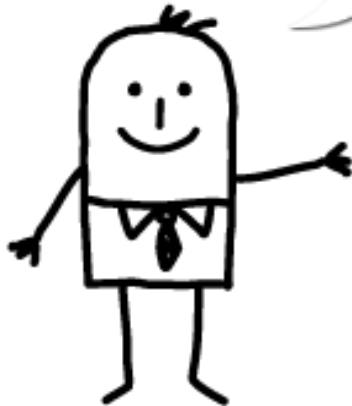


LES MÉTADONNÉES D'ARCHIVAGE

Il faut pouvoir retrouver ce qu'on a archivé. Il faut aussi le gérer. Heureusement, les métadonnées sont là!



Tout le monde a entendu parler des métadonnées, même les médias en parle parfois.

Pour **gérer** un document numérique, il est nécessaire de disposer d'informations sur ce document (auteur, date, thème, lieu...):

- pour pouvoir le **retrouver** (grâce aux mots-clés) ou
- pour piloter son cycle de vie, sa **consultation**, sa **conservation** et sa **destruction** à échéance de la durée de conservation.



D'où viennent les métadonnées?



Le terme *métadonnée* a fait son entrée dans la langue courante dans les années 2000. Le mot est d'origine grecque, composé :

du préfixe **méta**, qui signifie « ce qui dépasse » ou « ce qui englobe », comme dans métaphysique, métamorphose, métatarse ou métalangage ;

et de **donnée**, terme qui désigne tout élément d'information géré par un outil informatique.

Toute personne qui a utilisé une application de GED (gestion électronique de document) ou une solution d'archivage ou même un outil de photographie numérique est familier des métadonnées.

Pourtant le mot ne figure pas encore dans le *Dictionnaire de l'Académie française* (9^e édition). Dans le *Petit Larousse* en ligne, le mot est rattaché au domaine informatique [ce qui est restrictif] et est défini, au singulier comme « donnée servant à caractériser une autre donnée, physique ou numérique » avec comme exemple : « *Les métadonnées sont à la base de l'archivage* ».

Nota bene : le mot français métadonnées existe au singulier (une métadonnée) tandis que l'anglais *metadata* est un pluriel sans singulier; on utilise l'expression *metadata element*.



Des métadonnées pour quoi faire?



Les métadonnées ne sont pas une fin en soi, même si l'environnement numérique impose de produire des informations structurées sur les documents que l'on doit gérer.

Pour simplifier, on dira que les métadonnées ont **deux finalités** :



retrouver l'information contenue dans le document géré : de quoi ou de qui parle-t-il ? Quels lieux ou quelles personnes sont concernés ? À quelle date ?

gérer **le cycle de vie de ce document**, dans un premier temps pour **l'élaboration du document jusqu'à la validation finale** (différentes versions, contrôle de la qualité du document, identifiant unique) et dans un second temps pour **la conservation** (qui conserve ? Où ? Sur quel support ? Pendant combien de temps ? Quelles règles pour la destruction ? Etc.) et pour **l'accès** (qui a le droit de consulter ? Pour quel usage ? A-t-on le droit de faire des copies ? Y a-t-il des droits à payer, des précautions à prendre ?).



Il est donc important, quand on est en situation de créer des métadonnées de définir clairement le rôle qu'elles auront à jouer.

Comment faisait-on avant ?

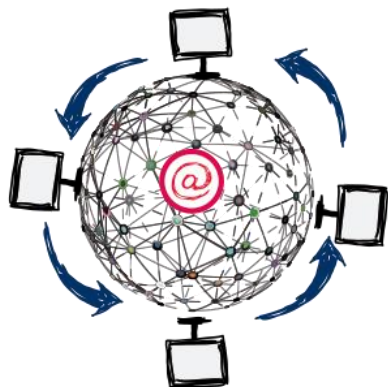


Question : si les métadonnées n'ont qu'une vingtaine d'années, comment faisait-on avant ?

Dans la pratique, les métadonnées sont effectivement associées à la gestion des documents numériques, mais le besoin de gérer des collections documentaires et archivistiques existe depuis de nombreux siècles. Le principe des « métadonnées » est donc très ancien au travers des index, des catalogues, des registres de bibliothèques, des inventaires d'archives, des thésaurus, etc. qui étaient le moyen d'organiser l'information essentielle sur les documents, pour les retrouver (contenu, type d'information) ou pour les gérer (localisation, droits d'accès, destruction...).

L'informatique, avec les possibilités de traitement automatique, avec l'impossibilité de voir directement le document numérique, avec la normalisation exigée pour échanger l'information dans les réseaux, a favorisé le développement et l'importance des métadonnées.

Les nuages de tags sont des nuages de métadonnées.



De la documentation à l'archivage



Le premier modèle de métadonnées, le **Dublin Core** (Dublin dans l'Ohio, USA), date de 1995 et concerne la description des ressources bibliographiques. Il définit les 15 métadonnées qui permettent de gérer un livre ou une ressource numérique :



**Titre - Créateur - Sujet et mots-clefs - Description - Éditeur -
Contributeur - Date - Type - Format - Identifiant - Source - Langue -
Relation - Couverture - Droits.**

Le Dublin Core (ISO 15836 :2003) est à l'origine de très nombreux modèles dont la plupart concerne les ressources documentaires et les documents patrimoniaux, notamment le modèle EAD (Encoded Archival Description) préconisé pour les archives historiques.

Cependant, ce modèle ne répond pas aux besoins de l'archivage.

En effet, il n'inclut pas des informations telles que le rattachement à un processus ou à une contrainte réglementaire, la durée de conservation et sa justification, le contrôle d'intégrité ou encore les modalités de destruction / versement aux archives historiques, qui sont spécifiques à l'archivage.



Quelques définitions



Il existe aujourd'hui plusieurs définitions du mot métadonnées. Voici quelques définitions extraites des normes et glossaires relatifs à l'archivage (*Métadonnées d'archivage / Records management metadata*).



Données décrivant **le contexte, le contenu et la structure** des documents archivés et leur gestion **dans le temps**.

ISO15489/MoReq2 (2001, 2008)



Métadonnées qui **identifient, authentifient et contextualisent** les documents engageants ainsi que les personnes, processus et systèmes de création, gestion, conservation et utilisation afférents, et les règles associées.

ICA-Req (2008)

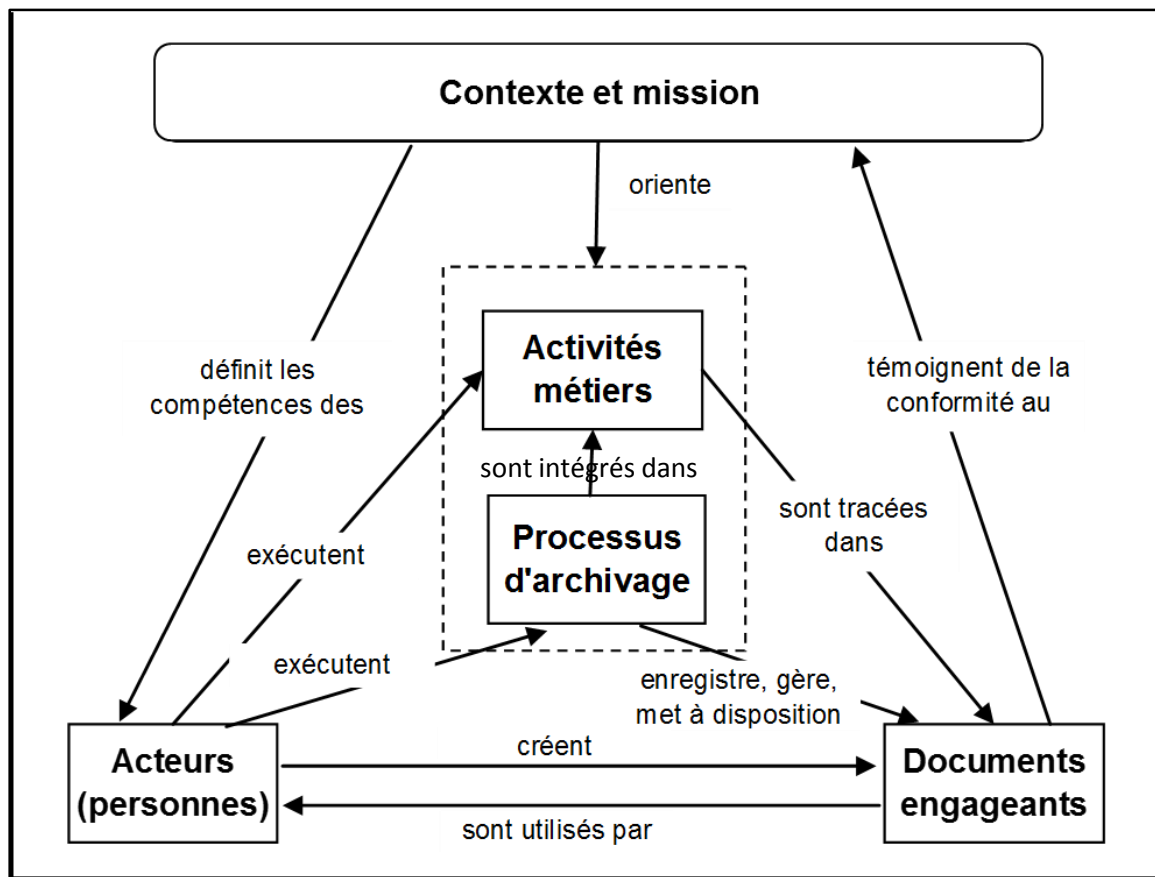


Données structurées ou semi-structurées qui permettent de qualifier et de gérer les documents archivés **tout au long de leur cycle de vie** : contexte de création, contenu, validation, règles de conservation / destruction, caractère vital, confidentialité, support, etc.

Nouveau glossaire de l'archivage de Marie-Anne Chabin (2010)

Un schéma pour les métadonnées d'archivage

Ce schéma (extrait de norme ISO 23081) met en évidence les différents rôles que jouent les métadonnées pour l'archivage. Elles relient le contexte de création (avec sa dimension réglementaire) aux acteurs et aux documents qui tracent leur activité et présentent un risque de non disponibilité dans le temps.



Trad MAC 2009

Production des métadonnées. Qui? Quand? Comment?

À quel moment les métadonnées sont-elles produites ? Comment sont-elles créées ? Par qui ? Où se situent-elles ? Les réponses sont très nombreuses car cela dépend de plusieurs facteurs. Voyons trois remarques pour éclairer la question.

Certaines métadonnées sont des éléments exprimés naturellement dans le document.

On trouve dans un document bien fait : le nom de l'auteur, le destinataire, la date du document, la référence ou l'objet, les thématiques abordées (mots-clés renseignés par l'auteur). D'autres métadonnées sont extérieures au contenu, comme par exemple une vérification de signature électronique.

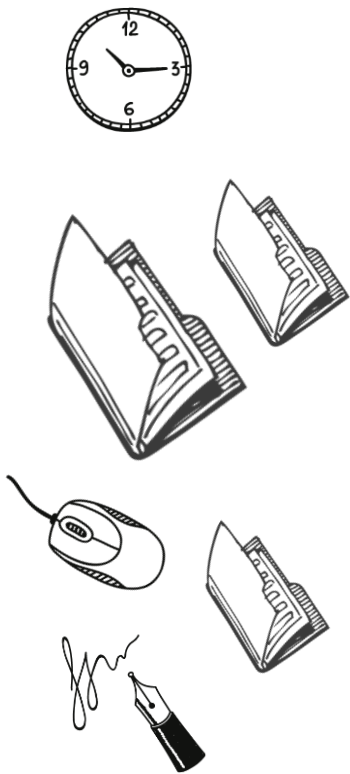
Certaines métadonnées peuvent être incluses dans le corps du document archivé parce qu'on les connaît dès la création du document.

C'est le cas du niveau de confidentialité car c'est souvent le modèle de document ou l'auteur du document qui le détermine. Cela peut également être le cas de la durée de conservation, bien que ce soit assez rare. Les métadonnées qui concernent la gestion du cycle de vie sont en général connues après la validation du document (localisation, migration de support, etc.).

La production est métadonnées peut être automatique ou manuelle.

Les métadonnées peuvent être extraites automatiquement du document si celui-ci a été modélisé ou normalisé, ou extraites de tables associées à la règle de conservation.

Si elles doivent être saisies par l'utilisateur, il faut tenir compte du caractère fastidieux et aléatoire de cette tâche.



Exemple



IMAGINONS UN DOSSIER DE PROJET DE RÉPARATION D'UN THÉÂTRE : on y trouve un rapport d'expertise technique qui existe sous forme papier originale et en copie numérique, quatre plans, une vingtaine de photos, une interview vidéo du directeur du théâtre et de deux acteurs avec des cessions de droit à l'image, une demande de subvention (accordée), deux comptes rendus en trois versions, plusieurs dizaines de mail.

Chaque pièce de ce dossier a ses propres métadonnées : auteur, destinataire, date, objet, format de données, mots-clés thématiques (personnes, lieux, équipements), utiles pour la **recherche** ultérieure.

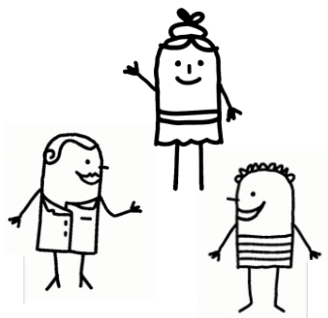
La durée de conservation, métadonnée capitale de l'archivage, peut être portée par chaque document ou par **l'ensemble du dossier** en tenant compte des dates d'ouverture et de clôture du dossier.

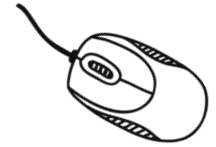
Que faut-il faire ? C'est une question de granularité : a-t-on besoin de gérer le cycle de vie de chaque pièce ou peut-on le faire au niveau du dossier ?

L'organisation des métadonnées dépend de la granularité des objets archivés.

La vidéo qui se rattache au dossier, avec les documents juridiques associés quant à l'exploitation des images, méritera une gestion à part si on envisage de valoriser cette interview dans une action de communication. De même, **les plans et les photos** pourront peut-être être réutilisés en dehors de cette opération de réparation.

L'organisation des métadonnées dépend aussi de l'usage ultérieur qui est souvent prédictible.





Modèles de métadonnées

1

Il existe des modèles de métadonnées, proposés par des normes ou d'autres documents élaborés par des institutions ou des associations.

Ce sont des listes de rubriques à renseigner pour bien décrire ou bien gérer les documents archivés.

Il faut bien distinguer :

la liste des **rubriques ou « champs »** qui définissent les types d'informations à collecter pour bien gérer/retrouver le document archivé (ville concernée, confidentialité...) ; et

les valeurs correspondantes possibles (Brest, Perpignan, Besançon... ; confidentiel, restreint, public, etc.).

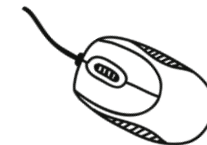
On peut normaliser les champs et laisser libre la valeur ; on peut aussi associer des tables de valeurs qui empêchent de saisir des valeurs non prévues et non contrôlées.

Quand les tables sont déjà prévues dans l'application de création du document, c'est autant de gagné pour la qualité de l'archivage.

Chaque projet doit définir son modèle de métadonnées et décider d'associer ou non des tables de valeurs, pour la meilleure articulation « saisie manuelle / automatisation » en termes de faisabilité. La faisabilité dépend des ressources humaines et technologies disponibles.



Modèles de métadonnées



2

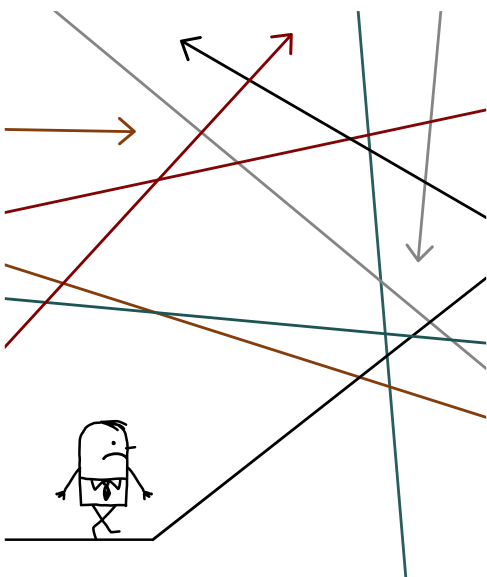
Un des modèles de métadonnées d'archivage les plus complets est celui de la norme européenne **MoReq2**, publié en 2008 (en anglais et en français).

Le modèle de métadonnées de MoReq2 identifie **158 métadonnées** possibles au sein du système d'archivage mais (heureusement!) elles ne sont pas toutes obligatoires et certaines sont exclusives.

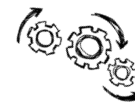
En se limitant à l'essentiel et en automatisant au maximum la capture des métadonnées, il ne reste, parmi ces 158 métadonnées possibles, que **4 à 5 données à renseigner par l'utilisateur**. Ouf !

Petit aperçu des types de métadonnées de MoReq2 :

- Métadonnées de description : auteur, mot-clé
- Métadonnées de capture : date et heure de capture dans le système d'archivage, identifiant système, code de classement, type de document
- Métadonnées de dossier : identifiant dossier, date ouverture, date clôture, nombre de documents
- Métadonnées de classement : entité de rattachement, existence de copie, existence d'extrait
- Métadonnées de signature électronique : date et heure de vérification, algorithme de chiffrement, date et heure de déchiffrement
- Etc.



Métadonnées, archivage et moteur de recherche



L'archivage numérique conduit à une granularité plus fine des documents archivés. Là où hier on décrivait deux ou trois dossiers par boîte d'archives papier, on aura aujourd'hui 200 à 300 fichiers versés dans un système. C'est 100 fois plus! Si chaque objet décrit dans le système est doté de 10 métadonnées, on voit tout de suite l'inflation qui s'ensuit. Si chaque entité décrite possède 50 métadonnées...

1

Quand la capture et la gestion des métadonnées est automatisée et que le système d'archivage est solide et bien dimensionné, la recherche de l'information est efficace et transparente pour l'utilisateur.

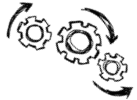
Les métadonnées capturées sont centralisées dans **un « entrepôt de métadonnées »** qui permettra de tout retrouver et de gérer le cycle de vie et la communication de toutes les entités documentaires archivés.

Dans le cas de saisie manuelle, confiée à des personnes extérieures au processus (archiviste, prestataire ou stagiaire), on observe souvent des difficultés à suivre le **rythme de la production**. Si la production des métadonnées est différée de plusieurs mois, la recherche d'information n'est plus fiable...

Attention à la mauvaise anticipation des ressources nécessaires pour atteindre l'objectif!



Métadonnées, archivage et moteur de recherche



2

Les métadonnées se trouvent généralement dans une **fiche de métadonnées** insérée ou associée au document archivé.

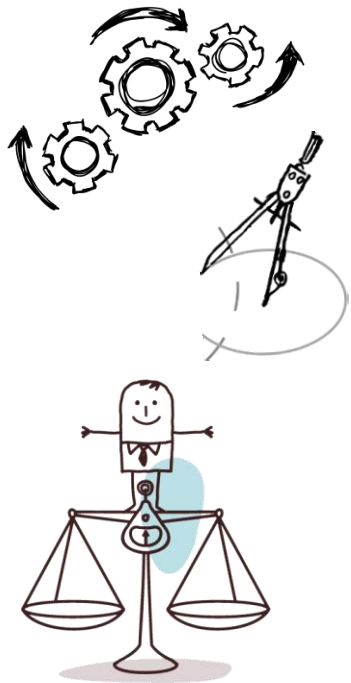
Elles peuvent aussi se présenter sous forme d'**objet de données au format XML**, ce qui facilite leur exploitation dans une base de données indépendante (l'entrepôt de métadonnées); l'accès au document décrit se faisant par le biais d'un identifiant prédéfini qui pointe sur la localisation du document ou du fichier.

On peut encore avoir une **formule mixte** (certaines métadonnées centralisées, d'autres non, accessibles dans un second temps).

Tout est possible !

Il est judicieux d'organiser son modèle de métadonnées en tenant compte des technologies numériques de recherche, notamment les moteurs de recherche, et de leurs performances.

Ainsi, pour un projet d'archivage modeste (PME, famille), en s'appuyant sur un moteur de recherche extérieur et des règles de nommages rigoureuses et descriptives, on peut pratiquement se passer de métadonnées.



En conclusion

Le choix d'un modèle de métadonnées dépend de la finalité de l'archivage de chaque catégorie de documents :

1. des **contenus riches** dont l'exploitation présente une grande valeur ajoutée pour les utilisateurs ;
2. la maintenance des **éléments probants** des documents à risque ou la gestion des destructions, etc.

IL FAUT AUSSI PENSER QUE TROP DE MÉTADONNÉES TUE LES MÉTADONNÉES !

Enfin, **ne pas oublier que les métadonnées sont d'abord des données** : elles doivent elles aussi être sauvegardées, conservées, contrôlées, migrées...

Priorité aux métadonnées nécessaires et suffisantes pour gérer le cycle de vie des documents archivés et pouvoir les retrouver!

