



HAL
open science

L'édition électronique scientifique, son impact dans le traitement de l'information d'un centre de documentation

Sylvie Gresillaud, Jacqueline Gillet, Catherine Morel-Pair

► **To cite this version:**

Sylvie Gresillaud, Jacqueline Gillet, Catherine Morel-Pair. L'édition électronique scientifique, son impact dans le traitement de l'information d'un centre de documentation. 2ème conférence Document numérique et Société, 2008, Ghislaine Charton; Évelyne Broudoux, Nov 2008, Paris, France. pp.279-298. sic_00359634

HAL Id: sic_00359634

https://archivesic.ccsd.cnrs.fr/sic_00359634

Submitted on 9 Feb 2009

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.

L'édition électronique scientifique, son impact dans le traitement de l'information d'un centre de documentation

Sylvie GRESILLAUD¹, Jacqueline GILLET¹, Catherine MOREL-PAIR¹
(1) INIST-CNRS, Vandoeuvre-lès-Nancy, France

Résumé :

L'INIST, l'Institut de l'Information Scientifique et Technique du CNRS, s'adapte continuellement aux nouvelles technologies. Acquérir un fonds documentaire, le gérer, l'exploiter et le valoriser demande, à toutes les étapes, des compétences orientées vers l'informatique documentaire. Nous verrons comment les différents métiers de l'INIST-CNRS ont été modifiés au cours de ces dernières années, en particulier pour l'acquisition de documents électroniques, de leurs données de signalement et la transformation de celles-ci, les négociations et les aspects juridiques concernant leurs droits d'exploitation imposés ou devant être discutés directement avec les éditeurs.

Ces pratiques d'ingénierie documentaire nécessitent une méthodologie favorisant la mutualisation des ressources et outils de transformation ; les techniciens et ingénieurs bénéficient actuellement d'un déploiement de compétences technologiques initié il y a déjà quelques années.

Mots-clés : chaîne documentaire, ressource électronique, fourniture de documents, compétences, INIST-CNRS.

Pour assurer ses missions de collecte, traitement et diffusion de l'information scientifique, l'INIST, l'Institut de l'Information Scientifique et Technique du CNRS, s'est doté depuis plus de 20 ans d'une chaîne documentaire adaptée. Les technologies évoluent, les acteurs changent ; l'INIST s'adapte continuellement.

Dans les années 90, ses principales activités étaient la constitution des bases de données bibliographiques PASCAL (Sciences, Technique et Médecine) et FRANCIS (Sciences Humaines et Sociales) et la fourniture de copies de documents. Pour réaliser ses missions l'INIST s'appuie sur un fonds documentaire de près de 20 000 titres de périodiques et de plus de 300 000 monographies (actes de congrès, ouvrages et thèses).

Dès 2000, avec l'arrivée de services de diffusion de versions électroniques de revues, l'INIST s'oriente vers l'acquisition et la mise à disposition de ces nouvelles formes de publication à travers des portails documentaires généralistes ou thématiques. Le service d'édition électronique est également né à cette époque, dans le cadre de la diversification des activités d'offres de service personnalisés et intégrés aux utilisateurs de l'Institut. En parallèle de ces décisions, une réflexion sur l'impact de l'édition électronique s'engageait, tant au niveau des métiers que de la structure organisationnelle et informatique. Cette évolution des services a servi de modèle à d'autres instituts francophones tels le CERIST à Alger et l'IMIST à Rabat.

Nous verrons dans cet article comment l'INIST s'est adapté à ce nouveau support de diffusion, en particulier les évolutions de la chaîne documentaire et des services, les nouvelles compétences mises en œuvre et les métiers créés. Cela concerne non seulement les aspects techniques mais aussi les aspects juridiques et organisationnels.

Nous commencerons par cadrer les activités autour de la chaîne documentaire, puis nous poursuivrons avec les activités du département Edition Electronique, qui, grâce aux compétences diverses développées au sein de l'équipe, propose à tout éditeur la production et la diffusion d'une version électronique, en accès libre ou réservé, et en complément ou non de la version papier. Nous verrons ensuite comment ces compétences ont été déployées au sein d'autres services directement concernés par l'acquisition et le traitement de ressources électroniques externes. Tout au long de la description de ces activités, nous situerons les évolutions des métiers des personnels concernés et les placerons dans le contexte des divers référentiels actuels.

1. La chaîne documentaire

A ce jour, 20 millions de notices sont gérées à l'INIST. Le système d'information constitué d'une multitude de réservoirs de références bibliographiques – presque un par activité – fut très vite centralisé pour permettre une mutualisation des données. Que ce soit pour la production des données bibliographiques indexées par les ingénieurs de l'information spécialistes de toutes disciplines ou pour la gestion et le traitement des commandes de copies de documents, un référentiel documentaire unique et centralisé permet l'intégration et la mise à disposition de références bibliographiques plus ou moins enrichies.

Vers une simplification de la circulation des flux d'information

D'abord en SGML, ce référentiel fut normalisé en XML. Son alimentation provient de données de catalogage générées lors de la réception des publications puis complétées au niveau analytique. Il est également alimenté par des données provenant de partenaires. Historiquement, l'INIST reçoit des références bibliographiques de coopérants participant à la production des bases PASCAL et FRANCIS. L'INIST fournit également des données à ses partenaires diffuseurs des bases de données. Lors de la refonte du système d'information, comme indiqué sur la figure 1, le cœur est le référentiel documentaire « RefDoc ». Il est alimenté par les flux issus de l'outil de gestion des acquisitions du fonds documentaire Millénium, complétés par ceux de la sous-traitance et partenaires produisant les données de catalogage. RefDoc alimente, en parallèle, les outils de la chaîne de traitement de la Fourniture de documents (nous en détaillerons quelques points plus loin), les différentes plateformes de production des bases de données, avec Mefisto et I-Station pour l'indexation puis la « base produit » pour la génération des flux destinés aux distributeurs de ces bases ainsi que les catalogues mis à disposition du public tel Article@INIST. Il est structuré de la façon la plus fine possible de manière à intégrer tous les niveaux d'information de tous les documents gérés. Les données relatives aux documents électroniques y sont également présentes. Cependant, cet affinement paraît complexe pour élaborer des échanges avec les différents partenaires. Une version simplifiée du modèle de données est alors écrite ; il s'agit de la dtd Exodic (<http://exodic.inist.fr>), profil d'application du standard de métadonnées Dublin Core simple.

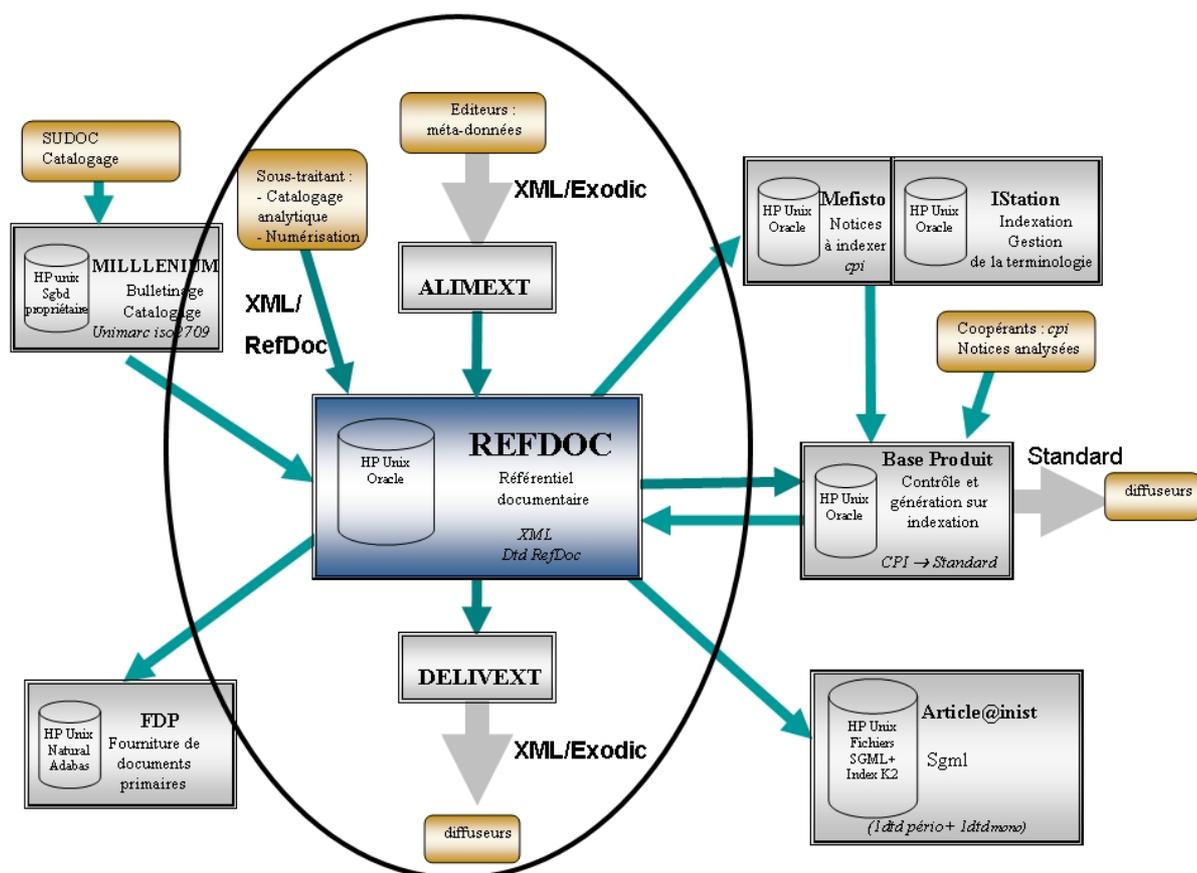


Figure 1 - La chaîne documentaire de l'INIST – CNRS

L'intégration de données d'éditeurs

Cette simplification s'est accompagnée d'un ajout d'informations permettant à l'utilisateur d'accéder plus facilement à la ressource primaire. Plusieurs solutions ont été retenues : développement par les distributeurs d'outils d'accès aux documents électroniques, par exemple Ovid qui propose ce service pour toutes les bases de données qu'il diffuse telles MEDLINE ou INSPEC, implémentation d'outils résolveurs de lien sur les plateformes hébergées à l'INIST dans la cadre des portails ou intégration d'un identifiant de type DOI ou url dans les notices bibliographiques avec interaction sur cette donnée. Une première coopération¹ de ce dernier cas a été développée en 2005 pour alimenter la base de données FRANCIS : la collection ÉRUDIT (www.erudit.org) rassemble un large éventail de disciplines des sciences humaines et sociales, ainsi que des sciences naturelles. L'accord prévoit d'établir des liens entre les notices des bases de données bibliographiques de l'INIST et les articles récemment parus dans les revues publiées en version électronique et diffusées sur le site d'ÉRUDIT. Il permet ainsi l'enrichissement des notices des bases de données par l'introduction du lien et des métadonnées correspondantes, vers le texte intégral de revues universitaires majoritairement francophones. Pour absorber un tel flux, le département des systèmes d'information conçoit une application (ALIMEXT – ALIMENTation EXTérieure) permettant le traitement des métadonnées de cet éditeur. Des coopérations sont en cours d'élaboration avec d'autres éditeurs.

¹ <http://www.inist.fr/article64.html>

La transformation se fait en deux étapes principales. La première est celle du format éditeur vers le format pivot (Exodic), la seconde consiste à convertir le format pivot en format cible (RefDoc). Les notices possèdent alors toutes les informations nécessaires conformes à la structuration de RefDoc ainsi que les (jeux de) caractères utilisés et les identifiants internes, externes (DOI, PII...) et les indicateurs de traitement des documents à l'INIST. ALIMEXT garantit, pour un éditeur donné, l'unicité des articles avant le transfert vers le référentiel RefDoc ; en cas de non-conformité, les notices sont mises en rejet en fonction du défaut constaté. L'application produit également une notice « série » selon le format du catalogue d'acquisition du fonds documentaire Millenium, si le titre n'appartient pas encore au fonds. Elle gère la correspondance entre ISSN papier et ISSN électronique si elle existe. Elle mémorise les informations d'identification nécessaires à tous les niveaux de description du document : série, fascicule et article. Ces informations permettent de faire à tout moment un état des collections numériques.

Ces travaux devant être reproduits avec d'autres éditeurs dans le cadre de l'activité de fourniture de documents, cette application a dû passer en phase d'industrialisation de façon à pouvoir absorber des flux hétérogènes et volumineux. De manière à rendre utilisable simplement l'application ALIMEXT, une organisation appropriée s'est mise en place. Une entité de spécialistes des métadonnées documentaires, connaissant l'environnement XML et en particulier le langage de transformation XSLT², analyse les métadonnées fournies par les éditeurs partenaires et exprime les spécifications de conversion du format éditeur vers Exodic sous forme de feuilles de style. Les membres de cette entité sont soutenus si nécessaire par des informaticiens pour les aspects de conception des feuilles de style. Comme pour tout projet informatique, l'entité est en charge de « recetter » les métadonnées alors transformées avant leur mise en production tout en s'assurant de l'exhaustivité de l'intégration des articles édités. Enfin, elle analyse les éventuels rejets et propose des solutions.

ALIMEXT répond à une démarche d'intégration des métadonnées paramétrable et spécifique, si besoin, à chaque type de flux. En effet, ces flux sont généralement propres à un éditeur.

De manière à appréhender les différentes étapes décrites plus haut, les spécialistes documentaires ont du(e)s être formé(e)s. Il existait au sein de l'INIST une première équipe qui avait l'expérience du traitement du document électronique et de ses métadonnées, celle du département Edition Electronique, dont l'activité est décrite ci-après.

2. La production de publications électroniques

Rappelons tout d'abord les acteurs de la chaîne de production éditoriale (figure 2) ainsi que les différentes étapes de production (figure 3).

² Nous décrivons plus précisément ces technologies dans le § Production éditoriale

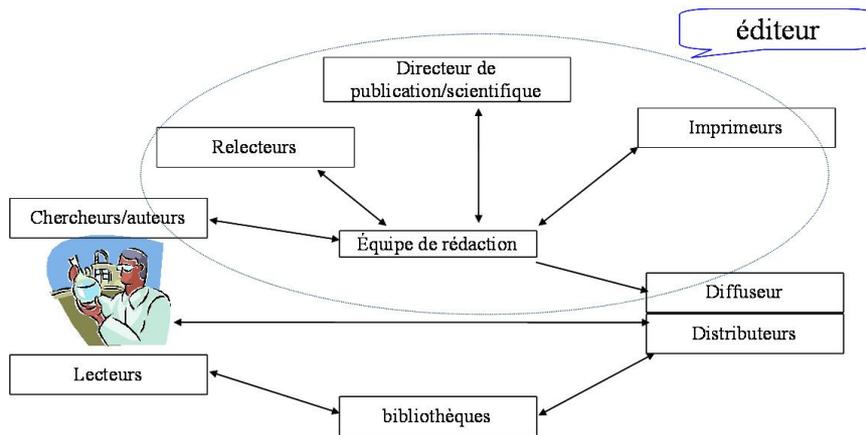


Figure 2 - les acteurs de la chaîne éditoriale

L'INIST avec le service I-Revues se place comme diffuseur. Nous verrons également qu'il existe un accompagnement de l'équipe de rédaction, et dans l'idéal des chercheurs auteurs. Les imprimeurs ont également un rôle important. Ils ont modifié leur chaîne de production de façon à prévoir le plus en amont possible la fabrication d'une version papier et d'une version électronique en parallèle. La figure 3 illustre l'exemple de JOUVE (Favreau, 2008). On peut citer également le cas de l'imprimeur nancéien BIALEC, prestataire de bon nombre de revues scientifiques.

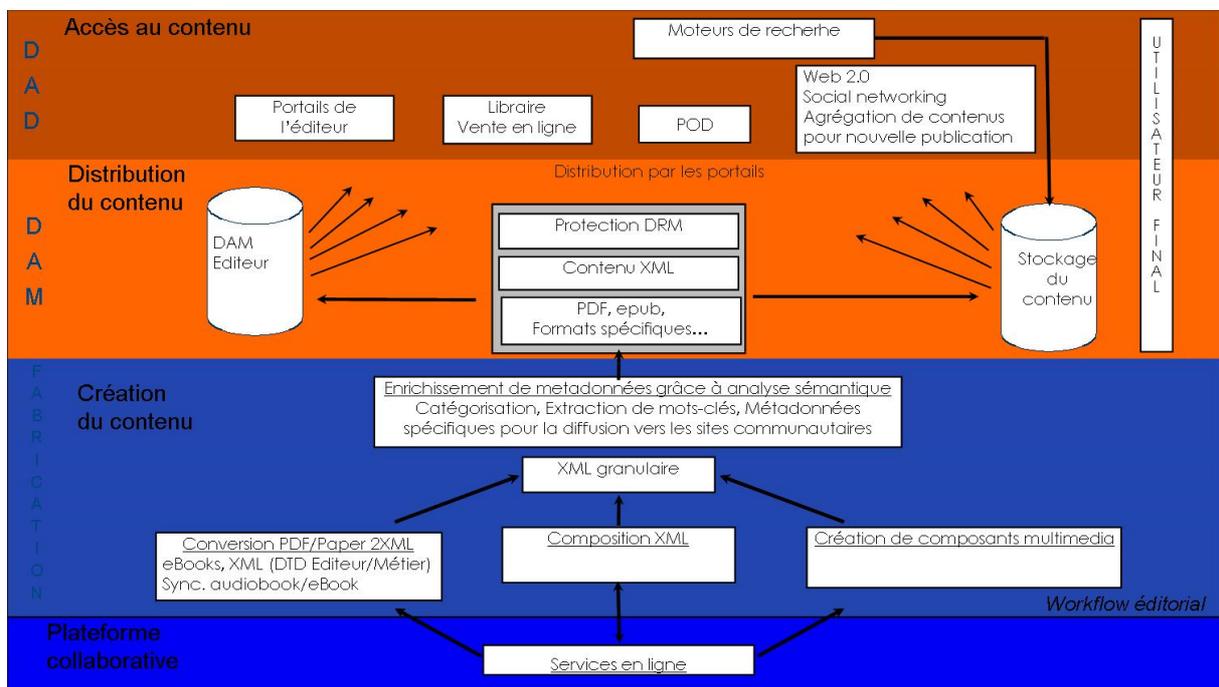


Figure 3 - Modèle de production (source V Favreau - JOUVE)

XML est au cœur de la chaîne. La production de données de référencement est prévue le plus en amont possible et celles-ci sont extraites du document lors de sa fabrication.

2.1.La structuration de contenus

Produite depuis très longtemps avec des outils informatiques, l'édition « papier » permet

aujourd'hui une production électronique sous forme de fac-similés, pdf en particulier, à la sortie de la chaîne d'impression.

Cependant, l'objectif d'une version électronique est de permettre une meilleure visibilité au niveau de la publication elle-même et de son contenu, et de proposer des processus de lecture plus diversifiés. Il n'est plus nécessaire de présenter dans ce but l'intérêt de la structuration des contenus et du langage XML.

Dès 2001, le projet de normalisation de la revue Alsic (Apprentissage des langues et systèmes d'information et de communication, revue créée en 2000 uniquement sous forme électronique)³, a permis à l'INIST d'appréhender la production de documents électroniques structurés. Les travaux ont porté sur la transformation de documents Word produits par les auteurs et validés par l'équipe de rédaction. Ils se sont basés sur les réflexions disponibles à cette époque, en particulier la chaîne Cyberthèses⁴ et le site spécifique de l'ENSSIB, <http://revues.enssib.fr/>. Le workflow portait sur la structuration du document initial en utilisant les styles de mise en forme proposés dans Word, sa transformation en XML via la suite Open Office© et enfin sa conversion en XHTML ou pdf pour la diffusion.

Toute une phase d'apprentissage démarrait dans le cadre de ce projet : le choix d'une dtd - c'est finalement DocBook qui a été retenu (Grésillaud, 2004) -, le concept XPath⁵ qui permet de désigner une partie d'un document XML et sa mise en application pour la génération de feuilles de style de transformation XSL-T et XSL-FO⁶. Les archives de la revue, disponibles uniquement sous forme HTML, ont été elles aussi normalisées avec d'autres feuilles de style. La structuration a été poussée au plus fin, de façon à pouvoir valoriser les contenus par des usages en recherche scientifique sur le texte et en infométrie. Les deux ingénieurs en charge de ce dossier ont établi et documenté des procédures pour leur mise en œuvre sur les articles de la revue Alsic. Il en a été de même dans le cadre du Projet SINEDIE⁷, partenariat avec le Ministère de la Culture et de la Communication.

Dans le cadre d'une modification en profondeur de l'organisation et des profils des services de l'institut, s'est alors construite toute une démarche de « mutation technologique ». Une grande partie du personnel a été sensibilisé aux nouvelles technologies et à l'environnement du document électronique. L'expérience assimilée au sein du service d'édition électronique et du service veille, associée à l'expérimentation pour la valorisation des contenus des documents électroniques et de leurs métadonnées, a été déployée lors de sessions de formation sur l'environnement XML et incluse dans le plan de formation des personnels de l'INIST puis dans le catalogue général des formations de l'INIST. Nous verrons plus loin que d'autres formations complémentaires sont nécessaires.

2.2. La nécessité d'un format pivot

XML, connu des responsables des systèmes d'information, est devenu le cœur des processus de production des publications. Nous l'avons rapidement constaté dans les nouvelles versions des outils de fabrication tels que Xpress© ou Indesign©. Les imprimeurs ont modifié leur

³ www.alsic.org

⁴ www.cybertheses.org

⁵ recommandation maintenue par le w3c <http://www.w3.org/TR/xpath> avec version française <http://xmlfr.org/w3c/TR/xpath/>

⁶ langages de transformation de documents XML pour obtenir respectivement du texte/html/xml et pdf

⁷ <http://sinedie.inist.fr>

chaîne pour qu'il devienne le format pivot (Favreau 2008). Ce format garantit également la pérennité de l'archivage.

Comme l'indique la figure 3, des étapes sont essentielles pour valoriser la diffusion de la revue, en particulier créer des métadonnées ou enrichir celles qui sont disponibles, et les transformer. En effet, celles-ci sont soit extraites des bases de données bibliographiques de l'INIST, soit fournies par l'éditeur de la revue, et dans ce cas, les formats varient : export XML ou ASCII délimité depuis des logiciels documentaires, fichiers excel... Il faut donc les convertir en fichiers compatibles avec la plateforme d'édition retenue, en utilisant notamment des feuille de style XSL-T, souvent associées à des programmes informatiques complémentaires.

Pour son site de diffusion de publications scientifiques I-Revues, l'INIST a retenu la plateforme de publication DSpace, développée par le MIT et utilisée aujourd'hui dans le monde entier (plus de 300 applications recensées). Les métadonnées des documents y sont en Dublin Core qualifié dans la base de données interne. Elles sont exposées en Dublin Core simple dans l'entrepôt de métadonnées OAI-PMH, conforme au protocole international du même nom qui garantit l'interopérabilité des archives ouvertes, entrepôt effectivement bien enregistré sur les sites officiels et moissonné par les moteurs spécialisés comme OAIster.

Diffuser une publication sur I-Revues implique que toutes les données puissent être valorisées à travers les divers produits et services de l'INIST. Pour cela, il a été décidé d'adopter comme format pivot, le format d'échange propre à la chaîne documentaire de l'INIST, Exodic qui a connu lui aussi une évolution progressive depuis 2000, en particulier sur les données identifiant le document telles le numéro handle⁸. Depuis le format Exodic, les métadonnées sont transformées en Dublin Core et associées au document numérique avant import de l'ensemble par lots dans le système DSpace. Un ensemble de programmes appelant des feuilles de style XST-T est disponible et adapté à chaque publication.

A côté de DSpace qui permet essentiellement l'acquisition, la publication et l'archivage pérenne de ressources numériques « statiques », le département Edition Electronique commence à utiliser la plateforme d'édition Lodel. Celle-ci permet la mise en ligne de documents XHTML à partir de documents Word stylés et de divers paramétrages de personnalisation, pour des collections de ressources plus évolutives et créées en partenariat avec l'éditeur. Lodel permet également l'exposition des métadonnées en OAI-PMH.

Compte tenu des cas étudiés et mis en pratique, une véritable expertise en ingénierie documentaire s'est ainsi développée ...

2.3. Une expertise reconnue à perdurer

Connaissance des standards, de la méthodologie pour l'automatisation du traitement des données, pratique de la transformation des documents et de leurs métadonnées, veille sur les technologies, participation à des forums spécialisés et groupe d'utilisateurs ... : toutes ces activités ont apporté progressivement une expertise aux membres du service d'édition électronique. Ce n'est pas terminé : des expérimentations sur la valorisation des contenus de vidéos, sur l'illustration d'un livre électronique en médecine, démontrent que l'exploitation du multimedia est indispensable et nécessite des spécificités dans les traitements. On peut citer

⁸ Identifiant unique et pérenne attribué par le cnri (www.cnri.org). Le DOI (Digital Identifier Object) est un « handle » particulier.

également les pratiques liées à l'édition papier « à la demande », ou les technologies de diffusion répondant aux divers modèles économiques que l'éditeur détermine. Les récentes recommandations de Bernard Prost du Ministère de l'enseignement supérieur et de la recherche (Prost, 2007) le confirment. Nous verrons plus loin qu'il ne suffit pas de maîtriser la technique : il est nécessaire de connaître et suivre l'environnement des acteurs et du marché et les aspects juridiques.

3. Le service de Fourniture de Document vers l'électronique : de nouveaux métiers ?

L'utilisation d'Internet a profondément changé la relation entre l'utilisateur final et l'information.

Les exigences des chercheurs ont changé, ils veulent commander et payer plus facilement, obtenir le document immédiatement sous format électronique, ne plus se déplacer à la bibliothèque, avoir une bonne qualité de documents (format PDF versus format TIFF image) (Rahali, 2007).

Pour répondre à ces exigences, l'INIST a décidé d'adapter la fourniture de documents traditionnelle (reproduction papier) aux nouvelles technologies dans l'ensemble des processus et sa stratégie est double :

- améliorer la diffusion de l'information en utilisant Google Scholar et Google et en développant la commande en ligne et le paiement monétique,
- proposer la livraison en format PDF d'un article directement sur le poste de travail du chercheur.

Pour mettre en place de ce nouveau service « la livraison électronique sécurisée » lancé en mars 2007, d'autres documentalistes de l'INIST ont dû acquérir de nouvelles compétences liées à l'utilisation de nouvelles ressources et de nouveaux outils (Gillet, 2007).

Revenons à la chaîne documentaire pour expliquer ce changement.

Aujourd'hui, le catalogue en ligne article@inist, qui permet de signaler les titres disponibles à la fourniture de documents, est produit à partir des acquisitions papier.

Trois cents fascicules sont traités chaque jour et intégrés dans le système de gestion de bibliothèque. Comme indiqué dans le §1, les notices séries sont alors reformatées en XML et versées dans un réservoir bibliographique interne. Le catalogage des sommaires est externalisé puis mise en phase avec les notices séries dans le réservoir bibliographique. Cinq jours après leur réception, les sommaires sont intégrés dans article@inist.

La production de ces 800 000 références bibliographiques par an nécessite une externalisation de certaines activités (catalogage des revues de sommaire), un budget important et une organisation particulière en interne ; dix documentalistes gèrent les flux de données et de documents, vérifient la qualité des notices et l'intégration dans notre réservoir bibliographique (Schopfel, 2003).

Les documentalistes collectent, gèrent et exploitent le flux de documents et d'information et sont formés aux techniques de traitement physique des documents et à la gestion de production.

Avec ce nouveau projet qui est une étape supplémentaire vers la dématérialisation, l'objectif est de signaler dans nos catalogues les articles électroniques disponibles pour la livraison

électronique sécurisée lorsque la revue papier n'a pas été acquise et cataloguée.

L'option choisie a été de projeter notre chaîne de production et notre organisation dans le monde numérique en intégrant les données des éditeurs et en faisant évoluer l'application ALIMEXT vers une industrialisation.

Ce traitement des métadonnées éditeurs est une activité qui nécessite de fortes compétences technologiques et un contrôle qualité constant.

Certains éditeurs fournissent déjà leurs métadonnées à des bases de données (Pubmed) selon la DTD fournie par la NLM – National Library of Medicine, d'autres exportent les métadonnées de leur chaîne éditoriale : le format des métadonnées diffère d'un éditeur à l'autre. Ces métadonnées sont encore quelquefois en format SGML et ne sont pas forcément contrôlées par une DTD, ce qui engendre une grande hétérogénéité des entrées.

Le reformatage et l'intégration des flux sont longs et un contrôle qualité après intégration dans notre référentiel documentaire est indispensable.

L'équipe de documentalistes chargée de la gestion des flux liés aux acquisitions papier a pris en charge cette nouvelle activité en développant de nouvelles compétences, aidée et formée également par le département Edition Electronique. Cette formation s'est réalisée par étapes, et portait essentiellement sur les jeux de métadonnées standards, l'analyse des données, les principes de bases en algorithmique et l'utilisation de XSL/XSLT pour la transformation de données souvent complexes

Actuellement cinq éditeurs internationaux sont passés en production, dix-sept sont en phase de développement. L'équipe de fourniture de documents a dû aussi se former à de nouvelles technologies.

En effet, pour une majorité des éditeurs, l'utilisation d'un DRM (logiciel de gestion des droits) est une condition imposée par les éditeurs. La British Library, précurseur dans la livraison électronique sécurisée, a dès le début, adopté un DRM développé par Adobe ((Braid, 2004). Le DRM choisi par l'INIST est le logiciel FileOpen utilisé par d'autres centres de documentation à l'étranger (CISTI⁹, SUBITO¹⁰). Ce DRM est composé de trois modules, un encrypteur, un plug-in à installer dans le lecteur de document (Adobe Reader) sur le poste de l'utilisateur et un serveur de droits.

Le principe est différent du logiciel ARIEL qui est un logiciel de transmission de documents scannés entre 2 PC communiquant entre eux.

Avec un DRM, le document en PDF crypté par FileOpen est déposé sur un serveur à l'INIST, le serveur de droits est paramétré selon les conditions des éditeurs (une ou deux impressions, pas de sauvegarde) le plug-in de FileOpen est installée sur le PC de l'utilisateur. Ces trois composants communiquent entre eux pour gérer l'affichage et l'impression d'un document (Figure 4).

⁹ CISTI : Centre d'information Scientifique et Technique Canadien : <http://cisti-icist.nrc-cnrc.gc.ca/>

¹⁰ SUBITO : Réseau des bibliothèques universitaires allemandes : <http://www.subito-doc.com/fr/>

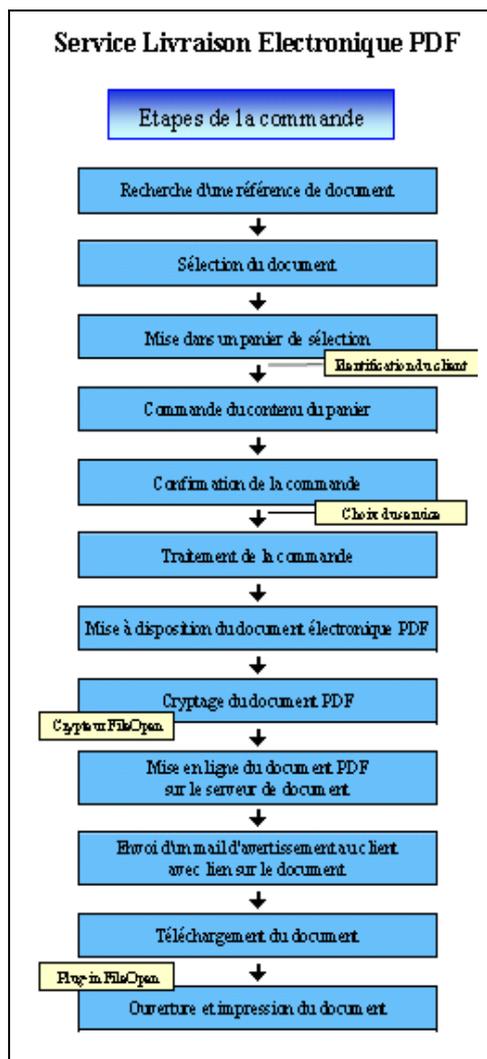


Figure 4 - Processus de traitement d'une commande

Avec cette nouvelle technologie, l'assistance technique devient de plus en plus informatique et le degré de technicité des équipes de documentalistes a largement progressé. Ils ont étendu leurs compétences à l'administration des PC, les serveurs de droit, et l'interopérabilité des systèmes de DRM.

4. Les relations éditeurs : des compétences juridiques à acquérir

Avant de mettre à disposition et de transformer ces nouvelles ressources, il faut bien sûr acquérir le droit de le faire.

L'INIST pour son activité traditionnelle de fourniture, s'appuie sur une vaste collection en sciences de 19 700 titres de périodiques dont 6 700 abonnements en cours, 125 000 thèses en sciences et techniques, 115 000 conférences, 75 000 rapports et 13 000 monographies.

L'INIST, soumis à la loi française, a contractualisé depuis 1995 avec le CFC¹¹ (Centre français du droit de copie) un accord permettant la reproduction par reprographie des copies de

¹¹ voir le site du CFC : <http://www.cfcopies.com/V2/>

documents (séries, ouvrages) moyennant une compensation financière fixée par le CFC chaque année

Si, aujourd'hui, certains éditeurs de presse pour les panoramas de presse diffusés sur des intranets confient au CFC la gestion de leur droit, le CFC n'a aucun mandat obligatoire pour percevoir et gérer les redevances pour les copies électroniques.

L'INIST devait donc négocier directement avec les détenteurs des droits de c'est-à-dire les éditeurs. Deux types de contrats sont alors négociés, les contrats autorisant la numérisation à partir de notre fonds papier (numérisation à la demande), ceux autorisant le téléchargement des articles à partir du site des éditeurs : pay-per-view (PPV).

Certains éditeurs ont proposé leur contrat et d'autres ont accepté le contrat type réalisé à l'INIST avec le service juridique.

L'acquisition traditionnelle via des agences d'abonnement a évolué ; analyser l'offre et le marché, prendre contact avec les éditeurs, proposer un modèle économique et négocier les droits d'auteurs demandent des compétences juridiques, des compétences en anglais et bien sûr leur combinaison pour négocier en anglais.

Le département Fourniture de Documents a une longue expérience dans le développement des partenariats notamment avec d'autres centres de documentation (British Library¹² et CISTI, ...) mais les aspects juridiques devaient être renforcés.

Un groupe d'autoformation sur les aspects juridiques liés à la fourniture de documents – droit d'auteur, copyright, contrat avec le CFC- a été animé par le responsable du département. Quinze personnes sont ainsi documentées, ont étudié et participé tous les 15 jours à des réunions qui ont permis l'enrichissement commun des connaissances sur des thèmes comme le copyright en France avec la loi DADVSI¹³, le CFC (fonctionnement, contrat), EUCD¹⁴ (loi de 2001) et l'évolution pressentie, le copyright dans d'autres pays (fair use, library privilege, les licences Editeurs, le droit de prêt).

Cette autoformation a permis d'étendre considérablement le champ des connaissances juridiques et de la gestion des contrats sans pour autant transformer les personnels en véritables juristes. Aujourd'hui 22 contrats ont pu être réalisés pour ce nouveau produit.

Le département Edition Electronique est également entré dans cette démarche ; un contrat est signé pour chaque prestation : outre le contenu de la prestation technique, les éditeurs autorisent l'INIST à utiliser les documents pour une diffusion électronique et garantissent qu'ils disposent de tous les droits nécessaires pour cette diffusion. Il a fallu acquérir des compétences pour répondre à toutes les interrogations des responsables des publications, qui, dans la plupart des cas, n'ont pas à leur disposition de structures juridiques et apporter ainsi une expertise.

5. Les enjeux de ces évolutions

Comme indiqué au début de cette communication, les bases de données et le catalogue d'articles en ligne de l'INIST sont constitués de notices bibliographiques stockées dans le réservoir du référentiel documentaire (RefDoc).

Le système de gestion de la fourniture de documents (application FDP) contient une base d'articles issus des revues du fonds INIST qui permet l'identification automatique et facilite

¹² British Library : voir le site web : <http://www.bl.uk/>

¹³ DADVSI : Droit d'auteur et droits voisins dans la société de l'information Droit d'auteur et droits voisins dans la société de l'information : <http://www.legifrance.gouv.fr>

¹⁴ EUCD : European Union copyright directive

l'identification manuelle des commandes.

La plus grande partie des données (catalogage analytique d'articles de revues) est issue d'une chaîne de production « papier » dont 99,5% est sous-traitée. (figure 1)

Même si cette chaîne de production a subi plusieurs modifications, sa nature est restée identique depuis plus de 10 ans : capture des données dans des documents imprimés, segmentation en champs bibliographiques, transformation des données en texte (OCR, saisie directe), livraison dans un format exploitable par les applications de l'INIST.

Actuellement, le délai moyen de la production d'une notice avant l'intégration dans RefDoc est de 8 à 10 jours ouvrés après la réception de la revue, plus long pour les monographies. La qualité de saisie se situe entre 99,95% et 99,98% pour les notices sous-traitées (OCR).

Avec le développement de l'édition scientifique et la part croissante des revues électroniques en ligne, notamment en STM, cette chaîne de production n'est plus en phase avec les nouvelles technologies et la production d'autres bases de données (Medline, SCI, INSPEC etc.) et ne permet pas le signalement des ressources électroniques négociées pour la fourniture de documents.

L'intérêt de l'adaptation de notre chaîne de production avec l'intégration des métadonnées est incontestable. Les enjeux sont économiques (plus de dépenses pour l'externalisation du catalogage) ; la qualité (pas d'erreurs d'OCR ou de saisie) est aussi concernée ainsi que la mise à disposition rapide du signalement (souvent, métadonnées et fichiers en ligne sont disponibles avant la diffusion par l'éditeur de la version imprimée).

Quant à la mise en ligne de revues et congrès par le département Edition Electronique, elle correspond à un service complètement nouveau à l'INIST, puisque l'utilisateur accède directement et immédiatement depuis son poste au texte intégral des publications, sans intervention d'un intermédiaire ; cette évolution est évidemment fondamentale aujourd'hui.

Lors du prototypage pour la mise en ligne, les relations avec les éditeurs qui nous confient leurs publications sont personnalisées, ce qui nous permet de mieux appréhender toute la chaîne éditoriale.

Le souci d'un service de qualité à la communauté scientifique, au travers d'une bonne visibilité, de conformité à des standards et de la pérennité de la mise en ligne nous a conduits aux choix techniques actuels, et notamment à entrer de plain-pied dans le mouvement du libre accès et des archives ouvertes.

Les activités de veille sur les formats de documents, les standards de métadonnées ou d'archivage et les outils de publication sont particulièrement importants dans cet environnement rapidement évolutif.

Ces évolutions techniques ne peuvent pas s'opérer sans une évolution des compétences. L'enjeu n'a pas été de trouver les formations adéquates ou de réfléchir à une nouvelle organisation mais parfois de motiver les personnels à faire évoluer leurs compétences métier - catalogage, développement des collections et référencement - vers des compétences appartenant à d'autres métiers, notamment l'informatique, la négociation, le juridique.

La mutation est en cours mais peut-on parler de nouveaux métiers car le cœur du métier reste, nous semble-t-il toujours inchangé : c'est acquérir, traiter, organiser, mettre à disposition l'information de manière adaptée à l'environnement.

Conclusion

Ce mouvement ne fait que commencer. L'INIST doit également s'intéresser de plus près aux autres types de ressources, images, multimédia, et continuer à veiller sur l'évolution des standards. Ainsi, les personnels vont pouvoir suivre dans les années à venir certains modules des enseignements de licence professionnelle donnés à Nancy.

Le métier de documentaliste devient de plus en plus transversal, en collaboration avec les informaticiens pour une mutualisation des compétences.

Ces transformations ont été en partie intégrées dans le référentiel des métiers REFERENS (REFérentiel des Emplois-types de la Recherche et de l'Enseignement Supérieur) répercuté sur les fiches « emplois type du CNRS de la branche d'activités professionnelles du CNRS (BAP) F Information¹⁵. Par exemple dans la fiche relative aux ingénieurs de recherche du secteur édition, ce qui a été décrit dans cet article fait partie des « savoirs généraux, théoriques ou disciplinaires ». Cependant dans la fiche relative aux emplois de Responsable de ressources documentaires, la notion de XML n'apparaît pas. Ceci peut apparaître comme une lacune pour les métiers de l'INIST situés à la fois sur les deux types.

De nouveaux stages apparaissent sur les plans de formation des chercheurs. Cela a déjà été évoqué lors de l'école thématique « Carrefour des acteurs de l'Édition en Archéologie »¹⁶. Quant aux documentalistes ou acteurs de l'édition, les axes sont de plus en plus spécifiques et orientés vers l'informatique documentaire. Nos collègues marocains utilisent pour désigner les métiers de la documentation la dénomination d'informatiste (spécialiste en science de l'information). Ce terme est approprié : on y entend la notion d'informatique ... La frontière devient plus en plus faible. Chaque année il devient de plus en plus difficile de trouver les acteurs pouvant répondre aux demandes exprimés dans les plans de formation. Certains personnels, comme nous l'avons évoqué tout au long de cette communication, deviennent formateurs pour d'autres membres de l'institut mais aussi pour des documentalistes, chargés d'édition et autres acteurs traitant de l'information scientifique. Il est donc nécessaire d'anticiper ces besoins.

¹⁵ Voir www.sg.cnrs/drh/publi/referens/referens.htm

¹⁶ Voir www.edsh.cnrs.fr/spip/IMG/pdf/SyntheseCarrefourEdition.pdf

Bibliographie

Accart, J.-P., Réthy, M-P., « Le métier de documentaliste », Electre-Ed. du Cercle de la Librairie, 3^e édition, juin 2008

Braid, A. (2004), « The use of a digital rights management system in a document supply service », *Interlending and Document Supply*, Vol 32 No 3, pp 189-191.

Favreau V, Quilbé JM (2008) « L'édition fait sa mue à l'ère du numérique : stratégies, organisation, "métiers", technologies, nouvelles offres éditoriales. », RPIST 2008 <http://rpist.inist.fr>, 16 juin 2008

Gillet, J. (2007), « Le service de livraison électronique sécurisé à l'INIST », 73rd World Library and information congress, 19-23 august 2007, Durban, South Africa ; <http://www.ifla.org/IV/ifla73/papers/096-Gillet-trans-en.pdf>

Goettmann, E., Gunet, M.C. (2005), "Portals enable national access", *Research Information*, No. 15, January/February, pp.21-2.

Grésillaud S., Chanier T., Barreau S., 2004, Semaine du Document Numérique la Rochelle, <http://www.univ-lr.fr/sdn2004/Documents/Presentation%20SDN.pdf>

Guichard, M. (1999), « L'INIST-CNRS: des services d'information payants et gratuits dans un cadre de service public », *Bulletin d'Information ABF*, No.184-5, pp.100-2.

Morel-Pair, C. (2007), « Métadonnées et XML. Des standards efficaces de l'environnement numérique », *Ingénierie des Systèmes d'Information*, Vol 12 No 2, pp 9-39

Prost B., (2008), « Rapport d'étude sur l'édition numérique de livres scientifiques et techniques : l'éditeur des années 2010 » Paris : La documentation française, pdf, 95 p. <http://www.ladocumentationfrancaise.fr/rapports-publics/084000335/index.shtml>

Rahali A., Bausenwein T., (2007) « FIZ Autodoc, a comprehensive solution for access to documents via corporate intranets and web portals », *Interlending and Document Supply*, Vol 35 No 3, pp 131-197.

Schöpfel, J. (2003), « INIST-CNRS in Nancy, France: a model of efficiency », *Interlending and Document Supply*, Vol 31 No 2, pp 94-103.

Schöpfel J., Ciolek A. (2005), « Au service de la recherche. L'Institut de l'Information Scientifique et Technique du CNRS », *Arabesques*, N° 40, pp 7