

Chaînes éditoriales XML

Table des matières

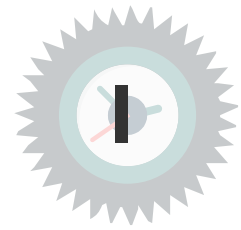
Objectifs	3
I - Introduction aux chaînes éditoriales	4
1. Définition des chaînes éditoriales XML	4
2. Illustration d'une chaîne éditoriale.....	5
3. Architecture des chaînes éditoriales XML	5
4. Écriture WYSIWYM	6
5. Polymorphisme et rééditorialisation	7
6. Les 6 "M"	8
7. Problématique : Motivations et contexte d'usage	10
II - L'environnement de conception de chaînes éditoriales Scenari	12
1. Présentation de Scenari	12
2. La suite logicielle Scenari	13
3. La communauté scenari.org.....	13
4. Exemples d'usages.....	14
Glossaire	15
Bibliographie	16
Webographie	17

Objectifs



- Comprendre le principe d'une chaîne éditoriale XML
- Découvrir le système de conception de chaînes éditoriales Scenari
- Expérimenter une chaîne éditoriale
- Apprendre à utiliser une chaîne éditoriale pour la formation

Introduction aux chaînes éditoriales



Une chaîne éditoriale numérique est un procédé de création de contenu multimédia qui, à la différence des outils bureautiques par exemple, ne fixe pas la mise en forme graphique du contenu a priori. La chaîne éditoriale permet en effet d'éditer une forme éloignée de la mise en forme (dite forme canonique), et d'ajouter au contenu des descripteurs de structure (la partie B est incluse dans la partie A) et des descripteurs sémantiques (la partie A décrit une procédure, la partie B est une définition, ...).

La chaîne éditoriale est alors capable de traiter automatiquement cette forme canonique, pour associer à chaque partie de contenu une mise en forme en fonction des descripteurs associés et du contexte de publication visé : support, public, modalité... Cette approche permet en particulier d'associer automatiquement plusieurs organisations et mises en forme différentes pour un même contenu.

Ce permet des gains substantiels en terme de productivité, de qualité et de diffusion de l'information.

1. Définition des chaînes éditoriales XML

Chaîne éditoriale XML



Définition

Une chaîne éditoriale XML est un système informatique permettant la production de documents structurés en XML, grâce à :

- un éditeur **WYSIWYM** ("*What You See Is What You Mean*" ou "ce que vous voyez est ce que vous voulez dire"),
- la publication **polymorphe** (Web, papier, diaporama, etc.),
- et la **ré-éditorialisation** sans recopie (dés-agrégation / ré-agrégation de documents).

Implémentations



Méthode

On peut distinguer trois grandes modalités pour implémenter une chaîne éditoriale :

- L'implémentation « sur mesure » consiste à mobiliser différentes briques logicielles et à programmer un logiciel spécifique pour un besoin particulier. Par exemple en couplant un éditeur du marché avec un outil de stockage et en réalisant des moteurs de publication sous la forme de feuilles XSL-XSLT.
- L'implémentation « modèle dédié » consiste à implémenter en dur un modèle de chaîne éditoriale pour un besoin assez transversal, tel que la publication de livres ou de modules de cours universitaires.
- L'implémentation « générique » consiste à réaliser un système paramétrable, indépendant d'un modèle en particulier, et à le configurer en fonction des besoins. Il s'agit alors de ce que l'on peut appeler un Système de Gestion de Chaînes Éditoriale par analogie aux Systèmes de Gestion de Bases de Donnée

Historique : de la recherche au développement économique



Le concept de chaîne éditoriale trouve son origine dans l'édition et dans la presse : il consiste à organiser les tâches de production et de publication, en séparant les métiers intervenant dans le processus. Une chaîne éditoriale est donc un processus avant d'être un outillage, qui est né d'un besoin d'industrialisation.

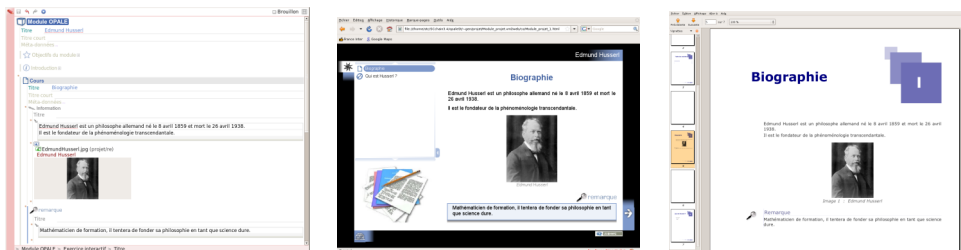
Historiquement, les deux technologies ayant permis l'implémentation de chaînes éditoriales numériques sont LaTeX (1982) et SGML (norme ISO depuis 1986). XML (standard W3C depuis 1998) est devenue la technologie de référence pour la réalisation de chaînes éditoriales. C'est sa maturité qui a permis de sortir ce procédé des domaines auxquels il était confiné jusque-là (documentation technique stratégiques dans l'aviation ou publication scientifique typiquement).

En 1999, l'Université de Technologie de Compiègne réalise Scenari, premier environnement intégré de conception de chaînes éditoriales XML proposant un langage déclaratif de modélisation et de publication de haut niveau.



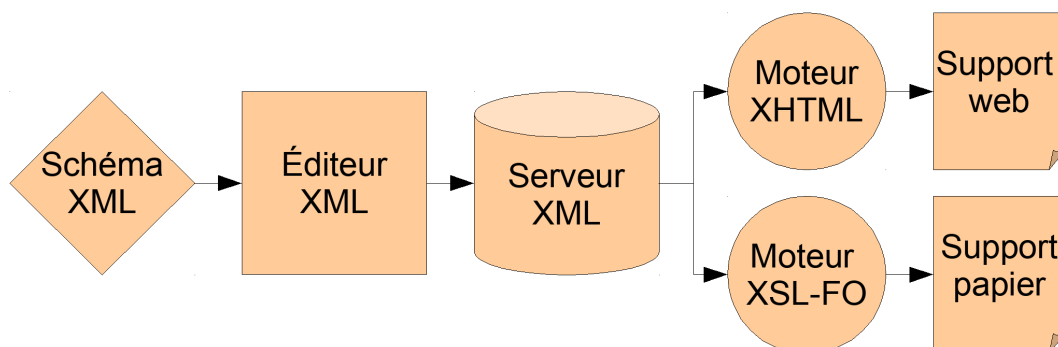
- Fondements théoriques :
*Ingénierie des connaissances et des contenus** (Bruno Bachimont, 2007)
- Fondements pratiques :
*Scenari, la chaîne éditoriale libre** (Stéphane Crozat, 2007)

2. Illustration d'une chaîne éditoriale



Exemple

3. Architecture des chaînes éditoriales XML



Exemple d'architecture d'une chaîne éditoriale XML

L'approche classique pour réaliser une chaîne éditoriale XML est la suivante :

1. Formalisation du schéma documentaire, avec un langage de modélisation de schéma XML (XML Schema, Relax NG, etc.)
2. Utilisation d'un éditeur XML standard et stylage de cet éditeur (il existe de très nombreux éditeurs XML, plus ou moins graphiques, qui se paramètrent automatiquement lorsqu'on leur fournit un schéma : Oxygen, XMetal, Arbortext, etc.).
3. Utilisation de serveurs de fichiers XML pour la gestion centralisée des contenus (pour les projets les plus importants surtout).
4. Réalisation de moteurs de transformation avec les technologies XSL-XSLT, combinées avec des langages de rendu (*XSL-FO** pour le papier, XHTML pour l'écran, etc.)

L'ensemble est en général intégré avec un langage applicatif tiers (Java, PHP, etc.).

4. Écriture WYSIWYM

La plupart des systèmes d'édition numériques se fondent sur le paradigme du WYSIWYG, « What you see is what you get », ou littéralement en français : « Ce que vous voyez est ce que vous obtenez ». Cette approche vise, en résumé, à permettre à l'utilisateur de créer un document tout en composant le rendu final, comme dans sur un support traditionnel non numérique.

La chaîne éditoriale se fonde, elle, sur le concept du WYSIWYM, « What you see is what you mean », c'est à dire : « Ce que vous voyez est ce que vous voulez dire ».

WYSIWYM



Définition

« Le WYSIWYM oppose sa démarche à celle du WYSIWYG en partant non pas du résultat graphique, mais de l'information à véhiculer, de sa signification et de l'intention auctoriale (Spinelli, 2006 *). »

« Le principe est pour un logiciel, de représenter les informations en fonction de leur sens, de l'information à véhiculer (par opposition à "représenter les informations sous leur forme finale, pour impression, avec mise en forme à l'identique...") (Wikipédia¹). »

L'édition WYSIWYM est l'instrument technique de la séparation entre format de création et format de publication et du formatage sémantique orienté métier.

Exemple WYSIWYM



Exemple

```
1 <exemple>
2 <texte> Un texte <important>structuré</important></texte>
3 </exemple>
```

Contre-exemple WYSIWYG



Exemple

Contre exemple : contenu et style sont **mélangés**.

Format de création orienté métier

Le format de création d'une chaîne éditoriale s'appuie sur une description métier, c'est à dire qui véhicule une sémantique liée au domaine documentaire et à l'**intention** de l'auteur, et non au traitement informatique à réaliser. Ainsi l'on n'inscrira pas dans le format de création qu'un paragraphe est en rouge et encadré (ce qui est une indication de mise en forme liée à la publication), mais l'on dira que c'est une définition (ce qui est une information liée au contexte d'usage, compréhensible par un être humain connaissant le domaine). Un algorithme informatique fixera par ailleurs les règles de transformation du langage métier vers des langages de mise en forme.

¹http://fr.wikipedia.org/wiki/What_you_see_is_what_you_mean

L'intérêt de cette approche est :

- de bénéficier d'un plus haut niveau de **sémantique** (il y a plus d'informations dans « définition » que dans « rouge », car on peut appliquer la règle « définition=>rouge », mais on ne peut pas déduire « rouge=>définition », rouge pouvant renvoyer à une autre intention).
- d'utiliser ce niveau de description pour réaliser des **calculs** (traitements informatiques) plus puissants sur le contenu
- d'assurer la **pérennité** de l'information en rendant indépendant le format de création par rapport aux évolutions technologiques (le format reste humainement interprétable)

Écrire autrement > Écriture WYSIWYM : "What you see is what you mean"



Complément

<http://aswemay.fr/co/010024.html>

5. Polymorphisme et rééditorialisation

Séparation des formats de création et de publication



Définition

La séparation entre les formats de création et de publication, également appelée séparation fond/forme ou fonds/formes consiste à utiliser un format informatique pour la **production** du contenu et un ou plusieurs autres pour sa **publication** (*Bachimont, Crozat, 2004**).

Dans le cadre d'une chaîne éditoriale numérique, le passage du format de création aux formats de publication est assuré automatiquement par un programme informatique.

L'intérêt de cette séparation est :

- de pouvoir utiliser des formats adaptés en fonction des objectifs opérationnels (créer, archiver, réutiliser, diffuser, etc.),
- de pouvoir disposer de plusieurs formats de publication à partir d'un seul format de création.

Polymorphisme



Définition

Le polymorphisme (ou séparation fonds/formes) consiste en la possibilité technique de disposer d'une **source unique** (*single sourcing* en anglais, voir *Wikipédia**) de contenu et de la transformer à volonté selon les supports et mises en formes désirés.

Le polymorphisme est un possible technologique qui reste limité dans la pratique : en effet il est rare que l'on souhaite présenter exactement la même information sous deux supports différents pour deux usages différents. Une nouvelle publication implique généralement la sélection du contenu (telle partie en plus, telle partie en moins), sa réorganisation (telle partie avant telle autre), sa remise en contexte (introduction, conclusions, transitions), etc.

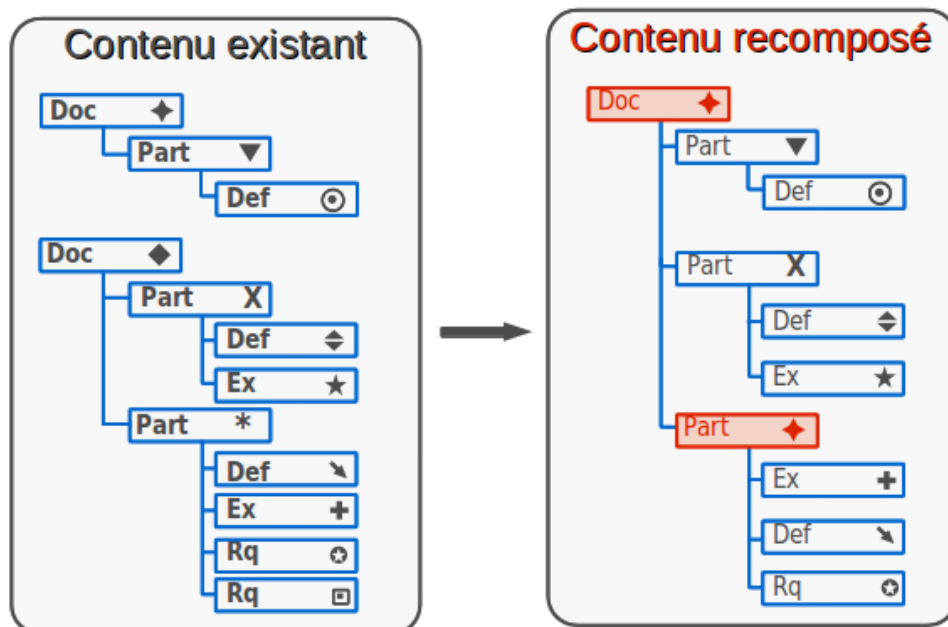
Réutilisation



Définition

La réutilisation (ou séparation scénario/contenu) consiste à profiter du découpage logique du contenu balisé avec un langage XML métier, pour appliquer des césures physiques (découpage de fichiers XML et utilisation de liens par référence ou **transclusion**).

Il devient alors possible de partager de mêmes fragments documentaires entre plusieurs documents, ce qui permet la réutilisation sans recopie.



Principe de la réutilisation : lors de la création de nouveaux documents des fragments de contenus existants peuvent être utilisés

Ré-éditorialisation



Définition

On appelle ré-éditorialisation (le terme anglais de *repurposing* étant encore plus adéquat) la remise en contexte de **fragments** issus d'un fonds documentaire, par leur **ré-agencement** au sein d'un nouveau document, leur **augmentation** par une création de contenus spécifiques et leur **publication** sur un nouveau support et/ou pour un nouveau public.

C'est une combinaison de la réutilisation et du polymorphisme.



Complément

« Finalement, le document numérique est un lieu de composition et d'assemblage de fragments documentaires différents, mais il est aussi une plate-forme à partir de laquelle il est possible d'obtenir des représentations matérielles polymorphiques. (Dupoirier, 1995, p.24) »

Chaînes éditoriales et rééditorialisation de contenus numériques



Complément

<https://hal.inria.fr/hal-00740268/document> (Crozat, 2012) *

Écriture numérique traditionnelle et écriture numérique computationnelle



Complément

<http://aswemay.fr/co/010023.html>

6. Les 6 "M"

Modèle

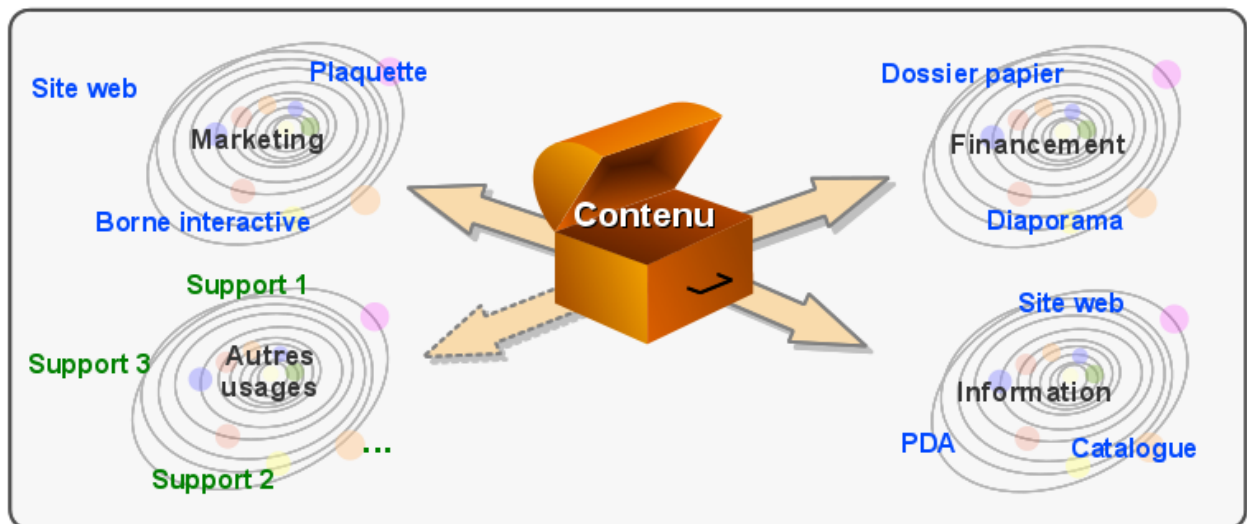
Le modèle sert de guide d'écriture en proposant à l'auteur les éléments de contenu pertinents dans le contexte de sa rédaction. Il assure une qualité standard de structuration, quelques soient les compétences éditoriales des auteurs. Il permet d'homogénéiser les publications, dans le respect des chartes graphiques et malgré les disparités d'auteurs.

Multisupports

Le multisupports permet de n'écrire qu'une seule fois ce qui est publié plusieurs fois. Il permet un gain de temps à la production du contenu, mais surtout il permet de ne maintenir qu'un seul fonds documentaire, sans redondance d'information.

Multiusages

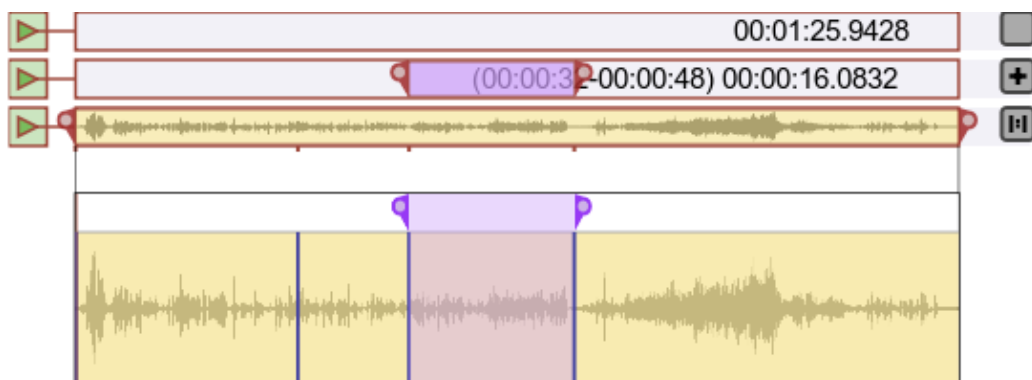
Le multiusages consiste à découper le contenu en unités documentaires autonomes et recombinaibles. Il permet de réutiliser ces unités documentaires sans recopie (par référence), afin d'adapter un discours à différents contextes.



Multisupport et multiusages : plusieurs publications en vue de plusieurs usages différents d'un même contenu

Multimédia

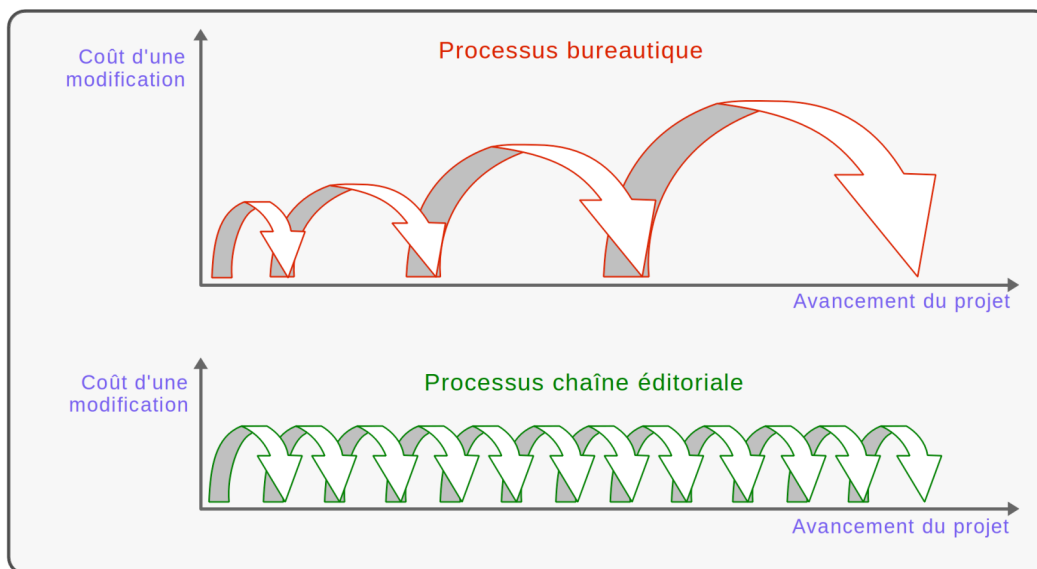
Le multimédia signifie la capacité à intégrer tous types de contenus (du texte à l'audiovisuel) et à structurer aussi bien les contenus textuels, mais aussi spatiaux (par exemple en définissant des zones dans des images) ou temporels (par exemple en définissant des segments dans un flux audio).



Exemple d'interface de structuration d'un document sonore (Scenari)

Maintenance

La maintenance des contenus est assurée dans le temps par l'indépendance technologique des formats de création (ils ne sont pas liés aux évolutions techniques comme les formats de publication, ils ne sont liés qu'aux évolutions du métier). La maintenance des publications est facilitée par le principe de génération automatique qui permet d'intervenir systématiquement sur la forme de la totalité du fonds (changement de charte graphique par exemple) sans retoucher aux contenus (la quantité de travail pour maintenir la publication est indépendante du volume de contenu).



Processus d'amélioration continue des contenus

Métier

Le modèle de document est contextuel et s'appuie sur le langage du domaine (et non sur un langage technique de publication), il est donc facilement appropriable par les acteurs. Cela permet également de configurer les interfaces aux justes besoins (plutôt que de disposer de fonctions génériques jamais toutes utilisées) et d'archiver les contenus dans un langage non technique et donc d'assurer leur accessibilité dans le temps.

Écrire autrement > Étude de cas



<http://aswemay.fr/co/010027.html>

7. Problématique : Motivations et contexte d'usage

Industrialisation : réduction des coûts et contrôle qualité

L'objectif d'une chaîne éditoriale numérique est d'instrumenter l'industrialisation d'une production documentaire. On retiendra dans le concept d'industrialisation les notions de :

- **Massification** : être capable de produire de grands volumes (plusieurs milliers de pages), malgré la rareté de compétences techniques (comme la capacité à mettre correctement en forme un document selon les canons du contexte et du support).
- **Économie d'échelle** : être capable de réduire les coûts de production et de maintenance
- **Contrôle qualité** : être capable d'assurer a priori un niveau de qualité requis (homogénéité, respect de règles éditoriales, graphiques, métiers, accessibilité, etc.)

On peut faire le parallèle de ce point de vue entre une chaîne éditoriale et une chaîne de production de produits manufacturés : l'objectif est de rationaliser pour massifier. L'approche s'oppose à une approche artisanale ou artistique (principe de l'œuvre unique).

Les limites du paradigme de la bureautique

La bureautique a permis la **démocratisation** de l'usage du numérique pour les pratiques documentaires, en rendant accessible les outils à tous. Mais dans son instrumentation, elle s'est majoritairement limitée à calquer les pratiques antérieures au numérique (comme la machine à écrire), en les améliorant, mais sans les repenser. Les raisons étant essentiellement opérationnelles, cette approche promettant de toucher le plus grand nombre le plus rapidement.

« L'approche bureautique n'apporte pas une satisfaction universelle pour la production de document en masse [...]. C'est ce constat qui est un des principaux facteurs de la motivation pour un public exigeant de chercher une alternative à la bureautique dans leurs cas d'usage (http://fr.wikipedia.org/wiki/Cha%C3%A9ne_%C3%A9ditoriale). »

La chaîne éditoriale propose de changer le paradigme fondateur, et plutôt que de copier les pratiques antérieures, elle propose une approche originale en symbiose avec les principes du numérique. Après l'objectif quantitatif atteint par la bureautique, l'enjeu est de remettre en avant des considérations qualitatives qui commencent à faire défaut dans les usages.

La séparation des métiers

Traditionnellement la production documentaire fait appel à plusieurs métiers (auteur, rédacteur, correcteur, éditeur, diffuseur, etc.). L'outil informatique, en facilitant certaines tâches (correcteurs orthographiques, outils simplifiés de mise en page, etc.), a tendu à fusionner tous les métiers en un seul « auteur-rédacteur-éditeur ». Mais, au delà de l'aspect technique, ces métiers sous-tendent des compétences qui font en général défaut à l'auteur (savoir écrire n'est pas savoir éditer), et la conséquence en est une dégradation importante des publications réalisées par les éditeurs « amateurs » que nous sommes (presque) tous.

L'objectif poursuivi par la chaîne éditoriale est de réintroduire ces métiers, en réorganisant une chaîne de production où chaque compétence est mise à profit pour ce qu'elle est.

« Une chaîne éditoriale numérique [...] est un outil ou une suite d'outils permettant d'accompagner un processus éditorial depuis l'écriture jusqu'à la publication finale. À l'inverse de la bureautique (qui fusionne l'ensemble des étapes du processus), la chaîne éditoriale les maintient séparées dans l'objectif d'offrir l'environnement le plus adéquat pour chaque type de tâche (*Scenari, la chaîne éditoriale libre* *). »



Fondamental

Ainsi la chaîne éditoriale vise d'une part à **rompre avec les techniques traditionnelles** de production prolongées par la bureautique, et d'autre part à **réhabiliter les processus professionnels** de production éditoriaux, mis à mal par la bureautique.

L'environnement de conception de chaînes éditoriales Scenari



1. Présentation de Scenari

Scenari



Définition

Scenari est un logiciel libre permettant de créer des contenus multimédia structurés selon une approche innovante : celle de la chaîne éditoriale XML.

Il permet de répondre aux enjeux actuels de la création et la gestion des documents numériques :

- réduction des coûts de production et de maintenance
- maîtrise de la qualité des publications
- multiplication des canaux de diffusion
- diversification des modalités de communication
- pérennisation de l'information

Téléchargement et documentation



Fondamental

Le site principal de Scenari est : scenari.software¹

Communauté et éditeur



Complément

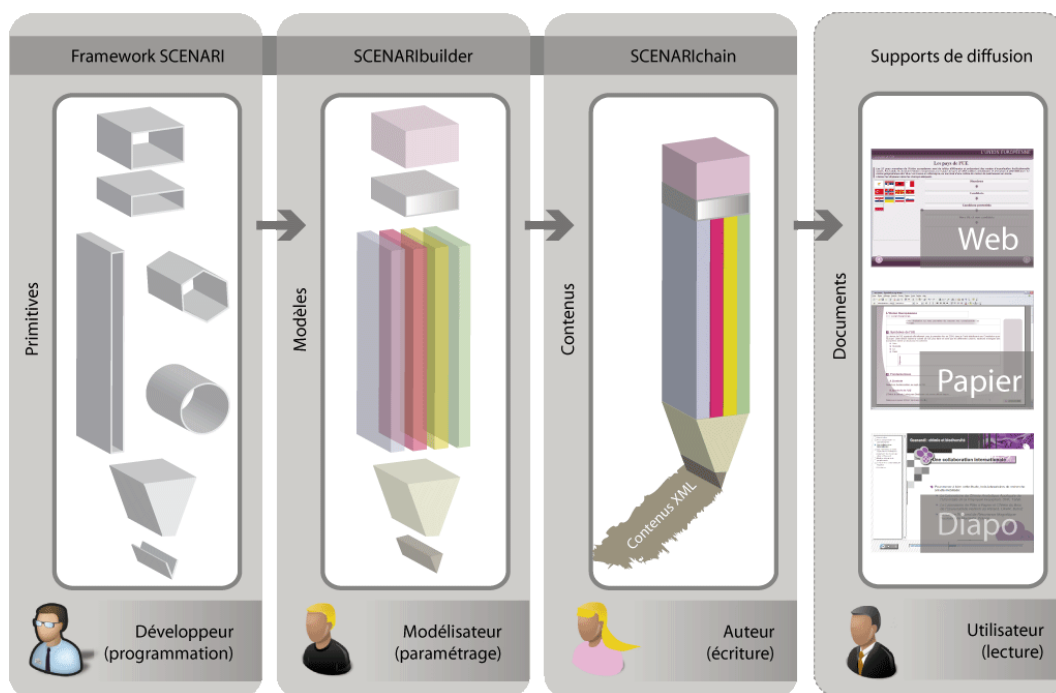
- scenari.org² est le site de la communauté.
- scenari.kelis.fr³ est le site de l'éditeur.

¹ <https://scenari.software>

² <https://scenari.org>

³ <https://scenari.kelis.fr>

2. La suite logicielle Scenari



Architecture générale de la suite Scenari

- SCENARIchain permet à des auteurs de créer des contenus et de publier des documents dans le cadre d'une chaîne éditoriale créée avec SCENARIBuilder.
- SCENARIBuilder permet à des modélisateurs de concevoir des chaînes éditoriales spécifiques à chaque métier, par l'agencement de composants paramétrables (primitives).
- Le framework Scenari est le système technique générique qui permet aux développeurs de mettre à disposition des modélisateurs les composants paramétrables pour concevoir une chaîne éditoriale dans SCENARIBuilder.

3. La communauté scenari.org

Constitution de la communauté

- Une association d'utilisateurs : scenari.org¹
- Des rencontres annuelles : [scenari.org/rencontres](https://scenari.org/rencontres2020)²
- Un forum : forums.scenari.org³

Indicateurs (2018)

- 500 inscrits à la lettre d'information hebdomadaire
- 15.000 visiteurs (en IP unique par mois)
- 3.000 téléchargements (par mois)
- 7.000 messages sur les forums
- 200 organismes utilisateurs et projets estimés

¹ <https://scenari.org>

² <https://scenari.org/rencontres2020>

³ <https://forums.scenari.org>

4. Exemples d'usages

Documents de formation

? Exemple

- Contenus pédagogiques académiques (universités, secondaire, professeurs, Canoprof)
- Contenus pédagogiques professionnels (centres de formation, grandes entreprises, armées, ministères)
- Formation-action orientée méthode (qualité)
- Formation à base de cas et jeux sérieux (topaze)

Documentation technique

? Exemple

- Documentation logicielle (dokieli)
- Documentation métier (service-public.fr)
- Catalogue (de formation, bases de connaissances)
- Documentation qualité

Audiovisuel et multimédia

? Exemple

- Webradio
- WebTV

Usages présentés aux Rencontres Scenari

? Exemple

scenari.org/rencontres2019¹

scenari.org/rencontres2018²

scenari.org/rencontres2017³

Voir aussi

+ Complément

scenari.org/co/adherents.html⁴

scenari.kelis.fr/co/references.html⁵

¹ https://scenari.org/rencontres2019/co/videos_supports.html

² <https://scenari.org/rencontres2018>

³ <https://scenari.org/rencontres2017>

⁴ <https://scenari.org/co/adherents.html>

⁵ <https://scenari.kelis.fr/co/references.html>

Glossaire



XSL-FO

XSL-FO (FO pour Formatting Objects) est la seconde partie du standard W3C « Extensible Stylesheet Language Family ». Il propose un langage XML de mise en forme de documents imprimables.

Exemple d'extrait de code FO :

- `<fo:block font-family="Times" font-size="12pt">Ceci est un paragraphe en police Times de taille 12pt</fo:block>`.

Bibliographie



Stéphane Crozat, *Scenari, la chaîne éditoriale libre*, Accès Libre, Eyrolles, 2007.

Bruno Bachimont, Stéphane Crozat, *Instrumentation numérique des documents : pour une séparation fonds/forme*, revue I3, vol 4, num 1, 2004. [http://archivesic.ccsd.cnrs.fr/sic_00001017.html]

Bruno Bachimont, *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*, Lavoisier, Hermès, 2007

Crozat Stéphane. 2012. *Chaînes éditoriales et rééditorialisation de contenus numériques*. in *Le document numérique à l'heure du web*. ADBS editions. pp179-220. <https://hal.inria.fr/hal-00740268>.

Webographie



Sylvain Spinelli, *WYSIWYM (what you see is what you MEAN)*, <http://scenari-platform.org/trac/scenari/wiki/WYSIWYM>, 2006.

Single Source Publishing, Wikipédia, consulté en mai 2008 [http://en.wikipedia.org/wiki/Single_source_publishing]