Standardisation des formats documentaires pour les chaînes éditoriales d'UNIT : un schéma pivot

Stéphane Crozat, Nicolas Delestre, Jacques Qyeyrut, Fabien Baillon, Fabien Gautron, Christine Vanoirbeek, Priscilla Velut, Xavier Hennequin

▶ **To cite this version:**

Stéphane Crozat, Nicolas Delestre, Jacques Qyeyrut, Fabien Baillon, Fabien Gautron, et al.. Standardisation des formats documentaires pour les chaînes éditoriales d'UNIT : un schéma pivot. TICE, 2006, Toulouse, France. 7 p. hal-00442860

HAL Id: hal-00442860
https://hal.science/hal-00442860

Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Standardisation des formats documentaires pour les chaînes éditoriales d'UNIT : un schéma pivot

Auteurs
Stéphane Crozat (UTC)
Contributeurs
Nicolas Delestre (INSA Rouen)
Jacques Queyrut (Bordeaux1)
Fabien Baillon (ENSTIMAC)
Pascale Gautron (ENSIETA)
Christine Vanoirbeek (EPFL)
Priscilla Velut (UTC)
Gestion et contact
Xavier Hennequin (UTC)

UTC
BP 60319
60203 Compiègne Cedex

stephane.crozat@utc.fr
nicolas.delestre@insa-rouen.fr
jacques.queyrut@math.u-bordeaux1.fr
fabien.baillon@enstimac.fr
gautropa@ensieta.fr
christine.vanoirbeek@epfl.ch
priscilla.velut@utc.fr
xavier.hennequin@utc.fr

.

Résumé

Depuis sa création, UNIT cherche à rendre identifiable et échangeables les ressources pédagogiques produites par les établissements partenaires. Un projet «Chaîne éditoriale XML » a pu voir le jour et a abouti notamment à la constitution d'un schéma pivot. Ce dernier doit permettre à plusieurs chaînes éditoriales d'échanger des ressources riches pédagogiquement. Cet article décrit la problématique de départ jusqu'aux solutions retenues pour le schéma pivot.

Introduction

Le problème de départ

La communauté UNIT cherche à produire des contenus pédagogiques qui soient :

- Echangeables (les membres de la communauté peuvent réutiliser tout ou partie des contenus des autres membres)
- Identifiables (les contenus doivent pouvoir être publiés selon une « charte » UNIT qui les définit et en prescrit l'usage)

Dans cette optique un problème d'interopérabilité se pose, qui peut être abordé de deux façons :

1. Soit l'on impose des outils de production qui assurent des formats nativement interopérables, ce qui se résume plus ou moins à adopter un outil unique avec un modèle unique (par exemple Word + un modèle de document donné) et impose de rester lié à cet outil ;
2. Soit l'on met à profit la puissance des technologies documentaires (XML) selon le schéma suivant :
   - Chaque outil produit selon ses propres schémas documentaires de production (schémas locaux[1]) et de publication (chartes locales) ;
   - Chaque outil s'assure que ses schémas documentaires sont compatibles avec un schéma pivot, qui assume le rôle de standard d'échange à l'échelle d'UNIT

1Le terme schéma est utilisé dans ce document au sens de modèle documentaire et non en tant que technique particulière d'expression de ce modèle. Un schéma peut être formalisé par une DTD, un W3C XML Schema, etc.

(standard local) ;
- Chaque outil propose les passerelles qui permettent de passer de ses schémas locaux au schéma pivot et du schéma pivot à ses schémas locaux.

Avant de détailler la solution 2 que nous préconiserons, évidemment, nous noterons que la solution 1 est une mauvaise solution car elle impose à chacun dans son contexte propre les règles générales qui, par définition, ne correspondent exactement à aucun contexte. Le numérique permet de faire évoluer la notion de format standard, puisque grâce au calcul, il autorise le passage d'un format à un autre. Et donc il n'est pas nécessaire que le format d'échange soit le format de travail, ni que le format d'échange à un niveau N, soit le même que celui au niveau N moins 1. Bien entendu on ne fait pas ces passages de standard pour le plaisir de faire calculer un ordinateur, mais parce que cela permet de traiter au niveau N moins 1 le cas général standard en même temps que les cas contextuels spécifiques.

Commentaires

Le format pivot doit être bien « délimité »
Il devra être suffisamment simple pour constituer un dénominateur commun acceptable pour chaque outil, en reléguant certaines fonctions au statut de fonctions spécialisées (donc non standardisées par UNIT, mais disponibles à travers une ou quelques chaînes éditoriales seulement et non toutes).

Les outils doivent être bien « soutenus »
Le travail d'import/export via un format pivot dans une chaîne éditoriale XML (sous réserve de la remarque précédente) reste relativement simple à implémenter. L'outil devra néanmoins disposer d'une capacité de maintenance/évolution suffisante pour assumer l'implémentation initiale du format pivot, ainsi que pour évoluer avec lui ensuite.

Pourquoi ne pas réutiliser un schéma « déjà standard » ?
Notre point de vue sur l'ingénierie documentaire est que les besoins documentaires sont toujours contextuels (aux pratiques) et donc qu'il est souhaitable de mettre au point des schémas locaux pour chaque besoin particulier. Néanmoins, afin évidemment de rationaliser le travail de modélisation puis de développement, il est souhaitable que ces schémas réutilisent au maximum certaines structures standard. En pratique nous proposons de définir des structures propres pour les éléments documentaires de haut niveau, faciles à abstraire (module, chapitre, exercices, etc.) et de réutiliser des structures standards pour les éléments de bas niveau (que nous appellerons ressources), tels que le texte, les formules de mathématiques, etc.. Sans en faire la démonstration ici, on posera que cette approche permet des économies notables par rapport à une redéfinition complète, et est sans surcoût par rapport à l'approche consistant à partir d'un schéma standard général pour chercher ensuite à l'adapter (cette position n'est que le reflet de l'une des origines de XML, à savoir une approche très localisée par la multiplicité des dialectes, versus l'approche SGML très centralisatrice).

Hypothèses

XML « structuré logiquement »

Les contenus pédagogiques sont destinés à des usages divers et évolutifs. Chaque pratique exige des formats de lecture différents et donc des mises en formes et des mises en support différentes. Or le numérique permet de calculer la transformation d'un format en un autre, et donc en particulier différentes mises en formes. Il n'est donc pas nécessaire de re-produire manuellement les contenus pour chaque support de lecture, à condition d'avoir respecté un cahier des charges technologique qui rende aisé les calculs de mises en forme variées.
Un tel cahier des charges est rempli par des contenus « structurés logiquement » et XML est le meilleur candidat à l'heure actuelle pour implémenter une telle structuration logique. On parlera alors de « XML structuré logiquement ».
Cette approche permet des économies considérables dans la mesure où une seule production peut couvrir de multiples usages en même temps, et dans le temps. Elle permet également d'éviter les erreurs inhérentes à tout processus de duplication de l'information. Enfin elle permet d'accéder à la pérennité, la portabilité, la séparation auteur/éditeur et le multi-supports.
Un contenu sera dit « XML structuré logiquement » si :

- L'ensemble de l'information qu'il véhicule est balisé en XML dans le respect d'une DTD ;
- Les balises dénotent la nature du contenu et de sa structure (structure logique) et non la façon de le mettre en forme (structure graphique). Le contenu ne prescrit donc pas la façon dont il sera mis en forme.

On notera que l'usage d'XML est une condition non suffisante pour la représentation logique de l'information. XHTML, qui est un langage XML, est néanmoins un langage de mise en forme, à l'instar d'HTML.

XHTML (XML structuré graphiquement)

Un format de publication est dédié à un support.

Il spécifie la façon dont l'information est présentée sur celui-ci.

```
<FONT color="red" font="Impact" size="24pt">
   Définition : Digital
</FONT>
<BR/>
<FONT color="black" font="Times" size="12pt">
   Relatif aux nombres.
</FONT>
```

Figure 7 : XML comme langage de mise en forme

XML structuré logiquement

Un format de stockage logique est indépendant d'un support.

Il spécifie la façon dont le contenu est intrinsèquement représenté.

Il est facile de générer un format de publication depuis un format de structure logique

```
<definition>
   <notion>Digital</notion>
   <explication>Relatif aux nombres</explication>
</definition>
```

Figure 7 : : XML structuré logiquement

Cahier des charges

Côté UNIT, nous proposons :

- D'établir les DTD de départ qui fixent le format documentaire pivot d'UNIT ;
- D'établir, pour chaque support de publication visé (papier, écran, web, etc.) la charte graphique et d'usage d'UNIT, qui valide le champ sémantique et fonctionnel des termes de la DTD ;
- D'établir les procédures de maintenance et d'évolution des DTD et des chartes de publication.

Côté outil, nous proposons que chaque chaîne éditoriale UNIT assume les fonctions :

- De production de contenu selon une logique « XML structuré logiquement » avec export dans le format UNIT (changement de DTD local vers pivot) ;
- De publication de contenu selon une logique de génération automatique de format de sortie depuis le format « XML structuré logiquement » dans les formats de publication de la charte UNIT avec import depuis le format UNIT (changement de DTD pivot vers local) ;
- D'évolution (avec un décalage de phase maîtrisé en terme de processus) en suivant les améliorations du format pivot UNIT ;

Donc pour chacun de ces outils, le standard UNIT constituera un sous ensemble de son champ fonctionnel en terme de capacités de création et de publication.

NB : On pourra envisager des outils capables soient de produire mais pas de publier (et inversement) selon les formats UNIT.

L'approche documentaire avec un format pivot permet d'autoriser la multiplicité des outils (chacun répondant à la fois spécifiquement à son besoin local) tout en conservant la cohérence du tout (via le format pivot).

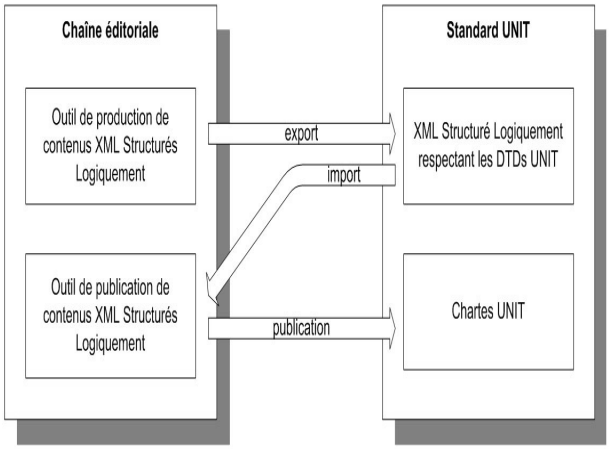

Figure 7 : Cahier des charges pour les chaîne éditoriales « labellisées » UNIT

Principes fondateurs

Abstraction des structures documentaires

Le premier principe que nous rappellerons rapidement, puisqu'il est un des intrants du projet, est celui de l'adoption d'un schéma propre qui définit une structure logique sans aucun attribut de mise en forme. Nous ne reviendrons pas ici sur les bénéfices de cette approche.

Modularisation des structures documentaires

Un second principe que nous proposons ici car il simplifie ensuite grandement la gestion des contenus (en particulier à des fins de maintenance et de réutilisation) et celui de modularisation des structures documentaires. Précisons tout de suite qu'il ne s'agit que d'un point de vue informatique de gestion (les documents sont éclatés en petits fichiers) et non d'un point de vue pédagogique sur la rédaction (il n'y a pas ici d'hypothèse sur l'autonomie d'une partie par rapport au tout). Il ne s'agit donc pas de poser « en dur » une approche dite « granulaire » (qui restera de la responsabilité éditoriale

de chaque institution participante).
La segmentation du document en multiples ressources informatiques indépendantes n'impose pas que ces ressources soient sémantiquement indépendantes.
On parlera de composant public lorsqu'il est réutilisable par différentes structures documentaires et inversement de composant privé lorsqu'il est informatiquement dépendant de ses structures mères (son cycle de vie lui est alors lié).
Modularisation des modèles (au niveau ressource)
Le dernier principe que nous proposons est que les schémas que nous construirons ne cherchent pas à déterminer ce que nous appellerons le niveau « ressource », et qui constitue en fait la modélisation des contenus à proprement parler, pour se concentrer sur la définition de structures de haut niveau. Nous proposons de faire appel pour le niveau ressource à des formats standard, ou à des sous ensembles de ces standards.

Solution proposée

Principes et terminologie

Nous proposons d'articuler la structure de la DTD pivot selon quatre types d'éléments documentaires :

- Unité d'Echange (UE) : Cette unité sert de base d'échange maximale du point de vue de la DTD pivot, cela signifie que si l'on désire décrire des assemblages plus « gros », cela doit être fait dans une LMS ou avec un EML, et cela signifie aussi qu'il est bien entendu possible de réaliser des échanges d'éléments plus « petits » que les UE. L'UE peut être plus ou moins « grosse », depuis un cours complet jusqu'à une seule activité isolée. L'UE permet de définir une organisation thématique.
- Unité d'Action (UA) : Cette unité rassemble un ensemble de contenus qui correspondent à une logique d'action cohérente pour un apprenant, comme par exemple du contenu à lire, annoter et résumer, des exercices à faire, une autoévaluation, etc. L'UA permet de définir une organisation didactique.
- Unité de Contenu (UC) : Cette unité agence un ensemble de contenu (texte, image, etc.) qui remplissent des rôles pédagogiques et constituent une unité documentaire : un concept, un exercice, etc. L'UC permet de définir une organisation documentaire.
- Ressource (Re) : Une ressource est un contenu inscrit selon une forme sémiotique (type de ressource) unique (texte ou image ou vidéo, etc.). La ressource est le contenu à proprement parler.

Le niveau Ressource

Nous proposons les types de ressources suivants :
Texte (qui contient les formules de mathématiques et les listings informatiques, plus éventuellement d'autres formats pour les formules de chimie, etc.) : le format retenu est un sous ensemble de la DocBook pour le texte, associé à Open Document Formula (ODF) pour les formules de mathématique, et à des fichiers ASCII (TXT) pour les listings.

- Tableau : le format retenu est Open Document Spreadsheet (ODS).
- Image bitmap : les formats retenus sont les formats PNG et JPEG.
- Dessin vectoriel : le format retenu est Open Document Graphic (ODG).
- Documents édités (au sens de fichier destiné à être maintenu dans une forme donnée) : le format retenu est PDF.
- Audiovisuel : le format retenu est MPEG-4 (AVC/H264, 1Mbit + ACC).
- Interaction (au sens de programme informatique autonome appelant des actions de l'utilisateur au cours de son déroulement) : les formats retenus sont HTML (un site sous forme de .JAR avec un packaging spécifique, Flash SWF, les formats Open Document et tout autre format (sous forme de .JAR avec un packaging spécifique)).

Gestion de droits

On pose que chaque objet public (UE, UA, UC et Re) est associé à une licence Creative Commons (by-nc-sa, by-nc-nd, by-nc, by ou publicdomain).

Exemple

Exemple de fichiers

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<AcademicUe xmlns="urn:unit-c.fr:schema:academic:v1"
xmlns:dc="urn:unit-c.fr:schema:dublincore:v1"
xmlns:db="urn:unit-c.fr:schema:docbook:v1"
xmlns:lom="urn:unit-c.fr:schema:lomfr:v1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:unit-c.fr:schema:academic:v1
unit-c.fr_schema_academic_v1.xsd">
   <id>/utc.fr/ics/stc/Ue001.xml</id>
   <title>
      <fullTitle>Modélisation des bases de
données</fullTitle>
      <shortTitle>Modélisation</shortTitle>
   </title>
   <dc:head>
      <dc:creator>Stéphane Crozat</dc:creator>
      <dc:publisher>UTC</dc:publisher>
      <dc:subject>Bases de données</dc:subject>
      <dc:language>fr</dc:language>
      <dc:spatial>fr</dc:spatial>


<dc:rights>http://creativecommons.org/licenses/publicdoma
in/</dc:rights>
      <dc:educationLevel>M</dc:educationLevel>
   </dc:head>
   <introduction>
      <db:text>
         <db:simpara>La modélisation est l'étape
fondatrice du processsus de conception de
<db:acronym>BD</db:acronym>. Elle consiste à abstraire
le problème réel posé pour en faire une reformulation qui
trouvera une solution dans le cadre technologique d'un
<db:acronym>SGBD</db:acronym>. Après avoir rappelé
succintement les fondements et objectifs des SGBD, ce
chapitre proposera les outils méthodologiques nécessaires
à la modélisation, à travers les formalismes Entité-
Association et Unified Modeling Language.</db:simpara>
      </db:text>
   </introduction>
   <division>
      <title>
         <fullTitle>Introduction aux bases de
données</fullTitle>
      </title>
      <CourseUa>/utc.fr/ics/stc/Ua001.xml</CourseUa>
      <CourseUa>/utc.fr/ics/stc/Ua002.xml</CourseUa>
      <AutoevaluationUa>/utc.fr/ics/stc/Ua003.xml</
AutoevaluationUa>
   </division>
   <division>
      …
   </division>
</AcademicUe>
```

Figure 7 : /utc.fr/ics/stc/Ue001.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<CourseUa xmlns="urn:unit-c.fr:schema:academic:v1"
xmlns:dc="urn:unit-c.fr:schema:dublincore:v1"
xmlns:db="urn:unit-c.fr:schema:docbook:v1"
xmlns:lom="urn:unit-c.fr:schema:lomfr:v1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:unit-c.fr:schema:academic:v1
unit-c.fr_schema_academic_v1.xsd">
   <id>/utc.fr/ics/stc/Ua001.xml</id>
   <title>
      <fullTitle>Vue d'ensemble</fullTitle>
   </title>
   <dc:head>
      …
   </dc:head>
   <ExpositionUc>/utc.fr/ics/stc/Uc001.xml</ExpositionUc>
   <PracticeUc>/utc.fr/ics/stc/Uc004.xml</PracticeUc>
   …
</CourseUa>
```

Figure 7 : /utc.fr/ics/stc/Ua001.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<ExpositionUc xmlns="urn:unit-c.fr:schema:academic:v1"
xmlns:dc="urn:unit-c.fr:schema:dublincore:v1"
xmlns:db="urn:unit-c.fr:schema:docbook:v1"
xmlns:bib="urn:unit-c.fr:schema:bibliography:v1"
xmlns:lom="urn:unit-c.fr:schema:lomfr:v1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:unit-c.fr:schema:academic:v1
unit-c.fr_schema_academic_v1.xsd">
    <id>/utc.fr/ics/stc/Uc001.xml</id>
    <title>
        <fullTitle>Vue d'ensemble</fullTitle>
    </title>
    <dc:head>
        …
    </dc:head>
    <definition>
        <title>
            <fullTitle>Bases de données</fullTitle>
            <shortTitle>DB</shortTitle>
        </title>
        <content>
            <db:text>
                <db:simpara>Une Base de Données est un
ensemble volumineux, structuré et minimalement
redondant de données, reliées entre elles, stockées sur
supports numériques centralisés ou distribués, servant
pour les besoins d'une ou plusieurs applications,
interrogeables et modifiables par un ou plusieurs
utilisateurs travaillant potentiellement en parallèle
</db:simpara>
            </db:text>
        </content>
    </definition>
    <example>
        <title>
            <fullTitle>Une compagnie aérienne</fullTitle>
        </title>
        <content>
            <audiovisual
refUri="/utc.fr/ics/stc/video001.mpeg"/>
        </content>
    </example>
</ExpositionUc>
```

Figure 7 : /utc.fr/ics/stc/Uc001.xml

(exemple limite)

Exemple d'organisation

L'exemple suivant est celui d'un cours complet de 100 H

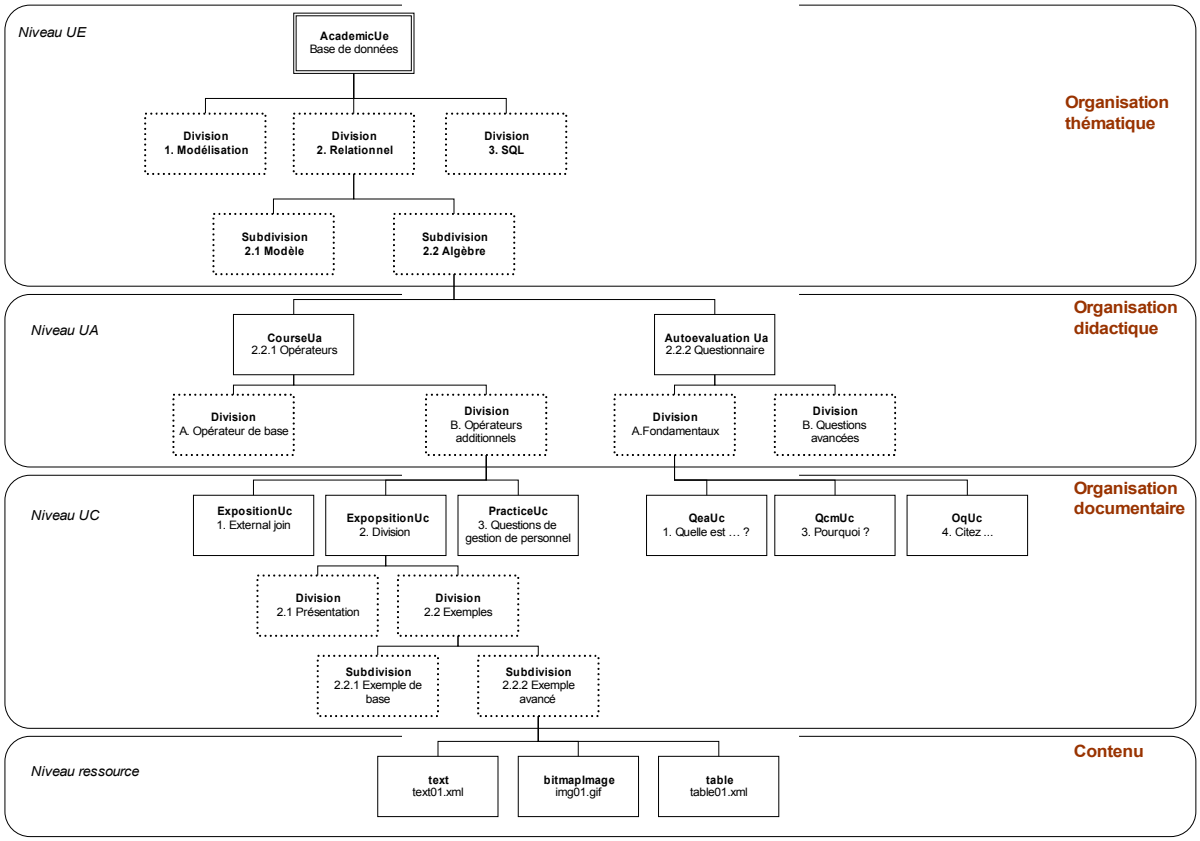

Figure 7 : Exemple d'organisation d'un document de cours de 100h selon le schéma pivot UNIT

Conclusion

Le schéma est implémenté sous la forme d'une implémentation W3C XML Schema. Ce schéma rend techniquement possible la construction d'une base de contenus échangeables et réutilisables par chacun des établissements. Des rapprochements avec des outils allant dans ce sens sont actuellement effectués.

Par ailleurs, il est essentiel d'accompagner ses futurs utilisateurs sur toute la chaîne éditoriale. Les auteurs des documents sont les premiers visés puisqu'il est en effet nécessaire de savoir rédiger en XML « structuré logiquement ». C'est pourquoi, parallèlement à la définition et à la construction du schéma pivot, se mettent en place des points conseils chargés de diffuser les bonnes pratiques du « XML structuré logiquement ».

Des retours d'expérience viendront compléter cet article au cours du colloque TICE 2006. Enfin un travail commun est lancé sur un schéma pivot inter-UNT.