

HIÉRARCHIE ET ANARCHIE

Dépasser l'opposition entre organisation centralisée et distribuée ?

Isabelle Boydens

Lavoisier | *Les Cahiers du numérique*

2010/3 - Vol. 6
pages 77 à 101

ISSN 1622-1494

Article disponible en ligne à l'adresse:

<http://www.cairn.info/revue-les-cahiers-du-numerique-2010-3-page-77.htm>

Pour citer cet article :

Boydens Isabelle , « Hiérarchie et anarchie » Dépasser l'opposition entre organisation centralisée et distribuée ?,
Les Cahiers du numérique, 2010/3 Vol. 6, p. 77-101.

Distribution électronique Cairn.info pour Lavoisier.

© Lavoisier. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

HIÉRARCHIE ET ANARCHIE

Dépasser l'opposition entre organisation centralisée et distribuée ?

ISABELLE BOYDENS

L'article se décline en deux parties. Un éclairage historique montre, dans un premier temps, que la coexistence des organisations centralisées et distribuées du savoir n'est pas neuve. Depuis l'occident médiéval à nos jours, en effet, l'information est traitée sur l'un ou l'autre mode selon sa nature et selon la valeur que la société lui attribue. Il en découle une apparente interaction entre hiérarchie et anarchie, interaction qui semble s'accélérer de nos jours avec le développement croissant des technologies de traitement de l'information. Toutefois, ces mêmes technologies peuvent fournir des solutions opérationnelles en vue de maîtriser l'incertitude générée. Dans une seconde partie, l'article présente un modèle temporel en vue d'évaluer et d'améliorer la qualité de l'information empirique, quel qu'en soit le mode d'organisation. Le modèle repose sur une approche herméneutique et sur un développement de la théorie en matière de modélisation conceptuelle. Plusieurs exemples concrets d'application de la méthode sont présentés dans le domaine des bases de données administratives et dans celui des systèmes documentaires culturels.

1. Introduction. La dialectique de la hiérarchie et de l'anarchie menace-t-elle la pérennité du savoir ?

Avec l'accroissement sans précédent de la performance des technologies de traitement de l'information selon une organisation sociale distribuée, une part du savoir actuel paraît échapper à tout processus de validation collectif, centralisé et maîtrisé. De ce fait, les modes de structuration des connaissances traditionnels qu'offrent l'archivistique, la muséologie ou la bibliothéconomie, par exemple, perdraient une part de leur efficacité.

La distinction entre les modes d'organisation hiérarchique centralisée et sociale distribuée de la connaissance¹ est tendancielle : ces modes ne définissent pas de partition discrète et figée entre les systèmes de production de l'information².

L'organisation hiérarchique centralisée s'apparente à un système contrôlé selon un savoir formalisé et reconnu institutionnellement à un moment donné de l'histoire. Nous y incluons la production d'informations et de documents³ traduisant un savoir dont la valeur et les enjeux sont officiellement reconnus (par exemple, les productions officielles de l'Église sous l'Occident médiéval ou, de nos jours, les bases de données déployées dans les institutions scientifiques, industrielles, financières, militaires ou les administrations publiques). Cela peut recouvrir les bases de données structurées⁴ mais aussi,

1. Par connaissance, nous entendons ici toute théorie formalisée à un moment de l'histoire en tant que savoir institutionnellement validé. À l'instar d'Escarpit, on peut considérer le savoir comme « une construction assez stable pour freiner l'écoulement du temps et rester disponible à l'échelle d'une vie humaine, assez mouvante pour accueillir sans cesse de nouvelles réponses et les insérer dans des structures lisibles » (Escarpit, 1991).

2. Par information (Ganascia, 1998), nous entendons la mise en forme d'une représentation du réel au moyen d'un code artificiel lisible par une machine ou d'un langage (humain, si l'on exclut les mondes animaux et végétaux) et requérant une interprétation au regard d'un ensemble de connaissances données (lesquelles sont elles-mêmes progressivement bâties sur la base de l'interprétation des informations qu'elles produisent).

3. Un document peut être considéré comme une information figée sur un support physique, quel qu'en soit le médium. Certains documents sont dits « non structurés » (une photographie, par exemple) ; ils peuvent en outre être associés à des collections de données structurées descriptives.

4. Une base de données est une collection organisée de données structurées. Une donnée structurée (Elmasri *et al.*, 2007) est un triplet (i, d, v) composé des éléments suivants : un intitulé (i), renvoyant à un concept (une catégorie d'activité, par exemple), un domaine de définition (d), composé d'assertions formelles (contraintes d'intégrité)

dans le domaine documentaire, les systèmes de repérage formels à base d'index, les représentations bibliographiques, les langages documentaires contrôlés (par exemple, classificatoires ou combinatoires) ainsi que les pratiques associées en matière d'analyse documentaire et de recherche documentaire (un système documentaire automatisé incluant une base de données structurée pour en gérer les métadonnées descriptives).

L'organisation sociale distribuée s'apparente à un processus producteur de savoir et d'information, issu d'un ensemble d'individus non régis *a priori* par une autorité quelconque. Cela ne l'empêche pas de véhiculer un savoir et de produire une information susceptible d'influer le cours de l'histoire (nous y incluons les mouvements Goliards ou les hérésies populaires au Moyen-âge ou, de nos jours, les forums de discussion sur internet ou les blogs diffusant une information constitutive de notre savoir actuel, tel que, par exemple, les mouvements d'opinions ou, plus spécifiquement, la correspondance des journalistes situés dans des pays en guerre). Contrairement aux systèmes d'information soutenant une organisation hiérarchique centralisée, dans une organisation sociale distribuée, le savoir repose sur un substrat informationnel et documentaire non codifié et informel : information orale, langage libre... Au regard du premier, ce second type d'organisation s'apparente à un mouvement créateur tendanciellement anarchique.

Dans cet article, nous proposons de montrer (section 2) qu'il y a, dans le temps, coexistence, et non rupture entre l'organisation hiérarchique centralisée et l'organisation sociale distribuée.

En effet, des modes hétérogènes d'organisation du savoir peuvent coexister bien que traditionnellement, événements et structures soient articulés sur la base du temps « physique », partitionnant le réel observé en une séquence linéaire de périodes distinctes et successives. La notion de période crée ainsi une unité factice entre des éléments hétérogènes et, nous le verrons, « en un même moment du temps, tous les éléments observés ne sont pas au même stade d'évolution » (Prost, 1996).

D'un point de vue historique, nous observerons que l'organisation hiérarchique centralisée n'est pas neuve et qu'on la rencontre dès notre Moyen-âge. De même, les folksonomies, ces classifications représentatives du web 2.0 reposant sous le seul contrôle de l'individu qui les crée et symptomatiques de l'organisation sociale distribuée, se retrouvent dans des pratiques médiévales. Si

spécifiant l'ensemble des valeurs admises dans une base de donnée pour ce concept (une liste contrôlée de valeurs alphabétiques, par exemple) et, enfin, une valeur (v) à un instant t (le secteur de la chimie, par exemple).

L'interaction entre hiérarchie et anarchie est inévitable, elle n'est toutefois pas uniquement le produit de la technologie.

Cet éclairage historique révélera en outre que l'information, quel que soit son mode d'organisation, n'est pas exempte d'incertitude. En fonction des enjeux qu'une société attribue au savoir à un moment donné de l'histoire, l'information associée relèvera tendanciellement tantôt d'une organisation centrale hiérarchisée, tantôt d'une organisation sociale distribuée. L'incertitude touche davantage la seconde mais elle affecte tout type d'organisation, quelle qu'en soit la solidité apparente. Les origines de cette fragilité conceptuelle tiennent à la nature empirique de l'information.

L'incertitude est latente au cœur de toute organisation. Hiérarchie et anarchie demeurent indissociablement liées dans le processus de construction de notre savoir. Toutefois, à l'heure où la technologie fait proliférer l'information, elle offre également les moyens associés en vue d'évaluer nos systèmes d'information et d'en améliorer la qualité.

Dans cet esprit, la section 3 aborde les perspectives offertes par les évolutions technologiques contemporaines. Notre tradition offre des instruments conceptuels (l'heuristique, la critique historique et l'herméneutique) pour interpréter l'information empirique, quelle qu'en soit l'organisation. Sur cette base, l'informatique fournit les outils opérationnels pour contribuer à gérer une part de l'incertitude induite par les pratiques d'organisation de l'information (Boydens, 1999, 2007).

2. Évolution historique des organisations du savoir

Certains auteurs affirment que la structure sémantique des notices bibliographiques est intrinsèquement pérenne et n'a évolué dans le temps que depuis le XX^e siècle, avec l'apparition de l'informatique et avec ses évolutions technologiques (Fox *et. al.*, 1996). Aucune étude historique en la matière n'aurait dès lors d'intérêt... Nous proposons de montrer, au contraire, que la sémantique des documents et de leurs représentations n'a cessé d'évoluer dans le temps et dans l'espace et que l'étude de cette évolution est fondamentale d'un point de vue épistémologique et opérationnel.

Les organisations hiérarchiques centralisées et sociales distribuées coexistent en dépit des inventions techniques. Celles-ci n'induisent pas de rupture brutale dans les modes de production de l'information. Ainsi, l'invention de l'imprimerie ne change pas la forme qu'a pris le livre avec le codex entre le II^e et le IV^e siècle (Labarre, 1989).

Nous envisagerons, pour aborder l'organisation centralisée, l'évolution des processus de repérage au sein du document (émergence des index) ainsi que de plusieurs épisodes de l'histoire des représentations bibliographiques et des langages documentaires (2.1.). Pour traiter l'organisation sociale distribuée, à l'ombre de la première (2.2.), nous considérerons l'histoire de la reproduction du document. Très vite, dans l'un et l'autre cas, l'histoire du document et celle de ses représentations sont étroitement liées (McKenzie, 1991). En effet, information et méta-information sont indissociables dans leur évolution, le langage naturel étant son propre métalangage.

En fonction des enjeux et de la valeur qu'une société attribue à un savoir donné et à l'information associée, ces derniers seront tantôt organisés selon un mode hiérarchique centralisé, tantôt social distribué et, de ce fait, plus ou moins sujets à l'incertitude (2.3.).

2.1. L'organisation hiérarchique centralisée est-elle invariante ?

2.1.1. Émergence des index

Au ^ve siècle de notre ère, après les invasions barbares, la tradition ne favorise pas la création d'outils de recherche documentaire (Mary *et al.*, 1989). Après la désagrégation sociale qui suit les invasions germaniques, les érudits avaient essentiellement pour tâche de conserver et d'assimiler le texte en profondeur. Les monastères bénédictins s'y attachent et ne questionnent pas la tradition biblique. Pour ces raisons politiques et idéologiques, on ne développe ni index, ni classification : « le Moyen-âge n'aimait pas l'ordre alphabétique qu'il considérait comme une antithèse de la raison » (Mary *et al.*, 1989). L'ordre alphabétique ou tout autre instrument de recherche artificiel aurait permis au lecteur de procéder à une lecture subjective, voire subversive.

Il n'est pas surprenant que les premiers index et dictionnaires (comme celui de Papias au ^{XI}e siècle) ne rencontrent pas de succès immédiat. C'est avec l'émergence des hérésies que se développent les systèmes de repérage autorisant une consultation plurielle du livre. Dans la foulée de la *devotio moderna*, notamment, les outils s'appuyant sur l'ordre artificiel offrent au clergé des instruments de repérage dans les textes sacrés afin de donner de fréquents sermons pour l'édification des laïcs.

Au ^{XIII}e siècle, l'Université parisienne applique ces techniques aux œuvres profanes jugées dignes d'intérêt dont Aristote (Mary *et al.*, 1989). Au ^{XIV}e siècle, s'organisent des professionnels de l'index. Le développement de l'index et de repères de lecture plus ou moins neutres se déploie dans un contexte de désacralisation du texte. Les résultats de ces pratiques d'indexation, souvent

ingénieuses, sont les précurseurs de nos « fichiers inversés » présents au cœur de tout système d'informatique documentaire contemporain.

Ce développement est toutefois progressif et hésitant. Le premier colophon est identifié à Mayence en 1457. Il inclut le nom de l'auteur, le titre, le lieu, la date et le nom de l'imprimeur (Labarre, 1989). Mais en 1473, 45 % des incunables répertoriés dans le catalogue général des incunables ne comportent pas de colophon : le filigrane du papier, la forme des caractères typographiques ou encore, les notes manuscrites anciennes contribuent à l'identification hypothétique du lieu et de la date d'impression. En décalage avec nos pratiques contemporaines, remarquons par ailleurs que la première bibliographie universelle de Conrad Gesner (1546) classe les ouvrages par ordre alphabétique des prénoms des auteurs (Martin, 1989b).

Sur le fond, de la fin du XIII^e siècle, avec Lulle, au XVII^e siècle, avec Bruno, les ouvrages sur la classification du savoir se multiplient. Ils sont parfois très critiques car l'émergence des procédés de repérage au sein du texte et, ensuite de classification de ceux-ci, s'accompagne de difficultés d'interprétation. Celles-ci sont liées à l'interaction empirique du contenu des textes et des conjonctures idéologiques et culturelles ambiantes. Ainsi l'éclectisme des ouvrages de la Renaissance rend périlleuse la classification du savoir selon un ordre rationnel; on parle à cette époque de *quodlibet* pour caractériser ces ouvrages qui semblent inclassables selon nos catégories de pensée moderne. Dans son *Pantagruel*, Rabelais compare d'ailleurs en 1532 l'encyclopédie à un abîme (Martin, 1989b).

Ces facteurs d'incertitude, qu'ils se situent au sein de la structure descriptive du texte ou au cœur même de leur interprétation, sont toujours latents dans nos classifications contemporaines.

2.1.2. *Bibliographie virtuelle (XVII^e et XVIII^e siècles)*

Un double phénomène aura plus tard un impact sur la qualité du contenu des bibliographies occidentales françaises (Moureau, 2006) : d'une part, sur le plan politique, Louis XIV impose la censure suite à la révocation de l'édit de Nantes en 1685 (Roche, 1990 ; Netz R., 1997). D'autre part, sur le plan économique, l'essor de la bourgeoisie marchande conduit les éditeurs à braver les interdictions en produisant des ouvrages prohibés ou contrefaits (Sauvy, 1990). Citons par exemple *Le traité sur la dîme royale* de Vauban (1707) qui fut prohibé et contrefait à plus de 1 000 exemplaires en deux ans (Sauvy, 1990). On observe alors un véritable capitalisme éditorial, la diffusion étant parfois facilitée par l'interdiction qui constituait un argument commercial. Pour contourner la censure, les éditeurs incluent des coordonnées falsifiées de lieux et de dates d'édition. Tantôt, cette littérature illicite s'accompagnait de fausses

adresses, tantôt, elle était dépourvue de toute mention de nom d'auteur, de libraire, d'imprimeur ou de lieu d'édition (Sauvy, 1990). Outre l'impact sur les métadonnées descriptives, à côté des renseignements bibliographiques, l'imprimé lui-même est souvent de piètre qualité et les différentes éditions pouvaient présenter des variantes.

On peut parler pour caractériser ce phénomène de « bibliographie virtuelle » car près de 50 % de la littérature d'alors échappe aux recensions bibliographiques officielles : on estime que la masse de livres contrefaits et prohibés a, aux XVII^e et XVIII^e siècles, une importance comparable à celle du livre permis (Sauvy, 1990). De nos jours encore, ce phénomène constitue pour la philologie et la critique historique un vaste chantier d'analyse critique, en dépit de la solidité apparente des modes de classification centralisée du savoir contemporain (Parinet, 2004).

2.1.3. De la CDD à la norme SKOS

Dans le contexte du positivisme du XIX^e siècle, apparaît la DDC (*Dewey Decimal Classification*) de Melvil Dewey. Elle est adoptée internationalement en 1895 par la Conférence bibliographique internationale de Bruxelles et donne le jour à la CDU (classification décimale universelle). Celle-ci reflète l'idéologie occidentale de son époque, avec ce qu'elle comporte d'arbitraire (Vanbinsbergen, 1994).

On a ensuite vu l'émergence de la classification à facettes de Ranganathan à la fin des années 1950 offrant une vision conceptuelle de l'univers. Elle propose une hiérarchie du concret vers l'abstrait (personnalité, matière, énergie, espace, temps), laquelle n'est pas exempte de philosophie indienne et a ensuite éclaté en codifications spécialisées (Escarpit, 1991).

Sont ensuite apparus les langages combinatoires, tels les thésauri (Hudon, 2009), normalisés par l'ISO. Ceux-ci, par définition, ne reposent pas sur une classification hiérarchique. Mais ils contribuent, de par leur rigueur, à la mise en ordre et à l'organisation centralisée du savoir. On assiste depuis cinq ans à l'émergence des normes du « web sémantique » qui se développent avec le soutien du *world wide web consortium* (W3C). Le pouvoir d'expression de ces normes est très riche : il inclut à la fois la sémantique d'un langage documentaire combinatoire, celle d'un modèle de base de données ainsi que la syntaxe d'un langage déclaratif, de type Prolog. Mais la prolifération, parfois anarchique, des langages documentaires requiert des normes d'interopérabilité. C'est ainsi que le 18 août 2009, la norme SKOS (*Simple Knowledge Organization System*) a été déclarée recommandation officielle du W3C. Elle a pour objet la conversion de tout type de langage documentaire contrôlé préexistant (thésaurus, classification, ...) en une syntaxe homogène. Toutefois, son usage

implique un investissement intellectuel important et rigoureux et sa généralisation dans l'espace ouvert d'internet est utopique (Boydens, 2009). Il est cependant raisonnable de l'appliquer dans le cadre d'environnements ciblés et contrôlés, comme c'est déjà le cas, par exemple, dans le cadre du thesaurus *Agronoc* de la FAO (*Food and Agriculture Organization of the United Nations*).

Le XX^e siècle voit par ailleurs naître la théorie en matière de modélisation conceptuelle fondatrice du caractère opératoire des bases de données structurées actuelles (Volle, 2006) : nous y aurons recours dans la suite de cet article, en vue de l'évaluation et de l'amélioration de la qualité des systèmes d'information structurés et non structurés contemporains (section 3). Une base documentaire inclut en effet un système de gestion de bases de données qui en structure l'information descriptive.

En résumé, l'idéologie dominante du début du Moyen-âge en Occident s'est opposée à la création de repères de lecture dans les textes sacrés. Par la suite, l'Église répond aux hérésies avec la production d'index qui rayonneront peu à peu dans les milieux universitaires laïcs. L'organisation hiérarchique centralisée du savoir est d'emblée parcourue d'une dynamique. À mesure qu'ils se consolident, les instruments de structuration et de description de l'information sont sujets à l'incertitude, se heurtant au caractère évolutif des réalités empiriques sous-jacentes (2.1.1). Un exemple remarquable de ce phénomène apparaît ensuite avec les bibliographies virtuelles aux XVII^e et XVIII^e siècles (2.1.2.) : face à la censure imposée par l'État, la solidité des repères bibliographiques est fragilisée par la réaction des éditeurs qui emploient des indications falsifiées pour contourner les interdictions officielles. La solidité des modes de catégorisation du réel semble restaurée avec le positivisme du XIX^e siècle (et, par exemple, l'apparition de la CDD) mais leur rigidité rend ces classifications inadéquates à la fluidité du réel empirique ; peu à peu, au XX^e siècle, apparaissent des langages documentaires plus riches et souples afin de rendre compte de l'élasticité du réel sous-jacent. Face à la richesse et à la diversité des langages documentaires institutionnalisés au fil du temps, apparaissent des standards visant à l'interopérabilité entre systèmes d'information. Toutefois ces standards sont par nécessité d'une complexité telle que leur mise en œuvre requiert des moyens humains et financiers importants (2.1.3.). On observe ainsi dans l'histoire des structures formelles et descriptives hiérarchiques centralisées une dynamique plus ou moins génératrice d'incertitude à mesure que celles-ci se déploient.

2.2. *L'organisation sociale distribuée, à l'ombre de l'organisation hiérarchique centralisée*

Parallèlement à l'émergence des index et des langages documentaires, se déroule une autre évolution, celle des modes d'organisation sociale distribuée. Plusieurs facteurs idéologiques, sociologiques, économiques ou politiques ont un impact sur la forme mouvante du manuscrit et de sa représentation bibliographique. On y retrouvera en germe les caractéristiques de l'organisation sociale décentralisée du web 2.0.

2.2.1. *La reproduction du texte*

La transformation de l'information est une préoccupation séculaire. Avant l'invention de l'imprimerie, les systèmes d'information se déployaient avec les générations de moines copistes qui recopiaient les manuscrits de siècles en siècles. Ces manuscrits nous sont souvent parvenus sous forme morcelée et partielle : il arrive que l'original soit perdu et que l'on ne dispose que d'un ensemble incomplet de copies divergentes (suite à des erreurs – volontaires ou non – commises par les copistes : passages transformés, omis ou ajoutés).

Dans le contexte de la littérature profane et de la goliardise, les copistes anonymes aspirent à l'individuel (Vezin, 1989). Ils racontent dans les manuscrits qu'ils recopient, entre les lignes et dans les marges, leur état d'esprit ou encore, la circonstance de la copie (Bourgain, 1989). Les marges des manuscrits pouvaient ainsi contenir de la mauvaise poésie, des blagues paillardes, des sentiments anticonformistes ou encore, des images obscènes. En outre, les copistes pouvaient inconsciemment commettre de nombreuses erreurs de retranscription ou, consciemment, modifier le texte. Le lecteur qui ne corrige pas le manuscrit dont il se sert est parfois taxé d'incompétence et de paresse. Dans l'Europe médiévale, s'agissant de la littérature profane et populaire, il eut été considéré comme impensable que le travail de quelqu'un soit parfait et ne puisse être amélioré (Bourgain, 1989).

Malgré le caractère téléologique du livre (« *incipit, finit* »), on observe ainsi des « manuscrits sans fin ». Le manuscrit, à l'instar de l'hypertexte aux XX^e et XXI^e siècles, est un médium fluide. La nature mouvante et fragile du texte est presque aussi manifeste que dans la prolifération des blogs. Cet irrespect du texte se retrouve dans les pratiques actuelles du *social tagging* (Boydens *et al.*, 2009). Le phénomène concerne surtout la littérature profane et beaucoup moins les textes sacrés ou les grands noms dont on dénature plus rarement les textes : Saint Bernard, Saint Thomas d'Aquin, ... (Bourgain, 1989). On verra plus loin (2.2.3) qu'une distinction analogue peut être effectuée de nos jours. En effet, selon les enjeux de l'information, à savoir la valeur que la société lui accorde à un

moment donné de l'histoire, on observera tantôt une production « mouvante », fragile et éphémère de l'information tantôt, une production organisée, à haute valeur ajoutée et « sous contrôle », à propos de laquelle nous avons évoqué l'émergence des index et des classifications précédemment (2.1).

2.2.2. De la littérature dite « éphémère » au web 2.0

Citons un phénomène ultérieur d'ordre politique et économique illustrant une autre tendance à la propagation sociale décentralisée de l'information. Il s'agit de l'émergence des périodiques, dont les précurseurs, appelés « occasionnels », datent de la fin du XV^e siècle (Chartier, 1989).

Sur le plan politique, ils prennent de l'essor avec les flambées pamphlétaires du XVI^e siècle dans la foulée des guerres de religion et de la propagande réformée. La littérature éphémère génère une dynamique de « questions-réponses » par pamphlets interposés à l'instar des forums contemporains. Celle-ci porte les débats sur la place publique : « elle diffuse des faisceaux de thèmes, des chaînes de textes liés à un réseau de réponses et d'échos, un engrenage polémique » (Chartier, 1989).

Sur le plan économique, la conjoncture du livre étant défavorable au XVIII^e siècle, la production imprimée tend à se périodiser sous des formes diverses : placards, billets... (Sgard, 1990).

L'apparition du périodique, indépendante de toute innovation technologique, illustre l'influence des cultures sociales sur le processus éditorial et institutionnalisera une restructuration des métadonnées bibliographiques qui devront ultérieurement prendre en compte de nouvelles caractéristiques formelles.

Cette brève esquisse de l'histoire de la reproduction du manuscrit explique que l'on parle de nos jours de « nouvelle esthétique médiévale » pour caractériser l'hypertexte sur internet, le lien hypertextuel étant un instrument de spoliation, dans l'esprit de l'irrespect du texte observé plus haut dans la reproduction du manuscrit. Dans le même esprit, les productions actuelles du web 2.0 (blogs, forums, folksonomies...) se caractérisent par un retour à l'individuel, une multiplicité des ouvertures et une atténuation de la frontière entre l'auteur et le lecteur, s'agissant de l'information de type « grand public ». Remarquons que les facteurs économiques précédemment évoqués sont plus présents que jamais. Avec les enjeux de la publicité contextuelle sur internet, les mots-clés se vendent aux enchères en fonction de leur popularité relative par sociétés marketing interposées. Le *google zeitgeist* (esprit du temps), baromètre continu de la popularité des mots-clés utilisés dans le monde lors des

recherches sur internet, illustre ce phénomène dont le marché atteignait déjà 5 milliards de dollars aux États-Unis en 2006 (Boydens, 2009).

Résumons. S'agissant de la reproduction du texte, alors que l'on tente de préserver l'intégrité des textes bibliques et des auteurs faisant autorité, les documents considérés comme mineurs pour la société d'alors sont sans cesse affectés par une intervention individualiste qui en fragilise le contenu initial (2.2.1). Cette forme d'irrespect du texte se retrouve de nos jours dans les forums et autres productions du web 2.0. Ponctuellement, toutefois, sous l'effet des évolutions politiques et économiques dominantes, on observe une institutionnalisation de ces pratiques, ainsi, avec l'émergence de la littérature dite « éphémère » (2.2.2). À l'ombre de l'organisation hiérarchique centralisée, l'organisation sociale distribuée fait elle aussi l'objet d'une dynamique dans l'histoire.

2.3. *Combinaison contemporaine des modes d'organisation*

La coexistence des organisations hiérarchiques centralisées et sociales distribuées se poursuit de nos jours. Les premiers instruments de repérage contrôlés se rapportent surtout à la production documentaire officielle (écritures saintes, auteurs illustres) alors que l'irrespect du manuscrit frappe plutôt la littérature profane. Le positionnement relatif des deux modes d'organisation dépend des valeurs et des enjeux que la société attribue au savoir de son époque (Boydens *et al.*, 2009).

Très populaire, le web 2.0 ne révolutionne pas les modalités de production du savoir dans le monde industriel. Le recours à la co-construction, à l'instar des *mashups*, combinaison de données issues de plusieurs applications ou services (par exemple, Google Maps, parmi les plus robustes et fameux), reste marginal dans le monde industriel ou de l'administration électronique. En effet, dans un contexte où les enjeux opérationnels et financiers sont stratégiques, il n'est pas envisageable d'avoir recours à des applications dont on sait a priori qu'elles sont volatiles et peu fiables. Les *mashups* souffrent en effet de l'absence de garantie quant à la pérennité des sources qui les alimentent. On parle souvent à leur sujet « d'applications jetables » (Groebbens *et al.*, 2009). De même, si les blogs sont potentiellement sans cesse alimentés, il est communément admis que leur contenu antérieur ne soit pas mis à jour et les « liens morts » ou informations désuètes y abondent. Applications et données (incluant aussi les « tags » constitutifs des folksonomies, relevant du seul contrôle de l'individu qui les crée) issus du web 2.0 sont appelés pour cette raison « situationnels ».

Au sein du monde scientifique, les applications de type web 2.0 enrichissent les contacts informels entre chercheurs et constituent un facteur de créativité incontestable. Toutefois, le processus de publication du savoir scientifique relève toujours d'institutions centralisées faisant autorité. Des serveurs internationaux tels que Dialog, dont les origines sont antérieures à internet, diffusent ainsi des centaines de bases de données bibliographiques : on y trouve, par exemple, dans le domaine de la médecine, la base Medline, produite par la *US National Library of Medicine* ou, dans le domaine des sciences humaines, l'*Educational Resources Information Center*, produite par l'*US National Library of Education*. À côté de ces ressources traditionnelles, dont l'accès est payant, il faut citer l'initiative des archives ouvertes (van de Graaf *et al.*, 2007) visant l'échange libre (en réaction au monopole financier des grandes maisons d'édition) d'archives via internet, et, notamment, des publications scientifiques validées. Ce mouvement est maintenant soutenu par des institutions telles que, en France, l'INSERM, l'INRIA ou le CNRS.

Nous concluons donc, au terme de ce survol historique, que la coexistence des modes d'organisation hiérarchique centralisée et sociale distribuée n'est pas neuve. Au fil du temps, en fonction des enjeux que la société attribue à un moment donné de l'histoire à la connaissance, ces deux modes d'organisation, l'un tendanciellement hiérarchique et l'autre, tendanciellement anarchique, interagissent. Sujets tous deux à l'incertitude, les systèmes d'information associés, de par leur nature empirique, doivent faire l'objet d'une analyse interprétative car l'information, même quand elle est générée dans un cadre hiérarchique centralisé, est intrinsèquement fragile. Les classifications descriptives et les valeurs qui leur sont associées portent sur une réalité empirique et évolutive toujours sujette à interprétation et à questionnement. Pour ces raisons, il est crucial de rechercher des solutions opérationnelles pour mieux interpréter l'information et en faciliter la gestion, quel qu'en soit le mode d'organisation.

3. Perspectives opérationnelles

L'effet des technologies sur la masse et la rapidité de transmission des informations amplifie l'interaction entre hiérarchie et anarchie. Mais ces mêmes technologies permettent aussi d'affronter les problèmes de qualité.

Notre démarche s'appuie sur deux piliers : d'une part, un cadre herméneutique temporel et, d'autre part, un développement de la théorie en modélisation de base de données. En effet, sur le plan opérationnel, la plupart des systèmes d'informations structurés ou non structurés incluent une base de données qui en gère les métadonnées ou les champs descriptifs. Après une

présentation de la méthode d'analyse (3.1.), nous évoquons plusieurs cas d'applications significatives généralisables à l'organisation hiérarchique centralisée (3.2.) et sociale distribuée (3.3.).

3.1. Évaluer et améliorer la qualité des bases de données et des systèmes documentaires : méthode

C'est au XV^e siècle, à partir de la fameuse étude philologique et comparative de Laurent Valla que l'on place traditionnellement les débuts de la critique historique (Valla, trad. Glard, 1993). Afin d'établir l'appareil critique d'un manuscrit, l'historien construit un *stemma codicum* (généalogie des données), technique d'analyse comparative empruntée à la philologie (Marichal, 1961). L'établissement du *stemma codicum* ainsi que l'analyse du lieu, de la date ou de l'auteur d'un manuscrit relèvent de ce qu'on appelle la critique externe (ou de provenance). À la critique externe succède la critique interne (ou de crédibilité), proche de l'herméneutique.

Peut-on appliquer la critique historique aux sources informatiques ? Nous avons produit une méthode d'analyse généralisable à l'évaluation et à l'amélioration opérationnelle de la qualité de vastes systèmes d'information administratifs contemporains (Boydens, 1999 ; 2007)⁵. Nous en évoquons brièvement les mécanismes principaux dans la mesure où ceux-ci peuvent être utiles en vue d'évaluer et d'améliorer la qualité de l'information empirique, qu'il s'agisse de systèmes d'information hiérarchiques centralisés ou sociaux distribués. L'exposé suit quatre étapes : la qualité des données, l'absence de référentiel, les dimensions temporelles de l'herméneutique et, enfin, les incidences opérationnelles des trois points précédents sur la théorie en matière de modélisation de bases de données.

3.1.1. Qualité des données

La qualité d'une base de données désigne son adéquation relative aux objectifs qui lui sont assignés (Boydens, 1999 ; Redman, 2001 ; Madnick *et al.*, 2009). La « qualité totale » n'existe pas car le concept est relatif : sur la base d'un arbitrage de type « coûts-bénéfices », les dimensions de la qualité les plus pertinentes (fraîcheur, rapidité de transmission, précision des données...)

5. Nous menons ces travaux au sein du centre de recherche de Smals, société informatique prestataire de services pour l'administration fédérale belge, dont nous coordonnons le *Data Quality Competency Center*, en parallèle avec nos recherches menées au Département « Sciences de l'information et de la communication » de l'Université Libre de Bruxelles.

devront être retenues dans un contexte donné. On parle de *fitness for use*, d'adéquation aux usages.

La question soulève des enjeux considérables lorsque l'information est un instrument d'aide à la prise de décision, voire d'action sur le réel. Ainsi, en mai 1999, pendant la guerre du Kosovo, l'OTAN a bombardé par erreur l'ambassade de Chine à Belgrade : les bases de données cartographiques utilisées alors pour guider les missiles répertoriaient un plan de la ville inadéquat. D'où l'attaque inopportune et l'incident diplomatique qui ont suivi (Boydens, 2002).

Afin d'aborder cette problématique, une approche herméneutique est indispensable. En effet, contrairement à ce qu'affirment certains (Wand *et al.*, 1996 ; Madnick, 2009), il n'existe pas, en l'absence d'erreurs formelles, de projection biunivoque nécessaire entre une base de données et les réalités observables qu'elle représente. Nous ne disposons d'aucun référentiel absolu en vue de vérifier la correction de l'information empirique, à savoir, l'adéquation de celle-ci à son objet.

3.1.2. Absence de référentiel

La question de l'absence de référentiel se traduit comme suit : afin de vérifier la correction de l'information répertoriée dans une base de données, il faudrait idéalement connaître a priori une réalité que seule cette base de données permet de connaître. Dans un autre domaine, la question a été exposée avec une grande clarté par Jean-Louis Besson, à propos de l'épistémologie des statistiques (Besson, 1992).

Par exemple, en médecine, comme dans toute discipline empirique, la théorie évolue au fil des expériences. Il est impossible dans ce cas, de vérifier la correction des systèmes d'information médicaux en les comparant au réel correspondant, la théorie et les observations correspondantes étant incertaines et évolutives.

C'est pourquoi, à la question « l'information est-elle correcte ? », il faut substituer la question suivante : « comment l'information se construit-elle progressivement ? ». Les faits empiriques (sujets à l'interprétation humaine) et les données correspondantes doivent être évalués par rapport à un principe unificateur, un horizon de similitude de sens. Afin d'appréhender ce processus de construction, l'approche herméneutique consiste en une mise en relation sans cesse renouvelée entre les observations et le contexte dans lequel elles s'insèrent (Aron, 1969 ; Koselleck, 1990). La question du temps est ainsi centrale. Nous recourons à deux modèles temporels qu'offre l'herméneutique : la notion de « temporalités étagées » (Braudel, 1976) et celle de « continuum évolutif » (Elias, 1996).

3.1.3. *Modèle herméneutique temporel*

Le concept de « temporalités étagées », théorisé par Fernand Braudel, est une construction permettant d'identifier, au sein d'un objet d'étude, une hiérarchie entre plusieurs séquences de transformation coexistantes. Dans le modèle de Braudel, les séquences relativement plus rapides (évolutions politiques, par exemple) sont conditionnées par les séquences relativement plus lentes (évolutions de la géographie et du climat, par exemple).

Appliqué à un système d'information empirique⁶, le concept de « temporalités étagées » clarifie le processus de construction de l'information et permet d'identifier plusieurs échelles de transformation interagissantes dont l'évolution est solidaire mais asynchrone. Par exemple, dans un système d'information administratif, nous pouvons distinguer le « temps long » de la législation, évoluant d'un trimestre ou d'une année à l'autre, le « temps intermédiaire » de l'appareil de représentation administratif et informatique dont les transformations s'opèrent d'une semaine ou d'un mois à l'autre et le « temps court » du réel observable, objet de la norme et de la représentation informatique et dont l'évolution est quotidienne. Régulièrement, en effet, des entreprises fusionnent, se scindent, d'autres disparaissent alors que de nouvelles professions ou catégories d'activité non prises en compte dans les nomenclatures officielles voient progressivement le jour, avec la diversification des métiers de l'informatique par exemple.

Mais la construction temporelle de Braudel doit se traduire par une stratification relative du temps au sein de laquelle les interactions entre séquences de transformation ne sont pas unidirectionnelles. Il arrive en effet que des séquences relativement plus rapides aient un impact sur des séquences de transformation relativement plus lentes, que les faits observés aient un impact sur les théories qui en ont guidé l'observation. Le modèle de Braudel peut alors être complété par le concept de « continuum évolutif » de Norbert Elias (Elias, 1996). Celui-ci montre que le temps est une construction résultant de la mise en relation de deux ou plusieurs séquences de transformations (chaque séquence s'apparentant à un « continuum évolutif ») dont l'une est normalisée en vue de servir aux autres d'étalon de mesure. Ce processus de construction est lui-même évolutif et intimement lié à l'état du « fonds de savoir » dont dispose la société dans lequel il s'insère. En d'autres termes, plusieurs « continnum évolutifs » coexistent, chacun d'entre eux étant à la fois l'objet normé et le référentiel normatif.

6. Le cadre temporel de Braudel a été sollicité dans différents domaines d'application ; certains l'ont utilisé en sciences de l'information à des fins descriptives, sans en examiner les prolongements opérationnels (Boyd, 1996 ; Ribes *et al.*, 2009).

Par exemple, dans le domaine des bases de données de la sécurité sociale, l'identification de la catégorie d'activité des employeurs est déterminante pour le calcul du taux de cotisations sociales qu'ils doivent payer à l'État. En Belgique, ces cotisations s'élèvent annuellement à quarante milliards d'euros environ. Les enjeux sociaux et financiers sont donc colossaux. Pour catégoriser les employeurs, la législation administrative utilise une nomenclature des activités européennes mise à jour selon une périodicité pluriannuelle. Mais entre chacune de ces mises à jour, la réalité économique évolue de manière quasi continue. Ainsi, quand se développèrent les « copy centers », ces boutiques mettant des photocopieuses à disposition de leurs clients, la nomenclature des activités européennes s'avéra rapidement inapte à leur recensement : elle proposait au mieux les catégories « imprimerie », « commerce de détail de livres » ou « secrétariat ». Afin de prendre en considération la catégorie « copy centers », issue du « temps court », il fallut d'abord modifier les textes réglementaires, puis adapter la structure des bases administratives en conséquence.

En section 2, nous avons observé le même phénomène de stratification temporelle s'agissant des repères de lecture et modèles bibliographiques. Ainsi, ces derniers apparaissent peu à peu pour structurer l'information associée au savoir de l'époque (« temps long »). Mais plus tard, avec les bibliographies virtuelles aux XVII^e et XVIII^e siècles, l'idéologie dominante qui impose la censure a un impact sur la qualité des représentations correspondantes (« temps intermédiaire »), le réel observable y étant représenté de manière partielle ou falsifiée suite aux efforts continus des éditeurs pour échapper, en réaction à la censure, à la répression tout en continuant de publier (« temps court »). *In fine*, ces réalités issues d'un temps relativement court, ont un impact sur le temps relativement long du savoir car nous ne disposons que d'informations partielles pour caractériser la production littéraire de l'époque. À des fins opérationnelles, il sera utile, de nos jours, de prendre ces interactions en considération afin de mieux interpréter et gérer les bases documentaires répertoriant le savoir de l'époque.

3.1.4. Développement en matière de modélisation conceptuelle

Pour des raisons opérationnelles, le fonctionnement d'une base de données reposant sur un modèle relationnel ou de type XML (parmi les plus courants à l'heure actuelle) repose sur l'hypothèse du monde clos, en vertu de laquelle toute valeur non incluse dans son domaine de définition est considérée comme fautive (Elmasri *et al.*, 2007). On parle alors de « violation de contrainte d'intégrité ». Toutefois, s'agissant des données empiriques, si l'on sort de ce cadre formel, il se peut qu'entre le moment où la structure de la base de

données a été formalisée et celui où l'information a été saisie, de nouvelles caractéristiques soient apparues au sein du domaine traité. Dans ce cas, il est impossible de vérifier la correction des valeurs de la base de données de manière automatique. En effet, lorsqu'une incohérence apparaît entre une valeur saisie au sein de la base et les tables de référence permettant d'en tester la validité, il peut s'avérer indispensable, selon les enjeux, de procéder à une vérification intellectuelle, en contactant les spécialistes du domaine, par exemple. L'impact de l'évolution du réel observable (séquence relativement plus rapide) sur la théorie et le système d'information associé (séquences relativement plus lentes) doit alors être pris en considération à des fins opérationnelles. Ce mécanisme n'est pas trivial puisqu'il implique des ressources humaines potentiellement importantes ⁷, tant dans le secteur administratif que dans le monde des banques ou, encore, celui des musées.

La théorie en matière de modélisation conceptuelle peut être enrichie si l'on distingue les erreurs des anomalies formelles associées aux données (Boydens, 1999 ; 2007). Les premières constituent une violation de contrainte d'intégrité certaine au sein d'une base de données (Elmasri *et al.*, 2007) : par exemple, une valeur numérique est présente dans un champ où sont attendues des valeurs alphabétiques. Les secondes s'apparentent à une présomption d'erreur formelle : une incohérence légale apparaît formellement (par exemple, entre la catégorie d'activité d'un employeur et le type de travailleur qu'il déclare) mais seule une interprétation humaine (avec une investigation sur le terrain, par exemple) permettra de détecter s'il y a erreur ou pas et quelle est l'information pertinente.

Afin d'en assurer une prise en compte semi-automatique, des prérequis s'imposent. Un système de détection des anomalies lors de la saisie mais aussi a posteriori doit être mis en place, de façon à disposer d'un historique des anomalies et de leur traitement ainsi que d'indicateurs de suivi latéraux. Des procédures claires quant à leur gestion doivent être établies, a fortiori lorsque la base de données s'inscrit dans un environnement fédéré.

Nous proposons maintenant plusieurs applications récentes et significatives de cette méthode. Nous présentons, pour l'organisation hiérarchique centralisée (section 3.2) un exemple d'application aux bases de données de la sécurité sociale belge (Boydens, 1999 ; 2007) et un autre appliqué à la base de données

7. « Des travaux, comme la présentation donnée par Simon Riggs à XML Europe 2003 ou le travail d'Isabelle Boydens (Informatique, normes et temps, Bruxelles, Éditions E. Bruylant, 1999) à propos de la qualité de vastes bases de données ont montré qu'environ 10 % des documents XML (ou des valeurs associées à des données structurées) contenaient au moins une erreur » (van der Vlist, 2003).

répertoriant les classifications du département ethnographique du Musée royal de l'Afrique Centrale à Bruxelles (van Hooland *et al.*, 2008).

Nous évoquons ensuite deux exemples d'application de l'approche aux ressources contemporaines ayant recours aux technologies 2.0 et reposant sur une organisation sociale distribuée (section 3.3). Il s'agit de la base du fonds photographique des *National Archives of the Netherlands* aux Pays-Bas (van Hooland, 2006) et du September 11th Memorial and Museum de New York (van Hooland, 2009 ; Boydens *et al.*, 2010).

3.2. Organisation hiérarchique centralisée

3.2.1. Bases de données de la sécurité sociale belge

À travers les indicateurs et l'organisation proposés section 3.1, le suivi statistique des violations de contraintes d'intégrité s'apparentant à des « anomalies formelles » permet de détecter non seulement les augmentations « anormales » (en fonction d'un seuil donné) d'anomalies mais aussi les augmentations de « validations » d'anomalies lors de la phase de traitement des bases de données administratives. Une opération de validation signifie qu'après examen, un agent a estimé que l'anomalie, qui est une présomption d'erreur, correspondait à une valeur pertinente. L'opérateur peut en effet « forcer » le système à accepter la valeur. Si le taux de telles validations d'anomalies est élevé et récurrent, la probabilité est grande que la structure de la base elle-même ne soit plus pertinente. Un algorithme émet alors un « signal » destiné au gestionnaire de la base afin qu'il examine si une modification structurelle de son schéma est requise. Ainsi, en Belgique, la mise en place d'une directive administrative en faveur du secteur « non marchand », a posé la question, au regard de la réalité progressivement appréhendée au sein de la base, de savoir s'il fallait inclure dans ce secteur les maisons de repos privées, a priori exclues car poursuivant des finalités lucratives. Initialement considérées comme des cas « erronés » au regard du domaine de définition spécifiant le secteur « non marchand », ces entreprises y ont finalement été intégrées, après interprétation juridique, sur la base de la méthode présentée ici. Ceci a donné lieu à une restructuration du schéma de la base de données. La restructuration de la base résulte d'une décision humaine tendant à rendre le modèle provisoirement conforme aux nouvelles observations. En l'absence d'une telle intervention, l'écart entre la base et le réel se creuse.

Si l'on n'adapte pas le schéma, les anomalies correspondant à ces cas vont continuer d'apparaître en masse, nécessitant un examen manuel potentiellement conséquent et ralentissant considérablement le traitement des dossiers

administratifs. Pour la sécurité sociale belge, la mise en œuvre de cette méthode a permis d'améliorer la précision et la rapidité de traitement des cotisations sociales, réduisant potentiellement de 50 % le volume d'anomalies formelles qui représentaient alors chaque trimestre de 100 000 à 300 000 occurrences à gérer manuellement (Boydens, 1999 ; 2007).

3.2.2. *Classifications du Musée Royal de l'Afrique Centrale à Bruxelles*

Dans le même esprit, une opération de *data profiling* (Olson, 2003), c'est-à-dire, d'audit formel de la base de données, a été appliquée aux classifications du département ethnographique du Musée Royal de l'Afrique Centrale à Bruxelles, (van Hooland, 2008 ; Boydens *et al.*, 2010). Ces classifications reposent sur une organisation centrale hiérarchisée, elles ne sont toutefois pas exemptes d'incertitude en raison de l'historicité des réalités empiriques sous-jacentes.

Il a été possible d'identifier, à partir de ce mécanisme, les écarts formels entre les définitions structurellement attendues et les valeurs saisies dans la base de données descriptive. Une analyse de structure (*pattern analysis*) permet par exemple, d'effectuer une typologie de tous les formats de date présents au sein du système et de quantifier chaque type : par exemple, champs vides, format de type 9999-9999 (par exemple : 1891-1912), de type AAA 9999 (par exemple : *mai 1938*), de type 99/99/9999 (par exemple : *04/08/1964*), ... Cette analyse est fondamentale en raison de l'incertitude associée à la datation en histoire.

Une automatisation partielle de cette approche permet d'aider les gestionnaires du système d'information à documenter la base de données. La prise en considération de la variété des formats de dates rencontrés et de leur fréquence contribue à évaluer et à améliorer l'interprétation de l'information. L'approche peut être enrichie en fonction, par exemple, de la typologie des objets datés (statuette, panier, masque, couteau...) et de leur contexte. Une interprétation a posteriori des modalités de datation passées pourra alors être effectuée : pourquoi a-t-on eu recours à un intervalle, à quoi sont dûs les champs vides (valeur incertaine, négligence lors de la saisie...)? Avec l'évolution des recherches et des idéologies, les modes de description des objets ethnographiques ont évolué dans le temps. La modélisation de la base de données a pu faire l'objet de contraintes d'intégrité plus ou moins rigoureuses, en fonction des fichiers disponibles et des gestionnaires de ceux-ci. Dès lors, il est utile, à des fins scientifiques, de prendre l'historique de ces évolutions en considération dans la définition des métadonnées descriptives. On observe à nouveau l'utilité opérationnelle du modèle temporel proposé : suite aux interactions entre les différents niveaux de la temporalité étagée, on peut être amené à enrichir et documenter la structure du système d'information avec l'interprétation des valeurs dont il a permis l'appréhension.

3.3. Organisation sociale distribuée

Dans certains domaines, les moyens financiers font structurellement défaut, les enjeux de l'information n'étant pas jugés par la société aussi stratégiques que dans les domaines militaires, industriels ou administratifs, par exemple. Citons le cas des institutions patrimoniales ou culturelles qui bénéficient généralement de subsides moins élevés et distribués de manière ponctuelle. Une approche de type web 2.0 est alors parfois envisagée.

3.3.1. National Archives of the Netherlands aux Pays-Bas

On peut citer les *National Archives of the Netherlands* aux Pays-Bas qui ont déployé pour leur fonds photographique un système de co-construction permettant aux utilisateurs de commenter et de compléter librement les métadonnées associées aux photographies (van Hooland, 2006 ; 2009). Alors que certains commentaires sont peu pertinents, d'autres sont extrêmement riches quant à l'identification de l'origine historique d'un document. Dans l'esprit de la méthode présentée dans cet article, ce fonds a fait l'objet d'une analyse critique spécifique (*data profiling*) couplée à une approche d'évaluation statistique) sur la base d'une grille d'analyse qualitative objectivable. Ainsi, à partir des fichiers de transaction, une analyse critique comparative a permis de confronter le questionnement sur les métadonnées aux commentaires associés à ces dernières et d'évaluer la pertinence relative de ces commentaires. L'analyse des corrélations statistiques entre commentaires et requêtes contribue à identifier les classes de commentaires les plus pertinentes sur le plan scientifique. Près de 50 % des requêtes se centrent sur les corrections de métadonnées décrivant des photographies (corrections formelles ou de fond relatives à l'identification d'événements, à leur localisation géographique ou à leur datation) ou à des éléments narratifs contextuels (van Hooland, 2006). Par contre, les commentaires relatifs à des jugements de valeur individuels occupent une place marginale (moins de 3 %) dans les requêtes. Les *tags* co-construits sont ainsi dotés d'un véritable appareil critique qui permet de guider l'interprétation des métadonnées correspondantes.

3.3.2. September 11th Memorial and Museum de New York

Une approche analogue a été initiée en 2009 au September 11th Memorial and Museum de New York (van Hooland, 2009 ; Boydens *et al.*, 2010). Un prototype y a été développé en vue d'adapter de manière dynamique la structure du modèle de représentation aux usages. Celui-ci s'inspire des modalités de *drag and drop* des blocs d'information qu'offre l'interface *iGoogle*. Il permet aux utilisateurs d'adapter les champs de métadonnées par défaut qu'il est possible

de personnaliser (par ajout, suppression ou restructuration). La consultation et les usages des différents champs descriptifs évoluent en effet avec la découverte de nouveaux objets (découverte de tel camion de pompiers calciné ou de tel nouveau type de débris d'armature métallique issu du *World Trade Center*) et la nécessité de les décrire. Le prototype a été développé en collaboration avec l'équipe du logiciel libre COLLECTIVEACCESS. Le suivi des fichiers de transaction associés à l'interface de recherche incluant les champs de métadonnées descriptives permet alors aux gestionnaires du système documentaire d'évaluer l'évolution des usages et d'optimiser de manière semi-automatique l'interface de recherche. Des utilisateurs « beta » ont testé le mécanisme : le caractère extrêmement évolutif et récent du *September 11th Memorial and Museum de New York* en fait un cas d'étude idéal et le module testé sera prochainement intégré en 2010 dans d'autres applications de métadonnées culturelles.

4. Conclusions

Dans de nombreuses sphères du savoir empirique, les modes d'organisation hiérarchique centralisée et sociale distribuée interagissent. Par exemple, le langage naturel : celui-ci est codifié dans nos dictionnaires par les institutions académiques, tout en évoluant avec l'usage social, lequel influence les définitions ultérieures. Ces modifications et les débats entre experts qu'elles suscitent varient dans l'espace et dans le temps selon l'héritage culturel des instances concernées. Ainsi plusieurs dictionnaires définissent-ils progressivement les nombreux termes issus des pratiques actuelles des utilisateurs d'internet : « surfer », « web », « blog »... (voir par exemple : *Petit Larousse 2010*, *Le Nouveau Petit Robert de la langue française 2010*).

Dans cet article, nous avons présenté un éclairage historique montrant que les organisations hiérarchiques centralisées et sociales distribuées coexistaient déjà au Moyen-âge. La frontière qui les sépare et les relie est mouvante et évolutive. Leur positionnement relatif jusqu'à nos jours varie en fonction de la valeur que la société attribue au savoir à un moment de son histoire. Dans l'un et l'autre cas, à des degrés divers, la connaissance empirique est intrinsèquement fragile : elle se transforme avec l'interprétation des informations qu'elle contribue à appréhender (section 2).

Que conclure de cette coexistence historique ? Nous observons, d'un côté, un effort permanent et nécessaire de l'esprit normatif pour classer et hiérarchiser. Cet effort contribue à rendre exploitables nos connaissances en tant qu'instrument épistémologique et outil d'action sur le réel. D'un autre côté, on observe une réalité sociale créative qui, en permanence, fait apparaître de

l'anarchie au regard des classifications hiérarchiques. Les technologies libèrent et multiplient potentiellement ces deux modes d'organisation du savoir qui semblent s'opposer tout en interagissant.

Face à ce phénomène, cet article a tenté de montrer l'utilité d'un modèle herméneutique et d'un développement de la théorie en matière de modélisation conceptuelle pour l'interprétation et l'amélioration opérationnelle des systèmes d'information empiriques incluant une base de données structurée. Nous en avons présenté les fondements théoriques ainsi que plusieurs exemples significatifs d'applications actuelles pour chaque type d'organisation (section 3). Si la puissance des technologies semble accroître l'anarchie induite par les modes d'organisation sociale distribuée, l'incertitude ainsi générée peut être partiellement maîtrisée par les instruments qu'offrent ces mêmes technologies.

Nous en approfondissons actuellement plusieurs axes dans le cadre de recherches en cours. Par exemple, la mise en place des stratégies de gestion proposées requiert que soit stocké dans un système documentaire le « know how » d'interprétation, de correction des anomalies et de restructuration des bases de données. La mise en place de tels systèmes devra être formalisée et, en particulier, il sera intéressant d'en examiner les aspects automatisables.

Entre la conformité au cadre opératoire de l'informatique et la question du sens, se déploient les arbitrages auxquels sont confrontés les gestionnaires d'un système d'information. L'herméneutique des systèmes d'information contribue à améliorer l'interprétation et la qualité de vastes bases de données documentaires au sein desquelles l'homogénéité des représentations formelles se heurte à l'hétérogénéité des catégories empiriques.

Remerciements

Nous remercions les lecteurs des versions préliminaires de ce texte pour leur patience et leurs commentaires critiques et constructifs. En particulier, notre reconnaissance s'adresse à Jean-Louis Besson, dont les travaux dans le domaine de l'épistémologie des statistiques ont guidé nos recherches.

Bibliographie

- Aron R., *La philosophie critique de l'histoire*, Paris, Vrin, 1969.
- Besson J.-L., éd., *La cité des chiffres ou l'illusion des statistiques*, Paris, Éditions « Autrement », 1992.
- Bourgain P., « L'édition des manuscrits », Chartier R. et Martin H.-J., éd., *Histoire de l'édition française. Tome I (q.v.)*, p. 53-94.

- Boyd R. W., « The history and historiography of information science : some reflections », *Information Processing and Management*, Vol. 32, No. 1, 1996, p. 3-17.
- Boydens I., *Informatique, normes et temps*, Bruxelles, Bruylant, 1999.
- Boydens I., « Déploiement coopératif d'un dictionnaire électronique de données administratives », *Document Numérique*, vol. 5, n°3-4, 2001, Paris, Hermès, p. 27-43.
- Boydens I., « Les bases de données sont-elles solubles dans le temps ? », *La Recherche*, novembre-décembre 2002, Paris, Sophia Publications, p. 32-34.
- Boydens I., *Qualité de l'information et e-administration : enjeux et perspectives*, in Assar S., Boughazala I., *Administration électronique : constats et perspectives*, Paris, Hermès, 2007, p. 103-120.
- Boydens I., *Documentologie*, Bruxelles, Presses de l'Université Libre de Bruxelles, 2009.
- Boydens I., Bruillard E., Caron, P.-A., Gallezot G., Schneider D. K., « Entretien », *Revue Distance et Savoirs. L'organisation et l'accès à l'information scientifique éditorialisée, vers de nouveaux agencements numériques*, coordonné par Timini I., Delamotte E., Peraya D., vol. 7, n°4, 2009, Paris, Hermès, p. 479-500.
- Boydens I., van Hooland S., « Hermeneutics applied to the quality of empirical databases », *Journal of documentation*, 2010 (à paraître ; parution en ligne, octobre-novembre 2010).
- Braudel F., *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*, Paris, Armand Colin, 1976.
- Chartier R., Martin H.-J., éd., *Histoire de l'édition française. Tome I. Le livre conquérant. Du Moyen-âge au milieu du XVI^e siècle*, Paris, Fayard-Cercle de la Librairie, 1989.
- Chartier R., Martin H.-J., éd., *Histoire de l'édition française. Tome II. Le livre triomphant 1660-1830*, Paris, Fayard-Cercle de la Librairie, 1990.
- Chartier R., « Pamphlets et gazettes », in Chartier R., Martin H.-R., éd., *Histoire de l'édition française, Tome I (q.v.)*, p. 501-526.
- Elias N., *Du temps*, Paris, Fayard, 1996.
- Elmasri R., Navathe S., *Fundamentals of Database Systems*, Reading, Addison Wesley, 2007.
- Escarpit R., *L'information et la communication. Théorie générale*, Paris, Hachette, 1991.
- Fox C., Levitin A., Redman T., « Data and data quality », in Kent A., Hall C. M., éd., *Encyclopedia of Library and Information Science*, New York, Marcel Dekker, 1996, vol. 57, suppl. 20.
- Ganasia J.-G., *Dictionnaire de l'informatique et des sciences de l'information*, Paris, Flammarion, 1998.
- Groebbens A., Hulstaert A., Ogonowski G., Vanhalst B., vom Berge K., Web 2.0., Enterprise 2.0, Government 2.0, Rapport de recherche, janvier 2009, Bruxelles, Smals.

- Hudon M., *Guide pratique pour l'élaboration d'un thésaurus documentaire*, Montréal, Éditions Asted, 2009.
- Koselleck R., *Le futur passé. Contribution à la sémantique des temps historiques*, Paris, Éditions de l'École des Hautes Études en Sciences Sociales, 1990.
- Labarre A., « Les incunables : la présentation du livre », in Chartier R. et Martin H.-J., éd., *Histoire de l'édition française, Tome II (q.v.)*, p. 228-266.
- Le Nouveau Petit Robert de la langue française 2010*, Paris, Dictionnaire Le Robert, 2010.
- Madnick S. E., Wang R.-Y., Yang W.-L., Hongwei Z., « Overview and Framework for Data and Information Quality Research », *Journal of Data and Information Quality*, Vol. 1, No. 1, 2009, p. 2-22.
- Marichal R., « La critique des textes », in Samaran C., éd. *L'histoire et ses méthodes. Encyclopédie de la Pléiade*. Paris, Gallimard, 1961, p. 1247-1360.
- Martin H. J., « La révolution de l'imprimé », in Chartier R., Martin H.-J., éd., *Histoire de l'édition française, Tome I (q.v.)*, p. 165-185. (Martin, 1989a)
- Martin H.-J., « Classements et conjonctures », in Chartier R., Martin H.-J., éd., *Histoire de l'édition française, Tome I (q.v.)*, p. 529-564. (Martin, 1989b)
- Mary A., Rouse R. H., « La naissance des index », in Chartier R. et Martin H.-J., éd., *Histoire de l'édition française, Tome I (q.v.)*, p. 95-108.
- McKenzie D. F., *La bibliographie et la sociologie des textes*, Paris, Éditions du Cercle de la Librairie, 1991.
- Moureau F., *La plume et le plomb : espaces de l'imprimé et du manuscrit au siècle des Lumières*, Paris, Presses de l'Université Paris-Sorbonne, 2006.
- Netz R., *Histoire de la censure dans l'édition*, Paris, Presses universitaires de France, 1997.
- Olson J., *Data Quality: The Accuracy Dimension*, Burlington, Elsevier, 2003.
- Parinet E., *Une histoire de l'édition à l'époque contemporaine (XIX^e-XX^e siècle)*, Paris, Éditions du Seuil, 2004.
- Petit Larousse 2010*, Paris, Éditions Larousse, 2010.
- Prost A., *Douze leçons sur l'histoire*, Paris, Seuil, 1996.
- Redman T., *Data Quality: The Field Guide*, Boston, Digital Press, 2001.
- Ribes D., Finholt T., « The Long Now of technology infrastructure : articulating tension in development », *Journal of the Association for Information Systems*, Vol. 10, 2009, Special Issue, p. 375-398.
- Roche D., « La censure », in Chartier R., Martin H.-J., éd., *Histoire de l'édition française. Tome II (q.v.)*, p. 88-98.

- Sauvy A., « Livres contrefaits et livres interdits », in Chartier R., Martin H.-J., éd(s), *Histoire de l'édition française, Tome II (q.v.)*, p. 128-146.
- Sgard J., « La multiplication des périodiques », in Chartier R., Martin H.-J., éd(s), *Histoire de l'édition française, Tome II (q.v.)*, p. 247-255.
- Valla L., *La Donation de Constantin (sur la donation de Constantin à lui faussement attribuée et mensongère)*, traduit et commenté par Glard J.-B., Paris, les Belles Lettres, 1993.
- van de Graaf M. et van Eijndhoven K., *The European Repository Landscape : Inventory Study into the Present Type and Level of OAI-Compliant Digital Activities in the EU*, Amsterdam, AUP, 2007.
- van der Vlist E., *Relax NG*, Cambridge, O'Reilly Media, 2003.
- van Hooland S., « Spectator becomes annotator: possibilities offered by user-generated metadata for image databases », Paper presented at *Immaculate Catalogues: Taxonomy, Metadata and Resource Discovery in the 21st Century*, 13-15 September 2006, University of East Anglia, UK.
- van Hooland S., Kaufman S., Bontemps Y., « Answering the call for more accountability: applying data-profiling to museum metadata », *Proceedings of the 2008 International conference on Dublin Core and metadata applications*, Berlin, 22- 26 Septembre 2008, Dublin Core Metadata Initiative, p. 93-103.
- van Hooland S., *Metadata quality in the cultural heritage sector: stakes, problems and solutions*, Thèse de doctorat, Université Libre de Bruxelles, 2009.
- Vanbinsbergen E. H. W., *La philosophie de la CDU*, Liège, Éditions du Centre de Lecture Publique de la Communauté française de Belgique, 1994.
- Vezin J., « La fabrication du manuscrit », in Chartier R. et Martin H.-J., éd(s), *Histoire de l'édition française, Tome I (q.v.)*, p. 17-52.
- Volle M., *De l'informatique*, Paris, Economica, 2006.
- Wand Y., Wang R.-Y., « Anchoring Data Quality Dimensions in Ontological Foundations », *Communications of the ACM*, novembre 1996, vol. 39, n° 11, p. 86-95.
- Wiener L. R., *Les avatars du logiciel*, Paris, Éditions Addison-Wesley France, 1994.