



VIAF : un pivot pour l'accès multilingue à diverses collections

Françoise Bourdon

et

Vincent Boulet

Bibliothèque nationale de France
Département de l'information bibliographique et numérique
Paris, France

Meeting:

79 — Tackling the challenges of multilingualism in the arts: catalogues, databases, digital collections and other material in the global context — Art Libraries Section

Résumé :

Avec le développement du Web sémantique, les données d'identification d'une personne, d'une collectivité ou d'une œuvre prennent une importance jusqu'alors inégalée dans le passé. Les données d'autorité établies par les bibliothèques pour gérer les points d'accès aux notices décrivant les documents recensés dans leurs catalogues ont une réputation de fiabilité, et sont considérées dans le cadre du Web sémantique comme étant des données de confiance.

Le texte ci-dessous explique en quoi le Fichier d'autorité international virtuel (VIAF), porté par des bibliothèques nationales et de grandes institutions culturelles, contribue à cette émancipation et en quoi il permet l'interopérabilité entre des données conçues selon des règles différentes, dans des langues et des écritures différentes ? Quelques exemples d'utilisation de VIAF dans des réalisations internationales, européennes ou nationales mettront en valeur des nouveaux usages des données d'autorité.

Introduction

Avec le développement du Web sémantique, les données d'identification d'une personne, d'une collectivité ou d'une œuvre prennent une importance jusqu'alors inégalée dans le passé. Établies au départ dans les notices d'autorité conçues par les bibliothèques pour gérer les points d'accès aux notices décrivant les documents recensés dans leurs catalogues, ces données connaissent aujourd'hui une nouvelle vie sur la Toile. Les données structurées contenues dans une notice d'autorité se prêtent aisément aux traitements automatiques : la forme autorisée du nom de l'entité concernée, ses variantes de forme, les formes qui lui sont associées, et d'autres éléments d'information comme la nationalité de l'entité ou ses dates (dates biographiques pour une personne, dates d'activité pour une collectivité ou date de création pour une œuvre), etc. Leur réputation de fiabilité place d'emblée ces données parmi les données de confiance.

Pourquoi et comment les "données" d'autorité doivent-elles et peuvent-elles se libérer du carcan des "notices" d'autorité pour vivre une nouvelle vie sur la Toile ? En quoi le Fichier d'autorité international virtuel, porté par des bibliothèques nationales et de grandes institutions culturelles, contribue-t-il à cette émancipation ? En quoi permet-il l'interopérabilité entre des données conçues selon des règles différentes, dans des langues et des écritures différentes ? Quels sont dès à présent les grands projets qui s'appuient sur cet outil pour développer de nouveaux usages des données d'autorité ?

1. Pourquoi un Fichier d'autorité international ?

Selon les principes du Contrôle bibliographique universel¹ exposés dans les années 1970, chaque agence bibliographique nationale doit non seulement recenser toute la production éditoriale du pays dans une bibliographie nationale officielle, mais elle doit aussi établir "la forme d'autorité du nom des auteurs nationaux". La diffusion de ces formes d'autorité dans les bibliographies nationales était censé inciter tous ceux qui réutilisaient les notices bibliographiques à réutiliser aussi ces points d'accès autorisés. C'était sans compter avec les usages nationaux qui affectent la structure des noms, leur langue ou leur écriture et qui rendent impossible l'utilisation de la même forme d'un nom à travers le monde entier. D'ailleurs dès 1963 l'IFLA publie la première édition de *Names of Persons : National Usages for Entry in Catalogues*². Aujourd'hui l'utilisateur final est au cœur de nos préoccupations et il est indispensable de lui permettre d'utiliser, lors de ses recherches d'information, la langue, l'écriture et les formes des noms qui ont sa préférence. Comment concilier cohérence et flexibilité des accès dans un même catalogue ? Le fichier d'autorité est l'outil le mieux adapté pour gérer cette apparente contradiction.

L'une des fonctions d'un catalogue est de rassembler pour chaque auteur toutes ses œuvres mais uniquement celles-ci. Il est donc nécessaire d'établir pour chaque auteur une forme de son nom qui « fait autorité » dans un catalogue donné, en tenant compte des attentes des usagers. En fait plusieurs facteurs viennent complexifier ce schéma. D'une part les usagers d'un catalogue constituent le plus souvent un groupe hétérogène aux besoins variés, voire contradictoires. D'autre part, privilégier la forme d'un nom suppose que les formes alternatives de ce nom puissent être aussi utilisées comme point d'accès au catalogue (sous forme de renvois d'exclusion « voir » ou d'orientation « voir aussi ») pour permettre à un plus grand nombre d'utilisateurs d'aboutir dans leurs recherches. Même si à l'origine de sa conception un catalogue donné a sa logique propre et des points d'accès cohérents, qu'en est-il après les opérations de conversions rétrospectives et les chargements divers qu'il peut connaître au fil du temps ? Les notices bibliographiques, et par conséquent les points d'accès qui y sont attachés, sont depuis plusieurs décennies dérivées d'un catalogue dans un autre, et ce changement de contexte s'accompagne aussi d'une relative perte de cohérence. Pour surmonter toutes ces difficultés dans la gestion des points d'accès aux catalogues des bibliothèques nationales et aux bibliographies nationales officielles, les agences bibliographiques nationales ont développé chacune un fichier d'autorité. La Bibliothèque du Congrès a été pionnière dans ce domaine en gérant un fichier d'autorité manuel dès la fin du XIXe siècle. Il a été informatisé depuis ! Aujourd'hui la grande majorité des fichiers d'autorité qui participent à VIAF sont nés au cours des 30 dernières années et sont bien sûr

¹ Anderson, Dorothy. *Universal Bibliographic Control: a Long Term Policy, a Plan for Action* (München: Saur, 1974)

² La 4ème édition publiée en 1996 est en cours de révision. Elle n'est plus disponible sous forme imprimée mais elle l'est sous forme d'un fichier PDF : <http://www.ifla.org/node/4953>

tous informatisés. Rappelons que les *Principes internationaux de catalogage*³ publiés début 2009 stipulent que "Des notices d'autorité doivent être créées pour contrôler les formes autorisées des noms, les variantes de formes des noms et les identifiants utilisés comme points d'accès" (§6.1.1.1). Le Fichier d'autorité international virtuel va plus loin, en mettant en relation le contenu de chacun des fichiers d'autorité nationaux des bibliothèques partenaires.

2. Qu'est-ce que VIAF ?

VIAF est un **fichier d'autorité**, dont le périmètre actuel couvre les noms de personne et les noms de collectivité. La prise en compte des noms géographiques est en test, et celle des titres d'œuvre est en projet. Par contre il semble exclu de couvrir les concepts, parce qu'ils relèvent d'une autre logique.

VIAF est un **fichier d'autorité international** :

Le premier contrat de partenariat a été signé en 2003, lors du congrès IFLA de Berlin, entre la Bibliothèque du Congrès (LoC), Die Deutsche Bibliothek et OCLC Online Computer Library Center⁴. La BnF signe le contrat en 2007. Depuis, les partenaires se sont rapidement multipliés. En 2010, chaque mois en moyenne un nouveau partenaire a rejoint le projet.

En juin 2011 on compte dix-neuf partenaires pour vingt-deux fichiers traités. Les partenaires se répartissent de la manière suivante :

Statuts des partenaires	Nombre
Bibliothèques nationales	12 (+ 2 en test ⁵)
Catalogues collectifs nationaux ou régionaux	3 ⁶
Bibliothèques de rang international	1 ⁷
Institution en dehors du monde des bibliothèques	1 ⁸

L'objectif premier de VIAF est de créer des liens entre les fichiers d'autorité des bibliothèques nationales. Un des considérants de l'accord de coopération actuellement en vigueur précise : « *Un tel fichier d'autorité international virtuel serait une extension concrète du concept de contrôle bibliographique universel qui serait établi à partir du travail effectué par chaque agence bibliographique nationale* ».

Depuis, VIAF a inclus dans ses rangs d'autres institutions ou catalogues d'importance nationale ou internationale, tels que des catalogues collectifs ou des grands établissements d'influence nationale et internationale. Il faut également noter que VIAF élargit son spectre de partenaires au-delà du cercle des bibliothèques. L'entrée du fichier du Getty est un jalon

³ http://www.ifla.org/files/cataloguing/icp/icp_2009-fr.pdf (consulté le 2011-06-01)

⁴ Le communiqué de presse annonçant la naissance du projet VIAF est daté du 26 août 2003 (<http://www.oclc.org/research/news/2003-08-26.htm>, consulté le 2011-06-01). Les racines de VIAF sont à chercher dans un accord datant de 1998, signé entre la Bibliothèque du Congrès et la Deutsche Nationalbibliothek permettant la réutilisation des notices d'autorité noms de personnes de la LC (en format USMARC à l'époque) par la Bibliothèque nationale allemande dans *Personennamendatei*.

⁵ La Bibliothèque du Congrès, les bibliothèques nationales d'Australie, de la République tchèque, de France, d'Allemagne, d'Israël, du Portugal, d'Espagne, de Suède, de Suisse, de Hongrie et la Bibliothèque vaticane. Bibliothèque et archives Canada et la Bibliothèque d'État de Russie sont en phase de test.

⁶ L'Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane et per le Informazioni Bibliografiche (ICCU, Italie), Norodowy Uniwersalny Katalog Centralny (NUKAT, Pologne) et le Réseau des Bibliothèques de Suisse occidentale (RERO, Suisse).

⁷ Bibliothèque d'Alexandrie.

⁸ Institut de recherche du Getty.

important dans l'élargissement en dehors du monde des bibliothèques *stricto sensu*, en direction des grandes institutions d'art.

L'origine géographique des partenaires actuels de VIAF est la suivante (en comptant les institutions en cours de test) :

Régions du monde	Nombre
Amérique du nord	3
Amérique du sud	0
Europe occidentale	4
Europe orientale	9
Afrique du nord, Proche et Moyen-Orient	2
Océanie	1
Asie	0

Le centre de gravité historique de VIAF est placé du côté des institutions « occidentales » (Amérique du Nord et pays anglo-saxons, Europe occidentale). L'ouverture vers d'autres régions du monde et, partant, d'autres traditions de catalogage est un défi dont les partenaires sont conscients. L'arrivée des fichiers d'autorité, souvent riches, d'Europe orientale, se poursuit. Mais l'enjeu principal est la capacité de VIAF d'intégrer des institutions du domaine oriental : Afrique du Nord, Proche et Moyen-Orient, Asie. L'absence de partenaires asiatiques et sud-américains est un défi important à relever.

Le noyau initial de VIAF s'est donc considérablement élargi et les partenaires sont aujourd'hui très divers, tant du point de vue de leur statut que de leur origine géographique.

3. VIAF : comment ça marche ?⁹

VIAF repose sur les relations qui existent entre un fichier bibliographique donné et le fichier d'autorité qui le pilote. Pour une même entité (personne, collectivité, etc.) la forme autorisée du nom dans la notice d'autorité est reportée à l'identique dans les notices bibliographiques décrivant les ressources sur lesquelles cette entité exerce une responsabilité intellectuelle, artistique ou commerciale. Dans la notice bibliographique, la forme autorisée peut être accompagnée de l'identifiant de la notice d'autorité d'où elle provient. Dans un catalogue donné, ces relations peuvent s'exprimer par des liens hypertextes permettant à l'utilisateur de naviguer du fichier d'autorité vers le fichier bibliographique et vice versa. Pour apparier les noms de personne ou de collectivité, VIAF s'appuie à la fois sur les fichiers bibliographiques et les fichiers d'autorité, et par conséquent, chaque partenaire doit fournir à OCLC ces deux types de fichier.

Pour chaque paire de notices formée par le tandem notice d'autorité/notice bibliographique, VIAF va créer une notice d'autorité dérivée. En effet, certains éléments de données appartenant à la notice bibliographique constituent des éléments complémentaires d'identification de l'entité décrite dans la notice d'autorité liée. Les informations suivantes sont extraites (quand elles sont présentes) de chaque notice bibliographique: titre, noms de tous les contributeurs, nom de l'éditeur commercial, ISBN, type de ressource, langue de la ressource, lieu et date de publication, indices de classification, identifiant de la notice bibliographique,

⁹ Rick Bennet, Christina Hengel-Dittrich, Edward T. O'Neill, Barbara B Tillett *VIAF (Virtual International Authority File) : un pont entre les fichiers d'autorité noms de personnes de Die Deutsche Bibliothek et de la Bibliothèque du Congrès*, 2006. http://archive.ifla.org/IV/ifla72/papers/123-Bennett_trans-fr.pdf (consulté 1er juin 2011)

etc. Ces informations sont stockées dans la notice d'autorité correspondante, dans des champs MARC 9XX définis expressément pour les besoins de VIAF. Le contenu de chaque champ 9XX est normalisé pour faciliter les traitements ultérieurs : pour les données textuelles, toutes les diacritiques et les ponctuations sont supprimées et les majuscules sont remplacées par des minuscules ; les années sont transformées en décennies (1963 devient ainsi 196X), etc. A l'issue de cette opération, on obtient une "notice d'autorité dérivée" contenant la notice d'autorité de départ (avec son identifiant) et les champs 9XX contenant les informations extraites de la notice bibliographique. Pour une entité donnée ce même traitement est appliqué à toutes les notices bibliographiques qui lui sont liées dans le fichier bibliographique du partenaire, et le processus est répété pour chaque personne ou collectivité figurant dans une notice bibliographique donnée, à l'exception des points d'accès sujet.

A terme, on obtient un lot de notices dérivées pour chaque entité et pour chaque partenaire. Ces notices sont fusionnées par fichier d'origine pour donner, pour chaque partenaire une notice d'autorité enrichie. Pour chaque élément d'information provenant des notices bibliographiques, le nombre d'occurrences dans le fichier du partenaire est indiqué dans une sous-zone \$9.

L'étape suivante consiste à apparier entre elles les notices d'autorité enrichies provenant de chaque partenaire, selon un algorithme mis au point par l'équipe de recherche d'OCLC. Cet algorithme s'appuie entre autre sur les occurrences figurant dans les \$9 mentionnés ci-dessus. Le résultat d'un appariement est une grappe de notices d'autorité enrichies : c'est ce nœud qui porte l'identifiant VIAF. A l'affichage, le nœud VIAF présente :

- l'identifiant VIAF,
- toutes les formes autorisées du nom de la personne ou de la collectivité, établies par les partenaires. Chaque forme est suivie de la mention des partenaires qui l'ont établie (le drapeau de leur pays est représenté). Un graphique permet aussi de visualiser de façon dynamique comment les appariements ont été faits.
- toutes les variantes de forme (les formes d'exclusion *Voir*) utilisées au moins par l'un des partenaires,
- toutes les formes associées (les formes d'orientation *Voir aussi*) utilisées au moins par l'un des partenaires,
- les pays où les œuvres de la personne ou de la collectivité ont été publiées. Un planisphère permet de visualiser le résultat global
- des statistiques sur l'étalement dans le temps des publications de l'entité, rendues facilement compréhensibles grâce à un graphique,
- les principaux éditeurs commerciaux qui ont publié les œuvres de l'entité,
- le sexe, la nationalité, la langue d'expression la plus utilisée par l'entité,
- des liens externes vers d'autres ressources telles que Wikipédia ou WorldCat Identities,
- des liens vers la représentation de la notice VIAF en UNIMARC, en MARC21 XML, et en RDF.

Tous les champs d'une notice VIAF sont interrogeables. Mais on peut limiter la recherche à un ensemble de champs comme Tous les noms, Toutes les collectivités, Toutes les personnes, ou encore Toutes les formes préférées, Nom exact, Titres des œuvres. On peut en plus limiter la recherche aux données d'un seul partenaire. A partir d'une notice VIAF on peut visualiser les notices enrichies de chaque partenaire à partir desquelles a été construite la grappe VIAF. Seules les données de la grappe sont déchargeables, et ce sont surtout les notices en RDF qui sont porteuses d'avenir. Une photographie de l'ensemble des notices VIAF à un instant donné,

un "dump", peut être fourni, sur demande auprès du Service Recherche d'OCLC, à ceux qui souhaitent l'incorporer dans une application qui doit elle-même être libre d'accès sur le Web.

4. VIAF et le dialogue des cultures

Le fait d'établir des nœuds entre différents fichiers d'autorité permet de dépasser les choix de catalogage locaux. Ces derniers dépendent du cadre culturel de chaque pays. De ce cadre culturel découlent les règles nationales ou régionales de construction des noms, le choix des formes du nom en fonction des publics de chaque pays, la langue et l'écriture des formes autoritées, le calendrier utilisé pour indiquer les dates biographiques et, de manière plus générale, la langue de catalogage. Chaque fichier d'autorité national est établi à partir des cadres culturels de la majorité de ses publics propres et des catalogueurs qui établissent les notices. VIAF permet de dépasser les particularismes régionaux ou nationaux en les faisant dialoguer les uns avec les autres.

Dépasser les usages culturels ne signifie pas les annihiler. Au contraire, en refusant d'établir une seule forme autorisée valable pour tous, VIAF respecte les cadres culturels d'origine et se garde de les hiérarchiser les uns par rapport aux autres. VIAF ne se place pas au-dessus des fichiers d'origine mais dresse des ponts entre eux.

Formes du nom : formes nationales, formes internationales ?

Une des manifestations du pluralisme culturel est la diversité des pratiques de catalogage. Cela implique d'importantes différenciations dans le choix et la structure des formes autorisées, voire des finalités des fichiers d'autorité d'origine.

L'exemple d'Andrej Rublev¹⁰ illustre une telle diversité parmi les partenaires du projet VIAF : nous prendrons l'exemple des notices de la BnF¹¹, de la Bibliothèque Nationale d'Israël et de la Bibliothèque du Congrès. Chacun de ces établissements utilise des pratiques de catalogage, des formats et une structure de catalogue différents.

	Notice BnF	Notice Bibliothèque du Congrès	Notice BN Israël (fichier cyrillique russe)
Format	Unimarc	MARC 21	MARC 21
Liens aux notices bibliographiques	Oui	Non	Oui
Formes autorisées	3 formes retenues : Andreï Roublev (Saint ; 1360?-1430?) <i>forme courante français</i> Andrej Rublev (Saint ; 1360?-1430?) <i>forme internationale translit.-ISO russe</i> Андрей Рублев (Saint ; 1360?-1430?) <i>forme internationale russe</i>	1 forme retenue translittérée selon le système LoC : Rublev, Andrej, Saint, d. ca. 1430	1 forme retenue en russe cyrillique Рублев, Андрей, Святой, ум. ок. 1430
Caractères non latins	Une des trois formes autorisées	Forme rejetée, par dérivation automatique	Forme retenue
Notes biographiques	En français	Indiquées dans les zones de source, en anglais	Pas de notes
Sources	En caractères latins	En caractères latins	Pas de sources

On voit donc que chaque pays choisit en première forme autorisée, ou en seule forme autorisée, celle qui correspond le mieux aux usages en vigueur chez lui. La BnF a choisi plusieurs formes autorisées, de statut équivalent. Ce sont des formes parallèles. La première forme autorisée est la forme traditionnelle en France, la plus conforme aux usages de la majorité des utilisateurs des catalogues de l'établissement. Les deux autres formes autorisées sont des formes russes. Elles sont données en caractères originaux, ici en cyrillique, et en translittération selon le système ISO 9:1995. Les formes parallèles permettent de gérer la multiplicité des usages possibles des données et des utilisateurs. Un étudiant ou un chercheur en histoire de l'art n'utilisera pas la même forme du nom qu'un slaviste.

¹⁰ Peintre russe d'icônes du début du XVème siècle.

¹¹ <http://catalogue.bnf.fr/ark:/12148/cb11950441k>

Le défi du multilinguisme

La gestion d'une grande diversité de formes du nom établies dans des bibliothèques de pays différents, selon des normes et des traditions de catalogage différentes pose la question du multilinguisme. C'est aujourd'hui une question centrale pour les catalogues des bibliothèques¹². La question du multilinguisme répond à trois objectifs fondamentaux des catalogues exprimés par le code international de catalogage¹³. D'une part, la question du « confort de l'utilisateur du catalogue » : « les décisions prises en matière de descriptions et de formes contrôlées des noms dans les points d'accès doivent l'être en tenant compte de l'utilisateur ». En second lieu, le principe d' « usage commun » : « le vocabulaire utilisé dans les descriptions et les accès doit correspondre à celui de la majorité des utilisateurs ». Et en troisième lieu le critère de « représentativité » : « les descriptions et les formes contrôlées des noms doivent reposer sur la façon dont une entité se décrit elle-même ».

Ces différents aspects entrent en résonance avec un des considérants de VIAF qui est de « permettre la coexistence de variations nationales ou régionales ce qui, par là, permet d'assurer les variations de langue, d'écriture et d'orthographe retenues selon les besoins des utilisateurs tout autour du monde ».

C'est ainsi que les langues de catalogage varient d'un pays à l'autre. Les qualificatifs de fonctions sont exprimés dans différentes langues. Le qualificatif « saint » d'Andrej Rublev est exprimé en anglais dans le catalogue de la Bibliothèque du Congrès, en français dans celui de la BnF et en russe (Святой) dans le fichier cyrillique de la bibliothèque nationale d'Israël. Les qualificatifs de date peuvent être introduits par des éléments linguistiques qui diffèrent d'un pays à l'autre. La Bibliothèque du Congrès indique en anglais « d. ca. 1430 » ; le fichier cyrillique de la bibliothèque nationale d'Israël précise « у.м. ок. 1430 » (traduction exacte de « d. ca. 1430 »). La BnF indique les dates incertaines par un point d'interrogation, sans élément linguistique (« 1430 ? »).

On voit que VIAF permet de relier ces données multilingues entre elles, sans établir de hiérarchie entre elles.

Le défi du pluralisme des écritures

Pour être traité entièrement, la question du multilinguisme doit inclure celle du pluralisme d'écritures. Les bibliothèques nationales occidentales, au début de leur informatisation, ont été incapables de gérer les caractères non latins. C'est dès lors la forme translittérée qui a fait office de forme originale du nom. Depuis, la BnF a progressivement introduit les caractères non latins dans les formes autorisées du fichier d'autorité. Ensuite, la Bibliothèque du Congrès a commencé à réinjecter les caractères non latins dans ses notices, en forme rejetée.

L'arrivée dans VIAF de partenaires dont le catalogage natif est en caractères non latins pose la question sous un angle nouveau. Il n'y a aucune raison pour que l'alphabet latin occupe une place centrale dans un projet international, plus qu'une autre écriture. En utilisant UNICODE pour le codage des caractères, VIAF est capable de prendre en compte un grand nombre d'écritures.

La diversité des calendriers

Les qualificatifs de date sont essentiels dans les fichiers d'autorité. Afin de respecter au mieux les usages culturels d'origine, VIAF doit être capable de traiter le pluralisme des calendriers qui apparaissent dans les différents fichiers d'autorité des bibliothèques partenaires.

¹² Voir la réunion plénière de la Section Catalogage, IFLA 2010 <http://www.ifla.org/en/conferences-sessions/216> (session 93) sur le thème "Multilingual bibliographic access : promoting universal access"

¹³ http://www.ifla.org/files/cataloguing/icp/icp_2009-fr.pdf (consulté le 2011-06-01)

Pour donner un exemple, Averroes apparaît dans la majorité des bibliothèques d'Europe et du monde anglo-saxon sous sa forme de notoriété en Occident avec des dates selon l'ère chrétienne (Averroës, 1126-1198). Cependant, à la Bibliothèque d'Alexandrie, il apparaît sous la forme arabe de son nom et avec un qualificatif de date construit selon le calendrier de l'Hégire (ابن رشد، محمد بن أحمد، 595-520 هـ). Le fichier d'autorité arabe de la Bibliothèque nationale d'Israël indique lui une forme arabe du nom plus complète que celle de la Bibliothèque d'Alexandrie et donne les dates selon l'ère chrétienne en graphie arabe des chiffres arabes (ابن رشد، محمد بن احمد، ابو الوليد، ١١٢٦-١١٩٨ م). L'algorithme de rapprochement de VIAF permet qu'une telle différence n'empêche pas l'appariement des notices d'autorité et donc de respecter les différentes ères en vigueur.

Le processus d'appariement de VIAF permet de dépasser les choix de catalogage dus aux identités culturelles tout en respectant ces dernières et surtout en ne les hiérarchisant pas les unes par rapport aux autres. Mieux, en permettant à des données si différentes de dialoguer, VIAF met en valeur les différentes traditions de catalogage, et, au-delà, les différentes manières d'attribuer un nom aux individus. Il s'agit donc d'un véritable outil de mise en évidence de la diversité culturelle qui apparaît dans les différents catalogues. Cela souligne la complémentarité des données de sources différentes, produites selon des règles culturelles différentes.

5. VIAF dans le web sémantique : un tremplin entre données

L'utilisation des données d'autorité de la BnF comme outil pour le web sémantique est présente dès les débuts du projet. Comme le précise l'accord de coopération initial signé entre les fondateurs de VIAF : *« Les propositions actuelles pour l'avenir du web décrivent l'utilisation d'ontologies pour rendre le web plus intelligible pour des processus automatisés par machines. Le VIAF pourrait être une des briques fondatrices pour un « web sémantique » s'il est combiné à d'autres vocabulaires contrôlés et d'autres fichiers d'autorité provenant de sources telles que des services de dépouillement et d'indexation, des archives, des musées, des éditeurs etc. Les bibliothèques ont désormais une opportunité pour apporter une importante contribution à cette perspective et pour contribuer faire de cette vision une réalité ».*

Les technologies du web sémantique

Les notices VIAF sont exposables et utilisables en RDF (*Ressource description framework*), qui est le format cadre du web sémantique. Le principe de RDF est d'exprimer une information sous forme de triplets. Par exemple : « Andrej Rublev est le sujet du film Andrej Rublev ». « Le film Andrej Rublev a été réalisé par Andrej Tarkovskij ». Les technologies du web sémantique sont fondées sur des identifiants pérennes assignées à des ressources (URI). RDF et les URI permettent de donner un langage commun à l'ensemble des ressources du web sémantique. Cela permet d'établir des liens entre elles en disant « Andrej Rublev exprimé dans la notice VIAF sous tel identifiant pérenne est équivalent à Andrej Rublev décrit dans DBpedia sous tel identifiant pérenne ». VIAF est donc utilisable comme rebond vers d'autres ressources.

Les utilisations de VIAF

Les opportunités offertes par les technologies du web sémantique permettent d'utiliser VIAF comme élément de contrôle d'autorité et d'identification des personnes et des collectivités dans différents projets. Il est également possible de concevoir VIAF comme une porte d'entrée vers les différentes ressources du web sémantique.

VIAF et l'initialisation de la base internationale des ISNI

A l'échelon international, VIAF va d'ici quelques semaines servir à l'initialisation de la base des ISNI (International Standard Number Identifier = Code international normalisé des noms). L'ISNI, défini par la norme ISO 27729, permet d'identifier tous ceux qui contribuent à un bien culturel – qu'ils soient artistes, créateurs, producteurs ou éditeurs. L'ISNI peut être attribué à toutes les parties prenantes qui créent, produisent, gèrent, distribuent ou figurent dans un contenu créatif, qu'il s'agisse de personnes, d'entités légales (comme les organismes) ou de personnages fictifs. Dans la base de données ISNI, la partie prenante est identifiée par son « identité publique », le nom sous lequel elle est publiquement connue. L'ISNI servira donc à identifier les ayants droit dans l'univers du numérique. C'est un code d'identification unique pour toute partie prenante impliquée dans différents domaines de création (musique, cinéma, arts visuels, littérature, etc.).

ISNI est conçu comme une passerelle entre d'une part les différents systèmes propriétaires d'identification d'ayants droit, tel que les Interested Party Identifiers (IPI) utilisés par les membres de la CISAC¹⁴ ou les International Performer Numbers (IPN) utilisés par l'IPDA¹⁵, et d'autre part des outils de découverte de ressources d'information, tel que VIAF. Les données présentes dans toutes ces bases de données, dont celles de VIAF, vont être utilisées pour créer la base de données initiale des ISNI. Les métadonnées ISNI relieront l'identité publique à toutes les autres manifestations qui lui sont associées dans les différents systèmes, permettant aux partenaires d'échanger des informations sur les contributeurs concernés sans révéler d'informations confidentielles. Dans quelques semaines, l'identifiant ISNI devrait figurer dans VIAF, à côté de l'identifiant VIAF et devrait être facilement récupérable dans les fichiers d'autorité des partenaires VIAF.

VIAF et le projet européen Europeana Regia

Europeana Regia est un projet mené par la Commission européenne dont le but est de reconstituer de manière virtuelle des collections de manuscrits aujourd'hui dispersées dans diverses bibliothèques européennes : les manuscrits carolingiens, les manuscrits de la Bibliothèque de Charles V au Louvre, les manuscrits des rois aragonais de Naples. Ces documents sont numérisés. Les formats de description utilisés par les bibliothèques partenaires sont divers (EAD, MARC, TEI). La nécessité pour un projet européen d'une description multilingue a incité à utiliser VIAF comme référentiel d'autorité pour les noms des auteurs des œuvres. Les langues de description dans Europeana Regia sont : l'anglais, le français, l'allemand, l'espagnol, l'italien, le catalan et le néerlandais. Chaque auteur sera identifié par l'identifiant pérenne qui est attribué à sa notice dans VIAF. Cet identifiant sera récupéré à partir de la représentation RDF de VIAF.

« Europeana Regia » est un exemple d'utilisation de VIAF pour la gestion d'un contrôle d'autorité multilingue.

¹⁴ Confédération Internationale des Sociétés d'Auteurs et Compositeurs

¹⁵ International Performers' Database Association

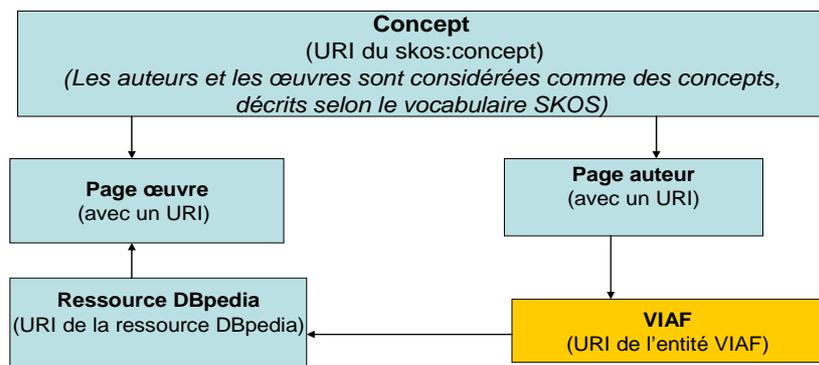
VIAF et un projet national : le projet *data.bnf.fr* de la Bibliothèque nationale de France

Le projet *data.bnf.fr*, dont un prototype sera en ligne à l'été 2011, est mené par la BnF. Son objectif est d'offrir un point d'accès unique à l'ensemble des ressources documentaires de l'établissement dispersées dans différentes bases (notices bibliographiques en MARC, description de fonds d'archives et de manuscrits en EAD, bibliothèque numérique « Gallica », expositions virtuelles...) et de faciliter leur référencement par les moteurs de recherche.

Le principe est de créer des « pages auteur » qui reprennent le contenu des notices d'autorité Personnes établies par la BnF et, de même, des « pages œuvre » qui reprennent le contenu des notices d'autorité Titres. Ces pages « auteur » et « œuvre » sont complétées par le signalement des manifestations, issues des descriptions bibliographiques, de manuscrits et de fonds d'archives.

Data.bnf.fr applique le modèle de données FRBR et les technologies du web sémantique, en RDF.

En outre, le projet utilise VIAF comme tremplin vers d'autres ressources, comme DBpedia, de la manière suivante :



Ce « chemin » permet d'intégrer dans les pages « auteur » ou « œuvre » des ressources issues de DBpedia, telles que d'autres formes du nom de l'auteur ou du titre de l'œuvre, ou encore des renseignements biographiques et une description de l'œuvre. Cela permet donc de compléter les données déjà présentes dans les notices d'autorité de la BnF.

Data.bnf.fr est un exemple d'utilisation de VIAF comme rebond vers d'autres ressources.

Ce ne sont que quelques exemples de l'utilisation de VIAF. Il reste que l'utilisation la plus courante est celle qu'en font les catalogueurs pour assurer au quotidien le contrôle d'autorité dans leur catalogue local. En 2010 il y a eu un peu plus de 171 000 visites sur le site et un million de pages vues. Pour un panorama plus complet des utilisations, il faudrait interroger les partenaires actuels de VIAF sur ce qui a motivé leur contribution au projet, et au-delà l'ensemble de ceux qui ont manifesté leur intérêt auprès du Service Recherche d'OCLC pour demander l'autorisation de réutiliser les données des nœuds VIAF dans leurs propres applications.

Conclusion

VIAF permet de construire des ponts entre des bases de données hétérogènes par les contextes culturels dans lesquels elles sont élaborées, par leurs architectures propres et par la diversité de leurs contenus, en servant de pivot pour les accès personnes et collectivités. VIAF permet de gérer à l'identique, sans les hiérarchiser, des données de langues et d'écritures différentes, respectant ainsi les contextes culturels d'origine des personnes ou des collectivités décrites aussi bien que les attentes des utilisateurs à travers le monde.

En confrontant les données des différents fichiers d'autorité sources, VIAF contribue à les consolider en faisant ressortir les incohérences qui doivent être corrigées à la source, et par là à accroître leur fiabilité. Les données de VIAF prennent ainsi une forte valeur ajoutée, très recherchée dans le Web sémantique où la notion de confiance est primordiale.

Les données de VIAF seront d'autant plus fiables que les contributeurs et les utilisateurs seront nombreux.

L'essayer c'est l'adopter !