



HAL
open science

Entre droit d'auteur et liberté des échanges: les métadonnées, objets informatiques, objets économiques.

Michel Arnaud

► To cite this version:

Michel Arnaud. Entre droit d'auteur et liberté des échanges: les métadonnées, objets informatiques, objets économiques.. La norme numérique. Savoir en ligne et Internet, CNRS Editions, pp. 90-101, 2010, CNRS Communication. halshs-00637071

HAL Id: halshs-00637071

<https://shs.hal.science/halshs-00637071>

Submitted on 29 Oct 2011

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.

Chapitre 4

Entre droits d'auteur et liberté des échanges Les métadonnées, objets informatiques, objets économiques

Michel Arnaud

Comme l'ont montré les trois précédents chapitres, l'univers des standards pour les formats et la circulation du savoir en ligne est d'une grande complexité. Ce chapitre nous aide à comprendre les oppositions majeures qui le structurent. L'industrie de la connaissance est désormais une entité économique en plein essor, ce qui explique les débats actuels et les hésitations du législateur. Un courant défend l'application universelle de la législation sur la propriété intellectuelle tandis que l'autre s'oriente vers la liberté et la gratuité des échanges. A terme des positions de compromis devraient surgir. Mais ici l'auteur attire notre attention sur un objet informatique qui est devenu objet économique : la métadonnée. Ce terme, qui effarouche, désigne en fait une réalité très simple. C'est un ensemble organisé de rubriques pour décrire un type de document donné. Que l'on songe, par exemple, à un cahier de textes pour le collège. L'ensemble des rubriques qui le constituent est un ensemble de métadonnées. Où se trouve l'enjeu économique ? Un système dominant, ou non, est supposé drainer le marché vers lui avec tous les échanges qui en découlent. Le contre ne viendra-t-il pas d'une application de la notion de bien public international ? J.P.

L'avènement de la société de l'information dans les pays développés place le savoir au coeur du processus de production tandis que les pays en voie de développement connaissent une croissance basée sur la production industrielle avec le mouvement continu de délocalisation d'entreprises attirées par le différentiel salarial. Les interrelations entre les deux modèles de production, matériel et immatériel, s'établissent de manière complémentaire par les réseaux de télécommunications. Parce que les travailleurs manuels augmentent en nombre dans les pays en voie de développement, il n'est plus vraiment de mise pour le citoyen d'un pays développé de gagner sa vie en vendant sa force physique de travail mais plutôt sa capacité à manipuler l'information. La production de la plus-value est une opération immatérielle, avec comme critères d'employabilité les capacités intellectuelles à analyser, trier et traiter l'information. Dans ce contexte, l'accès à l'information comme matière première et sa transformation en savoir deviennent l'acte essentiel du travail demandé aux salariés des services dans les pays développés. Le rapport au savoir devient beaucoup plus conflictuel que par le passé car il se trouve au coeur du processus de partage des richesses et des rapports d'exploitation. Maintenant, les savoirs sont fongibles et ne garantissent plus un statut social. Le cerveau devient

l'outil de travail en ce sens qu'il permet de s'adapter à des situations professionnelles où toujours plus est demandé en fonction du traitement de l'information qui s'accélère sans cesse, car la croissance économique est basée sur la plus-value immatérielle. La compétition entre nations dans la course aux connaissances nouvelles ne fait que commencer.

La propriété intellectuelle peut limiter l'accès au savoir

De même que les patrons capitalistes se basent sur la possession des moyens de production pour déclarer que les biens produits leur appartiennent, éliminant les prétentions des salariés au partage du titre de propriété sur le produit industriel mis en vente, le même mouvement se dessine pour contrôler le processus de génération de la plus-value issu de la production du bien immatériel cette fois-ci, en réduisant ou en achetant la part de l'auteur-producteur. Le processus de production du savoir tend à être séparé de la propriété du bien immatériel qui en résulte. La brevetabilité est une manière de garantir le régime de propriété intellectuelle et de justifier l'exploitation des travailleurs intellectuels en les cantonnant dans leur rôle de producteurs de savoirs sans leur laisser le contrôle de leurs droits de propriété.

La défense des droits de propriété de la chaîne de production et de diffusion des biens immatériels (que ce soit les constructeurs de terminaux, les éditeurs de systèmes d'exploitation, les opérateurs télécoms, les intégrateurs d'applications et de services, les fournisseurs de contenus, la régie publicitaire) fait que le prix de vente est calculé selon un prorata incluant les droits d'auteur mais surtout les droits voisins à répartir entre les membres du consortium. Toutefois, il est plus facile de dupliquer un bien immatériel sur support électronique que de voler un bien matériel, avec un manque à gagner certain non seulement pour l'auteur mais surtout pour la chaîne de production et de distribution. La tendance se fait jour de contrôler les outils de consommation des biens immatériels afin d'éviter les pertes sèches de plus-value avec la multiplication des copies libres. Est-ce judicieux d'appliquer un modèle inspiré de la fixation du prix pour un bien industriel à un bien immatériel ? Ne faut-il pas poser la question du bien public en relation avec les procédures d'accès et de diffusion libres de biens immatériels ?

L'EUCD (directive européenne sur le droit d'auteur et les droits voisins) adoptée en 2001 au niveau européen et en cours de transposition en France, interdit le contournement de mesures techniques de protection et prévoit des amendes et des peines de prison pour les contrevenants. L'acheteur a le droit de faire une copie privée mais ne peut le faire techniquement. Ayant réuni près de 100 000 signatures s'élevant contre le principe de marchandisation des échanges sur Internet, l'initiative EUCD.info a alerté les autorités françaises sur les risques que pose le projet de loi de transposition. La loi française DADVSI (Droits d'Auteurs et Droits Voisins dans la Société de l'Information) a pour effet principal de restreindre fortement les droits des particuliers achetant des produits culturels. Le DADVSI s'oppose ainsi directement à l'exception de copie privée en autorisant les producteurs de contenu à mettre en place des mesures techniques l'interdisant, et surtout en sanctionnant les acquéreurs enclins à contourner ces mesures. Ne pas respecter une mesure de protection, écrire ou diffuser un logiciel permettant de la contourner, ou faire connaître des moyens de le faire sont assimilés à des délits de contrefaçon, passibles de trois ans de prison et 300 000 euros d'amende. Cette position est préoccupante en ce qu'elle tend à créer de la rareté dans le cas de biens intellectuels sur support électronique. Dans ces conditions, l'accès à l'information, la capacité à la traiter et à la transformer en savoir, la protection des droits au sein du régime de la propriété intellectuelle sont des éléments stratégiques à négocier entre groupes sociaux car ils sont déterminants pour leur mieux-être.

Le bien public en tant que garantissant l'accès libre au savoir

La notion de bien public selon Joseph Stiglitz (1999) en tant que bien non compétitif et inusable ne s'applique pas directement au savoir, à l'inverse de ce qu'il a écrit. Car justement le processus d'extraction de l'information pertinente et de construction de savoir sont au cœur du processus conflictuel de partage des richesses entre classes sociales. Il n'est donc pas suffisant de décréter bien public tout ce qui a trait à l'éducation pour mettre en pratique cette approche. Joseph Stiglitz lui-même note que les entreprises n'ont aucun intérêt à divulguer leurs brevets. Les ressources en ligne peuvent rapporter gros si elles entraînent une dîme chaque fois

qu'un usager y a accès pour les consulter. Le contrôle de la reproduction des oeuvres de l'esprit fait partie de celui des flux d'information. La plus-value s'opère sur l'accès aux connaissances, ce qui pose la question de la privatisation des oeuvres de l'esprit. Sous couvert de préserver les droits d'auteur et les droits voisins, il se produit une privatisation de l'accès au savoir avec un risque de monopole des grandes multinationales sur ce marché comme un autre pour elles, au détriment de l'accès au savoir pour tous basé sur autre chose que le profit immédiat.

Le contexte du contrôle de l'accès des ressources en ligne au nom du droit d'auteur et des droits voisins pose la question du bien public. Il y a en effet une opposition entre la démarche consistant à donner à une population les moyens et méthodes pour mieux traiter l'information, et les intérêts des marchands visant à privatiser le plus possible la diffusion des savoirs afin d'en tirer une plus-value suffisante. La consommation à domicile et payante des savoirs en ligne ne peut être qu'un aspect, à ne pas privilégier davantage que la mise à disposition à titre gratuit dans le cadre du service public de contenus de savoir en ligne. Si certains abus de téléchargement de biens immatériels donnant lieu à des reventes doivent être réprimés, il n'en reste pas moins que le secteur public se doit de se positionner entre le producteur prenant la place de l'auteur et le consommateur qui subit un contrôle de plus en plus poussé de ses activités privées face à son ordinateur. Seuls les pouvoirs publics peuvent trouver un équilibre entre la juste compensation des auteurs et des diffuseurs et un projet de société cohérent, prenant en compte les besoins de chacun et ne pratiquant pas de discrimination sur la base du pouvoir d'achat.

Jusqu'à présent, la fonction des bibliothèques publiques est d'offrir au grand public le moyen d'accéder aux savoirs sur support papier : les livres achetés aux éditeurs étant mis à disposition des lecteurs. Avec l'avènement de l'informatique individuelle, ces savoirs deviennent disponibles en ligne, avec la possibilité de les dupliquer sans problème. Cette facilité peut jouer un rôle positif dans la diffusion de nouveaux savoirs et compétences nécessaires à l'amélioration du niveau d'employabilité d'une population confrontée au fonctionnement de la société de l'information, où les employeurs recrutent des salariés capables essentiellement de traiter l'information et de la transformer en savoir.

Métadonnées standardisées versus Google Print

De nombreuses ressources numériques publiques destinées à des usages pédagogiques sont produites par des enseignants, des chercheurs, des animateurs TICE, des corps d'inspection de l'Éducation Nationale, le SCEREN (réseau CNDP-CRDP-CDDP), etc. Ces ressources sont disséminées sur des sites Web institutionnels à tous les niveaux (national, académique, départemental, local), associatifs et personnels. Elles constituent un corpus foisonnant mais hétérogène de ressources en ligne qui resteront méconnues si un système commun d'identification et de repérage n'est pas mis en place. La question centrale est la manière dont il faudrait procéder pour que n'importe quel moteur de recherche puisse trouver toutes les ressources disponibles sur un sujet donné d'intérêt pour un apprenant potentiel, un chercheur, un travailleur intellectuel. Ceci suppose d'utiliser des métadonnées qui permettent d'identifier tout type de contenu, et plus spécifiquement dans notre cas, des contenus éducatifs, avec un format commun décrivant le type de contenu, son auteur, la meilleure manière de l'utiliser, etc. Ce qui distingue fondamentalement une base de données classique d'un système à base d'informations structurées avec métadonnées, réside dans le fait que les métadonnées sont produites à travers un schéma de description qui ne dépend pas d'un seul fonds et d'une seule collection. Ceci permet de concevoir des systèmes d'information en réseaux dans lesquels les métadonnées peuvent directement et systématiquement faire trouver les documents appartenant à des collections disparates. Ainsi un enseignant pourrait recenser tous les cours de swahili en ligne de par le monde grâce aux métadonnées. Par contre, le producteur de métadonnées sera conduit à produire un état d'information qui décrira exhaustivement toutes les conditions nécessaires à l'indexation des documents sans aucun implicite de contexte ou de localisation des dits documents.

Difficultés de l'édition des métadonnées

L'édition des métadonnées est conçue comme une aide au travail d'indexation et de recherche des

documents pédagogiques. Venant après la définition du profil français d'application des LOM (LOMFR)¹ proposé comme norme expérimentale par l'AFNOR et à laquelle le CRIS SERIES a fortement contribué, la normalisation de l'édition de métadonnées est la prochaine étape importante en vue de la facilitation de l'usage des métadonnées permettant d'identifier rapidement et efficacement les ressources pédagogiques dont un apprenant et un enseignant ont besoin.

La première question concerne les difficultés d'utiliser les formats de métadonnées : il faut un documentaliste pour pouvoir remplir les champs des métadonnées pour chaque ressource pédagogique. Ce travail est très coûteux en expertise documentaire. Les bases de ressources pédagogiques telles qu'ARIADNE, ELENA, EDUTELLA, etc. sont basées sur les IEEE LOM et demandent beaucoup d'investissements en termes de temps d'indexation.

L'autre question est la compatibilité des profils d'application nationaux basés sur les IEEE LOM. Les différences sont notables dans la définition des champs et surtout dans les thésaurus, les listes de termes de référence permettant de les remplir. Si chaque profil national d'application des LOM compte entre 15 et 25 champs obligatoires, ce ne sont souvent pas les mêmes si on se réfère à la grille de référence des IEEE LOM. Cette difficulté a conduit EUN à définir 13 champs obligatoires pour le profil d'application de métadonnées européennes CELEBRATE.

Les éditeurs de métadonnées utilisés à présent (Lompad, Formist, Ariadne, Reload) se contentent de respecter les différents formats et proposent des passerelles entre eux pour les conversions. Un premier travail a été réalisé sous la forme d'une interface grand public appelée *EducaNext* afin de permettre à un enseignant par exemple d'indexer lui-même des ressources pédagogiques et de les rendre compatibles avec la base de ressources du CNDP : *Educasources* (cf. annexe). Le développement réalisé par le CRIS SERIES avec *EducaNext*, s'inscrit dans la démarche du SCEREN/CNDP de développer un système commun de métadonnées permettant d'accéder à toutes les ressources pédagogiques en ligne rassemblées dans la base *Educasources*. *EducaNext* ne devrait pas se limiter à un convertisseur de formats de métadonnées (LOM, LOMFR, Dublin Core, *Educasources*) mais aborder aussi les procédures permettant de remplir les champs des métadonnées de manière automatique. En effet, du point de vue d'un enseignant qui n'a pas le temps matériel de remplir les champs de métadonnées pour tous les documents à caractère pédagogique qu'il met en ligne, un éditeur automatique de métadonnées basé sur des procédures de reconnaissance sémantique serait très utile.

La création automatique des métadonnées est l'étape clé qui en permettra un large usage en évitant de passer par l'étape fastidieuse du remplissage des 71 champs des LOMFR à la main ou du moins en allégeant la procédure. Pour ce faire, il s'agit d'identifier les meilleurs algorithmes d'analyse de texte pour extraire du document étudié les mots clés les plus significatifs avec des logiciels de cartographie tels que Kartoo ou Umap par exemple. Une autre étape est le remplissage des champs des LOMFR en utilisant des références bibliographiques, des thésaurus, des profils d'usage permettant de choisir les termes les plus significatifs à placer comme métadonnées. Dans cette démarche interviennent des analyses sémantiques afin d'ajuster au mieux le contenu et ses descripteurs.

Les moteurs de recherche basés sur les métadonnées

Les enjeux ne se limitent pas comme on l'a vu plus haut à l'édition automatique de métadonnées. La question posée concerne aussi le type de moteur de recherche mis en oeuvre. Si un format international de métadonnées réussissait à s'imposer, des moteurs de recherche utilisant ce format pourraient être envisagés pour concurrencer Google. Le célèbre moteur de recherche utilise aussi des métadonnées mais cette fois-ci propriétaires et non divulguées, ce qui supprime toute possibilité de libre accès au savoir en ligne. D'après Jean Véronis, professeur de linguistique et d'informatique à l'université de Provence, l'index de Google serait constitué de cinq milliards de pages Web référencées automatiquement, soit 60% de l'ensemble des premières pages de sites Web reconnues et classées par le système à partir du calcul du nombre de visites, la prime étant donnée aux sites les plus visités.

1 Voir dans cet ouvrage les chapitres 1 et 2

Le projet Google Print concerne la mise en ligne intégrale de quinze millions de livres sans droits d'auteur parce qu'au-delà des soixante-dix ans de protection des droits de propriété intellectuelle, en provenance des bibliothèques des plus prestigieuses universités américaines (Harvard, Stanford et Michigan), de la Bodleian Library d'Oxford et de la New York Public Library, soit 4,5 milliards de pages en six ans. Mais la démarche s'est diversifiée en devenant le service Recherche Google Livres (<http://print.google.com/>) qui propose à l'internaute de lire des extraits, de parcourir certains paragraphes et d'acheter en ligne en partenariat avec Amazon et autres distributeurs en ligne comme Yahoo et son projet de bibliothèque numérique OCA (Open Content Alliance). Le résultat est que non seulement les sites américains de livres ou de bibliothèques sont privilégiés au détriment d'autres cultures car moins référencées donc moins citées dans les requêtes, mais surtout que le service Recherche Google Livres participe fortement à la privatisation de l'accès au savoir, avec ses partenaires privilégiés Amazon et Yahoo autour desquels se rassemble la quasi-totalité des éditeurs de livres de langue anglaise. Par rapport à ce mouvement de fond important et quasi irrésistible, que proposons-nous ?

Conçu selon le principe de l'exception culturelle qui a permis au cinéma français de survivre face à Hollywood, le projet de bibliothèque européenne à l'instigation de Jean-Noël Jeanneney, président de la Bibliothèque nationale de France, rassemble vingt-trois bibliothèques nationales européennes sur vingt-cinq avec l'objectif de numériser entre un et deux millions d'ouvrages sur les six millions représentant la culture européenne. Des industriels soutiennent le projet en France (Thomson, Editis, les éditions Eyrolles). Au début, il s'agissait à la fois de numériser les livres mais aussi de faire pièce à Google en tant que moteur de recherche universel. Toutefois, vu l'ampleur des financements nécessaires, la priorité est mise sur la numérisation et non pas sur le moteur de recherche qui reste encore du domaine de la recherche. En effet, l'expérience de Gallica, le site numérisé de la BNF qui compte aujourd'hui quelque 80 000 documents et plus de 70 000 images, démontre la difficulté d'utiliser les métadonnées de format commun et ouvert qui exigent de remplir les champs correspondants manuellement en attendant une indexation automatique : trente documentalistes sont employés à temps plein à cet effet depuis cinq ans. A ce rythme, il n'y a aucune chance d'arriver à atteindre un objectif opérationnel dans les cinq ans tels que ceux fixés dans le cadre de la bibliothèque numérique européenne. Pour y parvenir, il y a deux obstacles de taille. Aucune compatibilité n'est possible entre les différents formats de métadonnées utilisés dans les bibliothèques nationales des vingt-trois pays européens participants au projet si le format standardisé de métadonnées bibliographiques Dublin Core n'est pas utilisé. A notre connaissance, cet aspect n'est pas abordé dans le projet de bibliothèque numérique européenne. Si cet aspect était résolu, la démarche d'automatisation du processus de remplissage des champs de métadonnées est encore du domaine de la recherche.

Google réapparaît comme partenaire potentiel de la bibliothèque numérique européenne avec ses algorithmes d'indexation automatique. La firme de Mountain View en Californie a les moyens de ses développements informatiques (estimés à au moins 250 millions de dollars pour Google Print) du fait de sa capitalisation boursière et des revenus tirés des bandes annonces. Dans ce contexte, la question de l'accès libre aux bases de connaissances européennes n'est pas tranchée, loin de là. Il nous reste à espérer que le consortium European Schoolnet EUN qui a développé un système d'indexation des ressources pédagogiques en ligne avec des métadonnées au format commun européen CELEBRATE, s'intéresse au développement de moteurs de recherche permettant de repérer un contenu en ligne sans avoir à passer par les fourches caudines de Google.

Cet objectif est loin d'être anodin puisqu'il permettrait de libéraliser l'accès à la connaissance cette fois-ci en permettant l'exercice de la libre concurrence entre fournisseurs de contenus, faciliterait la diversité culturelle, l'apprenant gardant le contrôle de ces parcours. De cette manière, il serait possible d'éviter que l'approche à la Google ne phagocyte toutes les autres possibilités en imposant une démarche basée sur les pages les plus consultées, renvoyant à une approche monocorde de la culture dominante. Une politique publique européenne de soutien à l'édition automatique de métadonnées sous forme de développement de routines standardisées paraît plus que jamais indispensable.

Annexe Description d'EducaNext

Ses fonctionnalités principales :

- L'édition des métadonnées sous différents formats (Fichier XML, HTML ou Notice).
- La conversion des métadonnées d'un profil d'application à un autre (les profils disponibles sont LOM, LOMFR, Educasources et DublinCore).

Public visé :

- Les enseignants ou webmasters qui veulent indexer des documents via cet éditeur ;
- Les documentalistes qui veulent décrire des ressources numériques pour se constituer leur propre base, ou qui veulent convertir des métadonnées d'un profil d'application (Educasources, LOM, LOMFR, Dublin Core) à un autre.

Description : En chargeant la page d'accueil, deux possibilités s'offrent à l'utilisateur :

- Conversion, correction et visualisation de métadonnées existantes.
- Création de nouvelles métadonnées.

La conversion

- L'utilisateur saisit le format actuel des métadonnées parmi les options : DublinCore, LOM, LOMFR ou Educasources.
- Il saisit le format de métadonnées vers lequel il veut faire la conversion (parmi les options : DublinCore, LOM, LOMFR ou Educasources).
- Il saisit le format de fichier qu'il veut obtenir parmi les options : fichier XML, fichier HTML ou notice.
- Il importe le fichier contenant les métadonnées, à partir de son poste de travail utilisateur (ce fichier peut être sous forme de notice, fichier XML ou HTML), en cliquant sur un bouton *Parcourir* qui lui permettra de chercher l'emplacement du fichier en question.
- Il génère le fichier converti, d'un schéma de métadonnées sch1 sous un format de fichier ff1 vers un autre schéma de métadonnées sch2 sous un format de fichier ff2.
- Il a la possibilité d'enregistrer le fichier ou notice généré sur son poste de travail.

La visualisation et la correction

L'utilisateur bénéficie de deux types de conversion en même temps, celle d'un profil à un autre et celle d'un format de fichier à un autre, ce qui lui donne les possibilités de correction et de visualisation dans les cas suivants :

- en gardant le même schéma de métadonnées, et en convertissant le fichier d'un format XML ou HTML vers le format notice qui est le seul format lisible et modifiable par un utilisateur ;
- en gardant le même format de fichier, en l'occurrence le format « notice » et en convertissant un schéma de métadonnées dans un autre, avec les champs qui n'existent pas dans le profil destination en sous-brillance.

Création de nouvelles métadonnées

- L'utilisateur saisit le profil d'application dans lequel il veut éditer les métadonnées parmi les options : DublinCore, LOM, LOMFR ou Educasources.
- Il saisit le format de fichier dans lequel il veut éditer les métadonnées parmi les options : fichier XML, fichier HTML ou notice.
- En validant son choix, il obtient un formulaire de saisie relatif au profil d'application choisi.
- Après avoir saisi les champs du formulaire, il clique sur *générer* pour éditer les métadonnées dans le format de fichier préalablement choisi.

Certains champs devront puiser les descripteurs à partir des thésaurus déjà définis, en l'occurrence Motbis et Dewey.