

1. La structuration de l'information

Objectifs:

- Montrer comment les professions traitant de l'information structurent (ou non) leur information.
- Montrer comment la structuration de l'information permet de la gérer
- Montrer comment la globalisation se manifeste par un processus complexe de partage de standards susceptible de supprimer ou d'atténuer les différences entre métiers dans la structuration de l'information.

| | |
|--|-----------|
| 1 L'âge des professions | 2 |
| 1.1 Les bibliothèques | 2 |
| 1.1.1 ISBD (international standard bibliographic description) | 2 |
| 1.1.2 MARC (Machine Readable Cataloguing) | 3 |
| 1.1.3 Un schéma émergent : les FRBR | 4 |
| 1.1.4 Accès par les contenus | 5 |
| 1.1.5 Les données bibliographiques : entrée / stockage / sortie | 6 |
| 1.1.6 Schéma sommaire d'une base de donnée bibliographique | 6 |
| 1.2 L'impression, l'édition | 9 |
| 1.2.1 A l'origine : le flou | 9 |
| 1.2.2 Structuration physique, structuration logique | 9 |
| 1.2.3 Les langages à balise (<i>Mark-up Languages</i>) : SGML, XML | 10 |
| 1.3 Le Web | 11 |
| 1.3.1 HTML | 11 |
| 1.3.2 XML | 11 |
| 1.3.3 Les feuilles de style : CSS | 11 |
| 1.4 L'auteur | 11 |
| 1.4.1 Indigence primitive | 11 |
| 1.4.2 Feuilles de style | 11 |
| 1.5 Une chaîne numérique complète ? | 12 |
| 2 Retour sur la logique de la structuration..... | 12 |
| 2.1 Structurer pourquoi ? | 12 |
| 2.2 Les logiques de structuration | 12 |
| 2.3 Ne pas confondre... .. | 12 |
| 2.4 Les métadonnées..... | 13 |
| 3 La globalisation ou l'âge des standards | 13 |
| Conclusion..... | 14 |

1 L'âge des professions

Pour être exploitée, diffusée, réutilisée, l'information doit être structurée, non dans l'ordre du discours (ce n'est pas le propos ici) mais dans celui de ses éléments.

Mais cette structuration ne s'est pas construite par objet (ex.: les données bibliographiques) mais par profession (l'économie, la sociologie et l'histoire peuvent expliquer cet apparent paradoxe).

1.1 Les bibliothèques

Objectifs :

- Décrire : produire de l'information sur l'information (information secondaire). C'est le **catalogage**.
- Accéder : donner des clés pour accéder à l'information secondaire. C'est l'**indexation**.

1.1.1 ISBD (international standard bibliographic description)

Ce schéma de catalogage précède l'informatisation des catalogues de bibliothèque mais demeure une référence.

8 zones subdivisées en sous-zones. Les séparateurs entre zones et sous-zones sont des signes de ponctuations conventionnels.

| Zone | Ponctuation | Principales sous-zones |
|---|-------------|--|
| 1. Titre et mention de responsabilité | | Titre propre |
| | : | Complément de titre |
| | / | Première mention de responsabilité |
| | ; | Autres mentions de responsabilité |
| 2. Edition | | Mention de l'édition |
| 3. Zone particulière à certains types de documents (publications en série, documents cartographiques, musique imprimée et documents électroniques) | | |
| 4. Adresse bibliographique | | Lieu de publication |
| | : | Nom de l'éditeur |
| | , | Année de publication |
| 5. Collation | | Nombre de pages ou d'autres unités (CD...) |
| | : | Mention d'illustration |
| | ; | Format |
| 6. Collection | (| Mention de collection |
| | , | ISSN |
| | ; | Numéro du volume dans la collection |
| |) | |
| 7. Notes | | |
| 8. ISBN et prix | | |

Exemple d'une notice bibliographique selon l'ISBD :

Le catalogage : méthode et pratiques. 1,
Monographies et publications en série / par Isabelle
Dussert-Carbone et Marie-Renée Cazabon. - Nouv.
éd. - Paris : Éd. du Cercle de la librairie, 1991. -
479 p. : ill. ; 24 cm. - (Collection Bibliothèques,
ISSN 0184-0886).

Bibliogr. p. 465-476. - ISBN 2-7654-0458-5



1.1.2 MARC (Machine Readable Cataloguing)

Famille de formats née aux Etats-Unis à la fin des années 1960 et toujours en vigueur.

Les formats MARC, dont il existe plusieurs variantes nationales ou internationales, permettent aux bibliothèques, au lieu de cataloguer chaque ouvrage, de récupérer les données chez un fournisseur ou dans un système coopératif.

Exemple présenté ici : **UNIMARC** (format universel très utilisé en France)

10 **blocs** comportant chacun des **zones** divisées en **sous-zones**, chacune étant **obligatoire** ou **facultative**, **répétable** ou **non répétable** :

| | |
|------------|--|
| 0XX | Numéros d'identification (zone 8 de l'ISBD) |
| 1XX | Informations codées (date de création de la notice, date de publication du document, type de document, langue du document, pays de publication, etc.) |
| 2XX | Informations descriptives (le " pavé ISBD ", zones 1 à 6) |
| 3XX | Notes |
| 4XX | Liens éventuels avec d'autres notices bibliographiques |
| 5XX | Titres associés (toutes formes du titre différant du titre propre figurant sur la page de titre et que l'on peut utiliser comme points d'accès, comme le titre uniforme) |
| 6XX | Analyse du contenu (accès sujet, classifications, etc.) |
| 7XX | Responsabilité intellectuelle (personnes physiques, collectivités auteurs, congrès) |
| 8XX | Usage international (organisme créateur de la notice, lien à une ressource électronique, etc.), |
| 9XX | Données locales (espace d'usage libre, non prescrit par le format, pouvant faire l'objet de recommandations nationales) |

Exemple de notice bibliographique en Unimarc, présentée pour plus de clarté sous forme de tableau (en bleu les éléments de l'ISBD, en vert les accès) :

| Guide | | 008866cam0022002653004500 | Guide |
|-------|----|--|--------------------------|
| 001 | | 01-0088058- | Numéros d'identification |
| 005 | | 20000129163844.0- | |
| 010 | 00 | \$a2-7654-0551-4\$bbr.\$d2950F- | |
| 100 | 00 | \$a19920512d19940000m00y1 a01030000 - | Données codées |
| 101 | 00 | \$afre- | |
| 102 | 00 | \$aFR- | |
| 105 | 00 | \$aayyyzzzz00 z - | |
| 106 | 00 | \$ar- | |
| 200 | 10 | \$a^Le %o catalogage \$eméthode et pratiques\$h1 Monographies et publications en série \$fpar Isabelle Dussert-Carbone et Marie-Renée Cazabon - | |
| 205 | 00 | \$aNouv. éd.- | |
| 210 | 00 | \$a Paris \$c Cercle de la Librairie \$d1994- | |
| 215 | 00 | \$a 493 p. \$c ill. \$d 24 cm - | |
| 225 | 20 | \$a Collection Bibliothèques \$x 0184-0886 -- | |
| 300 | 00 | \$a Bibliogr. p. 465-476 - | |
| 410 | 01 | \$aBibliothèques- | Liens |
| 461 | 00 | \$tLe catalogage\$v1- | |
| 606 | 00 | \$a Description bibliographique \$y France \$x Guides, manuels, etc. \$2Rameau- | Accès sujet |
| 676 | 00 | \$a 025.32 \$v20- | |
| 700 | 01 | \$a Dussert-Carbone \$b Isabelle \$4070- | Accès responsabilité |
| 701 | 01 | \$a Cazabon \$b Marie-Renée \$4070- | |

1.1.3 Un schéma émergent : les FRBR

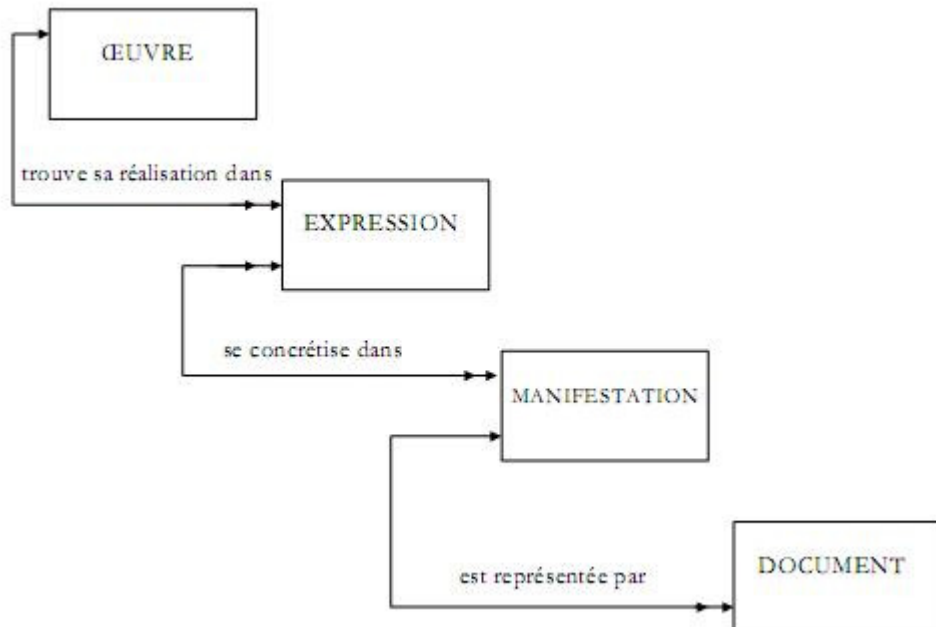
Une nouvelle structuration de l'information bibliographique est en cours de constitution à travers le monde à partir de ce qu'on nomme le schéma FRBR :

FRBR = Fonctional Requirements for Bibliographic records

Ce schéma identifie des entités entre lesquels il établit des relations

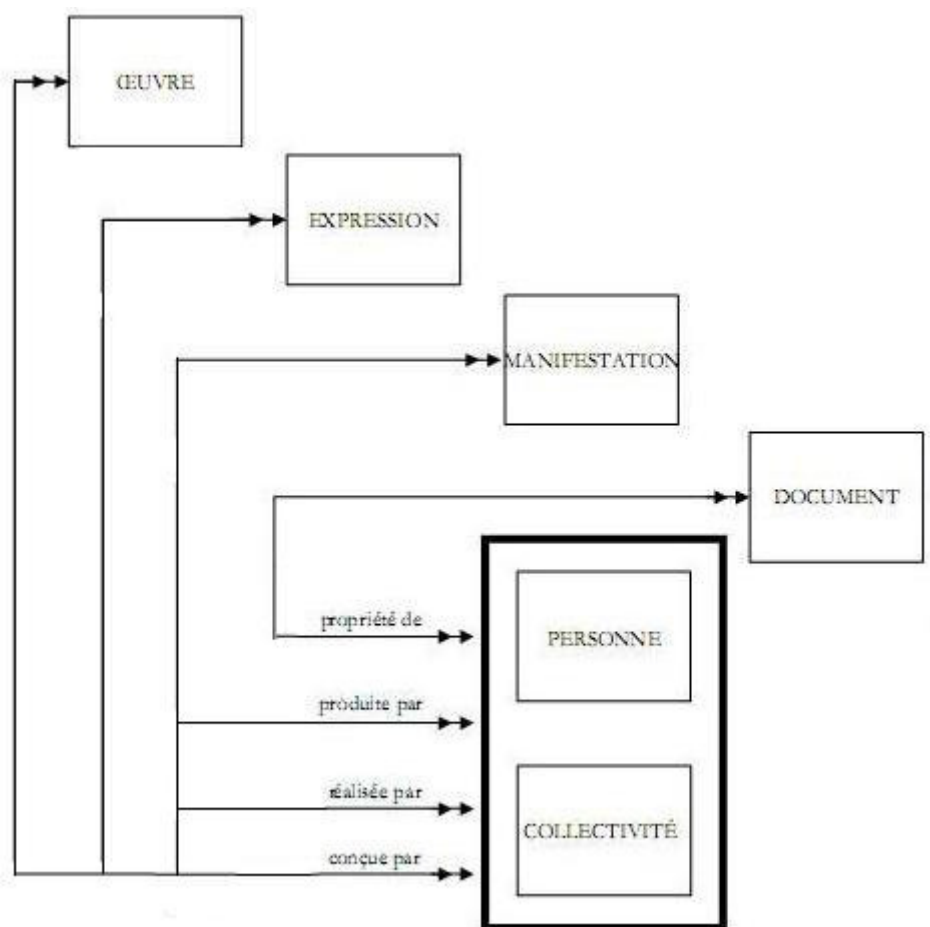
- Entités du 1^{er} groupe :

- œuvre
- expression
- manifestation
- document

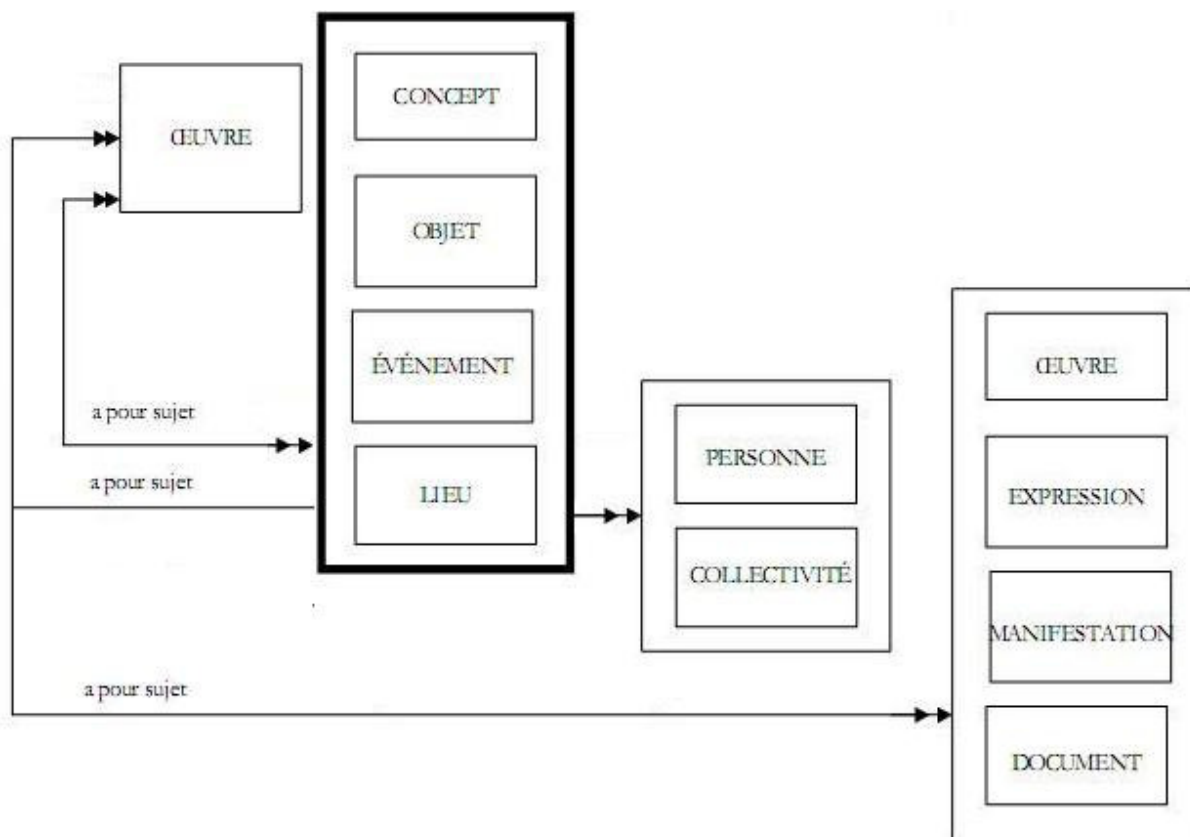


- Entités du 2^e groupe :

- personnes
- collectivités



- Entités du 3^e groupe :
 - concepts
 - objets
 - événements
 - lieux



1.1.4 Accès par les contenus

- Les langages documentaires peuvent comporter à la fois un vocabulaire et une syntaxe exprimant de façon conventionnelle le contenu d'un document. Les documentalistes parlent de mots clés et les bibliothécaires de mots matière ou d'autorités matière. Dans l'exemple ci-dessus emprunté au langage documentaire Rameau (Répertoire d'autorités matière alphabétique et encyclopédique unifié) en usage dans les bibliothèques françaises :

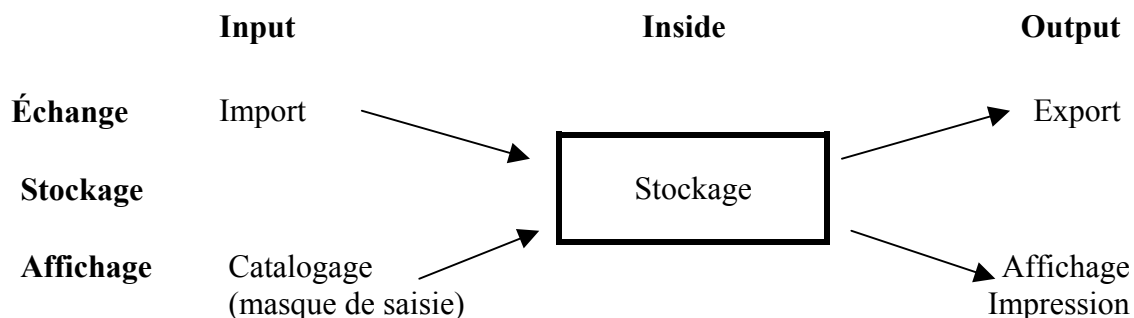
Description bibliographique – France -- Guides, manuels, etc.

- Les nuages de tags c'est (en gros) à la même chose



- Au-delà des langages documentaire, la tendance est à un **accès par tous les mots** d'un champ ou tous les mots de tous les champs. Par exemple : accès par les mots du résumé d'un ouvrage, qui peut (comme dans la base Electre) être contenu dans la notices bibliographique.

1.1.5 Les données bibliographiques : entrée / stockage / sortie



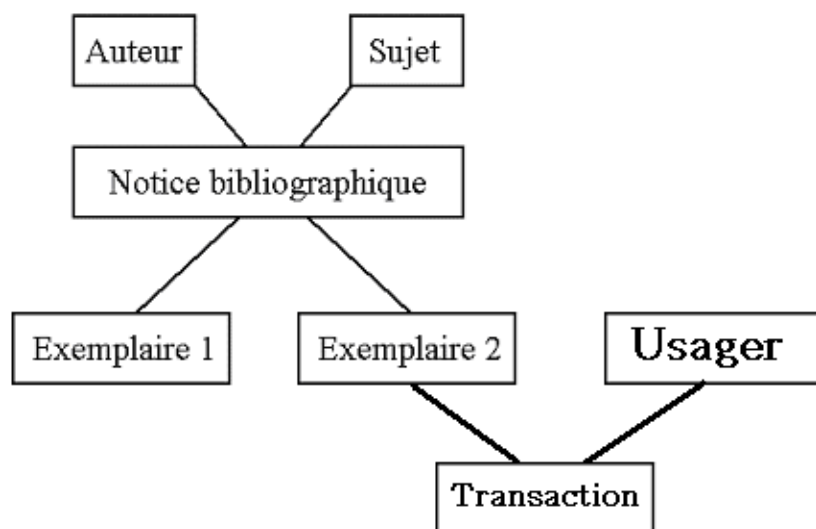
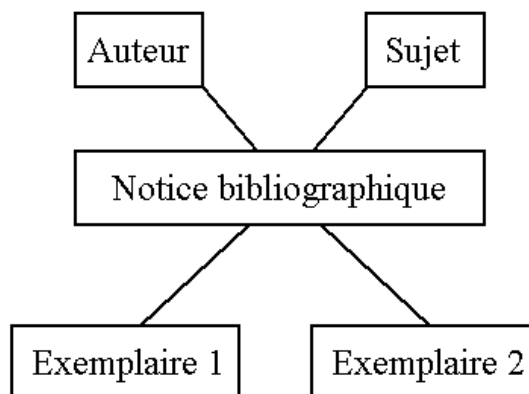
1.1.6 Schéma sommaire d'une base de donnée bibliographique

Schémas logiques

La notice bibliographique, comprenant au moins les éléments descriptifs du document, est liée à la description des exemplaires physiques possédés, dont le système va gérer le prêt éventuel.

Les éléments normalisés (par exemple les auteurs et les sujets) permettent un accès contrôlé par un vocabulaire et une syntaxe.

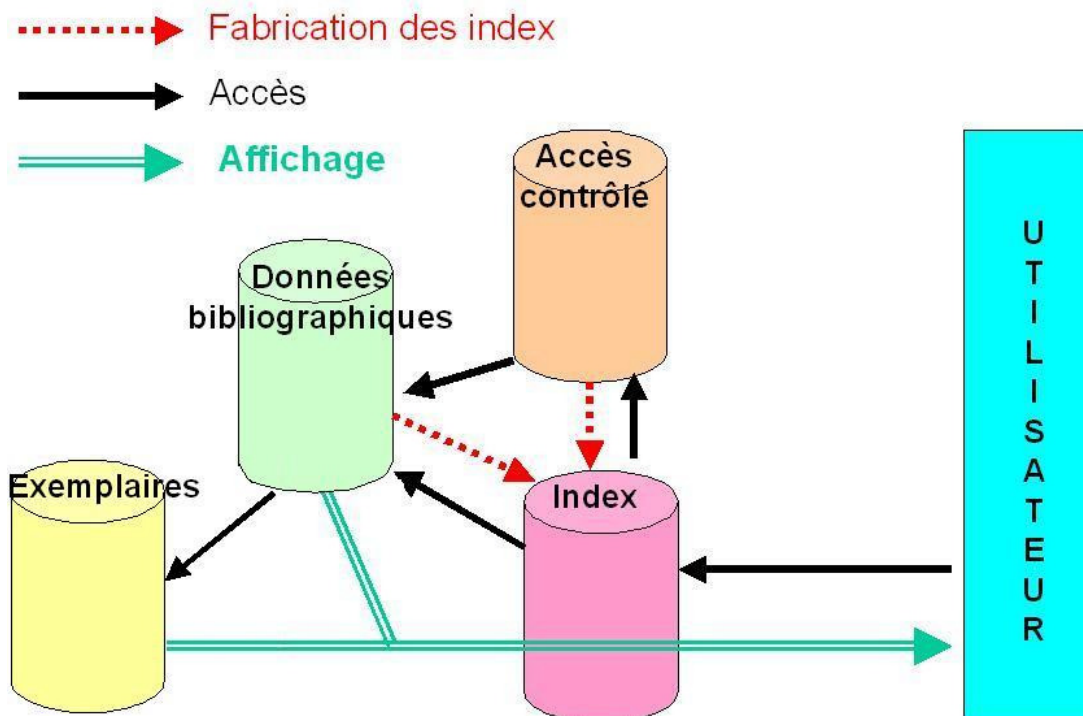
Ils peuvent être décomposés entre autorité (normalisée) et vedette (point d'accès éventuellement composite).



La gestion du prêt permet de lier le fichier des emprunteurs à celui des exemplaires via celui de des transactions.

Schéma informatique

Le système fabrique des index, fichiers permettant un accès rapide aux données. Ces index sont nourris par les données saisies pour les accès contrôlés par un langage documentaire, mais aussi, le cas échéant, par les mots contenus dans les données bibliographiques. Ils sont généraux (n'importe quel mots) ou spécifiques à certains champs (pour permettre par exemple une recherche par auteur). On dit que les champs des notices bibliographiques nourrissant les index sont **indexés**.



Exemple d'une recherche sur le catalogue collectif valdoisien RéVODOC :
<http://revodoc.valdoise.fr>

| Recherche | Recherche avancée | Ma sélection | Mon Compte | Déconnexion |
|--|-------------------|--------------|--|-------------|
| <p>Recherche avancée</p> <p>titre = <input type="text"/></p> <p>auteur = <input type="text"/></p> <p>sujet = <input type="text"/></p> <p>Publié entre <input type="text"/> et <input type="text"/></p> <p><input type="button" value="Rechercher"/></p> | | | | |
| <p>TYPE DE DOCUMENT :</p> <p><input checked="" type="radio"/> tous</p> <p><input type="radio"/> imprimé</p> <p><input type="radio"/> enregistrement sonore</p> <p><input type="radio"/> cédérom</p> <p><input type="radio"/> partition</p> <p><input type="radio"/> vidéo</p> | | | | |
| <p>TRIÉ PAR :</p> <p><input checked="" type="radio"/> auteur</p> <p><input type="radio"/> titre</p> <p><input type="radio"/> éditeur</p> <p><input type="radio"/> collection</p> <p><input type="radio"/> date</p> <p><input type="radio"/> type de document</p> | | | | |
| <p>Dans le catalogue collectif RéVODOC</p> | | | | |
| <p>DIVERS</p> <p>Tout cocher / Tout décocher</p> <p><input checked="" type="checkbox"/> Argenteuil</p> <p><input checked="" type="checkbox"/> Beauchamp</p> <p><input checked="" type="checkbox"/> Beaumont</p> <p><input checked="" type="checkbox"/> Bessancourt</p> <p><input checked="" type="checkbox"/> Bibliothèque départementale</p> <p><input checked="" type="checkbox"/> Goussainville</p> <p><input checked="" type="checkbox"/> Luzarches</p> <p><input checked="" type="checkbox"/> Montigny-lès-Cormeilles</p> <p><input checked="" type="checkbox"/> Saint-Leu-la-Forêt</p> <p><input checked="" type="checkbox"/> Sannois</p> | | | <p>FONDS SPÉCIALISÉS</p> <p>Tout cocher / Tout décocher</p> <p><input checked="" type="checkbox"/> Archives départementales</p> <p><input checked="" type="checkbox"/> CDDP</p> <p><input checked="" type="checkbox"/> Conservatoire à rayonnement régional</p> <p><input checked="" type="checkbox"/> Livres anciens</p> <p><input checked="" type="checkbox"/> Services archéologiques</p> <p><input checked="" type="checkbox"/> Université / ENSEA</p> | |
| | | | <p>CERGY-PONTOISE</p> <p>Tout cocher / Tout décocher</p> <p><input checked="" type="checkbox"/> Bibliothèque d'étude et d'information</p> <p><input checked="" type="checkbox"/> Cergy : L'Astrolabe</p> <p><input checked="" type="checkbox"/> Cergy : L'Horloge</p> <p><input checked="" type="checkbox"/> Eragny-sur-Oise</p> <p><input checked="" type="checkbox"/> Jouy-le-Moutier</p> <p><input checked="" type="checkbox"/> Menucourt</p> <p><input checked="" type="checkbox"/> Pontoise</p> <p><input checked="" type="checkbox"/> Saint-Ouen-L'Aumône</p> <p><input checked="" type="checkbox"/> Vauréal : Les Dames Gilles</p> | |

[Recherche](#)
[Recherche avancée](#)
[Ma sélection](#)
[Mon Compte](#)
[Déconnexion](#)

Recherche avancée

Résultat : 1 - 10 sur 130 - Précédent [Suivant](#)

trier par

 [Bons baisers de New York](#)

AUTEUR : Auster,
EDITEUR : Flammarion, 2003

 [Art de la faim : conversations avec Paul Auster](#)

AUTEUR : Auster, Paul
EDITEUR : Actes sud, 1991

 [Carnet rouge](#)

AUTEUR : Auster, Paul
EDITEUR : Actes Sud, 1993

 [Carnet rouge suivi de](#)

AUTEUR : Auster, Paul
EDITEUR : Actes Sud, 1994
COLLECTION : Babel

 [Constat d'accident : et autres textes](#)

AUTEUR : Auster, Paul
EDITEUR : Actes Sud, 2003

[Recherche](#)
[Recherche avancée](#)
[Ma sélection](#)
[Mon Compte](#)
[Déconnexion](#)

Bons baisers de New York 

Affichage :

| | |
|---------------|--|
| titre : | Bons baisers de New York Précédé de L'art de l'inquiétude Spiegelm |
| auteur : | auster FR_Elec |
| éditeur : | Flammarion , 2003 |
| sujet : | Spiegelm |
| description : | ill. en coul., 27 cm |
| note : | L'art d |
| identifiant : | ISBN : 9782080685612 |
| source : | Sal |



[Retour à la liste](#)
[Ajouter à ma sélection](#)
[Réserver](#)

Ce document se trouve dans les bibliothèques suivantes :

[Cergy : L'Astrolabe](#)
Cote : 741.5 SPI (prêt entre bibliothèques sous réserve d'acceptation)
Cote : B 741.5 SPI (prêt entre bibliothèques sous réserve d'acceptation)

[Pontoise](#)
Cote : 741.5 SPI (prêt entre bibliothèques sous réserve d'acceptation)

[Argenteuil](#)
Cote : BD SPI (prêt entre bibliothèques sous réserve d'acceptation)

1.2 L'impression, l'édition

Objectifs :

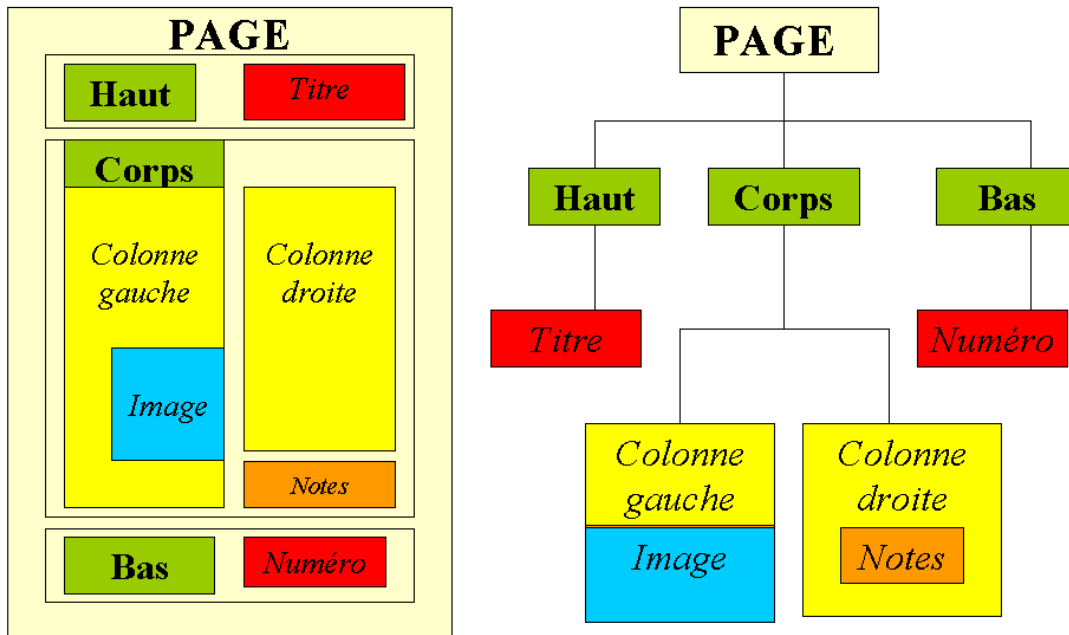
- Imprimer sur papier.
- Publier sous diverses formes (cédéroms, internet...).

1.2.1 A l'origine : le flou

En général, les éditeurs maîtrisent peu leurs références. Cette situation perdure largement aujourd'hui.

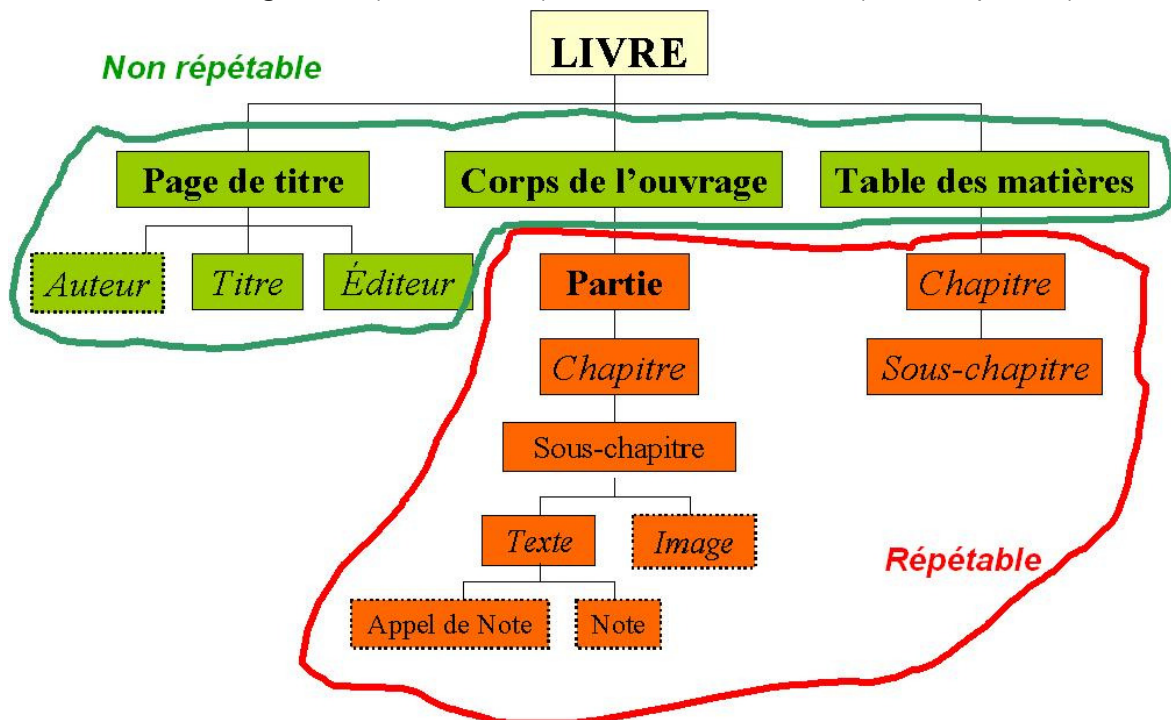
1.2.2 Structuration physique, structuration logique

Un livre présente une **structure physique**. Ainsi, les éléments d'une page (à gauche) peuvent être modélisés sous une forme arborescente (à droite) - *schémas établis par Pierre-Yves Duchemin*



Mais on peut identifier également une structure logique également exprimable sous forme d'arborescence, où nous distinguons :

- les éléments **uniques** des éléments **répétibles**
- les éléments **obligatoires** (cadre continu) des éléments **facultatifs** (cadre en pointillé)



Nous avons esquissés quelques notions de base permettant la formalisation d'une structure.

La distinction entre structure logique et structure physique permet

- de produire l'édition par combinaison entre les deux,
- de structurer un document indépendamment de sa forme physique, ce qui autorise une multiplicité d'éditions (impression en grand format et en format de poche, cédérom, internet...),
- d'associer à des éléments logiques des formes physiques propres à une édition (police, corps, couleur, gras, souligné, italique, etc.)

1.2.3 Les langages à balise (*Mark-up Languages*) : SGML, XML

La structure arborescente que nous avons représenté sous forme d'image peut s'exprimer dans un fichier « à plat » (c'est-à-dire un fichier qui sera lu par le système dans l'ordre séquentiel) grâce aux balises : <balise ouvrante>, </balise fermante>.

Exemple sommaire, où les retraits de texte ne sont là que pour aider à la lecture humaine :

```
<livre>
  <page de titre>
    <auteur>Isabelle Dussert-Carbone</auteur>
    [etc.]
    <titre>Le catalogage : méthode et pratiques. 1, Monographies et publications en série </titre>
    <editeur>Cercle de la Librairie</editeur>
  </page de titre>
  <corps de l'ouvrage>
    <partie>
      <chapitre>
        <sous-chapitre>
          <texte>xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx</texte>
        </sous-chapitre>
        [etc.]
      </chapitre>
      [etc.]
    </partie>
    [etc.]
  </corps de l'ouvrage>
  <table des matières>xxxxxxxxxxxxxxxxxxxxxxxx</table des matières>
</livre>
```

Dans les années 1980 a été développé sur ce principe le **SGML** (Standard Mark-Up Language), très complexe, qui n'a connu d'application que par de grands groupes d'édition ou industriels.

La fin des années 1990 a connu l'émergence d'**XML**, plus simple. C'est une structure générique qui permet de bâtir toutes sortes de modèles comportant une liste d'éléments et une syntaxe : livre, article de périodique, etc. Vous pouvez vous amuser à concevoir un modèle XML pour carte de visite.

XML est adopté par un nombre croissant de professions appelée à gérer de l'information structurée.

Un groupe européen d'éditeur appelé... EDItEur (<http://www.editeur.org>) a défini un modèle XML pour décrire un livre : ONIX

(voir <http://www.editeur.org/> et, en français, <http://www.editeur.org/ONIX2.0TradFr.Guide.pdf>).

On y retrouve des éléments comparables à ceux gérés par les formats MARC des bibliothécaires mais aussi d'autres informations utiles au monde de l'édition et de la librairie, comme la présentation de l'auteur :

```
<Contributor>
  <PersonNameInverted>Schuwer, Philippe</PersonNameInverted>
  <BiographicalNote> Philippe Schuwer a été secrétaire de rédaction dans la presse, sous-directeur
  de fabrication aux PUF, directeur aux éditions Tchou, directeur de département chez
  Hachette, Nathan et Larousse. Diplômé du British Institute et de l'Ecole des hautes études en
  sciences sociales, il a créé les premiers cours d'édition à l'Université Paris VIII.
  </BiographicalNote>
</Contributor>
```

ou des informations sur le matériel promotionnel (présentoirs, fac-similés agrandis de la couverture...), ou encore la largeur du livre, indispensable pour gérer le stockage (les bibliothécaires ne mentionnent que la hauteur). ONIX est un standard émergent dans le monde de l'édition.

1.3 Le Web

Objectifs :

- Publier

1.3.1 HTML

Le *world wide web*, qui a permis au début des années 1990 l'explosion d'Internet, repose à l'origine sur le langage HTML dans lequel sont écrit les pages. C'est un langage à balises simplifié. Pour voir du HTML, il suffit, quand vous affichez une page web, de cliquer dans la barre de menu sur *Affichage* puis de sélectionner *Source*.

HTML a surtout été utilisé pour décrire des éléments physiques de présentation (corps, police, gras, italique, etc.)

On peut voir le fichier source d'une page web en sélectionnant « Fichier source » dans l'onglet « Affichage ».

1.3.2 XML

XML est de plus en plus utilisé comme langage de description de pages web.

1.3.3 Les feuilles de style : CSS

La présentation physique d'un ensemble de pages web peut être gérée par un fichier extérieur dit feuille de style à chaque page HTML ou XML. Les pages comprendront alors des éléments logiques (titres de différents niveaux, corps de texte, etc.) et la feuille de style assignera des caractéristiques physiques à chaque élément. On peut ainsi les modifier en une seule fois pour toutes les pages. La syntaxe des feuilles de style est régit par le CCS (*Cascading Style Sheet*).

Chaque fichier HTML ou XML peut aussi comporter des éléments de style mais cela ne permet pas de gérer les changements de présentation dans plusieurs fichiers à la fois.

1.4 L'auteur

Objectifs :

- Créer
- Faire publier

1.4.1 Indigence primitive

L'auteur produit un texte manuscrit, puis de plus en plus tapé à la machine, et enfin saisi sur un ordinateur. Il ne se soucie généralement pas de la structure et saisit en fonction d'une présentation matérielle.

Pour l'éditeur, toutes les présentation matérielles saisies par l'auteur, si elles donnent des indications sur ses intentions de présentation, l'encombrent dans le traitement du texte : il doit généralement supprimer tous les enrichissements (gras, italiques, etc.)

1.4.2 Feuilles de style

Les logiciels de traitement de texte permettent généralement de gérer une structure par le système de la feuille de style. Sur la barre d'outils supérieure de Word, choisissez *Format* puis *Style* pour créer des éléments de style. A chacun d'eux est assigné un aspect matériel, que vous pouvez changer en une seule fois pour tout le document. Vous pouvez par exemple créer des titres de différents niveaux.

Dans Word, grâce à la feuille de style, une table des matières peut être créée automatiquement (*Insertion > Tables et index > Table des matières*).

Si vous avez affiché la présente page en format RTF, vous pouvez visionner le style de chaque séquence textuelle dans la barre d'outils *Mise en forme*.

Apprenons à structurer avec Word (ou un autre logiciel de traitement de texte)

(on peut visionner et modifier le style de chaque élément dans la barre d'outils *Mise en forme*)

Spontanément, les utilisateurs d'un traitement de texte travaillent pour la présentation et non pour la structure. Ce faisant, ils font un travail qui est toujours à recommencer.

Quelques conseils :

- Donner au fichier un nom significatif (qui permettra de le rechercher)
- Utiliser une feuille de style
- Afficher les caractères invisibles (espaces, saut de ligne ou de paragraphe, ...) en cliquant sur l'icône ¶ de la barre d'outils *Standard* pour maîtriser la structure.

- Gérer les retrait grâce à *Format > Paragraphe* et non par des tabulations
- Gérer des paragraphes grâce à *Format > Paragraphe*, en définissant les espacements avant et après en nombre de point, plutôt que de les séparer par deux sauts de paragraphes successifs.
- Pour présenter des tableaux (par exemple une colonne de mots et une colonne de chiffres), créer un véritable tableau plutôt que d'utiliser les tabulations (qui décalent les données si vous changez la taille de la police).

1.5 Une chaîne numérique complète ?

Dans le monde universitaire sont tentées des expériences pour permettre une chaîne numérique continue, de la saisie par l'auteur à l'édition sur papier ou en ligne.

Les auteurs de mémoires ou de thèses doivent alors saisir leur texte en respectant une feuille de style. Cela permet ensuite une conversion automatique de word en XML.

Sur une chaîne de traitement de thèses en ligne, voir <http://sourcesup.cru.fr/cyberdocs/>.

On voit ainsi se dessiner l'espoir d'une chaîne numérique complète qui irait :

- pour les produits imprimés, de l'auteur à l'imprimeur ;
- pour les produits numériques, de l'auteur à la publication sur le web ou un support électronique mobile (cédérom ; e-book).

On peut même imaginer une chaîne allant jusqu'à l'utilisateur final, qui pourrait récupérer et librement utiliser, manipuler, rediffuser l'œuvre, mais souvent les nécessités de la protection du droit d'auteur conduisent à briser la chaîne.

2 Retour sur la logique de la structuration

2.1 Structurer pourquoi ?

Structurer c'est séparer le logique du physique

C'est inscrire le contenu dans une forme pour pouvoir :

- le retrouver (grâce à des index spécifiques)
- le publier sous des formes diverses.

Qui structure l'information qu'il produit ou qu'il gère maîtrise ses contenus.

Qui structure l'information qu'il produit ou qu'il gère en maîtrise l'expression et la présentation.

2.2 Les logiques de structuration

Les modèles de structuration logique de l'information :

- définissent des éléments éventuellement hiérarchisés (le nom comprend le prénom et le nom de famille, etc.),
- assignent à chaque élément des attributs : répétitivité, obligation, format.

La structuration physique se fait par assignation à chaque élément de caractéristiques physiques.

2.3 Ne pas confondre...

- **identification** : élément non ambigu permettant de distinguer un objet d'un autre
- **structuration de l'information** : découpage de l'information en champ susceptible de se retrouver d'un enregistrement à l'autre (auteur, titre...)
- **classification ou classement** : mise en ordre d'objets pour pouvoir les retrouver ou d'information pour les regrouper selon un critère (par exemple leur contenu) – on peut distinguer le classement physique et la classification intellectuelle, des livres pouvant être classés physiquement (classement) d'après une classification intellectuelle

Combinaison entre les trois notions : un découpage de l'information peut comporter

- un ou des éléments d'identification (l'ISBN, le titre)
- un ou des éléments de classification (l'indice Dewey, la cote d'un livre)

2.4 Les métadonnées

Cette notion qui concerne un nombre croissant de métier désigne les « données sur les données » : classifications, catégories, mots clés...

Elles correspondent à ce que les bibliothécaires appellent « catalogage » mais peut porter sur des informations plus larges

Exemple : le DublinCore (utilisé pour les bibliothèques électroniques)

| Élément | (anglais) | Commentaire |
|---------------------------------|-------------|---|
| 1. Titre | Title | Titre principal du document |
| 2. Créateur | Creator | Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document |
| 3. Sujet ou mots clés | Subject | Mots-clefs, phrases de résumé, ou codes de classement |
| 4. Description | Description | Résumé, table des matières, ou texte libre. |
| 5. Éditeur | Publisher | Nom de la personne, de l'organisation ou du service à l'origine de la publication du document |
| 6. Contributeur | Contributor | Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément Contributor séparé |
| 7. Date | Date | Date d'un évènement dans le cycle de vie du document |
| 8. Type de ressource | Type | Genre du contenu |
| 9. Format | Format | Type MIME, ou format physique du document |
| 10. Identifiant de la ressource | Identifier | Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site |
| 11. Source | Source | Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. |
| 12. Langue | Language | |
| 13. Relation | Relation | Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc. |
| 14. Couverture | Coverage | Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle |
| 15. Droits | Rights | Droits de propriété intellectuelle, Copyright, droits de propriété divers |

3 La globalisation ou l'âge des standards

Les standards sont indispensables dans un monde où on échange l'information.

Il en existe deux sortes :

- les standards de droit ou normes, produits par des organismes habilités : en France l'Afnor (Association française de normalisation) sur le plan international l'ISO (International Standard Organisation),
- les standards de fait, émanant d'un organisme informel ou non habilité ou d'une entreprise.

Il existe des batailles de standard, et ce n'est pas forcément le meilleur qui gagne. Mais il est important d'utiliser le standard auquel la majorité s'est rallié.

Un organisme dont la mission est d'établir les standards du web, le W3C (World Wide Web Consortium), où siègent des organismes publics et privés, prend une place croissante dans la

définition des standards de structuration de l'information, au-delà) du seul Web. Il joue un rôle moteur dans le développement d'XML.

Alors qu'XML s'impose de plus en plus comme un standard universel, les standards propres à certaines professions survivent. On peut penser que leurs jours sont comptés.

La standardisation permet d'échapper aux frontières entre métiers, mais aussi entre logiciels. Les données transitent d'un logiciel à l'autre, les standards permettent leur conversion.

Comme le disait une revue informatique à la fin des années 1990 : « *Nous passons de la dictature du logiciel à la démocratie du document.* »

Conclusion

Vous arrivez dans un monde en mutation où les corporatismes contrarient l'unification, avec un va-et-vient entre unification et particularisme

Le professionnel de l'information structurée peut être au cœur de la stratégie des organisations.

A vous de jouer !