

## Structure de la séance

1. Indexation du savoir : *retour sur les prémices de l'indexation des connaissances qui ont participé d'une culture d'édition numérique*
2. Identités du livre : *présentation des différents systèmes d'identification du livre*
3. La structuration et la donnée : *présentation sur l'organisation des données numériques pour l'édition du livre et de l'impact du traitement de certaines données*
4. Partie pratique : Leaf Writer : *édition avec l'éditeur XML et RDF*

## 1. Indexation du savoir

### Le projet *Mundaneum*

- Centre d'archives mondial de Paul Otlet et Henri La Fontaine, 1898
- Réunion de l'OIB et du RBU de 1895 (Office International de Bibliographie et Répertoire Bibliographique Universel)
- Organisation des connaissances : corpus de 18 millions de fiches bibliographiques selon un système de classement (CDU) et de disposition physique (meubles-fichiers)

### Philosophie du projet

- Rêve de bibliothèque universelle : Bibliothèque de Babel, Projet Gutenberg, Wikipédia, etc.
- Visée pacifiste de science ouverte (voir *Essai de bibliographie de la paix*, 1891)
- Structuration et diffusion à l'international des connaissances

Rassembler et classer les savoirs du monde relève d'un projet global qui se construit dans le contexte spécifique des mouvements pour la paix, de la fondation de la Société des Nations (SDN) et des deux guerres mondiales. (Schafer, 2013)

### Contribution aux standards

- Classification décimale universelle (CDU, 1905) construit à partir de la Classification décimale de Dewey, 1876 (CDD)
- Norme des fiches bibliographiques (125 x 75 mm)

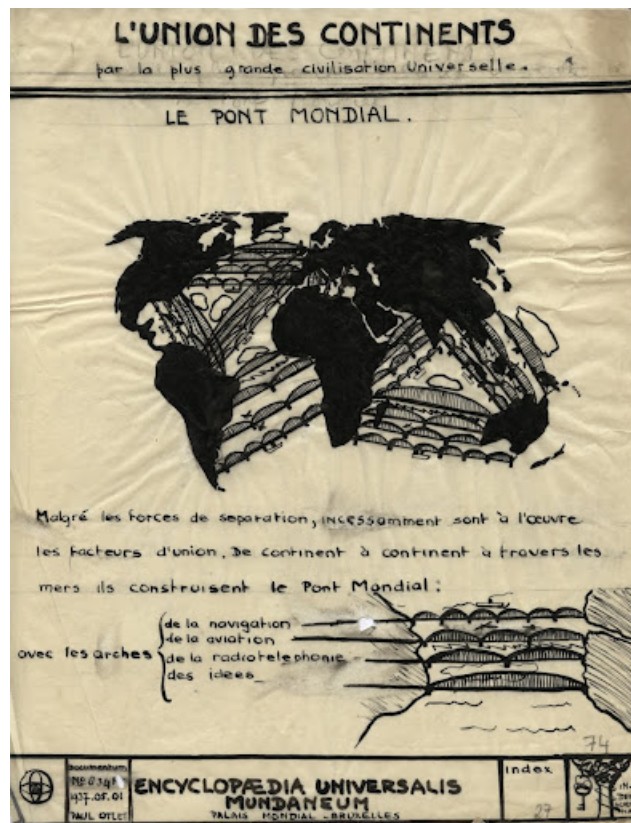
### Préfiguration de la culture numérique : *établissement d'un système d'organisation et d'échange d'information*

#### Fonctionnement du RBU :

- consultation sur place libre
- demande de recherche (préfiguration des navigateurs)

#### Accès à l'information multiple :

- répertoire par auteur (*onomastique*)
- répertoire par thématiques
- répertoire par titres de périodique et de livre



## Le pont mondial

### Suites du projet

- [HyperOtlet](#) : encyclopédie documentaire sur l'histoire de la documentation
- [OtletoSphère](#) : cartographie relationnelle des travaux d'Otlet
- [Cosma](#) : outil de visualisation des liens (Perret 2022) - [exemplaire d'un glossaire](#)

## 2. Identités du livre

**Identifiant pérenne : chaîne de caractères alphanumériques pour renseigner la ressource et/ou son emplacement.**

### Dépôt légal

- Loi sur Bibliothèques et Archives nationales du Québec : dépôt de 2 exemplaires par l'éditeur dans les 7 jours après publication

## ISBN

- Numéro international normalisé du livre (1970) pour identifier de manière unique chaque édition de chaque livre publié

*ISBN est pour les monographies; pour les publications en série : ISSN et e-ISSN pour la presse numérique*

## Composition

1. groupe linguistique : 2 pour les espaces francophones
  2. numéro d'éditeur : court si éditeur à grande publication annuelle (07 pour Gallimard ; 925226 pour Poètes de brousse)
  3. numéro de publication attribué par l'éditeur
  4. clef de contrôle
- 3 chiffres attribués de *Bookland* (978) pour les ISBN-13

## Fonctionnement

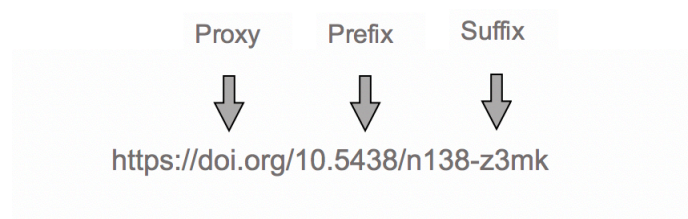
- non-obligatoire au Canada mais nécessaire pour le CIP
- recommandé pour l'auto-édition
- disponible via des agences d'ISBN (comme la BAnQ) avec un délai d'un mois

## URI

- *Uniform Resource Identifier* issue de la culture du World Wide Web
- deux types : URL et URN (URL : type *locator* pour l'emplacement et URN : type *name* pour le nom)

## DOI

- *Digital Object Identifier* : alternative au URI pour régler le problème des URLs changeants initiée au nom de l'Association of American Publishers et développée par l'agence d'enregistrement des DOI : Crossref



## Composition

### Fonctionnement

- Organisation autorisée comme CrossRef pour les publications académiques ou DataCite pour tout type de ressources
- Frais d'abonnement variable (250-300 \$/an et 1 \$/DOI article ; 2-3\$/DOI livre)

### ARK

- *Archival Resource Key* développée par la California Digital Library

**Composition** [https://NMA/ark:/NAAN/Nom\[Qualificatif\]](https://NMA/ark:/NAAN/Nom[Qualificatif])

1. *Name Assigning Authority Number* : obligatoire - identifiant unique de l'organisation
2. *Name Mapping Authority* : optionnel
3. Qualificatif : optionnel - variante de langues, versions, formats

exemple : <https://gallica.bnf.fr/ark:/12148/bpt6k1523583v/f227.item>

### Autres systèmes

- ASIN ou Amazon Standard Identification Number : utilisé et produit par Amazon
- LCCN ou Library of Congress Control Number par la Bibliothèque du Congrès pour catalogage
- ISMN ou International Standard Music Number pour les partitions musicales, livres musicaux

### Choix du système

- certains obligatoires comme le dépôt légal
- besoins de centralisation : DOI, parce que géré par des organismes accrédités, est plus centralisé ; ARK est libre, auto-hébergé, moins centralisé
- type de services : ARK pour institutions culturelles (archives) ; DOI pour publications scientifiques
- prévisions : ARK souple et extensible / DOI rigide et immobile

## 3. La structuration et la donnée

### Notice bibliographique

- description de manière unique un objet et le lieu de consultation (titre, auteur, date, ISBN, cote, etc.)
- normalisation selon les normes ISO

### RDF

- *Resource Description Framework* : modèle de données pour représenter/décrire les ressources sur le Web
- Fonctionnement sur le triplet (1 triplet = un sujet + prédicat + objet)

- Modèle flexible, extensible et interopérable (parfois écrit en XML)

Exemple : *Le ciel de Bay City* de C. Mavrikakis

- Sujet = ressource à décrire

```
<http: example.org="" books="" 1=""></http:>
```

- Prédicat = type de propriété

```
<http: purl.org="" dc="" elements="" 1.1="" title=""></http:>
```

- Objet = valeur de la propriété

```
"Le ciel de Bay City"
```

Résultat :

```
<http: example.org="" books="" 1="">  
<http: purl.org="" dc="" elements="" 1.1="" title="">  
"Le ciel de Bay City".
```

### Microdonnée

- Fonctionnalité du HTML5 pour encoder les métadonnées des pages web

### OAI-PMH

- *Open Archives Initiative Protocol for Metadata Harvesting* par Open Archive Initiative
- Protocole informatique pour récolter les métadonnées (ouvert, flexible, interopérable)
- Améliorer les informations et les recherches dans des systèmes de documentation

### Dublin Core (dc)

- Vocabulaire du Web sémantique pour description des documents numériques : par la Dublin Core Metadata Initiative (DCIM)
- Utilisé par les standards de structuration des métadonnées (Microdonnée, RDF, OAI-PMH, XML, etc.)

### Composition

1. *Dublin Core element set* : 15 éléments métadonnées (pour le contenu : titre, description, source, langue, relation, couverture; pour la propriété intellectuelle : créateur, contributeur, éditeur, gestion des droits; pour l'instanciation : date, type, format, identifiant)
2. *Dublin Core metadata terms* : éléments additionnels (catégories : résumé, format de fichier, collection, etc.; qualificatifs : nom de famille du créateur, etc.)

## Réflexion et ouverture sur la donnée

Croire que les données sont intrinsèquement quantitatives — évidentes, neutres sur le plan des valeurs et indépendantes de l'observateur — exclut la possibilité de les concevoir comme qualitatives, co-dépendamment constituées — en d'autres termes, de reconnaître que data sont des captas [...] Je suggère que nous repensions fondamentalement les données comme des captas en termes d'ambiguïté plutôt que de certitude, et que nous trouvions des moyens d'exprimer graphiquement la complexité de l'interprétation. (Drucker, 2011)

## La recommandation et la donnée

[...] les données qui peuvent être identifiées par un algorithme sont, par nature, des données standardisées, c'est-à-dire des données qui sont formatées et prêtes à être utilisées par l'algorithme, et ce, sans intervention humaine. Cela implique, par exemple, que les données erronées, ou absentes, d'un titre de livre sur une plateforme en ligne, feront en sorte que le titre en question ne sera pas sélectionné par l'algorithme. (Labbé, 2015)

## Les méthodes de recommandation

- *Collaborative Filtering* : regrouper les profils selon des patterns de préférence
- *Content-based Filtering* : utilisation des métadonnées communes pour recommandation

## 4. Partie pratique : Leaf Writer

### Leaf Writer

- Éditeur de texte en ligne de documents sensibles à des schémas (comme le XML-TEI)
- Ouvert et en libre accès ([répertoire Git de l'outil](#))
- Outil développé dans le cadre du portail [LINCS](#) (Guelph, Brown, Marin and Stacey, 2020)

### Exercice d'édition enrichie de *Mrs Dalloway in Bond Street*

1. Ouvrir échantillons (textes déjà préparés en XML)
2. Double-clic sur Virginia Woolf, "Mrs Dalloway in Bond Street" (prose)

### Décrypter l'interface

- 2 vues : vue de montage pour l'édition et vue de visualisation pour le rendu
- 3 espaces : centre : texte et édition avec outils ; gauche : table des matières, balisage, entités encodées ; droite : rendus xml modifiable, rendu visuel, schéma de validation continue
- Sauvegarde et Export (HTML, XML et Md.)

### Identifier les entités nommées

- Tags (balises TEI-XML éligibles)
- Personne (réelle ou non)
- Lieu
- Organisation ou société
- Titre ou texte
- Nom d'usage, étiquette (expression régulière ou redondante)
- Citation
- Date

### Identifier les interventions sur le texte

- Annotation (remarque, note de recherche)
- Correction apportée
- Mot-clef (thématique ou marqueur redondant)
- Lien ou URL
- Traduction : doit prendre en compte une div/section entière

*conseil : afficher les balises.*

### À éditer

1. 3 Personne : 2 fictionnelles : Mrs Dalloway et Shakespeare (avec person type + certainty) ; une réelle : Virginia Woolf (person type + certainty + role)

*Conseil : si vous ne parvenez pas à les trouver via le module, allez les chercher directement sur Wikidata ou essayez plusieurs graphies (minuscule)*

2. 4 Lieu : 1 ville, 1 quartier, 1 rue, 1 monument
3. 1 Organisation ou société : *the Parliament*
4. 2 Titre ou texte : le titre du texte (certainty + level) et les sonnets de Shakespeare (certainty + level)
5. 6 Tags : 2 balises forename/prénom ; 1 balise surname/surnom (Milly) ; 1 balise abbréviation ; 2 balises roleName/statut (King/Queen)
6. 1 Annotation libre
7. 1 Correction apportée (BY par by)
8. 1 Traduction

### À observer

- la validation tout au long de l'édition
- le RAW XML pour voir le changement/ajout de balises
- la liste des entités
- la structure du balisage

Pour aller plus loin : [Documentation de Leaf-Writer](#)

## Lexique

- *API (Application Programming Interface)* : ensemble de règles et de protocoles permettant à des applications de communiquer entre elles, facilitant l'intégration de fonctionnalités.
- *CDD (Classification Décimale de Dewey)* : système de classification bibliographique.
- *CDU (Classification Décimale Universelle)* : système de classification bibliographique issu du CDD.
- *CIP (Cataloging in Publication)* : système de catalogage utilisé pour les livres, qui fournit des informations bibliographiques détaillées sur un ouvrage avant sa publication.
- *dépôt légal* : obligation pour les éditeurs de déposer un exemplaire de chaque publication auprès d'une bibliothèque nationale, garantissant la conservation du patrimoine écrit.
- *DOI (Digital Object Identifier)* : identifiant unique attribué à un document numérique, permettant de le localiser et de le citer de manière permanente.
- *Dublin Core* : vocabulaire du Web sémantique utilisé pour structurer les données dans les standards de métadonnées comme RDF.
- *ISBN (International Standard Book Number)* : numéro unique attribué à chaque livre publié, facilitant son identification et sa distribution dans le commerce.
- *ISSN (International Standard Serial Number)* : numéro unique attribué aux publications en série (revues, magazines).
- *RBU (Référentiel Bibliographique Universel)* : système de classification et de gestion des références bibliographiques pour standardiser la citation et l'accès aux documents.
- *URI (Uniform Resource Identifier)* : identifiant qui permet de nommer une ressource sur Internet, incluant à la fois les URL et les URN.
- *URL (Uniform Resource Locator)* : type d'URI qui spécifie l'adresse d'une ressource sur Internet, permettant d'accéder à celle-ci via un navigateur.
- *URN (Uniform Resource Name)* : type d'URI qui fournit un identifiant unique pour une ressource, sans spécifier son emplacement.

## Ressources

[DOI](#)

[HyperOtlet](#)

[LINCS](#)



## Bibliographie

- Bachimont, Bruno. 2007. « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation ». In *L'indexation multimédia : description et recherche automatiques*. IC2. Paris : Hermès science publications-Lavoisier.
- Bourassa, Renée. 2021. *Le livre en contexte numérique : Un défi de design*. ARCANES, Québec. <http://livre-defi-design.arcanes.ca/abecedaire-hyperotlet/>.
- Drucker, Johanna. 2010. *Data as Capta*. First Edition. Druckwerk.
- Drucker, Johanna. 2011. « Humanities Approaches to Graphical Display ». *Digital Humanities Quarterly* 5 (1). <http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html#p1>.
- « [https://www.banq.qc.ca/sites/default/files/2022-09/guide\\_technique\\_extranet.pdf](https://www.banq.qc.ca/sites/default/files/2022-09/guide_technique_extranet.pdf) ». s. d. Consulté le 15 octobre 2024. [https://www.banq.qc.ca/sites/default/files/2022-09/guide\\_technique\\_extranet.pdf](https://www.banq.qc.ca/sites/default/files/2022-09/guide_technique_extranet.pdf).
- Irosoft, architecture de gestion de l'information législative-legal information management system. s. d. « - Loi sur Bibliothèque et Archives nationales du Québec ». Consulté le 15 octobre 2024. <https://www.legisquebec.gouv.qc.ca/fr/document/lc/B-1.2>.
- Jones, Steven E. 2016. *Roberto Busa, S.J. and the emergence of humanities computing : the priest and the punched cards*. New York : Routledge.
- La Fontaine, Henri. 1891. *Essai de bibliographie de la paix*. Bruxelles : T. Lombaerts.
- Labbé, Stéphane. 2015. « La médiation du livre à l'ère du numérique : l'enjeu des algorithmes de recommandation ». *Documentation et bibliothèques* 61 (1) :15-21. <https://doi.org/10.7202/1029000ar>.
- Lupovici, Catherine. 1998. « Le digital Object Identifier ». *Bulletin des bibliothèques de France (BBF)* 3 (janvier). <https://bbf.enssib.fr/consulter/bbf-1998-03-0049-007>.
- Schafer, Valérie. 2013. « Le Mundaneum, un patrimoine inclassable ». *Hermès, La Revue* 66 (2) :155-59. <https://doi.org/10.4267/2042/51569>.
- Schafer, Valérie, et Bernard Tuy. 2013. *Dans les coulisses de l'Internet : RENATER, 20 ans de technologie, d'enseignement et de recherche*. Le sens de la recherche. Paris [Meudon] : A. Colin Comité pour l'histoire du CNRS.
- Vitali-Rosati, Marcello, et Margot Mellet. 2021. « Éditorialiser l'Anthologie grecque. L'API comme livre numérique ». In *Le livre en contexte numérique : Un défi de design*. ARCANES, Québec. <http://livre-defi-design.arcanes.ca/editorialiser-anthologie-grecque/>.