



**HAL**  
open science

# Préparer un catalogue à la transition bibliographique : le cas des métadonnées des documents patrimoniaux de la Médiathèque de Roubaix

Margaux Lainé

## ► To cite this version:

Margaux Lainé. Préparer un catalogue à la transition bibliographique : le cas des métadonnées des documents patrimoniaux de la Médiathèque de Roubaix. Sciences de l'information et de la communication. 2023. dumas-04442084

**HAL Id: dumas-04442084**

**<https://dumas.ccsd.cnrs.fr/dumas-04442084>**

Submitted on 6 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Université  
de Lille**



**Margaux LAINÉ**

**Master 2 GIDE**

Gestion de l'Information et du Document en Entreprise

**MÉMOIRE DE STAGE**

Mission effectuée du 4 avril au 5 août 2023

à la

**MÉDIATHÈQUE LA GRAND PLAGE DE ROUBAIX**

**PRÉPARER UN CATALOGUE À LA TRANSITION BIBLIOGRAPHIQUE –**

**Le cas des métadonnées des documents patrimoniaux de la**

**Médiathèque de Roubaix**

**Sous la direction de :**

**Monsieur E. KERGOSIEN (tuteur universitaire)**

**Monsieur F. PICHENOT (tuteur professionnel)**

**Soutenu le 12 septembre 2023 à l'UFR DECCID-SID  
Université de Lille, Sciences Sociales et Humaines  
BP 60 149, 59 653 Villeneuve d'Ascq Cedex**

**Année Universitaire 2022/2023**



## Remerciements

*Je souhaite commencer par remercier toutes les personnes qui m'ont accompagnée au cours de mon stage et durant la rédaction de ce mémoire.*

*Je remercie sincèrement Monsieur François Pichenot, responsable du Pôle Société, Sciences et Numérique de la Médiathèque de Roubaix, de m'avoir accueillie dans son service.*

*Grâce à votre expertise et à votre pédagogie, j'ai pu gagner en confiance en moi et développer mes compétences malgré une mission technique et pointue.*

*J'en profite pour remercier également toute l'équipe de La Grand Plage pour son accueil chaleureux et sa bienveillance. Je remercie particulièrement Anne-Sophie, Julie et Mathilde de m'avoir intégrée dans leur bureau et fait découvrir leurs missions. Une pensée à Clara, Clément et Lucas pour leur bonne humeur et leur motivation.*

*Je tiens à remercier également Monsieur Éric Kergosien, mon tuteur universitaire, pour son aide précieuse tout au long de mon stage.*

*Enfin, je remercie l'ensemble de la promotion du Master 2 GIDE ainsi que l'équipe enseignante pour cette agréable dernière année d'études passée à leurs côtés.*

*Une dernière pensée à mes proches, pour les remercier de m'avoir toujours soutenue, tout au long de ma scolarité.*

## Résumé

Ce mémoire est l'aboutissement d'un travail de réflexion et de recherche autour de l'exposition des notices bibliographiques sur le web de données. Il a été réalisé dans le cadre du Master 2 Gestion de l'Information et du Document en Entreprise (GIDE) de l'Université de Lille et plus particulièrement d'un stage à la Médiathèque La Grand Plage de Roubaix. Dans le contexte de la Transition Bibliographique, les données bibliographiques et d'autorités des catalogues de bibliothèques ont vocation à être rendues davantage visibles pour les usagers, en adoptant les standards du web sémantique. Pour atteindre cet objectif, les bibliothèques françaises, encadrées par l'Agence bibliographique de l'enseignement supérieur (Abes) et la BnF (Bibliothèque nationale Française), doivent travailler sur leurs métadonnées afin d'arriver progressivement à un catalogage selon le code RDA-FR (Ressource Description Accès). Ainsi, le but du projet est d'aligner les métadonnées locales de la Médiathèque de Roubaix à celles des agences bibliographiques nationales mais aussi de mettre en place une méthodologie propre à la médiathèque, par rapport aux collections qu'elle conserve. En effet, la problématique du projet concerne particulièrement les métadonnées des documents patrimoniaux, plus complexes à traiter.

### *Descripteurs*

Transition Bibliographique ; ouverture des données ; web de données ; données liées ; métadonnées d'identification ; bibliothèque ; notices bibliographiques ; catalogue en ligne ; documents patrimoniaux ; usagers.

## Abstract

### **Preparing a catalogue for Bibliographic Transition – Focus on metadata for heritage documents at the municipal Library of Roubaix**

This dissertation is the result of a reflection and research work on the display of bibliographic records on the Web. It was produced as part of the Master 2 GIDE program at the University of Lille, and more specifically during an internship at the Médiathèque La Grand Plage in Roubaix. In the context of the Bibliographic Transition, bibliographic and authority data in library catalogs are to be made more visible to users, by adopting semantic web standards. To achieve this goal, French libraries, under the guidance of Abes and BnF, need to work on their metadata in order to gradually move towards cataloguing according to the RDA-FR code. The aim of the project is therefore to align the local metadata of the Médiathèque de Roubaix with that of the national bibliographic agencies, and also to implement a methodology specific to the library, in relation to the collections it holds. Indeed, the project is particularly concerned with the metadata of heritage documents, which are more complex to process.

#### *Keywords*

Bibliographic Transition ; open data ; web data ; linked data ; identification metadatas ; library ; bibliographic records ; online catalog ; heritage documents ; users.

## Table des matières

<b>Remerciements</b> .....	<b>3</b>
<b>Résumé</b> .....	<b>4</b>
<b>Abstract</b> .....	<b>5</b>
<b>Table des matières</b> .....	<b>6</b>
<b>Table des illustrations</b> .....	<b>8</b>
<b>Liste des tableaux</b> .....	<b>10</b>
<b>Liste des abréviations</b> .....	<b>11</b>
<b>Introduction</b> .....	<b>12</b>
<b>Chapitre 1 : Approche générale et contextualisation du sujet de recherche</b> .....	<b>15</b>
<b>Partie 1 : Présentation de l’environnement de stage</b> .....	<b>15</b>
1. Présentation rapide de la structure d’accueil : la médiathèque La Grand Plage de Roubaix .....	15
2. Mes missions au sein du Pôle Société, Sciences et Numérique .....	16
<b>Partie 2 : Qu’est-ce que la Transition Bibliographique ?</b> .....	<b>18</b>
1. Le volet théorique – état de la littérature .....	18
1.1 La genèse du programme : définition du projet, enjeux principaux.....	18
1.2 Ce qui est attendu des bibliothèques françaises .....	21
1.3 Un changement de paradigme : du document vers la donnée, quelle finalité pour les usagers ?.....	23
2. Le volet technique – comprendre la logique derrière le programme de la Transition bibliographique .....	25
2.1 L’ouverture des données, quels enjeux pour les bibliothèques ?.....	25
2.2 Sortir de l’isolement des données en adoptant la logique du linked data.....	26
2.3 La mise en place d’un nouveau code de catalogage adapté aux technologies du web de données .....	28
<b>Chapitre 2 : Mise en application et travail concret d’alignement des notices</b> .....	<b>30</b>
<b>Partie 1 : Méthodologie d’alignement des notices bibliographiques via le logiciel Bibliostratus</b> .....	<b>30</b>
1. Présentation rapide du corpus.....	30
2. Utilisation de Bibliostratus – les étapes à suivre pour réaliser l’alignement .....	30
2.1 Présentation des trois modules selon leur ordre d’utilisation .....	31
2.2 Focus sur le processus d’alignement – module blanc .....	35
3. Alignement du corpus test.....	36
3.1 La conversion des fichiers de notices Unimarc en tableaux .....	36

3.2	L’alignement des notices bibliographiques à celles de la BnF / du Sudoc.....	37
<b>Partie 2 : Étude des résultats obtenus après le premier alignement et manipulation des fichiers pour obtenir le fichier de notices alignées.....</b>		<b>38</b>
1.	Analyse des résultats obtenus suite au premier alignement et constat des problèmes rencontrés .....	38
1.1	Exposé des résultats obtenus.....	38
1.2	Analyse des résultats obtenus et constat des problèmes rencontrés.....	40
1.3	Observations complémentaires .....	47
2.	Manipulation des fichiers en entrée pour améliorer les résultats obtenus .....	48
3.	Bilan du travail d’alignement des notices .....	55
<b>Partie 3 : Discussion globale .....</b>		<b>56</b>
1.	Pistes de réflexion pour approfondir le travail .....	56
2.	Limites et obstacles rencontrés.....	56
3.	Ouverture. Quels enjeux concrets pour les bibliothèques ?.....	57
<b>Conclusion générale.....</b>		<b>59</b>
<b>Références bibliographiques .....</b>		<b>60</b>
<b>Ouvrages .....</b>		<b>60</b>
<b>Articles .....</b>		<b>60</b>
<b>Sites web.....</b>		<b>61</b>
<b>Références juridiques .....</b>		<b>62</b>
<b>Annexes .....</b>		<b>63</b>

## Table des illustrations

- Illustration 1 : Les étapes de la Transition Bibliographique.....p.18
- Illustration 2 : Liste des tâches utilisateurs à l'origine de la création du modèle IFLA-LRM.....p.19
- Illustration 3 : Le modèle de données de data.bnf.fr.....p.20
- Illustration 4 : Schéma organisationnel de la Transition Bibliographique – décembre 2022.....p.21
- Illustration 5 : Apports du web de données en bibliothèque.....p.27
- Illustration 6 : Organisation des sous-groupes « Normalisation » de la Transition bibliographique.....p.28
- Illustration 7 : Interface d'accueil de Bibliostratus et modules proposés.....p.31
- Illustration 8 : Visualisation du « module bleu » de Bibliostratus.....p.32
- Illustration 9 : Visualisation du « module blanc » de Bibliostratus pour l'alignement des données bibliographiques.....p.33
- Illustration 10 : Visualisation du « module blanc » de Bibliostratus pour l'alignement des données d'autorité.....p.34
- Illustration 11 : Visualisation du « module rouge » de Bibliostratus.....p.35
- Illustration 12 : Capture d'écran des fichiers obtenus après le passage dans le module bleu de Bibliostratus.....p.36
- Illustration 13 : Métadonnées d'identification associées au premier module de Bibliostratus.....p.36
- Illustration 14 : Métadonnées d'identification associées au module d'alignement de Bibliostratus.....p.37

- Illustration 15 : Extrait du tableau Excel contenant les résultats obtenus après l’alignement des notices des monographies.....p.37
  
- Illustration 16 : Notice bibliographique (Unimarc) sur le SIGB de la Médiathèque de Roubaix.....p.41
  
- Illustration 17 : Notice bibliographique (Unimarc) sur le catalogue de la BnF.....p.42
  
- Illustration 18 : Capture d’écran de la notice de l’ouvrage de A. Le Roux de Lincy, sur le catalogue de la médiathèque de Roubaix.....p.43
  
- Illustration 19 : Capture d’écran de la notice de l’ouvrage de A. Le Roux de Lincy, sur le catalogue de la BnF.....p.43
  
- Illustration 20 : Capture d’écran de la notice de l’ouvrage *Physiologie du calembourg*, sur le catalogue de la médiathèque de Roubaix.....p.44
  
- Illustration 21 : Capture d’écran de la notice de l’ouvrage *Physiologie du calembourg*, sur le catalogue de la BnF.....p.44
  
- Illustration 22 : Photographie de la première page de l’ouvrage *Les français peints par eux-mêmes*..... p.45
  
- Illustration 23 : Capture d’écran de l’ouvrage de Longus, sur le catalogue de la médiathèque de Roubaix.....p.46
  
- Illustration 24 : Capture d’écran de la page d’édition des préférences de Bibliostratus.....p.48
  
- Illustration 25 : Capture d’écran – extrait du fichier de résultat du second alignement des partitions.....p.54

## Liste des tableaux

- Tableau 1 : Résultats du premier alignement (monographies).....p.39
- Tableau 2 : Résultats du premier alignement (partitions).....p.39
- Tableau 3 : Résultats du premier alignement (périodique).....p.40
- Tableau 4 : Résultats du deuxième passage dans Bibliostratus (sans P.Destombes).....p.50
- Tableau 5 : Résultats du troisième passage dans Bibliostratus (sans P.Destombes, sans les tomes).....p.51
- Tableau 6 : Résultats du quatrième passage dans Bibliostratus (sans P.Destombes, sans les tomes, avec contrôle éditeurs).....p.53

## Liste des abréviations

Abes : Agence bibliographique de l'enseignement supérieur

Afnor : Association française de normalisation

ARK : *Archival Resource Key*

BnF : Bibliothèque nationale de France

BNR : Bibliothèque Numérique de Référence

FNE : Fichier National d'Entités

FRBR : *Functional Requirements for Bibliographic Records*

IdRef : Identifiants et Référentiels pour l'Enseignement Supérieur et la Recherche

IFLA : *International Federation of Library Associations and Institutions*

ISBN : *International Standard Book Number*

LRM : *Library Reference Model*

MARC : *MAchine Readable Cataloging*

PPN : *Pica Production Number*

RDA : *Resource Description Access* / RDA-FR : transposition française de RDA

RDF : *Resource Description Framework*

SIGB : Système Intégré de Gestion de Bibliothèque

SPARQL : *SPARQL Protocol and RDF Query Language*

SRU : *Search / Retrieval via URL*

Sudoc : Système universitaire de documentation

UNIMARC : *UNIversal MAchine Readable Cataloging*

WWW : *World Wide Web*

W3C : *World Wide Web Consortium*

XML : *Extensible Markup Language*

## Introduction

« *L'Humanité est à un tournant de son histoire. La masse des données acquises est formidable. Il faut de nouveaux instruments pour les simplifier, les condenser ou jamais l'intelligence ne saura ni surmonter les difficultés qui l'accablent, ni réaliser les progrès qu'elle entrevoit et auxquelles elle aspire* » (Otlet, 1934).

Depuis les années 1990, le développement d'Internet et des technologies de l'information et de la communication bouleverse le paysage informationnel de la société. Grâce à la numérisation des informations et au partage des données de plus en plus massives, on assiste à une véritable révolution numérique. Dans ce contexte d'infobésité, la recherche d'information sur le web est complexifiée. Il faut désormais réfléchir à l'organisation de cette quantité de données sur le web, afin de faciliter la recherche pour les utilisateurs. Tim Berners-Lee, inventeur du World Wide Web (WWW) et fondateur du World Wide Web Consortium (W3C) réfléchit à cette problématique et, en 2006, propose la notion de « *linked data* » (Berners-Lee, 2006) qui donnera naissance au web des données. L'idée est alors de penser le web comme un ensemble de « données structurées », liées entre elles, afin de « *constituer un réseau global d'informations* »<sup>1</sup>. Cette volonté de lier les données sur le web a surtout pour objectif de rendre les recherches des utilisateurs plus pertinentes. Pour cela, on assiste à un mouvement d'ouverture des données publiques, dans tous les pans de la société, y compris dans le monde de la culture et de la lecture publique.

Problématique au cœur du sujet actuel de l'ouverture des données, la Transition bibliographique vise à rendre davantage visibles les catalogues des bibliothèques sur le web, et donc leurs collections. Pour atteindre cet objectif, les bibliothèques françaises, encadrées par l'Agence bibliographique de l'enseignement supérieur (Abes) et la Bibliothèque nationale de France (BnF), doivent travailler sur leurs métadonnées afin d'arriver progressivement à un catalogage selon le code RDA-FR (Ressource Description Accès). La Transition bibliographique pose ainsi des questions techniques, d'alignement des métadonnées des bibliothèques locales, à celles des agences bibliographiques nationales.

---

<sup>1</sup> Vers de nouveaux SGB. Site web de la Transition Bibliographique. En ligne : <https://www.transition-bibliographique.fr/systemes-et-donnees/recommandations/>.

Afin de réaliser cet alignement, les bibliothèques disposent de différents outils et technologies produits par le programme de la Transition bibliographique. En effet, le groupe « Systèmes et Données » a développé le logiciel Bibliostratus qui permet l'alignement des métadonnées locales à celles de la BnF et de l'Abes et le groupe « Formation » propose son aide aux professionnels des bibliothèques afin de préparer leurs catalogues à cette transition.

Ainsi, en réalisant l'alignement d'un corpus test de notices, issues d'un fonds patrimonial de la Médiathèque de Roubaix, le but sera, d'une part, d'évaluer l'utilisation de ces outils, en particulier le logiciel Bibliostratus, afin de mettre en lumière son fonctionnement et ses failles, et d'autre part, de construire une méthodologie d'alignement à destination des bibliothèques locales.

Il s'agit donc de se demander comment préparer un catalogue local à la Transition bibliographique. Fil rouge de ce mémoire, cette problématique renvoie à différentes questions, notamment :

- *Quels sont les grands enjeux de la Transition bibliographique pour les bibliothèques françaises ? Comment peuvent-elles agir, chacune à leur échelle ?*
- *Comment utiliser Bibliostratus, logiciel permettant l'alignement des données ? Comment optimiser les fichiers de notices fournis en entrée pour améliorer la qualité des fichiers en sortie ?*
- *Quelles sont les difficultés liées à l'alignement d'un corpus de documents patrimoniaux ?*
- *Comment travailler sur la qualité des données bibliographiques ?*
- *Quelles sont les finalités de la Transition bibliographique pour les usagers des bibliothèques ?*

Le but de ce mémoire de recherche sera de répondre à ces différentes interrogations, afin d'éclaircir les réflexions autour de la Transition bibliographique et de l'alignement des données.

Dans le premier chapitre de cette étude, il s'agira d'adopter une approche générale et de contextualiser le sujet. Ce chapitre présentera d'une part l'environnement du stage ainsi que mes missions puis il exposera un état de la littérature sur les questions théoriques et techniques liées à la Transition bibliographique.

Le second chapitre se consacrera à la mise en application des principes observés et au travail d'alignement du corpus de notices test. Il démontrera à la fois la méthodologie d'alignement des notices via le logiciel Bibliostratus mais aussi une analyse des résultats accompagnée d'hypothèses pour tenter de résoudre les différents problèmes observés. Enfin, ce mémoire se terminera par une discussion globale afin de confronter les différents résultats obtenus.

# Chapitre 1 : Approche générale et contextualisation du sujet de recherche

## Partie 1 : Présentation de l'environnement de stage

### 1. Présentation rapide de la structure d'accueil : la médiathèque La Grand Plage de Roubaix

Située en plein cœur de la ville de Roubaix, la médiathèque La Grand Plage accueille le public dans son bâtiment actuel depuis 1979. Bibliothèque classée par l'État en 1972, elle est dotée de fonds documentaires conséquents et elle ne cesse de renouveler son offre de services à destination de ses usagers. La Grand Plage est en réalité composée de deux services : le service de la médiathèque, qui comprend la médiathèque centrale, le « Zèbre » (bibliobus) et le prêt aux collectivités ; et le service des archives municipales situé dans les bâtiments de la mairie, à deux pas de la bibliothèque.

Bibliothèque inclusive et dynamique, la médiathèque de Roubaix propose une action culturelle riche et variée et est à l'origine de nombreux partenariats, notamment avec des associations ou encore des centres sociaux de la ville. Son objectif est d'accueillir tous les publics, sans distinction, afin de les accompagner dans leurs démarches quotidiennes. La médiathèque possède une large amplitude horaire, afin d'être disponible et de répondre à la demande de ses usagers. La Grand Plage se distingue également par son offre de services numériques, qui comprend à la fois de la médiation numérique sur place pour aider les usagers dans leurs démarches quotidiennes, mais aussi une valorisation du patrimoine roubaisien en ligne, sur la [bn-r](#), bibliothèque numérique de Roubaix. Elle a d'ailleurs été mise en avant par le Ministère de la Culture en 2012, qui lui a attribué le label « Bibliothèque Numérique de Référence » (BNR).

Pour remplir ces missions et contribuer à son rayonnement, la bibliothèque de Roubaix est organisée en différents pôles qui permettent de rendre la médiathèque dynamique<sup>2</sup> :

- le service des archives ;
- le pôle administratif et financier ;
- le pôle patrimoine ;

---

<sup>2</sup> Voir Annexe 1 : Organigramme de la médiathèque de Roubaix

- le pôle jeune public ;
- le pôle loisirs, art et culture ;
- le pôle société, sciences et numérique ;
- le pôle services aux publics.

De plus, étant l'unique médiathèque de la ville de Roubaix, le bibliobus permet de rendre accessibles les collections à tous les usagers, en desservant 15 arrêts dans les quartiers de la ville, pendant toute la semaine. Cela permet aux usagers ne pouvant pas se déplacer jusqu'au centre-ville de pouvoir emprunter et rendre des documents.

Enfin, il est important de noter que la médiathèque de Roubaix se distingue également par sa démarche « Zéro Déchet », qui s'inscrit dans la politique « Zéro Déchet » de la ville de Roubaix. Dans ce sens, elle propose une sélection de documents, qui ne cesse de s'élargir, sur les thématiques de l'éco-responsabilité et des démarches durables à mettre en place au quotidien. Elle propose également régulièrement des ateliers et activités sur le thème du zéro déchet, allant de la confection d'objets réutilisables au troc de boutures par exemple.

## 2. Mes missions au sein du Pôle Société, Sciences et Numérique

La Médiathèque de Roubaix a pris le train de la Transition bibliographique en commençant par l'alignement de ses documents « récents ». En effet, pour ces derniers, la recherche d'identifiants pérennes qui vont permettre d'aligner les notices du catalogue local à celles de la BnF est assez simple puisque ces documents possèdent des métadonnées d'identification assez complètes.

Ainsi, ma mission au sein de la médiathèque concerne plutôt les documents patrimoniaux, qui posent davantage de difficultés pour l'alignement, principalement à cause du manque de précisions dans leurs métadonnées d'identification. L'objectif principal de mon stage est donc de réfléchir à l'alignement du catalogue local de la médiathèque, en particulier sur un fonds patrimonial qu'il s'agira d'aligner. Cela sous-entend également de creuser la question de l'importance de la qualité des métadonnées bibliographiques pour réaliser l'alignement des notices.

Concrètement, au cours de mon stage, il s'agira de travailler sur un corpus d'environ 3 000 notices issues du fonds patrimonial « Destombes », du nom de son donateur roubaisien. Pour réaliser cet alignement, j'utiliserai l'outil Bibliostratus, développé par le groupe « Systèmes et Données » de la Transition bibliographique puis j'analyserai les résultats obtenus après l'alignement. À travers ce corpus restreint, le but est de définir une politique d'alignement pour les documents patrimoniaux et des hypothèses pour parvenir à répondre aux difficultés posées.

Pour réaliser ma mission, différents prérequis sont nécessaires et occupent mes premières semaines de stage<sup>3</sup>. Il s'agit dans un premier temps de bien comprendre les enjeux de la Transition bibliographique, en réalisant une veille documentaire ainsi qu'un état de la littérature sur le sujet. Il est également nécessaire de comprendre les différents formats UNIMARC (*UNIversal MAchine Readable Cataloging*), l'articulation entre les notices bibliographiques et les notices d'autorités ainsi que de savoir manipuler les fichiers de notices pour en extraire des données et les analyser. Enfin, d'un point de vue technique, il était également important de prendre en main l'outil Bibliostratus, de comprendre son fonctionnement et ses mécanismes internes lui permettant de fournir les résultats.

---

<sup>3</sup> Voir Annexe 2 : Planning du projet

## Partie 2 : Qu'est-ce que la Transition Bibliographique ?

### 1. Le volet théorique – état de la littérature

#### 1.1 La genèse du programme : définition du projet, enjeux principaux

##### ➤ Qu'est-ce que la Transition Bibliographique ?

Dans le contexte du web de données et de la volonté de lier les données entre elles sur le web (*linked open data*), les bibliothèques ont un rôle à jouer afin de rendre leurs catalogues davantage visibles aux yeux de tous. C'est pour amener les bibliothèques vers cette logique de web de données que le programme de la Transition bibliographique a vu le jour, en 2015 (voir Illustration 1), sous l'égide des deux agences bibliographiques françaises, l'Agence bibliographique de l'enseignement supérieur (Abes) et la Bibliothèque nationale française (BnF). L'objectif principal de ce programme est donc de « *positionner les catalogues comme des outils efficaces pour améliorer la visibilité de la bibliothèque sur le web, par l'intermédiaire de ses collections* » (Toussaint, 2019).

Pour arriver à cette fin, il s'agit de s'appuyer sur les modèles conceptuels développés par l'*International Federation of Library Associations and Institution* (IFLA), IFLA-FRBR et désormais IFLA-LRM et sur un nouveau code de catalogage, le code RDA-FR (Ressource : Description Accès), pour prendre en compte le « *nouvel environnement numérique des catalogues* » (Cantie, 2015), ce qui implique notamment de travailler sur les métadonnées des catalogues de bibliothèques.

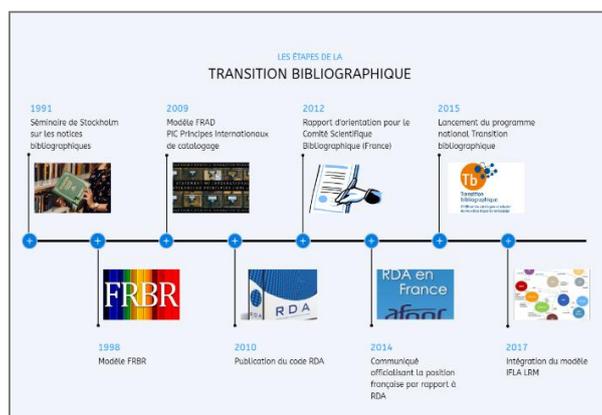


Illustration 1 : Les étapes de la Transition Bibliographique<sup>4</sup>

<sup>4</sup> Source : site web de la Transition bibliographique - <https://www.transition-bibliographique.fr/enjeux/position-francaise-rda/>

En lien avec ces nouveaux mécanismes, l'objectif est également de simplifier la recherche documentaire pour les usagers, en proposant des catalogues adaptés à leurs besoins et à leurs attentes. Cela repose notamment sur les «  *cinq tâches utilisateurs génériques*  » (voir Illustration 2) définies par l'IFLA lors de l'élaboration du modèle conceptuel IFLA-LRM, à savoir : « trouver », « identifier », « sélectionner », « obtenir » et « explorer / naviguer » (Riva *et al.*, 2021).

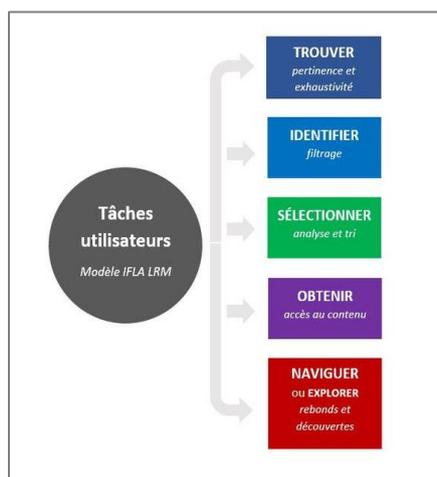


Illustration 2 : Liste des tâches utilisateurs à l'origine de la création du modèle IFLA-LRM<sup>5</sup>

On cherche désormais un modèle plutôt orienté utilisateur. Ainsi, la finalité du projet de la Transition Bibliographique pour les bibliothèques françaises est de réussir une « FRBRisation » de leurs données bibliographiques, en s'appuyant sur les travaux menés par la BnF et l'Abes. Pour avoir une idée de ce que peut représenter cette logique de *open linked data* en lien avec des collections de bibliothèques, il est possible de citer le projet initié par la BnF : data.bnf.fr. À travers ce projet, la BnF a construit une plateforme, pour laquelle elle expérimente le modèle de données inspiré d'IFLA-LRM qui permet une structure spécifique des informations, basée sur les métadonnées d'identification du document (auteur, œuvre, date, etc.). Ainsi, selon ce modèle, on part du principe qu'une « manifestation » matérialise une « expression », elle-même réalisant « l'œuvre »<sup>6</sup> (voir Illustration 3).

<sup>5</sup> Source : « Schéma Tâches utilisateurs IFLA LRM ». *Wikimedia Commons*. En ligne : [https://commons.wikimedia.org/wiki/File:Sch%C3%A9ma\\_T%C3%A2ches\\_utilisateurs\\_IFLA\\_LRM.jpg](https://commons.wikimedia.org/wiki/File:Sch%C3%A9ma_T%C3%A2ches_utilisateurs_IFLA_LRM.jpg)

<sup>6</sup> LRM Entity-Relationship Model. IFLA Standards Info. En ligne : <https://www.iflstandards.info/lrm/lrmer.html>

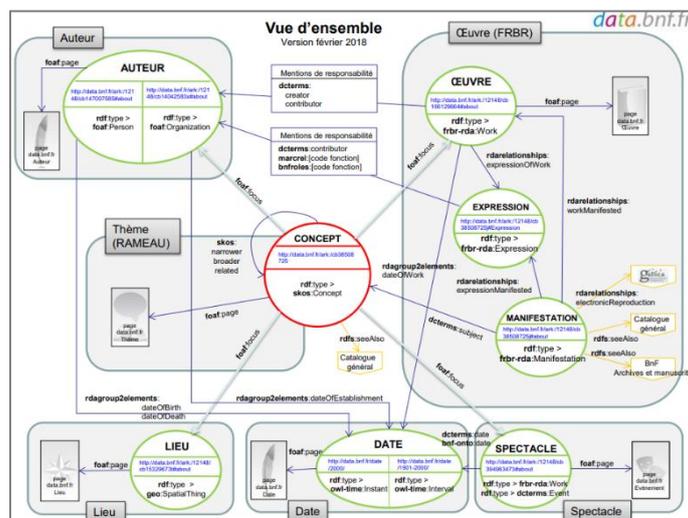


Illustration 3 : Le modèle de données de data.bnf.fr<sup>7</sup>

### ➤ Enjeux principaux

Le programme de la Transition Bibliographique répond à divers enjeux, à la fois du point de vue des usagers, mais aussi au niveau technique et professionnel. Dans un premier temps, il faut mettre en lumière l'enjeu d'efficacité, c'est-à-dire améliorer l'interopérabilité des données des bibliothèques et « adopter les standards du web pour proposer des données liées ouvertes et en favoriser la réutilisation » (Goubina, 2018). Il existe ensuite un enjeu de cohérence puisqu'il s'agit de mettre en lien la chaîne du livre en bibliothèque et le web de données, en réfléchissant aux différentes réutilisations qui pourront être faites des données mises à disposition sur le web.

Ensuite, la finalité du projet de la Transition Bibliographique, du côté usagers, est de positionner les données bibliographiques sur le chemin de l'utilisateur. Grégory Miura, directeur du Service Commun de Documentation (SCD) de Bordeaux 3, explique qu'il s'agit de procéder à une « dissémination des données sur de nombreux canaux » (Goubina, 2018), afin de les rendre le plus accessible possible aux usagers. L'enjeu est de faciliter la recherche documentaire sur les catalogues en ligne des bibliothèques.

Enfin, pour le Ministère de la Culture, à travers le web de données, il subsiste un enjeu de « co-production de connaissances » (Sajus et Leroi, 2016), à la fois entre les professionnels des institutions culturelles, les professionnels du monde du livre au sens large, mais aussi les

<sup>7</sup> « Web sémantique et modèle de données ». BnF Data. En ligne : <https://data.bnf.fr/semanticweb>

usagers qui peuvent faire avancer cette transformation en proposant des services innovants et utiles afin d'optimiser leurs recherches.

## 1.2 Ce qui est attendu des bibliothèques françaises

### ➤ Une approche collective du projet

La particularité du programme de la Transition Bibliographique est qu'il s'agit d'un projet organisé de manière collective, dans lequel les idées et réflexions des professionnels sont les bienvenues. Ainsi, le programme est organisé en trois groupes de travail, composés de représentants du monde des bibliothèques en France et pilotés par un comité stratégique bibliographique (voir Illustration 4). Les trois groupes de travail sont les suivants : le groupe Normalisation « RDA en France » ; le groupe Formation ; le groupe Systèmes et Données.

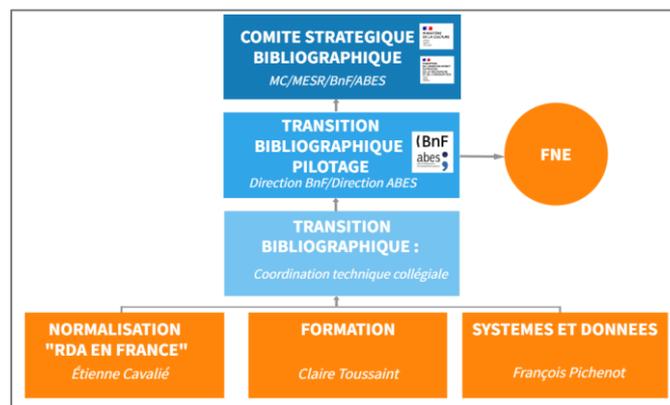


Illustration 4 : Schéma organisationnel de la Transition Bibliographique – décembre 2022<sup>8</sup>

Pour permettre aux professionnels des bibliothèques de faire leur transition bibliographique, les agences bibliographiques nationales mettent à leurs dispositions différents outils. Dans un premier temps, le site web du projet<sup>9</sup> est une mine d'informations à exploiter, qui regroupe l'ensemble des ressources disponibles pour comprendre le projet et le mettre en place au sein d'un établissement. Ensuite, les initiateurs et les membres du projet tentent d'être au plus proche des professionnels, en les encadrant sur le côté technique et organisationnel de la Transition bibliographique. Pour cela, une offre de formation complète leur est proposée, afin de bénéficier d'une expertise technologique. De plus, des journées professionnelles sont organisées

<sup>8</sup> Source : site web de la Transition bibliographique - <https://www.transition-bibliographique.fr/enjeux/position-francaise-rda/>

<sup>9</sup> <https://www.transition-bibliographique.fr/>

régulièrement pour annoncer les mises à jour et faire la synthèse sur l'avancement du projet. En effet, c'est un travail de réflexion qui est mené en lien avec les professionnels de tous les pans des métiers du livre et de la documentation, à la fois avec des bibliothécaires mais aussi les éditeurs de Système Intégré de Gestion de Bibliothèque (SIGB). Enfin, dans le but de mutualiser ce travail, un second projet, rattaché à la Transition Bibliographique, a été initié en 2017 : le Fichier National d'Entités (FNE). Le but de ce second projet est de permettre une co-production des données d'autorité, dans un premier temps par rapport aux travaux menés par la BnF et l'Abes, puis en mutualisant les données à l'échelle d'autres établissements. Les données d'autorité étant celles qui décrivent un document de « *manière univoque et pérenne* »<sup>10</sup> en se basant sur des métadonnées d'identification non ambiguës (concept et personne).

➤ *L'aspect technique de la Transition Bibliographique*

Outre l'aspect réflexif de la Transition Bibliographique, le cœur du programme est un travail de fond assez technique et complexe. En effet, pour réussir l'exposition de leurs données bibliographiques dans le contexte du web de données, les professionnels des bibliothèques doivent s'emparer des modèles conceptuels dédiés. Pour arriver à cet objectif, il faut réussir à travailler sur les métadonnées d'identification des documents, qui « *constituent le socle du web des données liées* » (Ministère de la Culture et de la Communication, 2014). Pour faire leur transition, les professionnels doivent s'appuyer sur le modèle conceptuel IFLA-LRM, publié en 2017 par l'IFLA et sur le nouveau code de catalogage RDA-FR, transposition du code de catalogage anglo-saxon RDA, paru en 2010. C'est d'ailleurs sur cette transposition que travaille le groupe « Normalisation » de la Transition Bibliographique, qui ne souhaite pas simplement traduire le code de catalogage, mais plutôt l'adapter au contexte français. Ainsi, le code de catalogage RDA-FR « *donne des indications sur l'identification et la description des ressources documentaires, en se fondant sur le modèle conceptuel IFLA-LRM* »<sup>11</sup>. Le but est de travailler sur la structuration des métadonnées d'identification des documents au sein des catalogues bibliographiques dès la source, grâce au code de catalogage RDA-FR. Il faut donc s'adapter au code pour améliorer la qualité des métadonnées, puis l'adopter pour en favoriser la réutilisation (Leresche et Lucerna, 2015).

---

<sup>10</sup> Données d'autorité à la BnF. En ligne : <https://www.bnf.fr/fr/donnees-autorite-bnf>

<sup>11</sup> Code RDA-FR. Site web de la Transition Bibliographique. En ligne : [Code RDA-FR - Transition bibliographique - Programme national \(transition-bibliographique.fr\)](http://www.transition-bibliographique.fr)

Les professionnels des bibliothèques doivent progressivement adapter leurs habitudes de catalogages à ces nouveaux modèles conceptuels mais également se doter d'un SIGB qui prend en charge la description conforme au modèle IFLA-LRM et qui permet de cataloguer les documents conformément à RDA-FR. Au cours de cette transition, des difficultés sont évidemment à prendre en compte et notamment la « *capacité à gérer les différences de temporalité entre d'une part les évolutions normatives et d'autre part les progrès techniques ou l'évolution des usages* » (Cantie, 2015).

Dans le but d'aider et d'orienter les bibliothèques françaises dans le prisme de la Transition Bibliographique, le groupe « Systèmes et Données » a développé le logiciel « Bibliostratus », pour leur permettre d'aligner leurs données par rapport à celles de la BnF et de l'Abes. En effet, depuis 2015-2016, les deux agences bibliographiques nationales ont travaillé sur des algorithmes de FRBRisation des données de leurs catalogues. Ainsi, les bibliothèques françaises peuvent utiliser Bibliostratus pour aligner leurs données à celles de la BnF et de l'Abes puis les importer dans leurs propres catalogues locaux. Dans le cadre de la Transition Bibliographique, des formations et des journées professionnelles sont organisées par le groupe « Systèmes et Données » pour présenter le logiciel Bibliostratus et collecter les retours d'expériences des professionnels, afin de faire évoluer l'outil de manière collective.

### *1.3 Un changement de paradigme : du document vers la donnée, quelle finalité pour les usagers ?*

#### *➤ L'impact sur les catalogues de bibliothèque*

Dans ce nouveau paysage informationnel, il faut désormais s'interroger sur les pratiques des usagers en bibliothèques, pour construire des catalogues adaptés à leurs besoins et à leurs attentes. En effet, l'utilisateur s'intéresse à des contenus, qu'il doit pouvoir retrouver facilement au sein du catalogue. C'est dans cette logique que la Transition Bibliographique va venir réorganiser l'information bibliographique afin d'aboutir à une interface optimale pour les usagers. Le catalogue en ligne d'une bibliothèque « *constitue un reflet de la collection physique, pour laquelle il joue le rôle d'outil d'accès* » (Bermès, 2013). Toutefois, dans le contexte du web de données, il faut trouver un moyen de valoriser les ressources des bibliothèques au sein de la masse d'informations présentes sur le web. En effet, cela s'explique par le fait que l'on passe désormais d'une logique de document à une logique de données, toujours dans l'optique de favoriser la recherche bibliographique par les usagers des bibliothèques. Pour construire un

catalogue répondant à ces exigences, les bibliothèques doivent donc réfléchir à une stratégie, en identifiant « *la place de l'information bibliographique dans les services que les usagers [...] sont en droit d'en attendre ou que l'on souhaite leur offrir* » (Leresche, 2016).

Dès lors, il s'agit de placer l'information bibliographique sur le chemin emprunté par l'utilisateur dans l'environnement du web. Pour cela, le catalogue de bibliothèque doit être pensé comme « *outil de découverte* » (Goubina, 2018) des collections, à la fois pour renseigner les usagers inscrits à la bibliothèque mais aussi pour toucher un public plus large, qui pourra découvrir les services de l'établissement à travers son catalogue en ligne. L'idée est également de mutualiser les données de la bibliothèque, pour que celle-ci apparaisse comme un service public culturel au cœur des questions du numérique et des données liées

➤ *La finalité du côté usagers*

Quant à la finalité de la Transition Bibliographique pour les usagers, à l'heure actuelle, il est encore difficile de savoir comment toutes ces évolutions seront visibles au niveau de l'interface-utilisateur des catalogues de bibliothèques. En effet, il n'existe pas encore de visualisation propre sur l'aboutissement de cette transition, en termes de conception et de mise en place dans les catalogues de bibliothèques. Toutefois, l'objectif vers lequel il faut se tourner est le fait de prendre davantage en compte les usagers et de construire des catalogues plus pertinents, capables de fournir des réponses plus précises à leurs requêtes.

Pour le moment, la société française *Tech'Advantage* développe un système de gestion documentaire nommé « Syrtis », basé sur le modèle IFLA-LRM qui permettrait de cataloguer en RDA-FR. L'objectif de ce système serait de proposer aux usagers une « balade intuitive » au sein des catalogues de bibliothèques, avec différentes visualisations des données bibliographiques, par exemple sous forme de graphes (Gallo, 2023). Cela permettrait à la fois de rendre plus aisée la recherche d'informations par les usagers mais aussi d'apporter un côté ludique aux catalogues des bibliothèques.

Enfin, l'idée est également de proposer des services innovants au sein des catalogues de bibliothèques, intégrant les technologies du web 3.0 et de l'intelligence artificielle. Afin de se rapprocher des usagers et de les guider dans leurs recherches, il pourrait par exemple être pertinent de construire un « chatbot » standardisé et unifié, c'est-à-dire un service de questions-

réponses automatique qui permettrait aux usagers de poser leurs questions et de recevoir une information précise. Par exemple, il est possible de citer le projet « Sparnatural Partitions » mené en collaboration entre la société française *Sparna* et la Cité de la Musique / Philharmonie de Paris qui propose une solution permettant, grâce à un chatbot, de faire des recherches intuitives sur les partitions musicales<sup>12</sup>. Ainsi, cela se révèle possible grâce aux évolutions apportées par la Transition Bibliographique, puisque les informations seraient désormais liées entre elles sur le web.

## 2. Le volet technique – comprendre la logique derrière le programme de la Transition bibliographique

### 2.1 L'ouverture des données, quels enjeux pour les bibliothèques ?

Le mouvement d'ouverture des données publiques n'est pas récent. Déjà à la fin des années 1970, en France, on se positionne dans la logique de l'ouverture des données (ou Open Data en anglais), suite à l'adoption de la loi du 17 juillet 1978 relative au droit d'accès des citoyens à tout document administratif<sup>13</sup>. Ce mouvement se poursuit au début des années 2010, avec la création par le gouvernement français de la mission Etalab qui « *coordonne la politique d'ouverture et de partage des données publiques* »<sup>14</sup> via la plateforme d'Open Data nationale [data.gouv.fr](http://data.gouv.fr). De plus, en 2016, avec la loi pour une République numérique<sup>15</sup>, émane un cadre juridique relatif à l'ouverture de l'information publique. Désormais, les collectivités et organes publics français doivent publier en accès libre leurs données publiques sur le web, dans l'objectif de pouvoir être réutilisées par tous. En plus d'un enjeu de transparence et de disponibilité des données ouvertes, le but est de rendre l'information plus facilement accessible pour les citoyens sur le web, en suivant la logique des données liées. Ainsi, dans tous les domaines de la société, les citoyens disposent donc d'un accès libre aux informations.

Dans le monde de la culture, l'enjeu est également fort pour les institutions culturelles puisqu'il s'agit de promouvoir leurs collections en facilitant l'accès et la diffusion de la culture.

---

<sup>12</sup> Sparnatural Partitions. En ligne : <https://sparnatural.eu/demos/demo-partitions/index.html>

<sup>13</sup> Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal. En ligne : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT00000339241>

<sup>14</sup> Etalab – Politique publique de la donnée. En ligne : <https://www.etalab.gouv.fr/qui-sommes-nous/>

<sup>15</sup> Loi n°2016-1321 du 7 octobre 2016 pour une République numérique. En ligne : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/>

En 2019, le Ministère de la Culture explique que l'ouverture des données publiques est une « *opportunité en matière de démocratisation culturelle et de transmission des savoirs en démultipliant les usages, favorisant ainsi la diffusion de la culture au plus grand nombre* » (Ministère de la Culture, 2019). En entrant dans la Transition bibliographique, les bibliothèques prennent donc le pas de l'ouverture de leurs données, le but de ce programme étant de rendre davantage visibles et accessibles leurs catalogues en ligne. À travers ce programme, et notamment avec le projet FNE (Fichier National d'Entités), l'objectif est de « *favoriser la production mutualisée de données de qualité, dans le respect des principes FAIR (Faciles à trouver, Accessibles, Interopérables, Réutilisables)* ». Pour parvenir à cet objectif et réussir le projet de la Transition bibliographique, il apparaît donc nécessaire pour les bibliothèques d'entrer dans une démarche d'ouverture de leurs données bibliographiques. Cela amène ainsi à des réflexions autour de la structuration de ces données et du format à adopter au moment de leur saisie dans les catalogues des bibliothèques.

## *2.2 Sortir de l'isolement des données en adoptant la logique du linked data*

Afin de rendre les données des institutions publiques visibles et accessibles, il faudrait sortir de l'organisation des données par « silos » et favoriser une structure standardisée. En effet, cela permettrait ensuite aux protocoles de requêtes de pouvoir apporter une réponse plus pertinente aux utilisateurs. Les données doivent être organisées et structurées de manière à être liées entre elles, grâce à des liens de navigation qui vont construire un chemin de découverte pour l'utilisateur. C'est à ces objectifs que devraient arriver les catalogues des bibliothèques dans le cadre du programme de la Transition bibliographique. Les liens entre les données bibliographiques permettront de renvoyer l'utilisateur vers des notices similaires et de découvrir de nouveaux documents, en lien avec leur recherche initiale. Ainsi, les apports du web de données pour les catalogues des bibliothèques sont nombreux (voir Illustration 5).



Illustration 5 : Apports du web de données en bibliothèque<sup>16</sup>

Pour structurer les informations sur le web de données, on se base désormais sur le modèle RDF (*Resource Description Framework*) qui permet de décrire formellement les ressources web et leurs métadonnées, modèle développé par le W3C à partir de la fin des années 1990. D'après ce modèle, les documents sont décrits sous la forme de « triplets ». Un triplet est la plus petite structure de description de ressources du modèle RDF qui suit la structure « sujet – prédicat – objet ». À noter que « *le sujet et l'objet peuvent eux même être identifiés par une URL décrivant un autre document* »<sup>17</sup>. Ensuite, pour exposer les données sur le web, les triplets sont reliés entre eux sous la forme de graphes, en fonction de leurs informations communes. Cette structuration organisée sur la base d'un modèle conceptuel standardisé va permettre pour la machine de relier les informations entre elles et de traduire ces liens de manière compréhensible pour les utilisateurs.

<sup>16</sup> « Sensibilisation à l'évolution des catalogues. 2. Web sémantique et web de données ». Support de formation du programme Transition Bibliographique. 2019. En ligne : [https://www.transition-bibliographique.fr/wp-content/uploads/2020/03/Evolution-des-catalogues\\_2\\_Le-web-semantique\\_dec-2019.pdf](https://www.transition-bibliographique.fr/wp-content/uploads/2020/03/Evolution-des-catalogues_2_Le-web-semantique_dec-2019.pdf)

<sup>17</sup> « Comment interroger la Bnf et Wikidata avec Sparql ? ». BibLibre. 2018. En ligne : [Comment interroger la Bnf et Wikidata avec Sparql ? - BibLibre - Services et logiciels libres pour votre bibliothèque - SIGB Koha, portail Bokeh, numérique, gestion](#)

### 2.3 La mise en place d'un nouveau code de catalogage adapté aux technologies du web de données

Pour les bibliothèques, la Transition bibliographique est donc le fait de « basculer d'un catalogue par document à un catalogue par entité »<sup>18</sup>. Et pour réussir cette transition, il a fallu revoir les méthodes de catalogage des professionnels afin de s'adapter aux technologies du web de données. Ainsi, est apparu en 2010 un nouveau code de catalogage, le code RDA (Ressource : Description et Accès) proposé par l'IFLA et basé sur le modèle conceptuel IFLA-LRM. Ce nouveau code de catalogage a pour ambition de permettre le liage des données, par des relations logiques. En France, dans le prisme du programme de la Transition bibliographique, l'objectif est de proposer une transposition française de ce code de catalogage, afin de proposer le code RDA-FR. C'est pourquoi le groupe « Normalisation RDA en France » est chargé de publier progressivement les nouvelles règles de catalogage. Le but est qu'une fois le code RDA-FR prêt à être utilisé, il remplace, à terme, les normes de catalogage établies par l'Afnor. La transposition française du code est un travail considérable, c'est pourquoi le groupe Normalisation est divisé en plusieurs sous-groupes de travail, répartis en fonction des différentes entités du modèle IFLA-LRM (voir Illustration 6).

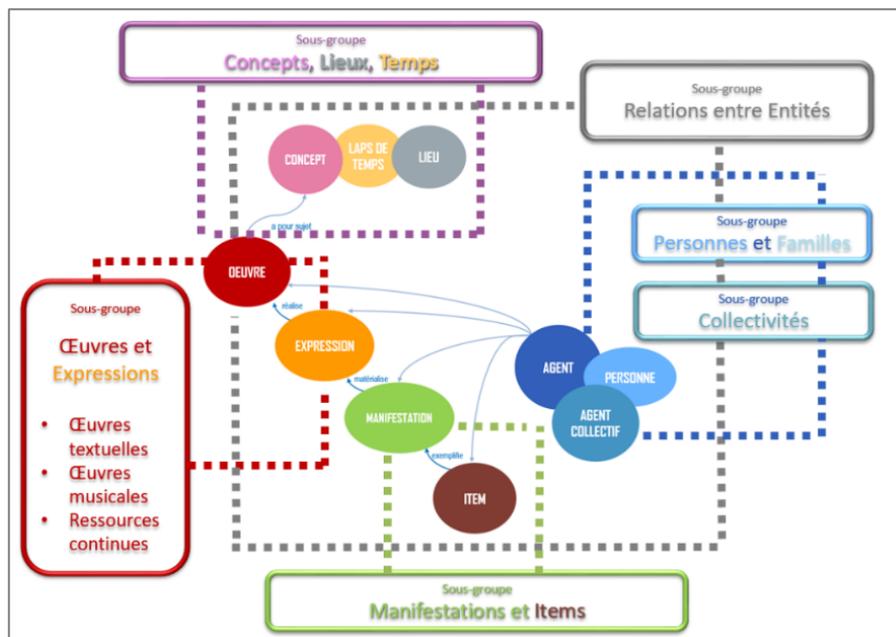


Illustration 6 : Organisation des sous-groupes « Normalisation » de la Transition bibliographique<sup>19</sup>

<sup>18</sup> La Transition Bibliographique en France. Site web de la Transition Bibliographique. En ligne : <https://www.transition-bibliographique.fr/enjeux/position-francaise-rda/>

<sup>19</sup> Code RDA-FR. Site web de la Transition Bibliographique. En ligne : <https://www.transition-bibliographique.fr/rda-fr/>

Concrètement, le code de catalogage RDA-FR a pour finalité d'améliorer la recherche des usagers puisqu'il «  *vise à ce que la description bibliographique et sa structuration répondent mieux aux différentes tâches des utilisateurs en ligne* »<sup>20</sup> (voir Illustration 2, p.19). À propos des requêtes des usagers, les technologies du web de données prises en compte dans le programme de la Transition bibliographique permettront de se baser sur le protocole SPARQL, qui est un langage de requête se fondant sur les graphes issus du modèle RDF, pour répondre aux besoins des utilisateurs. Grâce à la structuration des données au sein du catalogue en ligne des bibliothèques, les réponses aux requêtes des usagers seront plus pertinentes et plus complètes.

Ainsi, au cours de ce premier chapitre, nous avons pu comprendre le contexte et les enjeux de la Transition Bibliographique et, de surcroît, de ma mission au sein de la Médiathèque de Roubaix. Dans le prochain chapitre, il s'agira désormais de se concentrer sur la mise en application des concepts étudiés et sur la réalisation concrète de ma mission de stage.

---

<sup>20</sup> *Ibid.*

## Chapitre 2 : Mise en application et travail concret d’alignement des notices

### Partie 1 : Méthodologie d’alignement des notices bibliographiques via le logiciel Bibliostratus

#### 1. Présentation rapide du corpus

Le corpus de données sur lequel vont être réalisés les tests d’alignement est un corpus composé d’environ 3000 notices issues d’un fonds patrimonial de la Médiathèque de Roubaix, le fonds Destombes, du nom de son donateur. Le fichier de départ contenant ces notices est un fichier aux formats iso2709 et XML, exportées du catalogue de la médiathèque depuis le SIGB Koha.

Les notices de ce corpus se rapportent à différents types de documents, à la fois des monographies (3082 notices), des partitions (5 notices) mais aussi une notice de périodique. Aussi, ce qui fait la particularité de ce fonds est son aspect patrimonial et donc l’ancienneté des documents. Les métadonnées de ces documents peuvent donc apparaître spécifiques et il s’agira de les traiter avec prudence.

Pour réaliser l’alignement de ce corpus de notices à celles de la BnF, il faudra donc réussir à retrouver les identifiants pérennes associés, grâce aux différents modules de l’outil Bibliostratus. Les identifiants pérennes ARK (*Archival Resource Key*) attribués par la BnF ou PPN (*Pica Production Number*) attribués par l’Abes sur le Sudoc.

#### 2. Utilisation de Bibliostratus – les étapes à suivre pour réaliser l’alignement

L’outil Bibliostratus propose trois modules principaux pour parvenir à l’alignement des métadonnées des bibliothèques locales à celles des agences bibliographiques nationales. Deux modules (bleu et rouge) sont des outils d’accompagnement pour aider les professionnels à réaliser l’alignement de leurs données et le module blanc est le module central qui permet d’alignement des données bibliographiques ou d’autorités (voir Illustration 7).

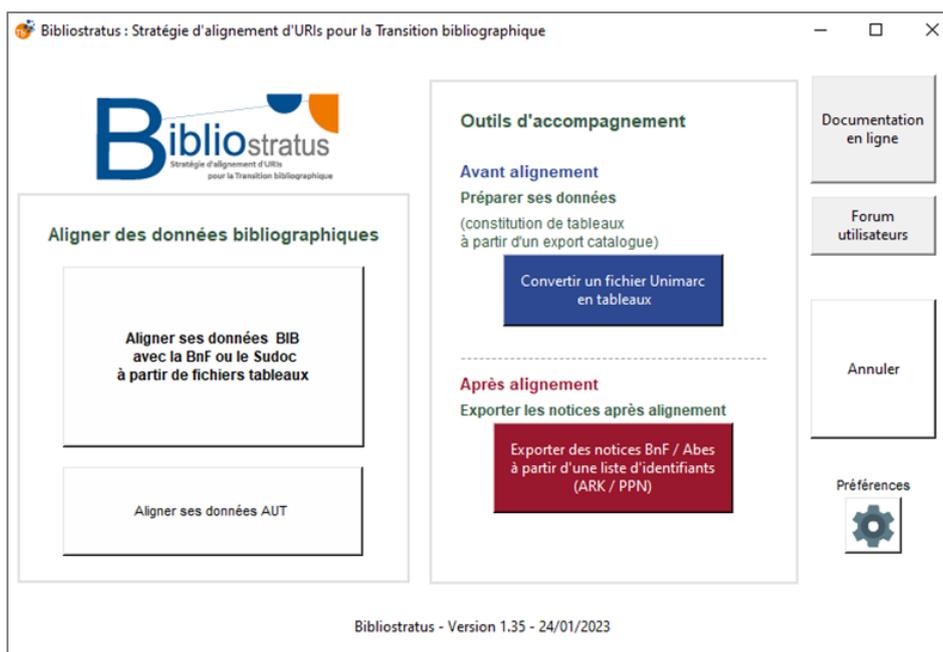


Illustration 7 : Interface d'accueil de Bibliostratus et modules proposés

### *2.1 Présentation des trois modules selon leur ordre d'utilisation*

Le **module bleu** est le premier module à utiliser sur Bibliostratus (voir Illustration 8). En effet, il permet de préparer les données avant de passer à l'alignement, en passant d'un fichier MARC en un tableau qui contient les informations nécessaires à l'alignement, qui renvoient aux métadonnées d'identification des documents. Pour cela, il suffit de fournir en entrée le fichier de notices Unimarc exporté depuis le SIGB de la bibliothèque, en sélectionnant le bon format de fichier : iso2709 encodé en UTF-8, iso2709 encodé en ISO-8859-1 ou MARC XML encodé en UTF-8. A noter que des liens vers la documentation technique sont à disposition des professionnels qui auraient besoin d'aide pour comprendre le fonctionnement de chaque module. Il faut également choisir en entrée le type de notices présentes dans le fichier, à savoir des notices bibliographiques ou d'autorités. Ensuite, il faut choisir le dossier de destination des fichiers en sortie. Après le passage du fichier dans le module bleu, Bibliostratus ajoute dans le dossier de destination choisi plusieurs fichiers, en fonction du type de documents (monographie, audiovisuel, périodique) selon les codes associés.

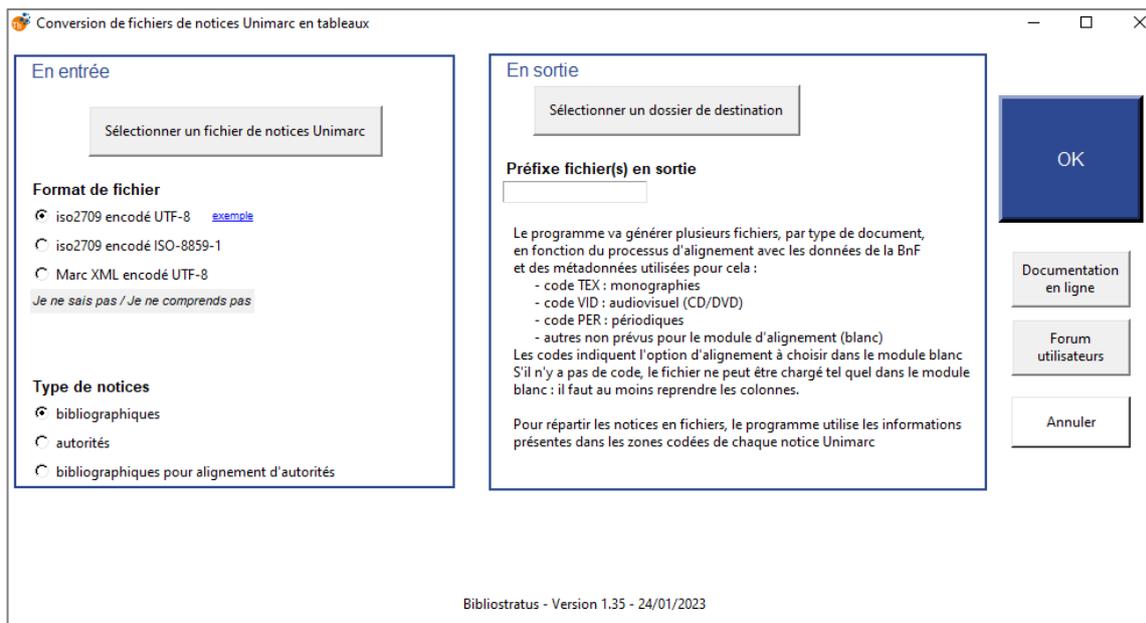
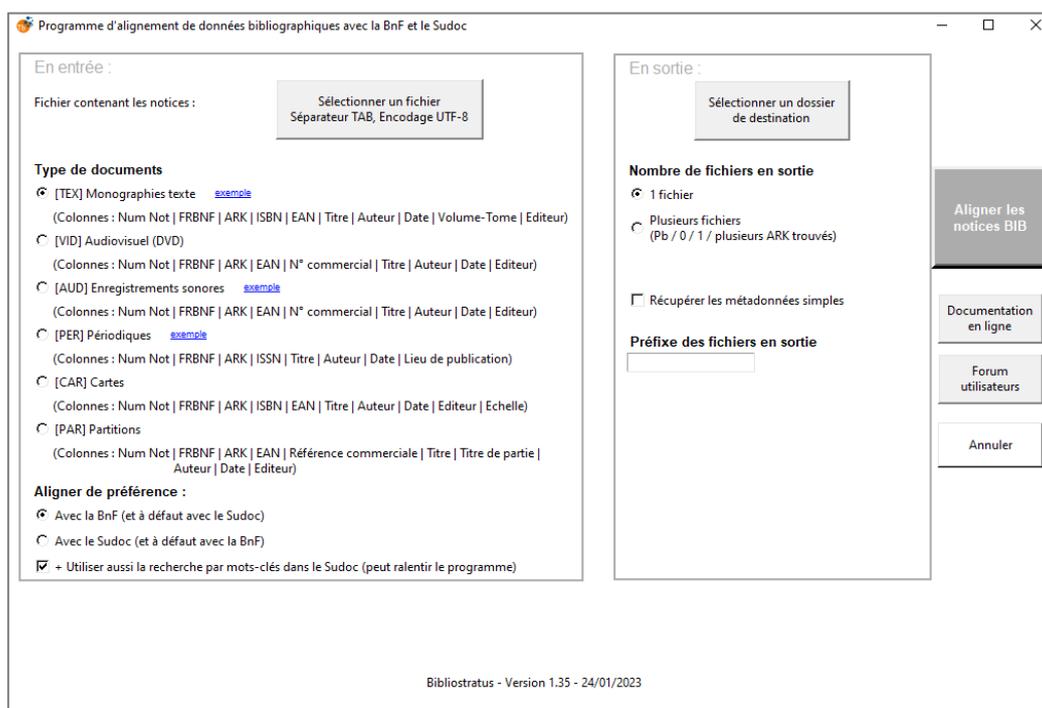


Illustration 8 : Visualisation du « module bleu » de Bibliostratus

Ensuite, après avoir obtenu les fichiers nettoyés, sous forme de tableaux, pour chaque type de documents, il est possible de passer au deuxième module, le **module blanc**, qui va réaliser l’alignement des données par rapport à celles de la BnF et/ou du Sudoc (voir Illustration 9). Pour cela, en entrée, il faut sélectionner les fichiers obtenus après le module bleu, en veillant à ce que le séparateur soit la tabulation et qu’il soit encodé en UTF-7, ce qui n’est pas forcément le cas par défaut pour les fichiers nettoyés du module bleu. Il faut également sélectionner le type de document qui renvoie aux notices à aligner, en fonction des métadonnées d’identification indiquées dans les colonnes du tableau. Enfin, en entrée, il est également possible de sélectionner l’agence bibliographique avec laquelle on souhaite aligner les notices, sachant que pour celle choisie, le choix se fera sur l’autre agence à défaut d’identifiant pérenne trouvé. Une fois le fichier sélectionné en entrée, Bibliostratus va interroger automatiquement le catalogue de la BnF ou du Sudoc, à partir de la liste de notices fournie. Si la notice de la bibliothèque locale est en conformité avec celle de l’agence bibliographique, alors le logiciel va enregistrer l’identifiant pérenne trouvé (ARK ou PPN). Ce module d’alignement va générer deux fichiers, pour chaque type de document, dans le dossier de destination choisie en sortie :

- un fichier {prefixe}-rapport\_stats ;
- un fichier {prefixe}-resultats.

Ainsi, dans le second fichier, on retrouve un tableau contenant le(s) identifiant(s) pérenne(s) trouvé(s) ainsi que la méthode d'alignement, c'est-à-dire sur quelles métadonnées d'identification du document le logiciel s'est basé pour retrouver l'identifiant, par exemple « Titre-Auteur-Date + contrôle Titre 200\$a ». Pour les notices dont la bibliothèque locale possède déjà un identifiant ARK ou PPN, le logiciel va simplement le vérifier et le mettre à jour.



**Illustration 9 :** Visualisation du « module blanc » de Bibliostratus pour l'alignement des données bibliographiques

En ce qui concerne l'alignement des notices d'autorité, la procédure est sensiblement la même (voir Illustration 10). Il faut sélectionner le type de données en entrée (liste de notices Personnes, liste de notices Organisation, liste de notices bibliographiques ou liste de notices Rameau) et l'agence bibliographique avec laquelle on souhaite aligner les notices, ici, la BnF ou IdRef (Identifiants et Référentiels pour l'Enseignement Supérieur et la Recherche).

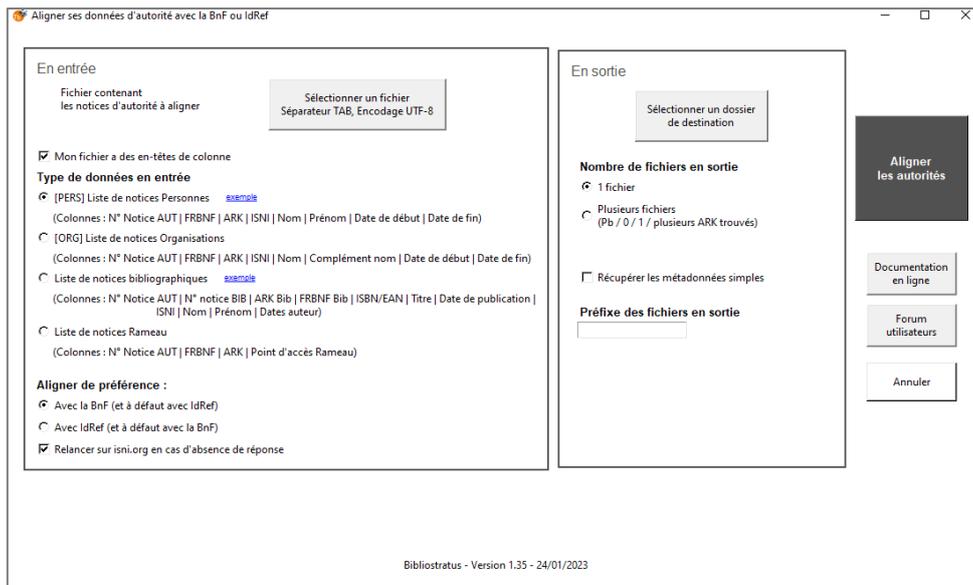


Illustration 10 : Visualisation du « module blanc » de Bibliostratus pour l’alignement des données d’autorité

Enfin, la dernière étape du processus d’alignement est le **module rouge**, qui permet de récupérer les notices alignées dans un format qui va permettre de les réinjecter dans le catalogue de la bibliothèque locale (voir Illustration 11). Pour cela, il faut sélectionner en entrée le fichier généré par le module blanc, pour chaque type de document, contenant les identifiants ARK et PPN des notices. Dans le dossier de destination choisi par le bibliothécaire, le logiciel Bibliostratus va ainsi pouvoir générer un fichier de notices alignées (en Unimarc ou InterMarc), encodé en iso2709 ou UTF-8, en fonction du SIGB dans lequel vont être réinjectées les notices alignées. Ainsi, une fois ces fichiers de notices alignées et analysées obtenus, les bibliothèques locales se tournent véritablement vers l’horizon de la Transition bibliographique.

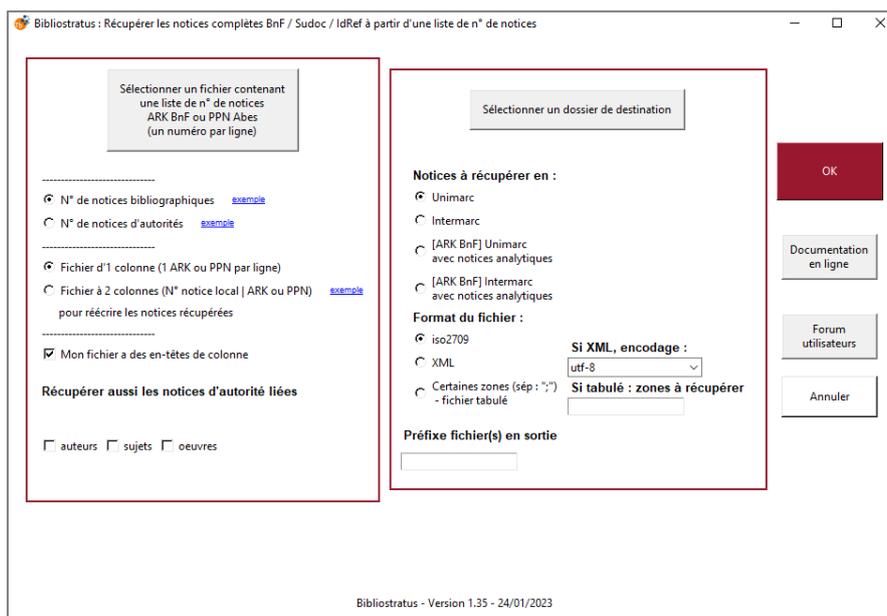


Illustration 11 : Visualisation du « module rouge » de Bibliostratus

## 2.2 Focus sur le processus d'alignement – module blanc

Le module blanc de Bibliostratus est le module central du programme car c'est celui qui va exécuter les recherches d'identifiants sur le catalogue de la BnF et/ou du Sudoc. Il s'agira de se concentrer ici sur les recherches menées dans le catalogue de la BnF, dans lequel se trouvent les identifiants pérennes ARK. Pour effectuer ces recherches, Bibliostratus se base sur les mécanismes de recherche des identifiants via le [formulaire de requête SRU](#) (*Search / Retrieval via URL*)<sup>21</sup> de la BnF. Il s'agit d'un service permettant de faire des requêtes sur le catalogue de la BnF, en remplissant des champs correspondant aux métadonnées d'identification des documents (auteur, titre, date, ISBN, etc.). Il est possible d'affiner la recherche en sélectionnant une relation (*all, any, adj*) ainsi qu'un opérateur booléen (*and, or, not*).

Ainsi, le mécanisme de recherche de Bibliostratus est simple : il va effectuer une recherche avancée dans le catalogue de la BnF, en utilisant les métadonnées indiquées dans le fichier fourni en entrée. Bibliostratus va réaliser une recherche par mot-clé, effectuée sur le titre, l'auteur et la date du document. Pour cela, le titre est nettoyé, c'est-à-dire que les accents et la ponctuation sont supprimés, ainsi que les mots de moins d'une lettre. Pour la date, Bibliostratus

<sup>21</sup> Formulaire de requête SRU de la BnF : <https://catalogue.bnf.fr/api/test.do>

se base sur la zone Unimarc 100\$a et prend en compte les années n-1 et n+1. Aussi, il est possible d'ajouter un contrôle sur l'éditeur dans les préférences du logiciel.

### 3. Alignement du corpus test

Après avoir pris connaissance du fonctionnement de Bibliostratus, le premier test d'alignement du corpus de notices issues du fonds Destombes de la Médiathèque de Roubaix a été réalisé. Pour cela, il a fallu simplement suivre les instructions des différents modules du logiciel.

#### 3.1 La conversion des fichiers de notices Unimarc en tableaux

Pour préparer le corpus de notices à l'alignement, il faut commencer par convertir le fichier de notices exportées du SIGB Koha en Unimarc vers des tableaux, en fonction du type de document. Le corpus de notices étant composé à la fois de monographies, de partitions et de périodiques, le module bleu de Bibliostratus ressort trois fichiers distincts en fonction de ces trois types de documents (voir Illustration 12).

Nom	Modifié le	Type	Taille
 TEX-texte-monographie	11/04/2023 14:05	Document texte	388 Ko
 PER-texte-périodiques	11/04/2023 14:05	Document texte	1 Ko
 PAR-partition-monographie	11/04/2023 14:05	Document texte	1 Ko

Illustration 12 : Capture d'écran des fichiers obtenus après le passage dans le module bleu de Bibliostratus

Ces trois fichiers de sortie du premier module sont des fichiers de données tabulaires au format .csv, qui contiennent en colonnes les métadonnées d'identification des documents (voir Illustration 13).

Num Not	FRBNF	ARK	ISBN	EAN	Titre	Auteur	Date	Volume-Tome
---------	-------	-----	------	-----	-------	--------	------	-------------

Illustration 13 : Métadonnées d'identification associées au premier module de Bibliostratus

Ainsi, les notices des trois types de documents sont désormais prêtes à être passées dans le deuxième module d'alignement.

### 3.2 L'alignement des notices bibliographiques à celles de la BnF / du Sudoc

Dans le cadre de ma mission, il s'agit de commencer par tester l'alignement des données bibliographiques du corpus test à celles des agences bibliographiques nationales. Une fois dans le module blanc, il faut donc sélectionner les paramètres d'entrée et de sortie des fichiers qui vont être analysés par le logiciel Bibliostratus. Après avoir sélectionné le type de document du fichier et choisi d'aligner les notices de préférences avec la BnF (et à défaut avec le Sudoc), je sélectionne le fichier tabulé obtenu grâce au premier module. Le logiciel Bibliostratus s'exécute, pendant quelques minutes (en fonction du nombre de notices à aligner). À la fin de l'exécution, le logiciel fournit dans le dossier de destination choisi les résultats de l'alignement, sous forme de fichiers tabulés également, pour chacun des types de documents. Ces tableaux possèdent en colonnes les métadonnées d'identification des documents, comme pour le résultat du module bleu, avec en plus l'identifiant ARK ou PPN trouvé et la méthode d'alignement, comme expliqué ci-dessus. Ainsi, pour les monographies, les métadonnées d'identification sont les suivantes (voir Illustration 14).

NumNot	Nb identifiants trouvés	ISBN	EAN	Titre	Liste identifiants trouvés	Auteur	Date	Méthode d'alignement	FRBNF	ARK	ISBN	EAN	Editeur	Type doc / Type notice
--------	-------------------------	------	-----	-------	----------------------------	--------	------	----------------------	-------	-----	------	-----	---------	------------------------

Illustration 14 : Métadonnées d'identification associées au module d'alignement de Bibliostratus

Pour plus de clarté et de lisibilité, il apparaît pertinent de copier les résultats obtenus dans un tableau type Excel, afin de pouvoir ensuite les analyser (voir Illustration 15).

	A	B	C	D	E	F	G	H	I	J	K
	NumNot	Nb identifiants trouvés	Liste identifiants trouvés	Méthode d'alignement	FRBNF	ARK	ISBN	EAN	Titre	Auteur	Date
1	2201	1	PPN113428057	Titre-Auteur-Date + contrôle Titre 2005a					La peinture au musée de Lille	Benoit François	1909
2	2202	1	PPN113427085	Titre-Auteur-Date + contrôle Titre 2005a					La peinture au musée de Lille	Benoit François	1909
3	2203	1	PPN113700822	Titre-Auteur-Date + contrôle Titre 2005a					La peinture au musée de Lille	Benoit François	1909
4				Actualisation ARK, Titre-Auteur-Date					Atlas pour servir à l'intelligence des campagnes de la Révolution française	Thiers Adolphe;Duv	1846
5	2390	1	ark:/12148/cb314568675	+ contrôle Titre 2005a		http://catalogue.bnf.fr/ark:/12148/cb3			Historique du 19e régiment de chasseurs 1792-1892	Lepage Paul Jules;Pi	1893
6	2409	0							Nouvelle géographie universelle la terre et les hommes	Reclus Elisée	1879
7	3149	1	PPN011425571	Titre-Auteur-Date + contrôle Titre 2005a							
8	5127	1	ark:/12148/cb30385926	Actualisation ARK, Titre-Auteur-Date + contrôle Titre 2005a		http://catalogue.bnf.fr/ark:/12148/cb3			Théâtre des marionnettes	Duranty Louis-Émile	1880
9	6436	2	ark:/12148/cb335598451	Titre-Auteur-Date + contrôle Titre 2005a, Titre-Auteur-Date + contrôle : Titre BnF/Sudoc contenu dans titre initial					Histoire des quatre fils Aymon très nobles et très vaillans chevaliers	Grasset Eugène;Gillk	1883
10	8938	0							L'histoire de France depuis les temps les plus reculés jusqu'en 1789 racontée à mes petits-enfants	Guizot François;Neu	1872
11	8941	0							L'histoire de France depuis les temps les plus reculés jusqu'en 1789 racontée à mes petits-enfants	Guizot François;Neu	1875
12	8942	0							L'histoire de France depuis les temps les plus reculés jusqu'en 1789 racontée à mes petits-enfants	Guizot François;Neu	1874

Illustration 15 : Extrait du tableau Excel contenant les résultats obtenus après l'alignement des notices des monographies

## Partie 2 : Étude des résultats obtenus après le premier alignement et manipulation des fichiers pour obtenir le fichier de notices alignées

Après ce premier alignement, réalisé en respectant les différentes étapes indiquées par le logiciel Bibliostratus, il est tout d'abord possible d'observer que pour certaines notices, aucun identifiant n'est trouvé. Il s'agira donc désormais d'analyser ces résultats obtenus après l'alignement, pour comprendre le fonctionnement et les possibles failles du logiciel, et de l'algorithme sous-jacent. Le corpus étant majoritairement composé de monographies, les études et réflexions menées concerneront plus particulièrement ce type de document. Une réflexion sur les corpus de partitions et de périodiques, qui posent d'autres problématiques, pourra toutefois être menée en parallèle.

### 1. Analyse des résultats obtenus suite au premier alignement et constat des problèmes rencontrés

#### 1.1 Exposé des résultats obtenus

##### ➤ Résultats obtenus pour les notices du corpus de monographies

Suite au premier passage du corpus de monographies dans Bibliostratus, voici les résultats obtenus pour l'alignement (voir Tableau 1).

Nombre d'identifiants trouvés	Nombre de notices concernées
0	2569
1	346
2	99
3	38
4	13
5	5
6	3
9	1
10	4
12	1
17	1
24	1
45	1

Tableau 1 : Résultats du premier alignement (monographies)

Pour rappel, le corpus de monographies est composé de 3082 notices. Suite à ce premier alignement, 2569 notices n'ont reçu aucun identifiant et donc, 513 notices ont obtenu au moins un identifiant, ARK ou PPN. Par conséquent, on peut admettre que, suite à ce premier passage dans Bibliostratus, seulement **16% des notices** ont été alignées. Ce résultat étant loin d'être satisfaisant, il faut désormais parcourir et analyser le fichier de résultats pour comprendre pourquoi Bibliostratus bloque sur certaines notices et ne trouve pas d'identifiant.

➤ Résultats obtenus pour les notices du corpus de partitions

Par rapport au corpus de partitions, composé de cinq notices, le résultat est très pauvre, puisqu'aucun identifiant n'a été trouvé, sur les cinq notices composant le fichier (voir Tableau 2).

Nombre d'identifiants trouvés	Nombre de notices concernées
0	5

Tableau 2 : Résultats du premier alignement (partitions)

Étant donné que le nombre de notice est succinct, j'ai tenté de rechercher manuellement les documents sur le catalogue de la BnF. Pour quatre notices sur cinq, je trouve l'identifiant ARK très facilement, grâce à une simple recherche sur le catalogue, et les métadonnées d'identification des documents sont similaires entre la BnF et la Médiathèque de Roubaix.

En observant plus attentivement les notices correspondant à ces cinq documents, il semblerait que quatre d'entre eux soient des monographies imprimées. Il s'agira donc d'une hypothèse à creuser au moment de la manipulation des fichiers pour améliorer les résultats de l'alignement. Pour la cinquième notice, le document n'est pas présent dans le catalogue de la BnF, raison pour laquelle Bibliostratus ne trouve pas d'identifiant ARK.

➤ Résultats obtenus pour les notices du corpus de périodiques

Concernant le périodique issu du fonds Destombes, après le passage du fichier dans Bibliostratus, le résultat est nul, l'identifiant ARK du périodique n'a pas été trouvé (voir Tableau 3). Pourtant, de même que pour les partitions, en effectuant une recherche manuelle sur le catalogue de la BnF, je retrouve facilement l'identifiant ARK.

Nombre d'identifiants trouvés	Nombre de notices concernées
0	1

Tableau 3 : Résultats du premier alignement (périodique)

De même que pour les partitions, le document présent dans le fichier 'périodiques' semble être une monographie, car il est indexé comme un livre imprimé dans le catalogue de la BnF. Afin de confirmer cette hypothèse, je suis allée voir le document physiquement, dans les magasins de la médiathèque et il s'agit bien d'une monographie.

### *1.2 Analyse des résultats obtenus et constat des problèmes rencontrés*

Pour parer au manque de performance du programme lors du premier alignement, il a été décidé de réaliser plusieurs sessions de recherche manuelle d'identifiants ARK, pour obtenir un fichier contenant les identifiants ARK non-trouvés lors du premier alignement par Bibliostratus. Cela pourra ensuite permettre de se référer à ce fichier de notices possédant des identifiants ARK, afin de mener des recherches et hypothèses et trouver des solutions aux problèmes

rencontrés à la suite du premier alignement. Ces sessions de recherches nous ont permis de relever les différents problèmes qui peuvent être des raisons pour lesquelles les notices n'ont pas été alignées.

Ci-dessous, il s'agira de dresser une liste des problématiques relevées, illustrées par des exemples et énonçant des hypothèses qui seront testées dans la partie suivante.

- **Problème n°1 :** identifiant ARK non trouvé lorsque Pierre Destombes (le donateur du fonds) apparaît dans la colonne des auteurs.

Pour un certain nombre de notices, le donateur du fonds Pierre Destombes apparaît dans la colonne des auteurs. En effet, lors du catalogage dans le SIGB de la médiathèque de Roubaix, le nom « Pierre Destombes » est indiqué dans le champ des auteurs secondaires, en zone Unimarc 702\$a (voir Illustration 16). Or, dans le catalogue de la BnF, Pierre Destombes n'apparaît dans aucune notice puisqu'il n'a pas de lien avec la BnF (voir Illustration 17). Ainsi, le logiciel Bibliostratus réalisant une recherche par exactitude dans la colonne des auteurs, il ne peut pas ressortir l'identifiant ARK des notices correspondantes.

Exemple : *Physiologie du chasseur*, écrit par T. Deyeux et publié aux éditions Aubert en 1841.

```
Prévisualisation MARC
000 01072nam a2200289 4500
001 91443
003 http://www.mediathequederoubaix.fr/ark:/20179/RH91443
005 20180720094828.0
035 0 0 _aH2_RBX152818
090 _a91443
091 _a0
099 _a1
_c2014-06-30
_d2014-06-30
_K0
100 _a19 1841 k y0frey5003 ba
101 _aTce
200 _aPhysiologie du chasseur
_fpar Deyeux
_gvignettes d'Eugène Forest
210 _aParis
_cAubert
_cAvisign
_d[1841]
215 _a122 p.
_cill.
_d15 cm
300 _a1841 d'après le "manuel de l'amateur du livre du 19ème siècle" de Vicaire
300 _aInventaire ENSRIT n.° 8087
300 _aEx libris impr. Pierre Destombes
310 _aDemi-rel. maroquin fauve, dos à cinq nerfs, tran che sup. dorée
339 _am
700 0 _9284279
_aDeyeux
_bTheophile
702 0 _9284280
_aForest
_bEugène
_f1808-1891
_4440
702 _9166363
_4320
_aDestombes
_bPierre
_cOubalsien
_f1846-1915
801 0 _aFR
_b99126101
_c00121103
_gAFNOR
930 _99126101:0000288234
_b99126101
_a0/50
_eFonds patrimonial: Legs Destombes
```

Illustration 16 : Notice bibliographique (Unimarc) sur le SIGB de la Médiathèque de Roubaix

Notice bibliographique

Notice Au format Unimarc ▼

```

000 cam 22 3 450
001 FRBNF303392230000009
003 http://catalogue.bnf.fr/ark:/12148/cb30339223w
039 ..So CRI Sa CG004001990111P
039 ..So OPP Sa 36380806 Sd 20070129
039 ..So OPP Sa 40356327 Sd 20070129
039 ..So OPP Sa 32027402 Sd 20070129
100 ..Sa 19970701d1841 m y0frey50 ba
101 0 .Sa fre
102 ..Sa FR
105 ..Sa |||z 00||
106 ..Sa r
181 0 $6 01 Sa i Sb xxxe
181 ..$6 02 Sc txt S2 rdacontent
182 0 $6 01 Sa n
182 ..$6 02 Sc n S2 rdamedia
200 1 .Sa Physiologie du chasseur Sb Texte imprimé Sf par Deyeux Sg vignettes d'Eugène Forest
210 ..Sa Paris Sc Aubert Sc Lavigne Sd [1841]
215 ..Sa 1 vol. (122 p.) Sc ill., couv. ill Sd 14 cm
316 ..$5 FR-751021001:30339223003001
321 1 .Sa Les physiologies / par Andr e Lh ritier, Jean Prinnet, Claude Pichols, [et al.], Paris, 1957, n  60 Sc Lh ritier, 60
606 ..S3 11931162 Sa Chasseurs S2 rameau
608 ..S3 11931282 Sa Ouvrages humoristiques S2 rameau
686 ..Sa 390 S2 Cadre de classement de la Bibliographie nationale fran aise
700 |S3 10432221 So ISNI0000000049426839 Sa Deyeux Sb Th ophile Sf 1789-1849 S4 070
702 |S3 12551701 So ISNI0000000118039884 Sa Forest Sb Eug ne Sf 1808-1891 S4 440
801 0 Sa FR Sb FR-751131015 Sc 19970701 Sg AFNOR Sh FRBNF303392230000009 S2 Intermerc
856 4 Su http://gallica.bnf.fr/ark:/12148/brv1b85303072
930 ..$5 FR-751131007:30339223001001 Sa 8-LI6-37 Sb 759999999 Sc Tolbiac - Rez de Jardin - Philosophie, histoire, sciences de l'homme - Magasin Sd 0
930 ..$5 FR-751131007:30339223002001 Sa 8-Z LE SENNE-4937 (27) Sb 759999999 Sc Tolbiac - Rez de Jardin - Philosophie, histoire, sciences de l'homme - Magasin Sd 0
930 ..$5 FR-751041002:30339223002002 Sa 8-NF-40277 Sb 759999999 Sc Arsenal - Magasin Sd N
930 ..$5 FR-751021001:30339223003001 Sa OA-295 (20)-BOITE FOL Sb 759999999 Sc Richelieu - Estampes et photographie - Magasin Sd N
930 ..$5 FR-759999999:30339223004001 Sa IFN-8530307 Sb 759999999 Sc Document num ris  Sd N

```

**Illustration 17 :** Notice bibliographique (Unimarc) sur le catalogue de la BnF

Une hypoth se appar it pour am liorer le r sultat de l'alignement : supprimer Pierre Destombes de la colonne des auteurs dans le fichier d'entr e, puis repasser le fichier dans Bibliostratus.

➤ **Probl me n 2 :** saisie des tomes

Pour certains documents comportant plusieurs volumes, au moment de la construction des notices bibliographiques, la m diath que de Roubaix a produit une notice par tomes, et ceci   la diff rence de la BnF qui regroupe l'ensemble des tomes en une seule et m me notice. Ainsi, lors du premier alignement, cela semble poser probl me   Bibliostratus, qui ne trouve pas forc ment l'identifiant ARK correspondant aux diff rents tomes.

**Exemple :** *Le livre des proverbes fran ais pr c d  de recherches historiques sur les proverbes fran ais et leur emploi dans la litt rature du moyen  ge et de la renaissance*,  crit par A. Le Roux de Lincy, publi  aux  ditions Adolphe Delahays en 1859.

➔ Sur le catalogue de la m diath que de Roubaix, il existe une notice pour le tome premier et une notice pour le tome second (voir Illustration 18).

**Le livre des proverbes français précédé de recherches historiques sur les proverbes français et leur emploi dans la littérature du moyen âge et de la renaissance / par M. Le Roux de Lincy . T. premier**

- Auteur principal : Le Roux de Lincy, Antoine (1806-1869)
- Auteur secondaire : Raparlier, Paul-Romain ( Relieur ) ; Destombes, Pierre, roubaisien (1846-1915) ( Donateur )
- Langue : français.
- Publication : Paris : A. Delahays, 1859
- Description : CXV-409 p. ; 16 cm
- Collection : Bibliothèque gauloise
- Sujet - Nom commun :
- Informations vendangeur : périmètre : aucune action

Illustration 18 : Capture d'écran de la notice de l'ouvrage de A. Le Roux de Lincy, sur le catalogue de la médiathèque de Roubaix

→ Sur le catalogue de la BnF, il existe une notice pour les deux tomes (voir Illustration 19).

Notice bibliographique

Notice Au format public

Type(s) de contenu et mode(s) de consultation : Texte noté : sans médiation

Auteur(s) : [Le Roux de Lincy, Antoine \(1806-1869\)](#)

Titre(s) : Le livre des proverbes français, précédé de recherches historiques sur les proverbes français et leur emploi dans la littérature du moyen âge et de la Renaissance, [Texte imprimé] / par M. Le Roux de Lincy

Édition : 2de éd. revue, corrigée et augmentée

Publication : Paris (4-6, rue Voltaire) : Adolphe Delahays, 1859

Impression : 75-Paris (8, rue Garancière) : Typ. de Henri Plon

Description matérielle : 2 vol. (CXV-409, 615-4 p.) ; 16 cm

Identifiant de la notice : ark:/12148/cb30798599n

Notice n° : FRBNF30798599

Illustration 19 : Capture d'écran de la notice de l'ouvrage de A. Le Roux de Lincy, sur le catalogue de la BnF

➤ **Problème n°3** : notice bibliographique mentionnant plusieurs auteurs

Lorsque la notice bibliographique comprend plusieurs auteurs (auteurs principaux, secondaires, préfaciers, traducteurs, etc.) cela semble poser problème dans la recherche exécutée par Bibliostratus. En comparant les notices des deux catalogues (médiathèque de Roubaix et BnF), on remarque que les auteurs présents dans les notices de la BnF ne sont pas tous présents dans les notices de la médiathèque de Roubaix, et inversement.

Exemple : *Physiologie du calembourg*, écrit par P. Bertrand (auteur principal), publié en 1841 aux éditions R. Bocquet.

→ Dans la notice de la médiathèque de Roubaix a été saisie la liste complète des auteurs et des co-auteurs (voir Illustration 20).

### Physiologie du calembourg / par un nain connu ; dessins de Henry Emy

- Auteur principal : Bertrand, Pierre
- Co-auteur : Carmouche, Pierre Frédéric Adolphe (Auteur) ; Duveyrier, Anne-Honoré-Joseph (Auteur) ; Lienhart, Frantz de (Auteur)
- Auteur secondaire : Destombes, Pierre, roubaisien (1846-1915) (Donateur)
- Langue : français.
- Publication : Paris : R. Bocquet, 1841
- Description : 116 p. : front., ill. ; 15 cm
- Informations vendeur : périmètre : aucune action

Illustration 20 : Capture d'écran de la notice de l'ouvrage *Physiologie du calembourg*, sur le catalogue de la médiathèque de Roubaix

→ Dans la notice de la BnF, seul l'illustrateur est indiqué dans le champ des auteurs - zone 702\$a Unimarc (voir Illustration 21).



Illustration 21 : Capture d'écran de la notice de l'ouvrage *Physiologie du calembourg*, sur le catalogue de la BnF

➤ **Problème n°4** : notice bibliographique avec dates de publication différentes

Pour certains documents, les dates de publication ne sont pas les mêmes sur la notice de la médiathèque de Roubaix et sur celle de la BnF. Après étude des résultats obtenus suite au premier alignement, nous pouvons constater que pour certains documents, cela semble être une erreur de saisie au moment du catalogage, par exemple lorsqu'il y a une erreur de 10 ans.

Exemple : *Les français peints par eux-mêmes*, publié aux éditions L Curmer en 1841.

→ Pour ce document, la date de publication sur la notice de la médiathèque de Roubaix est « 1851 » alors que celle sur le catalogue de la BnF est « 1841 ». Face à une différence de 10 ans, il est possible d'émettre l'hypothèse selon laquelle il s'agit d'une « erreur de frappe » de la part des catalogueurs de l'époque. Pour vérifier cette hypothèse, j'ai

effectué une vérification sur le document physique. La date de publication sur le document apparaît en chiffres romains, ce qui est possiblement à l'origine de la confusion lors du catalogage (voir Illustration 22).

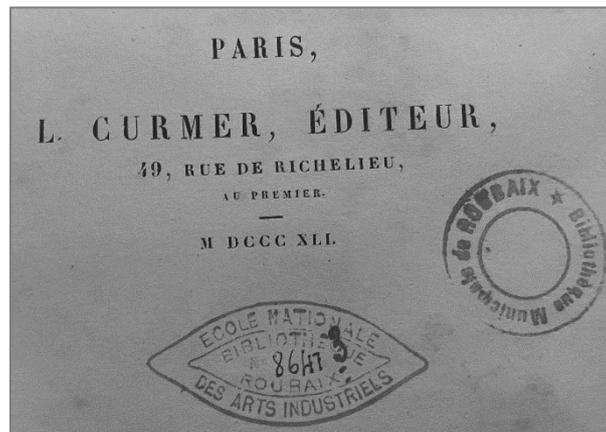


Illustration 22 : Photographie de la première page de l'ouvrage *Les français peints par eux-mêmes*<sup>22</sup>

➤ **Problème n°5** : documents propres à la ville de Roubaix

Certains documents sont introuvables sur le catalogue de la BnF car ce sont des « unica », c'est-à-dire des documents qui ne se trouvent qu'à la médiathèque de Roubaix. Le plus souvent, c'est le cas pour les documents officiels propres à la ville de Roubaix. Ainsi, puisque ces documents ne se trouvent pas dans le catalogue de la BnF, il est impossible pour le logiciel Bibliostratus de trouver un identifiant ARK. Finalement, dans cette situation, aucune solution n'apparaît pour améliorer la recherche automatique via Bibliostratus. Toutefois, si l'on attribue un identifiant ARK au document « unica » au sein du catalogue de la médiathèque de Roubaix, une fois visible sur le web de données, la notice permettra d'enrichir le catalogue général de la BnF.

Exemple : l'ouvrage *Exposition internationale du Nord de la France Roubaix 1911 rapport général*, publié en 1912 par G.Sayet est un unica de la médiathèque de Roubaix.

---

<sup>22</sup> Source : photographie personnelle.

➤ **Problème n°6** : erreur dans la saisie des informations lors de la création des notices

Étant donné que les documents du corpus à aligner ont été catalogués il y a déjà plusieurs années et sont des documents patrimoniaux, les notices ont forcément été saisies à la main et il n'était pas possible de récupérer des notices existantes pour ces documents. Ainsi, puisque l'erreur est humaine, il est possible que les notices possèdent des fautes de frappe, de syntaxe, d'orthographe dans le champ du titre ou des auteurs. De la même manière que pour le problème précédent, il apparaît compliqué de trouver une solution pour améliorer la recherche automatique de Bibliostratus. Il faudrait reprendre chaque notice pour vérifier si les informations ont correctement été saisies, et si ce n'est pas le cas, modifier chaque erreur manuellement. Or, cela prendrait énormément de temps.

Exemple 1 : *Daphnis et Chloé*, écrit par Longus (auteur principal), publié en 1878 aux éditions A Quantin.

→ Dans la notice bibliographique de la médiathèque de Roubaix, on peut constater une erreur d'orthographe dans le titre : « Daphnis et Cloé ». Bibliostratus réalisant une recherche d'exactitude pour le titre, il ne pouvait donc pas trouver la notice bibliographique sur le catalogue de la BnF et l'identifiant ARK correspondant (voir Illustration 23).



Illustration 23 : Capture d'écran de l'ouvrage de Longus, sur le catalogue de la médiathèque de Roubaix

Exemple 2 : *Les bourgeois de Molinchart*, écrit par Champfleury, publié en 1856 par la Librairie Nouvelle.

→ On peut constater également une faute d'orthographe dans le titre de cette notice, puisque l'écriture exacte est « Les bourgeois de Molinchart » et non « Les bourgeois de Monlinchart ». De même que pour l'exemple précédent, Bibliostratus ne pouvait pas trouver la notice correspondante sur le catalogue de la BnF en raison de cette erreur dans le titre.

### 1.3 Observations complémentaires

➤ Doublon de notices, nombre d'ARK trouvés pour un même document :

Au cours de l'analyse des résultats produits suite au premier passage dans Bibliostratus, d'autres observations sont faites. Dans un premier temps, il est possible de constater que pour certaines notices, le logiciel Bibliostratus ressort plusieurs identifiants ARK – entre 1 et 45 – pour le premier alignement. Il apparaît intéressant de se demander pourquoi un tel nombre d'identifiants ARK sont trouvés car lors de la dernière étape de l'alignement (l'import des notices dans le SIGB de la bibliothèque locale), il ne faut posséder qu'un seul ARK. La question a donc été posée lors d'un entretien avec Etienne Cavalié, adjoint au directeur du département des métadonnées à la BnF et fondateur de Bibliostratus. En réalité, ce dernier explique que pour construire le catalogue général informatisé de la BnF, on a repris tels quels les catalogues papiers de chacune des bibliothèques faisant partie du réseau de la BnF. Toutefois, certains documents étaient présents dans plusieurs bibliothèques et ont donc été ajoutés en double au moment de l'informatisation du catalogue. Au fil du temps, des campagnes de dédoublonnage du catalogue de la BnF ont été menées mais cela représente un travail rigoureux et chronophage, passant ainsi après d'autres projets, tels que la Transition bibliographique.

➤ La qualité des données en entrée, élément fondamental pour un alignement réussi :

Ensuite, il est important d'aborder la question de la qualité des données en entrée. En effet, la plupart des problèmes rencontrés sont dus à des faiblesses dans la saisie des données bibliographiques composant les notices. Or, pour que Bibliostratus soit performant et trouve un maximum d'identifiants ARK correspondant, il est essentiel que les informations soient bien saisies au sein de la notice bibliographique, dans les champs correspondants et de manière fiable par rapport aux documents physiques présents dans les fonds. Cela est plutôt complexe en raison de la nature patrimoniale des fonds traités car l'exemplarisation des documents à l'époque de leur acquisition n'était pas aussi cadrée et normalisée qu'aujourd'hui, ce qui peut expliquer les différences entre le catalogage de la BnF et celui de la médiathèque de Roubaix. Par exemple, c'est le cas pour la problématique rencontrée, liée au renseignement des volumes / tomes dans les notices.

➤ Intervenir sur les préférences de Bibliostratus pour un résultat plus performant ?

Au-delà des modules de nettoyage des données bibliographiques et d'alignement, il est possible de modifier certains paramètres dans Bibliostratus pour préciser ou apporter un certain contrôle sur la recherche effectuée par le programme (voir Illustration 24).

The screenshot shows the 'Editer les préférences' (Edit preferences) window. It contains the following settings:

- http\_proxy**: URL du proxy HTTP à utiliser dans le contexte d'un réseau professionnel. URL, login, Mot de passe.
- https\_proxy**: URL du proxy HTTPS à utiliser dans le contexte d'un réseau professionnel. URL, login, Mot de passe.
- display\_message\_in\_progress**: Afficher ou non le message en rouge quand on sélectionne un fichier à charger. Par défaut : 1 = message à afficher. 0 = ne pas afficher. Input: 1.
- marc2tables\_input\_format**: Format de données en entrée dans le module bleu (marc2tables). Par défaut : 'unimarc'. Autre valeur autorisée : 'marc21'. Input: unimarc.
- tsv\_csv**: Lors de la production de formats Colonnes en sortie, possibilité de choisir le séparateur : CSV (virgule) ou TSV (tabulations). Valeurs autorisées : csv, tsv. Valeur par défaut : tsv. Input: tsv.
- csv\_pref**: Pour les fichiers produits en format CSV, possibilité de choisir entre CSV (séparateur virgule) et TSV (séparateur Tabulation). Valeurs autorisées : csv, tsv. Par défaut : tsv. Input: tsv.
- stop\_words**: Si renseigné, nom du fichier où se trouve une liste de mots vides, à ignorer pour l'alignement. Par exemple, si la liste contient le mot 'Ste', le titre 'Notice historique sur Marizy-Ste-Geneviève' cherchera en réalité 'Notice historique sur Marizy Geneviève'. Input: (empty).
- timestamp**: Option (booléen True/False) pour injecter en dernière colonne des rapports d'alignement l'horodatage. Ceci pour évaluer la performance d'un alignement volumineux au fur et à mesure du traitement. Input: False.
- worldcat\_api\_key**: API Key OCLC pour interroger les web services Worldcat (recherche sur ISBN). Input: (empty).
- gmb**: Fichier d'alignement bibliographique : matrice Excel Gallica Marque Blanche. Valeurs : 0 [non] ou 1 [oui]. Input: 0.
- num\_parallel**: Nombre de requêtes réalisées simultanément pour aligner des notices avec la BnF ou IdRef/Sudoc. Input: 20.
- format\_ppn**: Mode de récupération des PPN : sous forme d'URI, ou simplement préfixé 'PPN'. Valeurs : 'ppn' ou 'uri'. Input: ppn.
- controle\_isbn2id\_date**: Paramètre (peut-être) temporaire pour tester le contrôle sur la date dans le cas d'alignements sur ISBN ou EAN. Input: True.
- controle\_editeur**: Paramètre (peut-être) temporaire pour tester le contrôle sur le nom d'éditeur pour les monographies (valeurs 0 : aucun contrôle ; 1 : valeur de l'un présente dans l'autre ; 2 : valeur strictement identique (sauf si l'un est vide et l'autre renseigné) ; 3 : valeur strictement identique). Input: 2.

Buttons: 'Enregistrer mes préférences' and 'Restaurer les paramètres par défaut'.

Illustration 24 : Capture d'écran de la page d'édition des préférences de Bibliostratus

Dans la seconde partie de cette analyse dédiée aux manipulations réalisées sur le fichier de départ, il sera question de modifier certains de ces paramètres, pour espérer aboutir à un meilleur résultat d'alignement. Il s'agira plus particulièrement de tester le paramètre « controle\_editeur » qui permettra à Bibliostratus de s'appuyer également sur la colonne des éditeurs pour faire la recherche avancée dans les catalogues des agences bibliographiques nationales.

## 2. Manipulation des fichiers en entrée pour améliorer les résultats obtenus

Afin de trouver des solutions aux différentes problématiques rencontrées suite au premier passage dans Bibliostratus, différentes itérations ont été réalisées, en modifiant le fichier de départ et en le repassant dans le programme. Ces différentes itérations correspondent aux

problématiques soulevées précédemment. Elles seront présentées ci-dessous, en détaillant les modifications et en exposant les résultats de chacune.

- Itération n°1

La première itération correspond au premier passage dans Bibliostratus, sans aucune modification du fichier d'entrée qui correspond, pour rappel, à l'export des notices depuis le SIGB de la médiathèque. Comme indiqué précédemment, à la suite de ce premier alignement, 16% des notices ont été alignées, c'est-à-dire qu'elles ont trouvées un identifiant pérenne ARK ou PPN. C'est à la suite de ces premiers résultats que découlent les observations et réflexions, afin de trouver des hypothèses de solutions pour améliorer le résultat.

- Itération 2

Suite au faible pourcentage de notices alignées après le premier alignement, je réalise avec l'aide de deux collègues une recherche manuelle d'identifiants ARK, pour les notices non-alignées lors du premier passage dans Bibliostratus. Ces recherches manuelles sont simplement des requêtes sur le moteur de recherche du catalogue de la BnF, grâce aux métadonnées d'identification des documents (titre, auteur, date, édition). Le but de ce travail de recherche est aussi de servir de base pour la suite, afin de posséder une liste des identifiants ARK pour les notices. Cela a permis également d'identifier de manière plus précise les différents problèmes rencontrés, en comparant les notices bibliographiques de la médiathèque de Roubaix avec celles du catalogue de la BnF. À la suite de ces recherches manuelles, 96 notices restent non-trouvées dans le catalogue de la BnF et ne possèdent donc pas d'identifiant ARK.

- Itération 3

Après avoir constaté que les notices comportant le nom du donateur (Pierre Destombes) dans la colonne des auteurs ne trouvaient pas d'identifiant lors de la recherche exécutée par Bibliostratus, je retire « Pierre Destombes » de la colonne des auteurs, en réalisant une suppression manuelle dans le fichier de départ. Je repasse ensuite ce nouveau fichier dans l'outil Bibliostratus. Suite à ce deuxième passage du fichier dans Bibliostratus, le résultat passe de 16% de notices alignées à 46% (voir Tableau 4).

Nombre d'identifiants trouvés	Nombre de notices concernées
0	1659
1	971
2	256
3	79
4	46
5	18
6	8
7	8
8	6
9	5
10	10
12	1
13	2
15	1
17	1
24	1
27	1
45	1
46	1
133	1
163	1
208	1
365	1
409	1
428	1
499	1

Tableau 4 : Résultats du deuxième passage dans Bibliostratus (sans P.Destombes)

En observant ce tableau de résultats, nous constatons que pour certaines notices, le programme trouve plus d'une centaine d'identifiants pour un même document. Cela s'explique par le manque de précisions dans les données pour certains documents. Par exemple, la notice qui possède 499 identifiants après l'alignement concerne l'ouvrage *Fables* de La Fontaine publié par Hachette. En allant vérifier physiquement le document dans le fonds, je constate qu'aucune date d'édition n'est indiquée, ce qui explique l'absence de date dans la notice et donc un certain taux de bruit dans les résultats.

- Itération 4

Après avoir constaté que les notices des documents possédant plusieurs tomes / volumes posaient problème dans la recherche effectuée par Bibliostratus, je réalise une nouvelle itération, en vidant le contenu de la colonne des tomes. Ainsi, je repasse dans le logiciel le fichier sans P.Destombes et sans les tomes. Suite à ce troisième passage du fichier dans Bibliostratus, le résultat passe de 46% de notices alignées 50,5% (voir Tableau 5).

Nombre d'identifiants trouvés	Nombre de notices concernées
0	1524
1	1025
2	287
3	103
4	43
5	25
6	6
7	19
8	7
9	5
10	10
12	1
13	7
16	1
17	1
24	1
27	1
45	1
46	1
64	5
85	2
133	1
163	1
208	1
365	1
409	1
430	1
499	1

Tableau 5 : Résultats du troisième passage dans Bibliostratus (sans P.Destombes, sans les tomes)

Suite à ce résultat, un autre constat est fait : pour certaines notices pour lesquelles un identifiant ARK était déjà présent dans le fichier d'entrée, Bibliostratus ne ressort aucun identifiant.

- Itération 5

Dans les préférences de l'outil Bibliostratus, il est possible d'ajouter certains paramètres pour renforcer le contrôle effectué lors de l'alignement (voir Illustration 22). Pour ce quatrième passage dans Bibliostratus, j'ajoute un contrôle sur l'éditeur, en choisissant le paramètre 1 : « valeur de l'un présente dans l'autre ». Je repasse le fichier de notices, sans P.Destombes, sans les tomes et avec le contrôle sur l'éditeur dans Bibliostratus. Suite à cette nouvelle itération, le résultat de l'alignement passe de 50,5% à 51% (voir Tableau 6).

Nombre d'identifiants trouvés	Nombre de notices concernées
0	1505
1	1037
2	289
3	105
4	45
5	26
6	6
7	19
8	7
9	5
10	10
12	1
13	7
16	1
17	1
24	1
27	1
45	1
46	1
64	5
85	2
133	1
163	1
208	1
365	1
409	1
430	1
499	1

Tableau 6 : Résultats du quatrième passage dans Bibliostratus (sans P.Destombes, sans les tomes, avec contrôle éditeurs)

- Itération 6

Concernant le fichier composé d'une notice de périodique, il a été remarqué précédemment que le type de support est erroné puisqu'il s'agit en réalité d'une monographie. Le document en question correspond à l'ouvrage *Les caresses*, écrit par Jean Richepin et publié aux éditions M. Dreyfous en 1882. Je repasse donc la notice dans Bibliostratus, en remplaçant les métadonnées

d'identification du document dans les colonnes correspondant aux documents texte. Ainsi, Bibliostratus trouve l'identifiant ARK du document. Il s'agissait donc simplement d'une erreur de la part du catalogueur de l'époque, au moment de la construction de la notice.

- Itération 7

De la même manière que pour le périodique, la nature des documents présents dans le fichier des partitions semblait inexacte, puisque quatre notices sur cinq semblaient correspondre à des monographies imprimées. Comme précédemment, je repasse le fichier des partitions dans Bibliostratus, dans un fichier correspondant aux métadonnées des monographies. Toutefois, le résultat est moins satisfaisant que pour l'itération précédente puisqu'une seule notice sur les cinq trouve son identifiant ARK.

Cependant, il semble y avoir un problème de correspondance au niveau des auteurs : les auteurs présents dans la colonne du fichier sont plus nombreux ou ne correspondent pas exactement aux auteurs présents dans la notice du catalogue de la BnF. Cela est surtout flagrant avec le document *Le musée pour rire dessins pour tous les caricaturistes de Paris* publié aux éditions Aubert en 1839. En effet, ce document apparaît en doublon dans le fichier de notice (voir Illustration 25). Toutefois, une des notices trouve un identifiant ARK et pas l'autre, il semble donc y avoir une différence dans les métadonnées d'identification. Effectivement, pour la notice qui ne trouve pas d'identifiant ARK, un auteur est présent en plus, qui n'est pas indiqué dans la notice du catalogue de la BnF. On peut donc affirmer au regard de ce constat que Bibliostratus ne renvoie pas d'identifiant lorsque d'un côté ou de l'autre (médiathèque de Roubaix ou BnF), des informations diffèrent.

B	C	D	E	F	G	I	J	K
Nb identifiar	Liste identifi	Méth	FRBNF	ARK	ISBN	Titre	Auteur	Date
1	ark:/12148/c	Titre-Auteur-Date + contrôle :				Le musée pour rire dessins pour tous les caricatu	Philipon Charles;Alhoy Maurice;Huart Louis	1839
0						Le musée pour rire dessins pour tous les caricatu	Philipon Charles;Alhoy Maurice;Huart Louis;Savigny A. de	1839
0						Chronique du règne de Charles IX	Mérimée Prosper;Lortic;Morin Edmond	1876
0						La double méprise	Mérimée Prosper;Bertrand;Chambolle René;Duru Alfred	1902
0						Ah les femmes	Jura Eddy;Bodart-Timal Charles;Rappart Francis	1920

Illustration 25 : Capture d'écran – extrait du fichier de résultat du second alignement des « partitions »

Enfin, pour la dernière notice de ce fichier, il semble que Bibliostratus ne trouve pas d'identifiant ARK sur le catalogue de la BnF puisque le document *Ah les femmes* de Jura Eddy

et Bodart-Timal Charles (chansonnier roubaisien) publié en 1920 correspond bien à une partition, mais la BnF ne la possède pas dans ses collections.

### 3. Bilan du travail d'alignement des notices

Finalement, l'alignement du corpus de notices issues du fonds Destombes aboutit, avec les différentes itérations réalisées, à 51% des notices alignées. Ce résultat n'est pas enthousiasmant mais cela conduit à se poser des questions, à la fois sur l'efficacité du programme et sur la qualité des données en entrée.

En effet, d'une part, la pauvreté des résultats de l'alignement avec Bibliostratus provient certainement de l'aspect patrimonial des notices à aligner et d'un manque de précision du logiciel pour ce type de documents. En effet, Bibliostratus effectue une recherche sur les catalogues des agences bibliographiques nationales par exactitude, en se basant sur l'opérateur booléen « ET » entre les différentes métadonnées d'identification. Pour améliorer la recherche et combler les lacunes de certaines notices, il faudrait que le programme réalise plutôt une recherche par similarité, notamment pour les champs du titre et de(s) auteurs(s). De plus, au sein des résultats, des faiblesses apparaissent puisque pour certaines notices, Bibliostratus ressort une grande quantité d'identifiants ARK correspondant au document, alors qu'à la fin du processus d'alignement, pour importer les notices alignées dans le SIGB local, il ne faut qu'un seul identifiant par notice. A noter que dans le cadre de ma mission au sein de la bibliothèque municipale de Roubaix, je me suis concentrée principalement sur les identifiants pérennes ARK présents dans le catalogue de la BnF plutôt que sur les identifiants PPN du Sudoc.

D'autre part, comme on a pu le remarquer à de multiples reprises au cours de l'analyse des résultats de l'alignement, la qualité des données en entrée joue un rôle essentiel dans la pertinence des résultats en sortie du logiciel. En effet, plus les notices bibliographiques contiennent des informations propres et fiables, plus le nombre de notices alignées sera important. Si les informations présentes dans les notices bibliographiques sont inexactes ou imprécises, cela augmentera le taux de bruit et de silence dans les résultats, et donc potentiellement le nombre d'identifiants fournis. Avant de passer à l'alignement des notices via Bibliostratus, il faut donc s'assurer de la qualité des données dont on dispose.

## Partie 3 : Discussion globale

### 1. Pistes de réflexion pour approfondir le travail

À travers cette étude du programme de la Transition Bibliographique et du travail d'alignement concret d'un corpus de notices, j'ai pu prendre conscience de la difficulté de ce projet, notamment car certaines zones d'ombres persistent. Ce mémoire apparaît donc comme une base de réflexion et de recherche pour aider les bibliothèques à faire leur transition et pour faire évoluer les outils à disposition des professionnels. En effet, à l'heure actuelle, outre Bibliostratus, il n'existe pas d'autre outil informatique permettant d'automatiser l'alignement des notices bibliographiques. Toutefois, on constate certaines failles dans le programme existant, qui permettront assurément de faire évoluer l'outil. Bibliostratus apparaît efficace pour les documents récents, catalogués selon des standards communs, mais rencontre des lacunes lorsqu'il s'agit de documents patrimoniaux, dont les métadonnées sont de moins bonne qualité.

Ainsi, les bibliothèques locales françaises, chacune à leur échelle, doivent s'appropriier les outils mis à leur disposition pour faire leur transition, en fonction du type de collections qu'elles possèdent et de leurs capacités informatiques. Pour le moment, le code de catalogage RDA-FR, qui permettra d'adapter les pratiques des catalogueurs aux standards du web sémantique, est encore en cours d'élaboration. Une fois que le nouveau code de catalogage sera publié et utilisable, il s'agira alors de trouver un moyen pour les bibliothèques de s'aligner et de rendre leurs données propres et bien structurées. Nettoyer les catalogues en amont de l'alignement des notices est un aspect essentiel. Pour les bibliothèques possédant d'importantes collections patrimoniales, il s'agirait de trouver une façon de nettoyer les notices, pour qu'elles soient conformes aux nouvelles normes qui découlent du mouvement de la Transition Bibliographique.

### 2. Limites et obstacles rencontrés

Les principales difficultés rencontrées au cours de ce travail ont été à la fois la manipulation des données mais aussi la compréhension du processus interne d'alignement des notices bibliographiques. Cela était notamment dû à la qualité insuffisante des données en entrée, puisque si les notices avaient été construites d'une manière propre et standardisée, elles auraient toutes trouvé un identifiant pérenne. Or, cette difficulté était liée à l'aspect patrimonial du corpus traité. Il s'agit donc pour les responsables de fonds patrimoniaux de réfléchir à une

solution qui permettrait d'adapter les notices aux standards actuels et cela passe par « *une nouvelle approche du signalement* » (Palluault, 2012). Ainsi, le programme de la Transition bibliographique, en plus d'avoir un impact sur la visibilité des catalogues des bibliothèques sur le web, remet également en question les pratiques professionnelles, dans chacun des pans du travail de bibliothécaire, notamment dû à l'évolution des modèles de description des documents numériques.

De plus, le manque de compétences techniques, plus particulièrement liées à l'informatique, a été une limite importante dans la réalisation de mon étude. En effet, davantage de compétences en programmation m'auraient permis de me réapproprier le script de Bibliostratus et de l'adapter aux différentes problématiques rencontrées. Aussi, une maîtrise plus poussée du langage Python m'aurait également aidée pour réaliser des analyses plus précises sur les données à traiter. C'est pourquoi il est important pour les professionnels des bibliothèques de posséder un minimum de compétences en informatique et en programmation, d'autant plus dans les perspectives à venir, plus particulièrement liées à l'exposition des catalogues sur le web de données.

### 3. Ouverture. Quels enjeux concrets pour les bibliothèques ?

A l'échelle actuelle, dans le cadre du programme de la Transition Bibliographique, l'enjeu pour les bibliothèques est de passer de la logique de documents à la logique de données, afin de commencer à envisager la présence de leurs catalogues en ligne sur le web de données. Le but est de lancer les bibliothèques locales françaises, chacune à leur rythme et selon leurs moyens, dans un processus de structuration et de standardisation de leurs données bibliographiques. Le nouvel objectif est d'arriver à un catalogue bien propre et structuré, en inculquant de nouvelles méthodes de travail dans les pratiques des bibliothécaires. C'est pour cela que le programme de la Transition Bibliographique œuvre sur les différents volets techniques, et notamment la transposition du code de catalogage RDA-FR et le travail d'alignement des données. Une fois que le travail d'alignement des données par rapport à celles des agences bibliographiques aura aboutie, il sera possible d'envisager leur liage et leur interopérabilité sur le web. En effet, il faudra plusieurs dizaines d'années avant que tout le travail mené par les bibliothèques dans le cadre de ce programme soit visible sur le web, mais l'objectif est d'aboutir de manière progressive aux exigences du web sémantique.

Outre cet aspect technique, il faut également commencer à réfléchir à « l'après Transition Bibliographique », du point de vue des usagers des bibliothèques et des bénéficiaires qu'ils pourront tirer de ce long travail de structuration des données. Quelques pistes de réflexion émergent des différents échanges entre professionnels, notamment par rapport à la modélisation des données sur les futurs catalogues en ligne, au renouvellement des interfaces et aux services innovants qui seront proposés aux utilisateurs. Le web de données permettra aux catalogues en ligne des bibliothèques de se renouveler et de prendre davantage de place sur la scène informationnelle, afin de démocratiser la culture et d'accrocher de nouveaux publics.

## Conclusion générale

Pour conclure, à l'échelle des bibliothèques locales françaises, le programme de la Transition Bibliographique peut paraître complexe et abstrait pour les bibliothécaires. Les bibliothèques doivent se projeter dans un travail progressif, en fonction des moyens dont elles disposent et à leur rythme. En effet, pour ce qui est du chantier d'alignement des notices des bibliothèques locales à celles des agences bibliographiques nationales, il s'agit d'un travail plutôt aride mais on peut espérer que les outils mis à disposition des professionnels continuent d'évoluer afin de rendre cette tâche un peu plus aisée. Pour cela, il est important de continuer à mener le programme de la Transition Bibliographique de manière collaborative et de suivre leurs travaux jusqu'à ce que les efforts des bibliothécaires soient visibles sur le web. Cet aspect collaboratif de la Transition Bibliographique n'est pas à négliger et chaque implication des professionnels a son importance. Les chercheurs et universitaires du domaine des sciences de l'information-documentation ont peut-être également un rôle à jouer dans la concrétisation de ce programme, afin d'apporter un regard scientifique au projet. De plus, il est indispensable d'axer les efforts sur la formation des personnels des bibliothèques aux enjeux de la Transition Bibliographique qui vont impacter leur travail au quotidien, à la fois dans le traitement intellectuel des documents mais aussi dans la médiation avec les usagers. Effectivement, il ne faut pas oublier que le but de la Transition Bibliographique est d'améliorer l'expérience utilisateurs des usagers et de rendre les collections des bibliothèques davantage visibles sur le web de données.

Ainsi, ma mission d'alignement des métadonnées d'un corpus de documents patrimoniaux, au sein de la médiathèque de Roubaix, n'a pas été simple tant d'un point de vue théorique que pratique. Cependant, je suis très fière d'avoir mené des réflexions utiles pour faire avancer les discussions sur les outils d'alignement des notices. J'ai également pu développer mes compétences en termes de méthodologie d'analyse et de travail en autonomie. Enfin, ce stage m'aura permis de concrétiser mon souhait de travailler dans le domaine des bibliothèques.

## Références bibliographiques

### Ouvrages

BERMÈS, Emmanuelle, ISAAC, Antoine et POUPEAU, Gautier, 2013. *Le web sémantique en bibliothèque*. Paris : Éd. du Cercle de la librairie. Collection Bibliothèques. ISBN 978-2-7654-1417-9.

LERESCHE, Françoise, 2016. La Transition bibliographique. In : *Vers de nouveaux catalogues*, pp.49-67. Éditions du Cercle de la Librairie. ISBN 978-2-7654-1513-8. DOI [10.3917/elec.berme.2016.01.0049](https://doi.org/10.3917/elec.berme.2016.01.0049).

OTLET, Paul, 1934. *Traité de documentation*. Bruxelles, Palais mondial.

PALLUAULT, Florent, 2012. 5. Décrire et signaler des documents patrimoniaux : enjeux, formats, perspectives. In : *Apprendre à gérer des collections patrimoniales en bibliothèques*. Presses de l'enssib. La boîte à outils. ISBN 979-10-91281-01-0. DOI [10.4000/books.presseenssib.666](https://doi.org/10.4000/books.presseenssib.666).

TOUSSAINT, Claire, 2019. La transition bibliographique. In : *Le métier de bibliothécaire*. Ed. du Cercle de la Librairie. ISBN 978-2-7654-1578-7.

### Articles

CANTIÉ, Philippe, 2015. Transition bibliographique : en avant marche ! *Bulletin des Bibliothèques de France (BBF)* [en ligne]. No 6. Disponible à l'adresse : [https://bbf.enssib.fr/tour-d-horizon/transition-bibliographique-en-avant-marche\\_65461](https://bbf.enssib.fr/tour-d-horizon/transition-bibliographique-en-avant-marche_65461) [consulté le 15 juillet 2023].

GOUBINA, Maya, 2018. La transition bibliographique au-delà de la normalisation : partage des données, impacts sur les métiers, enjeux pour l'offre de services. *Bulletin des Bibliothèques de France (BBF)* [en ligne]. No 16. Disponible à l'adresse : [https://bbf.enssib.fr/tour-d-horizon/la-transition-bibliographique-au-dela-de-la-normalisation\\_68598](https://bbf.enssib.fr/tour-d-horizon/la-transition-bibliographique-au-dela-de-la-normalisation_68598) [consulté le 15 juillet 2023].

HERNANDEZ, Claire, 2018. Comment interroger la Bnf et Wikidata avec Sparql ? *BibLibre - Services et logiciels libres pour votre bibliothèque - SIGB Koha, portail Bokeh, numérique, gestion* [en ligne]. 18 juillet 2018. Disponible à l'adresse : <https://www.biblibre.com/fr/blog/comment-interroger-la-bnf-et-wikidata-avec-sparql/> [consulté le 15 juillet 2023].

LERESCHE, Françoise et LUCERNA, Géraldine, 2015. L'adapter puis l'adopter : la position française par rapport à RDA. [en ligne]. 2015. Disponible à l'adresse : <https://experts.bnf.fr/node/38118> [consulté le 15 juillet 2023].

MINISTÈRE DE LA CULTURE, 2019. *Ouverture des données publiques culturelles. Guide pratique*. [en ligne]. Disponible à l'adresse : <https://www.culture.gouv.fr> [consulté le 15 juillet 2023].

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION - Secrétariat GÉNÉRAL, 2014. Feuille de route stratégique - métadonnées culturelles et transition Web 3.0 - Notice bibliographique | Enssib. [en ligne]. janvier 2014. Disponible à l'adresse : <https://www.enssib.fr/bibliotheque-numerique/documents/64776-feuille-de-route-strategique-metadonnees-culturelles-et-transition-web-3-0.pdf> [consulté le 15 juillet 2023].

RIVA, Pat, LE BOEUF, Patrick et ZUMER, Maja, 2017. Un modèle conceptuel pour l'information bibliographique. *International Federation of Library Associations and Institutions (IFLA)*.

SAJUS, Bertrand et LEROI, Marie-Véronique, 2016. Le développement du web des données culturelles: Les enjeux pour le ministère de la Culture et de la Communication. *I2D - Information, données & documents*. Vol. 53, no 2, p. 46. DOI [10.3917/i2d.162.0046](https://doi.org/10.3917/i2d.162.0046).

## Sites web

BERNERS-LEE, Tim, 2006. Linked Data - Design Issues. [en ligne]. 2006. Disponible à l'adresse : <https://www.w3.org/DesignIssues/LinkedData.html> [consulté le 15 juillet 2023].

Bibliogrill : la lecture publique doit-elle faire sa transition bibliographique ? - 18 avril 2023 - Bibliothèque publique d'information - Centre Pompidou (professionnels), 2023 [en ligne]. Disponible à l'adresse : <https://pro.bpi.fr/bibliogrill-lecture-publique-faire-transition-bibliographique/> [consulté le 15 juillet 2023].

Bienvenue sur le site de la Transition bibliographique. *Transition bibliographique - Programme national* [en ligne]. Disponible à l'adresse : <https://www.transition-bibliographique.fr/> [consulté le 15 juillet 2023].

Code RDA-FR. *Transition bibliographique - Programme national* [en ligne]. Disponible à l'adresse : <https://www.transition-bibliographique.fr/rda-fr/> [consulté le 15 juillet 2023].

*English: Users tasks for IFLA LRM*. [en ligne]. Disponible à l'adresse : [https://commons.wikimedia.org/wiki/File:Sch%C3%A9ma\\_T%C3%A2ches\\_utilisateurs\\_IFLA\\_LRM.jpg](https://commons.wikimedia.org/wiki/File:Sch%C3%A9ma_T%C3%A2ches_utilisateurs_IFLA_LRM.jpg) [consulté le 15 juillet 2023].

Données d'autorité à la BnF. *BnF - Site institutionnel* [en ligne]. Disponible à l'adresse : <https://www.bnf.fr/fr/donnees-autorite-bnf> [consulté le 15 juillet 2023].

La Transition bibliographique en France. *Transition bibliographique - Programme national* [en ligne]. Disponible à l'adresse : <https://www.transition-bibliographique.fr/enjeux/position-francaise-rda/> [consulté le 15 juillet 2023].

Le projet Fichier national d'entités. *Transition bibliographique - Programme national* [en ligne]. Disponible à l'adresse : <https://www.transition-bibliographique.fr/fne/fichier-national-entites/> [consulté le 15 juillet 2023].

LRM Entity-Relationship Model, 2023 *iflstandards.info* [en ligne]. Disponible à l'adresse : <https://www.iflstandards.info/lrm/lrmer.html> [consulté le 15 juillet 2023].

Qui sommes-nous – *Etalab* [en ligne]. Disponible à l'adresse : <https://www.etalab.gouv.fr/qui-sommes-nous> [consulté le 15 juillet 2023].

Sparnatural Partitions [en ligne]. Disponible à l'adresse : <https://sparnatural.eu/demos/demo-partitions/index.html> [consulté le 15 juillet 2023].

Vers de nouveaux SGB. *Transition bibliographique - Programme national* [en ligne]. Disponible à l'adresse : <https://www.transition-bibliographique.fr/systemes-et-donnees/recommandations/> [consulté le 15 juillet 2023].

Web sémantique et modèle de données, 2018 *BnF Data* [en ligne]. Disponible à l'adresse : <https://data.bnf.fr/semanticweb> [consulté le 15 juillet 2023].

## Références juridiques

« Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal » - Légifrance.

« Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique » - Légifrance.

## Annexes

### Annexe 1 : Organigramme non-nominatif de la médiathèque La Grand Plage de Roubaix (Source : documentation professionnelle fournie par le lieu de stage)

