

ÉTUDE

*L'indexation: luxe ou nécessité?**

Michel Lévesque

Il suffit aujourd'hui d'entrer dans une librairie, chez un marchand de journaux ou de naviguer sur Internet pour constater que l'information constitue la nourriture essentielle à tout être social et sociable. Afin de mieux faire circuler toute cette information, on construit même des autoroutes de l'information. Le cellulaire est à la mode et la mémoire collective se mesure en megaoctets. Risque-t-on les embouteillages et pertes de contrôle?

L'indexation de l'information institutionnelle s'avère probablement le moyen de contrôle privilégié. Elle n'est malheureusement pas la fonction la mieux exploitée dans un système de gestion des documents. Les archivistes minimisent l'importance des descripteurs pour le repérage et l'accès à l'information. Ils se préoccupent beaucoup plus de description en croyant que le repérage de l'information est assuré. Ils auraient intérêt à suivre et à adapter les travaux des spécialistes en indexation.

Notre article se présente en deux parties. La première est une mise en perspective et une analyse de concepts et des outils associés à la description des documents institutionnels, telles que la classification, l'inscription d'information aux bordereaux de documents, les métadonnées, les définitions de types de documents (DTD) et les *Règles de description de documents d'archives* (RDDA). En nous appuyant sur un exercice de description de documents, nous chercherons à démontrer que la description s'avère partielle lorsqu'il s'agit de repérer le contenu des documents.

Dans la deuxième partie, nous traiterons spécifiquement de l'indexation, du contrôle du vocabulaire, des thésaurus, de repérage plein texte et de stratégies de recherche. Tout en gardant un esprit critique, notre but est de sensibiliser les archivistes à l'importance de l'indexation comme moyen le plus efficace pour favoriser une gestion stratégique de l'information plutôt qu'un travail d'orfèvrerie inutile et coûteux.

* Ce texte est une version révisée d'une conférence donnée par l'auteur dans le cadre du 29^e congrès de l'Association des archivistes du Québec tenu à Montréal les 1^{er}, 2 et 3 juin 2000.

DE QUELQUES CONSIDÉRATIONS PRÉLIMINAIRES SUR LES PLANS DE CLASSIFICATION, LES BORDEREAUX DE VERSEMENT, LES MÉTADONNÉES, LES DÉFINITIONS DE TYPES DE DOCUMENTS ET LES RÈGLES DE DESCRIPTION DES DOCUMENTS D'ARCHIVES

Dans cette première partie, nous traiterons des outils qui habituellement sont associés au repérage de l'information. En quoi servent-ils véritablement au repérage et en quoi sont-ils des sources appropriées pour l'indexation?

Afin d'illustrer notre propos, nous avons décidé d'utiliser les documents et le dossier suivants et d'en faire une analyse de contenu:

- Un diplôme universitaire de 2^e cycle;
- Une lettre de la présidente du Groupe des responsables de la gestion des documents au gouvernement du Québec (GRGD) au président du Conseil du trésor;
- Un dossier sur le projet pilote en ingénierie documentaire au Greffe du Conseil du trésor;
- Une affiche sur le programme Allocation logement (Société d'habitation du Québec et ministère du Revenu).

1) Diplôme

Tel que défini dans le Robert, un diplôme est un «acte qui confère et atteste [...] un grade». Il s'agit dans le cas présent d'un acte émis par l'Université de Montréal pour attribuer à Michel Lévesque et lui attester le grade de maître en bibliothéconomie et en sciences de l'information. Il indique une mention relative à des études du programme de 2^e cycle. C'est un document sur support papier de dimensions 8 1/2" par 11". Il porte les signatures du recteur, du secrétaire général, du doyen et du secrétaire de la faculté, ainsi que les armoiries et le sceau de l'Université. Il est en langue française et est daté du 14 octobre 1986.

Transmis par le bureau du registraire, le diplôme est conservé dans une enveloppe qui ne comporte pas de titre, ni de cote.

2) Lettre

Telle que définie dans le Robert, une lettre est un «écrit que l'on adresse à quelqu'un pour lui communiquer quelque chose». Il s'agit dans le cas présent d'une lettre adressée à M. Jacques Léonard, ministre d'État à l'Administration et à la Fonction publique ainsi que président du Conseil du trésor, et signée par M^{me} Cécile LeBlanc, présidente du Groupe des responsables de la gestion des documents au gouvernement du Québec. Un acronyme est utilisé pour ce groupe: GRGD. Des copies conformes sont aussi adressées à la ministre de la Culture et des Communications, M^{me} Agnès Maltais, au ministre délégué à l'Autoroute de l'information et aux Services gouvernementaux, M. David Cliche, et finalement au conservateur des Archives nationales du Québec, M. Robert Garon.

En résumé, l'objet de cette lettre est d'obtenir du Conseil du trésor une confirmation, et les motifs qui la justifient, de sa position de ne plus reconnaître certains groupes d'intérêt, dont le GRGD. Une demande de reconnaissance officielle avait été

demandée et retirée par la ministre de la Culture et des Communications. Le GRGD évoque son importance en tant que regroupement, sa collaboration et son expertise pour justifier qu'il devrait être traité comme d'autres regroupements reconnus, tels le Groupe d'action en gestion financière (GAGF), le Comité consultatif en gestion du personnel (CCGP) et le Conseil des responsables de l'informatique du secteur public (CRISP).

Cette lettre a été préparée par le responsable du comité Avis, expertise et représentation du GRGD. Elle constitue une pièce supplémentaire au dossier sur le sujet. Le dossier a pour titre: *Reconnaissance officielle du GRGD* et ne comporte pas de cote. La lettre est datée du 19 mai 2000. Elle est sur support papier, dans un format 8 1/2" par 11" et est constituée de deux pages rectos.

3) Projet

Tel que défini dans le Robert, un projet est un «travail, rédaction préparatoire; premier état». Il s'agit dans le cas présent d'un ensemble de documents constitutifs d'un projet pilote en ingénierie documentaire, portant particulièrement sur la gestion des demandes d'autorisation et d'avis au Greffe du Conseil du trésor. Une équipe de recherche y est mentionnée: Danielle Bélanger, Lucie Dion, Claire Fecteau, Lise Readman et Michel Lévesque.

Un premier document est une fiche descriptive du projet intitulé *Gestion intégrée des documents papier/électroniques et circuit de production au Greffe du Conseil du trésor*. Ce document a été préparé par Nicole Boulet, Suzanne Couture et révisé par Michel Lévesque. Il établit le contexte et la problématique, la description sommaire du projet, les objectifs (résultats attendus et obtenus), les impacts budgétaires, le calendrier de réalisation, les impacts, les facteurs de risque, les remarques et commentaires. Il est daté du 23 septembre 1999.

Ce document est sur support électronique. Il a été sauvegardé dans un fichier appelé «Projet 1» du répertoire «CT». Le texte est sous format *Microsoft Word*.

Le deuxième document comporte quatre parties: 1) la page titre établissant un nouveau nom de projet: *Projet pilote en ingénierie documentaire/Gestion des demandes d'autorisation et d'avis au Greffe du Conseil du trésor*; 2) la description du processus d'affaires: l'identification du Greffe, le rôle du Greffe et les étapes du processus de la gestion des demandes d'autorisation et d'avis; 3) le modèle conceptuel, sous forme graphique, de la gestion des demandes par activités; 4) la description, sous forme de tableau, de chacune des activités. Il est daté du 6 décembre 1999 (bien que le tableau descriptif des activités se mette à jour automatiquement).

Ce document est sur support électronique. Il a été sauvegardé dans plusieurs fichiers appelés «Projet 2», «Projet 3» et «Projet 4» du répertoire «CT». Les projets 2 et 4 sont sous format *Microsoft Word* tandis que le projet 3 a été conçu avec un logiciel spécialisé dans la représentation de *workflow*.

Ces deux documents ont été élaborés par le chargé de projet (sauf le modèle conceptuel qui a été réalisé par l'informaticienne). Ce dernier relève de la Direction des ressources matérielles. Par contre, l'équipe précédemment nommée comprend une personne du Greffe, deux personnes du Sous-secrétariat à l'Inforoute gouvernementale

et aux Ressources informationnelles et une personne de la Direction des technologies de l'information.

4) Affiche

Telle que définie dans le Robert, une affiche est une «[f]euille imprimée destinée à porter quelque chose à la connaissance du public, et placardée sur les murs ou des emplacements réservés». Dans le cas présent, il s'agit d'une affiche produite conjointement par la Société d'habitation du Québec (SHQ) et le ministère du Revenu sur le programme Allocation logement. Elle a pour fonction de donner le numéro de téléphone du programme. Celui-ci est destiné aux personnes qui consacrent une part trop importante de leur revenu pour se loger et s'adresse particulièrement aux personnes de 55 ans ou plus et aux familles à faible revenu. L'image en couleurs montre un couple avec un enfant en train de lire un livre et deux images superposées en couleurs d'un homme et d'une femme ayant 55 ans ou plus. Le logo du gouvernement du Québec y apparaît ainsi que les noms des deux organismes.

Cette affiche est sur support papier et mesure 68,5 cm par 43 cm. Elle a été produite pour la Direction des communications par la firme Amalgame-Cargo, en janvier 2000.

Exercice de classification et identification des documents et des dossiers

1) Diplôme

En utilisant *Comment classer vos archives personnelles et familiales* (Héon 2000, 24-26), nous classerions ce document dans la série 6000 *Études et perfectionnement*, à la sous-série 6600 *Niveau universitaire* et à la subdivision 6600-10 *Activités administratives*.

Le titre du dossier devrait être: *Diplôme de 2^e cycle en bibliothéconomie et sciences de l'information*.

2) Lettre

En utilisant le *Guide de classification et de classement des documents des associations et autres organismes de même nature* (Casault et Lévesque 1997, 47), nous classerions cette lettre à la classe 9000 *Reconnaissance et développement de l'organisme*, à la sous-classe 9100 *Reconnaissance de l'organisme*.

Le titre du dossier devrait être: *Reconnaissance officielle du GRGD*.

3) Projet

En utilisant le *plan de classification* du Conseil du trésor, nous nous retrouvons devant un dilemme. La classe 1000 *Coordination et gestion interne de l'organisation*, sous-classe 1600 *Gestion des ressources informationnelles*, sous-division 1612 *Gestion documentaire*, subdivision 1612-01 *Système gestion documentaire*, pourrait être utilisée dans la mesure où il s'agit d'un exercice qui entre dans le cadre de la gestion documentaire (Québec 1999a, 15). Or, les catégories pour la classification sont larges, voire imprécises.

Comme ce projet s'inscrit dans la foulée des travaux réalisés en ingénierie documentaire, la classe d'exploitation 6000 *Activités liées aux technologies de l'information*, sous-classe 6200 *Gestion de l'information, des TI et de l'inforoute*, sous-division 6212 *Ingénierie documentaire* pourrait aussi convenir (Québec 1999a, 47). On remarque qu'ici encore les catégories sont larges.

Le titre du dossier devrait être: *Gestion intégrée des documents papier/électroniques en circuit de production au Greffe du Conseil du trésor - Projet pilote en ingénierie documentaire*.

4) Affiche

En utilisant le *plan de classification* de la Société d'habitation du Québec, nous nous retrouvons aussi devant un dilemme. La classe 6000 *Communications, relations publiques et externes*, la sous-classe 6300 *Relations publiques*, la division 6350, *Production graphique et visuelle*, subdivision *Affiches*, pourrait convenir dans la mesure où c'est le type de document qui prévaut (SHQ 1999, 6-43).

Or, comme elle est relative à un programme administré conjointement par la SHQ et le ministère du Revenu, nous pourrions envisager suivre la logique assimilant cette affiche à un document d'exploitation et privilégier la classe X5 *Programme d'aide aux personnes*, sous-classe X5-10 *Programme d'allocation logement unifiée*. Il est précisé dans le plan que les catégories de gestion peuvent être utilisées pour compléter le code d'exploitation. Ainsi, nous aurions le code suivant pour cette affiche: X5-10-6350-Affiches (SHQ 1999, 6-16, 6-43).

Le titre de l'affiche devrait être: *Programme d'allocation logement unifiée/Affiche*.

Problématique de la classification et de l'index des dossiers comme éléments de repérage

Plan de classification

On définit généralement le plan de classification comme «une structure hiérarchique et logique permettant [la classification], le classement et le repérage de pièces d'archives ou d'ensembles documentaires [...]» (ANQ 1996, 154). Mais de quel repérage s'agit-il?

Essentiellement, le plan de classification permet de réunir des documents et des dossiers de même nature en les associant à des catégories précises, préalablement codées. Ces catégories, qui auparavant représentaient la structure administrative, sont généralement développées par rapport aux fonctions et aux activités de l'organisme. Elles constituent l'essence même du plan. Comme elles sont représentées selon la logique allant du général au spécifique, on parle de hiérarchisation, bien que parfois nous croyions qu'il faille parler plutôt d'association, c'est-à-dire la liaison de concepts ayant des sens apparentés sans à proprement parler qu'un concept soit obligatoirement le subordonné d'un autre.

Ces fonctions et ces activités s'expriment par des termes formant des expressions: Gestion des ressources mobilières, Développement de prototypes de fabrication,

Gestion du programme X, Réunion du conseil d'administration, etc. Nous comprenons tous qu'il s'agit de la formulation de concepts ou de sujets. En ce sens, nous pouvons en déduire qu'un premier niveau de repérage est possible. Mais, il est insuffisant pour permettre de retrouver facilement des documents et des dossiers précis. Nous connaissons la fonction ou l'activité précise. Par contre, il ne nous permet pas d'avoir une connaissance des dossiers et des documents qui portent sur des sujets plus précis en lien avec cette fonction ou cette activité. Il est évident que le premier niveau pourra satisfaire certains chercheurs et administrateurs mais il demeure avant tout limitatif pour les besoins de la recherche poussée de documents et de dossiers, physiques ou virtuels.

La classification est effectivement «un processus intellectuel d'identification et de regroupement systématique d'[éléments] semblables, d'après des caractéristiques communes pouvant faire l'objet d'une différenciation si la quantité l'exige [...]», mais affirmer que son objectif ultime est de «faciliter le repérage de l'information que contiennent les documents et, ce faisant, de permettre de mieux appréhender le contexte de fonctionnement de la personne physique ou morale» (Couture et al. 1999, 18) ne nous semble pas suffisamment nuancé.

La classification sert surtout au repérage du contexte de fonctionnement. Comment pourrait-il en être autrement lorsque ce sont les fonctions et les activités qui président à cette opération? Ce faisant, elle contribue, en partie seulement, à la connaissance du contenu intellectuel de regroupement de documents. C'est d'ailleurs une des raisons pour laquelle on précise maintenant à l'aide des descriptions explicatives les catégories. Sans ces précisions, il était souvent difficile de classer des documents et des dossiers précis.

Quant au code de classification, il sert surtout d'élément de repérage des dossiers, et des documents qu'ils contiennent, dans un classeur ou sur une étagère. Il fonctionne comme une adresse civique. Plusieurs considèrent que le code est significatif. Mais de quoi? Il l'est tout simplement par un phénomène mnémotechnique qui permet entre autres de distinguer une catégorie générale d'une catégorie spécifique. De plus, la codification est une codification de fonctions et d'activités et non une codification de la connaissance (comme avec Dewey, Library of Congress, Raganathan).

La gestion électronique des documents électroniques nous amène à revoir le principe de la classification tel qu'il a toujours été conçu. Les documents électroniques font partie d'un autre univers, mais ils n'en conservent pas moins des caractéristiques communes avec ceux sur support papier. Ils participent tous au fonctionnement d'un organisme et contiennent des concepts et des sujets. Ils peuvent donc être traités selon un plan de classification mais celui-ci peut présenter une structure beaucoup plus souple tout en conservant un objectif de contextualisation. La souplesse de recherche que permettent les systèmes d'information nous amène à croire qu'une autre approche de la classification est possible. Il s'agit de la classification à facettes. Un groupe de travail au Conseil du trésor est en train d'explorer ce terrain!

Par ailleurs, nous pouvons rester perplexes devant un principe lié au traitement d'archives qui établit que lorsqu'on doute de la structure du fonds, on est autorisé à en changer l'ordre. Ce principe permet à la limite de douter de la classification

utilisée par le créateur d'origine et d'en établir une autre. Mais sur quelle base? À partir d'une connaissance du fonds obtenue comment? Les plans de classification qui en résultent ne sont-ils pas trop généraux, souvent marqués uniquement par une approche par type de documents? Pensons entre autres au plan de classification utilisé pour les archives d'individus. Nous croyons que même le premier niveau de repérage du contenu du fonds peut s'en trouver considérablement affecté, au point d'en arriver à une sous-utilisation du fonds.

Mais, malgré cela, les plans de classification ne sont guère différents de ceux utilisés à l'actif. Ils sont avant tout structurés selon les fonctions et les activités, se rapprochant ainsi de l'inévitable valeur de témoignage renseignant sur ce qu'était l'organisme, sans toutefois être vraiment précis quant à la nature même des sujets qui découlaient de ces fonctions et activités.

On assiste même à la limite à un constat d'échec sur la dimension informationnelle des documents qui, trop nombreux, malgré l'application de règles de conservation et donc de l'évaluation, ne peuvent être traités comme tel. On demeure ainsi à des niveaux structurels très généraux qui n'éclairent que très peu sur le contenu véritable des documents pour se centrer sur la dimension plus contextuelle. Les *Règles de description des documents d'archives* (RDDA), que nous aborderons plus loin, en témoigneront.

Index des dossiers

L'outil qui va permettre de mieux repérer l'information est l'index des dossiers. Mais tout dépend de quelle façon il aura été conçu. Grâce aux possibilités que procure maintenant l'informatique, nous pouvons constituer un index exhaustif de dossiers et de documents. Il ne faut pas se contenter toutefois d'inscrire seulement le titre du dossier. C'est insuffisant, même dans une perspective de recherche sur les mots du titre. Il faut ajouter une zone où sont véritablement circonscrits les sujets des documents, donc des dossiers qui le plus souvent les rassemblent. On comprendra qu'il s'agit véritablement ici de repérage du contenu du document et non du repérage contextuel du document.

Habituellement, nous retrouvons, dans les logiciels spécialisés en gestion des documents, des panoramas de gestion des dossiers comprenant des champs qui permettent de constituer cet index: le numéro de référence du dossier, le code de classification, le titre du dossier, le type d'exemplaire, le support, la date de création, la date de fermeture, l'unité administrative, les mots-clés, etc. Encore ici, on comprendra que tous ces champs, à l'exception du titre du dossier et des mots-clés, sont d'ordre plutôt contextuel que documentaire.

Or, nous minimisons trop souvent l'importance de l'analyse du contenu du ou des documents dans la détermination du titre du dossier et des descripteurs qui lui sont associés. Opération le plus souvent décentralisée dans l'organisme, c'est en règle générale la personne qui crée les documents, ou du moins qui les prend en charge, qui donne les titres et attribue les descripteurs. Comme elle n'est pas une spécialiste et qu'un contrôle trop rapide est fait par les responsables de la gestion des documents, nous nous retrouvons avec un ensemble de dossiers contextualisés et codés mais qui

ne révèlent qu'une infime partie de ce qu'ils contiennent. C'est d'ailleurs devant ce cul-de-sac que se retrouvent ceux qui obligent l'inscription obligatoire du titre de la catégorie classificatoire comme titre du dossier. Une personne, quelle qu'elle soit, est plus à l'aise lorsqu'elle est en mode inné qu'en mode acquis. Ne suffit-il pas de voir comment réagissent certaines personnes lorsqu'elles ont à classer des documents, et surtout à attribuer un code de classification? Nous comprendrons aussi pourquoi de grandes séries de dossiers homogènes sont devenues la matière première du spécialiste en système d'information en mal de banques de données documentaires, qui elles-mêmes auraient par contre avantage à être mieux décrites. Il faut insister davantage sur l'importance des titres et des descripteurs significatifs, et, pour ces derniers, en nombre suffisant pour exprimer l'ensemble des concepts présents dans le document. D'où la nécessité d'une formation adéquate, qui met l'accent sur ce qui est fondamental, et la production d'un vocabulaire d'indexation facilement utilisable. Nous y reviendrons plus loin.

Bordereaux de transfert ou de versement

Dans la mesure où l'index est maintenant constitué de façon électronique, il est possible d'en extraire des informations pour permettre entre autres la gestion des documents semi-actifs. Il n'est plus vraiment nécessaire de faire des bordereaux de transfert tels qu'on les faisait auparavant. Il suffit de joindre la liste issue du système à un bordereau abrégé contenant les indications nécessaires pour la gestion des semi-actifs. Le repérage des dossiers pourra se faire à même la banque des données des dossiers qui auront préalablement été identifiés comme semi-actifs.

Il est vrai que bon nombre d'institutions d'archives se sont servi et se servent encore de ces bordereaux comme d'un instrument de recherche temporaire avant le traitement final des archives versées. Mais encore faut-il que les titres des dossiers inscrits sur ces bordereaux soient significatifs et fassent l'objet d'une saisie dans un système informatique. Certains d'ailleurs les numérisent pour faire de la recherche plein texte sur les titres. Il est curieux que cette pratique tienne encore lorsque ces informations sont déjà enregistrées dans les systèmes de gestion documentaire des organismes producteurs. Il suffirait de conclure une entente et de récupérer les données relatives aux dossiers versés. Il faudrait éviter des situations où, tel que cela se faisait par le passé, on récupérerait d'importants ensembles de documents sans récupérer les index qui les accompagnaient. Serait-on plus préoccupé par la description du contexte dans lequel sont créés les documents et moins par la description de leur contenu?

Métadonnées

La production des documents électroniques risque encore une fois de modifier les approches. Nous comprenons tous que l'objet à traiter est le document et que celui-ci fait toujours partie d'un dossier, quoique virtuel. Les études et projets vont tous dans le même sens: l'établissement et l'association de métadonnées. Pour pouvoir repérer un document, celui-ci doit être accompagné de plusieurs métadonnées. Certaines sont intrinsèquement comprises dans le document lui-même, tandis que d'autres doivent y être ajoutées.

Les travaux du chantier en ingénierie documentaire réalisés par le Conseil du trésor, avec la participation du GRGD, des Archives nationales du Québec et de la Bibliothèque nationale du Québec ont permis d'établir trois catégories de documents: les documents de référence, les documents de transaction et les messages de courrier électronique.

Les premiers sont «utilisés afin de prendre connaissance d'une information générale pertinente à la réalisation d'un processus de travail (études, volumes, publications, informations publicitaires, communiqués, audiovisuel, formulaires vierges, spécifications techniques, tables, catalogues et répertoires)» (Québec 1999b, 1. Rapport synthèse, 8).

Les seconds établissent «certaines informations relatives à un échange entre deux parties et liant par signature une responsabilité de portée administrative, juridique ou financière entre les parties. Contrairement aux documents de référence, les documents de transaction s'adressent à un seul destinataire ou à un nombre limité de personnes. Il peut s'agir autant de documents produits que de documents reçus par une organisation» (Québec 1999b, 1. Rapport synthèse, 8).

Afin d'établir les profils de métadonnées pour ces catégories de documents, les responsables du projet du Chantier ont utilisé un document conçu par le GRGD: les pages vertes². Ces pages sont une adaptation à l'archivistique du modèle GILS (*Government Information Locator Service*) servant à l'accès aux documents de référence, ainsi que du *Dublin Core* et du format MARC. De ces pages, un nombre important d'attributs ont été sélectionnés pour devenir les métadonnées servant à décrire le document et à son repérage ultérieur.

Exercice avec le deuxième document constitutif du projet pilote

Selon les définitions données ci-dessus, le dossier du projet pilote présenté précédemment est considéré comme un document de référence. Nous traiterons le 2^e document de ce dossier en utilisant le profil de métadonnées issu des pages vertes.

Ainsi, en prenant la liste des métadonnées proposée pour les documents de référence, nous aurons:

Identification du document

- 1.1 Titre (obligatoire): Gestion intégrée des documents papier/électroniques en circuit de production au Greffe du Conseil du trésor – Projet pilote en ingénierie documentaire
- 1.2 Auteur: Michel Lévesque, Lise Readman
- 1.3 Autres agents de production: Danielle Bélanger, Lucie Dion, Claire Fecteau
- 1.4 Organisation source (obligatoire): Québec (Gouv.). Conseil du trésor. Secrétariat du Conseil du trésor
- 1.5 Date de publication (obligatoire): 1999-12-06
- 1.6 Codification: 1612-01 ou 6212
- 1.7 Langue du document (obligatoire): française

Description du document

- 2.1 Résumé: (voir texte du début)
- 2.2 Raison d'être: fournir de l'information sur l'avancement du projet en question
- 2.3 Programme gouvernemental: Gestion des demandes d'autorisation et d'avis au Conseil du trésor
- 2.4 Information supplémentaire:
- 2.5 Mots-clés: (Cet élément n'a pas été jugé obligatoire. Nous y reviendrons plus tard)
- 2.6 Période couverte par le contenu: 1999-12-06
- 2.7 Référence géographique par coordonnées: (aucune)
- 2.8 Référence géographique par nom de lieu: (aucune)
- 2.9 Sources des données: Greffe du Conseil du trésor
- 2.10 Méthodologie: (ne s'applique pas)

Accessibilité

- 2.11 Distributeur: (ne s'applique pas)
- 2.12 Caractérisation: (ne s'applique pas)
- 2.13 Procédé de commande: (ne s'applique pas)
- 2.14 Support technique: (ne s'applique pas)
- 2.15 Lien accessible: (ne s'applique pas)
- 2.16 Type de lien accessible: (ne s'applique pas)
- 2.17 Durée de disponibilité: (ne s'applique pas)
- 2.18 Conditions d'utilisation: (ne s'applique pas)
- 2.19 Point de contact: Service de la gestion documentaire, Direction des ressources matérielles, Conseil du trésor, ...

Renvois

- 2.20 Titre du renvoi: (ne s'applique pas)
- 2.21 Lien du renvoi: (ne s'applique pas)
- 2.22 Type du renvoi: (ne s'applique pas)
- 2.23 Relation avec l'entrée: (ne s'applique pas)

Administration du document

- 3.1 Limite d'accès: ce document est public
- 2. Règle de conservation: 888* - 5 - C (* Durée du projet)
- 3.3 Document institutionnel: oui
- 3.4 Document essentiel: non
- 3.5 Sous-classe d'objet: document de référence
- 3.6 Création horodatée du document: (inscription automatique pour un document électronique)
- 3.7 Modification horodatée du document: (même commentaire)
- 3.8 Identifiant séquentiel de document: CTXXXX

- 3.9 Identifiant de structure logique: (ne s'applique pas)
- 3.10 Nom modificateur du document: (inscription automatique)
- 3.11 Contrôle d'accès: (inscription automatique)

Administration de l'entrée

- 4.1 Création horodatée de l'entrée: (inscription automatique)
- 4.2 Modification horodatée de l'entrée: (inscription automatique)
- 4.3 Nom du créateur de l'entrée: (inscription automatique)
- 4.4 Nom du modificateur de l'entrée: (inscription automatique)
- 4.5 Identifiant séquentiel d'entrée: (inscription automatique)
- 4.6 Identifiant séquentiel d'origine: (inscription automatique)
- 4.7 Date de révision de l'entrée: (selon les règles internes)
- 4.8 Langue de l'entrée: française

Exercice avec la lettre du GRGD (pages vertes)

Selon les définitions données ci-dessus, la lettre de la présidente du GRGD doit être considérée comme un document de transaction. Pour les fins de l'exercice, nous décrivons la version électronique en utilisant le profil de métadonnées issu des pages vertes.

Ainsi, en prenant la liste des métadonnées proposée pour les documents de transaction, nous aurons:

Identification du document

- 1.1 Nom administratif courant (obligatoire): lettre du GRGD au Conseil du trésor
- 1.2 Créateur/signature (obligatoire): Cécile LeBlanc
- 1.3 Organisation source (obligatoire): GRGD
- 1.4 Codification: 9100
- 1.5 Catégorie d'affaires: Reconnaissance et développement de l'organisme
- 1.6 Langue du document (obligatoire): française
- 1.7 Destinataire de transaction (obligatoire): Jacques Léonard

Description du document

- 2.1 Cadre réglementaire: (ne s'applique pas)
- 2.2 Mots-clés: (Cet élément n'a pas été jugé obligatoire. Nous y reviendrons plus tard)
- 2.3 Méthodologie: (ne s'applique pas)
- 2.4 Fichier du document (obligatoire): fichier en Word Perfect
- 2.5 Exécution du document: (liste des standards techniques)
- 2.6 Type de transaction (obligatoire): (ne s'applique pas)

Administration du document

- 3.1 Limite d'accès: aucune
- 3.2 Règle de conservation: 888* - 2 - C (*Jusqu'à la décision)
- 3.3 Document institutionnel: oui
- 3.4 Document essentiel: non
- 3.5 Sous-classe d'objet: document de transaction
- 3.6 Signature numérique: (saisie des algorithmes de signature)
- 3.7 Création horodatée du document: (inscription automatique)
- 3.8 Modification horodatée de l'entrée: (inscription automatique)
- 3.9 Identifiant séquentiel d'entrée: (inscription automatique)
- 3.10 Identifiant de structure logique: (inscription du schéma XML et DTD)
- 3.11 Nom du modificateur du document: (inscription automatique)
- 3.12 Réception horodatée du document: (inscription automatique)
- 3.13 Lien(s) avec transaction(s) antérieure(s): aucun
- 3.14 Action subséquente: lettre à la ministre de la Culture et des Communications
- 3.15 Références de vérification: (ne s'applique pas)

Administration de l'entrée

- 4.1 Création horodatée de l'entrée: (inscription automatique)
- 4.2 Nom du créateur de l'entrée: (inscription automatique)
- 4.3 Identifiant séquentiel d'entrée: (inscription automatique)

Comme nous pouvons le constater, cet exercice est particulièrement complexe. C'est d'ailleurs pourquoi plusieurs réserves ont été exprimées sur l'attribution d'un tel nombre de métadonnées (49) pour, il faut bien le comprendre, chacun des documents créés. Il est vrai par ailleurs que plusieurs métadonnées doivent faire l'objet d'une inscription automatique qui est soit déjà effective dans l'utilisation de systèmes informatiques, soit prise en charge par les directions gérant ces systèmes.

Or, le constat que nous pouvons faire à la lecture de ces métadonnées est que nous sommes encore centrés sur une approche contextuelle plutôt que sur une approche de contenu documentaire. Il est vrai que nous travaillons ici avec des documents actifs. Nous pouvons donc supposer que la gestion administrative des documents doit être plus formelle. Il est par contre particulier de ne pas rendre obligatoire une zone de mots-clés, à tout le moins un seul, pour caractériser le document. Pour plusieurs, la recherche plein texte est suffisante et ne demande aucun investissement de la part des créateurs de documents.

Comme d'autres projets existent quant à l'attribution de métadonnées aux documents mais dans une optique de conservation à long terme, nous nous sommes intéressés à tenter le même exercice que précédemment, en nous servant des travaux du projet *Metadata for Digital Preservation: the CEDARS Projet Outline Specification*. Ce projet a l'avantage d'être tout récent et est en relation avec un autre projet tout aussi intéressant qui est le *Reference Model for an Open Archival Information System (OAIS)* développé par le *Consultative Committee for Space Data System*.

Exercice avec la lettre du GRGD (Metadata for Digital Preservation)

Information Package

1.1 Preservation Description Information

1.1.1 Reference Information

1.1.1.1 Ressource Description: GRGD - Version 1

1.1.1.2 Existing Metadata: Lignes directrices du chantier

1.1.1.2.1 Existing Records: oui

1.1.2 Context Information: Cette lettre fait partie d'un dossier concernant la reconnaissance officielle du GRGD

1.1.2.1 Related Information Objects: mémoire sur la reconnaissance officielle du GRGD

1.1.3 Provenance Information

1.1.3.1 History of Origin

1.1.3.1.1 Reason for Creation: cette lettre a été créée pour demander la reconnaissance officielle du GRGD

1.1.3.1.2 Custody History: le GRGD est responsable de la création et de la conservation

1.1.3.1.3 Change History Before Archiving: aucun

1.1.3.1.4 Original Technical Environments

1.1.3.1.4.1 Prerequisites: PC, Word Perfect.

1.1.3.1.4.1 Procedures:

1.1.3.1.4.1 Documentation: manuels d'utilisation des logiciels

1.1.3.1.4.1 Reason for Preservation: selon la politique relative à la conservation

1.1.3.2 Management History

1.1.3.2.1 Ingest Process History: transfert de la lettre sur CD-Rom

1.1.3.2.2 Administration History

1.1.3.2.2.1 Action History: le format a été changé: transformé en PDF

1.1.3.2.2.2 Policy History: un changement de format doit se faire lorsque nécessaire

1.1.3.3 Rights Management

1.1.3.3.1 Negotiation History: aucune

1.1.3.3.2 Rights Information: les droits de propriété appartiennent au GRGD

1.1.3.3.2.1 Copyright Statement: ne s'applique pas

(Cette section se subdivise en 5 sous-sections)

1.1.3.3.2.2 Actors: ne s'applique pas

1.1.3.3.2.3 Actions: ne s'applique pas

(Cette section se subdivise en 2 sous-sections)

- 1.1.4 Fixity Information
 - 1.1.4.1 Authentication Indicator: certificat de certification des signatures
- 1.2 Content Information
 - 1.2.1 Representation Information
 - 1.2.1.1 Structure Information
 - 1.2.1.1.1 Underlying Abstract Form Description: résumé de la lettre
 - 1.2.1.1.2 Transformer Object
 - 1.2.1.1.2.1 Platform:
 - 1.2.1.1.2.2 Parameters:
 - 1.2.1.1.2.3 Render/Analyse Engines:
 - 1.2.1.1.2.4 Output Format:
 - 1.2.1.1.2.5 Input Format:
 - 1.2.1.1.3 Render/Analyse/Convert Objects
 - 1.2.1.1.3.1 Platform:
 - 1.2.1.1.3.2 Parameters:
 - 1.2.1.1.3.3 Render/Analyse Engines:
 - 1.2.1.1.3.4 Output Format:
 - 1.2.1.1.3.5 Input Format:
 - 1.2.1.2 Semantic Information
 - 1.2.1.2.1 Platform:
 - 1.2.1.2.2 Parameters:
 - 1.2.1.2.3 Render/Analyse Engines:
 - 1.2.1.2.4 Output Format:
 - 1.2.1.2.5 Input Format:
 - 1.2.2 Data Object

Cet exercice nous a permis de constater encore une fois l'importance accordée à l'information contextuelle. En fait, la gestion des documents électroniques à court, moyen et long terme exige un réseau formel de métadonnées qui va permettre de conserver une description des paramètres de la structure logique et physique des documents. Il ne servirait à rien de connaître parfaitement bien le contenu intellectuel du document, si nous ne pouvions plus y accéder.

L'étiquette de métadonnées qui sera rattachée aux documents ne servira pas ou très peu au repérage de l'information contenue dans ceux-ci. Voilà qui est à tout le moins paradoxal dans la mesure où des descripteurs peuvent être assimilés aux métadonnées et justement faire partie de cette étiquette. Nous croyons qu'il faut rendre obligatoire l'indication de descripteurs comme métadonnées.

XML et DTD

Dans le monde électronique, où l'intérêt est bien entendu la relecture des documents créés, la notion de format de langage est primordiale; d'après nous cette notion l'est même plus que celle du support. Voilà quelques temps, le SGML apparaissait comme langage informatique formel permettant la description de documents. Probablement à cause de sa grande complexité, il ne s'est pas véritablement imposé. Un langage simplifié vient de faire son apparition et semble porteur d'avenir. Il s'agit du XML. Avec le HTML, qui est un langage d'affichage (d'étiquette de présentation), le XML permet de décrire le contenu des données au moyen des étiquettes électroniques, par exemple: une étiquette titre, une étiquette paragraphe, une étiquette note, etc.

Un utilisateur peut créer le nombre d'étiquettes nécessaires pour décrire la structure logique des documents, par exemple un formulaire. Il créera donc autant de DTD que nécessaire, c'est-à-dire des définitions de types de document qui seront transformées en langage XML, donc selon un balisage de règles de codification.

Le projet d'une norme de description EAD (*Encoded Archival Description*) est ni plus ni moins, comme le soulignait le Comité des normes du Conseil canadien des archives, que l'élaboration d'une DTD qui serait une série d'étiquettes pour les instruments de recherche.

Or, selon les discussions que nous avons eues avec un spécialiste en ce domaine, M. Yves Marcoux, de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal, il semble bien que les DTD sont insuffisantes pour servir à un véritable repérage d'informations. Elles balisent le texte. Il serait donc fondamental qu'une zone de description soit prévue et étiquetée pour permettre cette possibilité. Il faudrait aussi faire en sorte que les moteurs de recherche puissent reconnaître cette zone.

Règles de description des documents d'archives (RDDA)

À l'instar de plans de classification structurés pour les «archives historiques» en l'absence de plans d'origine, des règles pour la description des documents d'archives ont aussi été développées. Dans un but de normalisation, ces règles permettent de décrire un fonds d'archives et ses parties selon différentes zones.

Nous ferons encore une fois un exercice de description. Pour ce faire, nous utiliserons l'affiche présentée précédemment. Comme les règles exigent une description partant du général pour aller vers le particulier, donc du fonds à la pièce, nous nous contenterons de décrire la pièce en faisant comme si nous avions décrit les niveaux supérieurs.

Exercice au niveau de la pièce avec l'affiche

Fonds: Société d'habitation du Québec

Série: Programme d'aide aux personnes

Sous-série: Programme d'allocation logement unifiée

Dossier: Programme d'allocation logement unifiée-Production visuelle

Pièce: Affiche

Description de la pièce:

Programme allocation logement. – janvier 2000. – 1 affiche: coul.; 15,5 X 25 cm

Cette affiche annonce le programme d'allocation logement qui a pour objectif d'aider les personnes de 55 ans ou plus et les familles à faible revenu à payer leur loyer. Des photos de personnes âgées (homme et femme) et d'une famille (homme, femme et enfant) y apparaissent.

Le ministère du Revenu est co-responsable de la production de cette affiche.

Tentons finalement le même exercice avec la lettre du GRGD en appliquant les RDDA au fonds et à la série seulement.

Exercice au niveau du fonds et de la série avec la lettre du GRGD

Fonds: Fonds GRGD

Série: Reconnaissance et développement du GRGD

Description de la série:

Reconnaissance et développement du GRGD. – 1998-2000. – 3 cm de documents textuels

La série comprend des études, des mémoires et de la correspondance sur la reconnaissance officielle et le développement du GRGD.

Titre provenant du plan de classification.

Pour la zone Portée et contenu, nous avons fait sciemment une description générale dans le respect de l'esprit des RDDA. Nous constatons donc qu'il y a bel et bien une perte d'information. L'indexation qui en résultera sera appauvrie. Des concepts, tels que reconnaissance de groupes d'intérêts, ministre de la Culture et des Communications, Conseil du trésor, regroupements reconnus, etc., ne seront pas indexés.

Bien qu'informatives en soi, plusieurs zones nous apparaissent encore décrire avant tout un contexte plutôt que la valeur de l'information. La zone Titre (et Mentions de responsabilité) ainsi que la zone Portée et contenu deviennent celles qui constituent un résumé du contenu des documents à décrire.

Tous comprendront fort bien que la description la plus précise se situe au niveau de la pièce, surtout lorsqu'il s'agit de compléter les zones Titre et Portée et contenu. Nous sommes aussi fort conscients qu'en règle générale, il est rare que la description se fasse à la pièce, mais l'exercice en valait la peine.

Paradoxalement, on semble mettre beaucoup plus l'accent sur l'histoire administrative et la notice biographique que sur les titres et les portées et contenus. Assistet-on à un réflexe d'historien? Comme le plan de classification est fortement empreint de témoignage, la description est à l'avenant. Ce jugement est peut-être sévère, mais il est nécessaire lorsque le but de la conservation des documents, à quelque stade que ce soit, est la diffusion. Peut-on vraiment croire que seules les quelques rubriques d'un

plan de classification ou les quelques termes composant les titres suffiront à exploiter la richesse des archives?

D'ailleurs, il est aussi assez particulier de voir l'insistance mise sur le contrôle d'autorité des noms de personnes physiques et morales (moins sur les lieux qui pourtant sont aussi très complexes) qui révèlent la provenance mais sont aussi souvent des sujets. Le créateur d'une lettre parle d'un sujet mais indirectement parle aussi de lui-même.

Nous avons déjà exprimé lors de séminaires que les archivistes auront intérêt à mieux exploiter la zone Portée et contenu. C'est là que réside la véritable source pour l'utilisation des archives, qui dépassera le stade de l'emploi par des historiens chevronnés qui, semble-t-il, trouvent toujours ce qu'ils cherchent, et qui rayonnera vers d'autres publics intéressés à la connaissance.

En ce sens, il faudra donner plus de souplesse aux principes qui régissent les RDDA. Dans la mesure où il s'avérera difficile de faire des descriptions au niveau du dossier ou de la pièce, il faudra permettre que, dans la zone Portée et contenu, l'information y soit la plus complète possible. D'après nous, décrire un fonds ou une série de façon générale parce que le niveau est supérieur n'offre pas suffisamment de clés d'accès à la recherche. Le repérage en sera considérablement affecté. En rendant la zone Portée et contenu la plus significative possible, cela permet de tendre vers un équilibre et de satisfaire deux types d'utilisateurs: celui qui cherche un document précis et celui qui cherche tout.

DE QUELQUES CONSIDÉRATIONS SUR L'INDEXATION, LE CONTRÔLE DU VOCABULAIRE, LES THÉSAURUS, LE REPÉRAGE PLEIN TEXTE ET LES STRATÉGIES DE RECHERCHE.

À la lumière des constats établis dans la partie précédente, nous essaierons dans cette deuxième partie de démontrer l'importance de l'indexation tout en posant un regard critique sur les moyens pour la réaliser.

Analyse de contenu et indexation

À l'instar de plusieurs spécialistes, Dominique Maurel et Michel Champagne ont défini l'indexation comme «l'ensemble des opérations permettant la création de clés d'accès regroupées dans les différents types d'index, informatisés ou non. Il s'agit, d'une part, de la sélection des termes et des expressions d'indexation dans les documents originaux ou dans des résumés et, d'autre part, de leur transcription en langage documentaire» (Maurel et Champagne 1999, 318).

Pour enrichir leur propos, ils citent Jacques Chaumier qui précise que «du point de vue de son processus, l'indexation est l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans le document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse. La transcription en langage documentaire se fait grâce à des outils d'indexation tels que thésaurus, classification, etc.» (Maurel et Champagne 1999, 318).

L'AFNOR, citée par Michèle Hudon lors d'un séminaire sur l'indexation, définit l'indexation comme «une opération [intellectuelle] qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse» (Hudon et al. 1998, 1).

Il est tout à fait intéressant de constater la concordance entre ces définitions où la source principale permettant l'indexation est le document lui-même. Maurel et Champagne ajoutent une nuance en parlant aussi de résumés qui sont habituellement associés à l'étape suivant l'analyse de contenu qu'en bibliothéconomie on appelle la condensation.

En archivistique, cette orientation risque d'en faire réagir plus d'un. Est-ce qu'on doit pour chaque document à la fois faire une analyse de contenu et exprimer les résultats de celle-ci dans un résumé pour pouvoir indexer et ainsi favoriser le repérage? Imaginez la situation où vous vous retrouvez dans un organisme qui a accumulé au fil des ans des milliers de dossiers qui ont été titrés et classés sans être accompagnés d'index faits dans les règles de l'art. La situation est tout aussi comparable lorsqu'on se retrouve dans un centre d'archives en présence de versements de documents et de dossiers d'organismes, d'individus ou de familles. Il sera facile de trouver bon nombre d'objecteurs qui trouveront irréaliste de référer à chacun des documents, voire même à chacun des dossiers pour parvenir à en faire l'indexation. Il faut convenir qu'ils n'ont pas vraiment tort. Paradoxalement, par contre, ne sommes-nous pas une des sources de cette problématique? Quelles pourraient donc être les solutions que nous devons proposer aux créateurs d'information et aux chercheurs pour répondre à leurs besoins multiples de repérage?

Dans la mesure où, dès le stade actif, nous sommes en présence de documents, et par extension de dossiers, il est possible de procéder à l'identification/reconnaissance des concepts présents dans les documents par une rigoureuse analyse de contenu basée sur les critères reconnus suivants: l'exhaustivité, la spécificité, la cohérence et l'exactitude³. C'est à cet exercice que nous nous sommes prêtés au début avec nos quatre exemples. Y a-t-il un moyen simple et souple pour aider les créateurs ou les personnes qui prendront en charge la conservation des documents et des dossiers à réaliser cette analyse qui, il faut le comprendre, s'ajoute comme un travail supplémentaire? Faut-il obligatoirement réaliser des résumés des documents et dossiers, tels l'annotation, le résumé indicatif ou le résumé informatif, pour procéder à la constitution d'index?

Une révision de la philosophie du plan de classification est probablement une piste de solution. Il faudrait en arriver à faire coïncider des descriptions générales de contexte et des descriptions spécifiques de contenu des documents, et par extension des dossiers. Sans pour autant inclure les dossiers comme tels dans le plan de classification, nous devrions «éplucher» le contenu de ce qui se retrouve dans les classeurs et dans les répertoires informatiques et en faire des descriptions significatives qui seraient incluses aux rubriques correspondantes du plan de classification.

En complément à ces descriptions, il suffira que les concepts ou sujets principaux soient systématiquement et obligatoirement transférés dans la zone actuelle des

mots-clés ou de descripteurs lors de l'inscription des dossiers et des documents, zone comprise dans les systèmes informatiques de gestion documentaire (par exemple dans COBA) ou dans les systèmes de GED (gestion électronique des documents) avec métadonnées. Ce transfert impliquera une indexation en vocabulaire contrôlé (langage contrôlé). Nous y reviendrons.

Cette approche saura convenir à la fois aux documents sur support traditionnel et à ceux sur support électronique. Il s'agit en fait de regrouper dans une banque de données toutes les métadonnées descriptives d'un document, en y incluant particulièrement les descripteurs relatifs au contenu, ou bien, lorsque nous avons des DTD, d'utiliser une zone préalablement identifiée pour les descripteurs et accessibles aux fureteurs. Nous obtiendrons ainsi plusieurs niveaux d'indexation qui conduiront vers une indexation pluridimensionnelle.

Cette approche permet aussi de donner une véritable assise informationnelle à la description qui sera faite pour les documents qui constitueront les archives définitives. Les descriptions spécifiques établies au plan de classification pourraient servir à documenter la zone Portée et contenu. Une nouvelle zone pourrait récupérer les descripteurs, créés au stade actif, associés aux documents versés.

En arriver à une telle pratique dans les ministères et les organismes gouvernementaux semble réaliste si celle-ci est souple et bien documentée à l'aide de guides. Selon ces balises, nous croyons même que les entreprises de moyenne et de grande taille adhèreraient à cette pratique.

Il en va autrement lorsque nous songeons aux petites entreprises, mais plus particulièrement aux individus et aux familles. Comment leur demander de se constituer une banque de données documentaire composée de descripteurs représentant les concepts présents dans leurs documents et obtenus à la suite d'une analyse de contenu? Il sera toujours possible de partir en croisade et démontrer, à ceux où le sens de l'organisation est transparent, l'avantage d'une telle pratique.

Qui sait, peut-être qu'une telle action suscitera une prise de conscience de l'importance de traiter ses propres documents. Mais, de toute évidence, il semble bien que des spécialistes devront prendre en charge la constitution de banques relatives à des fonds versés aux institutions d'archives.

Et qu'en est-il des fonds existants?

Pour éviter de retourner au contenu spécifique des documents, aussi bien dans les archives courantes que dans les archives définitives non indexées, et dans la mesure où elles existent, nous pourrions nous référer aux listes de dossiers et aux listes des fichiers (et par extension aux répertoires et sous-répertoires) pour identifier un minimum de concepts et procéder à leur indexation. Le principal problème sera l'exactitude et la spécificité des titres des dossiers et des fichiers. Il arrive souvent que ceux-ci ne soient pas précis et significatifs.

Dans le cas des archives courantes et définitives décrites et indexées, nous pourrions nous référer aux descriptions et aux index existants afin de connaître et d'évaluer les procédures de description et d'indexation utilisées. Déjà signalés dans la première partie, les principaux problèmes seront ceux reliés aux niveaux de description, aux titres et aux descriptions apparaissant dans les portées et contenus. Cela

pourra impliquer un travail de révision assez important car il s'agira de retourner aux listes des dossiers, aux fichiers et aux index d'origine.

L'archiviste doit finalement comprendre qu'un document ne peut se résumer à n'être qu'un reflet des fonctions et des activités d'un créateur. Il est plus que cela: il est une source d'informations en soi qui se colore par le contexte dans lequel il est créé et par les liens qu'il a avec les autres documents. Une étude sur le marché de l'art au Québec, la norme sur l'utilisation de l'Internet au gouvernement, les recettes manuscrites de sœur Angèle, le dossier personnel de M., les états financiers de Bombardier, les procès-verbaux du Cercle des fermières de telle municipalité, etc., illustrent très bien que des documents peuvent servir à d'autres fins que celles pour lesquelles ils ont été créés. Il suffit d'en exploiter le contenu pour enrichir le domaine de la connaissance. La «polarité» du document sera ainsi véritablement reconnue et permettra de satisfaire à la fois les besoins des individus associés aux processus administratifs et ceux des chercheurs. Trop souvent encore l'accessibilité se voit réduite au seul accès par provenance. Pour la secrétaire dans un organisme, ce n'est certes pas de cet accès qu'elle a besoin. De même, à l'examen des questions des chercheurs aux Archives nationales du Québec (compilation fait par M. Rénald Lessard du service de référence), nous constatons qu'une majorité de demandes se fait par sujets.

Il y a bien sûr des dangers relatifs à l'analyse de contenu qui se répercuteront sur l'indexation. M^{me} Gracia Pagola, lors d'un séminaire en indexation, précisait que l'analyse de contenu, bien qu'elle doive être prévisionnelle, c'est-à-dire répondre aux interrogations futures des utilisateurs, reste «toujours valable [lorsqu'elle est] faite pour un certain type d'utilisateurs à un moment donné, en réponse à certaines préoccupations»? Elle ajoutait que «l'analyse implique toujours un jugement sélectif puisqu'il s'agit de donner une représentation condensée» (Pagola 1996, session du dimanche 17 novembre 1996, [12]).

Ainsi, même avec des critères d'exhaustivité, de spécificité, de cohérence et d'exactitude, l'indexation demeure un travail complexe qui implique le jugement. Nous ne croyons pas qu'il soit fondamental de connaître l'étendue des besoins des personnes voulant accéder aux documents de l'organisme ou à ceux des chercheurs. Il faut par contre y faire ressortir beaucoup plus les caractéristiques de ces documents: caractéristiques physiques, caractéristiques de contexte et caractéristiques de contenu. De cette façon, nous sommes convaincus que quiconque trouvera ce qu'il cherche.

Contrôle du vocabulaire

Lorsqu'on procède à l'indexation, on utilise un langage documentaire. «Le langage d'indexation peut être la langue naturelle (indexation en vocabulaire libre) [ou il] peut être contrôlé (indexation en vocabulaire contrôlé)» (Hudon et al. 1998, 2).

Dans le premier cas, nous pourrions procéder par extraction automatique ou par assignation des mots du document qui constitueront l'index. L'extraction est bien entendue maintenant très facile à réaliser à l'aide de procédés informatiques.

Dans le second cas, il s'agit d'une indexation par assignation. Elle est encore maintenant réalisée par l'humain, même si des expériences sont tentées pour faire faire ce travail par des automates intelligents. Il s'agit de traduire les concepts retenus

à l'aide de termes d'un langage documentaire. Ce langage est artificiel et s'oppose ainsi au langage naturel.

Depuis longtemps, nous soutenons que l'indexation devrait se faire en vocabulaire contrôlé. Nous maintenons notre opinion. Sans parler de la recherche plein texte sur laquelle nous reviendrons plus loin, nous sommes d'avis que la constitution d'un index fait à partir d'une extraction ou l'assignation de mots en vocabulaire libre présente trop de problèmes pour devenir la voie principale du repérage de l'information d'un fonds documentaire. Ces problèmes sont d'ordre morphologiques, lexicaux et sémantiques.

À titre d'exemple, nous pouvons, en prenant le diplôme précédemment décrit, extraire les mots suivants: études, programme de 2^e cycle, grade, bibliothéconomie, sciences de l'information, M.B.S.I., recteur, secrétaire général, doyen, maître, Université de Montréal, etc. Ces termes se retrouveraient dans un index et seraient ainsi repérables. Sans trop nous forcer, nous pouvons relever les problèmes morphologiques et sémantiques suivants: étude/études/étudier, grade/gradé, maître/maîtrise, programme de 2^e cycle/programme de maîtrise/programme d'études supérieures, UdeM/Université de Montréal, Maître (au sens de chef) /maître (au sens de diplômé universitaire), doyen (au sens de personne la plus âgée)/doyen (au sens d'administrateur d'une faculté universitaire).

En règle générale, en établissant et en utilisant un vocabulaire contrôlé, nous exerçons un contrôle flexionnel (études/étude), orthographique (UdeM /Université de Montréal), morphologique (grade /grader/gradation), lexico-sémantique au niveau de la synonymie (2^e cycle/maîtrise) et au niveau de la polysémie (maître (chef)/maître(diplômé universitaire)).

Hudon et Pagola (Hudon et al. 1998, 6; Pagola 1996, session du mardi 19 novembre 1996, [6]) ont d'ailleurs fait ressortir les avantages et les inconvénients d'utiliser les deux types d'indexation:

Indexation en vocabulaire libre	Indexation en vocabulaire contrôlé
Peu prévisible	Très prévisible
Plusieurs termes pour exprimer un même concept	Un seul terme pour exprimer un même concept
Stratégie de recherche complexe et coûteuse pour l'utilisateur	Stratégie de recherche compacte et efficace
Bruit au repérage	Peu de bruit au repérage
Très haute spécificité	Plus général
Plus dynamique	Moins dynamique
Investissement minimal (formation des indexeurs, préparation d'outil, etc.)	Investissement important (formation des indexeurs, préparation d'outils, contrôle de qualité, etc.)
Grande flexibilité au niveau de la traduction des concepts	Limites imposées au niveau de la traduction des concepts

À l'examen de ces différences, nous pourrions être séduits par la souplesse que présente l'indexation en vocabulaire libre. Mais encore faut-il retenir l'objectif principal qui est le repérage de l'information. Même si effectivement il est très simple d'extraire des mots, il n'est pas complètement vrai que nous obtenons une si grande spécificité. Celui qui utilise un terme précis pour sa recherche trouvera effectivement tous les documents dans lesquels apparaît son terme. Mais rien ne lui indiquera si d'autres documents contiennent des termes synonymes, tout comme il aura sûrement repéré quelques documents contenant son terme, mais pris dans un autre sens.

Il faut quand même garder à l'esprit l'investissement que représente la construction d'un vocabulaire contrôlé et son utilisation par des personnes qui peuvent être des non-spécialistes.

Un autre élément préoccupant est la question des noms de personnes physiques ou morales et de famille, ainsi que les noms de lieux. Selon les RDDA, ces noms constituent des points d'accès aux fonds d'archives en tant que noms de créateurs ou d'auteurs. Des règles conditionnent le choix et la présentation de ces points d'accès. Ainsi, les institutions d'archives doivent se doter de listes d'autorité ou de registres d'autorité pour contrôler ces points d'accès.

Toujours dans le but de contextualiser les fonds d'archives dans le respect du principe de provenance, il n'est pas surprenant que, comme pour les RCAA2 (*Règles de catalogage anglo-américaines, 2^e édition*), les vedettes d'accès aient été retenues. Est-il par contre utile de distinguer ces accès à des fins de repérage? Nous ne le croyons pas. Cette question était probablement très importante dans la constitution d'index manuels mais elle n'a plus vraiment la même teneur lorsqu'une banque de données documentaire est constituée. Par contre, il faut toujours exercer un contrôle sur les formes des noms de personnes physiques ou morales, des noms des familles et des noms de lieux. Mais comme ceux-ci peuvent être à la fois sujet et créateur (ou auteur), ce contrôle doit dépendre des mêmes règles.

Devrions-nous envisager de faire ce contrôle dès le stade actif? Dans la mesure où un nom peut se retrouver dans un index, il serait préférable de normaliser sa présentation. Il faudrait donc utiliser les règles proposées ou les adapter.

Thésaurus et autres langages documentaires

Toujours selon les spécialistes en bibliothéconomie, nous retrouvons quatre langages documentaires: les schémas de classification, les listes ou répertoires de vedettes-matière, les thésaurus de descripteurs et les listes de mots-clés. Les deux premiers sont des langages précoordonnés (termes associés à l'analyse: gestion des ressources humaines (schémas), Femmes - Histoire des - 1960-1980 (vedettes-matières)) et les deux derniers sont des langages postcoordonnés (descripteurs qui sont reliés lors de la recherche: «documents électroniques» et «plan» et «classification»).

À l'examen des plans de classification existant, il ne nous semble pas qu'ils forment une base suffisamment solide pour constituer le langage d'indexation souhaité. Comme nous l'avons précisé, les plans de classification ne sont pas semblables aux schémas de classification que nous retrouvons pour la classification des monographies et des périodiques.

Les listes ou répertoires de vedettes-matière réunissent des «terme[s] d'indexation composé[s] exprimant plusieurs concepts représentant le sujet d'un document» (Hudon et al. 1998, 5). Un réseau de liens hiérarchiques, associatifs et d'équivalence les font ressembler aux thésaurus. Plusieurs spécialistes ont émis de sévères critiques par rapport à ce langage. Ainsi, les connaissances y sont représentées par des vedettes-matières constitutives d'une «structure hiérarchique lâche» (Hudon) et aucune «règle syntaxique à suivre dans le cas des termes complexes exprimant un sujet pouvant être formulé de plusieurs façons différentes» (BCA 1992, 103).

Le thésaurus est un «langage documentaire fondé sur une structuration hiérarchisée d'un ou de plusieurs domaines de la connaissance et dans lequel les notions sont représentées par des termes de la langue naturelle et les relations entre notions par des signes conventionnels» (Hudon et al. 1998, 7). Il est constitué d'un vocabulaire d'accès, c'est-à-dire de descripteurs et de non-descripteurs, de définitions et de notes d'application et d'une structure classificatoire ou sémantique faite de relations: relation d'équivalence, relation hiérarchique et relation associative. Pour plusieurs spécialistes, le thésaurus est l'outil le plus rigoureux pour permettre l'indexation de concepts et leur repérage.

Le dernier langage documentaire est la liste de mots-clés. Il s'agit tout simplement d'une liste de termes, sans structure hiérarchisée. Les termes retenus ne sont pas précoordonnés, ce qui favorise la précision. Mais l'absence d'un réseau adéquat de relations ne facilite pas la recherche.

Quel langage faut-il choisir?

Ce choix est tributaire de ce que nous voulons ou plutôt de ce que nous pouvons nous offrir. De toute évidence, le thésaurus est l'outil le plus «scientifique». L'investissement en ressources humaines et financières pour en constituer un est par ailleurs très important, tout comme son entretien. Il peut être intéressant de considérer ceux existants, ou bien, dans la mesure du «réalisable», un amalgame de plusieurs. Nous pensons, entre autres, au thésaurus BIPA spécifiquement conçu pour l'information politique, économique et sociale. Tout comme il pourrait être stratégique de réunir, en fonction de secteurs particuliers ou groupes d'intérêt, des équipes de travail disposant de moyens financiers pour produire un thésaurus dans un domaine particulier et le maintenir à jour.

À cause des problèmes signalés, il ne serait pas approprié de créer son propre répertoire de vedettes-matières. Leur intérêt vient du fait que les vedettes-matières traduisent un sujet et que, pour le commun des mortels, un document, quel qu'il soit, porte sur un ou plusieurs sujets. Quand le chercheur veut retrouver un document, il précisera un sujet. Or, un sujet simple peut coïncider avec un concept, souvent par contre un sujet complexe ne coïncidera pas avec un seul concept. Prenons comme exemple le sujet: *Reconnaissance du GRGD*, nous avons deux concepts: reconnaissance et GRGD. Le problème est ici de savoir comment traduire ce sujet: *GRGD-Reconnaissance*, *Reconnaissance-GRGD* ou *Reconnaissance du GRGD*. Des répertoires de vedettes-matières existent déjà. Celui de la *Library of Congress* et celui de la Bibliothèque de l'Université Laval sont très connus. Certains pourront les employer mais il serait conseillé de faire un exercice à partir d'un corpus de documents avant de les

utiliser de façon systématique. Les Archives nationales du Québec s'en inspirent pour construire leur vocabulaire contrôlé.

La liste de mots-clés peut sembler un compromis intéressant. Dans la mesure où le contrôle flexionnel, orthographique, morphologique et lexico-sémantique au niveau de la synonymie et de la polysémie est pris en charge, la liste de mots-clés devient une véritable liste de descripteurs en vocabulaire contrôlé. Elle doit comprendre obligatoirement aussi des non-descripteurs. Afin de réaliser ce contrôle, d'autres langages peuvent être mis à contribution. Le problème majeur de cette liste consiste cependant en ce qu'il faut exercer un contrôle serré du nombre de termes constitutifs du langage. D'un autre côté, bien que la philosophie des langages documentaires l'interdise, une telle liste permet d'intégrer les noms de personnes physiques et morales, de familles et de lieux. Elle constitue donc une banque de données qui pourra à la fois servir aux utilisateurs/indexeurs et aux utilisateurs/chercheurs.

Que nous utilisons un thésaurus ou une liste de mots-clés, il serait nécessaire et utile de pouvoir automatiser le processus d'indexation lui-même. Ainsi, il faudrait concevoir une application informatique qui contienne les descripteurs de l'un ou l'autre langage, constituant ainsi un dictionnaire obligatoire pour l'enregistrement des documents. Un utilisateur procéderait à l'entrée de ses termes qui seraient analysés de façon électronique pour acceptation ou refus, dans le même esprit qu'un correcteur d'orthographe dans un traitement de texte. L'archiviste aurait la tâche formelle de développer et de contrôler le langage documentaire et d'être la personne-ressource pour la formation et l'aide ponctuelle auprès des utilisateurs.

À la limite coercitive, cette approche sera peut être difficile à faire accepter à des utilisateurs qui sont «dans le feu de l'action» et qui craignent que nous leur compliquions la vie. C'est seulement si l'application informatique se révèle peu contraignante que nous pourrions avoir du succès. La politique d'indexation et ses procédures devront aller de pair.

Repérage plein texte et moteur de recherche

Posséder un logiciel qui permette le repérage dans le texte d'un document, d'un résumé, voire à la limite d'un titre est pour plusieurs la solution au problème de repérage. Mais est-ce vraiment le cas?

Ces logiciels recherchent à partir de chaînes de caractères correspondantes aux termes recherchés. Ils ne font aucune distinction et retrouvent partout la chaîne, y compris dans d'autres termes qui peuvent la contenir. À titre d'exemple, en cherchant le terme PUBLIC, il retrouvera PUBLICS, PUBLICATION, PUBLICITÉ, PUBLIC (au sens de groupe de personnes). Il y a donc beaucoup de bruit au repérage et même la spécificité que nous retrouvions avec la recherche dans un index de termes en vocabulaire libre s'en trouve altérée. L'absence de contrôle peut donc conduire à des résultats parfois très douteux. De plus, dans un corpus de centaines de documents, le temps de réponse à une requête est long.

Avec l'arrivée d'Internet, un nouvel outil de repérage a fait son apparition: le moteur de recherche, tels Yahoo, Altavista, Copernic, etc. Plusieurs croient ainsi avoir résolu la problématique du repérage. Voir les pourcentages élevés de succès des recherches effectuées à l'aide de ces moteurs nous permet de comprendre cet engouement.

Nul n'est surpris lorsque dans une recherche simple sur Internet des milliers de documents sont référencés. Peu importe le temps considérable que prendra la consultation des références, nous avons alors le sentiment de tout avoir. C'est rassurant. Mais combien abandonnent à la quinzième référence? Le pourcentage élevé de succès l'est seulement à cause du nombre de références. Il n'indique en rien le nombre de documents non pertinents, ni la qualité des documents récupérés par l'utilisateur.

Le principal problème provient du fait que les moteurs de recherche sont lancés dans des sites indexés de différentes façons, avec une priorisation possible du vocabulaire libre qui est simple et peu contraignant. Même s'il existe des normes internationales concernant l'indexation, la normalisation semble difficile.

Ce constat vient démontrer que de «nombreuses expériences ont conclu [...] que le système d'information idéal était un système hybride, permettant la recherche plein texte et en langage naturel ainsi que la recherche contrôlée au moyen d'un langage documentaire [...]. Les études dans le domaine ont révélé que la recherche au moyen de termes du langage naturel permet régulièrement de repérer certains documents pertinents non repérés par la recherche contrôlée, et vice-versa» (Hudon 1997-1998, 81).

Nous souscrivons à ce point de vue uniquement si le recours au repérage plein texte ou en langage naturel est vu comme un filet de sécurité nécessaire pour parfaire une recherche. Il doit intervenir à la suite d'une recherche faite dans une banque de données de termes contrôlés. Les archivistes doivent conscientiser les utilisateurs aux bénéfices d'utiliser un langage documentaire pour leur repérage.

Stratégies de recherche

Autant l'accent a été mis sur les études d'utilisateurs et leurs besoins, autant les stratégies de recherche ne font l'objet que de courtes analyses, si analyse il y a. Pourtant, ne s'agit-il pas d'une fonction fondamentale intrinsèquement liée au repérage? Les «on veut tout», «on veut cela», «on veut cela, sans cela», «on ne sait pas exactement ce qu'on veut mais c'est fait par», «on veut la correspondance» signifient-ils vraiment que l'utilisateur/chercheur sait comment il ira chercher ce dont il a besoin?

L'archiviste a donc un double rôle à jouer: expliquer le système d'indexation implanté et aider l'utilisateur/chercheur à préparer sa stratégie de recherche, stratégie qui peut même à certaines occasions l'aider à mieux circonscrire ce qu'il veut et lui éviter ainsi bien des détours et frustrations. L'archiviste n'a pas à se substituer à l'utilisateur et procéder lui-même à la recherche. Il doit lui donner toutes instructions pour lui permettre de se débrouiller seul.

Il y a, d'après nous, un lien étroit entre la stratégie de recherche et le langage documentaire qui doit être utilisé. Les répertoires des vedettes-matières et les thésaurus ont comme particularité un réseau de relations: relations d'équivalence, relations hiérarchiques et relations associatives. En vue du contrôle du vocabulaire, les premières sont obligatoires car elles règlent les problèmes de synonymie et de polysémie.

Le caractère obligatoire des secondes ne fait pas l'unanimité. Certains jugent essentiel de produire la structure logico-sémantique des familles de concepts. Pour d'autres, ces relations sont en filigrane des termes recherchés. Il est difficile de trancher.

Il est évident que pour aider les utilisateurs/chercheurs, ces relations peuvent être nécessaires pour mieux circonscrire les termes à rechercher, généraux et spécifiques. Dans le même sens, elles peuvent aider l'indexeur. Le compromis acceptable pour ceux qui ne peuvent investir dans une telle structuration est au moins de commencer avec la constitution d'une liste de termes contrôlés et, après une ou deux années, de produire une structure plus complète. Les expériences d'indexation et de recherches étant faites et, de par la flexibilité de cette liste, il sera toujours possible de construire les relations hiérarchiques.

Les troisièmes relations devraient être facultatives même s'il s'agit d'interrelations entre les concepts. Constituer un réseau de relations associatives représente un travail important. Très souvent, à cause de l'ampleur que cela représente, on limite le nombre d'associations. Nous considérons que les relations associatives amènent une complexité supplémentaire et augmentent l'investissement en temps/personne et en argent. Un utilisateur devrait être amené à faire lui-même ces relations et à les introduire dans sa stratégie de recherche. Sans les minimiser, elles ne servent après tout qu'à donner des pistes vers des concepts complémentaires qui n'ont pas la même importance que les concepts généraux associés à des concepts spécifiques qui leurs appartiennent. Cette position peut paraître paradoxale mais elle est aussi de l'ordre du compromis.

CONCLUSION: VERS UNE POLITIQUE D'INDEXATION

Dans la mesure où nous voulions démontrer que le traitement descriptif de documents et de dossiers s'avère partiel lorsque nous utilisons par exemple, les plans de classification, les métadonnées, les DTD et les règles de description des documents d'archives, nous avons utilisé des documents que quiconque pourrait retrouver dans son fonds d'archives. Ces exercices sont suffisamment concluants pour démontrer que le contexte de création est le point de vue qui est privilégié. Quant au contenu, nous n'hésitons pas à croire que plusieurs semblent vouloir se contenter du premier niveau informationnel lié aux fonctions et aux activités. Il faudrait chercher les raisons qui conduisent à un pareil choix. L'archiviste doit aider le chercheur en lui procurant des outils de recherche où le contenu informationnel des documents et des dossiers aura été analysé et traduit de façon efficace et utile. L'indexation s'avère être une activité essentielle à ajouter aux autres activités plus traditionnelles de l'archivistique. Elle paraîtra contraignante sous plusieurs aspects mais elle contribuera à affirmer le rôle fondamental de l'archiviste comme spécialiste du traitement de l'information.

Sans avoir élaboré une véritable politique d'indexation, nous avons voulu énoncer quelques principes qui nous apparaissent particulièrement importants à considérer pour sa réalisation.

Ainsi, nous croyons que l'indexation doit s'inscrire dans un continuum évolutif qui respecte le cycle de vie des documents. De cette manière, l'indexation est évolutive et pluridimensionnelle. Bien que le sujet demeure important, le point d'ancrage de toute indexation est le concept. Les critères d'exhaustivité, de cohérence, de spécificité et d'exactitude doivent servir à donner les assises d'un véritable travail professionnel. Nous suggérons d'exercer un contrôle du vocabulaire qui, dans un premier temps, peut

se limiter aux relations d'équivalence (1^{er} niveau de contrôle) avec l'utilisation de descripteurs et de non-descripteurs. Nous croyons qu'une complète et satisfaisante indexation n'est possible que si nous arrivons à l'automatiser en nous assurant de la participation des utilisateurs, des chercheurs, des informaticiens et des indexeurs.

Michel Lévesque Chef du Service de la gestion documentaire, Directeur général des élections du Québec

NOTES

1. Les travaux à ce sujet ne sont pas encore terminés. L'idée première est de faire reposer la classification sur une structure plus souple et appropriée aux domaines d'activités de l'organisme. Des facettes, telles l'objet, l'agent, le domaine d'activités, l'action posée, permettraient par une approche combinatoire de décrire le document et de le contextualiser.
2. Les pages vertes font partie du *Répertoire gouvernemental québécois* conçu dans le cadre de l'infrastructure gouvernementale. Il comprend également des pages blanches et bleues qui colligent respectivement l'information permettant de connaître, situer et rejoindre une personne au gouvernement ainsi que l'information sur les programmes, activités et services des ministères et organismes gouvernementaux.
3. L'exhaustivité (profondeur) est liée au nombre de concepts identifiés/nombre de termes d'indexation issus de ces concepts. La spécificité est liée à l'exactitude avec laquelle une notion est représentée. La cohérence (ou uniformité) correspond à la capacité chez plusieurs individus, ou chez un seul, d'identifier les mêmes concepts, de les traduire de la même façon, pour les mêmes documents. L'exactitude est en lien avec la spécificité et correspond à la conformité de l'indexation par rapport au contenu des documents indexés (BCA 1992; ANQ 2000).

BIBLIOGRAPHIE

- ARCHIVES NATIONALES DU QUÉBEC (ANQ). [1996]. *Normes et procédures archivistiques*. 6^e édition. Québec, Publications du Québec/Ministère de la Culture et des Communications, 67-72.
- ARCHIVES NATIONALES DU QUÉBEC (ANQ). 2000. *Guide d'indexation des Archives nationales du Québec*. Sainte-Foy, ANQ, [106 p]. (Version non publiée).
- BUREAU CANADIEN DES ARCHIVISTES (BCA). 1990. Comité de planification sur les normes de description. *Règles pour la description des documents d'archives*. Version révisée. [Ottawa], BCA. Pagination multiple.
- BUREAU CANADIEN DES ARCHIVISTES (BCA). 1992. *L'indexation par sujets en archivistique*. Rapport du Groupe de travail sur l'indexation par sujet. [Ottawa], BCA, 154 p.

- CASAULT, Denis et Michel LÉVESQUE. 1997. *Guide de classification et de classement des documents des associations et autres organismes de même nature*. Sillery, Association des archivistes du Québec, 54 p.
- COUTURE, Carol *et al.* 1999. *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Presses de l'Université du Québec, 559 p.
- CEDARS PROJECT TEAM. 2000. *Metadata for Digital Preservation: The CEDARS Project Outline Specification*. [Royaume-Uni], CEDARS Project Team, 33 p.
- DIRECTION DE LA DOCUMENTATION FRANÇAISE. 1992. *BIPA: Thésaurus d'information politique, économique et sociale*. 2^e édition. Paris, La Documentation française, 317 p.
- EL HADI, Widad Mustafa, Jacques MANIEZ et Steven A. POLLITT, dir. 1999. *Structures and Relations in Knowledge Organisation*. Proceedings 5th international ISKO Conference, Lille (France), 25-29 August 1998. Germany, ERGON, 1999. 438 p. *Advances in Knowledge Organisation*, Vol. 6.
- HÉON, Gilles. 1999. La classification. In Carol Couture *et al.* *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Presses de l'Université du Québec, p. 219-254.
- HÉON, Gilles. 2000. *Comment classer vos archives personnelles et familiales. Vos papiers: supports et témoins de la vie quotidienne*. Sillery, Association des archivistes du Québec, 51 p.
- HUDON, Michèle. 1997-1998. Indexation et langages documentaires dans les milieux archivistiques à l'ère des nouvelles technologies de l'information. *Archives* 29, 1: 75-98.
- HUDON, Michèle *et al.* 1998. Séminaire sur les langages documentaires pour l'indexation des documents d'archives/6-7 février 1998. Sillery, Association des archivistes du Québec. Pagination multiple.
- MAUREL, Dominique et Michel CHAMPAGNE. 1999. La description et l'indexation. In Carol Couture *et al.* *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Presses de l'Université du Québec, p. 255-371
- NASA. 1999. Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-R-1, Red Book. [Washington], NASA, 132 p. (<http://www.ccsds.org/>)
- PAGOLA, Gracia. 1996. Analyse: condensation et indexation: survol théorique. In *Séminaire sur l'analyse en archivistique destiné à des formateurs*. Montréal, Association des archivistes du Québec, 1996. Pagination multiple.
- PITTI, Daniel V. 1999. Encoded Archival Description. An Introduction and Overview. *D-Lib Magazine* 5, 11, November, 9 p. (<http://www.dlib.org/dlib/november99/11pitti.html>)
- QUÉBEC (Province). Conseil du trésor. 1999a. *Plan de classification*. Québec, CT, 1999. Pagination multiple.

- QUÉBEC (Province). Conseil du trésor *et al.* 1999b. *Rapports des travaux du chantier en ingénierie documentaire*. Québec, CT, Pagination multiple. (13 rapports: 1. La gestion des documents adaptée à l'inforoute gouvernementale (rapport synthèse); 2. Les composantes d'un document électronique; 3. Choisir un environnement logiciel approprié au cycle de vie du document; 4. Conserver les documents électroniques: comment et pourquoi?; 5. La protection du document électronique: aspects techniques et juridiques; 6. Lignes directrices pour la description d'un document de référence aux fins de conservation et de diffusion; 7. Lignes directrices pour la description d'un document de transaction enregistré à des fins de conservation; 8. Lignes directrices pour la description d'un message de courrier électronique enregistré aux fins de conservation; 9. La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique; 10. Thésaurus de l'activité gouvernementale: orientations et principes directeurs; 11. Thésaurus de l'activité gouvernementale: liste alphabétique des termes; 12. Rôles professionnels et ingénierie documentaire; 13. La sécurité inforoutière)
- SOCIÉTÉ D'HABITATION DU QUÉBEC (SHQ). 1999. *Plan de classification*. Québec, SHQ. Pagination multiple.
- SOCIETY OF AMERICAN ARCHIVISTS (SAA). 1998. *Encoded Archival Description Tag Library: Version 1.0*. Chicago, SAA, pagination multiple. ([Http://lcweb.loc.gov/ead/tglib/tlhome.html](http://lcweb.loc.gov/ead/tglib/tlhome.html)).