

La minute du PSIR

12 mai 2021 - # 21

Les métadonnées

Cette semaine le PSIR vous propose de faire un point sur l'utilisation des métadonnées. La définition que nous donne Wikipédia : il s'agit d'une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique).

En quoi cela peut être utile ?

Les données que nous manipulons sont très diverses. Elles peuvent avoir de multiples supports (papier, fichier, base de données, ...), être décrites dans diverses langues, être stockées sous divers formats (pdf, xml, doc, jpeg, tiff, ...). Cette hétérogénéité complexifie grandement les recherches nous permettant de les atteindre, de les partager, ...

C'est pourquoi nous utilisons des métadonnées **normées** qui les décrivent et sur lesquelles nous pouvons **élaborer des outils de recherche**.

L'un des premiers projets basés sur les métadonnées est **OAI (*Open Archives Initiative*)** qui a démarré en 2001 avec la naissance du mouvement promouvant l'auto-archivage des publications.

LA BNF nous propose une classification des métadonnées en 3 catégories :

- les métadonnées descriptives
- les métadonnées structurelles,
- les métadonnées administratives

L'OAI utilise un set de métadonnées normé le « **Dublin Core** » :

Des applications appelées « moissonneurs » permettent d'agréger ces données descriptives au sein d'entrepôts et permettent donc **l'interopérabilité de données** construites et hébergées dans des environnements différents.

Le moteur de recherche ISIDORE permettant l'accès aux données numériques des sciences humaines et sociales (SHS) est un exemple d'outil utilisant les métadonnées et pour les curieux nous vous proposons la lecture de l'article <https://documentation.huma-num.fr/isidore/#comment-faire-pour-que-des-donnees-soient-referencees-par-isidore> qui montre la richesse des méthodes employées et la puissance d'une bonne utilisation des métadonnées et des standards et également ses limites pour valoriser vos données.

Nous signalons **également le cas particulier des images** où les métadonnées peuvent être intégrées au sein même des fichiers (lorsque le format utilisé est jpeg, tiff et png) en **utilisant les normes IPTC** (norme en anglais).

Des outils Open source comme ExifTools permettent de manipuler ces métadonnées.

NIVEAU DE DIFFICULTÉ



Données

PERSONNES CONCERNÉES



PDF

MINUTE_PSIR_#21. Pdf 307Ko.



université
Lumière
LYON 2

