

## ● ● ● QUALINCA ET IDREF : L'INTÉGRATION EST EN COURS !

**Disposer de descriptions d'entités fiables et uniques, les associer sans erreur et en masse à des ressources de tout type : tels sont les enjeux de Qualinca. Bilan des travaux menés à l'Abes.**

À l'Abes, l'automatisation du liage des ressources documentaires aux autorités n'est pas un sujet nouveau. Explorée dès 2010 dans le cadre de SudocAD, qui offrait une entrée en matière aux résultats prometteurs, cette question a été approfondie grâce à la participation à Qualinca (« qualité etinteropérabilité de grands catalogues documentaires »), projet Agence nationale de la recherche initié en 2012 et officiellement terminé depuis septembre 2016. Si le cheminement théorique fut long et sinueux, il se révèle aujourd'hui payant. Les premiers fruits sont mûrs pour la récolte, gageons que ce ne seront pas les derniers ! Aujourd'hui, les briques fonctionnelles issues du projet Qualinca se répartissent autour de trois axes de travail : alignement, diagnostic qualité et production assistée.

Le premier axe concerne **l'alignement** : grâce à un programme dédié, il est désormais possible d'aligner sur IdRef – ainsi que sur d'autres référentiels tels ISNI, VIAF, Orcid – des entités « personnes » issues de différentes sources (catalogues documentaires, archives institutionnelles, entrepôts OAI-PMH, bases d'articles, annuaires, référentiels eux-mêmes...). À ce jour, l'alignement à IdRef de 50000 autorités Persée constitue le plus bel exemple.

### VALIDER ET CORRIGER

Le deuxième axe de travail concerne le **diagnostic qualité**. L'objectif de multiplier les liens ne suffit pas, il est primordial de s'assurer de leur fiabilité. En matière de processus automatisés, le principe de prudence est bien sûr requis. Ainsi, chaque nouvel algorithme est testé en confrontant ses résultats à des « vérités terrain », autrement dit à des alignements déjà existants et validés par des experts humains. Néanmoins, que faire du rétrospectif ? Le fait que nos bases soient truffées d'anomalies (liens erronés, doublons, autorités mêlées) risque d'engendrer de la confusion pour les algorithmes qui se nourrissent de ces liens. Les conséquences ? Au mieux, moins

de liens créés ; au pire, malgré la plus grande prudence, la création de nouveaux liens erronés.

Comment sortir de ce dilemme ? La collaboration étroite avec des chercheurs de l'équipe GraphIK de Montpellier<sup>1</sup> a guidé l'Abes vers la résolution du problème. Il en a résulté SudoQual, un outil de validation et de correction des liens préexistants entre le catalogue Sudoc et IdRef. Produit d'une réflexion scientifique exigeante, SudoQual est un programme sophistiqué, au paramétrage complexe ; bien qu'il s'agisse encore d'un prototype, il est possible d'en exploiter certaines « pièces détachées », comme par exemple, un petit programme dérivé capable de détecter des anomalies grâce à deux règles logiques simples qui peuvent s'énoncer ainsi : « *Il existe une anomalie si : 1) au moins deux autorités distinctes ont des ressources liées dont les titres sont identiques ; 2) au moins deux autorités distinctes ont des ressources liées dont les co-contributeurs sont identiques.* »

### CONTRÔLER ET ASSISTER

La **production assistée** constitue le troisième axe de travail, considéré comme le volet le plus important : comment donner aux catalogueurs des moyens de contrôle et d'assistance à la mesure des enjeux actuels (gestion de la masse, qualité, exhaustivité) ? À partir d'une interface imaginée initialement pour visualiser les diagnostics produits par SudoQual, il a

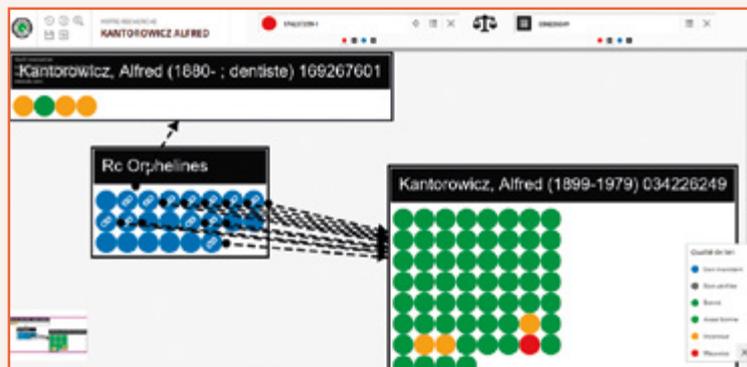
été décidé d'aller plus loin. À la représentation visuelle et synthétique des autorités enrichies de leurs liens bibliographiques grâce à IdRef, un principe essentiel a été ajouté : l'interactivité. En effet, l'utilisateur doit rester acteur du système, ses connaissances dépassant et complétant celles du programme informatique. Il en résulte un outil d'aide à la décision, conçu pour faciliter le contrôle qualité et fiabiliser la production des liens.

Reste à présent à définir les scénarios concrets d'intégration de cet outil dans les environnements de travail quotidien. Au vu de ces résultats, le projet Qualinca a déjà servi les objectifs portés par IdRef. En se concentrant sur les entités personnes, il a mis au point, éprouvé et validé des méthodes génériques qui pourront s'appliquer à d'autres types d'entités, comme les collectivités ou les œuvres. Au-delà du liage, des perspectives ont été ouvertes : création, fusion, scission d'autorités. De son côté, le réseau IdRef est plus mûr, renforcé, mieux armé qu'en 2012. Il semble désormais naturel de poursuivre la convergence et de promouvoir l'intégration de Qualinca et d'IdRef... ou l'intégration de Qualinca dans IdRef.

**ALINE LE PROVOST**

Expert métadonnées - équipe Qualinca  
le-provost@abes.fr

[1] GraphIK (Graphs for Inferences on Knowledge) est une équipe commune de l'Inria (Sophia Antipolis), de l'Université de Montpellier, du CNRS et de l'INRA <https://team.inria.fr/graphik/>



➔ Diagnostic posé par SudoQual pour un sous-ensemble circonscrit par l'appellation « Kantorowicz, Alfred » : les billes sont des points d'accès dans les notices du Sudoc, les boîtes sont des autorités IdRef. Catalogueurs, à vous de jouer !