

L'indexation des ressources pédagogiques numériques : un partenariat à créer entre les SCD et les services TICE au sein des universités

**Synthèse de la journée d'étude du 16 novembre 2004 à Lyon
rédigée par Monique JOLY à partir des notes prises en séance
Monique JOLY**

Note : tout au long de ce texte, nous utiliserons l'abréviation RPN pour désigner la notion de Ressource(s) pédagogique(s) numérique(s)

Cette présentation se concentre sur le concept d'objet d'apprentissage (Learning Object) qui focalise aujourd'hui un grand nombre d'attentes, que ce soit au niveau pédagogique, économique, politique ou culturel. Un des problèmes majeurs réside dans le manque de cohérence des différentes définitions, provenant principalement de la variété des points de vue adoptés pour aborder la question.

À travers différentes propositions (LOM, SCORM et IMS-LD), nous analysons trois des principales approches, apparues successivement : les langages d'indexation de données, les modèles de mise en œuvre informatique et enfin les langages de modélisation pédagogique. Ces derniers, qui font à l'heure actuelle l'objet d'études approfondies, présentent de réelles avancées en replaçant le point de vue pédagogique au centre du dispositif et en proposant de décrire l'ensemble des relations sémantiques liant les activités et les ressources de manipulation de connaissance. En particulier, cette nouvelle approche nécessite de s'intéresser aux usages effectifs des ressources d'apprentissage au sein de communautés de pratique (enseignants, formateurs et documentalistes), tant en termes de mutualisation, de réutilisation que de stratégies d'indexation.

Situer les enjeux de la journée dans les réseaux institutionnels et inter-personnels

Les discours d'ouverture de Messieurs **François DUPUIGRENET-DESROUSSILLES** (directeur de l'ENSSIB) et **Philippe GILLET** (président du PUL) situent la problématique de la journée :

- au sein des divers réseaux nationaux et locaux qu'ils animent à l'ENSSIB et au PUL,
- dans les expériences conduites par leurs établissements, en terme de démarches de mutualisation de la production des RPN,
- dans leurs préoccupations personnelles d'enseignants, en demande de RPN mutualisés, et, enfin, en signalant la convergence de la formation des professionnels de bibliothèque vers ce point.

Comprendre les enjeux de l'indexation des ressources pédagogiques numériques

Yolaine BOURDA, professeur d'informatique à SUPELEC et animatrice du groupe 4 de la CN36 (groupe de travail sur les métadonnées au sein de la commission de normalisation des technologies de l'information pour l'éducation, la formation et l'apprentissage à l'AFNOR), rappelle que nous sommes encore à une époque de pionniers, époque qui devrait rapidement déboucher sur des processus normalisés et simplifiés pour les technologies de l'éducation.

La production des RPN s'inscrit dans cette logique : coûteuses à produire et complexes à gérer dans leur cycle de vie, elles doivent être partageables et partagées (dans des environnements distribués), et pour cela elles doivent se conformer à certaines normes de description et d'indexation afin que :

- des « humains » apprenants et enseignants puissent les « re-connaître » intellectuellement,
- des ordinateurs puissent les traiter (en filtrage, en affichage, en échange) pour en augmenter l'accessibilité, la réutilisation et l'interopérabilité.

La description normalisée des RPN prend la forme de données à propos des données, autrement nommées métadonnées. Des formats déjà très élaborés existent comme LOM (Learning Object Metadata) et comme DC du DCMI (Dublin Core Metadata Initiative). Ces formats sont implémentés dans des applications concrètes sous la forme de profils d'application qui, dans certains cas, aboutissent à des interprétations divergentes. Il convient donc de continuer le patient

travail de normalisation pour fixer le plus possible un langage commun qui rendra le partage effectif et durable.

Laurent FLORY, ingénieur de recherche à l'université Lyon 1, dans le service PRACTICE en charge de la production et de la diffusion des ressources pédagogiques, décrit la RPN comme un objet pédagogique de dimension variable : le grain. Le grain peut être un très petit élément (métaphore du LEGO à une barre, comme par exemple une image ou un objet média simple) ou bien un élément beaucoup plus complexe (un assemblage de grains élémentaires, ou un gros grain indissociable en grains plus petits). L'étape d'assemblage transforme l'élément d'information (grain isolé) en une formation construite (grains assemblés). L'enjeu essentiel de la production des RPN se situe dans la personnalisation massive : il s'agit de faciliter le « réassemblage » de grains élémentaires pour créer de nouveaux cours (à l'image d'une construction LEGO).

Pour cela il faut être capable :

- de produire des RPN sous la forme de grains réutilisables,
- de mettre en place des systèmes permettant d'identifier et de retrouver ces grains,
- de construire un processus d'indexation qui prenne en compte une production nombreuse et foisonnante,
- d'indexer les RPN en temps réel pour rendre fluide le processus de production.

Vaste challenge car de nombreuses questions se posent :

- qui va indexer : des spécialistes de l'indexation ou des automates d'indexation ?
- selon quelle organisation : humaine, automatique, à la volée, avec quel contrôle qualité ? avec quel degré de réactivité ?
- avec quel langage partageable : un thesaurus local « maison », un vocabulaire plus universel mais nécessitant un important travail d'appropriation ?

Si les services TICE sont en attente des solutions que peuvent leur apporter les bibliothécaires dans le domaine de l'indexation des ressources pédagogiques numériques, le pragmatisme doit être de rigueur afin de répondre le mieux possible aux besoins des usagers, tout en n'obérant ni la réactivité ni la qualité.

Questions de l'assistance

- Qui cherche et qui trouve des RPN ? : les concepteurs de cours ? les apprenants de l'université ? d'autres apprenants ?
- Comme il y a plusieurs utilisateurs et plusieurs utilisations, n'y a-t-il pas plusieurs indexations ? À l'évidence, l'indexation doit être construite en fonction des destinations finales et probablement par des indexeurs distincts.

Concevoir des systèmes de description de ressources dans un processus de normalisation en cours d'avancement

État de l'art en matière de normes et approche critique

Elizabeth CHERHAL, de la Cellule MathDoc de Grenoble, apporte des éléments de contexte historique sur l'avènement des formats MARC et Dublin Core.

MARC, utilisé depuis 1965 pour indexer les documents des bibliothèques, existe aujourd'hui sous deux formats, après avoir connu une époque où chaque pays utilisait son propre format. Ce format ouvert permet des échanges industriels de données pour constituer des catalogues collectifs ; les migrations logicielles sont facilitées par la normalisation des données. Le contrepoint en est la lourdeur du format qui exige des compétences pointues pour la création des données (pas loin de 1000 éléments différents). Enfin ce format est inadapté pour décrire de grandes collections numériques.

DC (Dublin Core), qui a vu le jour en 1995 sous l'impulsion de bibliothécaires et d'informaticiens, a été conçu pour décrire simplement des ressources en ligne. Le revers de la simplicité est qu'il ne permet pas de tout décrire avec précision et qu'il nécessite donc des extensions qui le complexifient. Son succès est maintenant garanti : norme ISO, il est aussi reconnu par le protocole OAI-PMH de moissonnage (partage) de métadonnées.

MARC et DC cohabitent et savent se parler, c'est à dire échanger des données.

Jean-Philippe PERNIN, de l'INRP, se concentre sur l'évolution des représentations du rôle de l'enseignant :

- enseignant artisan qui est l'auteur du cours,
- enseignant éditeur qui conçoit et prescrit des ressources,
- enseignant documentaliste qui prospecte et référence
- et enfin, enseignant pédagogue centré sur l'activité pédagogique.

Ces divers rôles ont donné lieu à divers modèles :

- LOM axe tout sur la ressource et s'appuie sur un modèle de granularité des RPN ;
- SCORM lie les ressources aux activités, organise la traçabilité du parcours de l'apprenant et permet l'interopérabilité des RPN entre plates-formes ;
- IMS-Learning Design met l'accent sur les scénarios pédagogiques et non pas sur les RPN ; en conséquence il préconise la création de bibliothèques de ressources (numériques ou non, pédagogiques ou non) et la création de bibliothèques de scénarios pédagogiques. Cette nouvelle approche nécessite de s'intéresser aux

usages effectifs des ressources d'apprentissage, au sein des communautés, tant en termes de mutualisation, de réutilisation que de stratégies d'indexation.

Les normes et standards de description des ressources

Elizabeth CHERHAL, de la cellule MathDoc détaille le format Dublin Core, constitué de 15 éléments de base relatifs au contenu, à la propriété intellectuelle et à des instances techniques (date, type, identifiant...). Ces éléments peuvent être raffinés avec des « qualifieurs » qui sont soit des raffinements d'éléments, soit des schémas d'encodage (vocabulaires contrôlés). Il existe ainsi un raffinement « Educational Level ». Pour la compatibilité OAI-PMH seuls les 15 éléments de base sont concernés.

Rosa Maria GOMEZ DE REGIL, de Doc'INSA, présente le LOM (Learning Object Metadata), standard IEEE, détaille les 9 catégories d'éléments (General, LifeCycle, MetaMetadata, Technical, Educationnal, Rights, Relation, Annotation, Classification) et montre leurs parentés et leurs spécificités avec les formats MARC et DC.

De nombreux projets universitaires en Rhône-Alpes, en France et dans le monde utilisent la structuration LOM des métadonnées, mais souvent dans des profils d'application qui diffèrent. Dans le cadre du groupe 4 de la CN36, l'AFNOR élabore un profil d'application français en se basant sur le profil ManUel. La mise en œuvre de l'indexation est abordée avec le workflow d'indexation de l'INSA de Lyon, faisant intervenir à la fois des enseignants/auteurs/concepteurs, des développeurs et des documentalistes dans la chaîne de l'indexation. De nouveaux métiers se dessinent pour les bibliothèques, nécessitant beaucoup de relationnel pour coordonner l'ensemble du dispositif qui permet de constituer une bibliothèque de ressources numériques et d'assumer la gestion des droits associés aux ressources.

Gilles BERTIN, projet Ancoly du PUL, présente SCORM, standard IEEE, qui se propose de rendre les RPN interopérables sur diverses plates-formes pédagogiques (autrement nommées LMS : Learning Management System), tout en offrant un environnement d'exécution centré sur les ressources et sur les activités pédagogiques. Pour cela SCORM définit :

- un modèle d'agrégation de contenus qui permet de créer, de décrire et d'échanger des ressources indépendamment du système qui les a créées ; l'agrégation se définissant en trois niveaux : ressources, grain de formation, cours,
- un environnement d'exécution qui permet une communication avec les plates formes pour l'exécution de l'enseignement.

Il existe des outils logiciels permettant de « scormer » rapidement des RPN. SCORM complète LOM – qui continue d’être utilisé pour décrire le niveau d’agrégation de la ressource – et est bien supporté par les plates-formes, avec une bonne interopérabilité.

Les évolutions du format LOM

Yolaine BOURDA, SUPELEC, animatrice du groupe 4 CN36, présente les travaux d’évolution des normes de métadonnées pour les RPN. LOM, standard IEEE a été proposé en norme ISO en procédure accélérée avec appel à commentaires en 2002, mais l’IEEE ayant refusé de discuter une évolution du LOM en fonction des commentaires (dont ceux de la France), la norme ISO a été refusée. LOM est rarement utilisé tel que le définit le standard IEEE de 2002 ; de nombreux profils d’application viennent raffiner les définitions, ajouter des éléments, spécifier des valeurs, rendre obligatoires des éléments facultatifs – ce qui rend extrêmement difficile la préservation de l’interopérabilité avec le schéma de base. Plusieurs exemples d’interprétations divergentes du sens des éléments et de listes de valeurs non compatibles, sont montrés.

Il y a donc un réel besoin de revenir à une réflexion fondamentale sur la ressource et sa description : ceci fait l’objet du projet MLR : Metadata for Learning Resources de l’ISO. La France est co-éditrice du MLR avec la Chine. Deux niveaux de réflexion prévalent : conceptuel/abstrait et réalisation/implémentation. La norme de conception ISO/IEC 11179 est utilisée comme méthode de conception, afin d’éviter des confusions de sens dans l’élaboration du MLR, confusions qui n’ont pas été évitées dans LOM. Ce travail est de longue haleine, sur plusieurs mois.

En attendant le groupe 4 CN36 travaille sur un profil d’application français du LOM, dans le but d’élaborer une norme AFNOR expérimentale. Le profil d’application ManUel de CampuSciences sert de base à la réflexion. Un premier appel à commentaires est dépouillé en novembre et décembre 2004, pour une enquête élargie en janvier 2005 et une publication expérimentale en mars 2005.

Tous les participants à cette journée sont conviés à se manifester lors de cette consultation : c’est en recueillant les avis de tous que l’on tendra vers un profil d’application adapté au plus grand nombre.

L’indexation des RPN : questions transversales

Ghislaine Chartron, INRP, URSIDOC/enssib, s’interroge :

- Sur les niveaux actuels d’utilisation des standards pour les RPN, et confirme qu’ils sont utilisés :
 1. Identification des documents avec DOI,
 2. Description des documents avec DC norme ISO,

3. Échange de données avec des protocoles OAI-PMH.

- En ce qui concerne les RPN, l'indexation sert à décrire un contexte pédagogique d'usage (niveau, méthode, évaluation, ...) et un contexte technique pour la réutilisation dans la conception (plate-forme, format, liens avec d'autres objets...).
- Sur la création des métadonnées des RPN : elles peuvent se présenter selon 4 niveaux : métadonnées de repérage d'un contenu (notice bibliographique renseignées par un bibliothécaire), métadonnées de conception pédagogique (renseignées par un enseignant), métadonnées sur la conception technique (renseignées par les concepteurs/développeurs) et enfin les métadonnées permettant de gérer les droits (renseignées par les bibliothèques).
- Sur la réalité de la mutualisation des ressources pédagogiques : certaines études montrent que les RPN sont encore peu partagées.
- Sur la réalité de l'utilisation du LOM : d'après l'enquête OCLC, sur 35 projets, les métadonnées les plus renseignées sont celles de DC et les moins renseignées sont celles du bloc pédagogique.

L'indexation des RPN doit donc se concevoir comme une activité à pratiquer de façon raisonnable en tenant compte du facteur temps et du facteur coût, et en fonction des contextes d'usages.

Questions de l'assistance

- Quelle organisation mettre en place pour une indexation en « strates » avec plusieurs acteurs ? quelle dimension humaine ?
 - oui, il est important d'associer plusieurs acteurs à cette création,
 - oui il est important de passer de pratiques individuelles à des pratiques coopérantes,
- La convergence des divers formats présentés est frappante ; vont-ils tous converger ?
 - ces formats ne sont pas des sur-ensembles : ils doivent co-exister,
 - il semble que les spécificités du LOM soient peu utilisées,
 - LOM et DC Educational Level ont signé pour une convergence commune.

Observer les partenariats entre SCD et TICE : rôle et positionnement des acteurs

Cas de l'UTC : Annie BERTRAND, SCD, et Xavier HENNEQUIN, SCENARI

L'UTC vient d'ouvrir son portail documentaire, conçu comme un point d'entrée unique sur les ressources documentaires et pédagogiques, réalisation portée par la stratégie de l'UTC en matière de diffusion des ressources TICE. Le portail est enrichi avec des technologies de workflow permettant de matérialiser cette relation entre l'enseignant et la bibliothèque au travers des fonctions suivantes : déposer des documents, proposer des sujets et des mots clés, gérer les autorisations de consultation pour des groupes et des durées ; le SCD ayant la charge de finaliser la description et de la publier.

SCENARI est une chaîne éditoriale qui produit et gère des contenus ainsi que des métadonnées propres (pédagogiques, documentaires, d'indexation sujet, de formation). Les contenus et leurs métadonnées sont exportés pour être publiés sur le portail documentaire.

Il est prévu un renforcement du partenariat.

Cas de l'université de Lille 1 , avec Jean-Bernard MARINO, SCD

[NDLR : C'est avec une immense douleur que nous avons appris le décès de Jean-Bernard Marino quelques jours seulement après son intervention au sein de cette journée d'études.]

La bibliothèque numérique GRISEMINE propose 1100 documents issus de la production pédagogique et scientifique de l'université. Le retour d'expérience, de plusieurs années, est très encourageant sur la visibilité des ressources numériques. La description est actuellement basée sur la DTD Marc. Une évolution technique de l'application est souhaitée pour mieux répondre aux utilisateurs : recherche type Google, métadonnées DC, compatibilité OAI-PMH, diversification des contenus (évolution vers une politique d'archives ouvertes pour l'université), signalement de l'unité documentaire la plus fine, validation des contenus pas les responsables de communautés. Le SCD a choisi d'implanter DSpace du MIT qui sera décliné en deux parties : Ariane pour la recherche et Iris pour le SCD. Ce sont 48 éléments de métadonnées DC qui sont retenus. La Dewey a été implémentée.

Questions de l'assistance

- Les bibliothécaires deviennent-ils des documentalistes ?
 - oui, en utilisant des outils spécialisés comme Rameau, et d'autres Thésaurus,
 - en quoi l'utilisation de Rameau transforme-t-il des bibliothécaires en documentalistes ?
- Les documents sont-ils validés ? et si oui, se trouve-t-on dans une réappropriation du rôle éditorial de l'université ?
 - oui, c'est bien le cas, l'université se positionne comme un éditeur.

Identifier les enjeux du partenariat SCD-TICE – Point de vue des tutelles et des réseaux nationaux

Table-ronde animée par Jean-Michel SALAUN, ENSSIB avec :

- Sabine BARRAL, ABES,
- Gilles BRAUN, Direction de la Technologie, sous-direction des TICE,
- Michel ROLAND, Couperin,
- Jean-Émile TOSELLO-BANCAL, Direction de l'Enseignement Supérieur, Sous Direction des Bibliothèques.

Le débat s'articule autour de 3 thèmes : Qui doit indexer ? Quelle politique centralisée/distribuée ? Qui prend la responsabilité de cette édition ?

Thème 1 : Qui doit indexer ? avec quels moyens ? avec quelle motivation ?

Michel ROLAND, souhaite qu'un tel point soit discuté au sein des comités de pilotage des systèmes d'information des établissements, car il faut éviter que ce travail d'indexation des RPN ne devienne une tâche supplémentaire sans moyens supplémentaires ; il convient donc qu'une prise de conscience globale habite une telle décision.

Jean-Émile TOSELLO-BANCAL rappelle que les bibliothécaires sont déjà naturellement impliqués dans la pédagogie (glossaires, bibliographies...) et qu'ils doivent être indexeurs, mais en association avec les enseignants. Tout ceci doit être réalisé dans un schéma de système d'information d'établissement comportant des représentants des CRI, TICE et SCD.

Sabine BARRAL, nous rappelle, qu'il y a 8 ans, il n'y avait pas de bibliothécaires dans les colloques sur les TICE et que cela a changé. Pour le reste chacun fait son métier et si on veut mutualiser, la normalisation est très importante.

Gilles BRAUN note que l'indexation de la ressource est partie intégrante de la ressource et qu'elle en est même à l'origine ; il faut donc intégrer des bibliothécaires dès l'élaboration de la ressource. Des exemples très concrets comme ceux de la numérisation des ressources vidéos de l'INA sont là pour en témoigner.

Jean-Michel SALAUN prend acte de ce consensus commun et convient que les pouvoirs publics doivent prendre la mesure de l'extension des métiers des SCD à l'indexation des RPN, avec l'affectation de moyens qu'il convient de prévoir.

Thème 2 : Quelle politique centralisée/distribuée privilégier pour donner un moteur à toutes ces actions ?

Sabine BARRAL indique que la production des RPN et leur description se fait le plus souvent au niveau local, ce qui n'exclut pas la logique d'un partage dans des réseaux nationaux. Les réseaux nationaux pouvant eux produire et mettre à disposition des fondamentaux utilisables par tous.

Jean-Emile TOSELLO-BANCAL ne pense pas que le paradigme des grands réseaux bibliographiques de bibliothèques puisse être transposé aux RPN, car la production des RPN est plutôt locale et qu'elle doit garder un fort potentiel d'évolutivité, au gré des mises à jour des RPN et des changements de programmes.

Gilles BRAUN plaide fortement pour que les producteurs de RPN se conforment à la normalisation, ceci prenant la forme de mesures financières incitatives pour soutenir la production de RPN ; dans ce cas, l'intervention incitative est nationale.

Michel ROLAND considère que le modèle des réseaux bibliographiques ne peut être totalement transposé, car la RPN est un objet mouvant, avec des statuts de validation très divers. Le niveau d'intervention est donc plutôt local dans une logique de partenariat fort entre les services TICE et les SCD, qui seront à même de spécifier les meilleures méthodes de travail.

Thème 3 : Qui prend la responsabilité de la publication ? Qui gère les droits ?

Sabine BARRAL considère deux cas : validation locale du contenu par l'auteur enseignant et/ou l'auteur personne morale (un établissement), validation plus centralisée (nationale, régionale...) en cas de construction d'archives pérennes.

Gilles BRAUN considère qu'il faut évaluer les ressources publiées sur le web, et l'établissement doit donc prendre ses responsabilités d'éditeur pour valider les RPN et la description LOM des RPN.

Jean-Émile TOSELLO-BANCAL considère que la validation scientifique est plurielle et à faire par les acteurs impliqués dans la production des RPN. La validation pourrait aussi être

complétée en dehors de l'établissement pas un comité scientifique, à l'instar de ce qui se passe dans les comités scientifiques de numérisation.

Questions de l'assistance

- Les comités de pilotage des systèmes d'information sont-ils très présents dans les établissements ?
 - oui, l'on commence à en rencontrer et la CPU pousse fortement les universités dans ce sens.
 - oui, dans le cadre des contrats quadriennaux, certains projets d'établissements les signalent.
- Un témoignage sur l'évaluation des ressources pédagogiques : attention à la question de l'évaluation car évaluer la ressource conduit à évaluer indirectement la qualité des enseignements et c'est un sujet très sensible.
- Quelle est la synergie entre les portails du CERIMES et de l'ABES ?
 - Le CERIMES met en place des systèmes de distribution de contenus spécialisés comme des vidéos à destination des UNT.
 - L'ABES met en place un système de fédération de portail, donc un point d'entrée sur des ressources.

Synthèse

Monique JOLY souligne la qualité des interventions et des débats montrant :

- qu'il reste encore un important travail de spécifications pour concevoir des méthodes de travail homogènes, afin de rendre parfaitement interopérables les systèmes de partage de RPN,
- mais que cela ne doit pas retarder l'engagement des actions coopératives TICE-SCD car l'on peut considérer cette étape de finalisation comme un ensemble de détails de mise au point qui seront rapidement résolus,
- parce que les acteurs concernés (TICE, SCD, DT, DES/SDB) se réunissent dans une dynamique récente mais très active, sous la pression des utilisateurs de ces futurs systèmes.

Ce sont eux qui auront le dernier mot en jugeant la qualité, la fiabilité et la pérennité des systèmes que nous aurons su mettre en place à leur intention.