

Allemagne : le puzzle des réseaux sémantiques de données

Le fichier d'autorités intégré allemand offre aux institutions germanophones un vocabulaire fiable pour la création de réseaux sémantiques de données. Un travail de patience et de longue haleine.

La modernisation du fichier d'autorités intégré (GND dans son abréviation allemande) me fait penser à l'un de ces énormes puzzles constitués de milliers de pièces qui ne peuvent être résolus qu'au prix d'une extrême persévérance. En 2017, le *programme de développement GND* a permis de définir le cadre du puzzle et depuis lors, nous faisons tout pour remplir l'intérieur du cadre avec des secteurs qui s'intègrent lentement les uns aux autres et en ayant recours à de nombreuses activités parallèles ainsi qu'à des projets de coopération encouragés par la Fondation allemande pour la recherche (DFG). Parmi ces derniers figurent le projet GND4C¹ et le projet Orcid DE². Toutefois, nul ne sait si toutes les pièces seront présentes à la fin du travail ou si certaines manqueront. Il s'agit donc d'un puzzle destiné à des optimistes incorrigibles.

DES DONNÉES FIABLES POUR DES RÉSEAUX SÉMANTIQUES VALIDÉS

Le fichier GND offre à une communauté grandissante d'institutions culturelles et scientifiques de l'espace germanophone un vocabulaire fiable pour la création de réseaux sémantiques de données que nous considérons comme des passerelles culturelles et scientifiques. Ces passerelles stables, composées de notices d'autorité concernant des auteurs, des collectivités, Geographica, des événements, des créations intellectuelles et des termes techniques, sont un ingrédient déterminant de la colonne vertébrale d'un graphe de connaissances intersectoriel. Des concordances doivent lui permettre de relier par exemple des thésauri et des banques de données spécialisées des diverses disciplines scientifiques. Ce rôle déterminant du GND pour la mise en place d'une infrastructure de données intersectorielle est plus que jamais souhaité par les instituts de recherche. Parallèlement au lancement du programme de la Fondation allemande de la recherche concernant l'Infrastructure nationale des données de recherche (NFDI)³, la pression en ce qui concerne les attentes n'a cessé d'augmenter. De plus en plus de consortia se mettent au travail et le GND joue dans leurs projets un rôle central en tant que lien entre les différentes spécialités. Pour remplir ce rôle de passerelle, le GND doit se rapprocher des besoins de ces communautés, notamment dans le domaine des sciences humaines.

Toutefois, le GND est une œuvre collective de plus de 1000 institutions derrière laquelle se cachent des rédacteurs motivés qui contribuent, grâce à leur expertise et à leur souci de la qualité, à l'excellence et à la fiabilité des données. En cas de modifications, il faut donc convaincre de nombreuses institutions et personnes. Il faut notamment trouver un équilibre entre les intérêts des nouvelles communautés et les partenaires de longue date tout en conservant les points forts du GND : qualité, fiabilité et universalité.

ÉTAT DES LIEUX EN SIX SECTEURS CLÉS

Gouvernance : les représentants de nouvelles communautés viennent progressivement s'ajouter à la gouvernance et aux organes du GND. L'introduction de forums et de groupes de travail destinés à regrouper les besoins de chaque communauté en est un élément important. Seul l'organisme partenaire appartenant à telle communauté ou proche d'elle est à même d'en connaître les besoins et d'échanger avec elle. La mission de la Bibliothèque nationale allemande en tant que centrale GND consiste à proposer la plateforme et le cadre fiable pour cette coopération.

Rédaction commune : les nouvelles agences GND pour les communautés des archives et de la conservation et restauration des monuments historiques ont lancé leur pilote informatique. Elles se chargent des tâches de rédaction et de conseils pour ces différentes communautés.

Règles et format : le support pour des développements spécifiques aux communautés dans le modèle de données (dites *Domaines PLUS*) est sur le point d'entrer en phase de production. Les bases de données peuvent ainsi être enrichies de manière modulaire avec des informations dotées d'une fonction et d'une compétence particulières. Le noyau commun d'un enregistrement est ainsi libéré de certaines exigences trop spécifiques. En parallèle, la documentation GND est actuellement transférée sur une nouvelle plateforme technique entièrement révisée.

Base de données : la base de données est progressivement élargie en fonction des besoins des nouvelles communautés. Citons par exemple l'optimisation des normes pour les sciences littéraires grâce au projet « Titre de l'ouvrage en tant qu'espace de connaissances »⁴. Les expériences concrètes ont

[1] https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/GND4C/gnd4c_node.htm

[2] <https://www.orcid-de.org>

[3] https://www.dfg.de/en/research_funding/programmes/nfdi/index.html

[4] <https://www.klassik-stiftung.de/herzogin-anna-amalia-bibliothek/werktitel-als-wissensraum>

permis de développer un flux de travail et une liste de critères pour les notices et les enrichissements.

Mise en réseau des données et analyse : les liens à l'intérieur du GND et avec d'autres systèmes comme l'Orcid ont pu être renforcés par des procédures mécaniques et un Service des réclamations⁵. Dans le cadre d'un projet, nous travaillons à l'entretien de concordances croisées des termes du GND vers d'autres thésauri comme Rameau.

Applications et interfaces : avec le *GND-Explorer*, nous développons actuellement une application qui permettra de fouiller dans les données et de visualiser les hiérarchies et les liens sémantiques. Par ailleurs, afin de faciliter la coopération avec les nouveaux venus dans le cadre de projets de recherche, les possibilités de saisie ont été élargies grâce à de simples formulaires Web.

En outre, le logiciel Wikibase⁶ est élargi en tant que « deuxième lieu de résidence » pour le GND, ce qui va faciliter tout particulièrement la coopération avec les utilisateurs de Wikidata.

CRÉER UNE INTERFACE STANDARDISÉE

Afin que la vision d'un *graphe de connaissances intersectoriel* devienne réalité, il faut toutefois créer des conditions-cadres appropriées autour du GND. Ainsi, une infrastructure commune est nécessaire pour permettre l'utilisation facile des différentes sources de données. Plus précisément, une couche logicielle qui présente les sous-graphes (GND, base d'autorités spécialisée, banques de données spécialisées, etc.) est nécessaire, aussi bien pour l'utilisation que lors de la saisie des données. Nous imaginons un réseau de banques de données qui communiquent entre elles grâce à des interfaces standardisées. Le candidat auquel nous avons pensé pour de telles solutions est le logiciel *Wikibase*, le logiciel de banque de données derrière le projet Wikidata. Utiliser le même logiciel que la plus grande base mondiale de connaissances, nous semble logique. Malheureusement, Wikibase ne remplit pas encore toutes les exigences. Nous sommes toutefois convaincus de son potentiel⁷. C'est pourquoi nous essayons de nous mettre d'accord sur une éventuelle feuille de route avec des partenaires du programme NFDI et *Wikimedia Deutschland* (les développeurs de Wikibase).

OBTENIR DES FINANCEMENTS PÉRENNES

Afin d'aider les utilisateurs dans leurs recherches, un autre défi consiste à mettre sur pied des services de gestion des données pour les différents domaines spécifiques. Entretenir les notices d'autorité, développer des modèles spécifiques, créer des banques de données spécialisées, utiliser des méthodes d'analyse des données mécaniques : il s'agit-là de tâches complexes qui demandent de l'expérience et une expertise spécifique. La fiabilité des données



© Deutsche Nationalbibliothek - Frankfurt



ne peut être atteinte que par un financement permanent. C'est une mission sur le long terme qui, malheureusement, vitote aujourd'hui encore en grande partie dans les projets de financement du passage au numérique. Les bibliothèques universitaires et les centres de services aux bibliothèques en Allemagne se disent compétents pour apporter leur soutien aux projets de recherche concernant les notices d'autorité uniquement dans certains cas exceptionnels, ce soutien ne faisant pas explicitement partie de leur mission. Faire changer les mentalités est probablement la partie du puzzle qui va demander le plus de temps. Mais nous allons persévérer en ce sens.

L'excellent échange dans le cadre du projet très similaire de Fichier National d'Entités (FNE) de l'Abes et de la Bibliothèque nationale de France est une aide cruciale pour résoudre ce puzzle. Il existe indéniablement des points communs en termes de défis et d'ébauches de solution. Et qui sait : certaines pièces manquantes du puzzle se trouvent peut-être dans l'espace européen ?

↑ La Bibliothèque nationale allemande remplit le rôle de centrale du Fichier d'autorités intégré

[5] <https://www.dnb.de/EN/Professionell/ProjekteKooperationen/ORCID/orcid.html>

[6] <https://wikiba.se>

[7] <https://www.wikimedia.de/projects/wikilibrary-manifest>

JÜRGEN KETT
Bibliothèque nationale allemande
j.kett@dnb.de

DEUTSCHE
NATIONAL
BIBLIOTHEK