

ENTRE FORMALISATION ET DÉCONSTRUCTION : ÉTAT DE L'ART CRITIQUE DE L'APPLICATION DOCUMENTAIRE DES ONTOLOGIES ET FOLKSONOMIES DANS LE DOMAINE DE L'INDEXATION DU PATRIMOINE CULTUREL

Seth van Hooland
Doctorant

Université Libre de Bruxelles
Av. F.D. Roosevelt 50, CP 123
1050 Bruxelles
Belgique

Résumé

Faciliter la recherche dans le patrimoine culturel numérique est un défi actuel pour les gestionnaires des musées, archives et bibliothèques. En formalisant à l'aide d'ontologies les relations entre ressources culturelles, auteurs, lieux et dates de production, des nouvelles fonctionnalités de recherche pourraient émerger. Mais si le développement d'un thésaurus classique représente déjà une opération complexe pour la plupart des instituts culturels, la création des ontologies pour le même secteur ne risque-t-elle pas d'être trop ambitieuse ? Indépendamment de ces développements largement académiques, une autre tendance dans la gestion de l'information numérique a émergé récemment, sous la forme du social tagging ou des folksonomies. Cette indexation libre a gagné une popularité énorme parmi les utilisateurs de blogs et de gestionnaires de signets. Depuis peu de temps, des prototypes de ce genre d'outils ont été également mis en place sur des sites de musées et de bibliothèques, dans le but de donner plus de pouvoir aux utilisateurs. Mais quelle est la valeur documentaire de ces outils ? Et est-ce que ces systèmes d'information offrent réellement une approche plus démocratique à la gestion de l'information ?

Mots clés : patrimoine culturel, numérisation, méta données, ontologies, folksonomies

Between formalisation and deconstruction : a critical state of the art of the application of ontologies and folksonomies within the digital cultural heritage domain.

Abstract

Facilitating the access to our digital cultural heritage is one of the biggest challenges for the moment within the domain of museums, archives and libraries. The development of ontologies which would formalize the relations between cultural resources, creators, places

and time of production could offer new exciting possibilities. However, if the development of a thesaurus is already highly problematic for most cultural institutions, is the idea of implementing ontologies not all too ambitious ? Independent of the academic research on ontologies, another new approach regarding the indexing of digital documents has risen, under the name of social tagging or folksonomies. This form of uncontrolled indexing by the user community emerged from the blogosphere, but has recently been implemented within collection management systems of certain museums and libraries. But what is exactly the added-value of folksonomies and do they really offer a more democratic access to our heritage ?

Keywords : cultural heritage, digitalisation, metadata, ontologies, social tagging

Introduction

Les pratiques de l'indexation et du catalogage sont sujets à l'évolution technologique des outils documentaires et au statut changeant de ses pratiquants. Ces évolutions ont clairement un impact sur les descriptions formelles, sur le contenu et donc également sur l'accès et l'interprétation des documents par les usagers. Le domaine du patrimoine culturel nous offre des exemples intéressants de cette problématique. Lors de la création et l'institutionnalisation de nos musées, archives et bibliothèques au XIXe siècle, l'indexation et le catalogage d'œuvres d'art étaient considérés, dans la tradition de l'ekphrasis, comme une science ou un art. La naissance du métier du documentaliste au XXe siècle a professionnalisé les pratiques descriptives. Depuis une dizaine d'années, l'utilisation du terme «metadata» démontre l'implication croissante à la fois de l'informatique mais également des utilisateurs lors de la description du patrimoine. Ces évolutions font surgir des questions sur la qualité des méta-données, leur évolution dans le temps et à travers des différents systèmes documentaires. L'article aborde cette problématique en analysant deux nouveaux langages documentaires qui sont actuellement appliqués dans le secteur culturel : les ontologies et les folksonomies.

1. L'INDEXATION ET LE CATALOGAGE N'ONT JAMAIS ETE AUSSI EN VOGUE

1.1 Naissance de la notion de méta-données

Jadis le domaine des documentalistes, bibliothécaires et archivistes a été relativement fermé, les opérations de catalogage et d'indexation de l'information sont depuis quelques années de plus en plus exercées par d'autres intervenants. La prolifération des ressources électroniques sur le web a fait apparaître, il y a une dizaine d'années, la notion de méta-données. Dans un premier temps le terme faisait surtout référence aux standards de schéma de données descriptifs. L'exemple le plus connu est sans doute le Dublin Core¹, qui vise à standardiser la description minimale d'une ressource électronique afin de faciliter son interopérabilité. Mais graduellement le terme a été utilisé pour désigner tout type de données issues de la description formelle et du contenu de ressources électroniques, les deux étapes de la chaîne documentaire traditionnellement désignées catalogage et indexation. Ce changement de vocabulaire reflète l'influence croissante de la communauté de recherche en informatique. Malgré son caractère pluridisciplinaire et l'importance de l'implémentation technique sur le terrain, les sciences de

¹ <http://dublincore.org/>

l'information et de la documentation se sont retrouvées sous les ailes des sciences humaines et ont été influencées par leur tradition épistémologique. Mais les informaticiens ont de plus en plus « envahi » le domaine de recherche des sciences de l'information et de la documentation. Après la maturité de l'indexation full-text qui repose surtout sur la force brute de calcul des machines, les informaticiens se sont de plus en plus investis dans les méthodes de balisage de l'information à l'aide d'XML. Ce langage est particulièrement utile pour la gestion et l'échange de méta données et son succès a énormément fait avancer le domaine de l'indexation et de la recherche d'information. Ce langage, en raison de son caractère fédérateur, est particulièrement utile pour la gestion des méta-données. De ce fait, son succès a enrichi le domaine de l'indexation de l'information. Le travail actuellement entrepris au niveau du web sémantique essaie d'aller encore plus loin en accordant à l'information des descriptions formelles susceptibles à être exploitées automatiquement.

Parallèlement à ces évolutions, les producteurs et utilisateurs de l'information sur le web ont de plus en plus l'opportunité de permettre un accès autonome aux ressources en ligne, et ceci pour trois raisons. En premier lieu, les gestionnaires d'un site web peuvent inclure dans les balises HTML des mots-clefs descriptifs. Ensuite, la publication d'un lien vers un certain site augmente sa visibilité dans les listes de résultats de certains moteurs de recherche. Et enfin, depuis peu de temps, des outils en ligne permettent à chacun d'ajouter librement des mots-clés à l'ensemble des informations publiées sur le web. Un musée américain a même lancé des « tag parties », pendant lesquelles le public est invité, dans une ambiance informelle, à décrire la collection. Ces pratiques ont donné à la gestion de méta données une image quasi « sexy » et fait que le terme « metadata » fait maintenant partie du jargon pseudo scientifique qui entoure le web et l'information numérique.

Ces deux tendances prétendent chacune de leur côté pouvoir révolutionner les pratiques de l'indexation et du catalogage. Les acteurs traditionnels, comme les documentalistes, bibliothécaires et archivistes, ont du mal à se positionner vis-à-vis des outils documentaires novateurs comme les ontologies et folksonomies. Les promoteurs des nouvelles approches de leur côté critiquent les outils traditionnels comme les plans de classifications ou les thésaurus. Les informaticiens leur reprochent un investissement humain trop élevé et un manque d'automatisation. Les adeptes des folksonomies critiquent par contre la nature autoritaire et non démocratique des vocabulaires contrôlés. Cet article essaie de dépasser ce débat qui est souvent guidé par des arguments trop stéréotypés et vides de sens.

1.2 Domaine d'application du patrimoine culturel numérique

Dans le cadre de ma thèse de doctorat, je m'intéresse plus particulièrement à l'accès au patrimoine culturel numérique, et notamment aux collections d'images numériques. Les autorités régionales, nationales et européennes investissent depuis une décennie fortement dans la numérisation des collections des musées, des archives et des bibliothèques pour diverses raisons. Mais cette opération ne se limite pas à passer d'un support physique au format numérique : « La mobilité est en contradiction avec la stabilité, et la plupart des éléments rapportés ne supportent pas le déplacement – comme ces bons sauvages que les anthropologues ne se laissaient jamais d'envoyer en Europe et dont la plupart mouraient misérablement – ou deviennent pleins de vermine, comme ces grizzlis que les naturalistes empaillaient trop rapidement, ou s'assèchent, telles ces précieuses graines que les naturalistes ont planté dans un sol trop pauvre. Même les éléments qui résistent au voyage peuvent n'avoir plus de sens une fois qu'ils se retrouvent dans les sous-sols des musées, parce qu'on ne possède pas d'informations suffisantes sur le contexte d'où ils proviennent. » (Latour 2005). Comme ces objets exotiques ramenés par les scientifiques, les images numérisées perdent complètement leur pertinence en tant que documents historiques et culturels si elles ne reçoivent pas une description qui les replace dans leur contexte d'origine.

Mais le catalogage et l'indexation des images offrent des défis particuliers, par rapport aux documents textuels. En effet, les possibilités d'une indexation automatique sont très limitées. Au mieux, le sens dénotatif d'une image peut être identifié à l'aide d'un logiciel de reconnaissance de formes. Mais la richesse du sens connotatif ne se laisse pas automatiquement déduire. Par ailleurs, les images possèdent très peu de méta-données « intrinsèques », ce qui est en revanche le cas avec beaucoup de documents textuels qui donnent eux-mêmes déjà beaucoup d'informations sur leur auteur, date et lieu de création. Enfin, la profondeur de l'indexation est parfois extrêmement difficile à déterminer, dans le sens où une image peut être décomposée dans un grand nombre de sous-ensembles informationnels. La somme de ces caractéristiques engendre des pratiques d'indexation et de catalogage fortement complexes.

Mis à part ces caractéristiques connues, j'aimerais analyser deux aspects peu traités jusqu'alors, qui sont néanmoins problématiques : l'évolution de la notion du patrimoine et le

souci d'obtenir une interopérabilité entre ressources hétérogènes. D'une part, depuis les années 1970, nous avons assisté à un élargissement graduel de la notion de patrimoine. Des objets du quotidien et issus de la culture populaire peuvent aujourd'hui revendiquer leur place dans les musées, archives et bibliothèques (Gielen 2005). Cette évolution a également eu un impact sur les modes d'accès au patrimoine et donc à la description de celui-ci. Le public exerce de plus en plus de pouvoir sur la façon dont le patrimoine est présenté et rendu accessible. Pensons par exemple aux études qui sont faites dans le monde muséal sur le temps passé par les visiteurs devant certaines oeuvres pendant leur visite à une exposition. Les résultats sont ensuite utilisés pour modifier l'organisation du parcours de l'exposition. Ce pouvoir se manifeste également au niveau de l'indexation. Depuis quelques décennies, un nombre croissant d'instituts culturels font appel à des amateurs, des témoins ou d'autres personnes qui ne sont pas des experts pour décrire ou commenter des biens culturels historiques, matériels ou immatériels. Cette approche est surtout utilisée dans le cas où trop peu d'informations historiques sont disponibles ou lorsque la thématique sort du canon classique de l'historiographie, comme par exemple l'histoire culturelle de certaines minorités. Ces commentaires sont souvent diffusés sous la forme de narrations qui aident à construire une identité personnelle ou collective au travers de l'image. De ces pratiques résultent deux niveaux d'indexation différents dans le sens où la description normative de l'institution est confrontée à des descriptions subjectives venant du public. Toutefois, cette tension fait partie intégrale de la notion d'histoire, comme le démontre Aleida Assmann dans l'article «Funktionsgedächtnis und Speichergedächtnis – Zwei Modi der Erinnerung » (Assmann 1995). Assmann démontre à travers les pensées de Nietzsche, Maurice Halbwachs et Pierre Nora la dichotomie entre la mémoire « fonctionnelle » et la mémoire « factuelle ». La première notion correspond aux traditions, qui font revivre le passé par le moyen de l'actualité et la deuxième à l'historiographie scientifique. Les nouvelles caractéristiques des collections de photographies historiques qui sont publiées en ligne, comme leur malléabilité et leur facilité d'accès, ont fortement contribué à l'insertion de ces collections dans la mémoire fonctionnelle.

D'autre part, la volonté d'obtenir une interopérabilité entre différentes collections numérisées introduit également une problématique nouvelle. Cette ambition a été formulée dès les débuts de l'informatisation des instituts culturels à la fin des années 1960. Le développement du

protocole Open Archives Initiative (OAI)², qui permet l'agrégation de méta-données venant de plusieurs collections hétérogènes annonçait un premier grand pas vers cette interopérabilité. L'utilisation du Dublin Core permet de construire dans des portails conformes au protocole OAI une interface de recherche unique pour interroger plusieurs collections avec une seule requête. Quelques projets européens comme le European Visual Archives³ ont de plus tenté de permettre des recherches multilingues, en traduisant automatiquement les requêtes et les méta données. L'implémentation de standards et techniques du « web sémantique », comme dans le MultimediaN projet⁴, annonce une étape supplémentaire dans l'intégration de l'information hétérogène. Même s'il est évidemment intéressant d'implémenter des technologies innovantes, je doute parfois de la valeur ajoutée des portails qui veulent agréger un maximum de documents. Si la mise en commun de collections hétérogènes peut offrir une vision plus riche de l'information, elle peut, en contrepartie, impliquer une perte de contexte.

2. DEUX TRADITIONS DIFFÉRENTES DES SCIENCES DE L'INFORMATION ET DE LA DOCUMENTATION

Les ontologies et folksonomies proposent deux modèles qui se situent aux deux extrémités du spectre des vocabulaires documentaires. Mais tous deux ont l'ambition de révolutionner le paysage de la description des ressources en ligne. Les deux termes sont chacun évocateurs mais fort vagues, et peu de consensus existe sur leurs définitions. Les ontologies forment un sujet de recherche très populaire depuis quelques années dans les départements informatiques universitaires. Les folksonomies de leur côté sont surtout propagées par la communauté des blogueurs. Peu de recherches scientifiques ont été réalisées sur ce sujet. Les quelques travaux publiés récemment viennent surtout des champs de recherche pluridisciplinaires tels que les : « media studies », « social informatics » ou « humanities computing ».

L'intérêt d'une analyse conjointe de ces deux modèles documentaires vient surtout du fait qu'ils illustrent des visions et traditions épistémologiques différentes dans les sciences de l'information et de la documentation: « There has always been evident within the study and management of information a schism between two different world views. On the one hand is

² <http://www.openarchives.org/>

³ <http://www.eva-eu.org/>

⁴ <http://e-culture.multimedian.nl/>

the classical view of logic and digital computing which focusses on the underlying order of information. On the other hand is the view of truths as culture-bound, knowledge as culturally dependent and information as a manifestation of power. » (Tredinnick 2006) Les ontologies traduisent l'influence de la logique formelle et de l'informatique qui tentent, en se concentrant sur l'ordre sous-jacent de l'information, de formaliser et classifier le réel. Les folksonomies par contre représentent l'influence des sciences humaines telles l'histoire, la sémiologie et les études culturelles qui mettent l'accent sur ou s'intéressent à la relativité culturelle de l'information et l'importance du contexte dans lequel un document est construit et interprété.

2.1 Les ontologies

Le terme a été emprunté par les informaticiens à la philosophie, où l'ontologie est considérée comme un système de catégories décrivant la réalité. Le travail en informatique sur les ontologies doit être situé dans le cadre plus large de la quête vers un web sémantique. En représentant des ressources et leurs méta-données d'une manière fortement formalisée et standardisée, deux buts principaux sont visés : l'exploitation automatique et l'interopérabilité de l'information. Mais comme Fabien Gandon l'a noté dans « Le Web sémantique n'est pas antisocial », très peu de consensus existe sur les enjeux et la portée du web sémantique. Mais entrer dans le débat de savoir s'il est possible limiter ou non le web sémantique à un ensemble de formalisations, je constate que l'implémentation concrète de cette idée repose actuellement sur deux standards : Ressource Description Framework (RDF)⁵ et Ontology Web Language (OWL)⁶. RDF offre, comme son nom l'indique, un modèle pour la description d'une ressource, dans un format « sujet – prédicat – objet ». Ce modèle permet alors de construire des déclarations simples et univoques qui peuvent s'imbriquer les unes les autres et devenir ainsi des réseaux sémantiques. L'identification unique à l'aide d'un URL de chaque composant permet une gestion automatisée. Comme prédicat entre un livre et son auteur, le champ « creator » du Dublin Core peut être utilisé, auquel l'URL <http://purl.org/dc/elements/1.1/> est associé. La syntaxe du RDF ne permet toutefois pas une expressivité riche.

OWL, qui est à la base d'une ontologie à proprement parler, permet de préciser des contraintes complémentaires à celles proposées par RDF. OWL est composé de trois parties : tout d'abord de classes qui représentent des groupes d'individus partageant les mêmes caractéristiques,

⁵ <http://www.w3.org/RDF/>

⁶ <http://www.w3.org/2004/OWL/>

ensuite de propriétés qui permettent de définir des faits ou des relations entre ces classes et, enfin, des instances, c'est à dire des individus d'une classe qui peuvent prendre les caractéristiques définies par les propriétés.

Citons trois avantages de l'utilisation de RDF et OWL. Premièrement, les ontologies permettent la construction de schémas de méta données plus riches. L'ABC ontologie par exemple permet de formaliser la notion de temporalité dans la description d'une ressource (Hunter 2001). Les schémas de méta données traditionnels sont souvent construits autour de la ressource considérée comme stable dans le temps. Mais cette hypothèse devient problématique dans le cas de la numérisation d'une photo historique. Comment par exemple décrire une image qui se trouve dans le format d'une plaque de verre de 1890, un tirage d'entre deux-guerres et le fichier en haute résolution produit par un scan de la plaque de verre ? Deuxièmement, les ontologies permettent l'application d'un raisonneur automatique, ce qui permet non seulement de tester si l'ontologie créée est cohérente mais aussi d'inférer automatiquement la hiérarchie des classes. Troisièmement, en formalisant l'ensemble des relations et des contraintes entre les concepts d'un domaine d'application, une ontologie peut guider la création d'un schéma de méta-données spécifiques. Pour le secteur du patrimoine culturel, le CIDOC Conceptual Reference Model (CRM)⁷ prévoit des définitions et une structure formelle pour la description des concepts et des relations utilisées dans la documentation du patrimoine culturel. En offrant ce canevas dans lequel toute information issue de ce domaine peut être située, le CIDOC CRM espère promouvoir l'interopérabilité entre les collections.

Mais malgré ces avantages, un ensemble de contraintes limite fortement l'application des ontologies dans le domaine du patrimoine numérique. La création et la gestion à long terme d'une ontologie demandent des investissements humains et intellectuels importants. Chacun ayant une expérience de terrain relative à la création d'un thésaurus connaît les difficultés intellectuelles et organisationnelles qui se présentent lors de l'élaboration d'un vocabulaire contrôlé. D'après mon expérience, seules les grandes institutions qui peuvent assurer un financement sur le long terme atteignent cet objectif. Au regard du niveau de complexité supplémentaire des ontologies et du manque chronique de moyens financiers dans le secteur, il me paraît plutôt illusoire de croire à l'implémentation de cet outil documentaire. Les retours financiers sont trop faibles pour défendre l'investissement dans la création de méta-données

⁷ <http://cidoc.ics.forth.gr/>

nécessaires. L'histoire de l'implémentation des nouvelles technologies dans le secteur culturel nous montre aussi la faible adaptation de systèmes coûteux. Les bases de données relationnelles, qui existent depuis 30 ou 40 ans, pourraient offrir des avantages énormes au niveau de la gestion des collections, numériques ou analogues. Mais à l'exception des bibliothèques, très peu d'organismes culturels ont vraiment profité des possibilités offertes par cette technologie. La volonté politique d'investir dans la numérisation du patrimoine numérique existe, mais le manque de résultats visibles sur le court terme rend d'éventuels investissements nouveaux, pour augmenter la qualité des méta-données, peu attractifs.

Du point de vue des utilisateurs finaux, la question se pose aussi de savoir s'ils peuvent vraiment profiter des nouvelles possibilités de recherche offertes par les ontologies. Ces dernières permettent d'élaborer des requêtes complexes, mais en considérant les difficultés que la plupart des utilisateurs rencontrent déjà avec la rédaction de requêtes booléennes et la consultation de l'interface d'un thésaurus, l'utilité des ontologies pour le public est peut-être limitée. La diffusion d'une partie simplifiée de l'ontologie pour le grand public, en cachant la complexité, qui elle serait exploitée par des spécialistes, constitue peut-être une façon pragmatique de contourner cette difficulté.

Enfin, les ontologies sont souvent critiquées pour leur ambition de vouloir formaliser le réel, en réduisant des signes à des signaux univoques et de proclamer l'indépendance de la signification vis-à-vis du contexte de la production du signal.

2.2 Les folksonomies

Ce néologisme désigne la description de ressources en ligne pratiquée par la communauté d'internautes. La description se fait de manière libre, en accordant des mots-clés au choix à des sites-web, des articles scientifiques, de la musique, des photos, etc à l'aide d'une interface simple. L'ensemble de mots-clés attribué par les différents utilisateurs forme un vocabulaire documentaire, qui peut être interrogé par les mots-clés les plus souvent utilisés. Cette approche est surtout utilisée pour la gestion de signets, des billets de blogs et des collections de photos privées stockées en ligne. Cette manière d'indexer de l'information, par sa nature libre, diffère fortement des outils traditionnels et a engendré un débat sur la nature contrôlée des vocabulaires documentaires.

Quelques instituts culturels, surtout en Amérique, expérimentent en ce moment les folksonomies et font donc appel à leur public pour attribuer des mots-clés aux images de leurs bases de données. La technologie est relativement facile à implémenter et ne demande pas d'intervention du personnel. Dans le monde muséal, l'exemple le plus connu est le Art Museum Social Tagging Project⁸, auquel des musées importants comme le Metropolitan Museum of Art et le Guggenheim participent. Les Archives Normandie⁹ proposent de leur part sur le site Flickr une partie de leur collection et invitent le public à attribuer des mots-clés et des commentaires. Le système intégré de gestion de bibliothèque libre PMB¹⁰ a même déjà intégré une fonctionnalité qui permet aux utilisateurs de l'OPAC de la bibliothèque d'attribuer des mots-clés à des notices.

Mais quelle est la valeur sur long terme du « social tagging » pour le patrimoine culturel ? Le reproche traditionnel exprimé par les adversaires des folksonomies est souvent le manque de contrôle du vocabulaire, ce qui implique bien évidemment des problèmes de synonymie, polysémie ainsi que des niveaux de discours différents. Mais dans la mesure où une institution ne possède pas les moyens de décrire elle-même sa collection, la description par le public peut offrir une alternative valable, même si la qualité des descriptions n'est pas toujours constante.

Au-delà de ce reproche traditionnel, il me semble tout aussi important d'analyser la problématique historique de la production de méta-données par les utilisateurs sur le web. L'idée de donner à l'utilisateur le pouvoir de classer ou de décrire de l'information sur le web n'est pas nouvelle. Le mécanisme du Pagerank de Google qui guide le classement des résultats d'une requête repose (ou a reposé) entre autres sur le nombre de liens qui pointent vers le site en question. Donc en incluant un lien sur son site vers un autre, un gestionnaire d'un site web peut donner plus de poids et donc favoriser son accessibilité. Par ailleurs, l'auteur d'une page web peut inclure dans des balises méta du code HTML des mots-clefs qui décrivent le contenu du site. Mais ces deux systèmes ont été complètement détournés et rendus inutilisables par des pratiques de « spamdexing ». Rien ne garantit que le spam n'envahira pas les folksonomies, ce qui est par exemple déjà le cas avec le site de gestion de signets Del.icio.us¹¹, qui est un des premiers sites à incorporer les folksonomies.

⁸ <http://www.steve.museum/>

⁹ <http://www.archivesnormandie39-45.org/>

¹⁰ <http://www.sigb.net/>

¹¹ <http://del.icio.us/>

Ajoutons que, la description par le public change profondément le cycle documentaire traditionnel. Jusqu'à présent, la classification de l'information précédait l'accès. Mais comme Ertzscheid et Gallezot le remarquent : « Jusqu'ici, l'accès était globalement subordonné au classement. Or c'est désormais le classement qui pourrait être subordonné à l'accès. » (Ertzscheid 2006) Les « tagclouds » qui sont constitués par les mots-clés les plus utilisés offrent donc une interface de recherche qui est justement basée sur l'accès et la description de l'information par le public. Un phénomène semblable se produit avec certaines procédures de numérisation d'images. Des instituts qui n'ont pas les moyens de numériser leurs collections à une grande échelle préfèrent parfois numériser uniquement les documents qui sont demandés par le public. À court terme, cette approche pragmatique ne pose pas de problème. Mais laisser faire la sélection uniquement par le public me paraît délicat. C'est justement le rôle des instituts culturels de préserver et de faciliter l'accès à l'ensemble du patrimoine, donc également aux documents moins sollicités par le public.

Une autre critique y est fortement liée. Un des premiers principes de l'indexation et du catalogage est la tentative d'être objectif et d'éviter une interprétation personnelle du document en question. Les folksonomies incorporent par contre parfois des mots-clés intrinsèquement liés à l'indexeur, par exemple « to_read » (à lire). La notion sociale introduite par le caractère parfois subjectif du choix des mots-clés offre sans doute des potentialités et représente un des avantages des folksonomies. Mais une indexation narcissique n'est sans doute pas utile pour tout type de collection, ou n'intéresse pas chaque groupe d'utilisateurs. Les commentaires écrits par des lecteurs qui sont publiés sous les descriptions de livres vendus sur Amazon peuvent être utiles. Mais parfois leur sur présence noie dans un certain sens l'objet d'intérêt initial, le livre, surtout quand des commentaires sont formulés sur les commentaires. C'est d'ailleurs une illustration du danger général des méta-données : les couches méta peuvent s'accumuler jusqu'à l'infini. (Boydens 2001)

3. Conclusions

Cet article a démontré que l'évolution des outils documentaires illustre également un changement dans les cadres mentaux et les traditions intellectuelles qui guident les pratiques du catalogage et de l'indexation. Les ontologies et folksonomies sont dans ce contexte des

outils exemplaires, par leur caractère contradictoire et leur ambition vis-à-vis des outils traditionnels.

Reste la question de la pertinence des ontologies et folksonomies pour la gestion documentaire du patrimoine culturel. Les critiques formulées par ces deux nouveaux outils dans l'introduction de cet article sur de l'approche traditionnelle des thésaurus et plans de classifications peuvent être réfutées. La création et gestion des ontologies demande un investissement humain trop élevé pour que leur adaptation soit justifiable dans le domaine culturel. Et s'il est vrai que par exemple la Classification Décimale Universelle (CDU) reflète une vision du monde trop occidentale, les folksonomies ne peuvent pas non plus prétendre que les nuages de balises, les « tagclouds », n'offrent qu'une interface biaisée à l'information en ligne, quand le mot-clé « Paris » ne fait pas référence à la capitale française mais à l'icône du non-event.

Néanmoins, les intérêts de standards comme RDF ne peuvent pas être niés. Si l'application d'ontologies complexes dans le domaine du patrimoine culturel n'est pas pour demain, il existe peut-être un intérêt immédiat dans la conversion de thésaurus dans le format SKOS par exemple. Les folksonomies de leur côté ont le mérite d'avoir lancé le débat sur la description de ressources par le public. Les commentaires d'utilisateurs peuvent par exemple avoir une grande utilité dans des bases de données d'images historiques pour augmenter la qualité des méta-données existantes (van Hooland, 2006). De manière plus générale la mise en place de méthodes d'analyses critiques de méta-données, qu'il s'agisse d'ontologies ou de folksonomies, constitue une condition sine qua non en vue d'une exploitation adéquate de l'information. En effet, les deux systèmes de méta-données sont complémentaires : « La mémoire fonctionnelle sans interaction avec la mémoire factuelle résulte dans des phantasmes, mais l'absence de la mémoire fonctionnelle réduit la mémoire factuelle à une masse d'information sans relevance sociale. » (Assmann,1995)

Bibliographie

Assmann Aleida, 1995. Funktionsgedächtnis und Speichergedächtnis – Zwei Modi der Erinnerung. Generation und Gedächtnis. Erinnerungen und kollektive Identitäten. Platt Kristin et Dabag Mihran. Opladen: Leske+Budrich, p. 169-185.

Boydens Isabelle, 2001. Informatique, normes et temps. Bruxelles : E. Bruylant.

Ertzscheid Olivier et Gallezot Gabriel, 2006. Indexation sociale et continents documentaire, in Document numérique et société. Actes de la conférence DocSoc-2006, Chartron Ghislaine et Broudoux Evelyne. Paris : ABDS éditions, p. 291-306.

Gandon Fabien, 2006. Le Web sémantique n'est pas antisocial. Proceedings Ingénierie des Connaissances, Semaine de la Connaissance, Nantes, pp. 131-140, 20-30 Juin 2006.

Gielen Pascal et Laermans Rudi, 2005. Cultureel Goed. Over een (nieuw) erfgoedregiem. Leuven : LannooCampus.

Hunter Jane et Lagoze Carl, 2001. The ABC ontology and model. *Journal of Digital Information*, vol. 13, n° 2.

Latour Bruno, 2005. La science en action. Introduction à la sociologie des sciences. Paris : La Découverte.

Tredinnick Luke, 2006. Digital information contexts : theoretical approaches to understanding digital information. Oxford : Chandos.

Van Hooland Seth, 2006. User-generated metadata : possibilities for the cultural heritage sector. In: Immaculate Catalogues: Taxonomy, Metadata and Resource Discovery in the 21st Century, Proceedings of CILIP Conference, University of East Anglia, UK, 13-15 September 2006.

