

## A la recherche du temps perdu, retour vers le futur : CBU 2.0

### Gildas Illien

Département de l'Information bibliographique et numérique, Bibliothèque nationale de France, Paris, France.  
gildas.illien[at]bnf.fr

### Françoise Bourdon

Département de l'Information bibliographique et numérique, Bibliothèque nationale de France, Paris, France.  
francoise.bourdon[at]bnf.fr



Copyright © 2014 by Gildas Illien & Françoise Bourdon. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

---

### Résumé:

*Fruit d'une conversation entre une pionnière du Contrôle bibliographique universel (CBU) et un conservateur de la génération internet qui dirige l'agence bibliographique nationale française, ce texte célèbre à sa manière le 40<sup>ème</sup> anniversaire de la fondation du CBU. Nous avons voulu relire ensemble les déclarations aux origines du CBU et partager nos questionnements et nos intuitions sur son avenir.*

*Le CBU reposait sur quelques idées ambitieuses mais simples : le savoir qui fait la richesse et le patrimoine d'une nation s'exprime dans des publications, au départ imprimées sur support; chaque Etat doit se doter d'une agence chargée de collecter et de décrire les collections d'intérêt national ; cette description doit se faire dans le respect de normes bibliographiques internationales, dont la principale fonction est de faciliter l'échange de notices standardisées ; ces échanges visent à répartir les tâches à travers le monde et à rendre possible l'accès pour chacun, via les bibliographies et les catalogues, à un vaste ensemble de ressources encyclopédiques de portée universelle.*

*L'idéal fondateur du CBU était à mi-chemin de celui de la Bibliothèque d'Alexandrie et de celui de Google. Il embrassait à la fois le rêve antique d'une maîtrise de tous les savoirs du monde et la construction d'un accès global et distribué à l'information. Dans cette vision, antérieure à la propagation de l'internet, l'idée qu'un usager pût accéder par lui-même à la somme des connaissances était inconcevable : les bibliothécaires étaient encore perçus comme des acteurs incontournables de la diffusion des savoirs. Et le réseau mondial des bibliothèques imaginé par le CBU préfigurait en quelque sorte un « internet des bibliothèques ». Que faut-il retenir de cet héritage? Le CBU est-il dépassé? Selon nous, l'histoire du CBU contient dans le récit de ses*

*propres mutations les clés d'une possible refondation qui, sans qu'on s'en rende forcément compte, est déjà en marche.*

**Mots-clés:** contrôle bibliographique universel, agences bibliographiques nationales, dépôt légal, web sémantique, coopération internationale.

---

## INTRODUCTION : AUX ORIGINES DU CBU

En 1974, lorsque Dorothy Anderson publie le fascicule « *Universal Bibliographic Control: a Long Term Policy, a Plan for Action*<sup>i</sup> », le programme de l'IFLA sur le contrôle bibliographique universel (CBU) est en réalité déjà engagé depuis plusieurs années. Cette « politique de long terme » assortie d'un « plan d'action » était entrée en gestation dès 1961, quand les premières pierres de l'édifice furent posées à Paris, par la même Dorothy Anderson, lors de la première Conférence internationale sur les principes de catalogage<sup>ii</sup>. Ce qui marque l'année 1974, c'est la formalisation, dans un texte qui s'apparente à un manifeste, des principes fondamentaux qui définissent le CBU et de recommandations qui s'adressent aux bibliothécaires mais aussi aux gouvernements. Ce texte, qui synthétise les enjeux et les principes d'une collaboration internationale nécessitant la coordination de politiques et l'allocation de ressources publiques, fixe les priorités à respecter pour mettre en œuvre cet ambitieux programme. Le CBU des origines repose sur quelques idées généralement bien connues des bibliothécaires, que nous résumons dans cette introduction afin de planter le décor mais aussi replacer sa fondation dans une époque où les technologies comme les pratiques culturelles étaient différentes de celles d'aujourd'hui.

Au début des années 70, la notion de *publication* est centrale dans la perception culturelle qui préside à l'invention du CBU. Parce qu'elles sont issues d'un processus d'édition codifié et organisé, les publications sont considérées comme l'essence même de la production culturelle qu'une nation se doit de conserver à l'attention de ses contemporains et des générations futures. A cette époque où l'internet n'existe pas, publier quelque chose implique en effet d'en passer par les circuits de sélection et de validation établis pour l'édition de livres ou d'articles de périodiques, qui en garantissent un certain niveau de qualité. L'expression publique d'idées originales constitutives d'une œuvre et reconnues comme telles passe ainsi par la fabrication d'un objet imprimé : un livre, un journal, une revue. Si des collections « spécialisées » (muséales, archivistiques, iconographiques...) sont par ailleurs conservées depuis longtemps, la majorité des bibliothèques s'intéressent encore principalement aux objets imprimés. C'est l'essentiel de ce qu'elles collectent et de ce qu'elles cataloguent. Aux commencements, la discussion sur l'échange bibliographique entre Etats se concentre ainsi sur les publications imprimées.

Le CBU préconise que chaque Etat se dote d'une organisation dédiée à la collecte et au traitement des publications, une « agence bibliographique nationale ». Cette agence, généralement hébergée par la bibliothèque nationale, est chargée de collecter systématiquement (par dépôt légal, ou d'autres moyens lorsque ce dispositif législatif n'existe pas) et de décrire, dans une « bibliographie nationale », les publications produites sur son territoire ou d'intérêt national. On note que le principe qui préside à l'organisation imaginée est celui de l'Etat-nation : chaque Etat, supposé doté d'une certaine unité territoriale, culturelle et linguistique, est conçu comme le détenteur d'un sous-ensemble du savoir universel réparti entre les nations. La notion de bibliographie nationale incarne ainsi une certitude de l'époque : la somme des bibliographies nationales permettra à terme de

constituer une bibliographie universelle, décrivant toutes les publications issues de l'invention humaine. La motivation du CBU est également économique : si chaque Etat décrit sa production nationale d'une manière intelligible et réutilisable par les autres, tous réaliseront une économie. L'effort consenti par chacun pour se plier aux exigences des normes permettra en retour de bénéficier de la production des autres. L'histoire politique, économique, voire diplomatique des grandes bibliothèques nationales occidentales est bien la toile de fond culturelle qui a influencé l'invention du modèle du CBU dans les années 70.

La modernité du CBU réside surtout dans les recommandations qu'il formule en matière de normalisation de la description. Il préconise en effet que le catalogage se fasse dans le respect de normes internationales établies par consensus et dans le cadre de l'IFLA. Les normes sont omniprésentes dans le texte de Dorothy Anderson : une annexe entière, « *The requirements for UBC : international standards for the bibliographic record* » leur est consacrée. La tradition et la pratique de la normalisation descriptive dans les bibliothèques étaient naturellement antérieures aux années 70. Mais le CBU est venu consolider et étendre à grande échelle une culture professionnelle de la norme et de l'échange de notices, d'abord sur fiches cartonnées, puis par enregistrements informatiques. L'informatisation progressive des bibliothèques a conduit à travailler conjointement non seulement à la standardisation des pratiques, codes et instructions de catalogage, mais aussi à la conception et à la maintenance de formats informatiques permettant d'encoder l'information bibliographique pour qu'elle soit lisible et exploitable par des machines. L'aventure des formats MARC est, de ce point de vue, indissociable du mouvement de normalisation porté par le CBU et de ce qu'on n'appelait pas encore l'interopérabilité.

Après des heures glorieuses et des péripéties dignes de la Guerre Froide, la gouvernance du CBU a fait long feu. Les structures qui avaient été mises en place en 1974 pour le coordonner ont aujourd'hui disparu. Le Bureau du contrôle bibliographique universel (*Office for Universal Bibliographic Control*) devenu ensuite le Bureau du contrôle bibliographique universel et MARC international (*Universal Bibliographic Control and International MARC*) a été fermé en 2003. La coordination de la normalisation bibliographique a ensuite été confiée à d'autres structures de l'IFLA, l'*IFLA-CDNL Alliance for Bibliographic Standards* (ICAB), puis l'*IFLA-CDNL Alliance for Digital strategies*, dont l'activité a pris fin à son tour en 2011. En 2012, le Bureau professionnel de l'IFLA a adopté, sur proposition des sections Bibliographie et Catalogage, une déclaration<sup>iii</sup> réaffirmant la pertinence des principes du CBU, mais sans étayer cette position de préconisations opérationnelles. Cette déclaration réaffirmait la validité des fondements du CBU tout en prenant acte de la révolution numérique mais aussi de préoccupations légitimes d'envisager les échanges bibliographiques dans un dispositif qui ne soit plus centré sur une vision au demeurant très occidentale des choses. Il n'y a donc plus d'instance de gouvernance du dispositif, ce qui, dans la vie des organisations internationales, est généralement le signe d'un enterrement de première classe.

Le CBU a-t-il été vaincu par K.O par la déferlante numérique ? Le web a-t-il tué les catalogues et les bibliographies ? Probablement, mais il ne faudrait pas aller trop vite en besogne et réserver le même sort aux données structurées et de grande valeur que recèlent ces outils, ni aux compétences qui ont permis de les produire. Cette communication n'est pas une oraison funèbre. En prenant un peu de recul, et en acceptant de regarder la réalité des usages hors des cadres institués par les bibliothèques, il nous est apparu que, de toute évidence, l'histoire du CBU se poursuivait, mais sous d'autres noms, dans d'autres cadres, et avec d'autres acteurs que les seules bibliothèques. Grâce à la dynamique du CBU, les bibliothécaires ont peut-être été les visionnaires de quelque chose – le web – qui s'est

finalement construit sans eux, faute d'avoir su penser assez loin, assez vite, et surtout hors de leur communauté. L'enjeu, pour eux, est de retrouver aujourd'hui leur place et de promouvoir leur valeur ajoutée dans un environnement qu'ils ne peuvent plus prétendre contrôler. Nous allons essayer de comprendre comment les bibliothèques, en tant que communauté internationale, en sont arrivées là, et par quels leviers elles pourraient se relever.

Cette communication s'organise en trois parties : la première s'intéresse aux évolutions du périmètre et des formes du savoir qui sont l'objet de la description bibliographique ; la seconde porte sur les modalités de description de ces ressources et de leur normalisation aux fins d'interopérabilité ; la troisième invite à imaginer de nouvelles formes de coopération internationale dans le prolongement des idéaux du CBU.

## **1 DES PUBLICATIONS AUX RESSOURCES, DE LA NATION AU RESEAU : L'EVOLUTION RADICALE DES FORMES ET DU CHAMP DES SAVOIRS**

Dans cette première partie, nous allons essayer de prendre la mesure des changements qui ont profondément modifié les contours et les formes de la « matière première » du CBU, la production culturelle et scientifique que les agences bibliographiques nationales ont vocation à collecter puis à décrire. Nous nous concentrerons sur le concept de publication telle que le CBU l'abordait en 1974 : son lien à un support imprimé, à un circuit d'édition et à et à une culture nationale.

### **Du support à la ressource : comment l'information s'est abstraite des objets**

Comme on l'a vu, dans les années 70, on entendait encore par « publications » essentiellement les livres et les périodiques imprimés. Les deux premières normes internationales de description bibliographique, publiées en 1971, sont d'ailleurs l'ISBD (M-Monographies) et l'ISBD (S-Publications en série). Très vite, on constate cependant que l'imprimé n'est plus qu'un des moyens de transmission documentaire. La notion de « non livres » fait son apparition. Elle est sanctionnée en 1977 par la publication de l'ISBD (NBM-Non Book Material) et s'accompagne de la notion de « nouveaux supports » : les microformes, les disques, les bandes magnétiques, les diapositives, les documents audiovisuels, et les fichiers de données informatiques. En 1986, les documents électroniques sur support sont suffisamment répandus pour que l'on codifie leur description bibliographique qui aboutira à la publication de l'ISBD (CF-Computer Files) en 1990. L'histoire des normes bibliographiques encadrées par le CBU a ainsi suivi toute la chronologie de l'invention des supports jusqu'à l'apparition du multimédia interactif, des technologies optiques, des techniques de numérisation et, finalement, l'internet. En 1997, apparaît un changement conceptuel fondamental dans l'approche de la description normalisée : dans la publication de l'ISBD (ER-Electronic Resources) le mot « ressource » supplante le mot « document ».

Si l'on prend en compte le fait que la normalisation est toujours un peu en retard par rapport à la réalité qu'elle tente de maîtriser, l'évolution de la terminologie utilisée dans les documents normatifs témoigne du bouleversement fulgurant qui a touché la production de l'information. En une quinzaine d'années, l'on est passé de la notion de « publications sur support », à la notion de « ressource immatérielle ». En d'autres termes, en moins de quarante ans, la représentation du savoir ou de l'information s'est non seulement émancipée de l'imprimé pour s'étendre à d'autres supports, elle s'est aussi littéralement abstraite de l'objet

qui la contenait. Aujourd'hui une « ressource » est une entité, matérielle ou immatérielle, ayant un contenu intellectuel et/ou artistique, qui peut comprendre du texte, de la musique, des images fixes et animées, des représentations graphiques, des cartes, des enregistrements sonores et vidéo, des données ou des programmes électroniques. Les formes de publication ne s'excluent pas les unes les autres mais au contraire coexistent et s'accumulent.

### **Production et granularité des ressources : d'infinies combinaisons**

Pour les bibliothèques, la diversification des types de documents a eu un impact considérable sur leur mode de production, de recensement, de diffusion et de conservation. La grande variété de l'offre éditoriale va aujourd'hui des livres imprimés traditionnels qui sont en général des objets avec un contenu figé, aux livres numériques multimédia, consultables à distance dans des formats variés dont chacun peut supporter un ou plusieurs systèmes de gestion des droits numériques (DRM). Alors qu'une ressource sur support doit être reproduite en nombre pour être mise à disposition de son public, une ressource numérique en ligne, même si le fichier est unique, est accessible à distance à plusieurs utilisateurs simultanément. La notion de « tirage » en  $n$  exemplaires disparaît avec la publication en ligne. Différentes techniques de mises à jour du contenu des ressources coexistent aujourd'hui. La nouvelle édition reste une technique « traditionnelle » pour les ressources imprimées. La publication imprimée avec mise à jour en ligne est aussi courante. Si la publication imprimée à feuillets mobiles se fait rare, le nombre de publications intégratrices<sup>iv</sup> explose sous forme numérique : les sites web en sont le meilleur exemple.

Le même contenu informationnel peut être ainsi diffusé concurremment sous de multiples formes, chacune de ces formes répondant à un besoin et à un usage différent. Par exemple, l'éditeur d'un livre numérique peut le diffuser chapitre par chapitre, alors que le livre imprimé homothétique restera diffusé *in extenso*. Les deux formes de publication coexistent : un guide de voyage imprimé sur papier glacé trouvera un public, un chapitre de ce même guide consacré au logement des voyageurs peut être commercialisé en ligne pour être facilement mis à jour et satisfaire le même public, ou un autre. Inversement, des ensembles de ressources numériques, monographies ou périodiques, sont désormais regroupés en bouquets thématiques par des éditeurs et diffusés sous cette forme à des conditions différentes de la vente à l'unité. Et une même ressource peut appartenir simultanément ou successivement à plusieurs bouquets. Enfin, un document numérique, qu'il soit né numérique ou qu'il ait été numérisé, peut être re-matérialisé à la demande d'un usager sur le modèle de l'impression à la demande... La présentation du contenu informationnel peut donc se décliner dans une infinité de combinaisons qui posent des problèmes de description de plus en plus complexes : à quel niveau de granularité faut-il qualifier l'information ?

### **Production nationale : les limites du dépôt légal**

Le principe de répartition des tâches entre agences bibliographiques nationales institué par le CBU repose, on l'a vu, sur un partage du « monde bibliographique » entre nations. Cette idée semble aller de soi à l'époque, à telle enseigne que dans les textes fondateurs du CBU de 1974, on ne trouve pas de définition de la production nationale. On déduit des recommandations, et un peu par défaut, que c'est celle qui est collectée au titre du dépôt légal et qui est décrite dans la bibliographie nationale officielle de chaque Etat.

Tous les Etats sont encouragés à se doter d'une bibliothèque nationale et d'une législation de dépôt. Les Etats sont libres de choisir l'organisation la plus adaptée à leur contexte, et le partage des responsabilités avec d'autres opérateurs culturels nationaux est encouragé. La législation doit également préciser quels sont les types de documents collectés, qui doit faire le dépôt et en combien d'exemplaires. Par exemple en France, où le code du Patrimoine conditionne étroitement le dépôt légal à des critères territoriaux, ce sont principalement les éditeurs et les imprimeurs dont le siège social se trouve dans le pays qui sont soumis au dépôt. La législation doit évoluer au fur et à mesure que le contexte national l'exige. La loi française sur le dépôt légal, l'une des plus anciennes (1537), est ainsi en constante évolution pour prendre en compte les différents types de ressources à mesure de leur apparition. En voici quelques jalons : 1925 les photographies, 1938 les phonogrammes, 1975 les vidéogrammes, 1992 les documents audiovisuels et l'édition électronique sur support (progiciels, etc.), jusqu'à 2006 où a été introduit le dépôt légal des « *signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique* » (c'est-à-dire l'internet).<sup>v</sup>

Pour autant, la production d'un pays correspond de moins en moins à la production éditoriale réalisée sur son territoire. Le dépôt est de fait souvent complété d'acquisitions étrangères, dont certaines concernent en réalité des œuvres produites par des ressortissants nationaux à l'étranger. Le cas des importations et de la production à l'étranger des ressortissants nationaux est connu depuis longtemps, mais à l'ère de l'édition numérique, ce phénomène s'est considérablement compliqué et prend des proportions qui pourraient sérieusement ébranler les fondements géographiques d'un système pensé par et pour des Etats. On pense naturellement à l'enjeu que représentent la conservation et le signalement de ce qui est publié sur internet. Compte tenu des masses de données en jeu, il est tout simplement impossible pour les bibliothèques de réaliser des collectes exhaustives de la production nationale. Certaines ont commencé à archiver les sites web, par sélection ou échantillonnage, au moyen de robots moissonneurs, et à développer des stratégies d'identification automatique d'un domaine national dans l'immensité de la Toile (par filtrage des collectes automatiques aux domaines de haut niveau comme le .fr pour la France ou le .dk pour le Danemark, par exemple). Mais la nature même du web, média essentiellement transfrontalier, résiste à toute idée de collecte exhaustive d'un corpus « national ».

Des organisations internationales comme le Consortium international pour la préservation de l'internet<sup>vi</sup> travaillent à la réalisation de collectes collaboratives de l'internet (par exemple, pour la couverture documentaire d'événements sportifs planétaires, comme les Jeux Olympiques) et à une répartition des tâches qui est finalement assez proche, dans son principe, des préconisations du CBU. Ces efforts sont cependant encore balbutiants et reposent sur des techniques de collecte, de signalement et de conservation qui sont très différents des procédés jusque-là utilisés pour traiter les collections nationales et échanger des données entre Etats. Il n'existe nulle part de bibliographie de l'internet, sauf à considérer que les grands moteurs de recherche sont ce qui s'en rapprocherait le plus.

### **Les bibliographies nationales à l'épreuve des nouvelles frontières du numérique**

Une autre façon de définir la production nationale consiste à considérer que c'est la production recensée dans la bibliographie nationale officielle, autre pièce maîtresse du CBU. Dès 1954, l'UNESCO reconnaît que le recensement des documents reçus par dépôt légal et la publication de ce recensement sont une contrepartie naturelle à l'effort consenti par ceux qui sont soumis au dépôt légal (éditeurs, imprimeurs, auteurs selon les pays). Mais il n'y a

pas non plus totale adéquation entre ce qui est collecté par dépôt légal et ce qui est signalé dans la bibliographie nationale. Les bibliographies nationales officielles décrivent majoritairement les ressources éditées en nombre et relevant des circuits commerciaux de l'édition. Les ressources publiées en dehors de ces circuits restent souvent minoritaires (publications à compte d'auteur, publications d'associations, publications officielles, publications des chercheurs). Si certaines bibliothèques engagées dans la collecte des livres et périodiques électroniques envisagent (comme c'est le cas de la BnF) d'introduire la description de ces publications dans leur bibliographie nationale, cela ne concerne encore qu'une infime partie de ce qui est publié en ligne.

Plusieurs congrès internationaux consacrés aux services bibliographiques nationaux<sup>vii</sup> ont longuement discuté de la couverture de la bibliographie nationale officielle. La liste des critères à prendre en compte est longue et dépend bien sûr du contexte national. Aujourd'hui, il est communément admis que d'une part tout ce qui est reçu par dépôt légal n'a pas vocation à être systématiquement décrit à la pièce dans la bibliographie nationale, et que d'autre part la bibliographie peut recenser des ressources produites hors des frontières du pays par des ressortissants nationaux ou sur des sujets relatifs au pays (par exemple, la Bibliothèque nationale de Suisse recense toutes les publications sur la Suisse, et pas seulement celles qui sont produites en Suisse). Les contraintes économiques des bibliothèques les ont conduites à adopter des politiques sélectives et aussi pragmatiques que possible, dont les instances internationales de l'IFLA ont reconnu la nécessité : il est recommandé de procéder par échantillonnage de ce qui est collecté et de partager avec d'autres institutions le travail de recensement de la production nationale.

Si la bibliographie nationale contribue ainsi à définir la production nationale, elle ne la recense donc pas de façon exhaustive. On ne peut plus considérer non plus que la somme des productions nationales est égale à la production mondiale, car un nombre croissant de ressources échappent à tout recensement : les ressources supranationales comme par exemple les publications intergouvernementales ou celles des associations internationales, dont le signalement est rarement fait dans le pays du siège comme cela devrait être le cas. En outre, il faut s'interroger sur la production collaborative de ressources par les internautes eux-mêmes dans des cadres qui échappent totalement aux prérogatives des Etats et des agences bibliographiques. Des acteurs « globaux », nés du web et de son industrie, sont devenus des forces de production d'information au moins aussi importants que de grands Etats. Aucune bibliothèque n'a mandat de conserver les contenus de l'encyclopédie en ligne Wikipédia. Et si la Bibliothèque du Congrès s'est vue confier par Twitter la totalité des données issue de ce réseau social, ces données intéressent potentiellement les chercheurs du monde entier. La production « grise » issue de l'activité des réseaux sociaux et professionnels en ligne échappe ainsi à tout recensement alors qu'elle peut témoigner de la vitalité d'une nation à travers la construction du débat public, la formation des opinions, des controverses, sujets qui ont toujours intéressé les chercheurs, historiens et sociologues, qui s'appuyaient jusqu'alors sur des sources telles que la presse ou les revues d'opinion, qui tendent à se dématérialiser.

Plus préoccupant peut-être, dans le domaine de l'édition académique, on observe un double mouvement de dématérialisation et de concentration des acteurs commerciaux. Les nouveaux modes de gestion des ressources électroniques bousculent profondément les conditions de leur diffusion et conservation par les bibliothèques. Tant que les documents disposaient d'un support matériel, la bibliothèque pouvait se procurer ces supports auprès de l'éditeur et organiser à son goût le signalement et l'accès pérenne au contenu. Mais les nouveaux modèles économiques mis en œuvre par les éditeurs et distributeurs de ressources

numériques reposent de plus en plus sur une logique d'abonnement. Les bibliothèques paient pour des accès à ces ressources distantes mais, sauf dispositions particulières issues de négociation à grande échelle (dans le cadre, par exemple, de licences nationales), les éditeurs ne fournissent pas les fichiers eux-mêmes, qui restent stockés sur leurs propres serveurs. Peut-on dire que ces ressources pourtant acquises de manière onéreuse appartiennent à la bibliothèque ? Ont-elles vocation à être signalées dans les catalogues et bibliographies qui, par tradition, recensent ce que les bibliothèques conservent dans leurs emprises ?

### **Le CBU : une utopie à revisiter ?**

Au terme de ces mutations technologiques, commerciales et sociales, il résulte que toutes les notions qui permettaient de définir et décrire une publication dans le cadre du CBU en 1974 sont aujourd'hui sérieusement ébranlées. En toile de fond des difficultés et des nécessaires adaptations d'ordre conceptuel et normatif rendues nécessaires par ces transformations de la nature même de la « matière » dont le CBU se préoccupe, une réalité économique majeure s'est en outre imposée à tous : la masse.

Le volume des ressources à décrire en 2014 est sans commune mesure avec celui qu'on pouvait concevoir en 1974. Toutes les mutations que traverse la production éditoriale aboutissent mécaniquement à un accroissement vertigineux du nombre d'entités à décrire. Un nombre croissant d'acteurs disposent de moyens de publication infiniment moins coûteux qu'autrefois. Les pratiques culturelles et sociales nées de l'internet ont fait entrer dans la sphère du patrimoine public à conserver par les institutions quantité de ressources qui jusque-là relevaient de la sphère privée et de l'oralité. Dans l'édition numérique, toutes les publications peuvent se décliner, s'agréger ou se découper dans d'innombrables parties ou manifestations. Dans le même temps, les moyens humains et technologiques des bibliothèques n'ont pas progressé dans les mêmes proportions, loin s'en faut. Cette situation implique de revoir non seulement l'appareil normatif et technologique de leur intervention, mais aussi le cadre et la mesure de leur coopération : celle-ci doit nécessairement s'ouvrir à d'autres acteurs et penser l'économie du partage des tâches à une autre échelle.

Du point de vue de la description bibliographique, on a vu que ces évolutions avaient été suivies de près, tant que cela paraissait possible, par les communautés de normalisateurs issues du CBU. Et ce jusqu'à ce que l'on épuise un certain modèle de représentation de l'information, aujourd'hui dépassé par la vitesse, la diversification et la volumétrie des ressources, mais aussi par les changements dans les usages. Nous allons voir à présent comment le modèle initial de la *notice* bibliographique tel qu'appréhendé dans les premières éditions de l'ISBD a explosé au profit d'une nouvelle représentation des *données*. Cette représentation induit en retour une nouvelle approche de la normalisation, mais aussi de l'économie générale de la production et de l'échange des données.

## **2 L'ADAPTATION DE LA NORMALISATION BIBLIOGRAPHIQUE A LA NOUVELLE DONNE DOCUMENTAIRE : DE L'ECHANGE DE NOTICES AU PARTAGE DE DONNEES**

Dans cette seconde partie, nous allons passer en revue quelques grandes mutations d'ordre normatif et conceptuel qui ont été rendues nécessaires par la nouvelle donne de l'offre documentaire que nous venons de présenter, mais aussi des usages. Il n'est bien sûr pas question de retracer ici quarante ans de normalisation bibliographique. Notre propos se concentrera sur trois aspects qui nous paraissent très structurants pour penser l'avenir de la



collaboration internationale issue du CBU du point de vue de la description et du contrôle : l'uniformisation de la description afin d'échanger les données, la modélisation de l'information pour gagner en granularité de description, et enfin l'évolution du contrôle d'autorité.

### **Uniformiser l'information bibliographique pour l'échanger**

En 1974, l'idée du CBU est de diffuser au niveau international les notices établies localement dans chaque pays. L'objectif de la normalisation est alors *d'uniformiser la forme et la structure de l'information* dans le but de l'échanger au niveau international. Par « échanger » il faut comprendre ici « dupliquer pour diffuser ». On évite la duplication des opérations de catalogage, mais on duplique les notices pour les diffuser. Pour qu'elles soient exploitables en-dehors de l'environnement dans lequel elles ont été produites, il faut uniformiser leur présentation et leur contenu.

L'uniformisation des *points d'accès aux descriptions bibliographiques* est l'un des principaux objectifs des Principes de Paris élaborés en 1961 dans le cadre de la première Conférence internationale sur les principes de catalogage. Ces Principes seront déclinés par la suite dans une grande variété de recommandations internationales portant sur la forme et la structure des noms par type d'auteurs (collectivités<sup>viii</sup>, personnes<sup>ix</sup>) ou par type de titres (œuvres anonymes<sup>x</sup>, œuvres liturgiques, etc.). Diffusées comme vedettes des notices bibliographiques publiées dans les bibliographies nationales, ces formes normalisées ont vocation à être utilisées à l'identique partout dans le monde.

Jusque dans les années 80, les vedettes et les descriptions bibliographiques dont elles constituent les points d'accès, forment des tandems indissociables. Mais l'informatisation progressive des catalogues va conduire à une rupture au sein du couple vedette/description bibliographique et aboutir à la création d'un nouveau type de notice : *la notice d'autorité*. Conçue initialement comme une aide au catalogage, celle-ci regroupe toutes les informations nécessaires à l'identification de l'entité concernée. La notice tire son nom du fait qu'elle fait autorité par rapport au catalogue qu'elle sert à gérer, c'est-à-dire qu'elle définit LA forme d'autorité à utiliser obligatoirement comme point d'accès aux descriptions bibliographiques de ce catalogue. Toutefois, au fil du temps, la notice d'autorité est devenue de plus en plus autonome par rapport à son catalogue d'origine, pour faire autorité en soi. Sa valeur s'impose parce qu'elle véhicule des informations vérifiées par des experts, informations dont on découvre qu'elles sont en réalité réutilisables dans des contextes très divers, bien au-delà de la seule communauté des bibliothèques. Selon les principes du CBU énoncés en 1974, les notices d'autorité sont censées être « échangées » au niveau international, c'est-à-dire dupliquées et exportées vers d'autres systèmes où elles pourront être mises à jour localement, par exemple par l'ajout de formes locales du nom de l'entité plus appropriées aux attentes des utilisateurs.

Le principe d'une description bibliographique internationale normalisée (ISBD) est né en 1969 lors de l'International Meeting of Cataloguing Experts qui s'est tenu à Copenhague. Le premier ISBD apparaît, on l'a vu, en 1971 pour les livres et les publications en série. Il précise les éléments d'information requis pour identifier une publication, prescrit leur ordre de présentation, définit une ponctuation pour les délimiter, donne les règles de transcription des données et les sources d'information à utiliser.

Parallèlement à ces recommandations précisant la forme et la structure de la description bibliographique et des vedettes servant de points d'accès, se développent les formats d'encodage des éléments de données qui s'appuient sur la ponctuation prescrite par l'ISBD : les formats MARC (*Machine Readable Cataloguing*) conçus et utilisés par et pour les bibliothèques. Le format bibliographique MARC inclut, dans les zones codées de longueur fixe, des données de gestion attachées à chaque notice et qui permettent par exemple de faire des tris ou des sélections de notices : pays de publication, langue et écriture du document décrit, date de création de la notice, langue de catalogage, etc. Ainsi, l'ISBD permet de produire et d'exploiter des notices lisibles par des humains et par des machines. Des notices produites manuellement et d'autres produites par du catalogage informatisé peuvent coexister dans un même catalogue.

Pour échanger des notices en MARC les bibliothèques utilisent le protocole défini par la norme ISO 2709, dont la première édition en 1973 a pour titre « Format pour l'échange d'information sur bande magnétique ». Signe des temps, sa dernière édition, en 2008, s'intitule « Format pour l'échange d'information ». Ce format d'échange structure les différents formats MARC (parce que chaque pays a développé son MARC, on parle de « Babel MARC »). Elle permet l'interopérabilité entre les différents MARC par la mise en place de conversions (malheureusement peu de conversions sont maintenues au niveau international, même pas entre MARC21 et UNIMARC) et elle permet donc *les échanges de notices MARC entre bibliothèques*.

En s'alignant sur les normes internationales, les pays les moins avancés techniquement ont pu réellement participer à l'effort international et en bénéficier en retour. Mais à partir de 1990, l'environnement des bibliothèques évolue profondément comme on l'a vu précédemment. La technologie des communications internationales se développe (internet et les technologies du web), et le besoin est ressenti de travailler plus étroitement avec les producteurs de l'information, les autres institutions patrimoniales, les détenteurs de droits, autant de communautés qui ne connaissent pas les normes sophistiquées et les formats quasiment « propriétaires » conçus par les bibliothèques. C'est le début d'un réexamen des normes et des pratiques, car une tension est née entre la maturité des normes internationales utilisées au quotidien et le besoin grandissant de décrire des ressources numériques (numérisées ou nées numériques). L'item (le document) cesse d'être l'unique objet de description : il faut pouvoir, on l'a vu, *décrire la ressource indépendamment de son support*. Mais il faut également pouvoir *décrire les objets numériques dans le contexte de leur cycle de vie complet* : les métadonnées descriptives ne suffisent plus et doivent être articulées avec les métadonnées techniques et administratives utiles à la préservation ou à la gestion des droits numériques. Une révolution normative s'engage pour s'attaquer à ces nouveaux besoins.

### **Modéliser l'information bibliographique pour gagner en granularité**

En 1974, on recommandait de faire la description bibliographique de référence « livre en main » selon la formule consacrée. Il s'agissait de reproduire et, plus tard, de saisir informatiquement ou d'encoder, dans le respect des normes et de la structure des formats, des informations déjà présentes dans la publication. Cette description bibliographique représentait alors l'essentiel du travail des catalogueurs. Le choix et l'établissement des points d'accès, qui supposait une analyse du contenu du document selon des référentiels contrôlés, était encore une étape secondaire de leur travail.

Avec l'arrivée des ressources numériques, non seulement il s'est avéré impossible de disposer de la ressource « en main », mais en plus il est devenu très complexe de déterminer ce qui devait faire l'objet d'une notice descriptive. Comme on l'a vu, définir l'objet ou l'entité à décrire est devenu un vrai casse-tête. Comment faciliter l'accès au contenu informationnel des ressources décrites ? Quelle granularité viser dans les descriptions pour répondre aux besoins des utilisateurs ? Un chapitre d'un livre ? Un article dans un fascicule de périodique ? Un bouquet de monographies ? La version re-matérialisée d'un document numérisé ? Chaque version numérique d'un e-book définie par son format (PDF, ePUB, etc.), voire par le type de DRM qui s'y applique ? Où s'arrête la description bibliographique, et où commence la description de l'exemplaire ? Quelle norme utiliser pour décrire un document imprimé qui a été numérisé : celle qui s'applique aux monographies imprimées ou celle qui s'applique aux ressources électroniques ? Pour toutes ces raisons, il est apparu indispensable de distinguer le *contenu* du *support* de l'information, et c'est ce qui a motivé la dernière modification apportée à l'ISBD (création de la zone 0).

En 2000, Elaine Svenonius<sup>xi</sup> écrivait en substance « *la description bibliographique est un art, et donc supporte une part d'approximation. Il y a toujours des cas "hors-normes" qu'il faut traiter avec une part d'arbitraire. [...] Mais si le nombre des cas hors-normes n'est plus maîtrisable, c'est qu'il est temps de revoir les méthodes de travail* ». C'est ce qui s'est produit en 1990 lors du Séminaire sur les notices bibliographiques organisée par l'IFLA où il a été décidé de revoir le modèle de l'information bibliographique afin de mettre l'utilisateur au cœur du dispositif et utiliser les avancées techniques pour mieux répondre à ses attentes. Ces discussions s'engagent dans le contexte où, en écho aux évolutions précédemment évoquées, l'univers bibliographique ne se limite plus aux bibliothèques, et où avec l'expansion d'un internet dont les moteurs de recherche concurrencent très sérieusement les catalogues, les bibliothécaires prennent conscience de la nécessité de se recentrer sur les besoins nouveaux des usagers : sur le web, ceux-ci n'ont pas envie de rester à la bibliothèque mais veulent interagir avec d'autres ressources, d'autres communautés.

La modélisation de l'information bibliographique va permettre d'évaluer si le contenu des notices bibliographiques tel qu'il a été défini par le passé répond ou non aux besoins réels des utilisateurs et va contribuer à le faire évoluer. C'est l'objectif du désormais célèbre modèle *Functional Requirements for Bibliographic Records* (FRBR) publié en 1998<sup>xii</sup>. Tous les éléments de données présents dans la notice ISBD sont confrontés aux tâches des utilisateurs : trouver, identifier, sélectionner, obtenir. Les nouveaux principes internationaux de catalogage publiés en 2009<sup>xiii</sup> sont organisés autour de ces quatre tâches et en définissent une cinquième : « *naviguer* », qui sera reprise dans l'édition révisée des FRBR par le terme « *explorer* ». Grâce à l'analyse proposée par les FRBR, la notice descriptive cesse d'être conçue comme un bloc d'information compact pour s'apparenter plutôt à un réseau d'éléments de données ayant entre eux de multiples relations. On y distingue les éléments de données relatives à l'œuvre, à l'expression, à la manifestation et à l'item, et ceux relatifs aux entités telles que les personnes, les collectivités, les titres, les sujets, qui constituent des concepts (les gens, les choses, les lieux...) particulièrement utiles pour permettre à l'utilisateur de trouver mais aussi de découvrir les ressources d'une manière plus intuitive.

Pour ce qui concerne les notices d'autorité, c'est en 2001, avec la 2<sup>nd</sup>e édition des *Guidelines for Authority Records and Entries*<sup>xiv</sup> qu'apparaît dans un texte normatif international la notion de partage (« *sharing* ») des notices d'autorité et non plus celle d'échange (« *exchange* »). C'est un pas important vers l'abandon de l'uniformisation des formes des noms au niveau international, abandon entériné par la création du Fichier

d'autorité international virtuel (VIAF) en 2003, pour les entités personnes. VIAF est alimenté par les notices d'autorité produites par les institutions qui en sont membres, soit près d'une quarantaine aujourd'hui. Pour une même personne, une notice d'autorité VIAF regroupe les différentes formes du nom qui font autorité dans les fichiers des différents pays, quelle que soit la langue ou l'écriture de ces formes, sans définir de forme pivot. Toutes les formes nationales du nom sont ainsi mises sur un pied d'égalité. C'est leur confrontation qui va contribuer à renforcer l'identification de la personne au niveau international. VIAF traite aussi, mais avec un degré de fiabilité moindre pour l'instant, des noms de collectivités, des noms géographiques, des titres d'œuvres et d'expressions. Des humains ou des machines peuvent venir identifier dans VIAF les éléments de données dont ils ont besoin localement, et, la technique aidant, choisir de dupliquer l'information dans leurs applications locales ou bien de créer un lien entre une application locale et le sous-ensemble adéquat de VIAF. Le pragmatisme de VIAF va inspirer les auteurs des modèles *Functional Requirements for Authority Data* (FRAD)<sup>xv</sup> et *Functional Requirements for Subject Authority Data* (FRSAD)<sup>xvi</sup> respectivement publiés en 2009 et 2010. Dans leurs titres respectifs, on note que le mot *données* (data) remplace le mot *notices* (Record) et ce changement n'a rien de fortuit : il marque une nouvelle étape vers la prise en compte de la granularité de l'information produite par les bibliothèques et utilisée par les usagers.

### **Le contrôle d'autorité et le partage des données**

Qu'est que le contrôle d'autorité ? En 1990, Clack<sup>xvii</sup> écrit qu'« *il s'agit d'un ensemble de procédés techniques appliqués à un catalogue de bibliothèque pour le structurer. Unicité, normalisation et mise en relation sont les principes de base du contrôle d'autorité* ». Selon cette définition, le contrôle d'autorité s'exerce aussi bien sur les notices bibliographiques de référence (*definitive bibliographic records*) qui décrivent les documents que sur les notices d'autorité qui documentent les formes des noms des contributeurs et des titres d'œuvres et qui font autorité (*authoritative forms of names*) dans un catalogue donné. Au départ, l'objectif du contrôle d'autorité était surtout d'éviter les doublons, de normaliser le contenu, la forme et la structure des données qui constituent les notices et de mettre en relation ces notices ou ces données les unes avec les autres, au sein d'un catalogue particulier, afin de garantir sa cohérence et de faciliter son usage. Aux origines du CBU, la notion de contrôle d'autorité est ainsi indissociable de celle de catalogue. C'est grâce au contrôle d'autorité qu'un catalogue de bibliothèque rassemble sous le nom d'un auteur *toute* son œuvre mais *seulement* son œuvre et que chaque document fait l'objet d'une seule et unique notice descriptive. On peut comparer cette activité qui consiste à nettoyer, redresser et harmoniser des paquets de métadonnées à celle du contrôle qualité dans une chaîne de production industrielle.

Dans le cadre du CBU, le contrôle d'autorité est confié au niveau national à l'agence bibliographique nationale (recommandations 18.1.1 et 18.1.3). Ce travail est, en réalité, celui du bibliothécaire depuis des siècles. Et de fait, une grande expertise humaine est requise pour mener à bien cette tâche. Toutefois, au fur et à mesure des progrès informatiques et de la normalisation documentaire, des procédés techniques sont venus en aide à la décision ou ont permis de réaliser certains contrôles automatiquement qui trouvent aujourd'hui leur prolongement dans l'utilisation d'algorithmes de rapprochement. Surtout, l'usage et l'environnement technique des catalogues ont beaucoup évolué au cours des cinquante dernières années, et peu à peu le contrôle d'autorité s'est émancipé de la notion de catalogue : il a changé d'objet, d'envergure et est devenu résolument stratégique.

L'unité d'information bibliographique est aujourd'hui moins la notice que *l'élément de donnée*. Cette évolution a été rendue possible par l'apparition d'XML qui permet d'individualiser les éléments de données beaucoup plus facilement que ne le permettaient les zones et les sous-zones des formats MARC, et qui est un format compris et utilisé par beaucoup d'autres communautés que celles des bibliothèques. Il contribue à décloisonner les données. De la même façon, on constate que le contrôle d'autorité se déplace lui aussi de la notice vers l'élément de donnée. A côté du contrôle d'autorité traditionnel assuré par les catalogueurs (contrôle de doublons et de conformité des formes et structures des données), se développe ainsi *un nouveau niveau de contrôle* reposant sur la cohérence de l'analyse documentaire à proprement parler. Il consiste à apparier des vocabulaires issus de communautés différentes en établissant des relations justes et fiables entre des concepts similaires mais exprimés différemment par des communautés variées. Il n'est plus question d'uniformiser les données pour les échanger, mais de *mettre en relation des concepts similaires pour les partager*. Par exemple, entre le concept de « titre » pour un bibliothécaire, pour un archiviste et pour un éditeur : mettre en relation ces trois éléments de données par un lien informatique, c'est permettre à un usager de naviguer facilement entre, d'une part, le catalogue local où il a lancé sa recherche « titre » et d'autre part un inventaire d'archives et un site d'éditeur. Le contrôle d'autorité, qui sort du catalogue pour s'intégrer au web tout entier, est devenu primordial dans le travail du catalogueur et est au cœur de la valeur ajoutée métier qui peut, à notre avis, distinguer aujourd'hui la contribution des bibliothèques dans un plus large écosystème de production et de partage de données.

### **3 QUEL FUTUR POUR LE CBU? L'OUVERTURE DES BIBLIOTHEQUES AU RESTE DU MONDE PAR LE WEB SEMANTIQUE**

Quel horizon imaginer aujourd'hui en matière d'organisation de la coopération entre bibliothèques? A la lumière des évolutions retracées, quels pourraient être les axes de réflexion stratégiques pour une refondation du CBU ? Dans quelles instances, dans quels lieux, l'avenir bibliographique des bibliothèques se joue-t-il à présent?

Souvenons-nous d'abord que le CBU est né de l'étroite coopération entre trois organisations internationales : l'UNESCO, parce que ce sont les gouvernements qui planifiaient et finançaient les systèmes d'information nationaux ; l'ISO/TC46, qui fédérait les organismes nationaux de normalisation dans le secteur de l'information et de la documentation ; et l'IFLA, fédération internationale des professionnels des bibliothèques. La coordination assurée par les organisations internationales apparaissait alors indispensable pour éviter que les systèmes nationaux ne se développent isolément et pour veiller à un équilibre entre les différentes parties du monde inégalement développées. Dans un contexte géopolitique fort différent (la fin de la Guerre Froide) où le contrôle de l'information par un grand nombre d'Etats demeurait une réalité, ce dispositif était certes marqué par la forte implication des associations professionnelles, mais aussi par la prépondérance du rôle des Etats. Il s'appuyait sur des politiques publiques ambitieuses et interventionnistes dans un contexte où les grands acteurs privés de la globalisation de l'information que nous connaissons aujourd'hui n'existaient pas, ou sous des formes très différentes.

Dès le départ, les bibliothèques, et les bibliothèques nationales au premier rang, ont été placées au cœur de ce dispositif de coopération essentiellement public, sans doute parce qu'elles avaient déjà une longue tradition d'organisation du savoir et de sa mise à disposition auprès des usagers, tradition qui faisait autorité aux yeux de la puissance publique. Utilisant

les puissants leviers législatifs que constituaient le dépôt légal et les bibliographies nationales, elles se sont d'elles-mêmes positionnées de manière volontariste voire militante comme les opérateurs de confiance capables de mettre en œuvre le programme du CBU. En 1974, l'installation du Bureau du CBU comme un organe permanent de l'IFLA marque bien le fait que la coordination internationale en est confiée sans partage à la communauté des bibliothécaires. Les archives et les musées sont absents du dispositif, peut-être parce que leur patrimoine était essentiellement constitué de pièces uniques nécessitant d'autres modes de collecte et de gestion pour lesquels on voyait moins l'intérêt de mutualiser le travail. Les éditeurs, on l'a vu, auraient pu légitimement être parties prenantes du dispositif, notamment parce qu'ils étaient soumis au dépôt légal et qu'ils fournissaient déjà des éléments de données utiles au catalogage avant publication. Toutefois, la Fédération internationale des éditeurs, pourtant créée dès 1896, n'est pas citée parmi les parties impliquées dans la mise en place du CBU et les éditeurs sont restés en dehors d'un processus de normalisation étroitement contrôlé par des bibliothécaires généralement peu enclins à prendre en compte leurs besoins particuliers de signalement, notamment d'ordre commercial.

Le dispositif initial du CBU était donc essentiellement un dispositif public porté par les bibliothèques, qui excluait de sa gouvernance et de son réseau d'autres organisations publiques ou privées. Une telle organisation n'est plus concevable aujourd'hui. Si elles conservent des atouts évidents dans certains domaines, les bibliothèques n'ont plus le monopole de la structuration, de la maîtrise et de l'échange de métadonnées. D'autres opérateurs, privés ou associatifs pour la plupart, ont pris le pas sur leur profession et sur les Etats. Et comme on l'a vu, la production d'information implique désormais des acteurs transnationaux ou multinationaux majeurs qui échappent totalement au contrôle des Etats. La devise de la fondation californienne Internet Archive<sup>xviii</sup> « *Universal Access to All Knowledge* » pourrait être celle de nombre d'organisations publiques et privées aujourd'hui. Dans l'environnement devenu extrêmement compétitif de la production et de la recherche d'information, les bibliothèques doivent donc se concentrer sur ce qu'elles savent le mieux faire et trouver les modes d'interaction avec d'autres communautés professionnelles qui leur permettront de valoriser leur contribution spécifique. Selon quelles priorités et quelles modalités doivent-elles aujourd'hui intervenir et coopérer ? Nous proposons ici trois axes de réflexion et d'action qui font écho à quelques grands débats d'actualité.

### **Premier axe : maintenir l'interopérabilité entre bibliothèques**

Dans le prolongement des précédents épisodes qui leur ont permis de construire par consensus, bon an mal an, un remarquable édifice de normes et de formats communs ou convertibles, les bibliothèques doivent d'abord veiller à rester interopérables entre elles. Seul le maintien de cette interopérabilité leur permettra de conserver collectivement et durablement une masse critique de données capable de se mesurer à l'échelle du web. En écho aux grands débats bibliographiques qui animent actuellement la profession, on relèvera quelques points de vigilance sur lesquels nous devons rester attentifs et à l'écoute les uns des autres<sup>xix</sup>.

On note d'abord que toutes les grandes bibliothèques sont engagées dans une transition bibliographique de longue haleine, qu'on évalue généralement à dix ans et qui se répercute à au moins trois niveaux : l'implémentation effective du modèle FRBR, le changement de code de catalogage, et le remplacement des formats MARC. Cette transition est abordée de manière différente selon les pays et les contextes. Dans le sillage de la Bibliothèque du Congrès, plusieurs grandes bibliothèques issues de la tradition des règles de catalogage

anglo-saxonnes AACR2 et du format MARC21 ont lancé la mutation de leur dispositif de production avec l'adoption du code de catalogage RDA. D'autres, comme la Bibliothèque nationale de France (avec le projet data.bnf.fr) ou la Bibliothèque nationale d'Espagne (avec le projet datos.bne.es) ont préféré pour l'instant investir en priorité dans la transformation et l'exposition en RDF de leur production catalographique rétrospective sans faire profondément évoluer leurs modes de production. Certaines bibliothèques (comme la Deutsche Nationalbibliothek) travaillent parallèlement aux deux niveaux (production et diffusion), tandis que les moins armés restent prudemment dans l'expectative.

Il est aujourd'hui communément admis que l'idée d'un format unique de production ou de diffusion pour tous les types de données et pour tous les usages est illusoire. Bien au contraire, elle pourrait s'avérer contre-productive, voire dangereuse : il restera nécessaire de gérer l'information bibliographique dans des contextes très différents et à des niveaux de granularité ou de complétude variés, qu'il s'agisse de la gestion locale, des échanges professionnels au sein d'un réseau de production ou d'une diffusion et réutilisation par des publics aux attentes plus ou moins pointues ou génériques. Dans ce contexte, une pluralité d'approches est donc possible, sinon indispensable. Il convient néanmoins de continuer à faire converger les approches pour la définition des *formats d'échange* qui seront des pivots essentiels pour le référencement, l'échange ou le liage d'informations au niveau international et interprofessionnel. De ce point de vue, les propositions portées par les initiatives BIBFRAME ou schema.org doivent évidemment être suivies de très près et discutées au niveau international. La définition des nouveaux formats d'échange constitue, de notre point de vue, une priorité qui mériterait une concertation et une coordination du même ordre que ce que le CBU avait mis en place en matière de normalisation.

Parallèlement à ces travaux, les grandes bibliothèques et agences bibliographiques nationales doivent continuer d'alimenter régulièrement, par une production de qualité, les grandes bases de données bibliographiques internationales de nature collaborative. On pense naturellement à WorldCat<sup>xx</sup>, à VIAF<sup>xxi</sup>, à des registres plus anciens comme celui de l'ISSN ou beaucoup plus jeunes comme celui de l'ISNI<sup>xxii</sup>. Dans ce domaine, le point de vigilance porte plutôt sur les modalités de gouvernance du traitement et du reversement de l'information bibliographique entre, d'une part, les opérateurs locaux fournisseurs et ré-utilisateurs et, d'autre part, les opérateurs globaux agrégateurs et diffuseurs de données. Parmi les questions importantes qui doivent faire l'objet d'une concertation internationale, il y a la question du statut ou du *droit des données* (enjeu de l'ouverture des données et du régime juridique des licences), de la *traçabilité* de leur provenance (enjeu de la citation, mais aussi de la confiance), et, finalement, de leur *qualité*. On peut considérer que les nouveaux lieux de gouvernance qui ont succédé aux instances de coordination du CBU se sont pour partie déplacés vers les instances de pilotage et de coordination de ces différentes initiatives (le conseil de VIAF ou l'Agence internationale ISNI-IA, par exemple). Afin de conserver la maîtrise de leurs données en toute transparence, il nous semble particulièrement stratégique que les bibliothèques, en partenariat avec les opérateurs privés ou associatifs comme OCLC, sans lesquels ces grands projets n'auraient jamais vu le jour, s'impliquent activement et fassent valoir leurs intérêts et leurs positions dans la définition des règles de fonctionnement de ces nouveaux processus d'agrégation, transformation et diffusion des données au niveau global.

## Second axe : adopter les standards du web sémantique

Les bibliothèques doivent néanmoins faire attention à ne pas s'enfermer dans les seules coopérations et communautés qui réunissent depuis longtemps leurs organisations, ni à reproduire à l'ère moderne leurs différends historiques et culturels. Les rythmes de la normalisation bibliographique internationale ne sont plus adaptés à la vitesse d'évolution des technologies et des usages et surtout, le périmètre des acteurs producteurs ou agrégateurs de données et prescripteurs de standards s'est considérablement élargi et diversifié. Les bibliothèques ont tout intérêt à poursuivre ou à engager le dialogue avec les communautés les plus actives dans le web de données. Cela les aidera d'ailleurs probablement à aplanir leurs propres différends.

Le ticket d'entrée pour exister sur le web passe par les mutations dont il a largement été question précédemment. Le web sémantique ouvre aux bibliothèques des perspectives de visibilité et de rapprochement de données jusque-là inconcevables du fait du cloisonnement des communautés professionnelles. On peut aujourd'hui envisager de rapprocher les données de bibliothèques avec celles d'autres opérateurs culturels, mais aussi des données statistiques, commerciales, géographiques, linguistiques, scientifiques... le champ des possibles semble infini et constitue un levier d'innovation extrêmement ouvert en même temps qu'une très grande source d'économie potentielle pour tous les acteurs producteurs, diffuseurs et utilisateurs de données. Les conditions techniques et juridiques d'entrée et de visibilité dans le web sémantique sont bien connues. Nous voudrions toutefois insister sur le fait que certaines d'entre elles sont devenues des vecteurs absolument stratégiques de la présence comme de la coopération des bibliothèques sur le web. L'expression et la publication des données au format RDF est une condition minimale. L'ouverture juridique facilitant la réutilisation des données est un vrai plus. Mais pour les bibliothèques, nous pensons que l'attention doit se focaliser aujourd'hui sur deux aspects : *les identifiants* (URI) et *les référentiels d'autorité* (ou vocabulaires).

Dans l'architecture et la « grammaire » du web sémantique, le partage des données repose sur l'existence et la maintenance d'identifiants uniques, normalisés, pérennes et actionnables. Les identifiants sont en effet des passerelles sans lesquelles on ne peut rien localiser, rien lier, rien construire. La normalisation et la dissémination d'identifiants de confiance apparaît ainsi comme un enjeu de taille pour les bibliothèques comme pour leurs partenaires d'autres communautés. Aujourd'hui, elles doivent donc se préoccuper collectivement des conditions pratiques et économiques d'assignation d'identifiants à leurs ressources, et faire des choix leur garantissant à la fois une certaine autonomie de gestion pour l'identification de leurs collections et un maximum de visibilité et d'interopérabilité. De ce point de vue, l'essor de l'ISNI, pour les identités publiques, ou les discussions relatives à l'ISTC, pour les œuvres textuelles, sont des sujets qui appellent à une vigilance et à une concertation internationale. Il y a en effet derrière les questions d'identification, certes des enjeux de nature scientifique ou normative (quel identifiant pour quel type de ressource ?), mais également économique (à quelles conditions obtenir et distribuer un identifiant standardisé ?).

Parallèlement, et comme nous l'avons vu précédemment, la valeur ajoutée distinctive des bibliothèques s'est déplacée progressivement des données bibliographiques (qu'il est possible de récupérer en grande partie à la source, auprès des producteurs) vers les *données d'autorité* (qui sont notre trésor de guerre, car elles résultent de l'analyse proprement documentaire des fonds, et commandent le regroupement des informations dans l'ère FRBR).



Sur le web sémantique, et comme le montre bien le succès international du projet VIAF, ce sont les vocabulaires et les référentiels spécialisés issus du monde des bibliothèques qui sont les plus recherchés et réutilisés par d'autres communautés. A la BnF, nous avons observé que le jeu de données diffusé par le biais du service data.bnf.fr et publié en SKOS qui rencontre le plus de succès auprès des ré-utilisateurs était le répertoire d'autorités matière RAMEAU, qui présente par ailleurs un grand intérêt d'exploitation multilingue du fait des alignements réalisés avec ses équivalents anglo-saxon (LCSH) et allemand (SWD). La collaboration internationale entre bibliothèques dans ce domaine devrait donc également constituer un domaine de concertation et d'investissement prioritaire, qui soulève lui aussi des questions de gouvernance et d'économie. Beaucoup d'institutions ont en effet tendance, faute de moyens, à réduire leur capacité de production et de contrôle des autorités et se reposent de plus en plus sur des répertoires collaboratifs comme VIAF. Si toutes les bibliothèques suivent ce chemin, la valeur même des répertoires collaboratifs diminuera et la confiance accordée aux données de bibliothèques ne sera plus la même.

Ce n'est donc pas un hasard si les initiatives qui lient la problématique des identifiants à celle des autorités sont aujourd'hui parmi les plus discutées dans la profession : elles sont en effet véritablement stratégiques pour la présence collective des bibliothèques dans le web de demain. La constitution de *registres collaboratifs d'autorités associés à des identifiants standardisés* constitue de notre point de vue un pilier fondamental du nouveau contrôle bibliographique universel. Le statut, le fonctionnement, le modèle économique et de gouvernance de ces registres sont des points d'attention de niveau international.

Pour conclure sur la participation des bibliothèques à la construction du web sémantique, nous voudrions évoquer une dernière question stratégique qui mériterait à notre avis d'être approfondie : les modalités de *participation des bibliothèques aux instances de gouvernance et de standardisation de l'internet*. De ce point de vue, il nous paraît très important de réfléchir à la place qu'occupent les bibliothèques au sein du Consortium W3C. Pour la petite histoire, on peut rappeler que le terme de « métadonnées », désormais solidement ancré dans le jargon des bibliothécaires, a été introduit dans notre communauté par les acteurs clés de la standardisation de l'internet au milieu des années 90. Le terme a en effet été proposé pour la première fois en 1994 (vingt ans après la fondation du CBU) par Tim Berners-Lee, lors d'une conférence à Genève à l'origine de la fondation du W3C.

Il faut saluer la démarche des professionnels qui ont réussi à porter la réflexion et la concertation sur la place des bibliothèques dans le web de données liées au sein même du W3C dans le cadre du groupe d'incubation dédié à cet effort il y a quelques années. Le rapport<sup>xxiii</sup> issu de cette concertation est devenu en quelques mois une référence capitale pour les bibliothèques alors qu'il n'est pas issu des instances traditionnelles de normalisation bibliographique. Est-ce à dire que les normes des bibliothèques de demain devraient trouver leur place au côté de celles du web, qui est le principal environnement de destination des données que nous produisons? Il nous faut nous interroger sur cet aspect de notre coopération en matière de normalisation et sans doute chercher à nous insérer davantage dans les instances de standardisation du web plutôt que de normaliser en circuit fermé entre bibliothécaires. Le nouveau CBU est probablement à penser au niveau de la gouvernance du web sémantique.

### **Troisième axe : s'insérer dans d'autres réseaux pour coproduire et lier les données**

Parmi les axes de collaboration sectoriels dont nous pensons qu'ils représentent un potentiel particulièrement intéressant en termes d'économie de production et de qualité de service pour les bibliothèques, nous voudrions finalement proposer quelques pistes de concertation que les bibliothèques gagneraient à envisager et à mieux coordonner pas seulement au niveau national, mais aussi international.

En premier lieu, les autres opérateurs publics producteurs de données culturelles, archives et musées en particulier, nous semblent être des partenaires naturels des bibliothèques. On a vu dans la première partie que le contexte numérique conduisait notamment à rapprocher de plus en plus les processus de gestion et les métiers des bibliothèques de ceux des archives. On peut citer comme exemple intéressant de collaboration et d'interopérabilité un point de convergence qui concerne, une nouvelle fois, les autorités. Les archivistes travaillent très sérieusement à la normalisation d'autorités pour les noms de personnes, familles et collectivités (schéma XML EAD-CPF)<sup>xxiv</sup> et cette initiative pourrait converger avec les efforts engagés par les bibliothèques pour faire de leurs propres notices d'autorité des pivots dans l'organisation et le partage de l'information. On pourrait ainsi imaginer des stratégies nationales voire internationales de mutualisation ou de co-production de fichiers d'autorités pour l'identification des collectivités en particulier. Les ressources des bibliothèques, archives et musées sont souvent complémentaires du point de vue des usagers (ne serait-il pas judicieux, pour un même créateur ou une même œuvre, de proposer par exemple un accès unique en ligne aux références de toutes les ressources pertinentes conservées par ces différentes institutions ?) mais décrites selon des logiques et des impératifs de gestion très éloignés au départ. Le web sémantique pourrait permettre (par l'utilisation des autorités et des identifiants en premier lieu) de regrouper des ressources décrites dans des systèmes de production et des langages et standards documentaires différents. On entrevoit dans ce secteur un vrai potentiel de coopération pour partager davantage la production de données culturelles et enrichir par la même occasion l'offre de services aux usagers. Pour les pouvoirs publics qui financent et administrent ces différentes institutions culturelles ou scientifiques, ce serait une belle démonstration d'efficacité et de prise en compte des besoins des usagers.

En second lieu, il nous semble que les éditeurs, les opérateurs de l'interprofession du livre (sociétés de gestion de droits, distributeurs, agrégateurs...) et les producteurs d'information au sens large, devraient être beaucoup plus impliqués dans la production et la diffusion de l'information bibliographique qu'ils ne l'ont été jusqu'à présent. Certaines bibliothèques et agences bibliographiques nationales ont déjà très largement intégré des logiques de collaboration avec les éditeurs en mettant en place des organisations propices au catalogage avant publication. Ce n'est pas le cas en France, où cette collaboration n'en est encore qu'à ses débuts, à la faveur de la dématérialisation des déclarations de dépôt légal et de la mise en place, encore expérimentale, d'un dépôt légal des livres numériques. La numérisation des collections, que ce soit du domaine public (développement de bibliothèques numériques patrimoniales comme Gallica à la BnF) ou sous droits (croissance de l'offre commerciale, d'ePub en particulier) crée un contexte de collaboration nouveau entre bibliothèques et éditeurs, pour l'identification et la gestion des droits en particulier.

Les bibliothèques ont, pour des raisons économiques, besoin de récupérer davantage de métadonnées à la source. Les éditeurs ont, pour des raisons de visibilité et de valorisation de leur offre en ligne, besoin d'enrichir leurs métadonnées. Dit autrement, les bibliothèques ont

besoin de récupérer des données bibliographiques tandis que les éditeurs sont en demande de données d'autorité. Certains projets nouveaux (en France, la numérisation des livres indisponibles du 20<sup>ème</sup> siècle, et le projet ARROW au niveau européen) nécessitent en outre le croisement de bases de données issues des deux communautés, qui possèdent des informations complémentaires : référentiels structurés et riches en autorités pour les bibliothèques, informations commerciales et de disponibilité pour les éditeurs, ou encore informations relatives aux auteurs et à leur œuvre conservées par les sociétés de gestion de droits. Le développement des livres numériques et du format ONIX offre finalement un nouveau contexte de coopération où l'on voit converger un intérêt mutuel à rapprocher les logiques descriptives et les référentiels des bibliothèques et ceux des éditeurs. En France, un rapport public<sup>xxv</sup> a récemment préconisé la création de *registres ouverts de métadonnées* qui permettraient de rassembler et diffuser conjointement des métadonnées issues des mondes de l'édition et des institutions patrimoniales en charge du dépôt légal. Il faut pouvoir imaginer des flux nouveaux de métadonnées entre les secteurs publics et privés afin de rechercher des bénéfices mutuels.

Enfin, et comme plusieurs bibliothèques ont déjà commencé à l'envisager, pourquoi ne pas faire de nos usagers des co-producteurs de nos métadonnées ? Les grandes librairies en ligne n'ont pas attendu les bibliothèques pour enrichir leurs catalogues des avis, notes de lecture et commentaires des usagers. D'une manière générale, les internautes ont pris l'habitude à la fois d'adopter une posture active et participative face à leur écran (partager, aimer, mais aussi commenter). Ils s'attendent désormais à trouver des avis d'autres usagers en complément des ressources descriptives de référence produites par les entreprises commerciales et les organisations publiques. Le web, est, par essence, une entreprise de co-production de masse dont l'économie repose de plus en plus sur le *crowdsourcing*. La question est naturellement à aborder avec certaines précautions du point de vue de la qualité des métadonnées, car les bibliothèques souhaitent conserver leur image de marque et leur capital de confiance en n'exposant que des données certifiées de par leur provenance. Rien n'interdit pour autant de lier l'information de référence produite par une bibliothèque à des informations produites directement par les usagers pourvu que la provenance des données soit clairement tracée et indiquée à l'utilisateur afin qu'il puisse identifier par lui-même le statut des informations. Ainsi, pour son service data.bnf.fr, la BnF n'a pas hésité à créer dès 2011 de multiples liens entre les données de ses catalogues et celles de l'Encyclopédie en ligne Wikipédia, estimant que les internautes étaient en mesure de distinguer par eux-mêmes ces deux types d'information mais apprécieraient que la bibliothèque leur propose ces liens entre deux réservoirs de données très complémentaires. De telles approches contribuent à placer les données de bibliothèque au cœur de l'écosystème du web.

## CONCLUSION

En 1974, le CBU actait qu'une bibliothèque ne pouvait décrire à elle seule tous les savoirs du monde, et qu'elle avait besoin de coopérer avec les autres bibliothèques (le modèle mythique de l'Alexandrina avait vécu). En 2014, il faut acter que les bibliothèques, même unies, ne peuvent pas non plus décrire tous les savoirs. Elles ont impérativement besoin de coopérer avec le reste du monde. L'ambition initiale du CBU – faire connaître à chacun ce qui a été produit partout dans le monde moyennant un partage international des tâches – nous semble toujours d'actualité. Cette ambition rencontre un idéal de démocratisation culturelle et d'égalité d'accès aux ressources aussi bien que les nécessités d'une réalité économique de plus en plus contrainte pour les agences de l'Etat. Toutefois, cette ambition s'est largement répandue à la faveur de la révolution numérique et des possibilités nouvelles qui se sont

ouvertes du point de vue des technologies comme des usages. Avec l'avènement de l'internet, on peut dire que ce qui semblait un rêve de bibliothécaires il y a 40 ans est devenu le rêve de tous.

Dans l'état actuel des technologies et des standards, le web sémantique s'impose résolument comme le meilleur environnement possible pour repenser la collaboration et l'ouverture des bibliothèques. Les possibilités de partage ou de co-production de données se démultiplient en effet avec l'apparition de ce nouvel environnement. Le web sémantique permet à des organisations de travailler dans des systèmes et des formats de production très différents tout en mutualisant leurs ressources en ligne, sous forme de liens, lorsque des complémentarités apparaissent bénéfiques à toutes les parties. Ce nouveau paradigme ouvre la voie au décloisonnement de nombreux secteurs d'activité jusqu'ici maintenus en silos. Pour les bibliothèques, cette nouvelle opportunité est, selon nous, le point nodal depuis lequel les idéaux du CBU peuvent être réinventés.

## Remerciements

Les auteurs souhaitent remercier **Mélanie Roche** pour sa traduction de cet article en langue anglaise, ainsi que **Gordon Dunsire** et **Mirna Miller**, dont l'ouvrage<sup>xxvi</sup> référencé à la fin de la bibliographie a été une précieuse source d'inspiration et de documentation pour la préparation de cette communication.

## Références

---

<sup>i</sup> Anderson, Dorothy. *Universal Bibliographic Control: a Long Term Policy, a Plan for Action*. PullachbeiMünchen : VerlagDokumentation, 1974

<sup>ii</sup> « Principes de Paris » adoptés par la Conférence internationale sur les Principes de catalogage, Paris, Octobre 1961  
[http://imeicc5.com/download/french/Principes\\_de\\_Paris\\_French.pdf](http://imeicc5.com/download/french/Principes_de_Paris_French.pdf)

<sup>iii</sup> IFLA Professional Statement on Bibliographic Universal Control, Décembre 2012  
<http://www.ifla.org/files/assets/bibliography/Documents/ifla-professional-statement-on-ubc-en.pdf>

<sup>iv</sup> Ressource intégratrice : ressource augmentée ou modifiée par des mises à jour qui ne restent pas distinctes mais qui sont intégrées à l'ensemble de la ressource (cf . Glossaire de l'ISBD édition intégrée, 2013 pour la traduction française).

<sup>v</sup> « La situation du dépôt légal de l'Internet en France : retour sur cette nouvelle législation, sur sa mise en pratique depuis cinq ans, et perspectives pour le futur », ILLIEN G., SANZ P., SEPETJAN S., STIRLING P. In: *Actes du 77e congrès de la Fédération internationale des associations de bibliothécaires et d'institutions (IFLA)*, San Juan (Porto Rico), août 2011.  
<http://conference.ifla.org/past/ifla77/193-stirling-fr.pdf>

<sup>vi</sup> Consortium international pour la préservation de l'internet (IIPC)  
<http://www.netpreserve.org>  
Pour une présentation en français de l'histoire et des activités du consortium IIPC, voir aussi:  
Illien, Gildas. Une histoire politique de l'archivage du web. *Bulletin des bibliothèques de France*, n° 2, 2011.  
<http://bbf.enssib.fr/consulter/bbf-2011-02-0060-012>

<sup>vii</sup> Proceedings of the National Bibliographies Seminar, Brighton, 18 August 1987, held under the auspices of the IFLA Division of Bibliographic Control / ed. by Winston D. Roberts. -London : IFLA UBCIM Programme, 1988.  
International Congress on National Bibliographies, Copenhagen, 25-27 November 1998  
<http://www.ifla.org/publications/international-congress-on-national-bibliographies-icnbs-recommendations>

<sup>viii</sup> Form and Structure of Corporate Headings. London : IFLA, 1980

<sup>ix</sup> Names of persons : National Usages for Entry in Catalogues, compiled by A H Chaplin, Dorothy Anderson. Sevenoaks (Kent), International Federation of Library Associations, 1967

<sup>x</sup> International List of Uniform Headings for Anonymous Classics, compiled by R. Pierrot, 1964 (2<sup>nd</sup> ed. 1978)

<sup>xi</sup> Svenonius, Elaine. *The intellectual foundation of information organization*. Cambridge, Mass.: The MIT Press, 2000 (p.51)

<sup>xii</sup> [http://www.bnf.fr/documents/frbr\\_rapport\\_final.pdf](http://www.bnf.fr/documents/frbr_rapport_final.pdf)

<sup>xiii</sup> [http://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-fr.pdf](http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-fr.pdf)

- 
- <sup>xiv</sup> Guidelines for Authority Records and Entries. London : IFLA, 1984 ((2<sup>nd</sup> ed. 2001)
- <sup>xv</sup> [http://www.bnf.fr/documents/frad\\_rapport\\_final.pdf](http://www.bnf.fr/documents/frad_rapport_final.pdf)
- <sup>xvi</sup> [http://www.bnf.fr/documents/frsad\\_rapport\\_final.pdf](http://www.bnf.fr/documents/frsad_rapport_final.pdf)
- <sup>xvii</sup> Clack, Doris Hargrett. Authority Control: Principles, Applications, and Instructions. Chicago & London : American Library Association, 1990.
- <sup>xviii</sup> <http://www.archive.org>
- <sup>xix</sup> Pour une présentation plus détaillée des idées présentées dans cette partie, voir : Illien, Gildas. Interview with Gildas Illien, Director, Bibliographic and Digital Information Department, Bibliothèque nationale de France (BnF). *Information Standards Quarterly*, Winter 2013, 25(4): 22-29. [http://www.niso.org/apps/group\\_public/download.php/11938/IP\\_Illien\\_BnF\\_isqv25no4.pdf](http://www.niso.org/apps/group_public/download.php/11938/IP_Illien_BnF_isqv25no4.pdf)
- <sup>xx</sup> WorldCat  
<http://www.worldcat.org/>
- <sup>xxi</sup> Virtual International Authority File (VIAF)  
<http://viaf.org/>
- <sup>xxii</sup> International Standard Name Identifier (ISNI) International Authority  
<http://www.isni.org/>
- <sup>xxiii</sup> W3C. Rapport final du groupe d'incubation Bibliothèques et web de données. 25 octobre 201. Traduction française :  
<http://mediatheque.cite-musique.fr/MediaComposite/ARTICLES/W3C/XGR-ld-fr.html>
- <sup>xxiv</sup> Informatisation des noms de personnes, familles et collectivités : le schéma XML EAC-CPF (Contexte archivistique encodé - Collectivités, personnes, familles)  
<http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/eac/>
- <sup>xxv</sup> Sur la proposition de création de registres ouverts de métadonnées, voir la proposition n° 79 du rapport Lescure : Lescure, Pierre. Mission Acte 2 de l'Exception culturelle. Contribution aux politiques culturelles à l'ère numérique. Mai 2013.  
[http://www.culturecommunication.gouv.fr/var/culture/storage/culture\\_mag/rapport\\_lescur/index.htm](http://www.culturecommunication.gouv.fr/var/culture/storage/culture_mag/rapport_lescur/index.htm)
- <sup>xxvi</sup> "Bibliographic Information Organization: a view from now into the past". In Miller, Mirna & Dunsire, Gordon. *Bibliographic Information Organization in the Semantic Web*. Oxford : Chandos Publishing, 2013 (Chandos Information Professional Series), pp. 1-85