

Datos.bne.es : un service LOD et un accès aux collections de la bibliothèque fondé sur le modèle FRBR

French translation of the original paper: “Datos.bne.es: a LOD service and a FRBR-modelled access into the library collections”.

Translated by: Raphaëlle Lapôtre, Bibliothèque Nationale de France, Paris, France.

The text of this document has been translated into French and differences from the original text may occur. This translation is provided for reference purposes only.

Ricardo Santos

Bibliothèque Nationale d’Espagne
ricardo.santos@bne.es

Ana Machado

Bibliothèque Nationale d’Espagne

Daniel Vila-Suero

Groupe d’ingénierie ontologique – Université Technique de Madrid
dvila@fi.upm.es



This is a French translation of “Datos.bne.es: a LOD service and a FRBR-modelled access into the library collections” copyright © 2015 by Raphaëlle Lapôtre. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

Datos.bne.es a été lancé pour la première fois en 2011 en tant que service de données de la Bibliothèque Nationale d’Espagne. En novembre 2014, une nouvelle version beta a été élaborée, élargissant le périmètre et la richesse des données et construisant, par-dessus le modèle de données fondé sur des entités, une recherche et une interface innovantes orientées vers le grand public.

Le modèle FRBR a été choisi pour transformer, trier, ordonner et enfin exposer les données issues des notices bibliographiques et d’autorité en MARC21. Cet article expose les objectifs et les principes qui sous-tendent ce projet en cours, les processus de travail et les décisions prises dans la conceptualisation, les services de données développés et la technologie sous-jacente à tous ces procédés.

Enfin, sur la base des retours de la part des utilisateurs, on évaluera les avantages et les inconvénients de cette représentation selon le modèle FRBR et de cette expérience de diffusion en LOD, tirant quelques conclusions et dessinant les potentielles futures étapes du projet.

Introduction

Datos.bne.es 2.0 a été lancé en novembre 2014, en tant que version beta et banc d'essai expérimental ayant pour objectif de fournir une preuve de concept et de recueillir des retours. Ce projet phare de la Bibliothèque Nationale d'Espagne avec le soutien et la participation du Groupe d'Ingénierie Ontologique (de l'Université Technique de Madrid) est une nouvelle étape dans le domaine du Linked Open Data, qui débuta dès 2011, avec la première version de datos.bne.es, dont l'expérience a permis d'élaborer cette nouvelle version. Deux caractéristiques principales sont à souligner:

- l'utilisation du Linked Open Data pour construire un immense jeu de données, décrit en conformité avec les bonnes pratiques de la publication en LOD, afin d'organiser les données de bibliothèque selon des modèles, structures et vocabulaires appropriés à l'environnement du Web Sémantique, de les rendre plus interopérables, réutilisables et visibles sur le Web, et de lier et échanger efficacement nos données avec d'autres sources.
- L'utilisation, comme modèle de référence, des concepts et entités élémentaires du modèle FRBR pour ordonner et exprimer des données bibliographiques, et étudier les avantages et les inconvénients de ce choix pour le grand public et les chercheurs par rapport aux méthodes traditionnelles.

Contexte

La BNE et l'UPM (l'Université Polytechnique de Madrid – Université Technique de Madrid), se sont rencontrées une première fois en septembre 2010 pour étudier la possibilité de travailler ensemble à l'exposition de données culturelles et de bibliothèques sous forme de données liées.

L'OEG (Le Groupe d'Ingénierie Ontologique) est un groupe de recherche de l'UPM, pionnier en Espagne pour les projets liés aux données liées. L'OEG est largement reconnue en Europe dans les domaines de l'ingénierie ontologique, de l'Infrastructure Sémantique, des données liées et de l'intégration de données. Ce groupe a développé des applications dans un grand nombre de domaines et en particulier dans la publication en Linked Open Data (dans des secteurs comme le tourisme ou les données géographiques) ; en outre, l'OEG maintient la partie espagnole de DBpedia et participe aux nombreuses activités de standardisation du W3C comme la spécification Linked Data Platform ou l'ontologie de provenance PROV-O. Ce groupe désire très fortement orienter ses recherches dans le secteur du patrimoine culturel.

L'objectif de la BNE était d'essayer de publier des données à la fois selon le modèle conceptuel et les vocabulaires de l'IFLA mais aussi selon les principes du web sémantique. Au début du cycle de vie de ce projet conduit selon un processus incrémental et itératif, un premier jeu de donnée a été traité à titre de preuve de concept (Machado, 2011). Ce jeu, constitué autour de Cervantes, concernait toutes les notices bibliographiques et toutes les autorités qui lui étaient reliées (environ 8000 notices bibliographiques et 7300 notices d'autorité).

A partir de l'expérience acquise au bout de cette année-là, le premier jeu de notices bibliographiques (2.4 millions, comprenant des monographies modernes et anciennes ainsi que de la musique imprimée et enregistrée) et de données d'autorité (personnes, organismes et Œuvres), a été FRBrisé et publié (Vila Suero, 2013), à hauteur de 58 millions de triplets.

Les données ont été rendues accessibles grâce à des fichiers de téléchargement massif de données mais aussi à travers un Sparql end-point sous une licence CC0 du Domaine Public.

Datos.bne.es 2.0 : principales caractéristiques

C'est en novembre 2014 que fut lancé le nouveau portail datos.bne.es. La dernière version contient un certain nombre de nouvelles fonctions et d'améliorations par rapport à l'ancienne version.

Elle a en effet été équipée d'une interface visuelle avec moteur de recherche, conçue pour permettre à l'utilisateur d'expérimenter de nouvelles façons de naviguer et de découvrir les données.

La variété et le nombre des notices bibliographiques et d'autorité ont été augmentés et optimisés, pour couvrir désormais presque toutes les collections de la bibliothèque, grâce à l'ajout des manuscrits, des enregistrements vidéo, des photographies, des dessins et des cartes. Seuls les périodiques ont été laissés en dehors dans cette nouvelle version. Les données relatives aux exemplaires et aux états de collection ne sont pas non plus disponibles.

Le fichier de vedettes-matières a été modélisé en SKOS et intégré dans datos.bne.es. Plus de 350 000 vedettes-matières ont été modélisées et près de 30000 ont été liées aux LCSH (Library of Congress Subject Headings).

Ont également été inclus des liens vers les œuvres numérisées de la Bibliothèque Numérique Hispanique, qui donnent accès à près de 100 000 documents numériques de la bibliothèque.

Pour cette itération, une ontologie propre à la BNE a été élaborée, afin de ne pas risquer d'incohérences avec les vocabulaires déjà existants. Cependant, un travail d'alignement reste encore à faire.

Les principes fondamentaux de ce travail de mise à jour et de ces futures orientations tiennent en trois points :

- Offrir à l'utilisateur une plateforme intégrant toutes les ressources disponibles à la BNE, et un portail vers des ressources sélectionnées d'autres provenances, ainsi qu'une expérience de navigation nouvelle et innovante. L'architecture proposée doit répondre aux besoins de différents types d'utilisateurs, du chercheur à l'utilisateur courant.
- Offrir un modèle de données sous-jacent robuste et adaptatif, capable d'agréger de nouvelles données, soit par importation de sources externes, soit par les ajouts et enrichissements des catalogueurs par rapport à l'existant.
- Offrir toute cette richesse de données dans un environnement en Open Access, sous licence ouverte pour les rendre réutilisables par d'autres acteurs.

Interface et navigation

Datos.bne.es est toujours en version beta : la recherche et l’affichage d’information sont désormais en cours de révision, de sorte qu’il est possible que quelques-unes des fonctionnalités exposées ci-dessous aient été modifiées depuis la publication de cet article.

Datos.bne.es se présente comme une interface multi-fonctionnelle et très réactive, utilisable de ce fait par des ordinateurs, des tablettes, des téléphones ou tout autre appareil similaire.

L’interface a été construite autour de trois entités principales sous-jacentes au modèle de données, à savoir Auteur, Œuvre et Sujet. A partir de ces entités, seule la forme retenue de chacune de ces entités est indexée. Le titre propre des manifestations est également indexé pour pouvoir être retrouvé.

Les requêtes se font à partir de la barre de recherche centrale. Des suggestions sont proposées au moment d’entrer les termes, si ceux-ci correspondent à quelques-uns des champs indexés susmentionnés. Pour ce faire, l’ordre des éléments doit également être respecté (signes et lettres capitales ne sont pas pris en compte), c’est-à-dire que « James He » renvoie « James Henry », mais pas « Henry Ja ».



Suggestion automatique

Dans les suggestions le type d’entité auquel les résultats appartiennent est montré. Cela peut-être :

- Un *Auteur*, incluant les personnes physiques et les collectivités.
- Une *Œuvre*, au sens du modèle FRBR.
- Un *Sujet*.

- Une *Ressource* (i.e. Manifestations), à savoir tout document quel qu'il soit. Ce type d'« entité » n'est pas nommé ainsi sur le portail : à la place figure le type de support auquel appartient la ressource.

doñana en la cultura contemporánea	
Doñana en la cultura contemporánea	Registro sonoro musical
Doñana, Parque Nacional	Vídeo
Doñana nombre a nombre	Obra
Doñana, diversidad y ciencia	Libro
Doñana, las otras huellas	Música impresa
Doñana parque nacional :	Libro
Doñana :	Vídeo
Doñana	Música manuscrita

Exemple de suggestion de recherche (modifié pour la démonstration) montrant un type de support et une entité Œuvre [traduit de haut en bas : enregistrement son, vidéo, Œuvre, livre, musique imprimée, livre, vidéo, musique manuscrite]

Les suggestions peuvent être passées en pressant la touche Entrée, ce qui commande l'affichage en première ligne d'une liste de résultats contenant les mots recherchés répartis dans les trois entités et, en seconde ligne, les documents groupés par type de document.

Buscar un autor, obra o tema

Búsqueda: zaragoza (10771 resultados)

Filtrar por tipo

3861 Obras	1632 Autores	1304 Temas			
2642 Libros	414 Dibujos	224 Materiales cartográficos	142 Manuscritos	124 Registros sonoros musicales	116 Fotografías

Exemple de recherche directe. La première ligne montre les résultats à partir de trois entités [de gauche à droite : Œuvres, Auteurs, Sujets]. La seconde ligne affiche les résultats de documents, disposés par types (Livres, dessins, cartes, manuscrits, enregistrements sonores et photographies)

Un clic sur un des encadrés permet enfin d'appliquer un filtre, restreignant la recherche selon l'entité choisie. Le moteur de recherche utilise le graphe FRBR pour retrouver et classer les entités, selon leurs relations respectives.

Búsqueda: *zaragoza* (3861 resultados)

Filtro

Obras x

Filtrar por tipo

1632 Autores 1304 Temas

2642 Libros 414 Dibujos 224 Materiales cartográficos 142 Manuscritos 124 Registros sonoros musicales 116 Fotografías 100 Dibujos, grabados y fotografías

21 de 3861

The screenshot shows a search interface with a header 'Búsqueda: zaragoza (3861 resultados)'. Below it is a 'Filtro' section with a red 'Obras x' button. Underneath is 'Filtrar por tipo' with a grid of filters: '1632 Autores', '1304 Temas', '2642 Libros', '414 Dibujos', '224 Materiales cartográficos', '142 Manuscritos', '124 Registros sonoros musicales', '116 Fotografías', and '100 Dibujos, grabados y fotografías'. Below the filters, it says '21 de 3861'. The main content area displays six search results cards. Each card has a title, version count, a red 'Obra' button, and some have a 'Disponible en digital' button and a thumbnail image. The results include 'Todo Zaragoza (6 versiones)', 'Zaragoza (3 versiones)', 'Zaragoza (3 versiones)' by Pérez Galdós, Benito, 'Zaragoza (3 versiones)', 'Zaragoza (2 versiones)', and 'Zaragoza (2 versiones)' by Blasco, Justo.

Filtre par œuvres

La recherche fondée sur les entités ainsi que l'architecture du web permettent à l'utilisateur de se plonger plus précisément dans le sous-ensemble de données qui l'intéresse. Certes, cela oblige l'utilisateur à sélectionner d'abord l'entité désirée, ce qui rajoute un clic avant d'explorer les documents eux-mêmes, mais nous pensons que cet effort en vaut la peine.

La page Auteur est composée d'informations contextuelles extraites de DBpedia (résumé et vignette), quand elle est disponible, ainsi que des ressources liées aux auteurs, réparties en trois catégories :

- Les œuvres par et à propos de l'auteur : ces nœuds peuvent conduire à une page Œuvre ou directement à une page Ressource.
- Les autres ressources connectées à l'auteur (ayant un rôle secondaire autre que celui d'auteur), conduisant directement à une page Ressource :

The diagram consists of a light gray rectangular background. On the left, the text 'En la BNE' is displayed. To its right are three large, light-colored circles arranged horizontally. Each circle contains text in red: the first circle says 'Autor de 105 Obras', the second says 'Tema en 32 Obras', and the third says 'Participa en 124 Obras'.

De gauche à droite : Auteur de ; Sujet dans ; Contribue à

En dessous des ressources internes à la bibliothèque, s'affichent des liens externes équivalents aux jeux de données sélectionnés :

Enlaces

Ver más	http://dialnet.unirioja.es/servlet/autorExterno/BNE/871548
VIAF	http://viaf.org/viaf/98038975
Sudoc	http://www.idref.fr/02666304X
Biblioteca Nacional Francesa	http://data.bnf.fr/11886882
ISNI	http://isni-url.oclc.nl/isni/0000000121447774
DBpedia	http://dbpedia.org/resource/Julián_Marías
Biblioteca del Congreso de los EEUU	http://id.loc.gov/authorities/names/n79093050
Biblioteca Nacional Alemana	http://d-nb.info/gnd/118781898

Ces liens sont des liens « similaire à » par rapport à d'autres jeux de données LOD tels que (VIAF, LC, BnF, Sudoc, DNB, Libris, ISNI, DBpedia), mais aussi des liens « voir aussi » pointant vers d'autres sources non-LOD, pour la plupart issues des notices d'autorité en MARC21 (champs 856 et 670\$u).

Enfin sont affichées les sources consultées pour l'œuvre d'autorité, issues des notices d'autorité en MARC21.

La page Œuvre est aussi une des pierres de touche de datos.bne.es ; elle a un rôle de pivot dans le regroupement et le tri de données relatives aux différentes éditions d'une même création intellectuelle. Elle pose également les bases d'une future interconnexion à terme de ces composantes élémentaires de l'univers bibliographique à d'autres entités reliées.

La page Œuvre regroupe les éditions existantes en versions, à savoir, par langues de présentation de l'Œuvre en versions linguistiques, et montre également les Œuvres portant sur l'Œuvre (relation descriptive).

Tomás de Aquino, / Summa theologica

Summa theologica

En la BNE

Disponibles
7
Versiones

Tema en
20
Obras

Versions disponibles (7) et Œuvres portant sur (20) pour la Somme Théologique de Saint Thomas d'Aquin

Afin d'épargner des niveaux de navigation au public, la version qui contient le plus de notices s'affiche par défaut. Les notices bibliographiques apparaissent en version abrégée, la totalité des données pouvant être vue en cliquant sur « Voir plus ». Les éditions liées à du contenu numérique sont montrées en premier.

Enfin, la page édition (i.e. manifestation) contient les attributs ordinaires d'une notice bibliographique.

Darwin, Charles / On the origin of species / Español / El origen de las especies

El origen de las especies

[Acceder a la versión digitalizada](#) [Acceder a la entrada en el catálogo](#)

Título	El origen de las especies;Charles Darwin ; traducción de Aníbal Froufe ; prólogo de Faustino Cordón;
Lugar de publicación	Madrid
Editorial	Edaf
Fecha de publicación	2010
Tipo de recurso	Libro
Tema	Evolución
Participante	Froufe, Aníbal
Pertenece a versión	Español
Descripción física o extensión	695 p.
Forma de contenido	[Texto impreso]
Dimensiones	21 cm
Depósito Legal	M 5761-2010
ISBN	978-84-414-2501-9

Depuis cette page, l'utilisateur peut rebondir sur l'OPAC traditionnel pour trouver des informations sur le document et sa disponibilité pour le prêt. On trouve également un lien vers la bibliothèque numérique si l'édition est reliée à un contenu numérique (habituellement, une version numérisée).

Enfin, la page Sujet contient des liens aux sujets liés dans le thesaurus, le terme équivalent en LSCH et les liens à des Œuvres :

Evolución

Término aceptado	Evolución
Usado por	Evolución (Biología) Origen de las especies Evolución biológica Especies (Biología)--Evolución
Término específico	Cerebro--Evolución Evolución humana Plantas--Evolución
Término relacionado	Emergencia (Filosofía) Filogenia Selección natural Adaptación (Biología) Genética evolutiva Creación Evolucionismo Biogénesis
Nota de alcance	Úsase para obras sobre la descripción y explicación de los hechos que ilustran las teorías de la evolución. véaseEvolucionismo
Término equivalente en otros vocabularios	http://id.loc.gov/authorities/subjects/sh90004042

Page Sujet

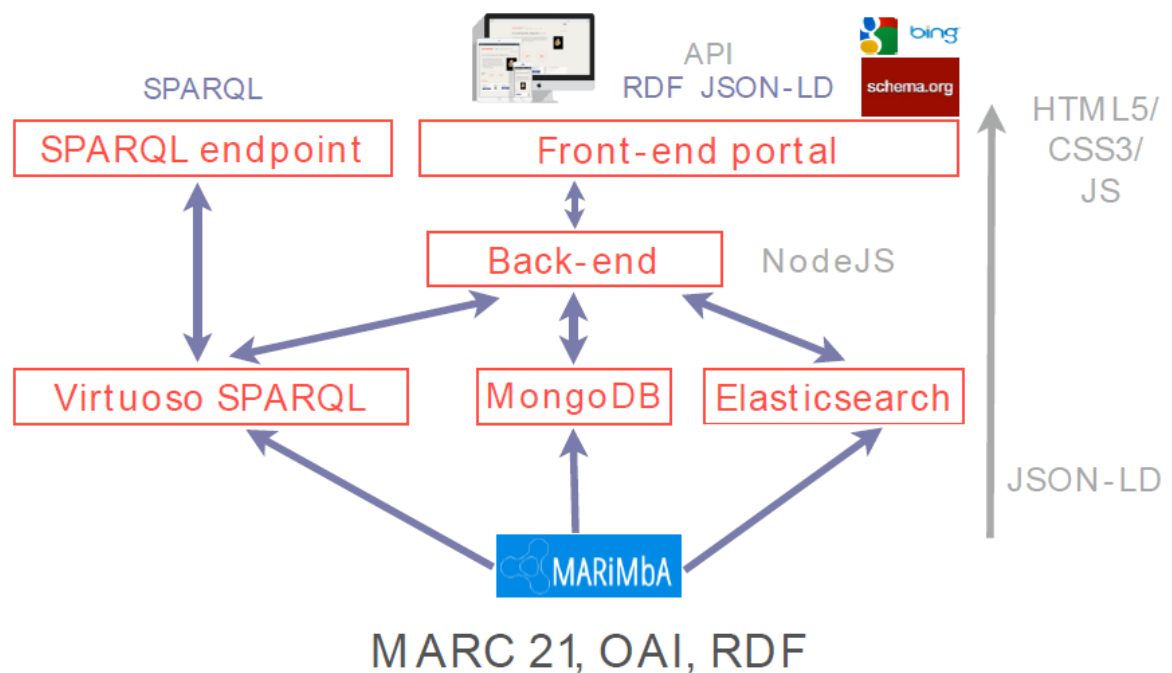
Futures améliorations possibles en ce qui concerne la visualisation et la navigation :

- Inclure plus de valeurs interrogeables.
- Permettre la recherche sur un seul type d'entité.
- Prévoir et gérer des recherches impliquant deux entités (par exemple, Shakespeare Hamlet).
- Améliorer l'affichage et le tri d'information. Mieux expliquer la différence entre les œuvres et les ressources.
- Expliquer ou raffiner le concept d'Auteurs (toutes les Personnes ne sont pas des auteurs, tous les auteurs ne sont pas des personnes).

Infrastructure technologique

La plateforme technologique sous-jacente au service datos.bne.es repose sur un grand nombre de technologies modernes du web. Afin de traiter, analyser, convertir et modéliser les données du catalogue en données liées, un outil souple appelé Marimba a été développé. Marimba permet à des bibliothécaires experts de participer à la conversion et à la production de données liées, dissimulant les complexités des modèles sémantiques et des algorithmes conduisant le processus de production (Vila-Suero, 2013). Une fois les données modélisées et liées entre elles, une chaîne de publication est établie. En premier lieu, les données liées sont indexées dans un moteur de recherche spécialement conçu pour prendre en compte la structure sémantique des données et en tirer parti. La forme de graphe des données rend

possible un mécanisme de classement clair directement fondé sur le modèle FRBR sous-jacent. En second lieu, les données sont également stockées dans deux systèmes différents : une base de données NoSQL pour soutenir l'interface, et un entrepôt de triplets pour faciliter les requêtes et la réutilisation par des applications sémantiques. A ce stade, il est important de mentionner que la sérialisation JSON-LD linked data est utilisée d'un bout à l'autre sur l'ensemble de la plateforme. Cette sérialisation constitue un avantage évident puisqu'aucun traitement de données n'est fait après le processus de production, et ainsi le moteur de recherche, les entrepôts de données et en dernier lieu l'interface traitent pour l'essentiel des mêmes objets de données. En outre, étant donné que le service est entièrement construit en JavaScript, tant pour les bases de code serveur que client, JSON est un choix naturel. Voici un aperçu de l'architecture technique :



Services de (Linked) Data

La plateforme décrite ci-dessus a été conçue pour fournir un service de données robuste, interopérable et adaptable. Fondé sur ces principes, un ensemble de services de données est en cours de développement. En particulier :

- Un SPARQL endpoint public est fourni pour rendre les données « requêtables » en utilisant le langage SPARQL. (<http://datos.bne.es/sparql>)
- Une API de recherche est proposée aux développeurs web pour développer des applications avec les données de la BNE. Cette API de recherche est basée sur les mécanismes de classement susmentionnés et peut également servir d'outil de réconciliation pour être utilisé avec des outils comme Open Refine. Par exemple, une recherche avec le mot-clé « Cervantes » fournirait comme premier résultat l'URI de Miguel de Cervantes Saavedra.
- Un mécanisme de négociation de contenu pour toutes les ressources avec les formats disponibles suivants : HTML, RDF/XML, Turtle et JSON-LD.
- Des annotations Schema.org au format JSON-LD pour chaque ressource web afin d'augmenter la visibilité et l'indexation sémantique par les moteurs de recherche des

ressources de datos.bne.es. Par exemple, les auteurs sont annotés en tant que schema:Person et les œuvres sont annotées en tant que schema:CreativeWork.

Outre les services listés ci-dessus, l'objectif est de développer les APIs et leur documentation afin de promouvoir la réutilisation par des tiers, ainsi que d'organiser des rassemblements de développeurs dans le cadre d'événements en lien avec l'Open data en Espagne.

Extraire et lier des entités FRBR à partir de structures MARC21

Elaborer une stratégie pour FRBRiser les données en MARC21 implique une réflexion sur les données culturelles et les procédures de codage de chaque institution (Boulet, 2013). Dans la conversion de datos.bne.es, les zones et sous-zones MARC21 ont été prises en compte, et la création, la délimitation et la relation ont été extraites en se fondant sur la présence ou l'absence des valeurs de codage.

De 1991 à 2007, le logiciel de catalogage ad-hoc en usage à la BNE a contraint les catalogueurs à rattacher explicitement chaque notice bibliographique à une notice d'autorité Œuvre préexistante, ou pour une traduction, à une notice d'autorité Expression. Cette multitude de notices d'autorité est souvent de niveau minimal, et ne permet parfois même pas d'identifier des Œuvres, mais malgré tout, nous avons à disposition une quantité remarquable de notices autorité pour les Œuvres et les Expressions à traiter et à exploiter.

Le modèle FRBR était la réponse logique à la question de savoir comment tirer le meilleur parti de ces données. En premier lieu, nous avons postulé les rattachements suivants entre données et entités :

Notice autorité Personne / Collectivité → Entité Personne / Collectivité
Notice autorité Œuvre (Personne/Collectivité + titre) → Entité Œuvre
Notice autorité Expression (Personne/ Collectivité + titre + langue) → Entité expression
Notice bibliographique → entité Manifestation
Notice autorité sujet → entité Sujet

En 2007, un nouveau logiciel est arrivé à la bibliothèque, et les liens des notices bibliographiques vers les notices autorité Œuvre/Expression ne peuvent plus être utilisés que dans les cas classiques de titres uniformes.

Lorsque aucune relation explicite n'était enregistrée (champs 240, 130), les notices de manifestations ont été rattachées à des Œuvre/Expression correspondante grâce à la combinaison de la zone 1XX + 245 \$a, en comparant la chaîne de caractères ainsi obtenue à la notice d'autorité correspondante. La langue de l'expression originale a été inférée de la zone fixe 008 dans les notices bibliographiques, quand aucune langue n'était explicitement mentionnée lors de ces appariements en 130/240 \$1. Bien sûr, de nombreuses erreurs de codage ont fait surface lors de ces attributions, et beaucoup de manifestations étaient des œuvres « orphelines », mais le jeu en valait la chandelle, puisque beaucoup de liens ont pu être reconstitués.

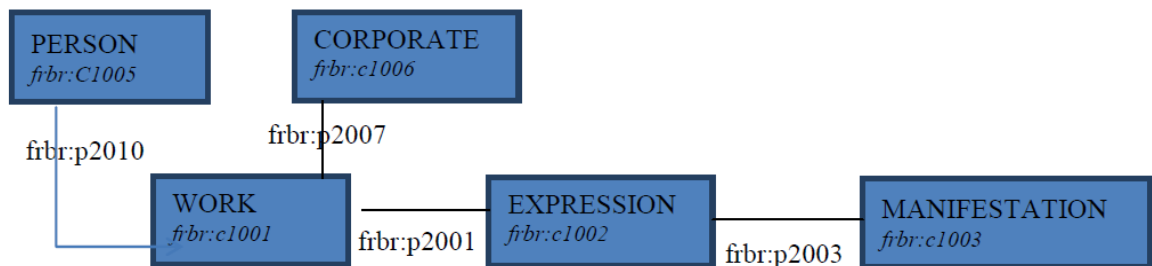
Après avoir établi ces entités à partir de l'existence de notices auto-contenues, le mappage des propriétés des éléments pour ces entités a été réalisé à partir de leurs sous-zones respectives. Par exemple :

Le mappage de propriétés a été établi à partir de sous-zones. De nombreux niveaux de granularité ont ainsi été atteints, selon les caractéristiques propres de sous-zones : des sous-zones ambiguës (la 245 \$b, par exemple) ont dû être représentées sous forme de propriétés générales, et les pratiques locales étaient utiles pour extraire du sémantique dans certains cas, mais pas dans d'autres. Cela nous a menés à la conclusion inévitable que les pratiques de catalogage classiques devaient être précisément repensées dans la perspective de faciliter le traitement de données par les applications de données liées.

FRBR comme modèle de référence

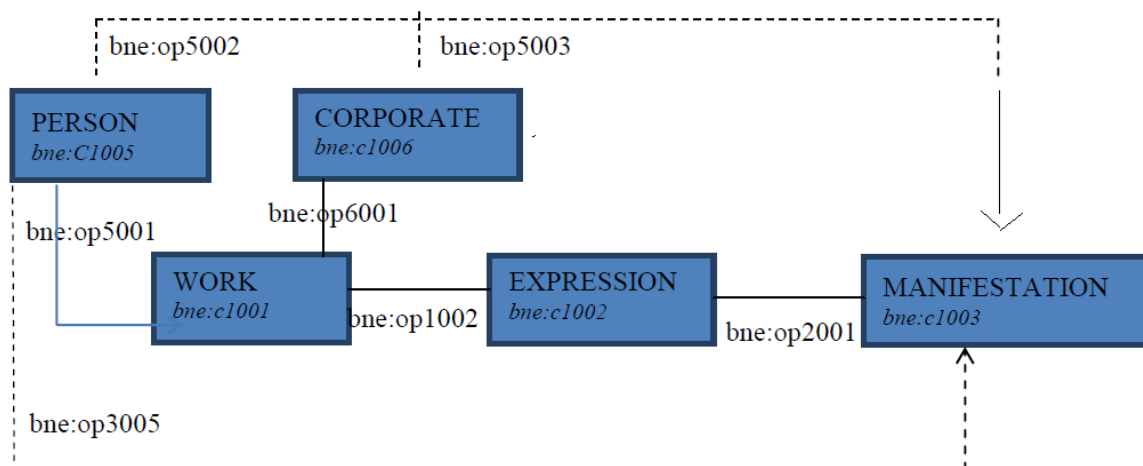
De notre point de vue, le modèle FRBR est le modèle le plus adapté pour représenter des relations, en particulier pour de grandes masses de données bibliographiques, telles qu'on en trouve habituellement dans une bibliothèque nationale de la taille de celle de l'Espagne. Cela permet de regrouper des ressources liées, de conduire les usagers vers le sous-ensemble de données qui convient plus précisément à leurs besoins. Cela contribue également à donner à l'utilisateur ce qu'il cherche véritablement : une recherche auteurs devrait retourner exclusivement des auteurs, par exemple. Au contraire, les dispositifs classiques de recherches des interfaces de bibliothèques ramènent immédiatement des ressources, quel que soit l'objet de la requête. Data.bne s'efforce de comprendre ce que l'utilisateur demande, et agit en fonction de cette demande.

Cependant, adopter le modèle FRBR dans sa totalité ne s'est pas avéré tâche facile. Les données en MARC issues de rétroconversion et la structuration en MARC sont bien plus anciennes que le modèle FRBR, de sorte que ce que nous pouvons extraire ou inférer de données existantes ne convient pas toujours à la structure FRBR. Comme nous l'avons dit plus haut, bien que certains mécanismes complexes d'analyse de données aient été appliqués à des notices dans la perspective d'identifier et de lier des entités FRBR, quelques données du catalogue de la BNE sont passées entre les mailles du filet. La version 1 de Datos.bne.es ignorait les données ne correspondant pas au modèle, de sorte que nous avons un modèle solide, mais une perte spectaculaire de données :



Dans ce modèle, les données qui ne correspondaient pas aux critères (à savoir, les manifestations non reliées à une Œuvre), n'ont pu rejoindre le réservoir des données exploitables par datos.

Pour datos 2.0, une vision plus pragmatique a été adoptée, introduisant davantage de souplesse et quelques échappatoires par rapport au modèle, permettant d'exprimer une plus grande variété de relations sans perdre l'esprit et les avantages des fonctionnalités FRBR :



Deux nouvelles relations significatives ont été implémentées dans le modèle entre les auteurs et les manifestations :

- OP5002 et OP5003 relient une personne ou une collectivité directement à une manifestation en tant que *créateur* ; ces relations ont été créées afin de constituer des liens entre auteurs et manifestations quand une œuvre ne pouvait être extraite ou inférée ; les propriétés étaient la plupart du temps utilisées pour contourner la chaîne logique des processus FRBR dans le cas d'un 1:1:1, à savoir des œuvres avec seulement une expression et une manifestation. Pour cette relation très ordinaire, aucune Œuvre /Expression n'est prise en compte.
- OP3005 relie une personne à une manifestation en tant que *contributeur*. Des auteurs mentionnés dans des notices bibliographiques comme entrées additionnelles sont liés avec cette propriété puisque rien de plus précis ne peut être constaté. Aucune relation entre auteurs et expressions n'est prise en compte.

Des relations plus complètes entre les Œuvres ou entre les Agents n'ont pas été pleinement prises en compte à ce stade. Des relations entre une manifestation et une Œuvre / Expression / Manifestation exprimées à travers un point d'accès dans les notices bibliographiques ont été exprimées à travers un « *a un lien* » (bne:OP3007) plus générique, avec le domaine *Manifestation*.

Pour les Sujets, nous avons trituré les données en MARC21 pour les convertir selon une structure FRBR. Dans les manifestations, des relations Sujets ont été modélisées en une relation « *a [entité] comme sujet* ». Mais des Sujets ont été rehaussés d'un niveau et rattachés à des Œuvres, chaque fois que celles-ci existaient :

Software libre

Término aceptado	Software libre
Usado por	Software de código abierto
Término genérico	Software
Término relacionado	Propiedad intelectual--Programas de ordenador
Nota de alcance	Úsase para obras sobre aquellos programas de ordenador que permiten al usuario ejecutarlos, copiarlos, distribuirlos, estudiarlos libremente
Término equivalente en otros vocabularios	http://id.loc.gov/authorities/subjects/sh99003437
Esquemas	http://datos.bne.es/scheme/conceptScheme http://datos.bne.es/scheme/topicalTerms

Obras

Implantació de sistemes de programari lliure

Albós Raya, Amadeu

(2 versiones)

Obra

Software libre: técnicamente viable, económicamente sostenible y socialmente justo

Mas, Jordi

(2 versiones)

Obra

E-business colaborativo

García Valcárcel, Ignacio

(1 versión)

Obra

Page sujet pour "Logiciel Libre" ; en dessous, les Œuvres connectées

Améliorations possibles pour le futur en matière de FRBRisation, des éléments de données ou des processus de travail :

- Envisager d'inférer et de créer des œuvres à partir de fouilles de données, en les reversant peut-être plus tard dans le catalogue, comme il a été proposé ailleurs (Ilien, 2015).
- Mieux représenter les œuvres complexes et les relations entre œuvres.
- Utiliser des enregistrements d'interprétation pour construire des expressions musicales.
- Reconsidérer et améliorer des relations de Sujet non exprimées comme concepts reliés dans le thesaurus, mais comme aspects (par exemple, relier *Bibliothèques* à *Histoire des Bibliothèques* ou à *Bibliothèque – Hongrie...*), ou entre des entités (relier *Poe*, *Edgar Allan à Poe*, *Edgar Allan- Adaptations musicales*).

Expériences et retours des utilisateurs :

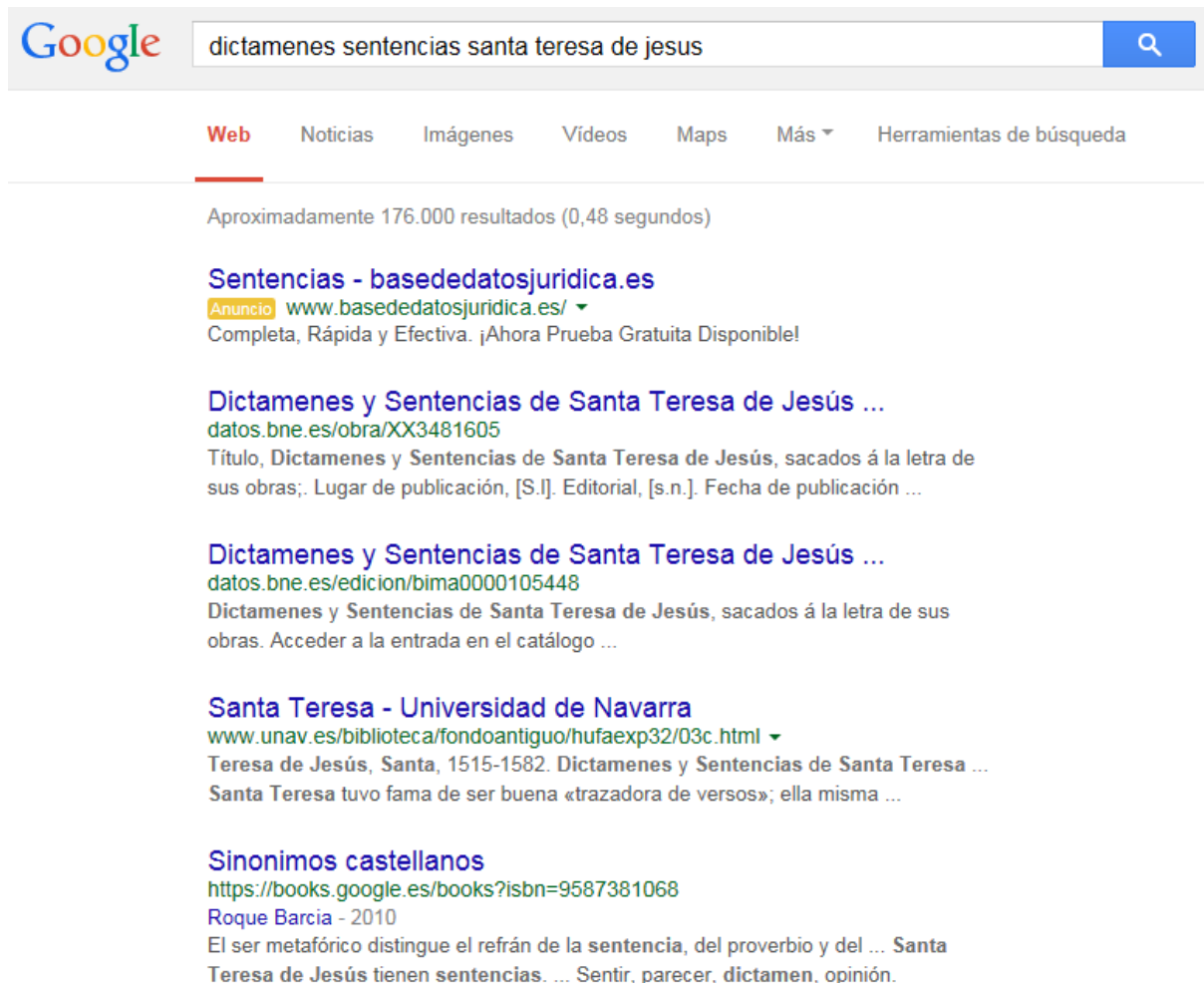
Alors que dans datos.bne version 1.0 le public visé était prioritairement des ordinateurs et des usagers spécialisés, la cible a été élargie dans la version 2.0 aux chercheurs et aux usagers courants. Nous nous interrogeons pour savoir comment les personnes habituées à la structure et l'affichage d'informations dans les OPAC allaient chercher et naviguer à travers des entités, et si elles seraient capables de sélectionner des entités en fonction de leurs préférences et descendre à partir de là jusqu'aux ressources. Nous étions plus particulièrement intéressés par la réception, la compréhension et l'usage du concept abstrait d'« Œuvre ».

En outre, nous voulions mesurer l'impact des bonnes pratiques du Web Sémantique et du Linked Open Data suivies dans ce projet sur le fait d'atteindre les non-publics des bibliothèques.

Datos.bne.es vit séparément de l'entrée principale du catalogue de la bibliothèque (<http://catalogo.bne.es>), de sorte qu'un lien direct a été établi entre cet OPAC et celui du fichier autorité (<http://catalogo.bne.es/uhtbin/authoritybrowse.cgi>), afin d'inciter les usagers à essayer une expérience alternative d'exploration des ressources de la bibliothèque. Le trafic principal a ainsi été canalisé par cette voie.

Bien entendu, les usages experts des données de datos par les communautés du Web Sémantique et Linked Open Data, les groupes d'intérêts et les spécialistes ont donné lieu à des expériences, projets et outils nouveaux utilisant les données culturelles (García, 2014).

La hausse du trafic en provenance directe des moteurs de recherche, rendue possible par le fait que chaque entité et ressource possède une URL permanente, a été particulièrement remarquable. Le classement dans la liste des résultats a été extraordinairement amélioré, particulièrement pour les ressources rares, les auteurs moins connus et les titres spécifiques d'œuvre rare.



The image shows a Google search interface. The search bar contains the text "dictámenes sentencias santa teresa de jesus". Below the search bar, there are navigation tabs for "Web", "Noticias", "Imágenes", "Vídeos", "Maps", "Más", and "Herramientas de búsqueda". The search results are displayed below, showing approximately 176,000 results in 0.48 seconds. The first result is "Sentencias - basededatosjuridica.es" with a URL "www.basededatosjuridica.es/". The second result is "Dictámenes y Sentencias de Santa Teresa de Jesús ..." with a URL "datos.bne.es/obra/XX3481605". The third result is "Dictámenes y Sentencias de Santa Teresa de Jesús ..." with a URL "datos.bne.es/edicion/bima0000105448". The fourth result is "Santa Teresa - Universidad de Navarra" with a URL "www.unav.es/biblioteca/fondoantiguo/hufaexp32/03c.html". The fifth result is "Sinonimos castellanos" with a URL "https://books.google.es/books?isbn=9587381068".

Capture d'écran des résultats de recherche dans Google pour une édition rare de Sainte Thérèse d'Avila

Dans la centaine d'e-mails reçus jusqu'à présent, on a pu lire des questions techniques portant sur l'ontologie et le vocabulaire, l'accès à Sparql, la modélisation, mais les requêtes les plus abondantes provenaient d'utilisateurs non professionnels. Ces dernières se sont avérées être une source d'information appréciable sur nos forces et nos faiblesses, et nous ont fourni quelques pistes pour des développements ultérieurs. Ci-dessous voici un condensé des conclusions les plus pertinentes :

- Exposition et reconnaissance des auteurs. Ces derniers se sont enfin vus attribuer le rôle central qu'ils méritent parmi les données de bibliothèque. Ils ont leur propre page, leur propre jeu de données portant sur eux et pouvant être potentiellement alimenté par de nouvelles données, des relations avec d'autres auteurs, ou d'autres sujets, etc. Leurs contenus sont bien délimités, à savoir pour les auteurs que leur corpus d'œuvre peut être accessible facilement. De sorte que des auteurs vivants (ou parfois leurs descendants et ayant-droits) se sont montrés désireux de fournir des données personnelles (lieu ou date de naissance, pages personnelles) ou de signaler des erreurs de paternité. Cela s'est montré également utile pour mettre en évidence les lacunes ou même impulser des donations.
- Exposition des ressources. Comme il a déjà été dit, le titre propre de ressources données ont amélioré leur classement par les moteurs de recherche, et conduit les usagers cherchant des ressources particulières directement dans les services de la bibliothèque. Ainsi, une des tâches principales du service de réponse par e-mail de `datos.bne.es` a été de rediriger efficacement les requêtes des usagers vers des services compétents (pour la musique, très souvent, mais également pour les livres anciens, les monographies modernes et les enregistrements vidéo). Ces requêtes portaient sur la manière d'obtenir un exemplaire d'un item, ou d'y accéder, ou sur les droits de reproduction.
- Surexposition. D'autre part, au cours de la très courte existence de `datos.bne.2.0`, quelques auteurs se sont plaints de voir ressortir leur nom « dans le cloud ». Des auteurs qui existaient déjà dans notre catalogue traditionnel depuis des années ont soudainement demandé que leurs données soient retirées du service, en raison du fait que leurs données apparaissaient lorsqu'ils entraient leur nom dans un moteur de recherche.
- Fausses représentation des services fournis par `datos.bne.es`. Peut-être en raison du nouvel affichage et de la disposition des données, assez éloigné d'un environnement traditionnel de bibliothèque, quelques usagers provenant directement de Google et d'autres moteurs de recherche n'ont pas compris où ils venaient d'arriver, et ont demandé comment acquérir un livre ou comment obtenir les coordonnées des auteurs. En outre, une incompréhension relativement fréquente a été relevée à propos de certains concepts utilisés ailleurs ou dans la section *à propos* ou dans l'aide en ligne, comme l'Open Data, la réutilisation des Données, les téléchargements de données, certains utilisateurs confondant les métadonnées et le contenu, et demandant ainsi comment télécharger un livre ou une chanson, ou même un film.
- Affichage et tris FRBR. D'après notre expérience, les concepts FRBR doivent être adoptés, mais il convient d'éviter la terminologie FRBR. Les retours d'utilisateurs reçus nous ont conduits à repenser et à clarifier l'usage du concept et du terme même

d'Œuvre. C'est un concept que nous estimons certainement utile, mais qui doit être présenté aux usagers de manière transparente. Ses bénéfices sont de deux ordres : nous épargnons du temps à l'utilisateur en rassemblant des éditions similaires, en particulier dans de grandes institutions bibliographiques où les universitaires ont l'habitude d'examiner et de comparer des documents entre eux afin de retracer une histoire bibliographique d'une Œuvre, et nous fournissons une base solide pour établir des relations à d'autres Œuvres. Par exemple, un usager a été très heureux de pouvoir découvrir rapidement quels étaient les traducteurs hispanophones du *Petit Prince* de Saint-Exupéry, pour les recherches qu'il menait à ce propos (il cherchait parmi les Expressions sans le savoir). Mais naviguer entre Œuvres et Manifestations s'est avéré plus compliqué pour certains usagers, incapables de faire la différence entre les premières et les secondes, notamment dans les relations 1-1, comme mentionné plus haut. Nous espérons combler ces failles dans la compréhension des relations élémentaires du modèle FRBR au cours des améliorations futures, soit au niveau de l'affichage, soit à celui des données, ou bien aux deux.

Conclusion et autres perspectives possibles de développement

Outre toutes les améliorations dessinées plus haut, certaines décisions stratégiques supplémentaires doivent être prises dans le futur. Ayant établi l'architecture qui sous-tend la structure générale de `datos.bne`, de plus en plus de lignes de recherche, d'expérimentation et de développement peuvent être envisagées dans de prochaines versions du service. En voici quelques-unes :

- Davantage de contenus, et d'entités, s'il vous plaît. Le modèle de données et la visualisation fondée sur des entités se sont avérés essentiels dans la construction d'une architecture de données robuste et adaptable. Il serait donc souhaitable de produire un plus grand nombre d'entités, et avec elles, un plus grand nombre de relations doivent être établies. Certains sont inhérents aux données de la bibliothèque, d'autres non. De toute évidence, la première étape concerne les données qui ne sont pas encore dans `datos.bne` (les périodiques, qui sont plus difficiles à intégrer au monde des FRBR, ainsi que les objets assortis d'informations pertinentes, comme leurs propriétaires et leur provenance, qui doivent être soigneusement liés avec le reste des entités). Mais nous sommes également intéressés par les Années, les Lieux et les Editeurs, plus difficiles à extraire et à manipuler.
- Extraire plus profondément. Des sources externes de données ne sont pas encore exploitées pleinement. Un examen précis des sources disponibles (à la fois dans le monde des bibliothèques et en dehors) est requis, il s'agirait de sélectionner des ensembles de données utiles, de concevoir et d'établir des portails pour échanger et extraire des données utilisables.
- Changer les esprits. Les catalogueurs doivent penser à la manière dont les données doivent être modélisées lorsqu'ils pratiquent quotidiennement le catalogage ; le web sémantique commence avec l'humain qui produit des données et facilite le liage lors de l'enregistrement des données. La formation et la sensibilisation sont des composantes absolument essentielles pour faire fonctionner correctement le dispositif.

L'architecture et l'affichage de données fondées sur l'entité profitent pleinement aux usagers, qu'ils soient experts, chercheurs, grand public ou bibliothécaires de référence. Bien que

datos.bne.es en soit encore à un stade précoce de développement, nous avons le sentiment d'être arrivés à un point de non-retour.

Références

GARCÍA, A. (2014). *Integración de contenidos semánticos en un portal web de científicos y humanistas valencianos: Vestigium*. Ph. D. Thesis Universidad Católica de Valencia San Vicente Mártir [Spanish].

Consulté le 20 février 2015 à l'adresse : <http://eprints.rclis.org/24649/>

BOULET, V. (2013). S'appuyer sur la structure des données et les liens - le format MARC comme tremplin pour le Web de données : l'exemple de data.bnf.fr. *79th IFLA World Library and Information Congress*

ILLIEN, G.; PEYRARD, S. (2015) Trust me if you dare: data quality in the age of Linked data. *Europeana Tech Conference* [Présentation Power point].

Consulté le 14 avril 2015 à l'adresse : <http://www.slideshare.net/Europeana/illien-peyrardeurtech2015?qid=52c1f6fb-70d3-4782-bf62-dac3620d107a&v=ql1&b>

MANCHADO, A.; VILA SUERO, D. (2011) Linked data: un caso de uso en la BNE [Power point slides, in Spanish]. *XII Jornadas Españolas de Documentación*, Málaga, 25-27 mayo de 2011.

Consulté le 12 mars 2015 à l'adresse : <http://www.slideshare.net/bne/un-caso-de-uso-en-la-bne>

VILA SUERO, D.; VILLALÓN-TERRAZAS, B.; GÓMEZ-PÉREZ, A. (2013) Datos.bne.es: A library linked dataset. *Semantic Web*, vol. 4, n° 3 (2013), p. 307–313