

Module 7B - section 3 : La description et l'indexation des documents numériques

Natasha Zwarich PIAF © AIAF

v1 25/07/2024

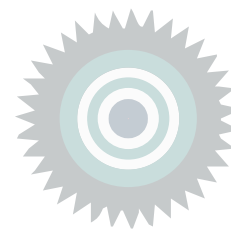


Table des matières

Objectifs	4
Introduction	6
1. Les métadonnées : définitions et typologie	7
1.1. Définition et caractéristiques des « métadonnées »	7
1.1.1. Définition	7
1.1.2. Le rôle des métadonnées	7
1.1.3. Où se trouvent les métadonnées ?	8
1.1.4. Quel est le mode de création des métadonnées ?	9
1.1.5. Quand saisir les métadonnées ?	9
1.1.6. Quel est le cycle de vie des métadonnées ?	9
1.2. Typologie des normes de métadonnées	9
1.2.1. Les métadonnées descriptives	10
1.2.2. Les métadonnées techniques	10
1.2.3. Les métadonnées administratives	10
1.2.4. Les métadonnées d'identification	10
1.2.5. Les métadonnées de provenance et de contexte	10
1.2.6. Les métadonnées d'intégrité	10
1.2.7. Les métadonnées de gestion des droits	11
1.2.8. Les métadonnées de structure	11
1.2.9. Les métadonnées sur la préservation	11
2. Normes et profils de métadonnées pour décrire les documents numériques	12
2.1. Les composantes des métadonnées	12
2.2. Les normes de métadonnées	13
2.2.1. Les normes de contenu	13
2.2.2. Les normes de structure	13
2.2.3. Les normes visant à encadrer le vocabulaire	13
2.2.4. Les normes visant à faciliter l'échange de données	14
2.2.5. L'archiviste doit accorder une grande importance à ces normes	14
2.3. Développement d'un profil de métadonnées	14
2.4. Enjeux des métadonnées	16
3. L'indexation	18
3.1. Définition de l'indexation	18
3.2. Les divers types d'indexation et modes de création	18
3.2.1. Indexation par assignation	18
3.2.2. Indexation par extraction (« automatique »)	18
3.2.3. Indexation en vocabulaire libre (langage naturel)	19
3.2.4. Indexation en vocabulaire contrôlé (langage documentaire)	19
3.3. Le cas particulier de l'indexation collaborative et participative	19

Conclusion	21
Glossaire	22
Bibliographie	24
Webographie	25

Objectifs



Alors que le numérique occupe une place très importante dans le quotidien de l'archiviste, ce module vise à porter un regard sur les opportunités et les défis que posent la gestion et la conservation des documents numériques. En effet, il traite des questions fondamentales relevant de la gestion des documents numériques ainsi que des interventions pouvant être réalisées par les archivistes. En tenant compte des aspects conceptuels, fonctionnels et techniques de la création, la description, la classification et la gestion au moyen de diverses plateformes, on y présente les principales méthodes d'analyse propres aux documents numériques et les principales normes nationales et internationales applicables. Ainsi, ce module permet de mieux comprendre les diverses facettes de la gestion des documents numériques.

L'apprenant doit être en mesure de :

- décrire les caractéristiques des documents numériques;
- évaluer différentes approches, méthodes et technologies pour assurer la gestion des documents numériques;
- prendre en compte les leviers ou les difficultés que pose l'environnement organisationnel et technique.
- avoir une connaissance des différents niveaux d'intervention des archivistes dans la gestion des documents numériques.

Il nous semble important de préciser que ce module se veut une initiation à la gestion des documents numériques. Par ailleurs, la gestion des documents numériques requiert l'intervention de divers acteurs, notamment les professionnels des technologies de l'information et les usagers. L'archiviste doit donc s'appuyer sur des compétences et des savoir-faire spécialisés qui sont complémentaires à l'archivistique. L'archiviste doit aussi disposer d'une vue globale complète et d'une bonne connaissance de tous les besoins afin d'être un acteur incontournable de la gestion des documents numériques. Les connaissances et les méthodes de gestion des documents numériques sont en constante évolution. Par conséquent, ce module vise à présenter des bonnes pratiques. Celles-ci doivent cependant être adaptées selon les contextes organisationnels, la technologie disponible et les acteurs impliqués.

Les notions abordées dans ce module peuvent être complétées avec d'autres modules de formation du PIAF. Comme les principes de l'archivistique demeurent valides avec le numérique, il est recommandé de consulter le module 5 – *Gestion et traitement des archives courantes et intermédiaires*. De plus, les problématiques liées à la conservation à long terme sont abordées dans la suite du module 7 (7C) – *Préservation numérique*.

Le module 9 – Section 2 : *Numériser les documents* est un complément de formation pertinent puisqu'il présente les techniques de base pour l'utilisation des moyens de numérisation et le transfert de support vers le numérique.

Le module, « *Un autre regard* » : *gérer les archives courantes : approche pratique* peut également offrir à l'apprenant des suggestions quant aux niveaux d'intervention à adopter dans la mise en œuvre de dispositifs de gestion des archives courantes.

Le module 14 - *Partenariat et coopération* fournit des suggestions pour mener un projet de gestion des documents ainsi que sur les façons de développer des collaborations entre professionnels.

Finalement, le glossaire du PIAF doit être consulté pour les définitions des termes employés dans ce module de formation.

Introduction



La description et l'indexation permettent de documenter le contexte d'un document et d'en faciliter le repérage. Cette fonction est d'autant plus importante dans un environnement numérique que la masse documentaire est souvent très volumineuse, les documents numériques sont créés et gérés sur divers types de plateformes et peuvent être consultés et modifiés par différents individus qui possèdent leurs propres méthodes de recherche. Dans cette section, nous aborderons la description et l'indexation de documents numériques au moyen de métadonnées.

1. Les métadonnées : définitions et typologie



1. Introduction

Cette section vise à donner une définition des métadonnées et à dresser une typologie à travers quatre grandes catégories : les métadonnées descriptives, les métadonnées techniques, les métadonnées de structure, et enfin les métadonnées administratives.

1.1. Définition et caractéristiques des « métadonnées »

1.1.1. Définition

Les métadonnées jouent un rôle important dans la gestion des documents numériques et leur préservation. Essayons d'en donner une définition.



Étymologiquement, « méta » provient du grec signifiant « après, au-delà de, avec » : « méta » données signifie « au-delà des données », « qui dépasse les données », « qui englobe les données ». Il s'agit donc de données sur les données, à propos des données, qui définissent, décrivent des données. La littérature regorge de définitions du terme métadonnée. Nous retenons ici cette définition qui réfère aux objectifs visés par les métadonnées : « Données utilisées pour décrire et organiser d'autres données. On les utilise entre autres pour définir les spécifications techniques, l'organisation intellectuelle, les conditions d'utilisation, la préservation, l'échange entre systèmes et l'administration des données. » (Turner, s.d. cité dans EBSI, 2018).

1.1.2. Le rôle des métadonnées

Les métadonnées visent plusieurs objectifs de gestion des documents numériques.

- Identifier et organiser les documents numériques;
- Repérer l'information, notamment grâce à l'utilisation de critères de recherche et en facilitant la localisation du document numérique;
- Assurer la préservation à long terme des documents numériques;
- Permettre la gestion des droits d'accès, d'utilisation, de copie;
- Certifier l'autorité intellectuelle du contenu;
- Contribuer à l'intégrité des documents numériques;
- Conserver le contexte de création du document;
- Informer sur le mode de lecture;
- Etc.



Il importe de préciser que les métadonnées sont des données soumises aux mêmes défis de création, gestion, conservation que les documents numériques eux-mêmes. Pour être en mesure d'effectuer une gestion adéquate des métadonnées, il importe de s'intéresser à certaines de leurs caractéristiques.

1.1.3. Où se trouvent les métadonnées ?

Les documents numériques possèdent un certain nombre de métadonnées internes. Les métadonnées internes sont intégrées dans le document numérique lui-même. Ainsi, si le document est déplacé, il transporte ses métadonnées avec lui.

Par exemple, les propriétés d'un fichier Word (Figure 1).

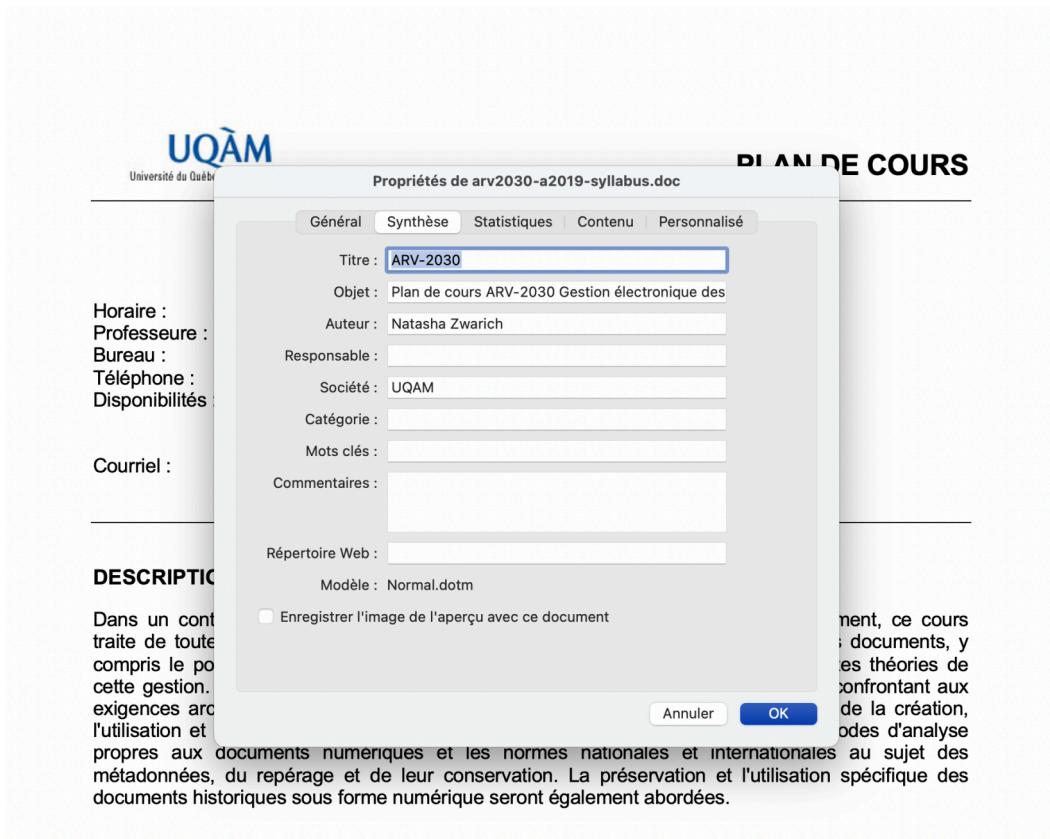
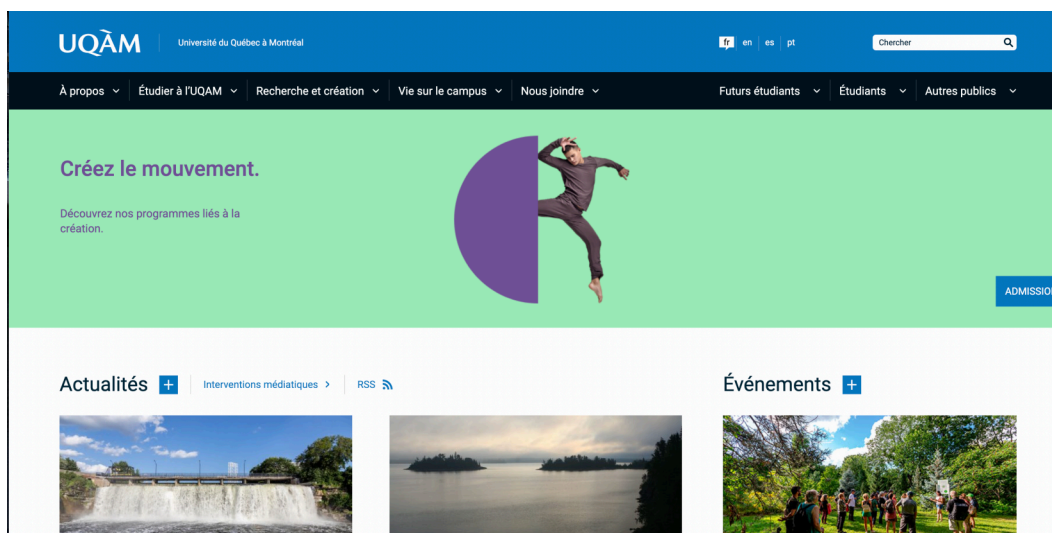


Figure 1. Exemple de métadonnées d'un fichier Word

Le code source d'une page Web (Figures 2 et 3)




```

<!-- front-page.php -->
<!doctype html>
<html lang="fr-CA">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
<meta name="web_author" content="Service Audiovisuel">
<meta name="format-detection" content="telephone=no">
<link rel="shortcut icon" href="https://gabarit-adaptatif.uqam.ca/2014/css/images/favicon_uqam.ico">
<!--[if IE]-->
<script src="https://cdnjs.cloudflare.com/ajax/libs/html5shiv/3.7.3/html5shiv.min.js"></script>
<!--[endif]-->
<meta name="robots" content="index, follow, max-image-preview:large, max-snippet:-1, max-video-preview:-1" />
<link rel="alternate" hreflang="fr" href="https://uqam.ca/" />
<link rel="alternate" hreflang="en" href="https://uqam.ca/en/" />
<link rel="alternate" hreflang="es" href="https://uqam.ca/es/" />
<link rel="alternate" hreflang="pt-br" href="https://uqam.ca/pt-br/" />
<link rel="alternate" hreflang="x-default" href="https://uqam.ca/" />
<!-- This site is optimized with the Yoast SEO plugin v20.11 - https://yoast.com/wordpress/plugins/seo/ -->
<title>UQAM | Université du Québec à Montréal | Accueil</title>
<meta name="description" content="L'UQAM Université du Québec à Montréal (UQAM) offre plus de 300 programmes d'études, dont plusieurs sont uniques au Québec, au Canada et en Amérique du Nord." />
<link rel="canonical" href="https://uqam.ca/" />
<meta property="og:locale" content="fr-CA" />
<meta property="og:type" content="website" />
<meta property="og:title" content="UQAM | Université du Québec à Montréal | Accueil" />
<meta property="og:description" content="L'UQAM Université du Québec à Montréal (UQAM) offre plus de 300 programmes d'études, dont plusieurs sont uniques au Québec, au Canada et en Amérique du Nord." />
<meta property="og:url" content="https://uqam.ca/" />
<meta property="og:site_name" content="Université du Québec à Montréal" />
<meta property="og:image" content="http://ressources-medias.uqam.ca/media/uploads/2020/03/6151917/campuscentral.jpg" />
<meta property="og:image:width" content="1080" />
<meta property="og:image:height" content="761" />
<meta property="og:image:type" content="image/jpeg" />
<meta name="twitter:card" content="summary_large_image" />
<meta name="twitter:site" content="uqam" />
<script type="application/ld+json" class="yoast-schema-graph">{"@context":"https://schema.org","@graph":[{"@type":"CollectionPage","@id":"https://uqam.ca/","url":"https://uqam.ca/","name":"UQAM | Université du Québec à Montréal"}]}

```

Figures 2 et 3. Exemple de métadonnées d'un site Web © Université du Québec à Montréal

Les métadonnées peuvent également être externes aux documents numériques. Elles peuvent être contenues dans une notice séparée du document, telle une description dans une base de données qui fournirait une localisation du document ou dans un instrument de recherche.

1.1.4. Quel est le mode de création des métadonnées ?

Les métadonnées contenues dans un document numérique sont souvent générées automatiquement lors de la création ou la gestion du document numérique. C'est ce que l'on nomme les « métadonnées système ». Par exemple, la date de création ou la date de modification d'un document numérique.

Les métadonnées peuvent également être assignées par un individu. En effet, certaines métadonnées, tel un titre, sont souvent attribuées par une personne afin de fournir des informations contextuelles précises sur le document numérique.

Par ailleurs, certaines métadonnées sont statiques, c'est-à-dire qu'elles ne changeront pas. Parmi les métadonnées statiques, pensons à la date de création. Alors que certaines métadonnées sont dynamiques et changeront durant le cycle de vie du document numérique, tels le format de fichier ou la date de modification d'un document.

1.1.5. Quand saisir les métadonnées ?

Selon leur mode de création, certaines métadonnées sont générées dès la création des documents numériques, d'autres au fil du cycle de vie. Elles permettent la gestion et la préservation des documents numériques durant tout leur cycle de vie.

1.1.6. Quel est le cycle de vie des métadonnées ?

Lors de la gestion et la préservation des documents numériques, il est important de considérer le cycle de vie des métadonnées qui peut être différent de celui du document. En effet, le cycle de vie des métadonnées peut être long, car les métadonnées peuvent être créées avant la création du document numérique, dans le cas d'un profil de métadonnées déjà existant (voir le chapitre 2 de la présente section) ou elles peuvent être conservées après la destruction du document pour en garder une trace.

1.2. Typologie des normes de métadonnées

Introduction

Les qualificatifs pour caractériser les métadonnées sont très nombreux : descriptives, administratives, contextuelles, structurelles, techniques, de pérennisation, statiques, évolutives, externes, de contenu... Certaines de ces données sur les données servent à organiser la connaissance et à utiliser et exploiter les

documents, d'autres servent à les maintenir accessibles dans le temps et à garantir et contrôler leur accès. Il est important de se rappeler que les métadonnées permettent la description des trois composantes d'un document numérique :

- **Contenu:** ce que le document contient; ce qui est intrinsèque.
- **Contexte:** informations extrinsèques aux documents.
- **Structure:** organisation du document.

Ainsi, divers types de métadonnées permettent la description de ces composantes. Nous allons ici présenter les grandes catégories de métadonnées.

1.2.1. Les métadonnées descriptives

Les métadonnées descriptives ^{P.22} servent à organiser la connaissance. Ce sont les métadonnées qui identifient, classifient, hiérarchisent l'information contenue dans le document numérique. Il s'agit typiquement d'un titre ou d'un nom, d'auteurs, de dates, de termes permettant la classification. Organiser le savoir ou du moins un domaine de connaissance est une activité potentiellement génératrice d'une quantité importante de métadonnées.

1.2.2. Les métadonnées techniques

Les métadonnées techniques ^{P.23} servent à identifier, caractériser, définir l'environnement technique des documents numériques. Ces données décrivent ce que peut contenir un format de représentation et comment l'exploiter.

1.2.3. Les métadonnées administratives

Les métadonnées administratives ^{P.22} servent à gérer la vie de l'objet numérique. Une partie de ces métadonnées administratives est également utilisée dans le processus de recherche d'information.

1.2.4. Les métadonnées d'identification

Les métadonnées d'identification ^{P.22} permettent l'identification univoque des objets archivés. Généralement, il doit s'agir d'un identifiant pérenne (à ce sujet, section 2, chapitre 2). L'objectif est d'adopter un identifiant qui résistera à l'épreuve du temps, aux évolutions de classement intellectuel des contenus, aux changements d'organisation physique des données. Il doit être adaptable.

1.2.5. Les métadonnées de provenance et de contexte

Les métadonnées de provenance et de contexte ^{P.22} informent sur la vie du document numérique. C'est le comment et le pourquoi : d'où vient le document numérique ? Comment l'information a-t-elle été colligée ? Par quel moyen technique (réception de signal, numérisation) ? Quels traitements ont été effectués ? Par qui ? Pourquoi ? Quand ? Quelles sont les raisons qui ont motivé sa création, sa collecte, sa réception ? Quelle confiance pouvons-nous avoir dans cette source ?

1.2.6. Les métadonnées d'intégrité

Les métadonnées d'intégrité ^{P.22} permettent de disposer d'informations relatives au respect de l'intégrité des documents. Ces informations sont à usage interne (surveillance et contrôle des documents) et/ou à usage externe (prévenir des litiges). Il s'agit par exemple d'empreintes des fichiers à conserver obtenues sur la base de procédés cryptographiques.

1.2.7. Les métadonnées de gestion des droits

Les métadonnées de gestion des droits ^{p.22} permettent de gérer le statut légal du document numérique. Rentrent dans cette catégorie, pour les archives publiques, les délais de communicabilité fixés par la loi ou la réglementation. Il s'agit des métadonnées qui vont permettre de définir la politique de diffusion selon les contraintes légales ou imposées par l'entité détentrice des droits. La concrétisation formelle de cette politique peut être une licence. On parle de droits d'usage et de droits d'accès. Ils définissent en fonction du contexte d'utilisation : Qui ? Avec quoi ? (télévision, ordinateur) ce qu'il est permis de faire : trouver, voir, imprimer, copier, modifier, détruire, et les contraintes d'usage : qualité, quantité, prix...

1.2.8. Les métadonnées de structure

Les métadonnées de structure ^{p.22} servent à connaître l'organisation de l'information contenue et des documents numériques. Il y a deux niveaux de structure :

- Le niveau logique définit les liens entre des éléments qui ont du sens pour l'utilisateur : numéro de page, de plages audio, titre de chapitres, d'articles, etc.
- Le niveau physique définit comment sont enregistrés les documents numériques : dans quel fichier ? Dans quel répertoire ? Sur quel support ?

1.2.9. Les métadonnées sur la préservation

Les métadonnées sur la préservation ^{p.22} contiennent certaines informations pour assurer l'archivage et la préservation à long terme des documents numériques. Il faut préciser que les métadonnées, même si elles contribuent à la conservation à long terme des documents, doivent être identifiées, consignées et gérées dès le stade actif des documents numériques.

2. Normes et profils de métadonnées pour décrire les documents numériques



Introduction

Ce chapitre présentera diverses normes utilisées pour décrire les documents numériques de diverses natures. Il présentera également les étapes de création d'un profil de métadonnées qui peut être utilisé au sein d'un organisme. Un profil permet d'encadrer les pratiques de création et de gestion des métadonnées.

2.1. Les composantes des métadonnées

Tout d'abord, pour décrire un document numérique, on doit reconnaître ses **propriétés** (attributs ou éléments). En effet, les métadonnées sont composées d'une liste de propriétés qui correspondent aux caractéristiques de l'objet numérique à décrire. Par exemple, un document numérique pourrait avoir les propriétés suivantes : un titre, un ou des auteur(s), un format. Tandis qu'un objet muséal pourrait avoir les propriétés suivantes : un titre, des dimensions, des matériaux, etc. Ainsi, certaines propriétés peuvent être communes aux objets numériques à décrire tandis que d'autres propriétés peuvent être différentes selon la nature de l'objet numérique à décrire.



Les propriétés pertinentes pouvant être utilisées pour décrire des documents d'archives numériques sont regroupées au sein d'un **vocabulaire** ou d'un **schéma de métadonnées** ^{p.23}. Ainsi, un schéma de métadonnées identifie et définit les propriétés. Il précise également les règles d'application de chaque propriété pour décrire les documents numériques.

Pour chacune des propriétés, des valeurs sont attribuées ou assignées. Les **valeurs** correspondent à la description des documents numériques (voir Tableau 1).

Nom de la propriété	Valeur
Titre	Procès-verbal du conseil d'administration
Format	.docx
Date de création	2022-01-01

Tableau 1. Association propriétés et valeurs. © Patrick Beaulieu révisé par Natasha Zwarich, ARV2030 Gestion électronique des documents, Université du Québec à Montréal, 2008-2020.



Cependant, pour être pleinement efficaces, les métadonnées doivent faire l'objet d'une normalisation. La sous-partie suivante présentera les principales normes applicables aux métadonnées.

2.2. Les normes de métadonnées

Introduction

Plusieurs normes de métadonnées existent. Elles servent à encadrer l'utilisation des différentes composantes d'un vocabulaire de métadonnées. Voyons les principales normes existantes. Nous nous limiterons aux normes spécifiquement utilisées pour les documents numériques. D'autres normes existent communes aux documents d'archives analogiques et numériques (Règles de description des documents d'archives - RDDA ; Norme générale et internationale de description d'archives - ISAD-G; Encoded Archival Description - EAD, Encoded Archival Context - EAC, Records in Context - RIC, etc.), mais elles ne seront pas examinées ici.

2.2.1. Les normes de contenu

Les **normes de contenu** fournissent des lignes directrices sur le contenu des descriptions, soit la rédaction des valeurs. Elles permettent d'uniformiser le format, la syntaxe et la sémantique des descriptions.

2.2.2. Les normes de structure

Les **normes de structure** identifient les propriétés ou éléments nécessaires à la description d'objets numériques. Les normes de structure sont nombreuses, compte tenu de la diversité d'objets numériques existants. Il s'agit de vocabulaire ou de schéma. Parmi les normes applicables à la gestion des documents actifs, mentionnons Dublin Core.



Dublin Core. Le vocabulaire Dublin Core est sous la responsabilité d'un organisme de normalisation, le DCMI (Dublin Core Metadata Initiative). La norme Dublin Core est conçue pour décrire plusieurs types d'objets numériques, tels des documents textuels, des images ou encore des ressources Web. Elle se veut simple à utiliser puisqu'elle compte peu de propriétés. Il s'agit d'une norme de portée internationale qui a été adoptée par l'ISO : ISO 15836-1:2017 Information et documentation — L'ensemble des éléments de métadonnées Dublin Core — Partie 1: Éléments principaux et ISO 15836-2:2019 Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes. Le schéma de métadonnées Dublin Core ne compte que 15 propriétés avec la possibilité d'ajouter des raffinements pour plus de précision. De par sa simplicité, il se veut une base pour la création de profils de métadonnées (voir section suivante) et pour faciliter l'interopérabilité entre les vocabulaires de métadonnées. C'est une norme surtout utilisée pour la diffusion mais le nombre restreint d'éléments peut limiter son efficacité dans la description archivistique.

2.2.3. Les normes visant à encadrer le vocabulaire

Ces normes visent à encadrer le vocabulaire utilisé dans une description à partir d'une liste de valeurs prédéfinies tel un thésaurus. À ce sujet, il est possible de consulter les référentiels du standard d'échange des données pour l'archivage (SEDA) créé et maintenu par le Service interministériel des Archives de France (<https://francearchives.gouv.fr/seda/>). Cet élément sera abordé plus en détail au chapitre 3 Indexation.

2.2.4. Les normes visant à faciliter l'échange de données

Les **normes visant à faciliter l'échange de données** permettent l'encodage de descriptions de façon à les transmettre via un système numérique tels XML, Metadata Encoding and Transmission Standard (METS) ou encore la norme ISO2709 : 2008 Information et Documentation – Format pour l'échange d'information.

2.2.5. L'archiviste doit accorder une grande importance à ces normes

L'archiviste doit accorder une grande importance à ces normes de façon à ce que d'un système numérique à l'autre, la signification des propriétés et des valeurs soient maintenues et ainsi permettre le développement de passerelles (« crosswalks ») entre les plateformes.



Un archiviste qui utilise la norme EAD aura une propriété qui s'intitule « creator » (pour créateur/producteur). Un archiviste pourrait également avoir sous sa responsabilité d'autres types de documents numériques pour lesquels on utilise une propriété qui s'intitule « auteur ». Or, les termes « auteur » et « créateur » ne sont pas des termes équivalents et pourraient engendrer des descriptions inexactes et interprétées de façon différente par des systèmes numériques.

Les normes encadrant l'usage des métadonnées sont nombreuses et touchent l'une ou l'autre des composantes des métadonnées. Le vocabulaire de métadonnées peut également contribuer, comme nous le verrons dans la sous-partie suivante, à développer un ou des profils de métadonnées.

2.3. Développement d'un profil de métadonnées

Un vocabulaire ou schéma de métadonnées permet de créer des profils de métadonnées qui serviront à décrire divers types de documents numériques. Un **profil de métadonnées**^{p.23} propose un ensemble de propriétés pouvant être utilisé pour décrire un type de document numérique ainsi que leurs règles d'utilisation. En effet, dans un profil de métadonnées, il faut identifier si une propriété doit être obligatoire pour décrire le document numérique (voir Tableau 2).

Propriété	Obligatoire	Répétable
Type	Oui	Non
Nom du comité	Non	Non
Date de la réunion	Oui	Non
Personne présente	Non	Oui

Tableau 2. Règles d'utilisation pour un profil de métadonnées © Patrick Beaulieu révisé par Natasha Zwarich, ARV2030 Gestion électronique des documents, Université du Québec à Montréal, 2008-2020.

Ainsi, à partir des propriétés contenues dans un vocabulaire, il est possible de créer plusieurs profils de métadonnées (Figure 4). Il pourrait y avoir un profil pour décrire les procès-verbaux, un profil pour décrire les contrats, un profil pour décrire des règlements, etc.

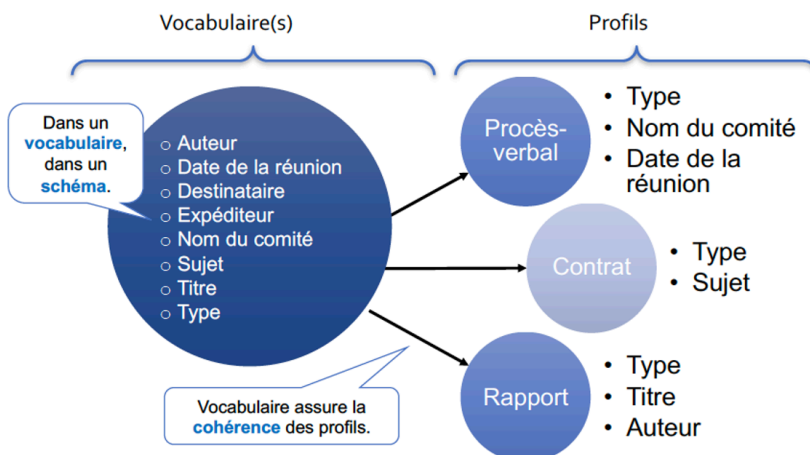


Figure 4. Création d'un profil de métadonnées à partir d'un vocabulaire. © Patrick Beaulieu révisé par Natasha Zwarich, ARV2030 Gestion électronique des documents, Université du Québec à Montréal, 2008-2020.

Sur la base de ces profils, il serait ainsi possible d'effectuer une description uniforme des grandes séries de documents numériques que possède une organisation (Figure 4).

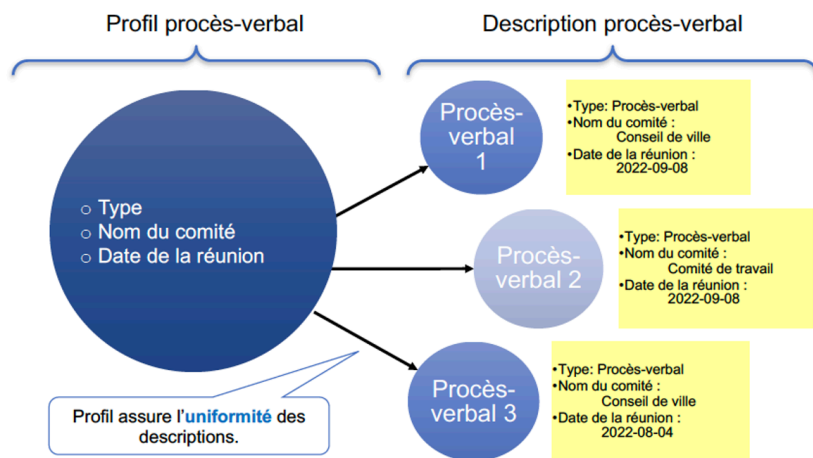


Figure 5. Utilisation d'un profil de métadonnées pour effectuer une description. © Patrick Beaulieu révisé par Natasha Zwarich, ARV2030 Gestion électronique des documents, Université du Québec à Montréal, 2008-2020.

Ainsi, pour élaborer des profils de métadonnées pour décrire les documents numériques de son organisme, il importe de :

- Identifier les types de documents numériques à décrire.
- Évaluer et prioriser les types de documents à décrire.
- Par type de document, analyser les caractéristiques et identifier les propriétés pertinentes pour la description.
- Si possible, réutiliser les propriétés déjà existantes dans le vocabulaire ou le schéma.
- Si nécessaire, créer de nouvelles propriétés et les ajouter au vocabulaire ou schéma.

- Mettre en application le profil de métadonnées.



La création de profils de métadonnées requiert une connaissance et une analyse fine des documents numériques et de leur utilisation. En effet, il faut tenir compte du niveau de détail requis, à savoir s'il est nécessaire de procéder à une description détaillée ou non. Il faut également considérer le mode de création des métadonnées, à savoir si les valeurs peuvent être générées automatiquement ou si elles doivent être saisies manuellement. Ce dernier élément est particulièrement important puisqu'une saisie manuelle peut considérablement alourdir le processus de description. La description des documents numériques doit être effectuée selon les besoins de recherche, d'utilisation et de gestion des documents. Elle doit également être effectuée en considérant les besoins futurs, notamment la préservation à long terme des documents numériques.

En France, le standard d'échange des données pour l'archivage (SEDA) donne un cadre à la création de profils et des outils existent pour faciliter leur élaboration comme le Profil d'archivage simple pour le traitement de l'Information en SEDA (PASTIS) (<https://download.programmevitam.fr/pastis-standalone/>) ou le Générateur automatisé de SEDA pour profils d'archivage Gaspar (<https://gaspar.mintika.fr/>).

En somme, la gestion des documents numériques requiert nécessairement des métadonnées. Or, différents enjeux doivent être considérés.

2.4. Enjeux des métadonnées

La description de documents numériques pose certains défis. Dans cette partie, nous ferons un survol des principaux enjeux que pose l'utilisation des métadonnées dans la description des documents numériques et auxquels l'archiviste doit porter une attention particulière.

- **Qualité.** Diverses catégories de métadonnées permettent la description, la gestion, l'utilisation et la préservation d'un document numérique. Par conséquent, il importe que ces métadonnées soient de qualité. En effet, elles doivent refléter le contenu du document, représenter les activités menant à sa création et générer des métadonnées système fiables.
- **Granularité et précision.** La granularité réfère au niveau de détail d'une description. Le niveau de détail des descriptions des documents numériques doit être de granularité similaire de façon à permettre un meilleur accès de ces ressources. Ainsi, plus une description est fine, plus l'accès sera précis. Cependant, ce niveau de précision peut causer du silence (c'est-à-dire l'absence de résultats pertinents) ou beaucoup de bruits lors du repérage.
- **Exhaustivité et pertinence.** Il s'agit de s'assurer que toutes les caractéristiques pertinentes des documents numériques sont bien documentées tout en tenant compte de la faisabilité de la collecte de ces métadonnées et de ce qui est nécessaire aux fins de gestion et de préservation.
- **Interopérabilité.** L'interopérabilité ^{p.22} est la « capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque. » (Office québécois de la langue française, 2007). L'interopérabilité est un enjeu très important au niveau des métadonnées. En effet, dans l'environnement numérique, il est nécessaire que l'archiviste s'assure de la prise en charge des métadonnées par les différentes applications et de la possibilité de partager les métadonnées par le biais des plateformes numériques. L'interopérabilité repose notamment sur l'usage de normes. L'archiviste doit également considérer les divers types d'interopérabilité.

L'interopérabilité sémantique assure une compréhension commune des informations échangées. Ainsi, pour atteindre l'interopérabilité sémantique, l'archiviste doit utiliser un modèle de référence commun. Le recours à des vocabulaires contrôlés communs pour saisir les valeurs rend ainsi les métadonnées compréhensibles à travers les systèmes.

L'interopérabilité syntaxique réfère à la forme des descriptions de façon à formater les métadonnées pour qu'elles soient traitées correctement par des systèmes informatiques. Par exemple, pour le nom du créateur (ou producteur) du document, il est important de déterminer si le nom de famille doit être inscrit avant le prénom ou si la forme prénom et nom est acceptée.

L'interopérabilité technique, quant à elle, réfère à la communication et à l'échange de données entre des systèmes différents au moyen de normes et de protocoles largement adoptés.

En somme, la description des documents numériques, dès le stade actif, s'effectue grâce aux métadonnées. Les métadonnées fournissent des informations sur les documents numériques et leur qualité influe sur la gestion, la préservation et le repérage de ceux-ci. L'uniformisation des métadonnées et le recours à des normes contribuent à améliorer leur partage et leur diffusion.

3. L'indexation



Introduction

L'*indexation* ^{p.22} permet de représenter de manière concise le contenu d'un document numérique pour améliorer la précision du repérage. Pour permettre une meilleure gestion et conservation à long terme des documents numériques, l'archiviste doit se familiariser avec les principes de l'indexation. Ce chapitre vise à présenter les principes et pratiques d'indexation des documents numériques **au stade actif**.

3.1. Définition de l'indexation

Le chapitre précédent a montré l'importance des métadonnées dans la description des documents numériques. Tel qu'indiqué, le recours à des normes, notamment pour la rédaction des valeurs, facilite le repérage et l'uniformisation des descriptions.



Définition

L'*indexation* ^{p.22} est l'« opération permettant de décrire et caractériser le contenu thématique d'un document à l'aide de représentations verbales » (Hudon, 2013). Elle consiste, d'une part, à analyser les documents numériques de façon à identifier les principaux concepts abordés dans le document et, d'autre part, à vérifier si ces concepts sont contenus dans une norme de façon à sélectionner les termes d'indexation appropriés. L'indexation vise, à partir de termes prédéfinis, à « décrire brièvement et clairement le contenu d'un document, faciliter le repérage à l'aide de représentations verbales et améliorer la pertinence du repérage » (Hudon, 2013).

3.2. Les divers types d'indexation et modes de création

3.2.1. Indexation par assignation

L'indexation par assignation est effectuée par un archiviste qui identifie les principaux concepts et thèmes du document numérique, ce qui permet une indexation fiable. L'archiviste considère les besoins de ses usagers de façon à optimiser le repérage des documents utiles dans l'exercice de leurs activités.

3.2.2. Indexation par extraction (« automatique »)

L'indexation automatique s'effectue au moyen d'algorithmes qui extraient les principaux concepts d'un document numérique.

3.2.3. Indexation en vocabulaire libre (langage naturel)

Les valeurs peuvent être rédigées en *vocabulaire libre*, c'est-à-dire qu'aucun « contrôle n'est effectué au niveau des valeurs saisies » (Hudon, 2013) et la personne choisit librement les termes de son choix. Ainsi, deux archivistes qui effectuent une description de documents numériques pourraient utiliser le vocabulaire de leur choix. Bien que le vocabulaire libre offre des avantages quant au choix des termes utilisés, le niveau de précision et le vocabulaire utilisé risquent de varier selon l'auteur de la description.

3.2.4. Indexation en vocabulaire contrôlé (langage documentaire)

Pour améliorer le niveau de précision, uniformiser le vocabulaire et faciliter le repérage, l'archiviste peut utiliser un *vocabulaire contrôlé*. Le vocabulaire contrôlé consiste à choisir les termes d'indexation dans une liste de valeurs possibles (*langage documentaire* ^{p.22}). Le vocabulaire contrôlé est habituellement sous la responsabilité d'un organisme ou association qui a la responsabilité de le faire évoluer. Par ailleurs, l'archiviste doit tenir compte des normes de syntaxe contrôlée, c'est-à-dire qu'il faut que certaines valeurs respectent une syntaxe précise.



Un langage documentaire est un « sous-ensemble structuré de termes extraits de la langue naturelle et dont on a contrôlé la forme et la signification » (Hudon, 2013).

Le langage documentaire permet notamment d'augmenter la précision lors de la recherche et de diminuer l'impact « des problèmes causés par la synonymie et l'homographie de la langue naturelle » (Hudon, 2013).



Il existe différents langages documentaires qui peuvent être pertinents pour l'archiviste. Parmi ceux-ci, mentionnons les normes visant à proposer des termes d'indexation : Library of Congress Subject Headings (LCSH) ou Répertoire de vedettes matières de la Bibliothèque de l'Université Laval (RVM). L'archiviste peut également utiliser un thésaurus pour identifier des termes d'indexation pertinents. En effet, un thésaurus fournit un lexique, des définitions, des notes d'application ainsi que des relations sémantiques avec d'autres termes apparentés. De nombreux thésaurus sont proposés par le ministère de la Culture français, comme le thésaurus pour l'indexation des archives locales ou les listes d'autorité typologie documentaire, contexte historique (<http://data.culture.fr/thesaurus/page/vocabulaires>).

Conclusion

Les différents types d'indexation peuvent coexister de façon à offrir davantage de précision dans la description et améliorer le repérage. La nature des documents numériques à décrire, les besoins des usagers ainsi que les plateformes numériques utilisées orientent le type d'indexation à privilégier.

3.3. Le cas particulier de l'indexation collaborative et participative

Depuis quelques années, nous assistons, en archivistique, à une participation accrue des usagers dans l'indexation de documents numériques. En complément de la description effectuée par l'archiviste, les usagers peuvent contribuer à la description par l'ajout de mots-clés. Ainsi, l'indexation collaborative s'effectue sur des petits ensembles de documents et à des niveaux plus fins.

La participation des usagers à la description peut contribuer à bonifier les outils de gestion documentaire ou les thésaurus déjà en place dans l'organisme. Les termes ajoutés par les usagers peuvent être validés par l'archiviste qui pourra confirmer leur pertinence. Par exemple, pour une photographie, l'archiviste peut la décrire en utilisant les normes applicables dans son organisme et l'utilisateur peut y ajouter des mots-clés qui lui semblent pertinents. Il est possible de consulter le site Colab de Bibliothèque et Archives Canada pour participer à une indexation collaborative entre institution et usagers (<https://colab.bac-lac.gc.ca/fra>). Ainsi, une telle démarche de collaboration peut non seulement améliorer le repérage, mais également favoriser la réutilisation qui, à terme, peut permettre de développer de nouveaux outils de repérage telle la folksonomie, qui sera abordée dans la section suivante. L'indexation collaborative est, pour les usagers, facile puisqu'ils peuvent indiquer les termes qui leur semblent pertinents pour décrire les documents.

Conclusion



La description et l'indexation des documents numériques sont des fonctions archivistiques très importantes dans l'environnement numérique. En effet, les métadonnées qui sont associées aux documents numériques permettent, non seulement leur description, mais également leur gestion et leur conservation à long terme. Sans des métadonnées de qualité, il sera difficile, voire impossible, de mettre en œuvre une gestion des documents numériques adéquate.

Glossaire



Indexation

L'indexation est l'« opération permettant de décrire et caractériser le contenu thématique d'un document à l'aide de représentations verbales » (Hudon, 2013). L'indexation vise, à partir de termes prédéfinis, à « décrire brièvement et clairement le contenu d'un document, faciliter le repérage à l'aide de représentations verbales et améliorer la pertinence du repérage » (Hudon, 2013).

Interopérabilité

L'interopérabilité est la « capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque. » (Office québécois de la langue française, 2007).

Langage documentaire

Un langage documentaire est un « sous-ensemble structuré de termes extraits de la langue naturelle et dont on a contrôlé la forme et la signification » (Hudon, 2013).

Métadonnées administratives

Les métadonnées administratives sont les métadonnées qui servent à gérer la vie de l'objet numérique.

Métadonnées d'identification

Les métadonnées d'identification permettent l'identification univoque des objets archivés.

Métadonnées d'intégrité

Les métadonnées d'intégrité permettent de disposer d'informations relatives au respect de l'intégrité des objets.

Métadonnées de gestion des droits

Les métadonnées de droits permettent de gérer le statut légal de l'objet numérique.

Métadonnées de préservation

Les métadonnées sur la préservation qui contiennent certaines informations pour assurer l'archivage et la préservation à long terme des documents numériques.

Métadonnées de provenance et de contexte

Les métadonnées de provenance et de contexte informent sur la vie de l'objet numérique.

Métadonnées descriptives

Les métadonnées descriptives sont les métadonnées qui servent à organiser la connaissance. Ce sont les métadonnées qui identifient, classifient, hiérarchisent l'information contenue dans l'objet numérique.

Métadonnées de structure

Les métadonnées de structure servent à connaître l'organisation de l'information contenue et des objets numériques.

Métadonnées techniques

Les métadonnées techniques sont les métadonnées qui servent à identifier, caractériser, définir l'environnement technique des objets numériques.

Profil de métadonnées

Un profil de métadonnées propose un ensemble des propriétés pouvant être utilisé pour décrire un type de document numérique ainsi que leurs règles d'utilisation.

Schéma de métadonnées

Un schéma de métadonnées identifie et définit les propriétés. Il précise les règles d'application de chaque propriété pour décrire les documents numériques.

Bibliographie



[01] COUTURE, Carol et Marcel LAJEUNESSE. (2014). L'archivistique à l'ère du numérique. Les éléments fondamentaux de la discipline, Sainte-Foy, Presses de l'Université du Québec.

[03] HUDON, Michèle. (2013). Analyse et représentation documentaires. Introduction à l'indexation, à la classification et à la condensation des documents, Sainte-Foy, Presses de l'Université du Québec.

[04] ISO 15836-1:2017 Information et documentation — L'ensemble des éléments de métadonnées Dublin Core — Partie 1: Éléments principaux

[05] ISO 15836-2:2019(en) Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes

[06] OBSERVATOIRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC (2017). État des lieux sur les métadonnées relatives aux contenus culturels, Québec, Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, 118 p., [En ligne]. [www.stat.gouv.qc.ca/observatoire].

[07] OTT, Florence. (2021). La gestion documentaire au cœur des processus d'affaires. Londres, UK, ISTE Éditions.

[08] OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE. (2007). Grand dictionnaire terminologique, Gouvernement du Québec. En ligne : <https://vitrinelinguistique.oqlf.gouv.qc.ca/>

Webographie



[02] École de bibliothéconomie et des sciences de l'information (EBSI). (2018). Terminologie en sciences de l'information. En ligne : <https://clip.ebsi.umontreal.ca/terminologie/sci/fr/>.