

Indexation et réutilisation de contenu

Unil

UNIL | Université de Lausanne

Nicolas Dunand – 13 juillet 2005



Lausanne - Suisse / 11.07 - 15.07 2005

T3FRUNIVERSITY 2005

De quoi faut-il tenir compte ?

- 1° : le contenu doit être **identifiable** sans équivoque
- 2° : le contenu doit pouvoir être **exporté** et **importé** facilement entre des environnements différents

1^{ère} partie : rendre le contenu identifiable

- ajouter des **métadonnées**, permettant d'**identifier** et d'**indexer** le contenu (p.ex. dans un vivier d'objets / de ressources) sur la base de différents attributs
- des métadonnées standard, propres aux objets d'apprentissage ont été établies : les LOM (**Learning Object Metadata**)

Structure des LOM (Learning Object Metadata)

Name	Value
Metadata	
General	
Identifier	IT global
Title	
Catalog Entry	
Language	fr
Description	
Keyword	
Structure	
Aggregation Level	
Life Cycle	
Status	
Contribution	
Contribution	
Technical	
Location	http://www2.unil.ch/infotheo/course_fr01/course-nonaw/
Requirement	
Installation remarks	
Other platform requirements	
Duration	
Educational	
Interactivity type	
Learning resource type	
Interactivity level	
Semantic density	
Intended end user role	
Context	
Typical age range	
Difficulty	
Typical learning time	
Description	
Language	fr
Rights	

Structure des LOM ... en XML :

```
<imsmd:lom>
  <imsmd:general> [55 lines]
  <imsmd:lifecycle> [120 lines]
  <imsmd:metametadata> [25 lines]
  <imsmd:technical> [36 lines]
  <imsmd:educational>
    <imsmd:interactivitytype>
      <imsmd:source>
        <imsmd:langstring xml:lang="x-none">LOMv1.0</imsmd:langstring>
      </imsmd:source>
      <imsmd:value>
        <imsmd:langstring xml:lang="x-none">Mixed</imsmd:langstring>
      </imsmd:value>
    </imsmd:interactivitytype>
    <imsmd:learningresourcetype>
      <imsmd:source>
        <imsmd:langstring xml:lang="x-none">LOMv1.0</imsmd:langstring>
      </imsmd:source>
      <imsmd:value>
        <imsmd:langstring xml:lang="x-none">Narrative Text</imsmd:langstring>
      </imsmd:value>
    </imsmd:learningresourcetype>
    <imsmd:interactivitylevel>
      <imsmd:source>
```

Pour simplifier la saisie : un masque de saisie

General	
Identifiant	<input type="text" value="IT global"/>
Title	<input type="text" value="Introduction à la Théorie de l'Information et ses Applications"/>
Catalog Entry	
Catalog	<input type="text" value="---"/>
Entry	<input type="text" value="---"/>
Language	<input type="text" value="fr"/>
Description	<input type="text" value="Ce cours présente une introduction à la Théorie de l'Information et ses applications. La Théorie de l'Information a été introduite par Claude Shannon comme moyen d'étudier et de résoudre les problèmes de communication ou de transmission de signaux à travers des voies. Bien que ce soit une vision relativement étroite de l'information qui se penche plus particulièrement sur la mesure du contenu de l'information, elle doit être incluse dans toute théorie -plus large- de l'information. Le cours de Théorie de l'information est un des projets soutenus par le Campus Virtuel Suisse, une initiative de la Conférence Universitaire Suisse (CUS) visant à diversifier et moderniser l'enseignement académique."/>
Keyword	<input type="text" value="information de Shannon"/>
Coverage	<input type="text" value=""/>
Structure	<input type="text" value="Collection"/>
Aggregation Level	<input type="text" value="4"/>
Life Cycle	
Contribution	
Role	<input type="text" value="Author"/>
VCard	<input type="text" value="BEGIN:VCARD N:Bavaud;François;;Professeur ADR;INTL;POSTAL;PARCEL;WORK;IMM-Lettres, IGUL-Géosciences et Environnement;Université de Lausanne, BFSH2;Lausanne;VD;1015;Switzerland TEL;PREF;WORK;VOICE:+41-21-692-3022 TEL;FAX:+41-21-692-3045 EMAIL;INTERNET:Francois.Bavaud@unil.ch ORG:UNI Lausanne;IMM-Lettres, IGUL-Géosciences et Environnement END:VCARD"/>
Date	<input type="text" value="2004-03-05"/>

Réutilisation de contenu : ... de quoi faut-il tenir compte ?

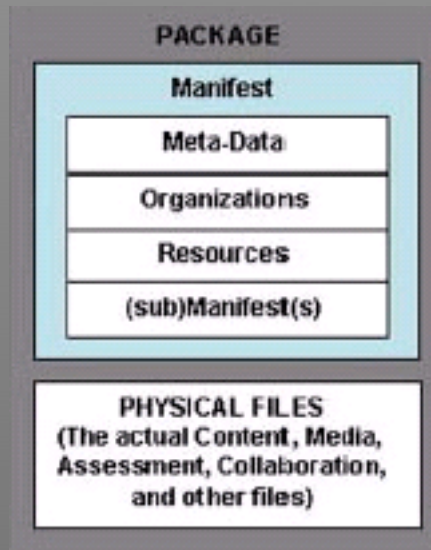
1° : le contenu doit être identifiable
sans équivoque

2° : le contenu doit pouvoir être
exporté et **importé** facilement
entre des environnements différents

2^{ème} partie : sous quelle forme exporter le contenu ?

- le contenu est généralement un ensemble de **plusieurs fichiers** – accompagnés de **métadonnées** – formant une entité logique **organisée**
- le tout doit être exportable en tant qu'**objet**
- il existe une spécification permettant la description d'un ensemble organisé de fichiers avec métadonnées : le **Content Packaging**

Content Packaging ?

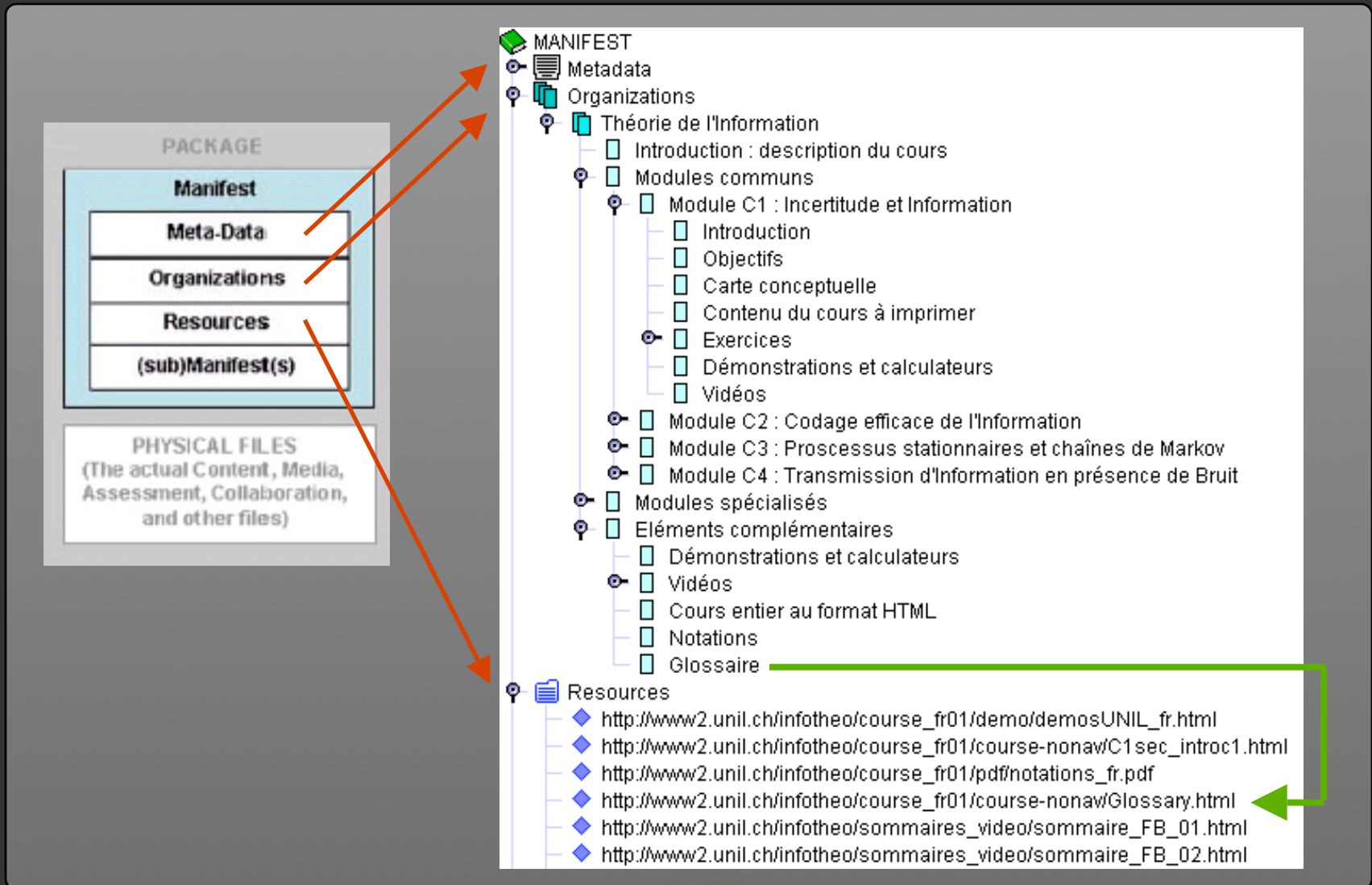


Content Package
(ZIP compressé)

=

manifeste
(fichier XML descripteur)
+
fichiers
(locaux / URLs)

Structure du manifeste :



Structure du manifeste en XML

```
<?xml version="1.0" encoding="UTF-8"?>
<manifest xmlns="http://www.imsglobal.org/xsd/imscp_v1p1" xmlns:imsmd="http://www.imsglobal.org/xsd/imsmd_v1p2" xmlns:xsi="
<metadata> [364 lines]
<organizations> [367 lines]
<resources>
  <resource identifier="RES-036" type="webcontent" href="http://www2.unil.ch/infotheo/course_fr01/demo/demosUNIL_fr.html" />
  <resource identifier="RES-7CD" type="webcontent" href="http://www2.unil.ch/infotheo/course_fr01/course-nonaw/C1sec_introc1.htm" />
  <resource identifier="RES-7A66" type="webcontent" href="http://www2.unil.ch/infotheo/course_fr01/pdf/notations_fr.pdf" />
  <resource identifier="RES-9FB" type="webcontent" href="http://www2.unil.ch/infotheo/course_fr01/course-nonaw/Glossary.html" />
  <resource identifier="RES-797" type="webcontent" href="http://www2.unil.ch/infotheo/sommaires_video/sommaire_FB_01.html" />
  <resource identifier="RES-85E" type="webcontent" href="http://www2.unil.ch/infotheo/sommaires_video/sommaire_FB_02.html" />
  <resource identifier="RES-7B5" type="webcontent" href="http://www2.unil.ch/infotheo/sommaires_video/sommaire_FB_03.html" />
  <resource identifier="RES-2AF" type="webcontent" href="http://www2.unil.ch/infotheo/course_fr01/course/" />
  <resource identifier="RES-6C5" type="webcontent" href="description.html">
    <file href="description.html" />
  </resource>
</resources>
<manifest identifier="MANIFEST-20"> [390 lines]
<manifest identifier="MANIFEST-913"> [401 lines]
<manifest identifier="MANIFEST-474"> [383 lines]
<manifest identifier="MANIFEST-4811"> [398 lines]
<manifest identifier="MANIFEST-59D"> [388 lines]
<manifest identifier="MANIFEST-0F"> [395 lines]
<manifest identifier="MANIFEST-068"> [389 lines]
<manifest identifier="MANIFEST-8F1"> [380 lines]
<manifest identifier="MANIFEST-1811"> [395 lines]
```

Content Package :

maintien de la structure d'ensemble

- Théorie de l'Information
 - Introduction : description du cours
 - Modules communs
 - Module C1 : Incertitude et Information
 - Introduction**
 - Objectifs
 - Carte conceptuelle
 - Contenu du cours à imprimer
 - Exercices
 - Enoncés
 - Solutions
 - Démonstrations et calculateurs
 - Vidéos
 - Module C2 : Codage efficace de l'Information
 - Introduction
 - Objectifs
 - Carte conceptuelle
 - Contenu du cours à imprimer
 - Exercices
 - Enoncés
 - Solutions
 - Démonstrations et calculateurs
 - Vidéos
 - Module C3 : Processus stationnaires et chaînes de Markov
 - Introduction
 - Objectifs
 - Carte conceptuelle
 - Contenu du cours à imprimer
 - Exercices
 - Enoncés
 - Solutions
 - Démonstrations et calculateurs
 - Vidéos
 - Module C4 : Transmission d'Information en présence de bruit
 - Introduction
 - Objectifs
 - Carte conceptuelle
 - Contenu du cours à imprimer

Introduction

Bienvenue dans cette première étape du monde de la *théorie de l'information*. De toute évidence, dans un monde qui est en passe de devenir une *société d'information*, la notion et le concept d'*information* devraient soulever une grande attention scientifique. En fait, bien que le développement de l'*information pragmatique*, dans les ordinateurs, sur Internet et dans d'autres réseaux informatiques, évolue à une allure extrêmement rapide, l'étude théorique et conceptuelle de ce qu'est l'information et de la façon dont elle devrait être traitée ne suit pas vraiment ce rythme frénétique.

La théorie de l'information, dans le sens technique qu'on lui donne aujourd'hui, remonte au travail de Claude Shannon. Elle a été introduite comme moyen d'étudier et de résoudre les problèmes de communication ou de transmission de signaux à travers des voies. Bien que ce soit une vision relativement étroite de l'information qui se penche plus particulièrement sur la mesure du contenu de l'information, elle doit être incluse dans toute théorie -plus large- de l'information. C'est pourquoi ce module est une introduction aux éléments de base de la théorie de l'information telle que Shannon et ses successeurs l'ont établie.

Mais, dans ce premier module déjà, nous tenterons d'élargir notre vision de l'information. Nous soulignerons qu'une information doit toujours être considérée par rapport à des questions précises. Une même information peut porter sur différentes questions, et son contenu d'information sera différent selon chaque question. Dans le cas de certaines questions, le contenu peut même être vide. La quantité contenue dans une information par rapport à une question donnée sera mesurée par la réduction de l'incertitude concernant cette question et induite par l'information ou, plus généralement, par le changement de l'incertitude. Nous suivrons Shannon en mesurant l'incertitude par l'*entropie*. Notre approche est dans l'esprit de celle de Shannon dans la mesure où l'information est mesurée par changement d'entropie, mais, en indiquant explicitement la question à laquelle l'information est appliquée, nous irons plus loin que lui.

Nous insisterons également sur l'importance de l'information préalable, par rapport à laquelle la quantité d'information sera mesurée. Bien que ceci soit implicite dans l'approche de Shannon, l'explicitation rend le concept plus clair. En effet, de ce point de vue, il devient évident que les probabilités sont elles-mêmes de l'information dont le contenu peut être mesuré par changement d'entropie.

Au fil des discussions il devient clair que l'information a aussi une structure algébrique : l'information peut être combinée ou assemblée et doit être concentrée sur des questions spécifiées. Cependant, nous ne traiterons pas cet aspect important en profondeur, cette tâche est réservée à d'autres modules. Ceci est également valable pour l'application de la théorie classique de l'information au codage, à la communication et à d'autres domaines.

Nous vous souhaitons beaucoup de plaisir dans l'étude de ce module.

Résultat : LOM + Content Packaging

- l'utilisation de **métadonnées (LOM)** permettent l'indexation et l'identification du contenu sur la base de différents attributs ;
- le **content packaging** permet la conservation de l'organisation des fichiers formant le contenu en un seul fichier facilement transportable ;
- Le **Content Package** ainsi créé contient toutes les informations nécessaires à la réimportation directe du contenu dans un nouvel environnement.