

La sémantisation des données à la Bpi

Sémantiser ses données :

Sémantiser, c'est transformer ses données selon les principes et les normes du web de données ou web sémantique.

Cela repose en particulier sur le [standard RDF](#), qui organise les données selon des triplets : sujet – prédicat – objet.

Objectifs :

Proposer des données plus largement interoperables et réutilisables sur le web.

Offrir de nouvelles possibilités d'exploitation permise par le format RDF (notamment les relations.)

Pourquoi sémantiser les données à la Bpi ?

- Anticiper les évolutions liées à la transition bibliographique.
- Tester et expérimenter les possibilités offertes par ces nouveaux formats.
- Valoriser les données spécifiques à la Bpi comme la classification ou le cinéma documentaire.

Quels critères de choix des corpus ?

- Des corpus bien circonscrits, faciles à identifier et cohérents.
- Des données produites par la Bpi.
- Des données de qualité.

Deux corpus traités :

- La classification : une classification adaptée de la CDU et propre à la Bpi, dont chaque indice fait l'objet d'une notice autorité. Notices simples et très homogènes avec peu de champs unimarc.

[Les Yeux doc](#) : un service d'offre de diffusion de cinéma documentaire, porté par la Bpi et accessible à toutes les bibliothèques partenaires, pour offrir l'accès au public à un catalogue de 450 films.

Des questions encore en suspens pour les autres données :

- Que faire des données bibliographiques non-produites par la Bpi ?
- Est-ce qu'on transforme les données depuis notre base en RDF ?
- Est-ce qu'on récupère directement des données sémantisées depuis la BnF ou du Sudoc ?
- Est-ce qu'il est possible d'exploiter les données sémantisées de la BnF/Sudoc par un simple alignement ?

Quels usages des données ?

- Offrir de nouveaux services au public exploitant les possibilités du nouveau format.
- Décloisonner nos sites, en utilisant une même donnée pour l'ensemble des sites administrés par la Bpi.
- Permettre de nouveaux modes d'interrogation des données beaucoup plus fins.
- Enrichir les données, en s'alignant sur des bases externes.
- Permettre l'ouverture et le partage des données.

La sémantisation de la classification et son usage dans l'outil de cartographie

Une notice d'indice et sa signification selon les règles de la classification.

Indice	Hierarchie	Élément 1	Élément 2	Description
284(44) Protestantisme . France	284 Protestantisme	284 Protestantisme	(44) France	<p>L'indice 284(44) Protestantisme. France un indice construit avec une subdivision géographique. Il a pour indice parent l'indice 284 Protestantisme.</p> <p>Il est composé de deux éléments : l'indice sujet 284 Protestantisme et la subdivision géographique (44)</p>

Une particularité de la Bpi : des notices autorités pour les indices en Unimarc.

001:# NOTICE	
152:REGL.CAT.IND	\$aBP
253:INDICE CDU	\$a284(44)\$cProtestantisme. France
900:DONNEES LOC.	\$aretl

- Le 253\$a indique le code de l'indice, le \$c indique son libellé.
- Le code en 900\$a indique que la notice est construite avec une subdivision de lieu.
- On peut parfois trouver des champs de forme rejetées (453\$a\$c) ou des notes de définition/d'usage.

Ces données ne permettent pas de décrire l'ensemble des règles hiérarchiques et de construction des indices !

En RDF, on peut exprimer l'ensemble des informations souhaitées :

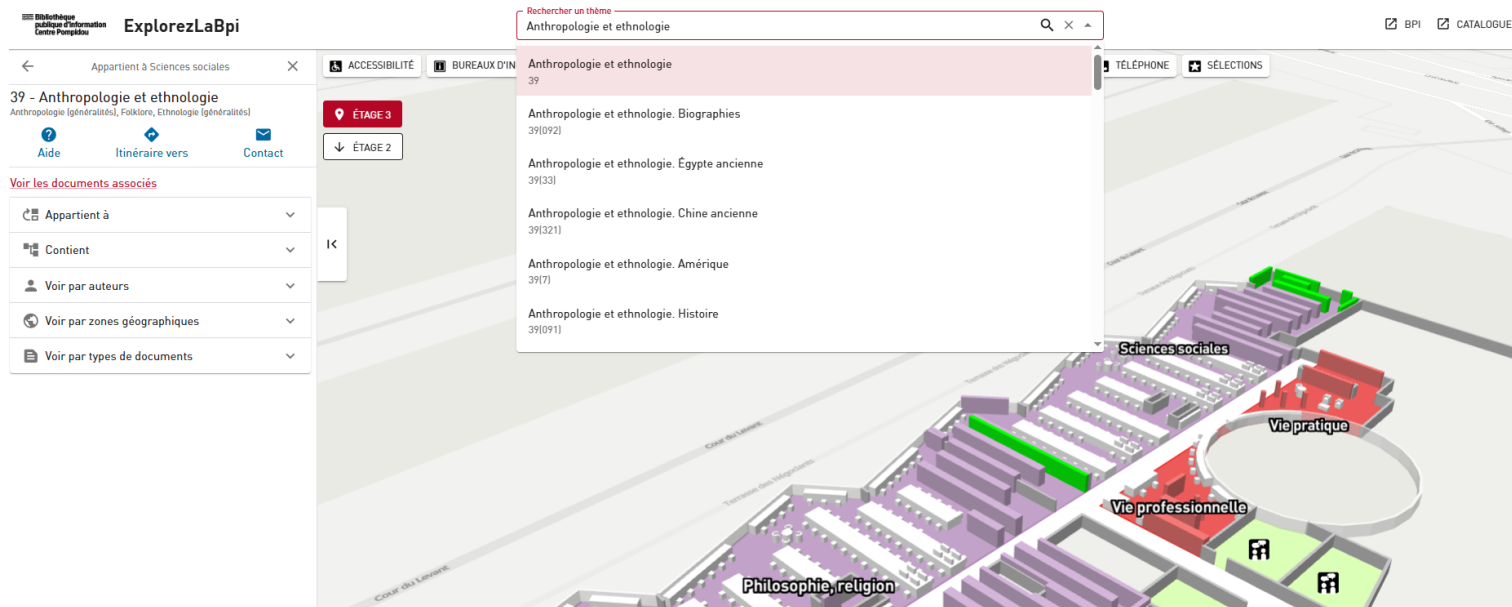
- En les formulant sous la forme d'un triplet Sujet - Prédicat - Objet. (l'équivalent informatique d'une phrase simple sujet - verbe - complément)
- En langage naturel :
- l'indice 284(44) a pour parent l'indice 284
- l'indice 284(44) est composé de l'élément de subdivision géographique (44) France
- Sous forme de triplet RDF :
- <http://data.bpi.fr/classification/0000572983> skos:broader <http://data.bpi.fr/classification/00000000652>
- <http://data.bpi.fr/classification/0000572983> skos:related <http://data.bpi.fr/classification/00000000082>

Les données en RDF de la Bpi

Outil de visualisation des données SparqlExplorer de Logilab

http://data.bpi.fr/classification/0000584680	dcterms:created	2002-10-30	default
http://data.bpi.fr/classification/0000584680	dcterms:identifier	584680	default
http://data.bpi.fr/classification/0000584680	dcterms:modified	2019-11-05	default
http://data.bpi.fr/classification/0000584680	dcterms:source	https://catalogue.bpi.fr/rautorite/ark:/34201/ptf0000038306	default
http://data.bpi.fr/classification/0000584680	dcterms:spatial	http://data.bpi.fr/location/299	default
http://data.bpi.fr/classification/0000584680	mads:authoritativeLabel	Protestantisme, Allemagne	default
http://data.bpi.fr/classification/0000584680	mads:code	284(43)	default
http://data.bpi.fr/classification/0000584680	mads:componentList	http://data.bpi.fr/classification/0000584680#componentList	default
http://data.bpi.fr/classification/0000584680	mads:elementList	http://data.bpi.fr/classification/0000584680#elementList	default
http://data.bpi.fr/classification/0000584680	mads:hasBroaderAuthority	http://data.bpi.fr/classification/0000628982	default
http://data.bpi.fr/classification/	mads:hasMADSSchemeMember	http://data.bpi.fr/classification/0000584680	default
http://data.bpi.fr/classification/0000628982	mads:hasNarrowerAuthority	http://data.bpi.fr/classification/0000584680	default
http://data.bpi.fr/classification/0000584680	mads:isMemberOfMADSScheme	http://data.bpi.fr/classification/	default
http://data.bpi.fr/classification/0000584680	rdf:type	mads:Authority	default
http://data.bpi.fr/classification/0000584680	rdf:type	mads:ComplexSubject	default
http://data.bpi.fr/classification/0000584680	rdf:type	skos:Concept	default
http://data.bpi.fr/classification/0000584680	rdfs:seeAlso	https://catalogue.bpi.fr/indice-cdu/ark:/34201/ptf00000584680	default
http://data.bpi.fr/classification/0000584680	skos:broader	http://data.bpi.fr/classification/0000628982	default

Exemple de l'exploitation de la classification dans la nouvelle [carte interactive de la Bpi](#) :



Une recherche thématique utilise les données skos comme règle de pertinence pour offrir en premier résultat les résultats les plus haut dans la hiérarchie.

La localisation des collections affiche l'emplacement d'un indice et de l'ensemble de ses enfants, pour permettre aux usagers de trouver tous les documents sur une thématique.

Le menu de gauche permet une série de rebond dans la classification grâce aux données RDF.