

## Les métadonnées dans le traitement bibliographique

---

Les métadonnées (« metadatas » en anglais) ont de nombreuses applications, mais il convient d'insister sur le fait que nous nous limiterons dans la présente fiche à la problématique des métadonnées dans le monde de l'information et plus précisément dans le traitement de cette dernière.

### 1. Définition et typologie

**Les métadonnées sont des données structurées qui servent à identifier, décrire, qualifier d'autres données permettant ainsi de définir, d'explicitier, de localiser une ressource informationnelle, quel qu'en soit le support, pour pouvoir y accéder.**

Elles sont apparues au milieu des années 90 avec le développement du web et sont indispensables à la structuration de l'information sur la toile car elles permettent de hiérarchiser et de relier les données entre elles.

Bien qu'il soit très difficile d'en fixer les frontières, les métadonnées peuvent être grossièrement classées selon la typologie suivante :

- Les **métadonnées descriptives** : elles visent à qualifier les ressources en fournissant des informations telles que le titre, l'auteur, les sujets, l'ISBN etc. afin de les identifier mais également de les localiser et de les lier à d'autres ressources connexes grâce à des identifiants uniques : les URI (Uniform Resource Identifier).
- Les **métadonnées structurelles** : elles placent les ressources dans leur environnement en explicitant les relations qu'elles peuvent avoir avec d'autres ressources (indications sur la collection, indication de l'appartenance à telle ou telle revue pour un article, titres en relations, etc.).
- Les **métadonnées de gestion** : elles fournissent des informations liées aux droits d'usage, d'exploitation et d'accès, aux données budgétaires et administratives.

A la lecture de cette typologie, on remarque que les métadonnées descriptives font partie du traitement des données bibliographiques depuis qu'existent les catalogues de bibliothèques. L'avènement des OPAC (Online public access catalog - catalogue de bibliothèque accessible en ligne) leur a ajouté des métadonnées structurelles et depuis maintenant plus de 15 ans les métadonnées de gestion s'y sont greffées. Ainsi aujourd'hui, ces 3 types de métadonnées sont devenus complémentaires et indissociables dans le travail des professionnels de l'information de qui elles sont cependant de moins en moins

l'apanage. C'est pourquoi la problématique de leur constitution et de leur gestion a un impact croissant sur l'évolution des métiers de la documentation.

## 2. Métadonnées et web sémantique

Les métadonnées sont intrinsèquement liées à l'évolution du web de données vers le web sémantique. Le travail effectué depuis plusieurs années sur la transition bibliographique<sup>1</sup> s'appuie justement sur l'harmonisation et l'évolution des métadonnées afin qu'elles s'approchent au plus près usages et contenus du web. L'objectif final étant d'optimiser la visibilité des catalogues de bibliothèques, la question des modalités d'accès aux documents et des liens entre les différentes ressources est essentielle.

Même si, selon Tim Berners-Lee, ce qu'on appelle « web sémantique » reste une utopie à réaliser car à l'heure actuelle l'objectif est encore de construire le web de données, on peut en schématiser les principes fondamentaux de la manière suivante :

1. **Toute ressource du web possède une étiquette (métadonnées).**
2. **Toute étiquette est lisible par les agents logiciels et par les êtres humains.**
3. **Toute étiquette représente fidèlement la ressource.**
4. **Toute étiquette est disposée dans un espace commun de lisibilité potentiellement explorable par l'ensemble des humains et des agents logiciels. La sélection de l'étiquette rend visible ou accessible la ressource.**

Les métadonnées correspondent ainsi à des marqueurs qui sont introduits dans les fichiers ou dans des langages de programmation appropriés, les langages de marquage XML. Ces marqueurs ont pour effet d'améliorer l'efficacité des recherches d'information par rapport aux recherches plein texte.

## 3. Métadonnées et traitement bibliographique

### 3.1. Un enjeu majeur

Etant un élément essentiel de l'architecture web, les métadonnées **favorisent le signalement des ressources (quelles qu'elles soient), ainsi que leur accès sur la toile**. L'utilisation des métadonnées dans le cadre du traitement bibliographique constitue donc un enjeu majeur : **plus l'information circule sur la toile, plus les bibliothécaires,**

---

<sup>1</sup> Pour approfondir : <https://www.transition-bibliographique.fr/> (consulté le 13/07/2024)

**professionnels du signalement et de l'accès à l'information, doivent s'emparer des outils de la toile.**

Les points suivants viennent étayer cette constatation :

- La troncature, les limitations chronologiques et géographiques ainsi que les opérateurs de proximité ou booléens des moteurs de recherche constituent certainement des bons outils de précision, mais ils ne suffisent pas toujours à augmenter la pertinence des résultats. L'inefficacité des moteurs lors d'interrogations sur le plein texte notamment est un constat que les professionnels de l'information, comme les usagers, font quotidiennement.
- Pour traiter les documents électroniques, le catalogage traditionnel, axé sur la description bibliographique, ne peut être suffisant. Les ressources visuelles constituent un exemple flagrant : sans l'apport des métadonnées, elles seraient tout simplement inexploitable et dépourvues de toute traçabilité.
- Les métadonnées sont internationales et multilingues, et font l'objet de constantes tentatives de standardisation (cf. les recommandations du W3C) ce qui garantit leur interopérabilité et augmente grandement les chances de repérage des ressources, quelle que soit leur origine.
- Les métadonnées permettant de diversifier les points d'accès à l'information ainsi que la présentation des résultats de recherche se révèlent d'une importance cruciale à l'heure où les usages de l'information sont de plus en plus multidisciplinaires et interprofessionnels.
- Le web compte plusieurs milliards de documents et s'enrichit à chaque instant. Si aucune règle ne structure leur classement, ces ressources perdent une grande partie de leur pertinence parce qu'elles ne sont pas repérées par les moteurs de recherche.

En outre, les métadonnées permettent de formaliser, de modéliser et de structurer les aspects juridiques, administratifs, communicationnels et techniques de ces mêmes ressources (droits d'usage, de reproduction, données économiques et budgétaires, liens avec d'autres ressources etc.). Elles conditionnent aussi la pérennité des ressources et leur accès à long terme. **Elles facilitent donc largement la gestion et le maintien des collections numériques.**

Par ailleurs, **liées aux données d'usage, les métadonnées peuvent fournir des statistiques précieuses** en vue d'évaluer l'adéquation d'une ressource avec les besoins des publics et d'aider ainsi à ajuster la politique documentaire d'un établissement.

### 3.2. Métadonnées et outils de description bibliographique

Ce sont à la fois la **dématérialisation des documents et l'évolution des pratiques de recherche d'information des usagers**, incluant de plus en plus des besoins de navigation qui **ont entraîné pour le catalogage traditionnel une véritable révolution** : il faut intégrer à la description des liens permettant d'élargir la recherche vers une tentative d'exhaustivité et même l'accès direct au document primaire.

Depuis l'avènement des catalogues en ligne, la description bibliographique est traditionnellement encodée dans des formats spécifiques à l'image des formats MARC (Machine readable cataloging). Ces formats permettent à la machine d'étiqueter, de reconnaître l'information bibliographique afin de pouvoir l'exploiter. Cependant, ces formats sont des formats dits « propriétaires », ils ne sont utilisables que dans le monde des bibliothèques : construits sur le modèle des zones de la description bibliographique, ils ne sont pas assez souples pour permettre une totale compatibilité avec les formats du web. C'est pourquoi Roy Tennant écrivait en 2002 « A mort MARC ! »<sup>2</sup>. En fait, les limites du format concernent essentiellement MARC 21 (qui, entre autres, ne permet pas de faire aisément des liens).

En revanche, INTERMARC et UNIMARC, utilisés respectivement à la BnF et dans le Sudoc de l'Abes, ont pu évoluer constamment pour prendre en compte les nouvelles exigences de description et d'accès à l'information. Bien qu'ils soient tous deux voués à être remplacés dans un avenir plus ou moins proche par des langages web, la BnF et l'Abes ont décidé de continuer à faire évoluer ces formats. En effet, les données et métadonnées recueillies dans les bases bibliographiques utilisant ces formats sont riches (fruits d'un travail mutualisé et porteuses de nombreux liens), structurées et normalisées (la modélisation y a été introduite avec succès), fiables (parmi les métadonnées, on trouve notamment des identifiants pérennes (ISBN, ISSN, ISNI, PPN, ARK...). Elles pourront donc être réutilisées pour une mutation/transformation.

**Aujourd'hui ces formats purement bibliographiques cohabitent donc avec d'autres formats de description qui utilisent des jeux de métadonnées** : schémas élaborés pour prendre en compte la diversité des documents en introduisant certains éléments spécifiques de description : configuration requise pour les documents électroniques par exemple ; l'établissement de soutenance, le directeur de thèse... pour la description des thèses ; le niveau, la nature de la ressource (examen, questionnaire, simulation, expérience, cours...) pour la description des ressources pédagogiques ; dans la description d'une archive, les informations sur le traitement par exemple.

---

<sup>2</sup> « MARC must die ! » *Library Journal*, 15/10/02

**Ces jeux de métadonnées sont basés sur des langages informatiques** tels que XML (Extensible Markup Language) ou RDF (Resource Description Framework), et principalement utilisés pour favoriser l'interopérabilité entre les différents systèmes.

Ainsi, parmi les schémas ou jeux de métadonnées utilisés, on peut citer TEF (Thèses Electroniques Françaises) pour les thèses, LOM FR (Learning Object Metadata version française) pour les ressources pédagogiques, ONIX pour le monde de l'édition ou encore EAD (Encoded Archival Description) pour les archives.

A ces schémas spécifiques s'ajoute le Dublin Core, format qui propose un jeu de métadonnées standard et plus généraliste permettant l'échange des données de description (cf. focus en Annexe).

« L'objectif du Dublin Core est de fournir un socle commun d'éléments descriptifs pour améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré. Le Dublin Core prévoit 15 éléments, tous facultatifs et tous répétables, qui portent sur la description du contenu, de la propriété intellectuelle, de l'instanciation. »<sup>3</sup>

La caractéristique principale de Dublin Core est sa grande simplicité et sa souplesse d'utilisation. Il est compatible avec différents langages (XML, HTML etc.).

Nous avons souligné ci-dessus que **l'interopérabilité et le moissonnage sont deux notions essentielles pour la circulation de l'information dans tout réseau** et sur la toile en particulier. L'interopérabilité est primordiale pour les établissements documentaires à l'heure du travail en réseaux et de leur présence sur la toile, **afin d'assurer la multi-accessibilité des données**, pour les professionnels comme pour les usagers.

Actuellement, le protocole d'interopérabilité pour les données adoptées par de nombreuses bibliothèques est le protocole OAI-PMH (Open archives initiative - Protocol for Metadata Harvesting = Protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes).

**Le protocole OAI-PMH est un moyen d'échanger sur internet des métadonnées entre plusieurs institutions**, afin de multiplier les accès aux documents numériques. Il est mis en œuvre dans de nombreux réservoirs de documents (par exemple la bibliothèque numérique Gallica, la base de données Medline ou STAR).

Il permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine.

---

3 <https://www.bnf.fr/fr/dublin-core> (consulté le 13/07/2024)

**Grâce au protocole OAI, une bibliothèque, en tant que fournisseur de données, a la possibilité d'offrir une visibilité accrue à ses documents**, notamment à ses publications électroniques ou à ses fonds spécialisés.

Réciproquement, en tant que fournisseur de service, une bibliothèque peut réaliser une base de données ou un portail documentaire dans son domaine de spécialité ou sur un thème défini, en collectant les données descriptives de ressources et documents de tous types, accessibles sur l'internet dans des entrepôts OAI.

### 3.3. Les métadonnées et le bibliothécaire

L'avènement du web de données et celui du web sémantique a rendu l'utilisateur acteur de la production de données et de métadonnées bibliographiques : toute personne qui publie de l'information sur la toile a en effet à cœur que celle-ci n'y soit pas perdue. L'utilisateur s'empare ainsi directement (s'il entre les métadonnées lui-même) ou indirectement (si cela est fait automatiquement) de techniques de description bibliographique (signalement, indexation...) qui ne sont donc plus l'apanage des seuls bibliothécaires.

Or on a souligné **l'importance cruciale de standardiser les métadonnées ou de recourir à des métadonnées compatibles avec les standards du W3C afin de renforcer la qualité et la pérennité de la circulation de l'information sur le web**. C'est là que doit se positionner le professionnel de la documentation. **Les bibliothécaires ont**, grâce à leur expertise en matière de contrôle bibliographique et de production de données structurées, **un rôle crucial à jouer dans le développement de métadonnées de qualité à l'échelle du web et non plus seulement du catalogue**.

D'une part, **côté technique, le rôle du catalogueur évolue de plus en plus vers le contrôle qualité des données, la vérification de la pérennité et de la fiabilité des métadonnées et des liens**. Rebaptisé « **cata-lieur** » au cours de Journées Abes, c'est lui « qui (re-)travaillera la donnée » selon l'expression de Dominique Cotte<sup>4</sup>. (Re-)travailler concrètement, c'est **aligner les métadonnées** pour traquer les doublons et les manques en élaborant des outils et en les utilisant pour corriger **et ajouter liens et identifiants pérennes** (par exemple Paprika développé par l'Abes, basé sur le programme SudoQual, cet outil est mis à disposition des catalogueurs du Sudoc<sup>5</sup>).

D'autre part, **côté médiation, le bibliothécaire se doit de mettre son expertise au service de l'utilisateur** : la **sensibilisation**, l'**accompagnement**, la **formation** sont des étapes incontournables à la saisie de métadonnées de qualité par la personne qui met en ligne l'information. Dans l'enseignement supérieur, cette médiation s'adresse par exemple aux

---

<sup>4</sup> Les métiers de l'information et la "donnée" : analyse d'un monde en mutation. *Documentaliste-Sciences de l'Information* 2013/3 (Vol. 50). 78 pages.

<sup>5</sup> <http://www.abes.fr/Autorites-et-referentiels/Services-disponibles/paprika.idref> (consulté le 13/07/2024)

enseignants chercheurs qui déposent les documents qu'ils produisent dans des réservoirs d'archives ouvertes (comme HAL – Hyper Articles en Ligne), car ce sont eux qui attribuent eux-mêmes des métadonnées à leurs ressources en rédigeant la description bibliographique au moment du dépôt. L'enjeu est à la fois politique (valoriser, rendre visible un laboratoire, un établissement, un pays, à travers l'avancée de la recherche), juridique (pouvoir identifier de manière fiable pour gérer les droits d'auteur) et scientifique (favoriser l'accès pérenne et la citation).

### Pour aller plus loin :

Nous vous recommandons les vidéos suivantes proposées par la BnF et le CNFPT qui reviennent sur des notions évoquées dans cette fiche.

- Web sémantique, web de données : définitions / Emmanuelle Bermès
- L'interopérabilité : définition et enjeux / Emmanuelle Bermès
- Les cinq étoiles de l'open data / Emmanuelle Bermès
- Les services de la bibliothèque et le web / Frédérique Joannic-Seta

Elles sont consultables via la plateforme vidéo du CNFPT : <https://video.wikiterritorial.cnfpt.fr> (consulté le 13/07/2024)

[Web sémantique, web de données: définitions/ Emmanuelle Bermès](#)

[L'interopérabilité: définition et enjeux / Emmanuelle Bermès](#)

[Les cinq étoiles de l'open data / Emmanuelle Bermès](#)

[Les services de la bibliothèque et le web / Frédérique Joannic-Seta](#)

Et pour les compléter :

- *Punktokomo*, le blog technique de l'Abes, et en particulier la série d'articles concernant la mise en réseau des données.  
<https://punktokomo.abes.fr/2016/05/16/mettre-nos-donnees-en-reseau-un-demonstrateur-1-introduction/> (consulté le 13/07/2024)
- *Arabesques* (pages consultées le 13/07/2024)
  - n° 83 (juillet – août - septembre 2016) dossier « Créer du lien, faire sens »  
<http://www.abes.fr/Publications-Evenements/Arabesques/Arabesques-n-83>
  - n° 85 (avril – mai – juin 2017) dossier « Autorités, identifiants, entités / L'expansion des référentiels »  
<http://www.abes.fr/Publications-Evenements/Arabesques/Arabesques-n-85>
  - n° 95 (octobre – novembre – décembre 2019) dossier « Les bibliothèques au service des chercheurs / Anticiper, irriguer, accompagner »  
<http://www.abes.fr/Publications-Evenements/Arabesques/Arabesques-n-95>

- *Journées professionnelles sur les métadonnées en bibliothèques*, journées professionnelles annuelles du groupe Système & Données (Transition bibliographique) qui offrent un espace de visibilité et de débat autour de solutions innovantes et qui permettent de rappeler les bonnes pratiques en matière de préparation des catalogues au modèle conceptuel IFLA LRM et d'accompagner au changement les professionnels chargés des données.  
<https://www.transition-bibliographique.fr/systemes-et-donnees/journees-professionnelles/> (consulté le 13/07/2024)

A suivre aussi sur X (Twitter) le #databiblio

#### 4. Annexe : Focus sur le Dublin Core

Créé en 1995 à Dublin (Ohio) par OCLC (Online Computer Library Center) et le NCSA (National Center for Supercomputing Applications), le Dublin Core est un format descriptif à la fois simple et générique, comprenant 15 éléments différents.

Propriété	Description
Title	Titre du document : il s'agit <i>a priori</i> du titre principal du document
Creator	Créateur du document : nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document
Subject	mots-clefs, phrases de résumé, ou codes de classement
Description	Description du document : résumé, table des matières, ou texte libre
Publisher	Editeur du document : nom de la personne, de l'organisation ou du service à l'origine de la publication du document
Contributor	Contributeur au document : nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document
Date	Date d'un événement dans le cycle de vie du document : il peut s'agir par exemple de la date de création ou de la date de mise à disposition
Type	Nature ou genre du contenu : grandes catégories de document
Format	Format du document : format physique ou électronique du document. Par exemple, type de média ou dimensions (taille, durée). On peut spécifier le matériel et le logiciel nécessaires pour accéder au document
Identifier	Identifiant non ambigu (par exemple : les numéros ISBN)
Source	Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question
Language	Langue du document
Relation	Lien vers une ressource liée
Coverage	Portée du document : la portée inclut un domaine géographique, un laps de temps, ou une juridiction (nom d'une entité administrative)
Rights	Droits relatifs à la ressource : permet de donner des informations sur le statut des droits du document, par exemple la présence d'un copyright, ou un lien vers le détenteur des droits. L'absence de cette propriété ne présume pas que le document est libre de droits

Un format descriptif un peu plus fin se fonde sur le Dublin Core : le Dublin Core qualifié (qualified Dublin Core). Il comprend trois éléments supplémentaires : audience, provenance, titulaires des droits liés à l'exploitation de la ressource, ainsi que des qualificatifs permettant d'affiner la description.

Le Dublin Core non qualifié (sans les raffinements) est devenu une norme internationale en 2003 : ISO 15836. La version actuelle, dite 1.1, correspond à la recommandation de 2007 RFC 5013 de l'IETF (Internet Engineering Task Force).

L'agence de maintenance du Dublin Core est la DCMI (Dublin Core Metadata Initiative), organisation non gouvernementale et non lucrative, enregistrée à Singapour et œuvrant au développement de formats de métadonnées interopérables.

Le Dublin Core est la principale initiative visant à la convergence des éléments de métadonnées à utiliser. Il a d'ailleurs été adopté par 9 pays (essentiellement anglo-saxons) comme référentiel de métadonnées de niveau gouvernemental.

En pratique, le Dublin Core peut s'appliquer à différents langages (XML, HTML etc.). En voici quelques exemples :

- **Titres d'objets (ici des textes) :**

```
<meta name = "DC.Title"
content = "Polycyclic aromatic hydrocarbon contamination">
```

```
<meta name = "DC.Title"
content = "Crime and Punishment">
```

```
<meta name = "DC.Title"
content = "Still life #4 with flowers">
```

- **Langue d'un objet + son titre** (ici la première partie d'une partition du « clavier bien tempéré » de Bach en allemand) :

```
<meta name = "DC.Title"
lang = "de"
content = "Das Wohltemperierte Klavier, Teil I">
```

- **Page créée en (X)HTML** (HTML enrichi de métadonnées) :

```
<html>
<head profile="http://dublincore.org/documents/2010/08/04/dc-html/">
<title>Un document en HTML</title>
<meta http-equiv="Content-type"
content="text/html; charset=iso-8859-1" />

<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
```

```
<meta name="DC.title" lang="fr" content="Un document en HTML" />
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2003-04-03" />
<meta name="DC.subject" lang="fr" content="HTML, document, Dublin Core" />
<meta name="DC.language" scheme="DCTERMS.RFC4646" content="fr-FR" />
<meta name="DC.description" lang="fr"
  content="Mon premier document HTML avec métadonnées" />
<link rel="DC.source" href="urn:ISBN:978-1-2345-6789-X" />
</head>
<body>
  ...
</body>
</html>
```