HAL open science

Histoire(s) de notices

Gautier Poupeau

To cite this version:

Gautier Poupeau. Histoire(s) de notices. Séminaire IST Inria : le document numérique à l'heure du web de données, Lisette Calderan, Pascale Laurent, Hélène Lowinger et Jacques Millet, Oct 2012, Carnac, France. pp.25-40. hal-00740295v1

HAL Id: hal-00740295

https://inria.hal.science/hal-00740295v1

Submitted on 9 Jul 2013 (v1), last revised 12 Jul 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Histoire(s) de notices

Gautier Poupeau

Titulaire d'un DEA en sciences de l'information de l'École nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB), Gautier Poupeau est consultant-architecte des données chez Antidot depuis juillet 2010. Après avoir géré le site web et les éditions électroniques de l'École nationale des chartes, il a été consultant chez Unilog puis Atos Origin. Il est l'auteur du blog Les petites cases. www.lespetitescases.net, gpoupeau@antidot.net

Le rôle du professionnel de l'information est de rendre accessible l'information pertinente au regard des activités d'une organisation et d'assurer la continuité de cet accès. Pour atteindre cet objectif, son activité se décompose en plusieurs phases :
- le choix des informations pertinentes en rapport avec les activités de l'organisation ;
- la modélisation de la description de l'information pour la rendre accessible dans le temps ;
- le classement et le catalogage de l'information suivant la modélisation établie ;
- le repérage de l'information suivant les demandes des utilisateurs.

Au cœur de chacune de ces étapes, il s'appuie sur le même outil : la notice, reflet des différents artefacts choisis et classés pour constituer une collection raisonnée.

La montée en puissance du numérique au sein de la documentation a peu à peu modifié le statut, la place et le rôle du professionnel de l'information et, avec eux, la notion même de notice de manière souvent implicite. Alors que le web social a constitué une phase de maturité dans l'appropriation du réseau, il semble qu'une nouvelle étape émerge avec la notion de web de données. À la différence de la précédente, qui avait trait à l'interaction des utilisateurs dans ce nouveau médium, cette dernière est certes plus technique mais semble aussi plus proche des pratiques et de l'activité des professionnels de l'information. Il nous paraît donc intéressant d'étudier cette nouvelle étape et ses conséquences sur leur travail et leur rôle en interrogeant ce qui constitue leur outil principal, la notice.

Nous nous attacherons tout d'abord à définir la notice, en particulier ce qui en fait un document, pour ensuite étudier son évolution à travers le temps, des premières étapes de l'informatisation jusqu'au web sémantique. Une histoire de la notice dont l'objectif est d'examiner comment les technologies du web sémantique permettent de retrouver l'essence de ce que celle-ci représentait dans le monde documentaire, pour mieux faire émerger l'apport spécifique des professionnels de l'information dans le web de données.

1 Le règne de la notice

Le terme de *notice* est essentiellement employé dans les bibliothèques. Or, si les traditions et le vocabulaire diffèrent d'un métier à l'autre, l'idée reste la même et nous utiliserons ici ce terme dans un sens plus générique, pour désigner tout outil ayant pour fonction de décrire un objet dans une collection suivant des règles définies par le contexte métier dans lequel se déroule l'activité de signalement : notice de bibliothèque, inventaire d'archive, fiche descriptive d'une œuvre dans un musée, etc. C'est dans cette perspective nous allons tenter de définir cet outil avant d'en tracer l'évolution dans le monde numérique.

1.1 La notice : essai de définition

Si on part du principe que la notice est un objet documentaire, il convient de l'analyser à la lumière de la définition du document donnée par Roger T. Pédauque [6]. Celui-ci définit le document comme un objet :
- dont les contours sont circonscrits par des impératifs physiques (forme) ;
- sur lequel est inscrit un ensemble d'informations porteur de sens (signe) ;
- qui permet de véhiculer une information dans un contexte social particulier (médium).

Une autre définition du document, celle de Suzanne Briet [2], reprise par Buckland [3] [4], fait résider l'aspect documentaire d'un objet dans son appartenance à une collection.

La fonction initiale de la notice, par rapport à l'objet qu'elle décrit, est de repérer son emplacement physique au sein de la collection. Pour cela, elle fournit, sous une forme circonscrite à l'origine par la fiche cartonnée, d'une part les informations de localisation de base, et d'autre part une vue synthétique de l'objet localisé, afin de s'assurer qu'il est également correctement identifié. C'est donc une deuxième fonction de la notice : fournir une description synthétique et formelle de l'objet, à travers la description de son intitulé, son créateur, ses caractéristiques physiques, sa date de création, etc.

Cette fiche prend son sens dans un contexte spécifique : celui de l'institution qui détient une collection. Il existe un contrat de lecture, lié à la fois à la forme de la notice (la fiche) et à l'institution qui l'a produite, qui permet à l'usager de comprendre la notice et de l'utiliser. Ce contrat de lecture détermine implicitement le fait que la notice est l'image de l'objet tel qu'il existe dans la collection. Cette convention permet la lecture de la notice et justifie donc son rôle de médium dans un contexte social particulier, ainsi que le définit Pédauque. On retrouve donc bien les trois dimensions qui font de la notice un document : la forme circonscrite, l'information porteuse de sens et le contexte social.

Au-delà de sa fonction individuelle, la notice se définit aussi par rapport au catalogue, c'est-à-dire à l'ensemble des notices disponibles. L'organisation des notices les unes par rapport aux autres a pour fonction de permettre de retrouver l'objet parmi l'ensemble des notices : en cela, cette organisation constitue le reflet du classement de la collection. Plusieurs organisations peuvent être proposées à l'usager : un classement alphabétique de titre, d'auteur... Mais l'organisation des notices peut aussi refléter une organisation thématique, suivant un choix raisonné qui s'appuie sur le contenu de la collection elle-même. On retrouve ici la définition du document de Briet / Buckland : la notice se définit par rapport à un ensemble plus vaste et organisé en collection, le catalogue.

Toutefois, en réalité, ce n'est pas le catalogue lui-même qui constitue une collection : il n'est qu'une image, un reflet de la collection physique pour laquelle il joue le rôle d'outil d'accès. En effet, la notice est un document particulier : elle décrit un autre objet qui, toujours selon la définition de Suzanne Briet, est lui-même un document, puisqu'il fait partie d'une collection.

Chaque notice reflète donc, par son classement au sein du catalogue, l'appartenance du document qu'elle décrit à cette collection dont la constitution provient d'un choix, d'une politique documentaire. C'est l'organisation des notices en catalogue qui rend lisible ce choix raisonné et, en cela, fonde la collection. Pour aller plus loin encore, dans la mesure où le catalogue et la notice constituent le principal, voire le seul truchement par lequel l'usager a accès à la collection (un document non décrit peut être considéré comme perdu), on peut dire que la notice est un métadocument qui fonde l'existence du document au sein de la collection parce qu'elle en est le témoin concret pour l'usager.

1.2 De la notice à la base de données informatique

L'informatisation, avec la mise en place de catalogues informatisés sous la forme de bases de données, a permis de développer de nouvelles fonctionnalités. En premier lieu, les catalogues informatisés consacrent une forme d'ubiquité de la notice : plusieurs utilisateurs peuvent la consulter en même temps, d'abord dans un même lieu sur différents terminaux, puis à distance par Internet. Comme elle est disponible sous la forme d'une seule fiche dans la base de données, il n'est plus nécessaire de la dupliquer pour refléter les différents modes d'organisation de la collection (par titre, par auteur, par thème, etc.). En lieu et place, on dispose de plusieurs champs de recherche qui permettent de trouver cette notice suivant différents critères qu'on appelle les points d'accès. On obtient ainsi des listes de résultats triés suivant le critère choisi (alphabétique, chronologique, etc.). La plus grande partie de l'information contenue dans la notice devient ainsi accessible dans le cadre d'une recherche structurée, permettant de croiser plus facilement les différents critères de recherche.

Rapidement vont se développer des référentiels communs, dans la continuité des points d'accès qui servaient à l'origine à classer les fiches pour fournir un catalogue ordonné. L'informatisation renforce le besoin de normaliser les formes, par exemple les formes des noms, dans la mesure où la recherche structurée, telle qu'elle est à l'origine proposée par les bases de données ne donne pas droit à l'erreur. Là où l'usager effectuait un feuilletage assimilable à une recherche approximative, dans un formulaire de recherche, il est nécessaire d'exprimer une requête précise et pertinente, même si les langages de requêtes offrent alors des possibilités nouvelles, mais au prix d'une technicité importante. Ainsi, en complément de la normalisation des formes, on est amené à créer de nouvelles fiches pour décrire non plus des objets documentaires, mais les concepts qui leur sont liés (des personnes physiques et morales, des lieux, des concepts, des sujets, etc.) [5]. Ce sont les notices d'autorité.

L'informatisation facilite les échanges de notices d'un catalogue à l'autre. Ces possibilités ouvrent des perspectives très intéressantes, notamment pour les bibliothèques qui conservent des objets multiples et qui, dès les années 1970, envisagent le développement du catalogage partagé grâce à l'échange de notices. Toutefois, pour échanger des informations, il est nécessaire de disposer d'une description formellement compatible afin qu'elle soit réutilisable d'un catalogue à l'autre : ce besoin d'interopérabilité implique la standardisation du modèle, que les bibliothèques mettent en place avec les formats MARC[1]. Comme les notices bibliographiques qui décrivent les documents, les notices d'autorité, qui jouent dans les catalogues de bibliothèque le rôle de référentiel, sont à leur tour normalisées et échangées. Pensée au départ dans une perspective d'échange, la standardisation laisse entrevoir la possibilité d'effectuer des recherches croisées sur plusieurs catalogues, ce qui entraîne l'apparition des catalogues collectifs pour les bibliothèques et la formalisation de la description archivistique avec ISAD(G)[2] [13].

Pourtant, malgré ces différents apports, l'informatisation des catalogues s'est inscrite dans la continuité de la définition de la notice donnée précédemment. En effet, par rapport à la notice sur fiche cartonnée, l'enregistrement dans la base de données est homothétique. L'informatisation a consisté essentiellement, dans un premier temps, à reproduire sous une forme informatisée le même paradigme documentaire : celui d'une fiche

[1] Machine Readable Cataloging

[2] International Standard Archival Description-General

qui synthétise un ensemble circonscrit d'informations, celles-ci n'étant exploitables que dans le contexte particulier de chaque notice. La forme reste donc la même : le catalogue se compose d'un ensemble de fiches informatisées qui restent juxtaposées. L'étape de modélisation nécessaire au passage à l'informatique s'est réalisée en prenant pour base non pas les entités décrites (les documents, les auteurs, etc.) mais les concepts documentaires qui régissaient la conception des catalogues, et en particulier le concept de notice. Cette approche a conduit à la pure et simple transposition de la forme documentaire de la notice dans le monde informatique.

2 Le temps des changements

À la suite de cette première phase d'informatisation, à partir de la fin des années 1980, la numérisation des documents eux-mêmes et l'augmentation des capacités des moteurs de recherche ouvrent une nouvelle étape qui va être à l'origine d'une période de déclin, voire de remise en cause du paradigme de la notice et du catalogue.

2.1 La remise en cause

La numérisation des documents, puis leur disponibilité sur le web, ouvrent de nouvelles perspectives. Si, dans un premier temps, on continuait à traiter le web de manière documentaire, à travers des annuaires de signets qui référençaient les sites et les classaient, l'irruption du modèle de Google, à partir de 1998, commence à installer l'idée que des outils, les moteurs de recherche, ont la capacité d'extraire les informations des documents et de permettre de les retrouver sans passer par cette étape d'identification, de description et de classement manuels.

Ce changement s'accompagne également d'un glissement dans le vocabulaire, puisque la notion d'indexation elle-même se transforme : on emploie le terme d'indexation pour désigner le fait d'extraire les contenus textuels des documents par les moteurs de recherche. Cette nouvelle indexation, que nous pourrions désigner par le terme de *computationnelle* par opposition à l'indexation manuelle, permet de fournir des fonctions de recherche dans les documents eux-mêmes, ce qui a pu laisser penser que l'indexation manuelle n'était plus utile.

Puisque les documents eux-mêmes sont numériques, et qu'ils sont directement accessibles via le moteur de recherche, la fonction de reflet de la collection qui était celle du catalogue et qui fournissait un intermédiaire entre le document et l'utilisateur ne semble plus utile. Le rôle de mise à disposition de l'information sous une forme numérique, qui était celui de la notice, est désormais délégué à la machine.

2.2 Une notion qui perdure

Néanmoins, la notion de notice parvient à se maintenir à travers différentes formes qui subsistent et se développent, dans le nouvel environnement que constitue le web.

Du côté des catalogues de bibliothèque, l'OPAC[3] se généralise, désormais accessible simplement en ligne à partir du site de l'institution. Les utilisateurs accèdent alors plus facilement au catalogue, ce qui en multiplie et transforme l'usage. De simple outil de repérage, il devient un outil de recherche bibliographique. Pour autant, il reste déconnecté de l'architecture du web, au sens où il constitue un silo d'information qui est invisible pour les moteurs de recherche, ceux-ci ne parcourant le réseau qu'en utilisant les liens hypertexte. Or, si chaque notice est représentée sous la forme d'une page, elle n'est pas forcément dotée d'une adresse stable permettant de poser un lien vers elle. Le formulaire de recherche, qui reste le passage obligé pour l'accès aux notices, est infranchissable par ces outils : on a utilisé l'expression de « web profond » ou « web invisible » pour désigner ce phénomène.

En effet, ce mode d'accès va à l'encontre du principe même du web, basé sur la mise en relation des différentes pages par des liens hypertexte, puisqu'il rend impossible la navigation. Cette situation s'explique par le fait que la conception de ces interfaces d'accès a été élaborée en s'appuyant sur le paradigme créé par la première vague de l'informatique, celui de la base de données, plutôt que de repartir du paradigme initial du catalogue.

C'est ainsi que, plus les habitudes de recherche d'information en ligne se normalisent autour des grands moteurs généralistes (recherche simple par mots, classement des résultats par tri de pertinence automatisé), plus les catalogues se marginalisent, détournant par leur austérité une partie de la population qu'ils visaient naturellement.

Peu à peu, l'application des principes du web aux interfaces du catalogue, en particulier de l'hypertexte, implique la création de liens et de « rebonds » entre notices à l'intérieur des OPAC et voit la réapparition d'un mode d'interaction avec le catalogue qui avait été un peu perdu avec l'informatisation : la navigation en son sein. En effet, le principe de l'hypertexte introduit par le web impose de nouveaux modes de navigation, sans structure linéaire ou hiérarchique, mais sous forme de graphe.

L'apparition de l'anglicisme « sérendipité » reflète bien cette évolution : c'est un nouveau terme en français utilisé pour désigner une fonction qui existait dans les catalogues papier et qui avait disparu dans la première phase de l'informatisation. La réflexion alors conduite concernant la navigation à l'intérieur des notices, cherchant à renouer dans la bibliothèque virtuelle avec d'anciens modèles de la bibliothèque physique, comme le fait de se promener

[3] Online Public Access Catalog

dans les rayonnages, est décisive pour faire évoluer les modes d'accès à l'information bibliographique. Ces avancées s'opèrent alors principalement à travers l'évolution fonctionnelle, ergonomique et graphique des interfaces des catalogues en ligne, sans remettre en cause les modèles de données définis précédemment.

Une autre forme de survivance de la notice, légèrement différente, apparaît avec les archives ouvertes, et le protocole OAI-PMH[4] [14] qui les accompagne et a vocation à permettre l'échange entre institutions. On retrouve ici la notion de *record* mais dans un sens différent, plus proche de celui des archives avec la notion de *records management*. Il s'agit d'une notion d'enregistrement qui intègre la dimension de trace ou de preuve : dans les archives ouvertes, le chercheur dépose son article, effectuant ainsi un acte engageant dont le *record* est le témoin. Dans ce contexte, la description de l'article qui est fournie joue davantage un rôle de trace du dépôt en tant qu'action que de reflet de la présence du document dans l'archive.

Enfin, en permettant d'associer à un même *record* plusieurs formats de description, le protocole OAI-PMH consacre un premier éclatement de la notion de notice, et on emploie désormais de manière préférentielle le terme de métadonnées, ce qui préfigure l'étape suivante.

2.3 Les prémices d'une ère nouvelle

L'emploi massif du terme *métadonnée* constitue un changement sémantique dont il ne faut pas négliger la portée. Ce terme acte le fait qu'on ne s'intéresse plus uniquement à la notice dans sa globalité, mais aux différentes portions d'information qu'elle contient à travers la notion de données.

Les raisons de ce glissement sémantique sont multiples. Les métadonnées sont définies notamment par le fait qu'elles ne sont pas forcément extérieures à l'objet qu'elles décrivent, comme l'était la fiche de catalogue, mais qu'elles peuvent être embarquées au sein même du document qu'elles identifient, comme c'est le cas pour certains documents numériques et en particulier les pages web. En outre, elles élargissent le champ du travail de signalement à d'autres informations que celles qui décrivent le contenu intellectuel du document, désormais connues sous le nom de *métadonnées descriptives*.

Ainsi, les métadonnées peuvent être :
- techniques lorsqu'elles renseignent en détail la forme du document, notamment en vue de sa conservation ;
- administratives lorsqu'il s'agit d'informer sur les droits d'accès, les modalités de gestion de la collection ou l'historique de l'objet ;
- de structure lorsqu'elles détaillent l'organisation interne des différents éléments qui composent un objet complexe.

Enfin, un formalisme de structuration est associé à cette notion nouvelle : le langage XML[5] [19]. En s'imposant de façon décisive, le XML contribue à éloigner progressivement le travail de description de la forme littéraire qu'il pouvait avoir auparavant. La normalisation croissante des formes, qui va de pair avec l'informatisation, est encore renforcée par la structuration en XML qui impose un formalisme plus rigide pour autoriser les échanges de données et leur exploitation technique. On l'observe, par exemple, pour les inventaires d'archives, dont la structure était à l'origine très littéraire et qui, alors que l'EAD[6] [11] se généralise, tendent de plus en plus à adopter la forme d'un ensemble de fiches organisées hiérarchiquement et conçues comme un outil de recherche.

Toutefois, l'emploi d'un néologisme pour désigner des concepts qui s'inscrivent pourtant clairement dans la continuité de l'activité de signalement reflète aussi un certain mal-être de la profession. En effet, le terme de métadonnées apparaît comme un moyen, pour les professionnels de l'information, de continuer à exister et de s'approprier le web.

Ainsi, ces évolutions débouchent sur une prise de conscience de la problématique de modélisation non pas des notices, mais des données qu'elles contiennent et des objets qu'elles décrivent. Au-delà du questionnement global issu de la place croissante qu'occupent le numérique puis le web dans les métiers de l'information, cette prise de conscience est également suscitée par un certain constat d'insatisfaction envers les services rendus par l'indexation « computationnelle » sous la forme que nous avons décrite plus haut. On s'aperçoit des limites du traitement opéré par les machines, mais aussi de celles des interfaces graphiques offertes aux utilisateurs qui s'appuient sur les données disponibles ; et il apparaît que, pour améliorer l'efficacité de ces traitements et leur valorisation, il est nécessaire de passer par une analyse plus fine de l'information.

Cette réflexion va alors rencontrer un effort de modélisation entamé dès le milieu des années 1990 en se détachant de l'activité de signalement elle-même pour revenir à une analyse de l'entité décrite, et conduisant à élaborer différents modèles conceptuels, qui incluent une réflexion sur les objets décrits et leurs relations :
- dès 1996, le CIDOC-CRM[7] [9], qui se donne pour objectif de fournir un cadre conceptuel de référence pour décrire toutes les interactions d'un objet avec d'autres entités dans un modèle orienté objet ;
- les FRBR[8] [12], publiés en 1998, qui définissent trois groupes d'entités pour décrire les documents, les personnes physiques et morales et les différents concepts qui peuvent être sujets des œuvres ;

---

[4] Open Archives Initiative-Protocol for Metadata Harvesting

[5] eXtensible Markup Language

[6] Encoded Archival Description

[7] Committee on Documentation - Conceptual Reference Model

[8] Functional Requirements for Bibliographic Records

- le Dublin Core qui définit en 2007 son modèle abstrait (DCAM[9] [10]).

Confrontés aux problématiques émergentes dans l'univers du web, ces efforts de modélisation démontrent toute leur importance et suscitent un regain d'intérêt. L'architecture du web fournit enfin les conditions technologiques pour mettre en place un milieu associé, c'est-à-dire mettre en relation, partager, distribuer l'information. Face aux deux modèles qui ont immédiatement succédé à l'informatisation, celui où le rôle des professionnels de l'information était délégué à des machines, et celui du *record* qui consacrait le paradigme traditionnel de la notice sous couvert d'une justification sociale autant que technique, la principale limite de ces modèles est l'absence de mise en relation. Les principes du web font exploser la notice et remettent en cause la vision qu'on peut en avoir en tant que métadocument.

3 Données structurées et web de données

3.1 La réconciliation

La période précédente se clôt donc sur une situation où la transformation de la notice et le recul sur les moteurs de recherche basés uniquement sur une indexation plein-texte ont provoqué la prise de conscience des limitations du paradigme apporté par la première génération de l'informatisation.

Ces limitations, ce sont d'abord celles de l'interopérabilité. La forme documentaire contrainte de la notice et le fait que l'architecture du web n'a pas été respectée et que les notices se trouvent dès lors dans le web profond obligent à utiliser des stratégies détournées pour mettre les données en relation. Le protocole OAI-PMH recherche le plus petit dénominateur commun entre des données très diverses, appauvrissant le service rendu à l'utilisateur. Les sites du web dit « 2.0 » mettent en place des API[10] complexes, mais qui sont propriétaires et ne fonctionnent que dans un contexte donné. On fait appel aux utilisateurs, avec les folksonomies, pour fournir aux machines le matériau textuel dont elles ont besoin pour créer ces relations mais, même ainsi, l'absence de structuration et de normalisation des formes débouche sur des impasses.

Finalement, l'application du principe des facettes [7] aux moteurs de recherche, en proposant des filtres sous la forme de classifications organisées, même si celles-ci sont parfois sommaires, consacre le besoin de fournir à la machine une forme de hiérarchisation pour qu'elle puisse rendre un service efficace. C'est le retour de la donnée structurée.

Cette étape marque l'apparition d'un nouvel équilibre entre la structuration de l'information par des humains, en utilisant les règles que les professionnels de l'information ont toujours su maîtriser, et les performances de la machine pour traiter ces informations.

En 2008, le modèle OAI-ORE[11] [15], qui s'inscrit dans la continuité des travaux de l'Open Archive Initiative, l'organisme qui normalise le protocole OAI-PMH, constitue un premier effort de la communauté des professionnels de l'information pour s'approprier non seulement les standards du web, mais surtout ses principes fondamentaux et son architecture. Puis, en 2010, s'appuyant sur les FRBR, les bibliothèques reviennent à la démarche de signalement et actent les nouveaux modèles dans de nouvelles règles de catalogage, les RDA[12] [16].

Cette réconciliation témoigne de la maturité nécessaire, à la fois du côté des professionnels de l'information et de ceux du web, pour que les pratiques de ces deux communautés se rencontrent et convergent. Les technologies du web sémantique, parce qu'elles impliquent un traitement de la donnée structurée pleinement compatible avec l'architecture du web, vont être la cheville ouvrière de ce nouvel équilibre.

3.2 Un nouveau modèle

L'appropriation des principes de l'architecture du web par les professionnels de l'information voit l'aboutissement de l'évolution de la notion de notice. L'inventaire, le catalogue ou le système documentaire ne sont plus simplement des systèmes locaux offrant une porte d'entrée sur le web via une interface graphique, mais ils s'intègrent dans une partie d'un ensemble plus vaste qu'est le web.

À partir de ce moment, il est nécessaire de gérer au niveau le plus fin l'interopérabilité des données. Il ne s'agit plus de relier des documents entre eux, mais des entités du monde réel sous une forme abstraite dans le web. En effet, le point essentiel de la théorie de l'architecture du web réside dans la séparation entre un objet abstrait et sa représentation sous la forme d'un flux documentaire. La notion fondatrice de ressource distingue ainsi l'identifiant de la ressource et sa représentation sous forme d'un fichier informatique, imposant de fait une séparation complète entre le support et le contenu. On ne réfléchit pas en termes de documents, mais en termes d'objets.

Le web fournit un protocole qui ne présage pas de la forme ou de la nature de la représentation ; il s'agit uniquement d'un langage de communication. Sur le web, ce n'est donc pas l'inscription d'une information sur un

[9] Dublin Core Abstract Model

[10] Application Programming Interface

[11] Open Archives Initiative – Object Reuse and Exchange

[12] Resource Description and Access

support qui fonde l'existence d'une entité, mais le fait de l'identifier. Ainsi, dans l'architecture du web [8], la ressource peut être n'importe quel objet du monde réel, transposé dans le monde numérique par le fait d'être identifié par une URI[13]. On ne construit pas une notice, on décrit un objet lui-même. Le métadocument qu'est la notice n'acte plus l'existence de l'entité qu'elle décrit : c'est l'identifiant, l'URI, qui représente le signifiant de l'objet dans le web.

Au-delà de cette dimension abstraite de modélisation, la technologie disponible permet de placer les objets en question dans un espace commun d'interopérabilité afin que tout le monde puisse y faire référence. Les objets peuvent être partagés sur le réseau. On dépasse donc le modèle où il était nécessaire de produire des notices compatibles pour les échanger, pour passer à un modèle où toute l'information est disponible en réseau dans un espace commun, global, décentralisé.

Ainsi, la notice laisse la place à un graphe d'entités abstraites reliées entre elles. Le web constituant un milieu associé, toutes les entités décrites sont au même niveau, reliées par des liens typés : l'ancienne distinction existant entre la notice bibliographique et la notice d'autorité tend à disparaître mais, encore au-delà, ce ne sont pas seulement les objets mais les activités, les organisations, les compétences qui sont mis en relation et peuvent désormais converger.

3.3 Le retour du flux documentaire ?

Le nouveau modèle que nous venons de décrire, qui est celui du web de données, se voit toutefois ajouter un niveau supplémentaire d'abstraction qui est, lui, de nature documentaire. D'un point de vue pratique, pour que les machines puissent exploiter les objets exposés et reliés dans le web, il faut que l'information soit sérialisée, c'est-à-dire inscrite sur un support, afin d'être mise à leur disposition. Ainsi, une machine récupère la description d'une entité sous la forme d'un flux informatique : c'est la notion de *représentation* dans l'architecture du web.

Selon Dan Brickley, « le *Linked Data* est l'utilisation des standards du web pour échanger des documents qui encodent des données structurées[14] ». Le web de données sépare la ressource abstraite, qui ne dispose d'aucune représentation, de la ressource documentaire qui en a une.

De ce point de vue, la représentation n'est ni plus ni moins qu'une notice descriptive de l'entité selon un formalisme informatique (le RDF[15] [17]) pensé par les humains à destination des machines. Ce flux contient des informations sur les entités de toute nature dont il est la représentation documentaire et suit une structure précise :
- l'entité elle-même est le sujet ;
- le verbe ou prédicat indique sa relation avec l'objet ;
- ce dernier peut être lui-même une entité à laquelle le sujet est relié ou une chaîne de caractères.

Il s'agit de donner sous la forme d'un flux une description synthétique et formelle d'un objet : nous retrouvons ici notre définition initiale de la notice.

Pour autant, peut-on assimiler la ressource documentaire dans le web de données à un document au sens où nous l'avons défini s'agissant de la notice ? Si on retrouve bien la forme du document, avec un flux fini d'informations, et la notion de signe puisqu'il contient des informations qui ont du sens, la question se pose quant au médium et donc à la dimension de contexte social.

En effet, si la notice disposait d'un contexte social induit par sa production dans le contexte d'une institution et d'un catalogue, le flux documentaire dans le web de données est confronté à la nécessité, pour permettre le traitement par les machines, d'expliciter formellement les informations de contexte. Ainsi, la représentation documentaire de la ressource est elle-même une ressource sur laquelle on peut faire des assertions pour gérer notamment la temporalité (un même flux documentaire peut contenir des informations différentes suivant le moment où il est consulté) et l'autorité (l'identification de l'institution productrice de l'information). Le web de données n'a pas été pensé par rapport à un contexte social parce que, par définition, il ne présage pas des usages des données ; mais le fait de fournir ces informations de contexte est nécessaire pour parvenir à établir la confiance et la preuve qui sont essentielles pour exploiter ces données.

Si ces notions de confiance et de preuve ont bien été définies dès l'origine, ainsi qu'on peut le voir dans le schéma du *Semantic Web Stack*, la feuille de route [1] établie dès 1998 par Tim Berners-Lee prévoyait de ne s'intéresser à ces questions qu'à l'issue de la mise en place des briques fondamentales. Pragmatiquement, on a donc commencé par créer les conditions techniques de la réalisation de ce nouveau paradigme, pour le théoriser dans un deuxième temps. Les travaux sont en cours au W3C[16] pour apporter les moyens de décrire ces informations de confiance et de preuve dans le web de données, à commencer par les informations de provenance [18] dont la normalisation est assurée par un groupe de travail dédié.

Au final, dans le web de données, ce n'est pas cet aspect de contexte social qui prime. La représentation documentaire y est une conséquence de l'existence de l'objet, et non ce qui le fonde ; elle a un caractère accessoire par rapport à la description abstraite des objets eux-mêmes et leur mise en relation sous forme de

[13] Uniform Resource Identifier

[14] http://lists.w3.org/Archives/Public/public-lod/2012Mar/0255.html

[15] Resource Description Framework

[16] World Wide Web Consortium

graphe. Finalement, le web de données permet de dépasser les contraintes documentaires pour mettre les informations en relation à un niveau plus fin et créer un milieu associé de métadonnées et, à travers elles, des activités des organisations.

En guise de conclusion...

De la même manière que l'ensemble des notices, à travers le catalogue, constituait un reflet de la collection, l'idée du web sémantique est de constituer un reflet du monde réel, en fournissant aux machines une vision abstraite du monde qu'elles soient en capacité de traiter. Dans ce contexte où la notice semble avoir disparu en tant que forme documentaire, la notion de représentation, dans l'architecture du web, réintroduit celle de flux documentaire sans lui donner la même importance fondatrice. Il s'agit dès lors de dépasser le cadre documentaire de la notice, non pas parce qu'il n'est plus utile de légitimer l'objet dans la collection, mais parce que cette démarche de légitimation n'a plus une forme documentaire.

Le rôle des professionnels de l'information doit évoluer à l'image de celui de la notion de notice. Avec la multiplication de l'information au sein des organisations, les différentes facettes de leur activité (choisir, modéliser, décrire, classer, repérer) sont plus que jamais nécessaires et leur expérience est un apport indéniable à cette évolution des systèmes d'information dans laquelle les données sont placées au centre de la réflexion. Les technologies du web sémantique offrent non seulement un cadre technique, mais aussi un cadre conceptuel dans lequel les professionnels de l'information doivent trouver une place naturelle. Dans un modèle abstrait où ce sont les objets eux-mêmes, et pas seulement leurs descriptions, qui sont mis en relation sous forme de graphe dans le monde numérique, ils doivent dépasser le cadre strictement documentaire de leur activité pour prendre en compte tous les objets et toutes les données des organisations, dans une modélisation globale qui permettra d'atteindre un niveau d'interopérabilité inédit.

Références bibliographiques

[1] Tim BERNERS-LEE. *An attempt to give a high-level plan of the architecture of the Semantic WWW*. Septembre 1998. http://www.w3.org/DesignIssues/Semantic.html

[2] Suzanne BRIET. *Qu'est-ce que la documentation ?* Paris : Éditions documentaires, industrielles et techniques, 1951. 48 p.

[3] Michael K. BUCKLAND. « What is a “document”? ». *Journal of the American Society of Information Science*, September 1997, vol. 48, n° 9, p. 804-809. http://people.ischool.berkeley.edu/~buckland/whatdoc.html

[4] Michael K. BUCKLAND. « What is a “digital document”? ». *Document numérique*, 1998, vol. 2, n° 2, p. 221-230. http://people.ischool.berkeley.edu/~buckland/digdoc.html

[5] Etienne HUSTACHE, Annie MEYER, Christine STOTZENBACH. « Salam LAMECH ! Liste d’autorités de matières encyclopédique, collective et hiérarchisée ». *Bulletin des bibliothèques de France*, 1986, t. 31, n° 3. http://bbf.enssib.fr/consulter/bbf-1986-03-0248-005

[6] Roger T. PEDAUQUE. *Le document à la lumière du numérique. Forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité.* Caen : C&F éditions, 2006. 218 p.

[7] Shiyali Ramamrita RANGANATHAN, *Colon Classification.* Madras : Madras library association, 1952

Référence des normes

[8] *Architecture of the World Wide Web, Volume one*. W3C Recommendation, 15 December 2004. http://www.w3.org/TR/webarch

[9] ISO 21127:2006 *Information and documentation – A reference ontology for the interchange of cultural heritage information*. [Version 5.0.4 disponible en ligne http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf]

[10] Andy POWELL, Mikael NILSSON, Ambjörn NAEVE, Pete JOHNSTON, Thomas BAKER. *DCMI Abstract Model*. 4 June 2007. http://dublincore.org/documents/abstract-model

[11] SAA ENCODED ARCHIVAL DESCRIPTION WORKING GROUP. *Encoded Archival Description Tag Library. Version 2002*. http://www.loc.gov/ead/tglib/index.html

[12] GROUPE DE TRAVAIL IFLA SUR LES SPECIFICATIONS FONCTIONNELLES DES NOTICES BIBLIOGRAPHIQUES. *Spécifications fonctionnelles des Notices bibliographiques.* Bibliothèque nationale de France, 2001. http://www.bnf.fr/documents/frbr_rapport_final.pdf

[13] CPBS-SUB-COMMITTEE ON DESCRIPTIVE STANDARDS. *ISAD(G): General International Standard Archival Description – Second edition.* 1 September 2011. http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html

[14] Carl LAGOZE, Herbert VAN DE SOMPEL, Michael NELSON, Simeon WARNER. *The Open Archives Initiative Protocol for Metadata Harvesting. Version 2.0*. 14 juin 2002. http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

[15] Carl LAGOZE, Herbert VAN DE SOMPEL, Pete JOHNSTON, Michael NELSON, Simeon WARNER. *ORE Spectification – Abstract Data Model*. 17 octobre 2008. http://www.openarchives.org/ore/1.0/datamodel.html

[16] THE JOINT STEERING COMMITTEE FOR DEVELOPMENT OF RDA. *RDA: Resource Description and Access*. http://www.rdatoolkit.org

[17] *Resource Description Framework (RDF): Concepts and Abstract Syntax*. W3C Recommendation, 10 February 2004. http://www.w3.org/TR/rdf-concepts

[18] *PROV-O: The PROV Ontology*. W3C Working Draft, 24 July 2012. http://www.w3.org/TR/2012/WD-prov-o-20120724

[19] *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C recommendation, 26 November 2008. http://www.w3.org/TR/2008/REC-xml-20081126