



HAL
open science

Chaînes éditoriales et rééditorialisation de contenus numériques

Stéphane Crozat

► **To cite this version:**

Stéphane Crozat. Chaînes éditoriales et rééditorialisation de contenus numériques. Lisette Calderan and Pascale Laurent and Hélène Lowinger and Jacques Millet. Le document numérique à l'heure du web, ADBS, pp.179-220, 2012, Sciences et techniques de l'information, 978-2-84365-142-7. hal-00740268

HAL Id: hal-00740268

<https://inria.hal.science/hal-00740268>

Submitted on 12 Jul 2013

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.

Chaînes éditoriales et rééditorialisation de contenus numériques

Stéphane Crozat

Ingénieur et docteur en informatique, Stéphane Crozat est enseignant-chercheur à l'Université de technologie de Compiègne (UTC) depuis 2002. Il assume la direction technologique de l'unité Ingénierie des contenus et savoirs, et l'enseignement des bases de données, des datawarehouses et de l'ingénierie documentaire au Génie informatique. Co-inventeur en 1999 de l'environnement de conception de chaînes éditoriales Scenari et animateur de la communauté scenari-platform.org des utilisateurs de cet outil, il a publié Scenari, la chaîne éditoriale libre (Eyrolles, 2007), premier livre grand public sur cette thématique. Spécialiste de l'ingénierie des documents numériques, il assure le lien entre la recherche théorique, les usages réels et les développements informatiques. stephane.crozat@utc.fr

1 As we may write

Pendant la seconde guerre mondiale, Vannevar Bush coordonne la recherche américaine et se trouve confronté à la gestion d'une colossale production scientifique. En 1945, au sortir de la guerre, il écrit un article devenu célèbre, *As we may think*, dans lequel il réfléchit aux moyens techniques permettant d'améliorer le stockage et la consultation des gigantesques volumes documentaires auxquels il a été confronté [11]. Dans la lignée de cette anticipation fondatrice, le numérique a d'abord été la réponse aux problèmes de la *gestion* documentaire, fondamentalement le stockage et la restitution.

Notre propos, dans ce chapitre, sera de montrer que le numérique est aussi devenu le support d'une autre façon de *produire* du document. En effet, le support numérique impose à l'écriture ses propriétés techniques et transforme la nature fondamentale du document. Nous proposerons en particulier d'étudier la *rééditorialisation*, au sens de la création de documents nouveaux à partir de documents existants. Nous défendrons qu'elle est une *tendance* inhérente de l'écriture numérique ; que la rééditorialisation est finalement *la façon dont nous pourrions écrire* avec le numérique.

1.1 Raison graphique et raison computationnelle

1.1.1 L'invention de l'écriture a donné naissance à la rationalité moderne : la raison graphique

Goody a montré que l'invention de l'écriture a modifié les schèmes de représentation de la connaissance, jusque-là orale, donnant naissance à une raison graphique. Le passage de l'oral à l'écrit n'est pas simplement un changement de support, c'est une reconfiguration du système technique de production et de manipulation qui agit sur la nature même de la connaissance. « Même si l'on ne peut pas réduire un message au moyen matériel de sa transmission, tout changement dans le système des communications a nécessairement d'importants effets sur les contenus transmis. [21, p. 46] »

« L'écriture a une importance décisive, non seulement parce qu'elle conserve la parole dans le temps et dans l'espace, mais aussi parce qu'elle transforme le langage parlé : elle en extrait et abstrait les éléments constitutifs ; ainsi la communication par l'œil engendre des possibilités cognitives nouvelles par rapport à celle qu'offre la communication par la voix. [*ibid.*, p. 221] » Goody développe des exemples concrets comme la liste ou le tableau et montre que ces structures sont essentiellement graphiques, elles n'existent que parce qu'un support leur donne une permanence dans le temps et une possibilité de spatialisation. Or le travail intellectuel d'écriture et de lecture via de telles structures spatiales permet de faire émerger de nouvelles connaissances, à partir de représentations qui ne peuvent être formulées oralement.

Prenons un exemple trivial avec le tableau ci-après. La visualisation graphique de ce tableau montre la corrélation entre la baisse du chiffre d'affaires et la hausse du bénéfice. Une lecture orale du tableau ne permet pas l'émergence aussi évidente de cette information.

Année	Chiffre d'affaires	Bénéfice
2004	123.315	5.154
2005	115.247	7.156
2006	114.265	8.245
2007	112.250	8.300

À l'autre extrême de cet exemple un peu simpliste, on observera que la tradition critique et la pensée scientifique se construisent grâce à l'écrit, à la possibilité d'accumuler le savoir, du retour critique, du scepticisme ; que la logique (raisonnement formel sur des lettres), l'algèbre et le calcul (raisonnement sur les nombres) n'existent qu'avec l'écrit ; ou même que la philosophie dépend de la possibilité de raisonnements logiques inhérents à l'écriture ou du retour philologique aux textes [*ibid.*, p. 96-97]. Les mathématiques sont inconcevables sans la formalisation écrite, et avec elles la science moderne, dont le développement se révèle lié à la naissance et aux évolutions de la communication écrite : l'écriture (Babylonie), l'alphabet (Grèce), l'imprimerie (Europe occidentale) [*ibid.*, p. 106].

Goody montre donc que, au-delà des conséquences herméneutiques et cognitives liées au changement de support, c'est l'ensemble du rapport à la connaissance qui en est transformé : « Lorsque je parle de l'écriture en tant que technologie de l'intellect, en particulier, je ne pense pas seulement aux plumes et au papier, aux stylets et aux tablettes [...], mais aussi à la formation requise, l'acquisition de nouvelles compétences motrices, l'utilisation différente de la vue, ainsi qu'aux produits eux-mêmes, les livres qui sont rangés sur les étagères des bibliothèques. [22, p. 194] »

1.1.2 Le numérique est le ferment d'une nouvelle rationalité en construction : la raison computationnelle

De même que l'écrit a permis le passage du temporel au spatial par projection de la parole, le support numérique apporte de nouvelles formes de représentation de l'information, basées sur le calcul : l'ordinateur ne traite que des séquences binaires qui, par des manipulations calculatoires, deviennent des signes sur un support tel que l'écran. C'est cette propriété du support numérique qui est fondamentale en tant qu'elle propose de nouvelles modalités d'inscription. Et ces nouvelles modalités induisent également la constitution de modes de représentation nouveaux, comme la raison graphique en leur temps. Bachimont donne en exemple la couche ou le réseau [4, p. 104].

« L'hypothèse que nous formulons est que l'informatique, sous la forme des systèmes formels automatiques, fournit précisément un nouveau type de support, les supports dynamiques, auquel doit correspondre un type spécifique de synthèse, et par conséquent une rationalité spécifique, que nous proposons de baptiser "raison computationnelle". [3, p. 16] » « [...] Les anthropologues, en particulier Jack Goody, évoquent une raison graphique pour expliciter le fait que l'écriture induit un mode de pensée particulier et un rapport au monde spécifique. Nous parlerons, quant à nous, d'une raison computationnelle pour expliciter le fait que nous pensons différemment avec les outils numériques. [5, p. 71] »

Il y a une relation fondamentale entre les supports que nous utilisons pour penser et la pensée elle-même. Goody l'a montré pour l'écrit, cette technique de synthèse spatiale de l'information est constitutive de la pensée scientifique rationnelle. Bachimont montre quand à lui que le support numérique implique une nouvelle forme de synthèse, dans l'espace du calcul formel.

Par conséquent, une nouvelle forme de rationalité est en émergence, nos sociétés actuelles en vivent la constitution en ce moment même, et il y a des enjeux scientifiques et pragmatiques – de savoirs et de pouvoirs, au sens de Goody [22] – à comprendre et anticiper cette nouvelle forme de rationalité.

1.2 Tendance technique et ingénierie documentaire

1.2.1 État de littératie numérique restreinte

Nous avons vu avec Goody et Bachimont que le numérique est une évolution technique qui reconfigure les moyens et les modes de penser. Mais leur caractère émergent limite les possibilités d'une approche anthropologique d'observation et de description. Goody propose la notion de « littératie restreinte » pour désigner l'état relatif à une écriture restreinte, notamment lorsqu'elle en est à ses débuts, que le système n'est pas pleinement exploité [22, p. 21]. Si le concept a été formulé pour désigner les périodes précoces des systèmes d'écriture, il nous semble raisonnable de considérer que l'état précoce de l'écriture numérique nous place dans un état de *littératie numérique restreinte*.

Nous faisons alors l'hypothèse qu'il n'est pas encore possible de décrire la nature d'une raison computationnelle en genèse, en observant les pratiques ou les traces d'écriture numérique, comme les anthropologues ont pu le faire *a posteriori* avec l'étude de l'apparition de l'écriture dans les sociétés orales.

1.2.2 L'évolution technique répond à une tendance propre

Leroi-Gourhan montre que la technique possède une dynamique propre qui s'impose aux sociétés humaines, qu'il nomme *tendance*. Cette tendance est universelle, elle relève des lois de la physique, et tout objet technique se façonne, s'architecture, s'équilibre, en réponse à des contraintes exogènes. En interagissant avec la culture intérieure du groupe ethnique inscrit dans son milieu extérieur (situation géographique, géopolitique, etc.), la tendance conduit à des faits techniques singularisés par les propriétés contextuelles des milieux intérieurs et extérieurs : « La tendance qui, par sa nature universelle, est chargée de toutes les possibilités exprimables en lois générales, traverse le milieu intérieur [...], elle rencontre le milieu extérieur [...], et au point de contact entre le

milieu intérieur et le milieu extérieur se matérialise cette pellicule d'objets qui constituent le mobilier des hommes. [23, p. 338] »

Stiegler explique ainsi que la technique dispose d'un *dynamisme* fonctionnant selon une logique propre. L'évolution technique est alors analogique à l'évolution génétique des espèces vivantes, elle s'inscrit dans une combinatoire et une sélection autonomes, et non plus dans une construction déterminée intentionnellement par l'être humain, qui se trouve relégué à un rôle d'*exhumation*. « Le concept de tendance technique s'oppose à cette illusion ethnocentrique [...] il n'y a pas de génie de l'invention, ou du moins, il ne joue qu'un rôle mineur dans l'évolution technique. [32, p. 57] »

Simondon [31] radicalise cette autonomisation de la genèse technique par les concepts de *processus de concrétisation* et d'*individuation* de l'objet technique. « Dans l'explication de l'évolution technique par le couplage de l'homme à la matière, traversée par la tendance technique, une part essentielle de celle-ci, provenant du milieu intérieur ethnique comme intention, reste anthropologiquement déterminée. Chez Simondon, ce milieu intérieur se dilue. Il n'y a plus de source anthropologique de la tendance. L'évolution technique relève pleinement de l'objet technique lui-même. L'homme n'est plus l'acteur intentionnel de cette dynamique. Il en est l'opérateur. [32, p. 80] »

1.2.3 Inventions et innovations documentaires

Leroi-Gourhan, Simondon et Stiegler nous disent que toute évolution technique dispose d'une dynamique propre, et que les processus d'invention sont des processus de découverte de ces tendances évolutives. Un enjeu de la recherche en ingénierie du document numérique est alors d'inventer – pour les exhumer – des formes documentaires qui ouvrent un nouveau champ du possible, pour tenter d'en anticiper les incidences cognitives et sociétales en genèse : « Il est apparu que le milieu intérieur produisait des objets, non d'une manière automatique [...], mais par des intentions successives, qui se traduisent en objets de plus en plus perfectionnés, en inventions progressives. [23, p. 397] »

Notre démarche consistera à concevoir des systèmes techniques de production documentaire adressant ces formes nouvelles - on parlera de *chaînes éditoriales numériques* -, pour être en mesure de les mettre en usage. L'idée est que les propositions en phase avec la tendance technique auront tendance à survivre et à se généraliser, tandis que les autres disparaîtront. « L'invention est un acte d'intelligence qui coïncide avec la tendance. [*ibid.*, p. 377] »

1.3 Documentarisation de la société industrielle

Le collectif Pédaque a montré que le document avait connu deux révolutions [28]. Chacune de ces révolutions se construit dans un contexte dual d'invention technique et de modernisation sociale, dont elles sont constitutives, à la fois conséquence et cause. La première révolution est liée à la révolution industrielle et conduit à une documentarisation de la société. La seconde révolution est liée à l'avènement du numérique et conduit à une redocumentarisation de la société, c'est-à-dire une réinvention de la nature documentaire.

1.3.1 Première révolution documentaire et organisation documentaire

La première révolution naît avec l'imprimerie, qui rend possible la massification documentaire, et le siècle des Lumières, qui prône la démocratisation de la culture et de la connaissance. La révolution industrielle et l'avènement de la société moderne prolongent ce mouvement en dotant le document d'une fonction utilitaire, dont la vocation est de rendre un service, et en en systématisant le recours dans tous les pans de la société. « L'imprimé serait directement associé à la première modernisation, celle qui a permis l'esprit scientifique, la rupture avec les traditions de l'Ancien régime, l'expérimentation et sa validation à travers des comptes rendus détaillés comme critère de la scientificité, celle aussi qui débouche progressivement sur la reconnaissance des autorités et en même temps des auteurs et des États-nations. Une bonne part des relations dans les sociétés dites "modernes" sont fondées sur et cimentées par la stabilité du document papier et sa reproductibilité industrielle à l'identique (effet de série) ou encore sa permanence sécurisant les contrats, les règles et les identités. [28, p. 3-4] » « L'écriture est maintenant condition interne de toute organisation politique et économique et non seulement de la vie intellectuelle. [21, p. 82] »

Le document devient un objet comme un autre, partie prenante de la société de production. On assiste alors à une généralisation de la *documentarisation* : l'information – scientifique, technique, juridique, sociale, etc. – est constituée systématiquement en documents. La massification atteint des proportions telles, la dépendance documentaire devient si urgente, qu'elles conduiront à la naissance au XX^e siècle des problématiques d'*organisation documentaire* que Otlet [26], Bush [11], Briet [9] ou Buckland [10] ont pratiquée et théorisée (archivage, identification, classification, recherche, etc.).

Définition : Document

On définira avec Bachimont un document comme une inscription de contenus sur un support *pérenne*, établie dans un *contexte éditorial* [6].

Un contenu est une forme d'expression pourvue d'une valeur culturelle associée à un véhicule matériel, il exprime une signification et suscite une interprétation ; une inscription est un contenu fixé sur un support matériel, tel qu'il lui apporte une permanence dans le temps ; un contexte éditorial est l'association d'un contexte de production et d'un contexte de réception.

Définition : Documentarisation

1. La documentarisation définit d'abord le traitement singulier permettant de constituer un contenu en document : « Elle consiste à doter [les] supports d'attributs spécifiques permettant de faciliter (i) leur gestion parmi d'autres supports, (ii) leur manipulation physique [...] et enfin, (iii) l'orientation des récepteurs » ; « ces attributs doivent permettre au document de circuler à travers l'espace, le temps, les communautés d'interprétation, pour tenter de prolonger les transactions communicationnelles initiées par ses réalisateurs. [33] »

2. Par extension la documentarisation désigne aussi chez Pédaque l'omniprésence documentaire dans l'organisation sociale moderne [28]. On parlera de *documentarisation de la société* pour se différencier du premier sens.

1.3.2 La gestion documentaire : d'Otlet à la gestion électronique des documents

La documentarisation de la société a conduit à la gestion documentaire. Paul Otlet, au début du XX^e siècle, construisait le Mundaneum, bibliothèque universelle ayant vocation à archiver la connaissance du monde, et le système de classification décimale universelle destiné à la cataloguer [24]. On voit ensuite émerger les nouvelles disciplines de la documentation [9] ou *librarianship* [10], puis avec l'arrivée de l'ordinateur celle de *l'information science*, au sens du regroupement de l'informatique et du documentaire [*ibid.*, p. 14]. La vision première d'Otlet reste prégnante sur l'informatique documentaire, l'ordinateur est d'abord un outil de catalogage, grâce aux bases de données documentaires, puis de stockage, par la numérisation progressive des fonds.

La gestion électronique des documents (GED) fusionne ces deux fonctions, et se déploie dans de nombreuses organisations comme moyen d'organiser les fonds documentaires devenus électroniques [18]. Si, à ce stade, l'informatique reste relativement périphérique à l'acte d'écriture lui-même, la généralisation du traitement de texte – et plus largement des outils bureautiques – comme moyen de production documentaire est le virage qui conduit à une nouvelle ère documentaire. En effet, même si cette première génération d'outils cherche à calquer les techniques d'écriture classiques (typiquement la machine à écrire et l'imprimerie), elle ouvre sur la possibilité d'une autre écriture, *numérique*.

1.4 Redocumentarisation de la société numérique

La seconde révolution documentaire naît avec le numérique et les transformations sociétales de la seconde moitié du XX^e siècle, Pédaque adoptant pour les qualifier le terme de « postmoderne » [28, p. 3]. Le document devenu numérique est fortement réinterrogé, dans sa nature intrinsèque ainsi que dans ses usages, le web étant un avatar prégnant de son évolution.

La première rupture fondamentale imposée par le numérique au document est la séparation entre la forme d'inscription, une ressource binaire sur un support d'enregistrement, et la forme de lecture, une manifestation sémiotique sur un dispositif de restitution, la seconde étant *calculée* à partir de la première par l'intermédiaire de l'exécution d'un programme. *Ce qu'on lit n'est plus ce qui a été écrit*. « Autrement dit, un document numérique n'a pas de mémoire. Il est d'emblée falsifiable et possiblement falsifié. Ainsi, l'essence du numérique, ce que, à l'instar de Roland Barthes, nous appelons le noème du numérique, est-elle : "ça a été manipulé" [5, p. 34] »

« La manifestation la plus évidente du changement est donc la perte de la stabilité du document comme objet matériel et sa transformation en un processus construit à la demande, qui ébranle parfois la confiance que l'on mettait en lui. [27, p. 2] »

La seconde rupture est liée aux libérations de l'acte auctorial grâce aux machines à écrire que sont les terminaux numériques, et de l'acte éditorial grâce aux machines à publier que sont les réseaux. Le numérique désacralise le concept de document tel qu'il a été fondé par la première révolution. Une part grandissante de nos pratiques conduit à la création et la consommation de productions numériques, que nous qualifions volontiers de documents.

On assiste donc à une nouvelle documentarisation, une *redocumentarisation*, qui est à la fois une *surdocumentarisation* et une *dédocumentarisation* : une production exponentielle de ressources numériques, qui ne sont plus vraiment des documents, privées de leur pérennité par leur nature calculatoire et au contexte éditorial relâché par leur nature ordinaire. « Nous emploierons le terme de "redocumentarisation", le préfixe "re" suggérant à la fois un retour sur une documentarisation ancienne et une révolution documentaire. [28, p. 4] »

Définition : Document numérique

1. L'une des définitions du document numérique proposées par Pédaque est : « Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs. [27, p. 10] » Cette proposition tend à maintenir les principes de pérennité et de contexte éditorial. Mais la trichotomie entre données, structures et règles est en décalage fondamental avec l'inscription pérenne qui définissait le document traditionnel et assurait le contrat entre contextes de production et de réception : la médiation du calcul introduit une incertitude théorique. Un document numérique serait donc un triptyque « données, structures, règles » qui cherche à se doter des caractéristiques du document, tout en ayant admis qu'il n'en est fondamentalement pas un.

2. Une autre définition, insistant plus radicalement sur la coupure calculatoire, est proposée par Bachimont : « Le document numérique est [...] un complexe documentaire composé de ressources enregistrées, d'un dispositif de reconstruction du contenu pour l'afficher dans une forme perceptible et intelligible, et finalement des vues reconstruites. [5, p. 223-224] » La ressource binaire n'est pas le document, car elle n'est pas un contenu, elle est vide de sens *a priori* [7, p. 203-204]. La forme calculée n'est pas non plus le document, car elle n'est pas une inscription, sa construction dynamique n'est pas pérenne. Si l'on considère une fixation de cette forme, par exemple une impression, on a bien un document, mais qui n'est plus numérique.

Le document numérique finalement n'existe pas, la locution est oxymorique. Il ne peut exister que des constructions numériques dont le traitement calculatoire permet de simuler un ordre documentaire. Et plus généralement il existe des constructions numériques, qui ne sont pas des documents, mais qui sont propres à engendrer des situations interprétatives homologues à celles rencontrées avec des documents. Nous proposerons dans la seconde partie le concept de *document-dossier* pour instrumenter ces constructions qui se substituent au document.

Définition : Redocumentarisation

1. Le terme de redocumentarisation peut être employé, avec Salaün, pour désigner la transformation d'un document non numérique afin qu'il puisse être traité par un système informatique : numérisation, description, modélisation... [30]. Il s'agit alors d'une informatisation du document, au sens de Bachimont [5].

2. La redocumentarisation désigne surtout chez Pédaque la mutation documentaire survenue suite à l'avènement du numérique [28] ; elle est donc à entendre comme une seconde documentarisation sociétale concomitante de la post-modernité. On parlera de façon analogue de *redocumentarisation de la société*. Étant encore en construction, cette redocumentarisation n'est pas précisément caractérisable aujourd'hui.

3. Redocumentarisation est aussi utilisée dans une acception différente par Zacklad : « Redocumentariser, c'est documentariser à nouveau un document ou une collection en permettant à un bénéficiaire de réarticuler les contenus sémiotiques selon son interprétation et ses usages. [34, p. 3] » Nous n'utiliserons pas le terme dans ce sens, pour lui préférer alors celui de *rééditorialisation*, que nous définirons par la suite.

Le rôle joué par le numérique dans nos pratiques documentaires s'est essentiellement cantonné, jusque ces dix dernières années, à prolonger la documentarisation de la société, par les moyens de massifier et d'organiser la production documentaire. Mais la tendance du numérique conduit à une redocumentarisation, et une réinvention de nos modes d'écriture. *Tout comme l'écriture ne s'est pas limitée à être une transcription de l'oral, l'écriture numérique ne s'arrête pas à être une numérisation de l'écrit.*

Nous proposons d'étudier par la suite la notion de rééditorialisation, dont nous pensons qu'elle est fondamentalement au cœur de cette nouvelle écriture en genèse.

1.5 Rééditorialisation des contenus numériques

1.5.1 La tendance technique du numérique : fragmentation / recombinaison et désémantisation / resémantisation

Dans la filiation de Leroi-Gourhan, Bachimont propose d'étudier la tendance technique du numérique, dont la caractérisation permet d'anticiper l'avenir : « Comprendre l'impact que peut avoir le numérique dans un domaine donné revient dès lors à savoir caractériser sa tendance technique. Si l'on demeure bien évidemment incapable de prédire l'avenir, il devient possible de l'anticiper dans le sens où l'on s'attend à une évolution des choses conforme à certaines contraintes ou propriétés des éléments intervenant dans cette évolution. [5, p. 37] »

Bachimont caractérise l'essence du numérique par les propriétés de *discrétisation* et de *manipulation*, puisque le numérique rapporte le contenu à un codage manipulable formellement. Il en déduit que la tendance technique du numérique s'exprime par un couple *fragmentation / recombinaison* d'une part et *désémantisation / resémantisation* d'autre part.

resémantisation d'autre part. Tout contenu numérique est vidé de son sens par le principe même du codage, fragmentable par le caractère discret de ce codage et recombinaison par les possibilités de manipulation ouverte : « Puisque le numérique repose sur la discrétisation et la manipulation, il en résulte que toute réalité touchée par le numérique sera réduite en unités vides de sens sur lesquelles des règles formelles de manipulation seront appliquées. La discrétisation fragmente la réalité considérée, la manipulation la recombine [...] De même, la discrétisation implique une rupture avec la sémantique propre aux contenus manipulés. Il en résulte donc une désémantisation, une perte de sens, qu'il faudra gérer à travers un processus de resémantisation explicitement assumé. [*ibid.*] »

Cette tendance technique se manifeste dans les pratiques d'écriture numérique par la généralisation des processus de rééditorialisation, c'est-à-dire incluant une sélection et un ré-agencement de fragments documentaires existants, avec ou sans transformation associée.

1.5.2 La rééditorialisation est une tendance de l'écriture numérique

De nombreux processus d'écriture font traditionnellement appel à la réécriture au sens où ils s'inspirent de documents antérieurs, voire les copient. Dans les systèmes éditoriaux non numériques, la rééditorialisation est essentiellement un travail de recopie, fondamentalement peu différent d'un travail de production originale.

Le numérique a amplifié et *naturalisé* ces pratiques dans la mesure où les processus de fragmentation / recombinaison et de désémantisation / resémantisation lui sont constitutifs. Les systèmes d'écriture numériques proposent des fonctions d'écriture, qui répondent à des fondamentaux, ou *tropismes*, de l'écriture numérique – comme la manipulabilité, l'abstraction, l'adressabilité, l'universalité et le clonage [14] – et qui *automatisent* la pratique de la rééditorialisation : ce qui était marginal devient la norme.

Définition : Rééditorialisation

Le terme de rééditorialisation est un néologisme qui a émergé dans le domaine du document numérique pour désigner le processus consistant à reconstruire un nouveau document à partir d'archives. La construction de ce mot tente une première synthèse entre les concepts d'*édition* au sens de publication d'une œuvre, d'*éditorialisation* au sens d'expression d'un point de vue propre, de *réédition* au sens de nouvelle proposition de lecture. Elle tente une seconde synthèse entre les fonctions d'*éditeur*, celui qui met en forme et diffuse, et d'*auteur*, celui qui écrit, fonctions qui tendent à se mêler dans le contexte du numérique. La rééditorialisation est donc la publication d'une œuvre originale dans son point de vue, sa forme, sa scénarisation, à partir de contenus qui ne le sont pas tous.

Rééditorialisation est la traduction admise du terme anglais *repurposing*, « donner un nouvel objectif », que l'on pourrait traduire par « réobjectivation », au sens commun de redonner un objectif, et au sens de Bachimont de poser une marque auctoriale objective à laquelle va se confronter l'appropriation par le lecteur [5, p. 166]. Le *Grand Dictionnaire terminologique*¹ traduit *repurposing* par « réorientation des données » et le définit comme la « réutilisation des données dans une autre application ou sur un support différent ». On pourrait parler de « réorientation documentaire » pour spécialiser cette définition dans notre contexte, ou de « *réobjectivation documentaire* ».

Un processus de rééditorialisation mobilise en général plusieurs étapes, dont essentiellement :

- une *sélection* des archives : recherche et sélection de documents ou extraits de documents ;
- une *déconstruction* de ces archives : découpage et sélection des fragments pertinents pour le nouvel objectif de communication ;
- une *transformation* de ces fragments : réécriture de partie de ces fragments ;
- une *production* originale : ajout de fragments spécifiques au nouveau contexte ;
- une *reconstruction* documentaire ou scénarisation : ordonnancement des fragments et articulation (introduction, transitions, conclusion, etc.) ;
- une *publication* : homogénéisation de l'ensemble dans une mise en forme unifiée et adaptée au nouveau contexte, ajout des métadonnées de publication (date, auteur, etc.).

Définition : Chaîne éditoriale numérique

1. Nous avons antérieurement défini déjà la chaîne éditoriale numérique comme un procédé technologique et méthodologique consistant à réaliser un modèle de document, à assister les tâches de création du contenu et à automatiser la mise en forme [13, p. 2]. Cette définition tendait à opposer traitement de texte WYSIWYG et chaîne éditoriale WYSIWYM [*ibid.*, p. VI], le premier cherchant à reproduire les pratiques d'écriture traditionnelles tandis que la seconde vise à proposer de nouvelles formes d'écriture propres au numérique.

2. Nous en proposons ici une définition élargie : une chaîne éditoriale numérique est un système de production documentaire cherchant à instrumenter des fonctions d'écriture originales (dans ce que le numérique lui apporte et lui impose de spécifique) en prise avec la tendance du numérique, en particulier afin d'automatiser la rééditorialisation.

¹ Québec, <http://www.granddictionnaire.com>

2 La domestication de la rééditorialisation sauvage

L'objet de l'ingénierie documentaire, après s'être surtout intéressé à la gestion électronique des documents en réponse à la documentarisation de la société, est à présent de faire émerger de nouvelles formes d'écriture, afin de comprendre *comment nous pourrions écrire* avec le numérique. Il est nécessaire de concevoir de nouveaux systèmes de production documentaire, des *chaînes éditoriales numériques*, afin d'inventer et de critiquer ces nouvelles formes. Nous avons montré que l'écriture numérique obéit à une tendance technique qui se manifeste notamment par la généralisation des pratiques de *rééditorialisation*.

L'enjeu est alors de chercher à *domestiquer* ces pratiques, encore sauvages, afin de comprendre, anticiper et s'inscrire au mieux au sein de la raison computationnelle qui se construit avec elles, et avec nous.

Préambule : clonage et rééditorialisation sauvage

Le numérique permet de reproduire à l'infini, sans perte ni coût (autre que celui du stockage), toute séquence de code. Le tropisme qui a certainement marqué le plus immédiatement l'écriture numérique est le clonage [14], matérialisé par la fonction du copier/coller, qui nous intéressera ici surtout au sens de *dupliquer et modifier* : « L'écriture numérique propose de dupliquer un contenu ou une partie d'un contenu en vue d'en produire un nouveau qui lui ressemble². »

Or, utilisé comme technique de rééditorialisation, le clonage pose deux problèmes : la redondance et l'incohérence, et l'oubli de la source.

Le problème posé par la *redondance* est bien connu en informatique, il est notamment à l'origine des bases de données. Une même information numérique (modifiable), inscrite à plusieurs endroits, tendra à diverger lors des processus de maintenance du contenu et, inévitablement, deux fragments sensément identiques porteront, à un certain moment, des informations différentes. La conséquence théorique est la perte de la valeur de vérité associée à ces fragments : le système informatique ne sait pas décider quel fragment est correct (celui qui porte l'information à véhiculer effectivement) et quel fragment ne l'est pas (celui dont l'information n'a pas ou mal été mise à jour). Une intervention humaine experte est alors nécessaire (choix du bon fragment, correction, etc.). C'est le problème de l'*incohérence* des données.

Un second problème est lié à la nature documentaire des informations clonées. Un document est lié par définition à un contexte éditorial ; or la copie d'un fragment de document conduit à une *coupure avec la source* originelle, et donc à la perte de ce contexte éditorial. Des métadonnées peuvent en théorie être conservées avec le fragment pour conserver un lien à la source, mais les formats ne l'imposent pas et les outils le favorisent peu.

Nous présentons dans cette partie quatre techniques d'écriture propres au numérique destinées à favoriser la rééditorialisation tout en évitant ou contrôlant le clonage :

- le polymorphisme : possibilité de donner plusieurs formes au contenu en fonction du support de restitution ;
- la transclusion : possibilité de réutiliser des fragments documentaires par référence, sans recopie ;
- la dérivation : possibilité de réutiliser des fragments par duplication et modification, mais en contrôlant les évolutions conjointes de la source et de ses copies ;
- la déclinaison : possibilité de programmer *a priori* des variations dans le contenu en fonction de paramètres éditoriaux.

Ces quatre techniques seront illustrées par des exemples d'implémentation avec la chaîne éditoriale XML Scenari [13].

2.1 Polymorphisme et multisupports

2.1.1 Problématique : le multisupports

Un premier cas de mobilisation massive du clonage qui s'est répandu avec le numérique répond à la problématique du multisupports. Afin de diffuser un même contenu sur plusieurs supports différents (papier et web typiquement), il est nécessaire de copier l'information afin de la formater techniquement et graphiquement en fonction de chaque support.

Le multisupports s'accommode mal d'une simple copie du contenu, il est nécessaire de procéder à des *transformations*, afin de l'adapter aux exigences matérielles du support. On citera par exemple le transcodage (format vidéo, niveau de résolution des images, etc.), la mise en page (l'organisation du contenu dépend du format du support de restitution), les logiques de navigation (tables des matières, plans interactifs, etc.).

² <http://precip.fr/map>

Définition : Multisupports

On appelle multisupports la documentarisation d'un même contenu sur plusieurs supports différents, en mobilisant des formats techniques différents et/ou des mises en forme graphiques différentes, par exemple : un format PDF pour l'impression, un format et un style HTML pour la diffusion web, un second style HTML pour la présentation diaporama...

2.1.2 Principe du polymorphisme

Le polymorphisme est une technique d'automatisation de la transformation du codage d'un contenu en un autre codage afin de remplir des objectifs éditoriaux différents. Le principe du polymorphisme est de se baser sur un *format pivot*, qui n'a pas en général de vocation documentaire (il ne sera pas documentarisé tel quel, il ne sera pas produit ni consulté tel quel), mais qui a vocation à être *transformé*. Ce format pivot est donc choisi pour ses bonnes propensions à être transformé, pour sa disponibilité manipulative. Il est donc pour cela plus efficace qu'un format documentaire qui n'a pas été pensé en ce sens. En particulier ce format pivot s'appuiera sur un langage de description permettant d'*abstraire* les formes visées, intégrant et généralisant leurs différentes caractéristiques. Ce langage d'abstraction reposera sur des propriétés de structuration dites logiques des contenus [1], et s'implémentera grâce à un schéma XML.

On nommera *forme génératrice* (FG) le format pivot et *formes publiées* (FP) les formats documentaires obtenus via des fonctions de *transformation* automatique (T) de FG, telles que : $FP_i = T_i(FG)$. On proposera également de poser des *formes éditables* (FE) et des *formes de référence* (FR) telles que $FE_j = T_j(FG)$ et $FR_k = T_k(FG)$. Une FE est une forme qui permet la modification de FG. Une FR est une forme permettant de relire l'ensemble des informations structurées par la FG (contenu, métadonnées, structure, typage, etc.) en vue de sa *validation* [15].

Si l'on pose des rôles d'auteur (celui qui produit le contenu), d'éditeur (celui qui documentarise) et de lecteur (celui qui consomme le contenu) alors on peut dire que les FE s'adressent aux auteurs, les FR aux éditeurs et les FP aux lecteurs [figure 1].

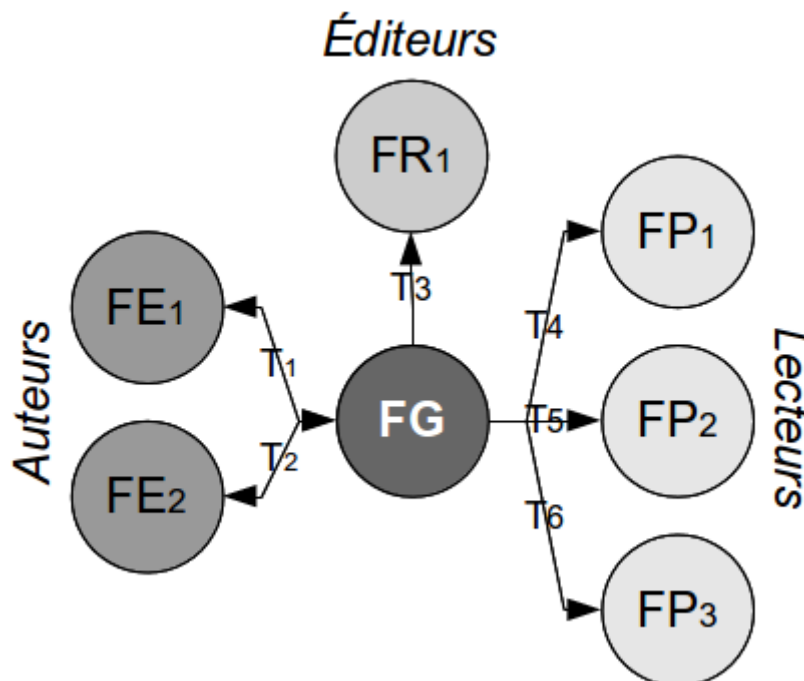


Figure 1 – Illustration du polymorphisme

Un contenu formalisé par une FG sera transformable en FE1 pour les besoins de rédaction des auteurs et FE2 pour les besoins de correcteurs, FR1 pour la validation par l'éditeur, et FP1 pour l'impression d'un livre, FP2 pour une diffusion web et FP3 pour une diffusion simplifiée sur terminaux mobiles.

2.1.3 Exemple d'implémentation du polymorphisme dans Scenari

The screenshot displays the SCENARichain 4.0 editor window titled "Opale - SCENARichain 4.0". The main content area is titled "Définition du XML" and is categorized as a "Grain de contenu". The interface includes a toolbar with various editing tools and a "Brouillon" (Draft) button. The content is organized into sections:

- Grain de contenu**:
 - Titre: Définition du XML
 - Titre court: (empty)
 - Métadonnées: (empty)
 - Définition**:
 - Titre: XML
 - Texte: L'eXtensible Markup Language <XML> est un <méta-langage> permettant de définir des langages à <balises>. Il standardisé comme une recommandation par le <W3C> depuis <1998>.
 - Exemple**:
 - Titre: (empty)
 - Liste à puces:
 - LaTeX et HTML sont des langages à balises (mais ce ne sont pas des langages XML)
 - XHTML est un langage XML
 - Fondamental**: (empty)
 - Complément**: Versions de XML
 - Complément**: Voir aussi
 - Partie**: (empty)

At the bottom, a breadcrumb trail reads: > Grain de contenu > Définition > Texte > Paragraphe. The footer indicates the version: Opale Advanced 3.3 (fr-FR) (3.3.100).

Figure 2 – FE1 : Éditeur SCENARichain

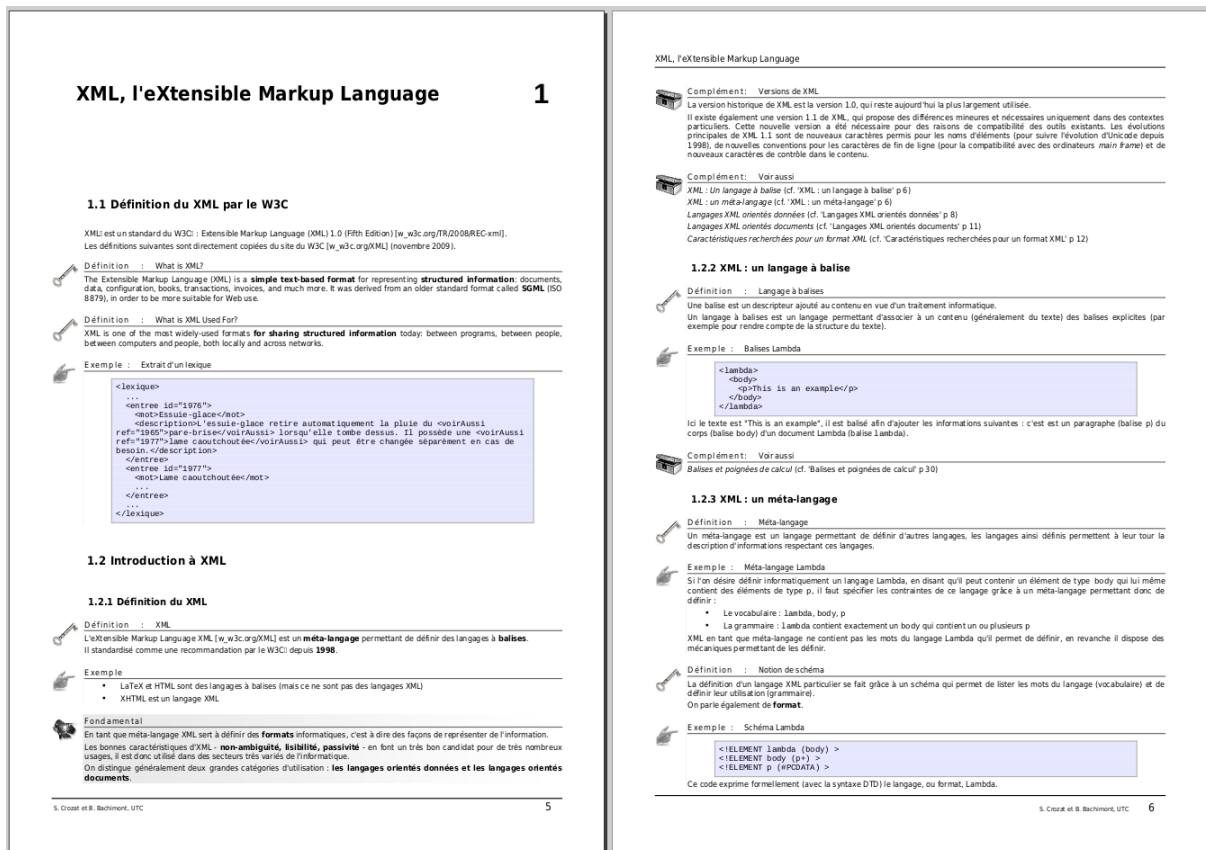
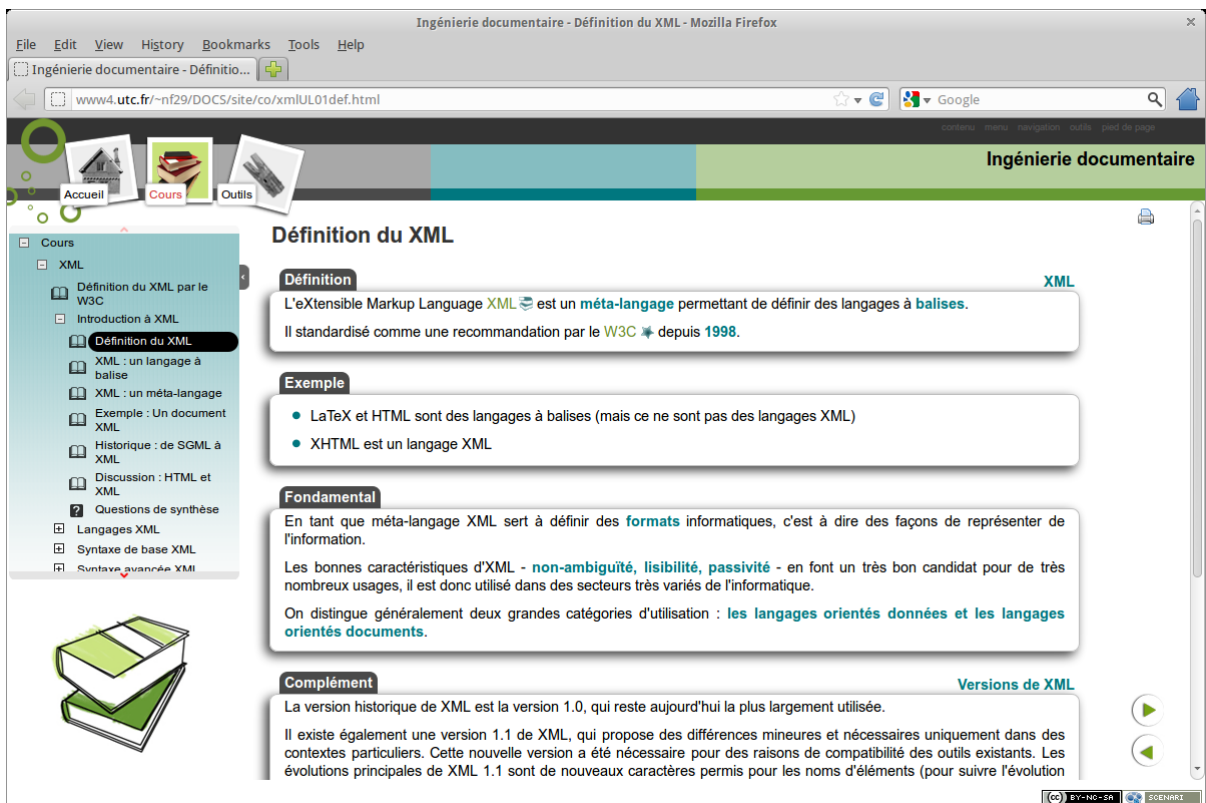


Figure 3 – FP1 : PDF pour impression de photocopiés



Définition du XML

XML

L'eXtensible Markup Language XML est un **méta-langage** permettant de définir des langages à **balises**.

Il standardisé comme une recommandation par le W3C depuis **1998**.

9/142

Stéphane Crozat et Bruno Bachimont, Université de Technologie de Compiègne

Figure 5 – FP3 : HTML pour diaporama vidéo-projeté en amphithéâtre

The screenshot shows an iPad displaying a presentation slide titled "Définition du XML". The interface includes a top status bar with "Carrier", "12:51 PM", and "100%". The application title "eXtensible Markup Language" is in the top right. A left sidebar contains a table of contents with items like "Définition du XML par le W3C", "Introduction à XML", "Définition du XML", "XML : un langage à balise", "XML : un méta-langage", "Exemple : Un document XML", "Historique : de SGML à XML", "Discussion : HTML et XML", "Questions de synthèse", "Langages XML", "Syntaxe de base XML", "Syntaxe avancée XML", "Introduction aux schémas XML", "Quelques bonnes questions sur XML", "Bibliographie", "Compléments", and "Exercices". The main content area is divided into sections: "Définition" (with a key icon), "Exemple" (with a hand icon), "Fondamental" (with a gear icon), and "Complément" (with a book icon). The "Définition" section contains the text: "L'eXtensible Markup Language XML est un **méta-langage** permettant de définir des langages à **balises**. Il standardisé comme une recommandation par le W3C depuis **1998**." The "Exemple" section lists: "• LaTeX et HTML sont des langages à balises (mais ce ne sont pas des langages XML)" and "• XHTML est un langage XML". The "Fondamental" section explains that XML is used to define formats and lists characteristics: "non-ambiguïté, lisibilité, passivité". The "Complément" section discusses XML versions 1.0 and 1.1. At the bottom, there is a footer with the authors' names and Creative Commons license icons.

Figure 6 – FP4 : Paquet ZIP pour application OpaleReader sur iPad

2.2 Transclusion et réutilisation sans variation

2.2.1 Principe de la transclusion

La réutilisation par clonage s'apparente au principe de programmation dit de *passage par valeur*, c'est-à-dire que la valeur de l'original est dupliquée dans un nouvel espace mémoire de l'ordinateur, sans qu'aucun lien ne soit maintenu entre l'original et la copie. Techniquement, il ne s'agit plus du même objet, même si sa valeur – l'information véhiculée – est strictement à ce stade identique (mais elle ne le restera pas si la copie est ensuite modifiée).

L'alternative connue en programmation au passage par valeur est le *pointeur*, c'est-à-dire un objet contenant l'adresse d'un autre objet. Tandis qu'un passage par valeur consiste à occuper plusieurs espaces mémoires identiques pour y inscrire autant de fois la même information, un *passage par référence* consiste à occuper un seul espace mémoire pour y stocker une information, puis à utiliser l'adresse de cet espace chaque fois que l'on souhaite utiliser cette information. Le passage par référence permet à deux ensembles de partager un *même objet* informatique (contenant l'information réutilisée), et non seulement la valeur de l'objet comme dans le passage par valeur.

Dans le contexte de la réutilisation, on parlera de *copie* pour le passage par valeur et de *transclusion* pour le passage par référence [25].

Définition : Fragment

On appelle fragment (F) une ressource numérique qui peut être intégrée par transclusion dans d'autres fragments et ainsi participer à la constitution d'un document. Un fragment est matérialisé par un ou plusieurs fichiers, typiquement des fichiers XML décrivant des structures logiques, des codages binaires d'image, sons, vidéos, etc.

L'ensemble des fragments F_i forment un réseau au sein duquel certains fragments partagent des sous-fragments communs par transclusion : $F_1 \supset F_2, F_3, F_5, F_4$; $F_6 \supset F_4, F_5, F_7$; $F_3 \supset F_5$ [figure 7].

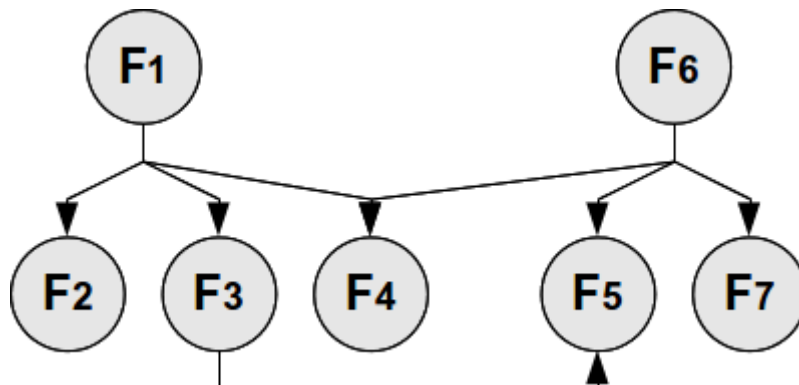


Figure 7 – Illustration de la transclusion

Définition : FG en contexte transclusif

Une forme génératrice (FG) est, en contexte transclusif, un arbre de fragments (F) acceptant des transformations (T) en vue d'obtenir des formes publiées (FP).

2.2.2 Exemple de syntaxe XML pour la transclusion

Dans la figure 8, le fragment *f1.xml* incorpore par transclusion le fragment *f2.xml*. Notons que d'autres fragments pourront également réutiliser *f2.xml* en utilisant la même référence.

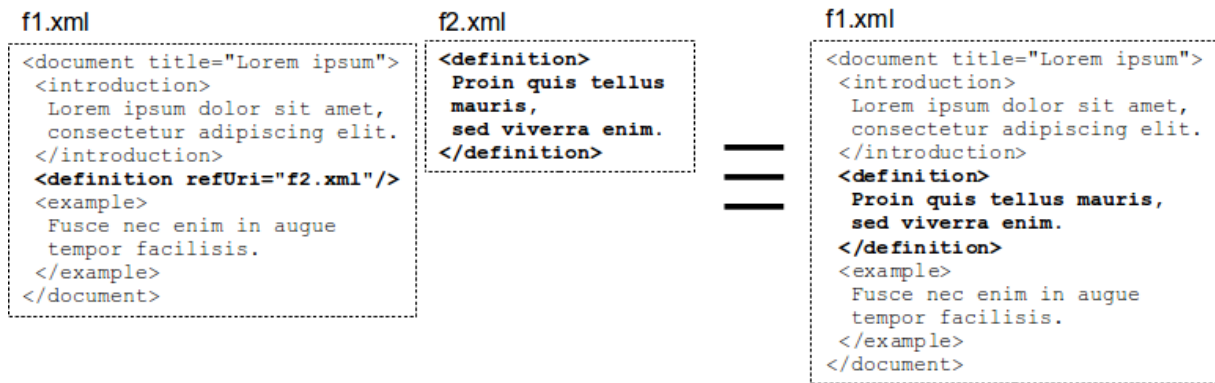


Figure 8 – Exemple d'implémentation du clonage et de la transclusion en XML

2.2.3 Exemple d'implémentation de la transclusion dans Scenari

Le module « Au delà des bases de données relationnelles » [figure 9] est composé de trois *activités* (comprenant elles-mêmes des *grains*) et d'un *grain* à la fin. On observe que :

- la première activité « Introduction aux datawarhouses » est récupérée par transclusion depuis l'espace *nf26* ;
- la seconde activité « Introduction à XML » est directement dans le module, mais elle réutilise elle-même deux grains par transclusion depuis l'espace *nf29* ;
- la troisième activité est une réutilisation par transclusion de l'activité « Introduction aux bases de données XML » de l'espace *nf17* ;
- enfin le grain à la fin est un contenu original directement écrit dans le module.

Le module permet donc de composer un document original, qui contient une part de contenu nouveau et une part de contenu réutilisé par transclusion. Le document résultant est strictement identique à un document qui aurait été obtenu par clonage des fragments réutilisés, comme en témoigne le plan à droite de l'interface d'édition [figure 9] ou la publication [figure 10], qui ne différencie pas les fragments transclus des fragments inclus.

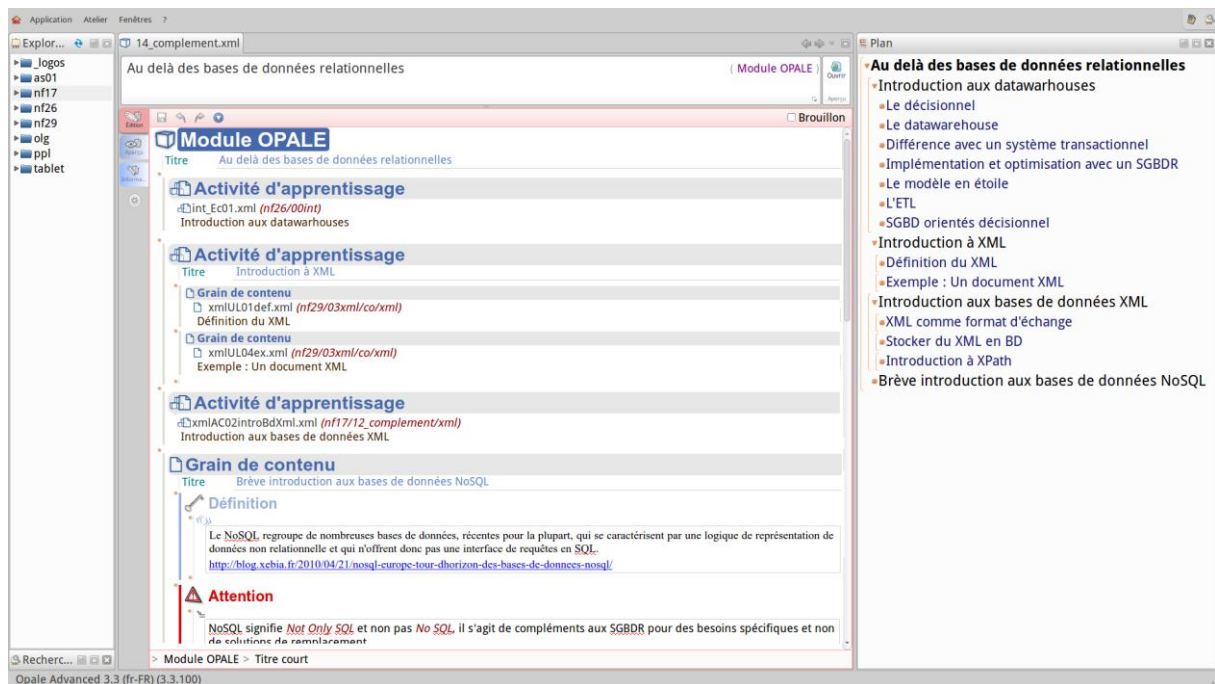


Figure 9 – Édition de l'arbre de fragments composant la FG « Au delà des bases de données relationnelles »

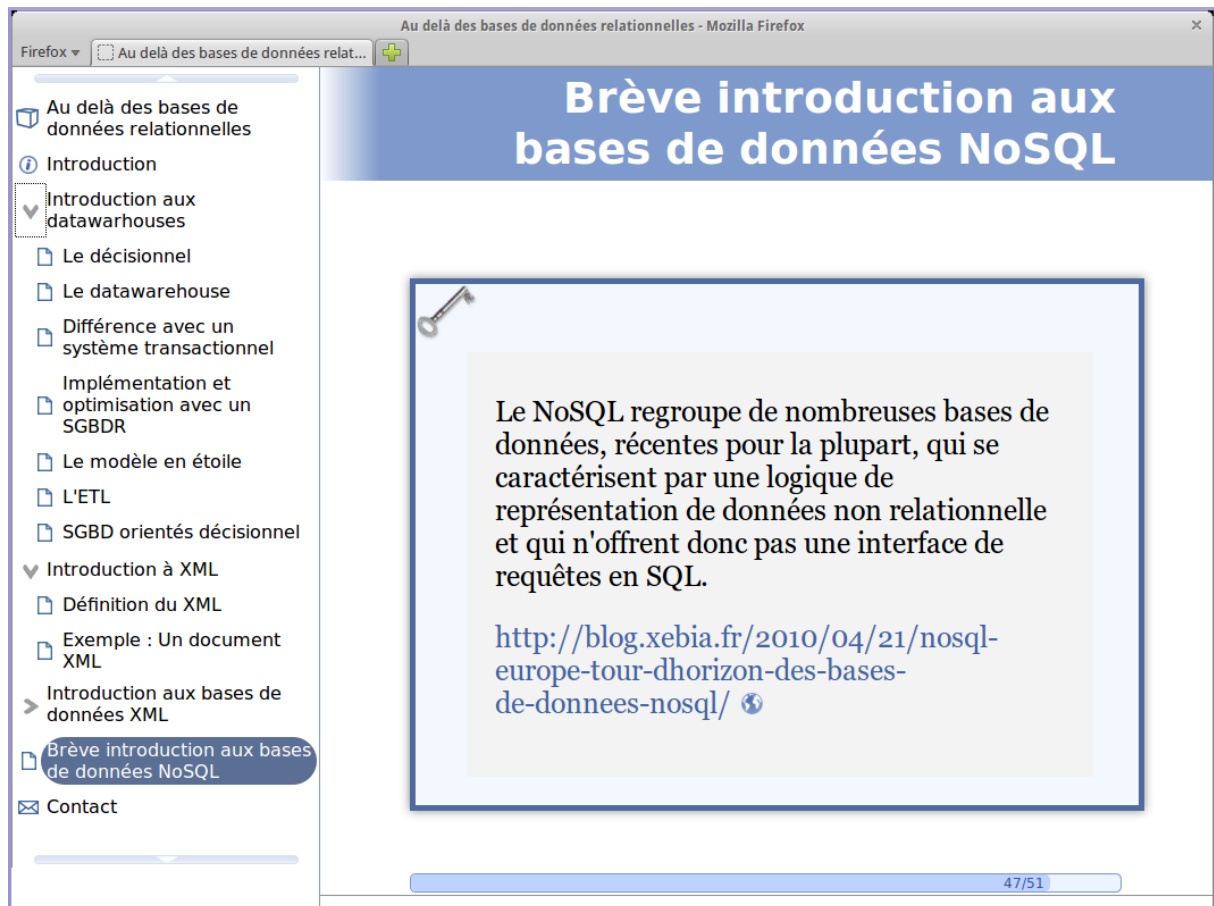


Figure 10 – Publication diaporama de la FG
« Au delà des bases de données relationnelles »

2.2.4 Réutilisation sans variation

La transclusion n'autorise que la réutilisation sans modification contextuelle du fragment réutilisé. En effet le même fragment étant mobilisé au sein de plusieurs FG, toute modification de ce fragment est nécessairement propagée à l'ensemble de ces FG.

2.3 Dérivation et réutilisation avec variation

2.3.1 Problématique : la réutilisation avec variation

Si la réutilisation par transclusion est parfaitement adaptée à certains usages (par exemple la réutilisation de contenus de référence qui ne doivent pas être modifiés quel que soit le contexte éditorial), elle peut s'avérer insatisfaisante lorsque, au contraire, l'adaptation au contexte cible est requise. Cette forme de réutilisation avec modification du contenu entre la source et la cible sera nommée dérivation, en référence au sens grammatical : « procédé qui consiste à former de nouveaux mots en modifiant le morphème par rapport à la base » [*Trésor de la langue française informatisé*], par exemple *sauter* et *sautiller*.

La dérivation renvoie à des usages typiques de copier/coller suivi de variations légères (petites modifications) et ciblées (partie du contenu impactée seulement).

En dehors des cas où la transclusion est souhaitable justement parce qu'elle interdit la modification, la réutilisation sera en général plus qualitative moyennant une adaptation, et c'est uniquement le coût de gestion de la dérivation qui restreindra son usage [cf. partie 3]. Prenons l'exemple de l'adaptation d'une présentation commerciale d'entreprise. Tandis que les parties relatives au contexte de l'entreprise seront à conserver telles quelles, celles relatives à l'argumentaire commercial seront à adapter au client visé : modification des exemples choisis, suppression de certaines parties jugées secondaires dans ce contexte, au contraire ajout de compléments...

2.3.2 Principe de la dérivation

L'acte de modification implique un retour au clonage du contenu. L'enjeu est alors d'une part de circonscrire au maximum ce clonage et d'autre part de conserver le lien entre la source et la cible de façon à ce que la cible soit informée lorsque la source évolue (afin d'évoluer avec elle).

Afin de limiter le clonage, on pose le principe de fragment dérivé F' obtenu par application d'opérations sur les éléments de F : la conservation, la suppression, le remplacement et l'ajout [figure 11].

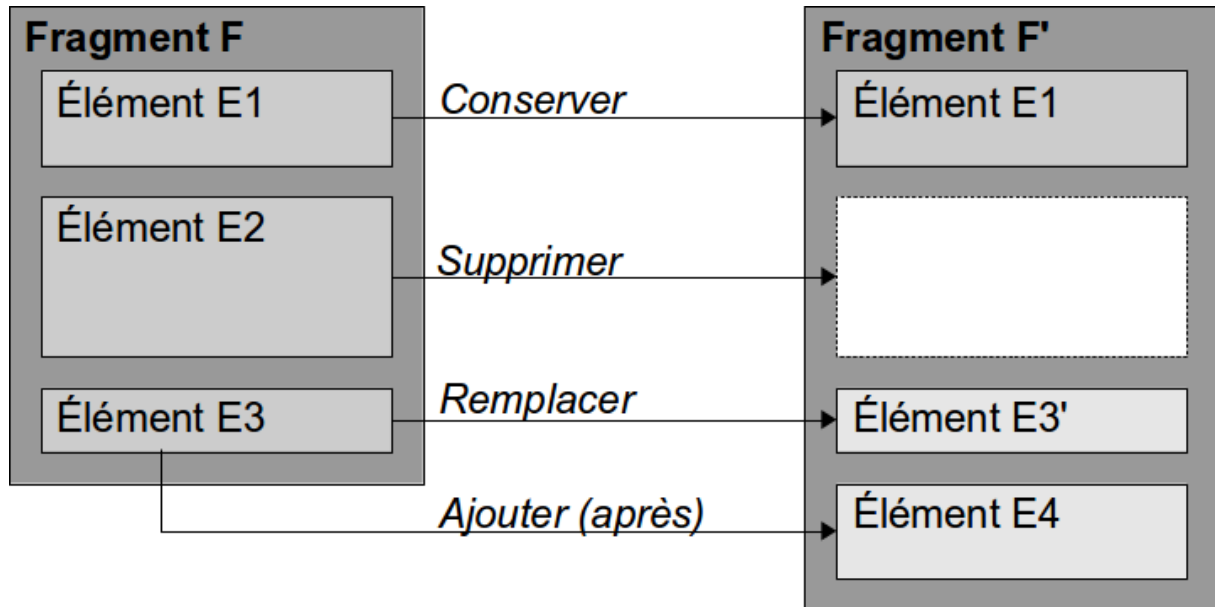


Figure 11 – Illustration de la dérivation

Le premier avantage de cette approche est que seul le strict minimum est cloné (ce qui diffère). Le second avantage est que la référence de F' à F est conservée intrinsèquement, que ce soit à des fins d'authentification de la source ou à des fins de maintenance du contenu. L'auteur du fragment F' pourra être prévenu par un système adéquat des modifications survenues dans F , afin de choisir de faire évoluer F' en symbiose.

2.3.3 Exemple de syntaxe XML pour la dérivation

Dans l'exemple proposé en figure 12, F est obtenu par dérivation de F , en remplaçant l'introduction, conservant la définition et l'exemple, ajoutant un second exemple et supprimant la conclusion.

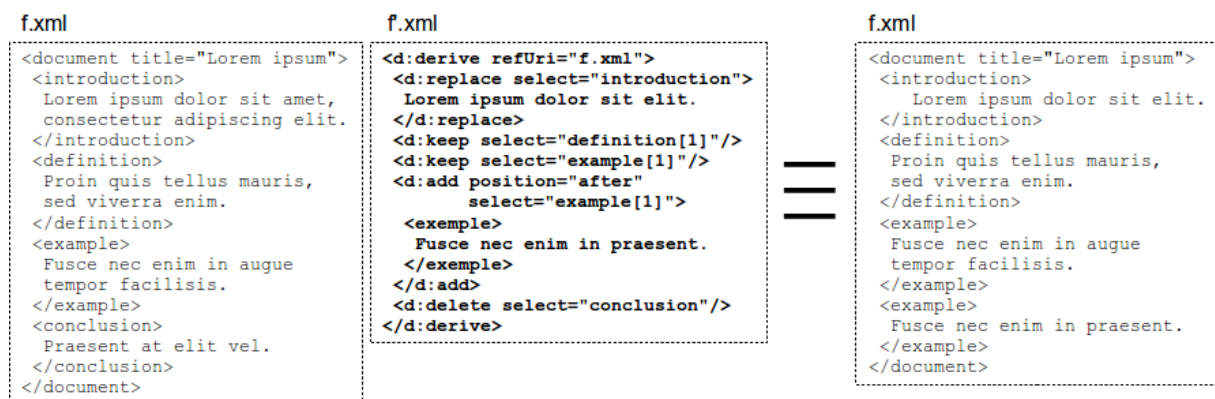


Figure 12 – Exemple d'implémentation de la dérivation en XML

2.3.4 Exemple d'implémentation de la dérivation dans Scenari

Scenari ne permet pour le moment que d'implémenter partiellement la dérivation ; en particulier, seule la surcharge (remplacer) est possible, et seulement à l'échelle du fragment entier. Dans l'exemple présenté en figure 13, le fragment « Définition du XML » est surchargé (ce qui est indiqué par l'icône bleu foncé dans la fenêtre de gestion à gauche, qui se distingue des icônes bleu clair des fragments non surchargés). En

l'occurrence l'exemple d'origine a été supprimé, ainsi que les compléments situés à la fin, et un nouvel exemple a été ajouté.

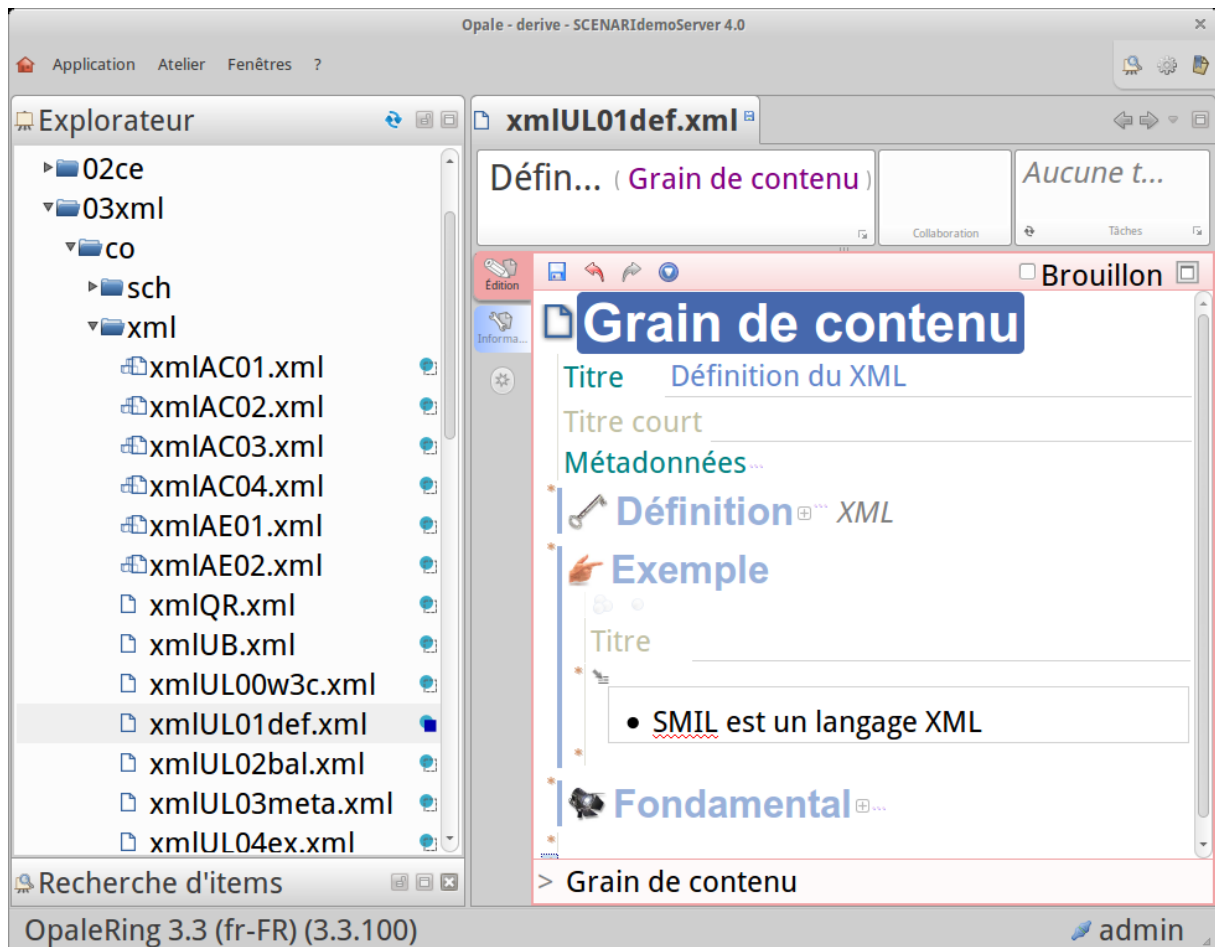


Figure 13 – Édition d'un fragment dérivé

Le système permet ensuite de gérer les cycles de vie de la dérivation en fonction de la source. Ainsi, lorsque la source est modifiée, l'icône du fragment dérivé passe en rouge pour inviter l'auteur à mettre à jour son contenu (l'icône deviendra vert quand la modification aura été validée) [figure 14]. Il est également possible à tout moment d'annuler la surcharge et de restaurer la version non dérivée du fragment.

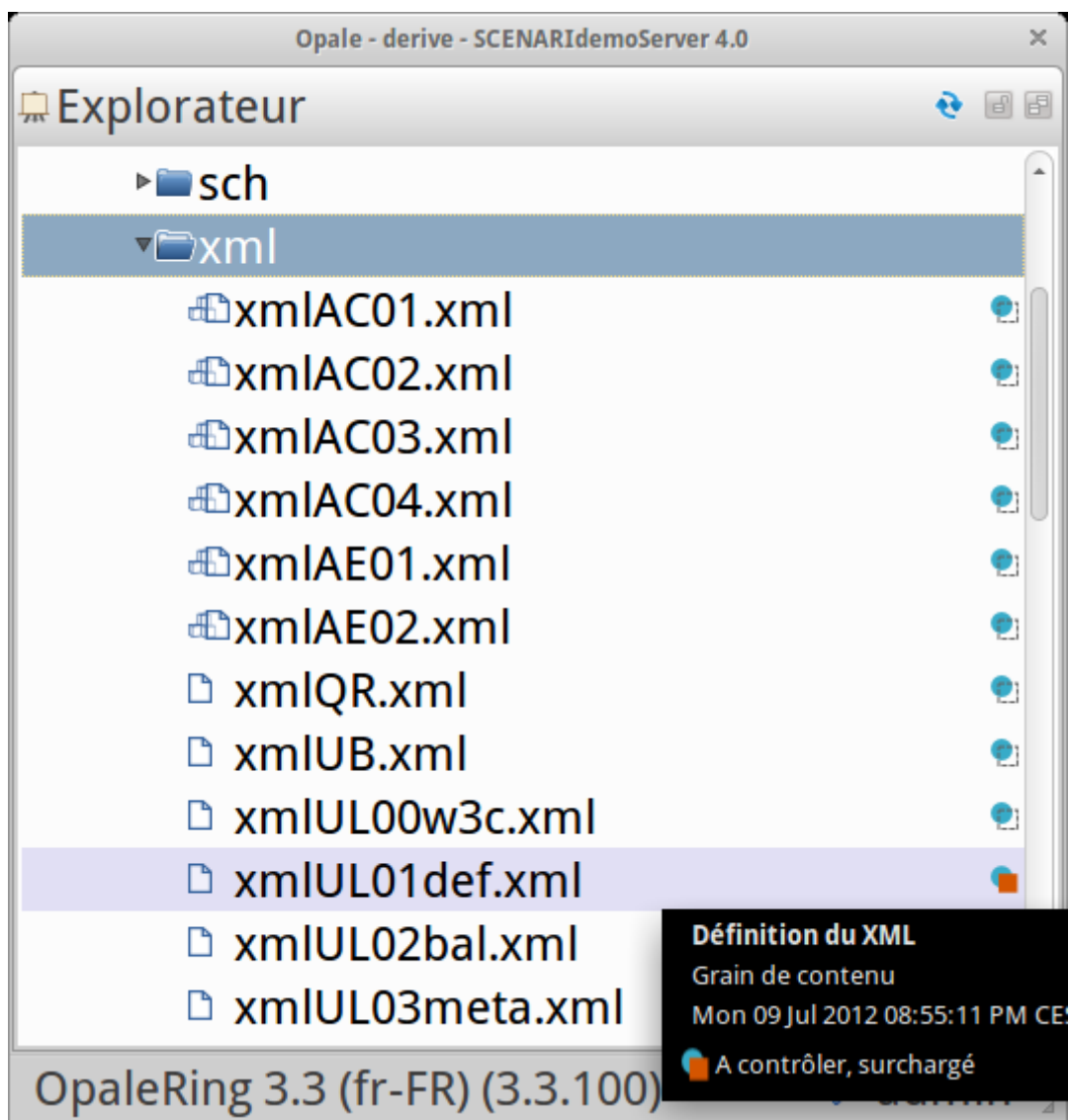


Figure 14 – Gestion des fragments dérivés

2.4 Déclinaison et programmation de variations

2.4.1 Problématique : automatisation des variations *a priori*

La dérivation permet la variation par enregistrement de modifications différentielles manuelles entre le fragment source et le fragment réutilisé. Bien que permettant de répondre à tous les cas possibles de variation, cette technique pose des problèmes de gestion, que nous étudierons dans la troisième partie. Or il existe des cas où la variation peut être programmée, c'est-à-dire que l'on peut en prévoir tous les cas *a priori* et automatiser les transformations associées.

On parlera alors de déclinaison, toujours en référence au sens grammatical : « action d'énoncer selon un paradigme les formes variables pourvues d'affixes que peuvent prendre les différents constituants d'un syntagme » (*Trésor de la langue française informatisé*), par exemple *gentil* et *gentilles*.

On pourra citer, par exemple, l'utilisation de variables dans le texte : un contrat-type est rédigé avec des variables *partie1* et *partie2* à la place des noms des contractants, les variables peuvent ainsi être instanciées en fonction d'un fichier de paramétrage qui en fixera contextuellement les valeurs.

2.4.2 Principe de la déclinaison

La déclinaison s'instrumente en deux étapes liées : l'auteur d'un fragment déclinable *F*- insère des variables au sein de son contenu, en prévision des variations qu'il a identifiées *a priori*; l'auteur d'un fragment décliné *F1* exploite ces variables pour fixer les variations correspondant à son *contexte* parmi celles possibles. Le fragment décliné *F1* pourra alors être obtenu automatiquement par une transformation de *F*-. Nous nommerons *F*- pseudo-

fragment ; en effet F^- n'est pas vraiment un fragment, au sens où il ne peut être utilisé que s'il est associé à une instantiation des variables qu'il contient. F^- n'existera en tant que contenu lisible qu'à travers ses déclinaisons F_i . Le fragment F_1 est construit par déclinaison de F^- , en fixant la valeur de la variable $\$var1$ utilisée dans le texte à `xxxxx` et en posant la règle d'exclusion des éléments dont le paramètre $\$var2$ est fixé à λ [figure 15].

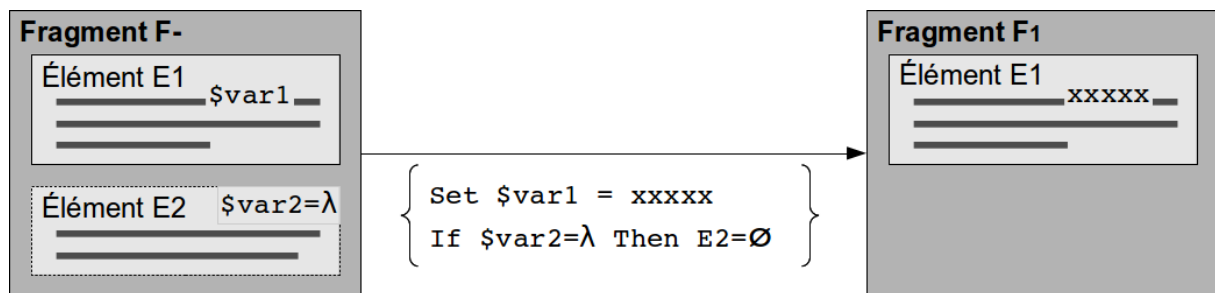


Figure 15 – Illustration de la déclinaison

2.4.3 Déclinaison versus dérivation

La dérivation est un processus auctorial :

- *maîtrisé par la cible* : l'auteur d'un fragment dérivé F' s'appuie sur une source F et peut la transformer à volonté, l'auteur de F n'est pas impliqué dans le processus ;
- *a posteriori* : les variations possibles ne sont pas déterminées *a priori*, les contextes de réutilisation adressables sont théoriquement infinis.

La déclinaison est un processus auctorial :

- *maîtrisé par la source* : l'auteur d'un pseudo-fragment F^- détermine toutes les variations possibles à travers son acte d'écriture ;
- *a priori* : l'auteur d'un fragment décliné F_1 ne peut choisir que parmi les variations qui ont été rendues possibles par l'auteur de F^- ; les contextes de réutilisation adressables sont limités à ce qui a été prévu à la source.

2.4.4 Exemple d'implémentation de la déclinaison dans Scenari

L'édition du fragment « Définition du XML » permet d'associer à certains éléments les valeurs *aucune version*, *version courte*, *version longue* ou *version courte et longue*, symbolisées graphiquement par les icônes *une sphère* et *trois sphères* pouvant chacun prendre les valeurs 0 ou 1 (barrée ou non). Puis l'édition de la transformation de la FG « eXtensible Markup Language » – incluant le fragment « Définition du XML » –, permet de sélectionner uniquement les éléments de type version longue, uniquement les éléments de type version courte ou les deux, selon le type de publication visée [figures 16 à 18].

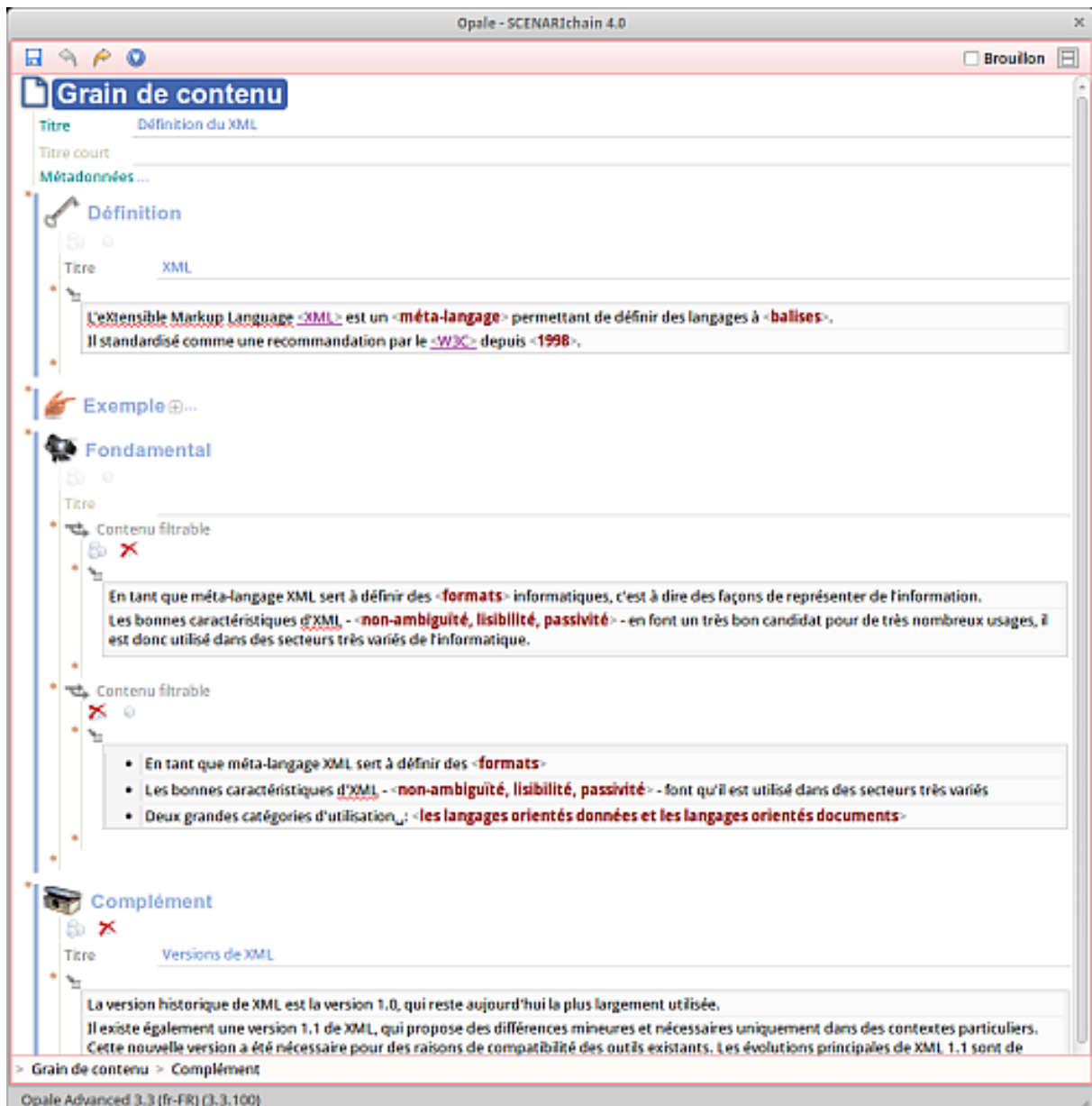


Figure 16 – Édition paramétrée du pseudo-fragment « Définition du XML »

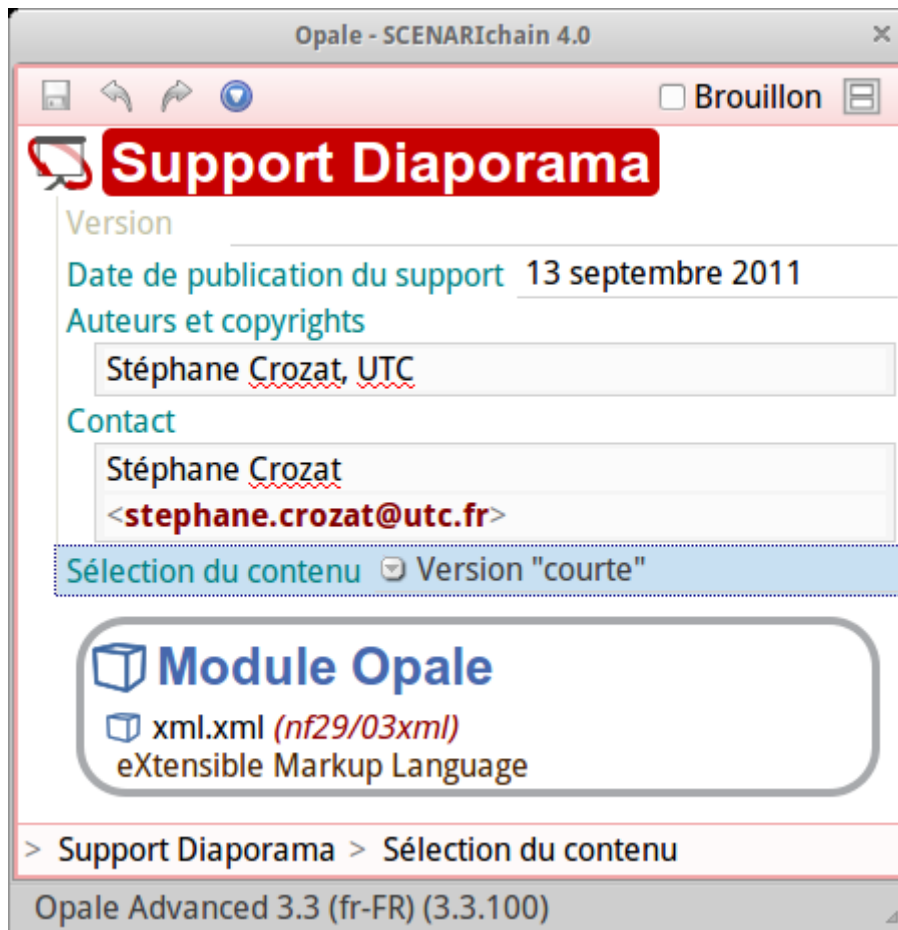


Figure 17 – Déclinaison « version courte » pour diaporama

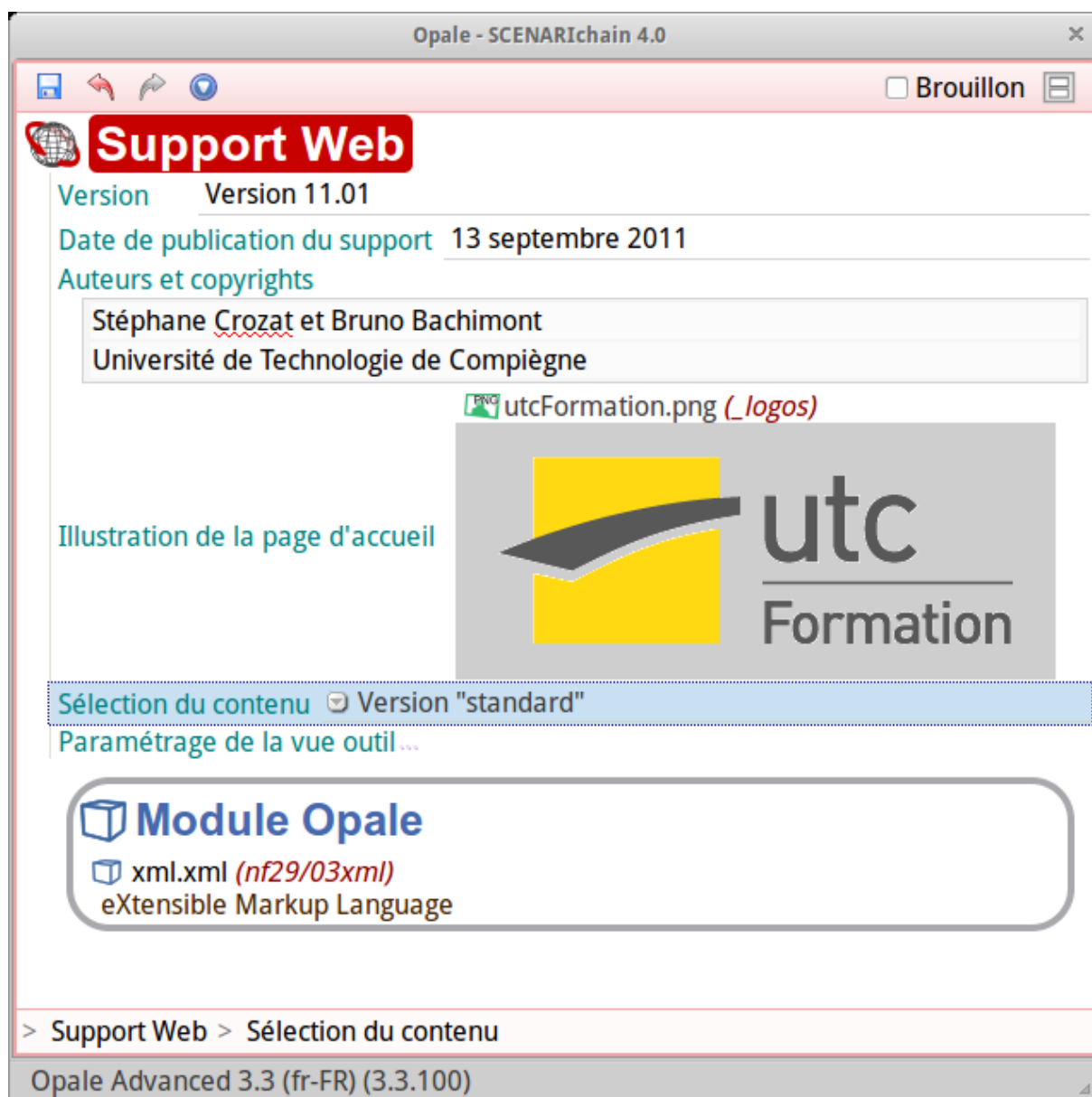


Figure 18 – Déclinaison « version longue » pour diffusion web

3 Enjeux actuels et prospectifs des chaînes éditoriales

Après avoir présenté différentes techniques de gestion – de domestication – de la rééditorialisation, nous présentons dans cette partie différents enjeux associés aux pratiques d'écriture qui en émergent. Dans les deux premières sections nous étudierons les concepts de *modélisation* et de *document-dossier* qui viennent compléter les techniques de rééditorialisation pour répondre aux problèmes de *complexité* et de *dédokumentarisation* qui en découlent. Dans les troisième et quatrième sections nous étudierons les extensions nécessaires à la prise en compte des contextes particuliers du *multimédia* et du *collaboratif*.

3.1 Complexification de l'écriture et modalisation contextuelle

3.1.1 La rééditorialisation entraîne une complexification des actes auctoriaux

Chacune des techniques présentées dans la partie précédente est un facteur de complexification de l'acte d'écrire. Au niveau *conceptuel*, l'auteur devra assimiler les principes de polymorphisme, transclusion, dérivation et déclinaison. Au niveau *logique*, l'auteur devra être capable d'intégrer les mécanismes techniques associés afin d'anticiper le traitement manipulateur qui sera effectué sur ses contenus. Au niveau *pratique*, il devra faire évoluer – voire réinventer – ses pratiques d'écriture pour s'exprimer au mieux au sein de ce nouveau dispositif. Il

développera ainsi une nouvelle littérature, numérique, au sens de sa capacité à mobiliser l'écriture pour atteindre ses objectifs de communication. On se propose d'illustrer cette complexification par des exemples liés à chaque technique.

L'écriture de contenus polymorphes implique d'abord de renoncer à s'exprimer dans une forme unique et concomitante de son écriture pour accepter l'idée de formes multiples et successives. C'est une rupture fondamentale dans la mesure où l'écriture s'inscrit de fait dans une forme que l'on voit, en cela l'écriture numérique n'est déjà plus une écriture par principe, mais la programmation d'une écriture de fait. Si le WYSIWYG tend à masquer cette différence en simulant l'identité entre forme écrite et forme lue, le WYSIWYM et le polymorphisme consomment la rupture et imposent à l'auteur d'abstraire son écriture de sa – ou plutôt de ses – formes. En effet la seconde difficulté, une fois la première surmontée, est de parvenir à projeter son écriture sur tous les supports visés. La qualité des formes éditables (FE) est alors fondamentale pour aider l'auteur dans ce travail d'abstraction et de projection.

La mobilisation de la transclusion pose un problème similaire de projection dans la forme documentaire, dans la mesure où les contenus ne sont plus présents ensemble. La FE peut également aider, par exemple en affichant le contenu transclus à la place de sa référence. Une seconde difficulté consiste ensuite à gérer correctement la rédaction, puis les évolutions de la rédaction, d'un fragment utilisé au sein de plusieurs documents, afin qu'il respecte ses différents contextes. Typiquement, lors de la rédaction d'un tel fragment, il ne pourra plus être fait référence aux fragments antérieurs ou postérieurs qui risquent de ne pas être présents au sein de tous les documents l'utilisant. La transclusion conduit à la *granularisation* de l'information, et à l'émergence d'une écriture délinéarisée.

La dérivation impose essentiellement un surcoût de gestion. En effet, il est plus facile de dupliquer/modifier que d'exprimer explicitement des différentiels. D'autre part, il sera nécessaire de gérer l'évolution du contenu dérivé avec sa source. En particulier, la modification d'une source, dans un contexte d'écriture donné, entraîne des actes de gestion – au minimum, voire d'écriture – au niveau de fragments dérivés, dans d'autres contextes. L'auteur passe ainsi d'une écriture spécifique à un contexte à une écriture multicontextes.

La mécanique de déclinaison impose quant à elle un surcoût à la création, dans la mesure où l'acte d'écriture paramétrable se rapproche de l'acte de programmation. L'auteur doit anticiper toutes les déclinaisons et trouver la meilleure solution pour les atteindre ; il n'est plus seulement auteur, mais *programmeur*.

On voit à travers ces exemples que le support numérique impose ses règles à l'acte d'écriture. Sa dimension calculatoire, et la complexité qui l'accompagne, ne peuvent être totalement ignorées de l'auteur.

3.1.2 La contextualisation des chaînes éditoriales permet de réduire la complexité

Cette complexification pose deux problèmes potentiels : elle réduit le spectre des individus capables d'écrire, ceux qui accèdent à la littérature numérique ; elle augmente la charge cognitive des auteurs et donc leur efficacité dans l'écriture. Si l'on peut supposer que le développement culturel de la littérature numérique tendra à réduire la portée de ce problème via des processus d'alphabétisation, l'on peut également supposer que de nouvelles formes complexes liées au numérique continueront d'émerger et auront un impact inverse.

Une observation générale est que chaque contexte d'écriture pris isolément n'a pas besoin de mobiliser l'ensemble des techniques de gestion de la rééditorialisation – pas de façon systématique et pas à tous les niveaux de l'écriture.

L'idée est alors de *spécialiser* les chaînes éditoriales en fonction des contextes afin de réduire le nombre et le spectre des techniques mobilisées, et de les limiter au juste nécessaire. Typiquement, la transclusion n'est en général pas utile à tous les niveaux de l'arbre documentaire. On peut donc décider *a priori* de niveaux de réutilisation (la diapositive pour un diaporama, par exemple), ce qui d'une part réduit les cas où l'auteur est impacté, et d'autre part permet éventuellement de faire coïncider la transclusion avec des niveaux de structuration qui la rendent plus naturelle d'un point de vue documentaire (la rédaction granulaire d'une diapositive est déjà inscrite dans les pratiques).

3.1.3 La modélisation documentaire permet d'instrumenter la contextualisation

La modélisation documentaire permet de spécialiser une chaîne éditoriale en fixant *a priori* les possibilités d'écriture qui seront activées, ainsi que la façon dont elles le seront. Cette approche est généralisée dans Scenari à travers l'outil SCENARIBuilder [2]. Ainsi chaque contexte auctorial verra un spectre fonctionnel différent de Scenari, dépendant du modèle qu'il utilise. Les exemples mobilisés depuis le début de ce chapitre sont en fait issus du modèle Opale de Scenari [20] ; d'autres modèles auraient pu montrer des usages dénués de polymorphisme, de transclusion...

Dans la figure 19, le modèle *diaporama.model* permet d'intégrer des diapositives (définies par *diapositive.model*) par transclusion, car le paramètre *internalized* est fixé à *never*. Les autres valeurs possibles sont *always* pour ne pas activer la transclusion à ce niveau et *userDependent* pour laisser le choix à l'auteur (ce qui est le plus complexe pour lui).

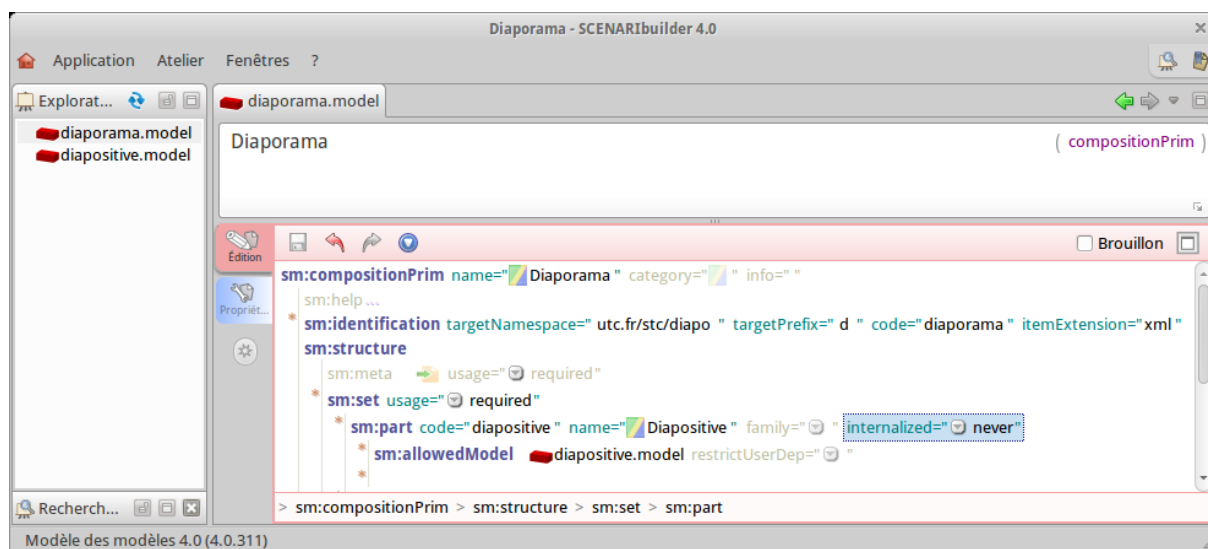


Figure 19 – Modélisation d'une transclusion dans un diaporama avec SCENARbuilder

3.2 Le document-dossier, à la recherche du document perdu

3.2.1 Document mort versus FG vivante

Par définition, un document est un contenu figé – mort – qui n'évoluera plus en tant que tel une fois publié (une seconde version pourra émerger, mais elle constituera un autre document). Le document n'est consulté par les lecteurs qu'à l'issue du processus éditorial, qui se termine par un acte de publication qui fixe le contenu. Cette stabilité du document est la condition de son appropriation.

Les techniques de rééditorialisation que nous avons présentées, au contraire, tendent à constituer des réseaux de fragments composant des FG toujours vivantes. En effet toute modification d'une FG impacte potentiellement plusieurs FP, liées par polymorphisme ou issues de FG mobilisant des fragments liés par transclusion, dérivation ou déclinaison.

3.2.2 La dédocumentarisation accrue par la rééditorialisation

La rééditorialisation est donc fondamentalement un facteur de dédocumentarisation. Les FG toujours vivantes sont instables et accroissent la perte du caractère pérenne du document, déjà théoriquement affaibli par le numérique. Par ailleurs la rééditorialisation efface par construction la relation 1:1 entre le contexte de production et le contexte de réception, pour la transformer en relation 1:N, voire N:M avec la transclusion [16]. Ainsi la chaîne éditoriale, finalement, ne connaît plus d'objet directement assimilable au document. Et l'auteur devient le producteur et le gestionnaire d'un réseau de fragments vivants, dont il doit maintenir intelligibilité et cohérence.

3.2.3 Le document-dossier comme réponse à la dédocumentarisation

Si la finalité d'une chaîne éditoriale numérique n'est plus de produire des documents au sens traditionnel, elle est bien en revanche de permettre l'émergence de formes numériques lisibles, qui ne sont plus des documents mais dont les propriétés permettent des approches interprétatives homologues (voir la définition proposée dans la première partie), en particulier la pérennité et l'existence d'un contexte éditorial.

Nous proposons le concept de document-dossier (ou *foldoc* pour *folder-document* en anglais) pour répondre à cette problématique [*ibid.*]. Un document-dossier est une construction permettant : d'une part de fixer une FG indépendamment de ses relations existantes à d'autres FG ; d'autre part pour un usage, un contexte éditorial donné. En tant que forme fixée et contextualisée, il tend à reproduire les propriétés d'un document.

Un document-dossier contient [figure 20] :

- une forme génératrice figée FG', qui est une copie d'une FG dans laquelle tous les liens dynamiques entre fragments ont été résolus ;
- le modèle (M) sous-jacent de FG' permettant de l'interpréter humainement et informatiquement, en particulier pour être en mesure d'exécuter des transformations ;
- des FP permettant la lecture directe du contenu ;
- des métadonnées (MD) de documentarisation.

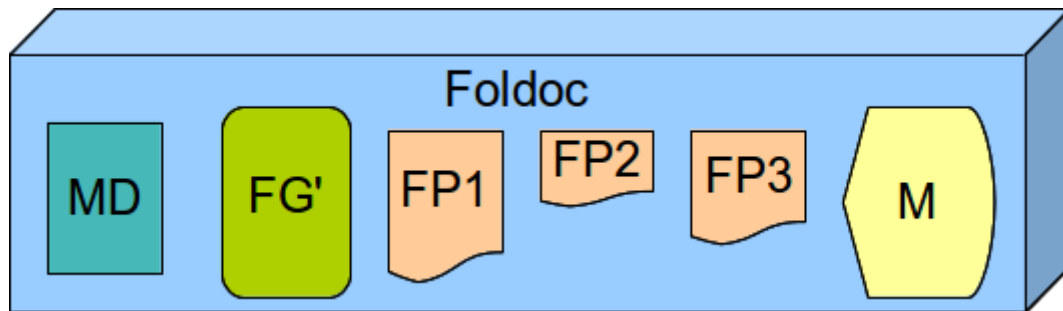


Figure 20 – Structure d'un document-dossier

3.3 Rééditorialisation de contenus audiovisuels et multimédia

3.3.1 L'audiovisuel est un domaine pionnier dans l'étude de la question de la rééditorialisation

Le domaine de l'audiovisuel et du multimédia est historiquement un champ privilégié de la rééditorialisation dans la mesure où le coût de la production est très élevé, et donc la réutilisation une voie économique de valorisation d'archives déjà existantes.

« Progressivement, les outils de gestion audiovisuelle ne permettront pas seulement de retrouver des contenus et de les rejouer dans leur intégralité, mais ils proposeront aussi de sélectionner des parties pour en faire des ressources pour d'autres productions. Autrement dit, on passe de l'indexation, qui a pour but de retrouver un contenu, à une éditorialisation, qui a pour but de produire de nouveaux contenus à partir d'éléments pris arbitrairement (c'est-à-dire comme l'on veut, et non pas au hasard !). [5, p. 166] » « L'étape suivante consiste en une réexploitation qui ne serait plus un simple catalogage des ressources disponibles mais une véritable éditorialisation des contenus. [19, p. 3] »

Si les techniques présentées précédemment s'appliquent à l'audiovisuel, elles posent néanmoins deux questions nouvelles : la gestion de la dimension temporelle et l'abstraction de contenus fondamentalement graphiques.

Définition : Multimédia

On appelle multimédia la mobilisation de plusieurs formes sémiotiques différentes (texte, image, son, etc.) au sein d'un même document.

3.3.2 La structuration logique de contenus temporels

Afin de pouvoir manipuler le contenu dans toute sa dimension multimédia, il est nécessaire de proposer une représentation logique qui permette la structuration de flux temporels. On mobilise pour cela des mécanismes qui permettent d'associer des représentations XML temporalisées à des ressources binaires audiovisuelles (à l'instar de ce que propose le langage standard W3C SMIL, par exemple) [figure 21].

```
<segmentation>
  <stream refUri="av/media.mp4"/>
  <segment>
    <title>Loret ipsum</title>
    <timeCode begin="0" end="5"/>
  <segment>
  <segment>
    <title>Dolor sit amet</title>
    <timeCode begin="5" end="13"/>
  <segment>
</segmentation>
```

Figure 21 – Représentation XML de flux temporalisés

3.3.3 L'abstraction de contenus essentiellement graphiques

Une seconde question, plus délicate, est celle de l'abstraction graphique. En effet les contenus audiovisuels sont par essence plus intimement liés à leur forme que le texte, et sont en ce sens moins transformables. Il est ainsi plus difficile d'abstraire la mise en forme lorsque l'on travaille sur de tels matériaux, et la dimension *esthétique* trouve une place plus importante dans le processus d'interprétation. Une piste à explorer consisterait à permettre une post-édition des formes publiées, afin de mieux en maîtriser le rendu graphique [figure 22]. La difficulté principale est alors de gérer l'évolution de cette post-édition lorsque la FG source évolue. Une solution pourrait être de faire remonter au niveau des FP des mécanismes de dérivation similaires à ceux existant au niveau des FG.

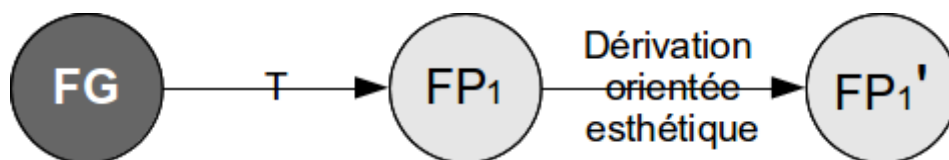


Figure 22 – Post-traitements de FP orientés esthétique

3.3.4 Un cas d'application : le modèle Webmedia

Historiquement conçu avec l'Institut national de l'audiovisuel (Ina) pour la valorisation d'archives radiophoniques [29], le modèle Webmédia2 de Scenari a évolué pour permettre aujourd'hui l'éditorialisation de contenus sonores et audiovisuels. Outre son exploitation à l'Ina, il est mobilisé pour la valorisation de conférences³ [figure 23] ou la promotion d'activités culturelles (avec le réseau des médiathèques de Valence).

³ Voir par exemple la publication du séminaire Precip sur le site de Skhole.fr : <http://skhole.fr/dossier-seminaire-precip-enseigner-l-écriture-numérique>

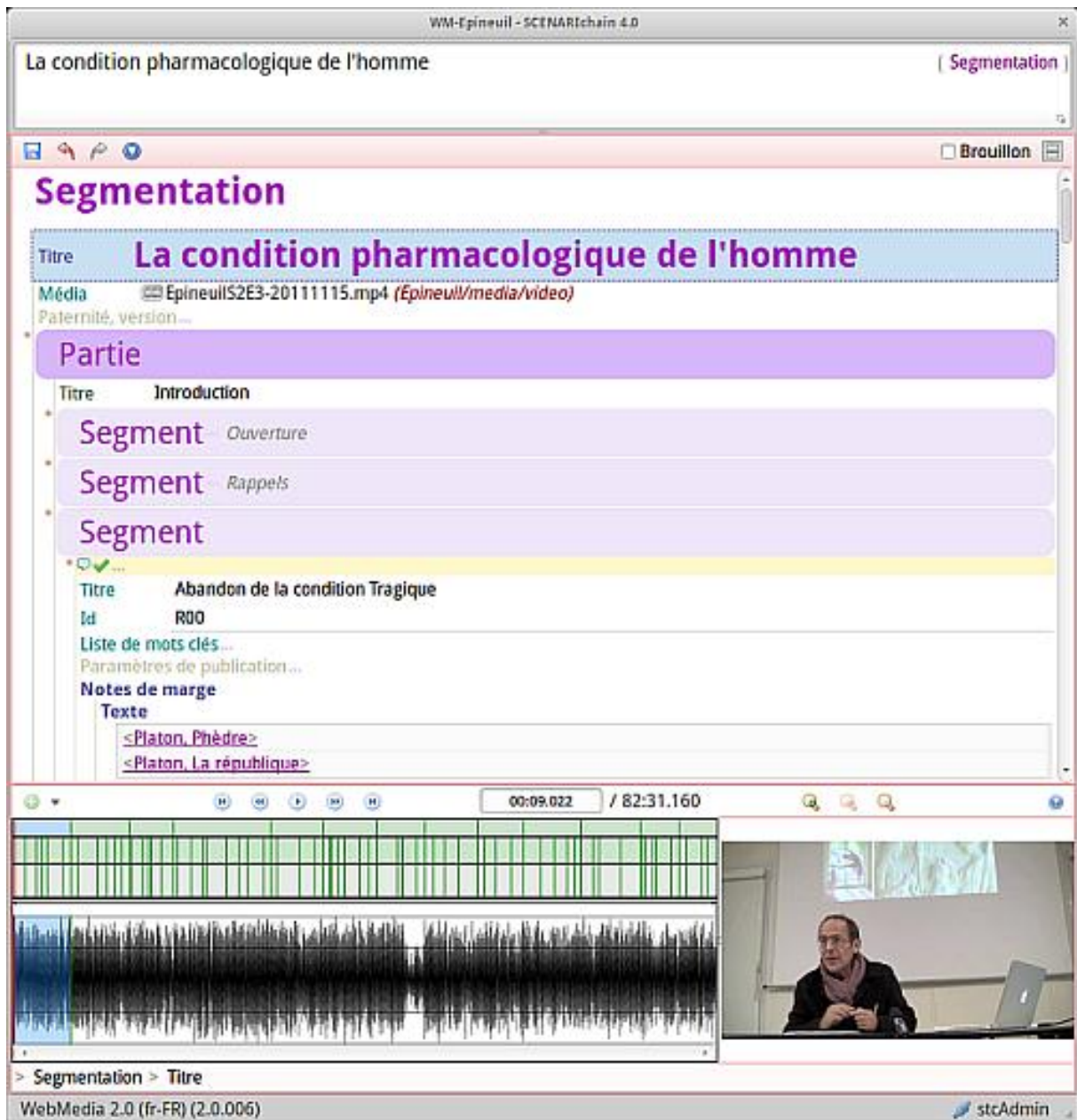


Figure 23 – Exemple d'interface d'éditorialisation de conférences de l'école de philosophie d'Epineuil-Le-Fleuriel

3.4 Rééditorialisation en contexte collaboratif

3.4.1 Le collaboration accentue la complexité éditoriale

Le projet ANR Chaînes éditoriales Collaboratives Multimédia (C2M)⁴, qui s'est terminé en mars 2012, avait pour objet principal l'étude des chaînes éditoriales et de la rééditorialisation en contexte collaboratif. L'introduction de la collaboration, c'est-à-dire l'intervention de plusieurs acteurs (auteurs, éditeurs, rédacteurs, etc.), au sein d'un système transclusif introduit une nouvelle couche de complexité pour l'auteur. Non seulement il doit appréhender un réseau vivant, mais celui-ci prend en quelque sorte son autonomie, dans la mesure où ses modifications peuvent provenir d'actes dont l'auteur n'a pas connaissance (car réalisés par d'autres acteurs). Outre les questions de gestion des transactions et des droits – assez classiques en gestion documentaire, bien que réinterrogées en contexte transclusif [16] –, on peut illustrer ce problème à travers la question de la propagation des modifications.

Lorsqu'un auteur modifie un de ses fragments, il influe potentiellement sur des FG dont il n'a pas connaissance. Dans l'exemple présenté en figure 24, la modification de 3 en 3' par A1 aura une influence sur 4, mais A1 n'a

⁴ <http://utc.fr/ics/c2m>

aucune raison *a priori* d'avoir connaissance de cette transclusion. En conséquence il ne peut pas la prendre en compte dans la gestion de sa mise à jour. Seul Bob sera en mesure de gérer la réutilisation transclusive, il devra pour cela être informé de la modification effectuée par AI, et le système devra lui laisser le choix entre référencer 3', continuer de référencer 3 en tant que version antérieure de 3', dériver une version 3'' adaptée à ses besoins...

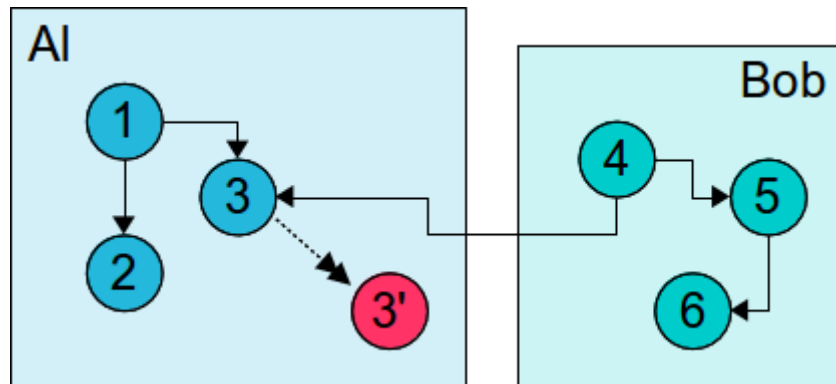


Figure 24 – Fragmentation en contexte collaboratif

3.4.2 Atelier de formes génératrices et bibliothèque de documents-dossiers

La réponse générale proposée par le projet a été de s'appuyer sur la notion de document-dossier pour proposer d'organiser la collaboration intensive, difficile à gérer par essence, au sein d'ateliers permettant la gestion dynamique des FG, tandis que des documents-dossiers peuvent être extraits pour être gérés dans des bibliothèques [*ibid.*] [figure 25].

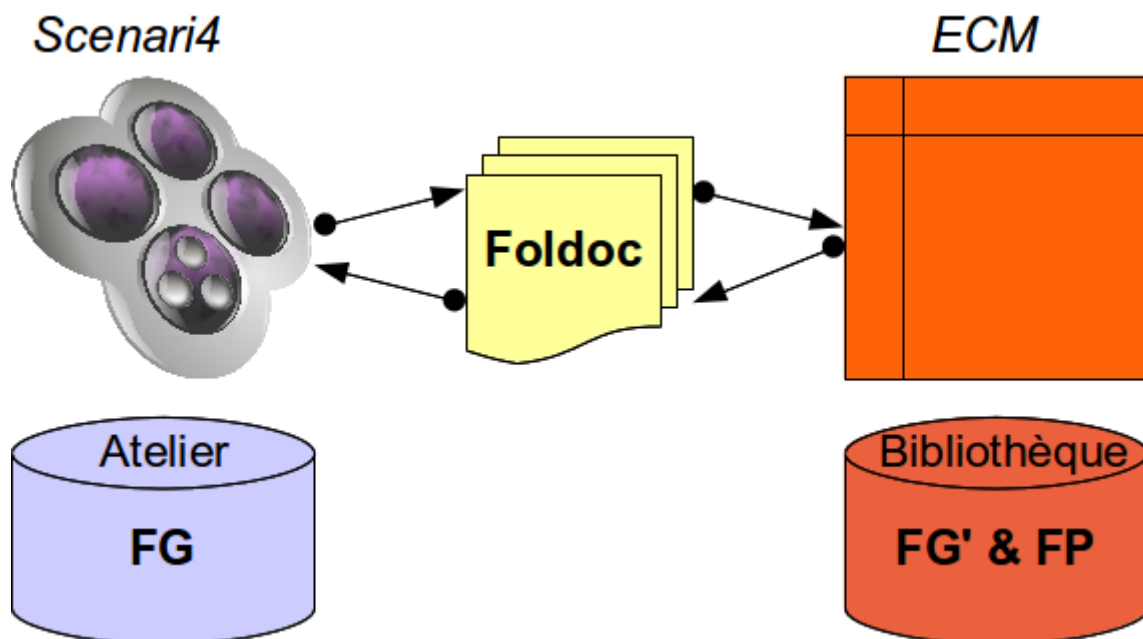


Figure 25 – Organisation atelier-bibliothèque

L'intérêt de cette séparation est de permettre de recréer une phase artificielle de publication, qui restreint à l'atelier les possibilités de rééditorialisation dynamique, pour réintroduire le clonage lors du passage à la bibliothèque. L'atelier sera réservé à un contexte de production homogène – un projet, une équipe – et la complexité pourra être gérée au niveau organisationnel. Le logiciel Scenari4 issu du projet C2M implémente ainsi des fonctions de gestion collaborative au sein de l'atelier, adaptées au contexte transclusif, telles que la sécurisation des contenus par l'historisation systématique, la planification de tâches, l'annotation... [*ibid.*]

4 Conclusion. Entre documents et données

En conclusion de ce chapitre, nous souhaitons ouvrir la réflexion sur l'articulation entre données et documents. Les champs de l'ingénierie documentaire et de l'ingénierie des données tels qu'ils sont appréhendés aujourd'hui sont traditionnellement séparés en informatique. Cette séparation remonte historiquement aux inventions parallèles chez IBM du modèle relationnel par Codd en 1970 et du langage de balisage GML par Goldfarb en 1969, qui donneront respectivement naissance au début des années 1980 aux systèmes de gestion de bases de données relationnelles et aux premières chaînes éditoriales SGML et LaTeX. Cette dichotomie technologique historique tend à s'atténuer progressivement depuis l'apparition d'XML en 1998, dont les usages sont devenus massifs aussi bien du côté des données que des documents.

Nous avons proposé dans la première partie de ce chapitre que le document numérique n'existait pas réellement en tant que document, que n'existaient que des ressources – des *données* – manipulées de façon à reproduire des formes ayant des propriétés assimilables à celles d'un document. Les principes de gestion de la rééditorialisation que nous avons décrits ne font que renforcer cette tendance de réduction du document numérique à de la donnée. Génération de formes publiées à partir de formes génératrices abstraites, relations de transclusion entre fragments, dérivation, déclinaison, ne sont finalement que des organisations particulières de données adaptées au contexte documentaire. Le fragment n'est que données du point de vue de la FG qui le transclut, la FG n'est que données du point de la transformation qui calcule une FP, la FP n'est que données du point de vue du programme qui permet sa lecture ou son impression – seul finalement le papier imprimé, coupé de son héritage numérique, pourra être pleinement considéré comme un document.

La frontière entre document et donnée est artificielle, elle est une construction intellectuelle qui permet de mieux penser des technologies orientées vers la production et l'interprétation de contenus par des hommes.

Il est intéressant de prolonger ce rapprochement pour étudier certains champs du documentaire en émergence à l'aune de champs déjà bien labourés dans le domaine des données, et ainsi continuer d'interroger le spectre fonctionnel des chaînes éditoriales. Nous proposons de terminer sur quelques exemples emblématiques de cette *tendance*.

EDI ou l'échange de documents et de données

L'EDI (*Electronic data interchange*) s'intéresse à l'échange de données entre systèmes informatiques, l'enjeu étant l'interconnexion de systèmes qui ne peuvent pas être intégrés informatiquement, car trop éloignés. L'approche principale est de standardiser la structure des données en fonction des domaines afin de favoriser l'import/export.

Dans le domaine documentaire, la même problématique émerge dès lors que la formalisation des contenus les rend plus facilement manipulables. Le standard OASIS DITA [17] a ainsi pour vocation de favoriser les échanges de documentation technique entre les organisations. Le schéma pivot élaboré pour UNIT [12] a pour vocation de permettre les échanges de fragments documentaires entre universités utilisant des schémas différents à l'échelle d'une université numérique thématique (UNT), tandis que le schéma inter-UNT élaboré par la suite permet les échanges entre UNT. Le projet SUP-E-educ (2012-2014) réunit des éditeurs logiciel et des universités françaises pour élaborer une chaîne de création, de gestion et de diffusion de contenus pédagogiques ; un de ses enjeux est d'instrumenter l'échange de contenus réutilisables entre les établissements de l'enseignement supérieur.

L'ETL ou la migration de documents et de données

L'ETL (*Extraction, Transformation, Loading*) s'attache à la migration des données entre bases de données, par exemple pour l'alimentation d'entrepôts de données constitués à des fins d'analyse décisionnelle à partir de systèmes opérationnels d'entreprises. L'approche consiste à récupérer les données des systèmes sources dans leurs formats d'origine, puis à les convertir afin qu'ils respectent les contraintes des formats cibles des systèmes à charger.

La question de la migration des contenus est fondamentale dans le domaine des chaînes éditoriales, surtout si l'on considère la masse de documents bureautiques qui sont potentiellement à récupérer. Cette question se pose dans une part significative des projets menés avec Scenari : comment reprendre l'existant dans les formats d'une chaîne éditoriale ? La difficulté est qu'un processus d'ETL suppose en général des niveaux de structuration différents de l'information, qui impliquent potentiellement des rejets, c'est-à-dire des cas que l'on ne sait pas traiter automatiquement. Dans le domaine documentaire, un problème typique est que le niveau de structuration disponible dans les traitements de texte bureautiques n'est pas assez détaillé, ni assez rigoureux, pour inférer automatiquement une structuration logique pertinente dans tous les cas. La difficulté est de repérer les cas problématiques pour ne donner que ceux-là à traiter aux opérateurs humains, et limiter le recours à des experts du contenu.

À noter que, dans le cadre des projets Scenari, la totalité des processus de récupération mis en place à ce jour ont été manuels, à savoir le copier/coller par des opérateurs de saisie depuis les documents bureautiques vers la chaîne éditoriale. Cette approche efficace pour la récupération de volumes relativement modestes (jusqu'à plusieurs milliers de pages néanmoins), trouvera ses limites dans des contextes de migration qui se rapprocheront de ceux traditionnellement gérés dans le monde des données.

La documentarisation des données

Les bases de données rejoignent les chaînes éditoriales dès lors que l'on souhaite publier des états, c'est-à-dire des extractions de données mises en forme pour une lecture humaine. Il n'y a pas de différence fondamentale entre un état calculé à partir de tableaux de données et une FP calculée à partir d'une FG, c'est-à-dire un arbre XML.

Or les bases de données rencontrent les mêmes problèmes de documentarisation de ces publications, en particulier comment les pérenniser et les contextualiser afin d'en rendre l'usage pertinent, tout en assurant que les contenus ainsi mis à disposition sont à jour eu égard à la dynamique de production.

Dans le domaine du PLM (*Product Lifecycle Management*), les données de conception gérées dans des PDM (*Product Data Management*) doivent être documentarisées au sein de multiples documents de travail et de référence : cahier des charges, dossiers de conception, nomenclatures de fabrication, documentation utilisateur, etc. Les problèmes de gestion des mises à jour de ces données recoupent alors celles du documentaire [8]. Les concepts de document-dossier, d'ateliers et de bibliothèques pourraient trouver des extensions intéressantes dans la façon dont est gérée la mise à disposition de l'information dans ce type de systèmes.

La « donnéeisation » des documents

Pour finir nous évoquerons le web sémantique et le web de données, dont l'objectif commun consiste à donner aux machines, sur le web, accès aux informations existantes afin d'automatiser les traitements documentaires (indexation et recherche typiquement). L'enjeu ici est donc le complément exact du précédent : il s'agit de transformer un contenu destiné à l'interprétation humaine sous la forme de structures de données manipulables par des machines.

La voie que nous avons défendue dans ce chapitre, consistant à définir des structures favorisant la manipulation automatisée, poursuit finalement un objectif assez proche, et sera particulièrement favorable à des traitements de ce type, traitements que, en réponse à la documentarisation, nous pourrions appeler « donnéeisation ».

Remerciements

Mes remerciements à Sylvain Spinelli pour sa contribution à la mise au point de ces concepts en général et à la relecture de cet article en particulier.

Références

- [1] Jacques ANDRE, Richard FURUTA, Vincent QUINT. *Structured documents*. Cambridge University Press, 1989
- [2] T. ARRIBE, Stéphane CROZAT. « Chaînes éditoriales numériques : allier efficacité et variabilité grâce à des primitives documentaires ». In : *Actes du Colloque international sur le document électronique, CiDE.15, « Métiers de l'information, des bibliothèques et des archives à l'ère de la différenciation numérique »*, Tunis, Tunisie, 1-3 novembre 2012. [À paraître]
- [3] Bruno BACHIMONT. « L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle ». In : Jean Petitot, Paolo Fabbri, éd. *Au nom du sens*. Paris : Grasset, 2000
- [4] Bruno BACHIMONT. *Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'HDR, Université de technologie de Compiègne, 2004
- [5] Bruno BACHIMONT. *Ingénierie des connaissances et des contenus*. Paris : Hermès Science Publications : Lavoisier, 2007
- [6] Bruno BACHIMONT. « Audiovisuel et numérique : la reconstruction éditoriale des contenus ». In : Lisette Calderan, Bernard Hidoine, Jacques Millet, coord. *Métadonnées : mutations et perspectives : séminaire Inria*. Paris : ADBS éditions, 2008
- [7] Bruno BACHIMONT. « Archivage audiovisuel et numérique : les enjeux de la longue durée ». In : Corinne Leblond, dir. *Archivage et stockage pérennes*. Paris : Hermès Science Publications : Lavoisier, 2009
- [8] Matthieu BRICOGNE, Louis RIVEST, Nadège TROUSSIER, Benoît EYNARD. « Towards PLM for mechatronics system design using concurrent software versioning principles ». In : *Proceedings of the 9th International Conference on Product Lifecycle Management, PLM'12, Montréal, Canada, July 9-11, 2012*
- [9] Suzanne BRIET. *Qu'est-ce que la documentation ?* Paris : Éditions documentaires, industrielles et techniques, 1951
- [10] Michael K. BUCKLAND. *Library services in theory and context*. Oxford : New York : Pergamon Press, 1983. 2nd ed. 1988
- [11] Vannevar BUSH. « As we may think ». *The Atlantic Monthly*, 1945

- [12] Stéphane CROZAT, Nicolas DELESTRE, Jacques QUEYRUT, Fabien BAILLON, Pascale GAUTRON, Christine VANOIRBEEK, Priscilla VELUT. « Standardisation des formats documentaires pour les chaînes éditoriales d'UNIT : un schéma pivot ». In : *Actes du colloque « Technologies de l'information et de la communication dans l'enseignement supérieur et l'entreprise (TICE 2006) »*, Toulouse, France, octobre 2006
- [13] Stéphane CROZAT. *Scenari, la chaîne éditoriale libre*. Paris : Eyrolles, 2007
- [14] Stéphane CROZAT, Bruno BACHIMONT, Isabelle CAILLEAU, Serge BOUCHARDON, Ludovic GAILLARD. « Éléments pour une théorie opérationnelle de l'écriture numérique ». *Document numérique*, 2011, vol. 14, n° 3
- [15] Stéphane CROZAT. *Notion de document-dossier. Rapport L2a du projet C2M, version 1.2*. 2011. http://scenari.utc.fr/c2m/DOCS/L2a/pdf/c2m_L2a_20110414.pdf
- [16] Stéphane CROZAT. « Structured and fragmented content in collaborative XML publishing chains ». In : *Proceedings of the 12th edition of the ACM Symposium on Document Engineering, DocEng 2012, Paris, France, 4-7 septembre 2012*. [À paraître]
- [17] DITA. *Darwin Information Typing Architecture (DITA) Version 1.2. OASIS standard*. 2010. <http://docs.oasis-open.org/dita/v1.2/spec/DITA1.2-spec.html>
- [18] Gérard DUPOIRIER. *Technologie de la GED : techniques et management des documents électroniques*. Paris : Hermès Science Publications : Lavoisier, 1995
- [19] Ludovic GAILLARD. *Modélisation rhétorique pour la publication de discours multimédias : applications audiovisuelles*. Thèse, Université de technologie de Compiègne, 2010
- [20] Audilio GONZALES-AGUILAR, Maria RAMIREZ-POSADA, Stéphane CROZAT. « Scenari–Opale : cadena editorial digital para la producción de contenidos e-learning ». *El profesional de la información*, 2012, vol. 21, n° 4
- [21] Jack GOODY. *La raison graphique : la domestication de la pensée sauvage*. Paris : Éditions de Minuit, 1979
- [22] Jack GOODY. *Pouvoirs et savoirs de l'écrit*. Paris : La Dispute/SNÉDIT, 2007
- [23] André LEROI-GOURHAN. *Milieu et techniques*. Paris : Albin Michel, 1945. 2^e éd. 1973
- [24] Françoise LEVIE. *L'homme qui voulait classer le monde : Paul Otlet et le Mundaneum*. Bruxelles : Les impressions nouvelles, 2006
- [25] Theodor. H. NELSON. *Literary Machines*. Mindful Press, 1981
- [26] Paul OTLET. *Traité de documentation : le livre sur le livre. Théorie et pratique*. Bruxelles : Van Keerberghen, 1934. [Reprod. en fac-sim. : Liège : Centre de lecture publique de la Communauté française de Belgique ; Bruxelles : Ed. Mundaneum-Palais mondial, 1989]
- [27] Roger T. PEDAUQUE (collectif). *Document : forme, signe et médium, les re-formulations du numérique*. 2003. http://archivesic.ccsd.cnrs.fr/sic_00000511
- [28] Roger T. PEDAUQUE (collectif). *Document et modernités*. 2006. http://archivesic.ccsd.cnrs.fr/sic_00001741
- [29] Dominique SAINT-MARTIN, Stéphane CROZAT. « Écouter, approfondir : perspectives d'usage d'une radio interactive ». *Distances et savoirs*, 2006, vol. 5, n° 2
- [30] Jean-Michel SALAÜN. « La redocumentarisation, un défi pour les sciences de l'information ». *Études de communication*, 2007, n° 30, « Entre information et communication, les nouveaux espaces du document »
- [31] Gilbert SIMONDON. *Du mode d'existence des objets techniques*. Paris : Aubier, 1958
- [32] Bernard STIEGLER. *La technique et le temps*. Tome 1 : *La faute d'Épiméthée*. Paris : Galilée, 1994
- [33] Manuel ZACKLAD. « Processus de documentarisation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées ». In : *Actes du colloque EBSI-ENSSIB « Le numérique : impact sur le cycle de vie du document pour une analyse interdisciplinaire »*, Montréal, Québec, 13-15 octobre 2004 [document numérique]. Villeurbanne : ENSSIB, 2004
- [34] Manuel ZACKLAD. « Réseaux et communautés d'imaginaire documédiatisées ». In : R. Skare, W. L. Lund, A. Varheim. *A Document (Re)turn*. Frankfurt am Main : Peter Lang, 2007