

FNE - preuve de concept Wikibase juin - novembre 2019 : synthèse

Table des matières

Déroulement de la phase de preuve de concept	1
Synthèse du rapport des prestataires sur la preuve de concept FNE	2
Conclusions du rapport des prestataires du POC FNE	5
Décisions prises sur les prochaines étapes	7

Déroulement de la phase de preuve de concept

Le [Comité stratégique bibliographique](#) (CSB) dans sa réunion du 7 juin 2018 avait conclu qu'une phase d'instruction supplémentaire était nécessaire pour valider le choix de [Wikibase](#) en tant qu'infrastructure du Fichier National d'Entités.

Ni l'Abes ni la BnF ne disposant de l'expertise suffisante pour tirer le meilleur parti de Wikibase, le choix a été fait de lancer un appel d'offres pour travailler en relation étroite avec un prestataire maîtrisant l'outil et impliqué dans la communauté Wikibase.

Les premières semaines de l'année 2019 furent donc consacrées à la rédaction d'un cahier des charges pilotée par les deux chefs de projet (Anila Angjeli pour la BnF et Benjamin Bober pour l'ABES) et impliquant largement les équipes de chacun des établissements. Les documents qui en résultent décrivaient notamment les principes de modélisation et de structuration des données, les différentes opérations qui peuvent leur être appliquées, mais surtout 154 fonctionnalités (recherche, création, imports, exports, administration des données,...) que l'on souhaitait étudier. Par ailleurs, l'ABES et la BnF mirent à disposition un échantillon de données représentatif des entités retenues dans cette phase, à savoir les personnes, les collectivités, les lieux, les œuvres et les concepts (860 notices pour l'ABES, 2903 pour la BnF).

- Afin de pouvoir tester la prise en charge des relations entre entités, un des critères de sélection était la présence des liens entre bon nombre de notices (par exemple liens entre personnes, entre collectivités mais aussi lien entre une notice d'œuvre et la notice de

personne - auteur de l'œuvre, notice de collectivité et notice de personne - membre de la collectivité ...).

- Comme un autre des objectifs de l'expérimentation était de tester les fusions de données provenant de sources différentes, plusieurs entités étaient par ailleurs communes aux échantillons ABES et BnF.

Au terme de la procédure de marché public, c'est un binôme de développeurs (Maxime Lathuilière et Vincent Jumeaux) qui a été retenu, notamment du fait de la réalisation très réussie de inventaire.io, application web utilisée pour dresser et partager l'inventaire de ses livres et s'appuyant sur les données fournies par Wikidata.

La prestation s'est articulée en trois étapes :

- Une formation d'une journée permettant aux équipes de s'approprier le vocabulaire et les principes fondamentaux de Wikibase. 6 personnes furent formées côté ABES, 4 côté BnF.
- L'installation d'une instance Wikibase, le chargement des notices fournies, et le test de fonctionnalités décrites dans les documents du marché.
Cette phase, s'échelonnant de juin à fin octobre 2019, a été conduite suivant une méthode itérative, en sept "sprints" au cours desquels les spécifications des fonctionnalités que l'on souhaitait tester ont été précisées et où une priorisation stricte fut appliquée, tenant notamment compte également des résultats du POC Wikibase mené en parallèle par la BnF dans le cadre du développement de son nouvel outil de production NOEMI. En plus des prestataires et des chefs de projets jouant le rôle de "product owner" suivant la terminologie AGILE, l'ABES mit à disposition pour cette étape un dev-ops (personne qui fait le lien entre les développements et l'infrastructure informatique) et un expert fonctionnel. Côté BnF, 2 expertes fonctionnelles et un développeur furent impliqués.
- La rédaction par les prestataires d'un rapport documentant la phase précédente et plus généralement visant à établir des pistes générales d'implémentation des fonctionnalités demandées ou de résolution de problèmes identifiés. Le rapport remis par les prestataires et dont une synthèse est fournie ci-dessous, a permis à la BnF et à l'ABES de disposer d'une base solide de réflexion sur la pertinence d'utiliser Wikibase comme cœur du prochain FNE.

Synthèse du rapport des prestataires sur la preuve de concept FNE

Le rapport sur la preuve de concept Wikibase-FNE remis par Maxime Lathuilière et Vincent Jumeaux constitue avant tout un rapport technique dont la bonne compréhension nécessite parfois un aller-retour avec les informations présentes sur l'espace Github consacré au projet. Il est par ailleurs accompagné d'un glossaire et d'une bibliographie permettant de mieux capitaliser l'information. Le cœur du rapport est organisé en trois volets, correspondant aux étapes suivies au cours de cette phase :

- *Installation de l'outil Wikibase*

Le prestataire recommande fortement d'installer Wikibase et les services associés en utilisant Docker (<https://github.com/wmde/wikibase-docker>), un logiciel libre d'automatisation et de virtualisation applicative. L'équipe est partie de l'image Docker proposée par Wikimedia Deutschland.

“L'utilisation de Docker nécessite un certain temps d'apprentissage et ajoute une certaine complexité. Il eut également été possible de se passer de cette couche logicielle, comme cela a été fait dans le cadre du projet NOEMI - projet mené par la BnF, visant à remplacer son actuel outil interne de catalogage. NOEMI s'appuie sur le modèle IFLA LRM et sur l'infrastructure logicielle de Wikibase. Nous pensons néanmoins que la valeur apportée par l'usage de Docker est supérieure à son coût en complexité [...]. Cela facilite également la collaboration inter-projets sur l'installation des différents services, la gestion de versions des services et l'intégration d'outils externes.”¹.

A noter que l'utilisation de Docker est en complète adéquation avec les pratiques récemment adoptée à l'ABES et que les compétences nécessaires à son utilisation sont présentes en interne.

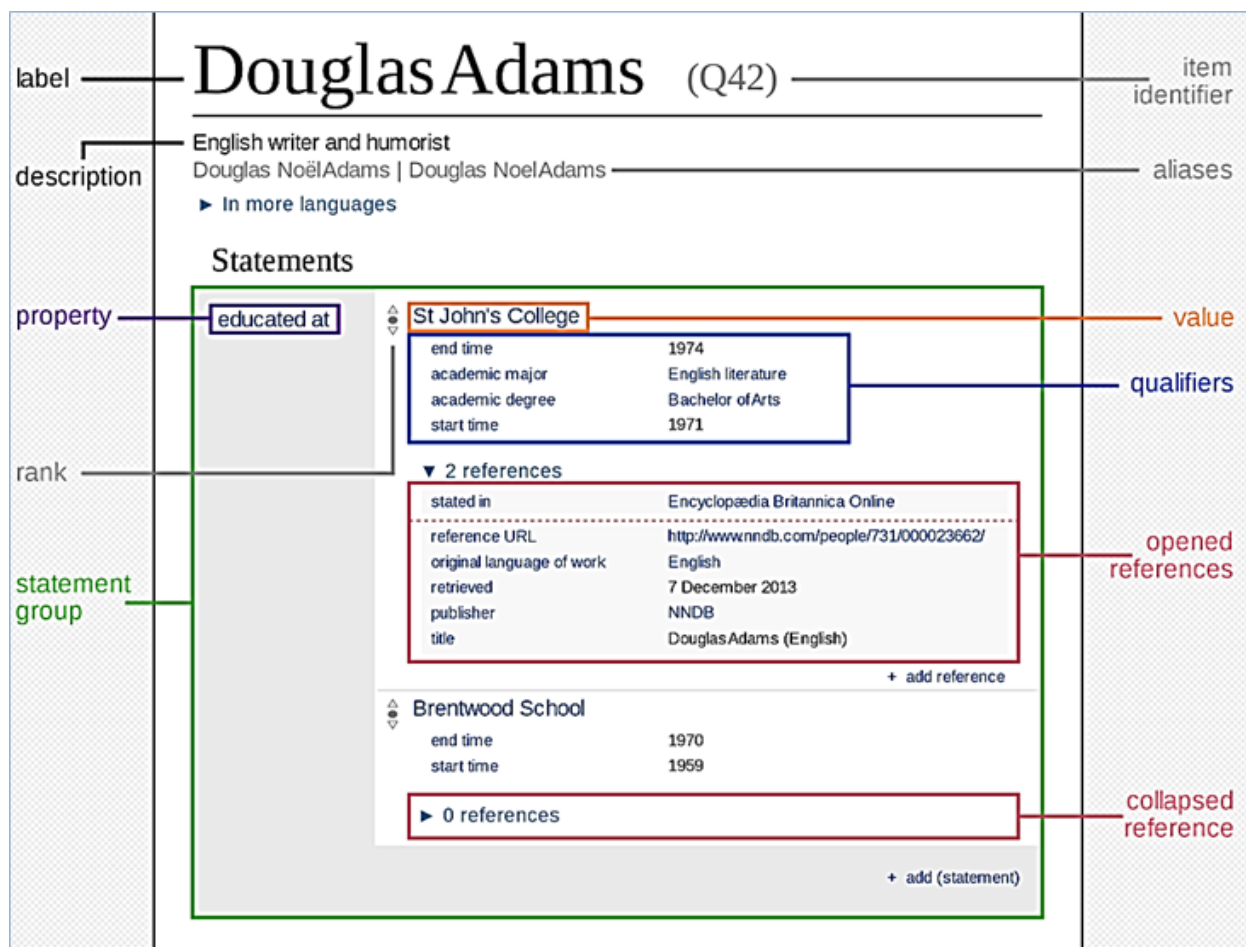
- *Chargement des données*

Les notices de personnes physiques, collectivités, œuvres, lieux géographiques, concepts (RAMEAU) de l'échantillon fourni, furent livrées dans des formats d'exports différents, à savoir Unimarc pour l'ABES et InterMarc pour la BnF.

La preuve de concept a montré qu'il était **possible de convertir les formats MARC en des données respectant le méta-modèle Wikibase**, à savoir un modèle entité-relation.

Voir le schéma du méta-modèle de Wikibase dans le diagramme ci-après (source <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>)

¹ Lathuilière, Jumeaux. *Rapport final de la preuve de concept du Fichier National d'Entités*. p 11



Par ailleurs, le prestataire a proposé une méthode permettant de faire passer des valeurs de sous-champs MARC en des propriétés communes-pivot ABES-BnF utilisant les vocabulaires RDA-FR. Ces propriétés ont été définies par la BnF et l'Abes : type d'entité /URL pérenne / Identifiant ISNI pour les personnes et collectivités / nom / prénom pour une personne / titre de l'œuvre / dates (naissance/mort) / type d'activité / langue.

Cette méthode permet de tester l'hypothèse qu'il n'est pas obligatoire que RDA-FR - convenu comme colonne vertébrale de la modélisation du FNE - soit complètement abouti ni que tous les choix de modélisation dans Wikibase soient entérinés pour commencer la migration des données.

Il a été montré également qu'il était possible d'utiliser les outils Wikibase (service de requêtes SPARQL) pour s'appuyer sur les alignements existants dans les données source (alignements effectués via VIAF, ou présence du même identifiant ISNI dans les notices BnF et Abes pour les mêmes entités) et fusionner des entités identiques provenant de sources différentes.

- *Confrontation de la liste des fonctionnalités souhaitées avec ce que peut faire Wikibase, ses extensions, ou des outils tiers exploitant ses API.*

Les 154 fonctionnalités listées dans le Backlog FNE, annexé au cahier des charges du marché, ont été regroupées en quatre familles qui ont été étudiées spécifiquement, faisant l'objet de discussions et de tests ou d'analyses poussées :

- Recherche (41 fonctionnalités)
- Règles et contraintes (17 fonctionnalités)
- Rôles et restrictions (10 fonctionnalités)
- Alignement et fusions (4 fonctionnalités)

L'étude a montré que les fonctionnalités élémentaires de recherche sont présentes nativement ou via des extensions aisément paramétrables. En revanche, elle met en lumière la nécessité de développements particuliers pour réaliser des fonctionnalités professionnelles comme la recherche par index, les tris, facettes, sélections... Il en va de même pour la question des règles et contraintes, comme les contrôles a priori, les formulaires d'aides à la saisie, les listes de valeurs contrôlées, et dans une moindre mesure pour celle des alignements et fusions. Globalement, le développement de ces fonctionnalités, bien qu'il ne soit pas trivial, ne présente aucun obstacle technique majeur, tout en respectant l'économie générale du projet. **La présence d'API riches permet d'utiliser le langage de programmation de son choix, sans avoir à développer des extensions nécessitant des connaissances approfondies des composantes logicielles de Wikibase.** Par ailleurs, les demandes exprimées étant parfois également émises par d'autres membres de la communauté Wikibase, il n'est pas exclu que l'équipe de développement de Wikibase en prenne à sa charge tout ou partie, dans la durée du projet.

La gestion des rôles et des restrictions est potentiellement plus problématique. L'outil tel qu'il est conçu peut gérer assez finement les différents rôles, mais ne peut pas aller jusqu'à la restriction en lecture ou en écriture d'une assertion particulière (date de naissance d'une personne, par exemple). **Les développements nécessaires sont beaucoup plus complexes et nécessiteront obligatoirement un investissement de la part de l'équipe de développement de Wikibase et des logiciels dont il dépend,** si l'on souhaite véritablement apporter un niveau de sécurité et de stabilité suffisant.

Les tests sur les alignements et fusions ont montré qu'il était possible d'utiliser les outils Wikibase (service de requêtes SPARQL) pour s'appuyer sur les alignements existants (VIAF, ISNI) et fusionner des entités identiques provenant de sources différentes.

Conclusions du rapport des prestataires du POC FNE

Les enseignements de la preuve de concept FNE, enrichis par les échanges avec la communauté Wikibase et l'expérience du projet NOEMI à la BnF, ont permis d'arriver aux conclusions suivantes.

Points forts

- Logiciel libre, utilisant les standards du web de données
- Architecture technique ouverte à l'extension et à la personnalisation via les extensions et des services tiers.

- Capacité de mise à l'échelle
- Coût et délai réduits en réalisation et en maintien en condition opérationnelle
L'utilisation de la technologie de containerisation s'est avérée très pertinente, permettant d'avoir rapidement à disposition une instance en production qui reste évolutive (ajout d'extensions par exemple)
- Bénéficie de tout l'écosystème Wikibase :
Même si à l'origine Wikibase était très lié à Wikidata, la tendance est à la montée en puissance de nombreuses extensions et outils génériques (par exemple wikibase-edit, utilisé dans le cadre de cette preuve de concept, initialement connu sous le nom de wikidata-edit)
- Modèle de données souple et granulaire. Compatibilité avec LRM vérifiée.

Points de vigilance

- L'implémentation du modèle de données compatible LRM pourrait se faire sans même avoir besoin que ce modèle soit complètement stabilisé. Tout cela reste néanmoins à analyser plus en profondeur.
- Fonctionnalités de recherche professionnelle à développer fortement
- Restrictions en lecture et contrôles sur la gestion de certaines informations difficiles à implémenter
- Outils de collaboration (tableau de bord, forum, messagerie) devront sans doute reposer sur des outils tiers.
- Pour les évolutions nécessaires liées aux besoins du FNE :
 - dépendance des décisions et du calendrier des développements de Wikimedia Deutschland et de la communauté Wikibase
 - Si les besoins formulés auprès de la communauté sont rejetés, les développements devront se faire en interne ou au sein d'une communauté réduite (dans et pour l'instance Wikibase FNE), avec le risque de s'éloigner du standard et la nécessité de prendre en charge toute la maintenance.
- Montée en compétence technique à prévoir
- Risque de confusion entre identité des projets Wiki (Wikipedia et Wikidata) et l'identité du FNE : les travaux du Groupe réseaux et communication du FNE proposent de mettre l'accent sur Wikibase uniquement auprès de la communauté Wikibase et non auprès de nos communautés d'utilisateurs. Jusqu'à présent, nous n'avons pas communiqué largement auprès de nos réseaux sur les aspects très techniques de stockage des données ou des technologies utilisées. Il n'y a pas de raison objective de changer de façon de procéder.
- Utilisation d'un outil de coproduction par des acteurs ayant des modes de fonctionnement distincts : cet aspect fondamental du FNE n'est pas lié à l'usage de Wikibase. Bien que non traité en tant que tel dans le POC-FNE, il est à mettre en relation, d'une part avec la gestion des rôles (abordés dans le POC), d'autre part avec la gouvernance des données dans le FNE.

INTERNE	FORCES	<ul style="list-style-type: none"> - Logiciel libre, utilisant les standards du web de données - Architecture technique ouverte à l'extension et à la personnalisation - Coût et délai réduits en réalisation et en maintien en condition opérationnelle - Modèle de données souple et granulaire. Compatibilité avec LRM vérifiée. 	FAIBLESSES	<ul style="list-style-type: none"> - Fonctionnalités de recherche professionnelle à développer fortement - Restrictions en lecture et contrôles sur gestion de certaines informations difficiles à implémenter
	EXTERNE	OPPORTUNITES	<ul style="list-style-type: none"> - Place grandissante des institutions culturelles et en particulier des bibliothèques dans la communauté d'utilisateurs 	MENACES

Synthèse des conclusions de la preuve de concept FNE sous forme de matrice SWOT (Strengths (forces), Weaknesses (faiblesses), Opportunities (opportunités), Threats (menaces))

Décisions prises sur les prochaines étapes

Au vu des éléments donnés ci-dessus, les recommandations sont :

- Développer le Fichier National d'Entités en s'appuyant sur Wikibase
- Limiter au maximum tout développement spécifique qui entraînerait une divergence avec la souche logicielle Wikibase
- Participer activement aux groupes collaborant avec les équipes techniques de Wikibase afin d'orienter les développements du coeur de l'outil dans le sens qui nous convient le mieux
- Reverser à la communauté Wikibase les développements que nous ferions, dans l'esprit du logiciel libre.