'utilisateur.

⁴⁶ *Idem*.

Dans le cas des vitraux, les photographies munies de leur localisation sont des *témoignages* de ce qui est présent dans la baie et l'édifice correspondants. Bien entendu, il y a peu de risques que les photographies soient volontairement falsifiées. Par contre, il arrive fréquemment qu'il y ait des erreurs de localisation voire que la localisation ait temporairement ou définitivement changé (démontage pour restauration ou préservation, déplacement d'une baie à l'autre, etc.). Sans mécanisme complémentaire, la fiabilité de ce témoignage ne serait pas garantie.

L'idée est de transposer le principe de *croisement des sources*, non seulement en multipliant les corpus de photographies d'auteurs et de dates différents, mais surtout grâce à une fonctionnalité de commentaires pouvant incorporer des liens et des pièces jointes. En effet, la présence de tels ou tels vitraux dans les baies est attestée par de nombreux inventaires, monographies, articles de recherche, et même par les photographies prises par un simple visiteur avec son téléphone portable. En accueillant des contributions (y compris de l'extérieur) et en engageant une discussion, il devient possible d'éclairer la situation et de réviser si nécessaire la localisation d'une photographie⁴⁷.

6. Révision

Un système documentaire numérique – destiné à soutenir l'activité d'un collectif – doit souvent permettre l'évolution d'un document tout en garantissant la valeur de preuve de différents états de ce document. En cela, il s'éloigne de la pratique des archives qui ne conservent d'ordinaire que les documents dans leur état final, validé. Pour cette nouvelle dimension du document, nous devons nous inspirer d'autres disciplines auxiliaires de l'histoire, notamment la philologie et la *génétique des textes*⁴⁸.

La génétique des textes étudie les manuscrits, les brouillons, afin de comprendre le processus de création et de maturation d'une œuvre (littéraire, philosophique...). L'observation minutieuse de ces brouillons (couleur, épaisseur et texture des traits) permet d'identifier les différentes couches d'écriture (superposition des traits pour les surcharges et les ratures). Cette image des *couches d'écriture*⁴⁹ fournit à la fois une

⁴⁷ Concernant l'authenticité (qui d'une certaine manière constitue l'objet dont on parle), on notera le souhait d'aboutir à un consensus, là, ou au contraire, concernant la datation du vitrail ou la détermination d'une scène, on pouvait se satisfaire d'un sens en suspens ou en dialogue, dans la cohabitation de “points de vue” opposés (cf. section 4).

⁴⁸ Daniel Ferrer, Critique génétique et philologie : racine de la différence, *Genesis 30*, 2010. <https://doi.org/10.4000/genesis.98>

⁴⁹ À rapprocher également des “repentirs” que l'analyse aux rayons X des peintures permet de révéler.

métaphore (utilisée notamment dans les interfaces des traitements de texte) mais aussi un ensemble de modèles, comme nous le verrons par la suite par contraste avec la philologie.

La visée et l'approche de la philologie diffère de celle de la génétique des textes : afin d'établir le texte de référence, de restituer une origine toujours inaccessible, le philologue compare les exemplaires d'un texte pour en déceler les variantes⁵⁰. La classification des variantes permettant celle des manuscrits en "familles", il est alors possible de construire le *stemma codicum* ("guirlande des codex"), arbre généalogique d'une œuvre dont les nœuds et notamment la racine sont des manuscrits perdus (ou supposés tels)⁵¹. La forme arborescente de la filiation n'est pas toujours suffisante : certains exemplaires d'une famille présentent des variantes caractéristiques d'une autre famille. Là où la philologie traditionnelle parlait de "contamination", filant la métaphore de la corruption par rapport à une origine considérée pure, les philologues contemporains reconnaissent dans le geste des copistes, se basant sur plusieurs exemplaires (de lignées différentes) pour en produire un nouveau, la préfiguration de leur propre travail. Ces branches qui se rejoignent (dans une structure dite "hiérarchique") sont indispensables pour donner une matérialité au travail intellectuel de fusion de révisions qui ont évolué indépendamment l'une de l'autre (sur des supports et souvent du fait de contributeurs différents).

Par contraste, ces deux approches de l'histoire d'un document dessinent l'espace des possibles pour la gestion des révisions :

1. Pour le stockage des révisions, nous avons le choix entre un modèle dit "différentiel" – où chaque révision est une couche posée sur la précédente – et un modèle par "instantané" (*snapshot*) dans lequel chaque révision est un exemplaire complet.
2. Pour la structure de filiation des révisions, nous avons dans l'ordre de complexité croissante : le modèle linéaire dans lequel on considère que chaque révision est basée exclusivement sur la révision précédente, le modèle arborescent qui permet de rendre compte de "lignées" évoluant en parallèle (notamment lorsque les contributeurs sont multiples), le modèle

⁵⁰ *Ibid.*

⁵¹ Lene Schøsler, Historical corpora : Problems and methods. *Linguistica Computazionale* 20-21. 2004.

hiérarchique qui permet de rendre visible, en outre, l'intégration dans une révision de plusieurs "lignées" (pour produire un œuvre commune).

À ces modèles structurels hérités des sciences historiques, l'informatique, quant à elle, a adjoint des modèles de collaboration : le modèle dit "optimiste" dans lequel on fait l'hypothèse qu'une révision tient forcément compte des révisions plus anciennes, le modèle par "verrouillage" dans lequel les coauteurs se coordonnent *a priori* en réservant l'édition d'un document (ou d'une de ses parties) à l'un d'entre eux pour un temps donné, le modèle, enfin, du "registre de sortie" (*checkout/checkin*) permettant de détecter que des révisions ne tiennent pas compte l'une de l'autre (on parle alors de "conflit d'édition"). Pour résoudre ce conflit, une "fusion" est nécessaire après comparaison et combinaison des différences entre les deux révisions.

Étude de cas n°1

Dans notre cahier de laboratoire numérique, comme nous l'avons vu dans la section précédente, le contenu d'un document est horodaté dans trois situations différentes : lors du dépôt explicite d'un état du document, lors de son partage avec des collaborateurs, mais aussi par scrutation des modifications de certains documents rédigés sur une longue période (mémoire, thèse, article long, etc.). Dans les deux premières situations, l'état du document est explicitement considéré comme significatif. Dans la deuxième, il est même une référence partagée. Il semble indispensable d'archiver ces états du document comme autant de révisions. Dans la troisième situation, par contre, le seul but de l'horodatage est d'obtenir la date la plus précoce possible au cas où cet état serait par la suite considéré comme significatif. Ces états non-significatifs *a priori* n'ont donc pas vocation à être archivés en tant que révisions.

Certains documents révisables sont co-écrits. Idéalement la structure de filiation des révisions devrait donc être hiérarchique et le modèle de collaboration devrait être celui du "registre de sortie". Cependant, la complexité de mise en place et d'usage de ce genre de structures est telle qu'il semble préférable (au moins dans un premier temps) d'adopter, pour un outil de bureautique comme le nôtre, une structure linéaire de filiation⁵² avec un modèle de collaboration optimiste.

⁵² Les usagers habitués à travailler avec de nombreux coauteurs (comme c'est le cas des développeurs de code source) utilisent déjà des structures de filiation hiérarchique et pourront continuer à les utiliser sur les mêmes documents conjointement avec notre système.

Étude de cas n°2

Dans le cas des vitraux, la plupart des photographes contribuant régulièrement au projet retouchent leurs clichés. Ils les recadrent systématiquement afin de ne garder que le sujet traité (panneau, ensemble de panneaux ou baie). Certains modifient la luminosité ou le contraste. D'autres encore appliquent des algorithmes de correction des distorsions dues à la perspective (la lentille de l'appareil étant positionnée d'ordinaire plus bas que le vitrail).

D'un point de vue mathématique, chaque retouche est une perte d'information⁵³. Si par la suite, une nouvelle chaîne de traitement est à substituer à l'actuelle, il est grandement préférable de partir du cliché d'origine. Ainsi même si la plupart des photographies ne seront retouchées qu'une fois, il est conseillé de garder le cliché d'origine et ses éventuelles retouches selon un modèle linéaire de révision.

Conclusion

Un regard sur le chemin parcouru au fil de cet article, nous permet d'apprécier comment six points d'attention sur la matérialité des documents et des corps qui les manipulent nous ont permis de poser les choix de conception de systèmes documentaires numériques. Les deux systèmes documentaires présentés comme “en devenir”, pour mettre le lecteur en situation d'écrire un cahier des charges, ont en fait été implémentés ces dernières années et sont aujourd'hui utilisables. Elles illustrent à elles deux l'ensemble des projets de recherche sur lesquels nous avons appliqué cette méthode, que ce soient des projets au long cours, souvent dans le domaine des humanités numériques (enquêtes qualitatives en sociologie, traduction participative de textes culturels, archive ouverte d'un laboratoire de recherche), ou des projets plus ponctuels portant sur des documents d'entreprise (gestion du cycle de vie des produits d'ABB, documents de recherche et développement chez EDF, documents d'entretien des équipements d'Air Liquide). En tant que méthode, cette grille n'est ni vraie ni fausse – elle ne se prête pas à une expérimentation scientifique – mais on peut en évaluer la fécondité. Aux projets, portés par nos propres équipes, il faudrait ajouter ceux de nos étudiants des 14 dernières années, égrener les métiers intellectuels (architecte, agent immobilier, agent du cadastre, enquêteur de police, avocat, notaire, journaliste, conservateur de musée, archéologue, historien, astronome, botaniste, médecin, épidémiologiste, généalogiste, musicien, développeur de logiciel...) pour lesquels ils ont su montrer

⁵³ Cette perte d'information est d'autant plus importante quand il y a une alternance de transformations géométriques et de compressions avec pertes (comme en JPEG).

comment un système documentaire conçu selon cette méthode, permettait d'instrumenter leur activité. Cette méthode invite le concepteur à laisser un temps les choix techniques en suspens. Elle le plonge dans l'épaisseur historique de la pratique des archives, des bibliothèques, de la philologie, de la diplomatique. Elle le confronte à la matérialité de la page, des feuillets, de leur rangement. Ce n'est qu'au prix de cette plongée que peuvent émerger les choix de modèles informatiques au service, non d'une vaine virtualité, mais de gestes de métier à réinventer, entre tradition et innovation.

Aurélien Bénel

Maître de conférences en informatique

Université de technologie de Troyes, LIST3N/Tech-CICO

aurelien.benel@utt.fr

Document numérique : L'informatique en quête d'un corps

L'influence de l'informatique sur la pratique des archivistes est indéniable. On imagine moins celle que l'archivistique peut avoir sur l'ingénierie informatique (notamment pour les systèmes d'information destinés aux professions intellectuelles). La méthode, présentée dans cet article, invite le concepteur à laisser un temps les choix techniques en suspens. À travers six points d'attention (identification, description, indexation matière, annotation/citation, authenticité, révision), elle le plonge dans l'épaisseur historiques de la pratique des archives, des bibliothèques, de la philologie, de la diplomatique. Elle le confronte à la matérialité de la page, des feuillets, de leur rangement. Ce n'est qu'au prix de cette plongée que peuvent émerger les choix de modèles informatiques au service, non d'une vaine virtualité, mais de gestes de métier à réinventer.