



PDF, CSV, RSS et autres acronymes : redéfinition des services bibliographiques à la Bibliothèque nationale d'Allemagne

Lars G. Svensson
&
Yvonne Jahns
Deutsche Nationalbibliothek
Frankfurt, Germany

Traduit de l'anglais vers le français par :
Mathilde Koskas
Bibliothèque nationale de France
Paris, France

Meeting: 91. Bibliography

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY
10-15 August 2010, Gothenburg, Sweden
<http://www.ifla.org/en/ifla76>

Résumé :

En janvier 2010, la Bibliothèque nationale d'Allemagne a interrompu la version imprimée de la bibliographie nationale, et l'a remplacée par un journal en ligne. C'était la première étape d'un plus vaste processus de redéfinition des services bibliographiques de la Bibliothèque nationale, pour quitter le domaine des supports traditionnels (par exemple le papier ou les bases de données sur CD-ROM) et se concentrer sur la publication des données sur le Web. Un nouveau modèle économique a été mis en place : toutes les ressources Web sont maintenant publiées dans une série supplémentaire de la bibliographie, et les données bibliographiques sont disponibles gratuitement. Graduellement, les prix des autres services bibliographiques seront également réduits.

Le deuxième stade du projet se concentre sur les services à valeur ajoutée fondés sur le catalogue de la Bibliothèque nationale. Le but principal est l'introduction de services d'alerte fondés sur les critères de recherche de l'utilisateur en offrant plusieurs moyens d'accès comme des fils RSS, la compatibilité avec Zotero, par exemple, ou l'export de données bibliographiques sous forme de fichier CSV ou PDF.

Les normes de catalogage actuelles demeurent les directives pour offrir une récupération de qualité à l'utilisateur final, mais elles seront complétées par des procédures d'indexation automatique pour trouver et parcourir des documents en nombre croissant. Une politique de catalogage transparente et des menus de sélection bien organisés sont les objectifs visés.

1. Un nouveau modèle économique

Un nouveau modèle économique a été mis en place en 2010 : toutes les séries de la bibliographie étaient disponibles gratuitement en ligne via le catalogue de la bibliothèque, mais en plus à présent les données bibliographiques en PDF sont gratuites pour nos utilisateurs.

1.1 La collecte par la Bibliothèque nationale

Le mandat de collecte de la bibliothèque remonte à 1912, et s'appuie aujourd'hui sur la Loi concernant la Bibliothèque nationale d'Allemagne¹. Cette réglementation légale du dépôt légal en Allemagne a été révisée en profondeur en 2006. Les spécifications du mandat de collecte sont expliquées plus en détail et précisées par le Règlement de Dépôt légal (Pflichtablieferungsverordnung 2008) et nos directives pour les collections (Sammelrichtlinien)². Le rôle de ces directives est de préciser quelles publications allemandes et étrangères sont à collecter, la quantité toujours croissante de publications et le grand nombre de publications en ligne causant des restrictions par critères formels. Pour donner une idée du volume, environ 100 000 nouveaux titres ont été mis sur le marché du livre par les éditeurs en 2009. En tout, la Bibliothèque nationale a enregistré environ 640 000 publications l'année dernière.

Depuis 1913, la Bibliothèque nationale d'Allemagne collecte :

- les ouvrages publiés en Allemagne
- les ouvrages publiés en allemand à l'étranger
- les traductions, dans d'autres langues, d'ouvrages allemands publiés à l'étranger
- les Germanica, c'est-à-dire les ouvrages sur l'Allemagne publiés à l'étranger dans d'autres langues
- les imprimés écrits ou publiés entre 1933 et 1945 par des émigrés germanophones

La collecte comprend des ouvrages sur divers supports, physiques et non physiques, c'est-à-dire des publications sur papier mais aussi des microformes, des enregistrements sonores, des ouvrages physiques sur des supports de données électroniques et des publications en ligne. La base de données en ligne de la bibliothèque, <https://portal.d-nb.de/>, fournit l'accès au catalogage original et remplit ainsi la mission légale de la bibliothèque de donner la description bibliographique de toutes les publications sorties en Allemagne. Notre situation est assez spéciale sur le plan international, puisque la collection de la bibliothèque et son catalogue en ligne sont presque équivalents à la bibliographie nationale.

En plus du catalogue en ligne, la Deutsche Nationalbibliothek publie la Bibliographie nationale allemande (Deutsche Nationalbibliographie) qui décrit périodiquement les nouvelles acquisitions de la bibliothèque. La Bibliographie nationale est constituée de manière traditionnelle de différentes séries aux périodicités variées et avec différents types d'accès sujets³ :

- Série A – Monographies et périodiques du marché de l'édition. Livres, magazines, enregistrements non musicaux, autres supports audiovisuels, microfiches et publications électroniques. À partir de 2004, cette série comprend aussi les traductions et les Germanica qui jusque là paraissaient séparément dans la Série G. Publication hebdomadaire.

¹ <http://bundesrecht.juris.de/dnbg/index.html>

² <http://files.d-nb.de/pdf/sammelrichtlinien.pdf>

³ Pour les accès sujets, voir partie 3.

- Série B – Monographies et périodiques hors marché de l'édition. Livres, magazines, enregistrements non musicaux, autres supports audiovisuels, microfiches et publications électroniques. Publication hebdomadaire.
- Série C – Cartes. Publication trimestrielle.
- Série H – Travaux universitaires. Thèses et travaux postdoctoraux des universités allemandes et thèses et travaux postdoctoraux en allemand des universités étrangères. Publication mensuelle.
- Série M – Musique imprimée. Publication mensuelle.
- Série O – Publications en ligne. Publication mensuelle.
- Série T – Enregistrements musicaux. Publication mensuelle.

1.2. La Bibliographie nationale allemande sous forme de journal en ligne à partir de 2010

Jusqu'en 2009, la Deutsche Nationalbibliographie était publiée sous forme de journal imprimé et plus tard sur CD-ROM. La Deutsche Nationalbibliothek ne publiait pas la bibliographie toute seule ; cela était fait par le MBV (Marketing- und Verlagsgesellschaft des Buchhandels) qui était chargé de la publication, des ventes et de la distribution. En raison des coûts de production élevés de la bibliographie imprimée et de la baisse significative du nombre d'abonnés – dont une des causes était l'augmentation de la fourniture électronique de données bibliographiques par la DNB – le MVB a décidé d'interrompre la publication sous forme imprimée après un historique de publication de plus d'un demi-siècle ; c'est également le cas pour la version sur CD-ROM. Ainsi, les abonnements à des séries particulières de la Deutsche Nationalbibliographie sous forme imprimée ou CD-ROM ont cessé en 2009. De nouveaux services électroniques ont remplacé ces versions et reflètent maintenant l'évolution des besoins des utilisateurs.

À partir de 2010, la base de données de la Deutsche Nationalbibliographie est universellement accessible via le catalogue en ligne <https://portal.d-nb.de/>.

The screenshot shows the search interface of the Deutsche Nationalbibliothek. At the top, there is a navigation bar with links for English, Kontakt, A-Z, Förderer, Datenschutz, Impressum, Hilfe, and Mein Konto. The main header reads 'KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK'. Below this, there are tabs for 'Gesamter Bestand', 'Deutsches Musikarchiv', and 'Deutsches Exilarchiv'. The search area includes several dropdown menus for 'Titel', 'Person', 'Nummer', and 'Schlagwörter', each followed by a search box and a 'und' dropdown. There is also a 'Jahr' field with a text input and a '?' icon, with an example: 'zum Beispiel: 1948-1980 oder 1976- oder 1955'. Below this, there is a section for 'Neuerzeichnet in der Deutschen Nationalbibliografie in den Reihen:' with checkboxes for A, B, C, G, H, M, T, O, N, and a '→ Alle → Keine ?' option. At the bottom, there is a 'von' field, a calendar icon, 'bis (optional):', another calendar icon, and a 'Finden →' button. On the right side, there is a note: 'Browsen in der SWD S Beta Version.'

Fig.1 <https://portal.d-nb.de/> Écran de recherche avec les options de filtrage sur les séries et les années de la bibliographie.

La Bibliothèque nationale espère que cela satisfera les besoins des clients qui souhaitent une livraison à intervalles précis des données ajoutées récemment ou qui n'ont pas besoin d'une forme d'édition particulière. Le catalogue comporte des options supplémentaires comme l'application de filtres sur chacune des séries de la bibliographie, la recherche de volumes de la bibliographie, la sélection de types de supports particuliers, comme les publications en ligne ou les thèses, la sauvegarde de profils de recherche ou l'exportation de listes de résultats dans des fichiers HTML. Des fils RSS et des services d'alerte sont également prévus.

Certains groupes de clients ont toutefois demandé expressément la poursuite de la publication périodique d'abrégés des nouveaux supports dans les collections de la bibliothèque. Pour faciliter la transition, nous avons refait l'ancienne édition imprimée de la bibliographie sur le modèle d'un journal purement en ligne publié au format PDF. Par rapport à la version précédente, son contenu a été quelque peu simplifié. La division en séries A, B, C, H, M et T et le nouveau service de livraison demeurent inchangés. La périodicité de publication n'a pas non plus été altérée. Une nouvelle série O pour les publications en ligne a été créée (voir 2).

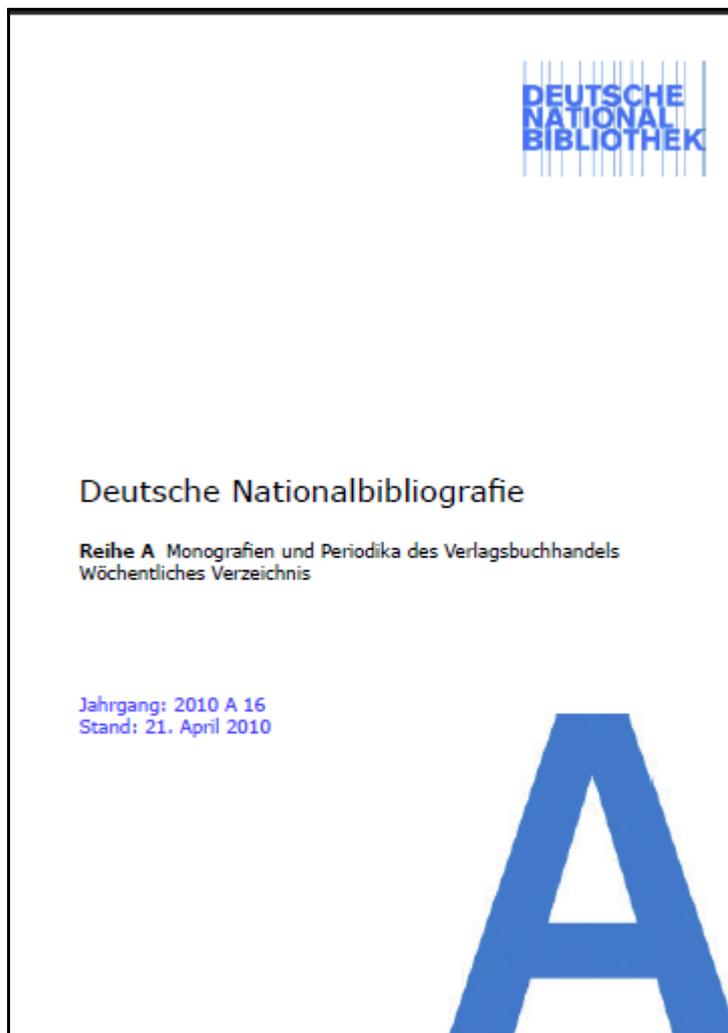


Fig.2 Maquette du journal en ligne

L'utilisation des possibilités offertes par la publication sous forme électronique était une condition importante de la réorganisation de la bibliographie sous forme de journal en ligne. Par conséquent, toutes les descriptions bibliographiques contiennent des liens hypertexte vers les notices correspondantes du catalogue en ligne. Ces liens hypertexte fournissent l'accès à la description bibliographique en ligne mais aussi à du contenu à valeur ajoutée tel que tables des matières, résumés, jaquettes ou plein texte.

2010, A16	<000>	Deutsche Nationalbibliografie
<p>000 Allgemeines, Wissenschaft</p> <p><000> XA-DE-BY http://d-nb.info/995823227 ∞ 09.N35</p> <p>Großer, Hartmut: Top secret : Geheimprojekte und Technologien der Supermächte / Hartmut Großer. - Marktberdorf : Argo, 2010. - 271 S. : Ill., graph. Darst., Kt. ; 23 cm. - Literaturangaben. - <i>Inhaltsverzeichnis</i> - ISBN 978-3-937987-87-3 Pp. : EUR 24.00 - EAN 9783937987873 <i>SW: UFO ; Geheimunternehmen</i> <i>DDC: 001.942</i></p> <p><000;610;150> XA-CH http://d-nb.info/1001405730 ∞</p> <p>Huter, Carl: Der Gebildete auf der Höhe : gesundes Urteil und starke Nerven / Carl Huter. Bearb., eingeleitet, mit Ill. vers. und hrsg. von Fritz Aerni. - Zürich : Carl-Huter-Verl., 2010. - 182 S. : Ill. ; 20 cm. - Literaturangaben - ISBN 978-3-03741-122-3 Pp.</p> <p>004 Informatik</p> <p><004;650;S> XA-DE-NW http://d-nb.info/958233527</p> <p>Basiswissen IT-Berufe [Medienkombination]. - Troisdorf : Bildungsverl. EINS. - Teilw. im Verl. Stam, Köln <i>SW: Informationstechnik ; Beruf ; Berufsschule ; Schulbuch</i> <i>DDC: 004.6</i> http://d-nb.info/999106864 10,N03</p>	<p>nitz: Tagungsband / Chemnitzer Linux-Tage 2010 : 13. und 14. März 2010 / Team der Chemnitzer Linux-Tage ; Technische Universität Chemnitz. - Chemnitz : Univ.-Verl., 2010. - 136 S. : graph. Darst. ; 21 cm. - Literaturangaben. - <i>Inhaltsverzeichnis</i> - ISBN 978-3-941003-10-1 kart. : EUR 15.10 <i>SW: LINUX ; Kongress ; Chemnitz <2010></i> <i>DDC: 005.432</i></p> <p><004;330> XA-DE-HE http://d-nb.info/999102338 ∞ 10,N03</p> <p>Deck, Klaus-Georg: Java-Grundkurs für Wirtschaftsinformatiker : die Grundlagen verstehen - objektorientierte Programmierung - fortgeschrittene Konzepte kennenlernen - betriebswirtschaftlich orientierte Beispiele ; [mit Online-Service] / Klaus-Georg Deck ; Herbert Neudorf. - 2., aktualisierte und verb. Aufl. - Wiesbaden : Vieweg + Teubner, 2010. - XII, 456 S. : graph. Darst. ; 24 cm. - (Studium) - Literaturangaben. - <i>Inhaltsverzeichnis</i> - <i>Inhaltstext</i> - ISBN 978-3-8348-1222-3 kart. : EUR 31.95 - EAN 9783834812223 <i>SW: Java <Programmiersprache> ; Wirtschaftsinformatik</i> <i>DDC: 005.133</i></p> <p><004> XA-DE http://d-nb.info/1000571130 ∞</p> <p>Grill, Ortrun: Überzeugend präsentieren mit PowerPoint 2007 / Ortrun Grill ; Charlotte von Braunschweig. - 1. Ausg. - Bodenheim ; Wien : Herdt-Verl. für Bildungsmedien ; Dübendorf : Herdt-Verl. Schweiz, 2010. - 172 S. : Ill., graph.</p>	<p><004> XA-DE http://d-nb.info/1000058611 ∞</p> <p>Peyton, Christine: Internet & E-Mail : [mit Windows 7 ; Sehen & Verstehen - das visuelle Praxisbuch] / Christine Peyton. - Düsseldorf : Data-Becker, [2010]. - 284 S. : Ill. ; 24 cm. - (Genial einfach!) - <i>Inhaltsverzeichnis</i> - ISBN 978-3-8158-1904-3 kart. : EUR 14.95 <i>SW: Internet ; Rageber tt E-Mail ; Rageber</i> <i>DDC: 004.678 tt 004.692</i></p> <p><004> XA-DE http://d-nb.info/1000571491 ∞</p> <p>Raith, Michael: Windows 7 : Systembetreuer : Workstation / Michael Raith. - 1. Ausg. - Bodenheim ; Wien : Herdt-Verl. für Bildungsmedien ; Dübendorf : Herdt-Verl. Schweiz, 2010. - 189 S. : Ill. ; 29 cm. - (Herdt Classics) - <i>Inhaltsverzeichnis</i> - kart. <i>SW: Windows 7</i> <i>DDC: 005.44</i></p> <p><004> XA-DE http://d-nb.info/1000876144 ∞</p> <p>Reuscher, Dominik: PC & Windows 7 für Einsteiger : [Sehen & Verstehen ; das visuelle Praxisbuch] / Dominik Reuscher. - Düsseldorf : Data-Becker, [2010]. - 288 S. : Ill. ; 19 cm. - (Genial einfach!) - <i>Inhaltsverzeichnis</i> - ISBN 978-3-8158-1902-9 kart. : EUR 14.95 <i>SW: Personalcomputer ; Windows 7</i> <i>DDC: 005.446</i></p> <p><004> XA-DE-HE http://d-nb.info/999819429 ∞ 10,N05</p> <p>Binmale, Holmar: Smart clients mit Eclipse</p>

Fig.3 Maquette et structure des notices bibliographiques dans le journal en ligne avec les URI et les liens hypertexte vers les tables des matières et les résumés

Nous avons pris en considération d'autres points : l'accessibilité pour les utilisateurs handicapés visuels et la conformité pour la préservation numérique à long terme. Lors du développement de la version PDF, la possibilité de consommer les contenus en utilisant des lecteurs d'écran ou d'autres technologies d'assistance a été un critère important. Les volumes produits sont conformes à la plupart des critères d'accessibilité et le processus est actuellement en cours de révision afin de le rendre entièrement accessible. La conformité pour la préservation numérique à long terme était un autre critère important. Sous le nouveau mandat, la Deutsche Nationalbibliothek collecte toutes les publications numériques publiées en Allemagne, y compris les siennes. Pour garantir la conformité des documents à la préservation à long terme, la version PDF de la bibliographie nationale respecte la norme PDF/A-1⁴. Cette norme précise la structuration requise pour que les documents PDF soient reproductibles même dans un futur éloigné.

En plus de ce journal en ligne accessible gratuitement au format PDF, qui s'adresse plus particulièrement aux utilisateurs finaux, les données bibliographiques sont proposées sous forme de fichiers de données en MARC 21 et MAB, de fichier titres imprimé et via le magasin de données. Le magasin de données permet aux utilisateurs de sélectionner et de télécharger des données bibliographiques dans des formats structurés et s'adresse en

⁴ <http://www.pdfa.org/>

particulier aux clients de la bibliothèque qui veulent réutiliser les données bibliographiques fournies par la Deutsche Nationalbibliothek.

Jusqu'ici, il semble que le nouveau journal en ligne soit bien accepté par nos clients. Nous avons été assez surpris qu'il y ait encore une demande importante pour la version imprimée. Pour y répondre, nous offrons maintenant des tirages papier des fichiers PDF. À part cela, il n'y a eu que des demandes secondaires. On nous a demandé d'augmenter la taille des caractères dans la section comportant les données bibliographiques, ce qui n'est pas vraiment nécessaire puisqu'un des avantages des publications électroniques est la possibilité pour les utilisateurs de le faire eux-mêmes lors de leur lecture sur écran. Une autre question a été de savoir pourquoi nous avons retiré la numérotation continue des entrées de chaque volume. La numérotation aidait à l'origine à retrouver une notice bibliographique au moyen d'un index (cumulatif). Dans la mesure où l'index cessé de paraître il y a un certain temps et où les utilisateurs peuvent rechercher les informations bibliographiques dans le catalogue en ligne, une numérotation explicite n'est plus nécessaire.

D'autres services, complémentaires, ont été élaborés spécialement en prenant en compte les besoins des clients.

1.3. De nouveaux services d'accès aux données bibliographiques

En plus des accès traditionnels aux données bibliographiques, la Deutsche Nationalbibliothek étudie également de nouveaux canaux d'accès et de distribution. Ces nouveaux services se concentrent sur les chercheurs et les étudiants qui ont besoin de trouver, de rassembler et de conserver des données documentaires et qui veulent être tenus au courant des nouvelles publications dans leur domaine de recherches.

Un des points d'attention est la compatibilité avec des logiciels de gestion bibliographique. Pour la première étape, les descriptions bibliographiques du catalogue en ligne contiendront à partir de l'été 2010 des métadonnées encodées en format COinS afin de permettre à nos usagers d'utiliser l'extension Zotero de Firefox pour télécharger et conserver les données. La compatibilité avec d'autres services, par exemple RefWorks et EndNote, est prévue.

Autre nouveau service : la possibilité de mémoriser une recherche et de souscrire au fil RSS des résultats de cette recherche. Cela permettra aux clients spécialisés d'utiliser une recherche personnalisée comme « publications de la série A sur la chimie (SG 540) » ou « notices contenant de la musique de Beethoven » pour recevoir des mises à jour sur les publications récentes directement dans leur agrégateur de flux. Le service ne fournit pas des notices bibliographiques complètes mais se concentre sur le cœur de la description bibliographique en utilisant le format OAI_DC⁵.

De plus, il sera bientôt possible d'exporter un résultat de recherche sous forme de fichier PDF similaire à la version PDF de la bibliographie nationale. Ce service est destiné aux bibliothécaires et aux autres clients qui veulent travailler avec un sous-ensemble de données bibliographiques particulier et préfèrent lire l'information par titres dans un format semblable à l'ISBD. Pour les bibliothécaires qui recherchent les nouvelles publications sur un sujet précis, ce service est particulièrement utile combiné aux fonctions de recherche avancée qui permettent à l'utilisateur d'effectuer une recherche dans une ou plusieurs séries bibliographiques, tout en la limitant à une fourchette de dates précise. Cette option déduit automatiquement de l'intervalle de dates le bon volume de la bibliographie nationale, en prenant en compte le fait que toutes les séries de la bibliographie n'ont pas la même

⁵ <http://www.d-nb.de/hilfe/katalog/datenshop.htm#daten>

périodicité : par exemple, pour rechercher des monographies sur Stravinsky et des partitions de lui publiées entre janvier et mars 2010, l'utilisateur devrait prendre en compte le fait que les séries A et B sont hebdomadaires, alors que la série M (partitions) n'est que mensuelle ; la recherche doit donc couvrir les volumes 1 à 13 (2010) des séries A et B, et les volumes 1 à 3 (2010) de la série M, ce qui peut être difficile à formuler dans une recherche. L'option de recherche par date soulage l'utilisateur de cette tâche.

1.4. Coûts et conditions actuels des données bibliographiques

La Deutsche Nationalbibliographie sous forme de journal en ligne au format PDF est disponible via le catalogue en ligne. En d'autres termes, le nouveau modèle économique donne la possibilité à chacun de se procurer toutes les séries et les données bibliographiques gratuitement en PDF.

Les utilisateurs ne paient les données qu'en cas de services de fourniture ou de sélection particuliers.

On peut acheter et télécharger les fichiers PDF sur le Web ou sur le serveur FTP de la Deutsche Nationalbibliothek. Le coût s'élève à 5 € par téléchargement.

Nous fournissons également le contenu PDF sur papier (relié et sur fiches perforées) pour répondre à la demande des usagers. Le prix dépend de la taille moyenne d'un volume, qui varie entre les séries (la série A est en général beaucoup plus volumineuse que la série C, par exemple), et va de 4 à 14 € par livraison selon les séries (y compris les frais de port pour l'Allemagne).

Les données bibliographiques sont fournies en MAB, UNIMARC ou MARC 21. Les séries de la Deutsche Nationalbibliographie et le Nouveau Service de Publication sont également disponibles en version HTML en ISBD.

Après six mois de mise en œuvre, nous nous rendons compte que ce système de prix est accepté par nos clients. La Deutsche Nationalbibliothek ouvre progressivement l'accès gratuit à ses données, et à terme les données par titre et les autorités seront accessibles gratuitement, sauf quelques exceptions comme l'usage commercial ou les ensembles de données préparés par le personnel de la DNB selon des critères de recherche personnalisés lorsque cela requiert un travail d'exploitation. Dans la mesure où la Deutsche Nationalbibliothek ne publiait pas la bibliographie imprimée elle-même, il est difficile de mesurer l'impact budgétaire de son interruption. La bibliothèque renonce à sa commission sur les ventes de la bibliographie imprimée ; en raison de la baisse des ventes dans les derniers temps, celle-ci ne représentait pas une somme importante.

En considérant le modèle économique des données dans sa totalité, il faut souligner que l'acquisition des données d'autorité est également gratuite depuis mars 2010. Les notices d'autorité sont livrées en MARCxml et MABxml-1 par protocole SRU (recherche et récupération par URL). Cela permet la réutilisation facile de toutes les données. À partir de mai 2010, la DNB distribue aussi les données d'autorité en Linked Open Data selon les formats et les protocoles du Web sémantique.

2. Les ressources Web dans la bibliographie nationale

Pour la première fois, il existe une série supplémentaire de la Bibliographie nationale allemande pour les ressources Web. La création des données bibliographiques est entièrement automatisée.

2.1 Collecte des publications en ligne à la Bibliothèque nationale allemande

La Bibliothèque nationale collecte les publications en ligne depuis 1998 sur la base du volontariat, en particulier les thèses de doctorat publiées en ligne. Lors de l'entrée en vigueur de la nouvelle loi concernant la Deutsche Nationalbibliothek en 2006, le mandat de la bibliothèque a été étendu. En plus de son devoir de collecter les imprimés, elle a reçu la tâche de collecter, cataloguer, indexer et archiver les ouvrages dématérialisés.

La collecte concerne tous les ouvrages texte, image et sons mis à disposition sur des réseaux publics. L'obligation de dépôt concerne à la fois les publications sur le Web qui ont des versions imprimées et les ouvrages dont le support est sur le Web. Cela inclut par exemple des magazines électroniques, des e-books, des thèses universitaires, du contenu numérisé, des fichiers musicaux et aussi des sites Web. Les prépublications temporaires, les outils purement logiciels ou applicatifs, les programmes de radio et de télévision ne sont pas collectés. Les publications en ligne à but purement privé ou commercial non plus.

Les procédures adaptées à la collecte, au catalogage et à l'archivage à grande échelle des publications en ligne seront développées progressivement. L'objectif de l'étape actuelle est la collecte des publications isolées en ligne comme les e-books. La procédure actuelle comprend par exemple l'extension des canaux de distribution existants, la mise en oeuvre d'interfaces de dépôt différenciées pour le dépôt de publications en ligne à la pièce et l'extension des formats actuellement acceptés.

La Bibliothèque nationale vise à rendre le dépôt de publications en ligne aussi pratique que possible. Il y a actuellement deux possibilités pour déposer des publications en ligne : dépôt par un formulaire Web et dépôt utilisant un système de moissonnage OAI-PMH. Le formulaire Web convient aux publications en ligne en petit nombre, alors que le moissonnage en OAI-PMH est la meilleure solution pour les plus grandes quantités.

Pour l'instant, aucun site Web ne peut être collecté et conservé numériquement de façon régulière. Cependant, nous avons récemment entrepris un projet spécial d'analyse des méthodes et des flux de moissonnage pour la collecte des sites Web et des applications dynamiques.

Des ensembles de données de base ont été définis pour les différents types de publications⁶. Les définitions contiennent un ensemble minimal d'éléments de métadonnées pour la livraison de métadonnées de documents électroniques. On donne ici l'ensemble de base des éléments de métadonnées pour les monographies qui sont déposées via formulaire Web. Pour le moissonnage automatique, l'ensemble minimal sera moins étendu.

- URL de la ressource Web, lorsqu'elle est accessible (obligatoire)
- Information sur le contenu : code de la Bibliographie nationale, adapté de la classification décimale Dewey ou code du système du marché du livre (obligatoire)

⁶ http://www.d-nb.de/eng/netzpub/ablief/kernset_metadaten.htm#header

- Type de ressource Web (monographie, thèse, périodique) (obligatoire)
- Date de publication (obligatoire)
- Information sur les droits : accès libre ou réservé à la publication originale et à l'exemplaire de conservation (obligatoire)
 - Titre (obligatoire)
 - Éditeur (obligatoire)
 - Lieu de publication (obligatoire)
 - Information sur le contenu : autre classification ou thésaurus (facultatif)
 - Nom du numéro/de l'édition (facultatif)
 - Auteur (facultatif)
 - Notes (facultatif)
 - Organisations (facultatif)
 - Relations hiérarchiques entre œuvres (œuvre complète) (facultatif)
 - Note de thèse et université (facultatif)
 - Langue de la publication (facultatif)

2.2 Série O de la bibliographie

En ce qui concerne les publications en ligne, la Bibliothèque nationale d'Allemagne a deux objectifs principaux. Le premier est le catalogage automatique des publications en ligne d'après les métadonnées collectées (voir 3.3). Le second est l'utilisation d'identifiants pérennes pour assurer la permanence de l'adresse et la citation à long terme des publications en ligne. La Bibliothèque nationale d'Allemagne attribue et administre les URN pour l'espace de noms « urn:nbn:de » et offre un service de résolution d'URN pour l'Allemagne, l'Autriche et la Suisse. Toutes les ressources en ligne archivées par la Bibliothèque nationale d'Allemagne reçoivent une URN de l'espace de noms « urn:nbn:de ». Par rapport aux URL, les URN offrent une identification unique et permanente des objets numériques, indépendamment de leur lieu de stockage. C'est un facteur important pour maintenir la préservation à long terme.

Depuis 2010, tous les nouveaux dépôts de publications en série ont été enregistrés dans la Bibliographie nationale, dans une série supplémentaire O. En avril 2010, la série O contenait déjà environ 20 700 titres. En raison de ce nombre important, il n'y a pas de version PDF pour ces nouvelles publications. On peut toutefois y faire des recherches par le biais du catalogue, et les usagers pourront, s'ils le préfèrent, créer un livret personnalisé avec les descriptions au format ISBD.

3. Normes de catalogage et politique d'indexation

Les normes de catalogage actuelles seront les lignes directrices pour offrir aux utilisateurs finaux des possibilités de récupération de haute qualité. Une politique d'indexation transparente et cohérente est aussi importante que l'application de méthodes d'indexation automatique pour trouver des documents en nombre croissant.

3.1 Traitement des supports et politique d'indexation

Comme nous l'avons déjà évoqué ci-dessus, la Deutsche Nationalbibliothek remplit son rôle de centre national de l'information bibliographique pour l'Allemagne sur la base des nouvelles acquisitions par dépôt légal. Selon le principe de la répartition des tâches, le traitement des supports a lieu une seule fois pour les deux sites, Francfort et Leipzig. Les éditeurs fournissent deux exemplaires de dépôt des monographies et périodiques au site qui en est chargé. La responsabilité est déterminée selon une logique régionale : le site de Leipzig

de la Bibliothèque nationale d'Allemagne traite toutes les publications des états fédéraux de Berlin, Brandebourg, Mecklembourg-Poméranie occidentale, Rhénanie du Nord-Westphalie, Saxe, Saxe-Anhalt et Thuringe. Le site de Francfort a la responsabilité des états de Bade-Wurtemberg, Bavière, Hesse, Basse-Saxe, Rhénanie-Palatinat, Sarre et Schleswig-Holstein. Après traitement, le second exemplaire est fourni « prêt au magasinage » à l'autre bibliothèque pour utilisation et archivage. De plus, les exemplaires uniques d'ouvrages publiés en allemand à l'étranger les traductions d'originaux allemands non soumis comme exemplaires de dépôt légal, et les ouvrages en langues étrangères concernant l'Allemagne sont traités à Leipzig. Le traitement de la musique imprimée et des enregistrements sonores est effectué exclusivement au département Deutsches Musikarchiv (DMA) – encore à Berlin en août 2010 ; à partir de novembre 2010, le DMA fera partie du site de Leipzig de la Deutsche Nationalbibliothek.

Toutes les publications sont cataloguées selon les Règles allemandes de catalogage alphabétique (RAK). Le catalogage sujet s'appuie sur les Règles allemandes pour le catalogue sujets (RSWK) et la Classification décimale Dewey (CDD). De plus, la Bibliothèque nationale d'Allemagne utilise pour les sujets, les personnes et les organisations, des fichiers d'autorités pour lesquels elle joue un rôle capital de développement et de maintenance. La future utilisation de Ressource Description and Access (RDA) est prévue et changera les règles actuelles comme la mise en œuvre du format international de données MARC21 l'a déjà fait.

La politique d'indexation actuelle est publiée en ligne⁷. Chaque publication est affectée à au moins une classe sujet (Sachgruppe) de la Bibliographie nationale allemande. On utilise pour le catalogage des publications diffusées sur le marché de l'édition (série A de la bibliographie) des mots-clefs fondés sur les règles RSWK. On utilise pour cela les fichiers d'autorité sujets personnes et organisations (SWD, PND et GKD)⁸. En plus de l'indexation sujets, il y a aussi depuis 2007 une classification fondée sur la CDD. Presque toutes les publications reçoivent une annotation CDD complète. Les manuels scolaires, la fiction et les livres pour enfants ne sont pas inclus. Les thèses médicales reçoivent une classification CDD de niveau intermédiaire.

Exemple de notice de la Bibliographie nationale		Groupement sujet (d'après les Cent Divisions de la CDD)	
Numéro titre Identifiant ISO 9706 Identifiant ISO 3166-1/-2 (provenance/éditeur)	070 Nachrichtenmedien, Journalismus, Verlagswesen ...	Numéro du Nouveau Service de Fourniture	
Nombre de pages, dimensions	0110 ∞ XA-DE-BW 03,N07,0075 Lau, Matthias: Pressepolitik als Chance : staatliche Öffentlichkeitsarbeit in den Ländern der Weimarer Republik / Matthias Lau. – Stuttgart : Steiner, 2003. – 441 S. ; 24 cm – (Beiträge zur Kommunikationsgeschichte; Bd. 14) ◇ Zugl.: Berlin, Freie Univ., Diss., 2002 ◇ ISBN 3-515-08071-6 kart.: EUR 78,00 sfr 124,80 – EAN 978515080712	Auteur: titre	
ISBN, reliure	SW: Deutschland; Pressepolitik; Geschichte 1918-1933 (213) ◇ Weimarer Republik; Pressepolitik DBN 96.231958.9	Éditeur, lieu et date de publication	
Numéro d'identifiant	Prix, EAN (European Article Number)	Vedette sujet

Fig. 4 Exemple de notice bibliographique

⁷ http://www.d-nb.de/eng/wir/ueber_dnb/inhaltser schl.htm

⁸ <http://www.d-nb.de/eng/standardisierung/index.htm>

3.2 Structure de la bibliographie nationale

Depuis l'année bibliographique 2004, la Bibliothèque nationale d'Allemagne structure les différentes séries de la bibliographie nationale conformément à la Classification décimale Dewey (CDD).

La structure des séries A, B, C et H (voir 1.1) est fondée en grande partie sur les deux niveaux supérieurs de la CDD (les Cent divisions ou Second sommaire). Quelques différences sont faites par l'intégration du troisième niveau (les Mille sections ou Troisième sommaire) de la CDD et même du quatrième niveau à chaque fois que c'est nécessaire pour maintenir des groupes utilisés précédemment comme les sciences militaires (classe 355) ou les bandes dessinées, dessins humoristiques et caricatures (classe 741.5). En 2010, nous avons ouvert une nouvelle classe, 333.7, pour les ressources naturelles, l'énergie et l'environnement pour répondre au besoin des utilisateurs de rechercher des ouvrages sur la protection et la préservation de l'environnement.

Dans deux disciplines, plusieurs classes du Second sommaire ont été combinées pour conserver le contexte d'une discipline bien connue comme la philosophie (classe 100, comprenant les classes 100, 110, 120, 140, 160, 170, 180 et 190 de la CDD) et la théologie et le christianisme (classe 230, comprenant les classes 230 à 280 de la CDD).

Les manuels scolaires et les livres pour la jeunesse sont exemptés de structuration CDD, ce qui les séparerait et les disperserait dans beaucoup de classes différentes. La fiction peut être sélectionnée séparément, mais fait partie des classes 800 à 890 selon la CDD. Ainsi trois classes supplémentaires, appelées S, K et B, peuvent être interrogées.

Les séries M (musique imprimée et publications musicales) et T (enregistrements musicaux) de la bibliographie sont également organisées selon la CDD. Des niveaux plus précis de la CDD, jusqu'à six chiffres, sont utilisés pour offrir des possibilités de récupération semblables aux structurations antérieures.

3.3 Procédures d'indexation automatique

En 2009, la Bibliothèque nationale a commencé à tester des procédures d'indexation automatique dans le but de relier le traitement conventionnel des supports et les procédures automatisées. Pour remplir la mission légale de traiter et d'indexer *toutes* les publications, il n'y a pas d'autre possibilité que de se fier aux applications logicielles. La croissance rapide du nombre de titres imprimés, aussi bien que le nombre énorme de ressources Web sous le domaine .de, nécessitent une nouvelle organisation du travail et une accélération des mécanismes de catalogage habituels. Dès que cela est possible, des procédures informatiques devraient aider les catalogueurs et permettre d'économiser les ressources humaines pour ce qui est vraiment nécessaire à l'amélioration de la récupération en ligne.

Pour le moment, nous sommes en train de tester et de comparer de nouveaux systèmes logiciels d'indexation et de classification. Au sein de différents scénarios, les bibliothécaires analysent différents supports et types de documents, le catalogage requis et la qualité de la recherche pour ces publications indexées automatiquement. Un des scénarios vise à la création automatique de la structuration par sujet de la bibliographie. C'est une sorte de système de classification automatique. Les méthodes d'indexation automatique par mots sont analysées dans un autre scénario.

Une étape supplémentaire du traitement automatique des supports est la réutilisation de données des auteurs, éditeurs ou d'autres bibliothèques et catalogues. Cela devrait être fait automatiquement par le chargement de données structurées en différents formats et l'utilisation de mappings. De plus, les données bibliographiques de différentes versions d'une œuvre, par exemple la version imprimée et la version en ligne correspondante, peuvent être copiées automatiquement.

Nous avons déjà modifié la politique de catalogage des publications en ligne en 2010. La récupération du plein texte étant possible, et les métadonnées fournies étant réutilisées, nous n'effectuons pas d'autre catalogage intellectuel pour ce type de support.

Dans le cadre de ces activités pourra apparaître un système d'indexation modulable et extensible, qui pourra faire la distinction entre différents supports et être contrôlé par des critères de qualités bien définis.