

# data.idref.fr : un référentiel d'autorités dans le web sémantique pour l'ESR et au-delà

Lancé en 2018, le triplestore data.idref.fr a progressivement étendu son périmètre au-delà des seules données de l'Abes et compte aujourd'hui plus de 6 millions de données.

**Approfondissant la stratégie d'exposition des données sur le web sémantique, le triplestore data.idref.fr a été lancé en 2018**, permettant d'interroger en SPARQL et en RDF les données d'IdRef, alimentées par les réseaux d'utilisateurs des applications de l'Abes. Dès le départ data.idref.fr a été conçu comme un miroir de la base IdRef elle-même, et non pas comme un *dump* mis à jour périodiquement comme c'est fréquemment le cas : chaque modification d'IdRef y est répercutée en temps réel. Cinq ans plus tard, il est temps de faire un bilan du chemin parcouru par ce service.

## Une base d'entités en plein essor...

Un nombre donne une première idée de l'évolution de la base : alors qu'à son ouverture elle contenait 110 millions de « triplets », c'est-à-dire de données, elle en compte désormais, fin 2023, près de 275 millions. Dans le détail, les entités sont passées de 3,5 millions à plus de 6 millions en 2023, dont près de 4 millions pour les seules personnes. Cette progression reflète pour partie les créations d'autorités dans IdRef. Mais elle est aussi le résultat d'enrichissements successifs de la modélisation. Par petites touches, toujours plus d'informations de l'Unimarc natif d'IdRef ont été extraites : libellés, genre pour les personnes, notes, identifiants externes, relations entre entités (pour les organisations notamment)<sup>1</sup>.

## ... qui est aussi une base bibliographique

D'autre part, data.idref.fr n'est pas seulement un pur référentiel d'autorités mais aussi un réservoir de références bibliographiques. À l'origine, on n'y trouvait que celles issues du Sudoc, ainsi que les thèses. Leur description, tout en restant succincte, a également été enrichie pour faciliter les recherches : précision des types de documents, dates de publication, nombre de localisations dans le Sudoc

(holdings). IdRef étant devenu au fil du temps le pivot des applications de l'Abes, data.idref.fr se devait de refléter cette centralité. Cela a été fait en intégrant les références issues d'autres applications de l'Abes : Calames, le catalogue des archives et manuscrits de l'enseignement supérieur, et SciencePlus, autre triplestore hébergeant une sélection des références d'articles et chapitres de documentation électronique provenant d'éditeurs ou diffuseurs, et où les auteurs sont identifiés à chaque fois que possible à des entités IdRef.

## data.idref.fr dans IdRef!

Plusieurs informations proposées dans IdRef illustrent le type de service que peut apporter un triplestore : pour chaque auteur les listes de ses coauteurs, des collectivités associées et de ses champs disciplinaires, ou pour une organisation les autres collectivités liées, sont fournies par des requêtes SPARQL envoyées de manière dynamique à data.idref.fr, utilisées donc comme des webservices.

## Une intégration dans un écosystème plus large

Parallèlement, IdRef et data.idref ont progressivement étendu leur périmètre au-delà des seules données de l'Abes, par des alignements (réalisés par l'Abes ou produits par les professionnels) avec des gisements documentaires extérieurs : BnF, HAL, Cairn, OpenEdition, Erudit... Ces alignements permettent d'intégrer dans le triplestore les références bibliographiques à partir de toutes ces sources, réunies autour des identifiants IdRef, et donc de les interroger en une seule requête, pour les réutiliser<sup>2</sup>.

Les documents signalés dans data.idref sont ainsi passés de 11 à 16,2 millions depuis 2018, et sont liés aux entités IdRef par 57 millions de liens distincts, dont plus de 28 millions de relations de contributions.

Au-delà de ces sources documentaires, d'autres alignements d'entités intègrent toujours davantage data.idref.fr dans un écosystème plus vaste : ISNI, VIAF, Wikidata, ORCID, ROR, permettant de rebondir vers d'autres environnements, notamment via le web sémantique<sup>3</sup>.

MICHAËL JEULIN

Gestionnaire de métadonnées, Service  
Outils et Méthodes de l'Abes  
jeulin@abes.fr

[1] Documentation du modèle de données :  
<https://documentation.abes.fr/aideidrefdata>

[2] Voir l'article consacré à SoviSu+ pages 14 et 15.

[3] Une sélection de triplestores « amis » est proposée en page d'accueil de data.idref.fr, et parmi les exemples de requêtes proposés : <https://data.idref.fr/yasgui.html>, des requêtes « fédérées » avec data.bnf.fr, data.persee.bnf.fr, <https://query.wikidata.org>

## LES CATALOGUES ET GISEMENTS DOCUMENTAIRES DANS DATA.IDREF.FR, en nombre de liens document - contributeur

Sudoc : 24,4 millions  
BnF : 6 millions  
HAL : 3,6 millions  
theses.fr : 2,4 millions  
SciencePlus.abes.fr : 580 000  
Cairn.Info : 208 000  
OpenEdition Journals : 120 000  
Calames : 21 000\*  
Erudit : 2 106\*

\* pour Calames et Erudit, des chargements du rétrospectif sont prévus.