



**Mise en place d'une chaîne d'édition structurée aux éditions
Quae : aspects techniques**

Perraud G. & Dandurand C.

Cahier des Techniques de l'INRA, Numéro spécial 2012

pp. 91-95

Cet article est tiré du numéro spécial 2012 du Cahier des Techniques de l'INRA :

Marchoux E. (Coord.) & Hologne O. (Dir.) **L'Information Scientifique et Technique à l'Inra, des compétences au service de la recherche. Retour d'expérience sur des projets, services, outils et méthodes.** *Cahier des Techniques de l'INRA* Numéro spécial 2012. Paris : Inra, 2012. 141 p.

Mise en place d'une chaîne d'édition structurée aux éditions Quæ : aspects techniques

Guillaume Perraud¹, Caroline Dandurand¹

Résumé. Les évolutions récentes des supports de lecture et des habitudes des lecteurs, notamment pour les documents scientifiques, modifient le métier d'éditeur. L'édition devient multisupport (ordinateur, tablette, smartphone, papier), et la chaîne graphique classique se modifie afin de s'adapter à la multiplication des supports de lecture. L'utilisation du langage XML permet de décrire la structure du document, ce qui élargit les possibilités d'exploitation avec tous les enrichissements qu'autorisent les différents formats. Le développement du langage XML affranchit en outre les fichiers obtenus des aléas industriels et commerciaux des éditeurs de logiciels, rendant ces fichiers utilisables dans le long, voire le très long terme. Les éditions Quæ ont intégré cette technologie et ont adapté leur chaîne de production, avec comme objectif de considérer un ouvrage à partir de sa structure, le livre papier étant l'une des formes possibles de l'ouvrage.

Mots clés : édition structurée, xml, fichier pivot, diffusion, multisupport

Introduction

L'apparition de nouveaux supports et de nouvelles habitudes de lecture transforment progressivement le monde de l'édition. Le livre électronique se développe en complément du livre papier qui, contrairement aux craintes premières, ne pâtit pas de ce nouveau mode de diffusion mais bénéficie de la nouvelle impulsion qu'il donne à l'édition. Les éditeurs doivent s'adapter à ces nouveaux formats et anticiper les demandes du lecteur. Dès leur création en 2006 par l'Inra, le Cemagref (maintenant Irstea), le Cirad et l'Ifremer, les éditions Quæ ont cherché à anticiper les évolutions que le livre numérique allait apporter. Le rapport remis par B. Prost sur demande des éditions Quæ recommandait la mise en place d'une chaîne d'édition structurée et l'utilisation de fichiers XML (Prost, 2007). Le langage XML s'est en effet rapidement imposé comme étant au cœur de la transition vers le numérique, et la mise en place d'une telle chaîne de fabrication a débuté en 2008.

L'édition structurée

Principes

L'édition structurée consiste à associer le contenu d'un document à sa structure au sein d'un même fichier : le fichier pivot. Alors que dans la chaîne d'édition traditionnelle, le moyen de rendre compte de la structure n'est assuré que par la mise en forme d'un document au travers des enrichissements typographiques (graisse, corps, espaces avant/après...), le langage XML permet de décrire informatiquement cette structure, rendant possible sa transcription formelle par un simple export dans le format retenu — pourvu que les outils aient été développés.

XML

Le langage XML (*extensible mark-up language*) a été élaboré en 1996 à l'initiative d'un consortium international : le W3C (World Wide Web Consortium) avec, notamment et grossièrement, les objectifs suivants :

- disposer d'un outil informatique permettant de décrire la structure d'un document (au moyen de balises, repérées par les fameux chevrons <>);
- obtenir un fichier exploitable aussi facilement que possible (un éditeur de texte suffit pour lire et modifier un fichier .xml);

¹ Editions Quæ, RD 10, F-78026 Versailles Cedex, France ; guillaume.perraud@versailles.inra.fr

- produire un fichier dans un format non propriétaire (affranchi des aléas industriels et commerciaux des éditeurs de logiciels).

Si Word remplit la première condition grâce aux feuilles de styles, il pose incontestablement problème pour les deux autres : non seulement un simple bloc-note (notepad) ne permet pas d'ouvrir un fichier au format .doc, mais le format lui-même est lié au devenir (voire au bon-vouloir) de son éditeur, Microsoft.

XML est, au contraire, non seulement un outil de description de structure, mais aussi un format d'archivage pérenne. À l'intérieur de la norme XML, différents formats existent, qui décrivent la place et la manière dont s'articule chaque élément avec les autres. Le format retenu par les éditions Quæ est un dérivé de la TEI (*text encoding initiative*²), un standard de l'édition scientifique.

Exemple sommaire

Un exemple de texte structuré en XML est donné ci-dessous :

```
<div1>
<titre>L'édition structurée</titre>
<p>L'édition structurée s'appuie sur le langage XML.</p>
<div2>
<titre>Principes</titre>
<p>Le XML, <hi>extensible markup language</hi>, est une émanation du W3C.</p>
</div2>
</div1>
```

Cet exemple, simplifié à l'extrême, donne l'aperçu d'une structure. Les titres de niveaux 1 et 2 sont caractérisés (la présence d'une barre de fraction indique la fermeture d'une balise), la balise <p> indique un paragraphe de texte sans valeur particulière. À l'intérieur de ce paragraphe, la balise <hi> révèle un groupe de mots mis en évidence ; la manière dont cette mise en évidence (ici la traduction d'un sigle) doit être formalisée sera évidemment conforme à ce qui sera prévu par la maquette du format d'export (par exemple gras pour un epub, italique pour une impression papier, couleur pour un site web...).

Si tout un document est ainsi décrit, le seul format d'archivage sera le fichier XML, qui permettra ultérieurement de réutiliser le contenu selon les opportunités éditoriales.

Changement dans la chaîne graphique, évolution du métier

Du fait des habitudes de travail parfois solidement ancrées et par souci d'ergonomie, il n'est pas raisonnable de demander à un éditeur ou à un assistant d'édition de traiter un manuscrit en apposant des balises XML manuellement. Il a été décidé de faire évoluer la chaîne éditoriale des éditions Quæ vers un modèle remettant le moins possible en question des méthodes de travail éprouvées, et en toute hypothèse en conservant MS Word comme outil de base (Figure 1).

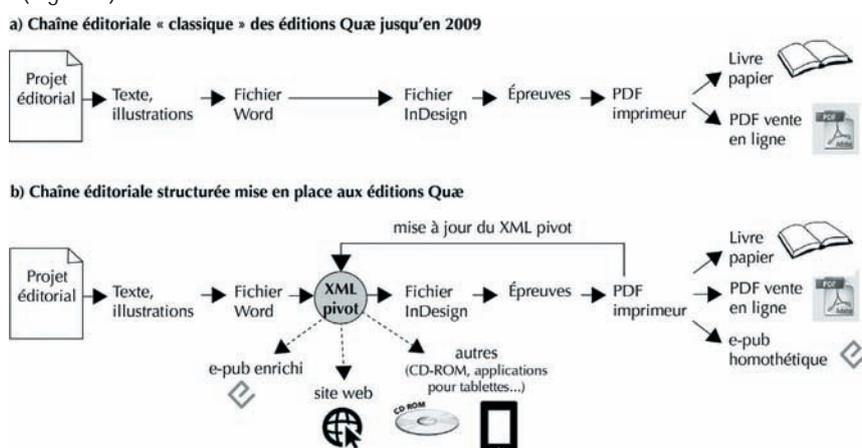


Figure 1. Chaîne d'édition d'ouvrage « classique » et chaîne d'édition structurée.

Jusqu'en 2009, seul le schéma a) s'appliquait. Actuellement, les deux chaînes cohabitent, l'objectif étant d'appliquer le schéma b) à l'intégralité de la production en 2015.

Production du XML

La solution retenue a été de baliser les différents éléments d'un manuscrit à l'aide des styles Word, puis de transformer ce fichier stylé en un fichier XML reprenant les balises créées grâce à une feuille de transformation XSLT³ (communément appelée « moulinette »), écrite par un prestataire extérieur, en passant par Open Office Writer — MS Word 2003 ne permettant pas d'exporter un fichier au format XML⁴. Le fichier obtenu est celui à partir duquel peut être exporté l'ouvrage. L'essentiel de la production des éditions Quæ étant en 2008 sous forme de livres imprimés, une seconde moulinette a été développée, qui crée un nouveau fichier XML au moyen d'un éditeur XML — distingué du pivot par l'ajout du suffixe « aid » dans son nom — qui sera coulé⁵ dans InDesign par le maquettiste. Ce fichier *aid.xml contient les balises de l'XML pivot traduites en balises interprétables par InDesign.

Les changements pour les éditeurs de Quæ ont été les suivants :

- la nécessité d'un stylage du manuscrit avec une plus grande rigueur par les éditeurs sous MS Word ;
- l'utilisation d'une fonction d'Open Office Writer pour créer le fichier XML pivot ;
- le contrôle de la validité du fichier XML au moyen d'un logiciel, XML Editor, et sa transformation en un XML exploitable par InDesign.

Le fichier XML destiné à la mise en page papier présente en outre le grand avantage de permettre une automatisation de la mise en page dans Adobe InDesign : pour peu que les styles soient correctement créés dans le gabarit de la collection accueillant l'ouvrage, tous les éléments décrits trouveront un style et donc une mise en forme (police, graisse, corps, espaces...), et tous les éléments iconographiques seront placés. La personne chargée de la PAO n'aura plus qu'à ajuster les tableaux, les figures et photos, et les manipulations fines — veuves et orphelines, césures, alignements, bas de pages, etc.

Tenue du XML pivot à jour

La mise en page des ouvrages papier est toujours réalisée sous Adobe InDesign, par le biais à présent d'un dérivé du fichier XML pivot, ce dernier étant le cœur de l'ouvrage à archiver. Il est donc impératif que le XML pivot soit le reflet exact de la version définitive de l'ouvrage — après relecture, correction, etc. Or, si la partie pré-PAO est articulée autour du contenu et de la notion de structure, toute la partie du travail éditorial qui se situe après le coulage dans Adobe InDesign demeure traditionnelle : la mise en page mais aussi les dernières coquilles sont corrigées à partir de la première épreuve, et sont donc reportées directement dans le fichier Adobe InDesign. Le texte imprimé n'est donc jamais le même que ce qui existe à ce moment-là dans le fichier XML supposé pivot !

La tenue des fichiers XML à jour est le point critique de la chaîne éditoriale structurée : reporter les corrections de la première, de la deuxième, éventuellement de la troisième, etc., épreuve est un travail potentiellement long, toujours fastidieux, et souvent décourageant. Un outil est donc en développement, qui permettra de récupérer les modifications apportées dans le fichier Adobe InDesign pour les insérer dans le fichier XML pivot.

Résultats

Format de diffusion

La structuration du fichier Word suivant la feuille de styles de l'éditeur puis sa conversion en XML permettent l'exportation vers différents formats. Pour la diffusion sous un format papier, le XML InDesign sera utilisé. Concernant la diffusion au format électronique, le XML permet de produire des ebooks. Le format epub (acronyme de *electronic publication*) est un format ouvert, normé, qui permet d'afficher le contenu en l'ajustant au type d'appareil de lecture quel qu'il soit. Lorsque la mise en page de l'ouvrage électronique est identique à celle de l'ouvrage papier, on parle de livre électronique homothétique ou *fixed layout* — très semblable à un PDF. La dernière version epub3, attendue pour l'automne 2012, permet l'ajout dans le texte de métadonnées et de contenu enrichi, développant ainsi l'interactivité de l'ouvrage. On parle alors de livre électronique enrichi (par rapport à ce qu'est ou pourrait être

3 Extensible stylesheet language transformations.

4 Les versions plus récentes de Word (à partir de 2007) permettent de produire des fichiers XML dont la structure est profondément différente de celle des fichiers issus d'Open Office. Or, les moulinettes utilisées par Quæ ont été développées à partir des XML Open Office. La production de fichiers XML à partir de Word imposerait par conséquent de réécrire l'ensemble des feuilles de transformation XSLT — solution qui n'a pas été retenue.

5 Le coulage désigne l'importation d'un texte et des éléments graphiques dans un logiciel de mise en page à partir d'un autre format — généralement .doc pour le texte, *aid.xml pour les éditions Quæ.

le même ouvrage au format papier). Robert Darnton imaginait déjà en 1999 une organisation du livre électronique enrichi sous forme pyramidale (Figure 2).



Figure 2. Proposition d'organisation pyramidale du livre électronique (d'après Darnton, 1999).

Une limite existe toutefois : le format epub ne permet pas de transmettre une mise en page « travaillée » : les collections dont la mise en page papier est complexe prennent alors une forme plus classique. Il semble donc très adapté pour l'édition scientifique classique, mais demande plus d'interventions dans le cadre de collections « grand public ».

Difficultés

La mise en place d'une chaîne d'édition structurée au sein d'une maison d'édition entraîne des modifications conceptuelles et organisationnelles. Les auteurs doivent, lors de l'écriture, considérer différents supports de diffusion possibles. Dès sa soumission à l'éditeur, le manuscrit n'est plus envisagé uniquement sous sa forme papier mais également au format numérique, qu'il soit enrichi (d'extraits vidéo, sonores, de liens, etc.) ou non. L'éditeur en charge d'un ouvrage doit s'attacher à la structuration correcte du contenu, cette étape étant le préalable indispensable à toute diffusion quel que soit le support. Concrètement, il s'agit pour l'éditeur de vérifier que l'auteur a stylé de façon précise le manuscrit sous Word à l'aide de la feuille de styles Quæ. Cette feuille de styles est intégrée dans un fichier modèle (.dot) disponible sur le site de l'éditeur.

La transition vers l'édition structurée rencontre quelques réticences de la part des auteurs et des éditeurs (O'Keefe, 2008). L'installation d'une feuille de styles dédiée à l'éditeur et le stylage du document Word présentent parfois quelques difficultés pour les auteurs. L'enrichissement du contenu peut également être complexe à concevoir : quels documents, quels formats peuvent venir « enrichir » l'ouvrage ? Par ailleurs, les épreuves de la version papier ne correspondront pas à la version numérique de l'ouvrage au format epub. On parle alors de WYSIOO (*what you see is one option*), a contrario du WYSIWYG (*what you see is what you get*), auquel les auteurs sont habitués. Certains logiciels proposent une interface WYSIWYM (*what you see is what you mean*), qui permet d'organiser l'information suivant l'idée de l'auteur. Cette approche serait intéressante dans le cas d'un ouvrage en diffusion électronique seulement, mais ne semble pas convenir à une diffusion multisupport incluant une version papier au contenu linéaire.

Le métier d'éditeur évolue. Certaines maisons d'édition se considèrent maintenant comme des éditeurs multimédia (Ferrand, 2012). Elles devront à terme concevoir des projets transmédiés pouvant notamment être interactifs. Les frontières entre les différentes activités d'édition deviennent floues et de plus en plus de développeurs et de professionnels du jeu vidéo participent à la conception d'ouvrage.

Conclusions et perspectives

Grâce au fichier XML pivot, les éditions Quæ peuvent aujourd'hui produire les formats epub des ouvrages publiés. Pour cela, il est indispensable que les fichiers XML pivot soient le plus propres et définitifs possible. Les ouvrages devront être envisagés avec des enrichissements afin de proposer au lecteur des ebooks enrichis.

Ces évolutions technologiques devront s'accompagner en interne de transformations non seulement au niveau éditorial mais également du marketing et commercial. De nouveaux liens avec le lecteur et de nouvelles méthodes de promotion restent à imaginer.

Références bibliographiques

Darnton R (1999) Le nouvel âge du livre. *Le Débat*, 105, 176-184.

Ferrand C (2012) Entretien avec Catherine Lucet. *Livres Hebdo*, 895, 20-21.

O'Keefe S (2008). Paradigm shifts are never pretty: advice on making the move to XML authoring. www.thecontentwrangler.com.

Prost B. (2007). Rapport d'étude sur l'édition numérique de livres scientifiques et techniques — L'éditeur des années 2010. http://medici.in2p3.fr/fr/images//Rapport_Prost.pdf

