

Le projet EuDML, une bibliothèque numérique européenne de mathématiques

Thierry Bouche

▶ **To cite this version:**

Thierry Bouche. Le projet EuDML, une bibliothèque numérique européenne de mathématiques. Armelle Jézéquel. L'IST au prisme de l'Europe, Centre Emile Durkheim, pp.47-56, 2013, 978-2-7466-6304-6. hal-00815683

HAL Id: hal-00815683
https://hal.science/hal-00815683

Submitted on 19 Apr 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.

Le projet EuDML, une bibliothèque numérique européenne de mathématiques

Thierry Bouche

Université de Grenoble I, Institut Fourrier (UMR 5582) & Cellule Mathdoc (UMS 5638) BP 74, 38402 St-Martin-d'Hères cedex, France
thierry.bouche@ujf-grenoble.fr http://www-fourier.ujf-grenoble.fr/~bouche/

1. Les enjeux spécifiques de la documentation en mathématiques

Les résultats mathématiques constituent l'outil primaire de nombreuses disciplines scientifiques. De fait, la formalisation et le savoir mathématiques jouent un rôle fondamental dans presque tous les domaines scientifiques et technologiques. Une remarque importante est qu'il n'y a pas de corrélation évidente entre la date à laquelle un résultat mathématique est publié et celle où il trouvera une application. En effet une même théorie peut d'ailleurs trouver des applications dans des champs et à des époques très éloignés les uns des autres. Comme le dit Jean-Pierre Serre : « Les mathématiciens se contentent de mettre leur production à la disposition de tous, comme sur des étagères où l'on peut venir se servir. » [1]

On constate donc que la bibliothèque est non seulement la principale production et l'outil de travail quotidien du mathématicien, mais aussi une ressource essentielle pour les autres sciences. Les articles anciens jouent pour cette discipline scientifique le rôle des données de la recherche, ailleurs : ce sont les briques à partir desquelles la recherche s'élabore. Aussi, les mathématiciens sont les héritiers d'une longue tradition d'édition de qualité. La relecture par les pairs avant la publication d'un article peut durer plusieurs années, si cela s'avère nécessaire pour valider les arguments. Ainsi, l'ensemble des résultats publiés forme un corpus relativement fiable et de dimensions réduites. Un résultat nouveau ne remplace pas les résultats valides précédemment connus, mais au contraire s'appuie sur eux en se contentant d'expliciter les références nécessaires.

La question de la conservation et de l'accès sur des durées très longues se pose de façon aigüe pour le corpus mathématique.

L'équipement principal des laboratoires de maths à travers le monde est une bibliothèque spécialisée. On peut considérer que le réseau mondial des bibliothèques universitaires et spécialisées formait une bibliothèque (virtuelle) de

[1] Michel Broué, « Jean-Pierre Serre et le métier de mathématicien », *Images des mathématiques*, CNRS, 2006. <http://images.math.cnrs.fr/Jean-Pierre-Serre-et-le-metier-de.html>, consulté le 2 juin 2012.

référence à l'époque où le papier était l'unique support de diffusion et d'archivage. Depuis une vingtaine d'années, la localisation d'un document à l'aide des catalogues fusionnés et la pratique du prêt entre bibliothèques (PEB), voire les services de fourniture de documents, permettent au chercheur de satisfaire la grande majorité de ses besoins documentaires, l'identification des besoins étant elle-même considérablement facilitée par la conversion en bases de données des traditionnelles revues recensant la production mathématique mondiale, selon une classification par sujet *ad hoc* (MSC, *mathematical subject classification*). Pour l'essentiel, les questions cruciales de la conservation, de la fouille, et de l'accès étaient résolues.

Il n'en est pas de même dans le paradigme numérique actuel. La conversion rapide de l'édition classique au numérique, au moins pour les revues, est *a priori* un atout formidable pour la facilité et la rapidité d'accès aux textes. L'indexation plein texte et la navigation hypertexte ont démultiplié la visibilité et l'accessibilité des collections. Cependant, un certain nombre de questions fondamentales n'ont pas obtenu de réponses satisfaisantes.

Conservation. Les bibliothèques ne possèdent plus les collections physiques et n'assurent donc plus cette fonction. Si elles reçoivent un exemplaire papier, il est de plus en plus souvent issu d'une impression laser impropre à une préservation à long terme. Dans le cas d'une faillite ou du rachat d'un éditeur, il n'y a aucune garantie que les collections resteront disponibles.

Accès. La propriété de l'exemplaire papier mis à la disposition des lecteurs, y compris hors les murs grâce au PEB, n'a pas d'analogue dans le monde numérique, où les éditeurs conservent l'intégralité de leur production sur leur serveur, et contrôlent tous les accès. Les quelques systèmes d'archives plus ou moins noires mis en place soit au niveau national, soit par des sociétés comme Portico, soumettent toute communication d'un article au contrôle de son éditeur sur des périodes d'embargo extrêmement longues. Pour la première fois de l'histoire, un chercheur peut se trouver dans la situation où il n'a pas de moyen légal de se procurer un texte nécessaire à son travail. L'auteur qui se réjouissait de communiquer aux quatre coins du monde ses tirés à part a généralement signé un contrat qui lui interdit d'expédier le PDF de l'éditeur par courrier électronique.

Je ne traite pas ici des aspects économiques et contractuels qui accompagnent ce passage au tout électronique et à la commercialisation des revues par bouquets, lesquels aboutissent à vider de leur sens d'autres fonctions essentielles d'une bibliothèque (notamment la capacité de sélectionner les acquisitions sur des critères de qualité et de pertinence).

2. Une bibliothèque numérique mondiale et exhaustive ?

Spectateurs de ces changements qui ont un impact considérable sur leur environnement de travail quotidien, mais aussi acteurs (notamment à travers les sociétés savantes qui sont impliquées dans l'édition mais aussi dans la production de livres, de revues, mais aussi de bases de données recensant leur production : MathSciNet est un service de l'*American Mathematical Society*, ZBMath est supervisé par la Société mathématique européenne [2]), les mathématiciens de la fin du XX[e] siècle se sont mis à rêver d'une bibliothèque

[2] MathSciNet, *Mathematical Reviews* on the Web : <http://www.ams.org/mathscinet/> ; ZBMATH, the *Zentralblatt MATH* database : <http://www.zentralblatt-math.org/zbmath/en/>, consultés le 2 juin 2012.

numérique mondiale et exhaustive. Cette bibliothèque numérique (dite DML : *Digital mathematical library*[3]), conçue à l'origine, et de façon un peu naïve comme la transposition électronique de la bibliothèque que chaque mathématicien fréquente dans son laboratoire, comporterait dans ses rayonnages virtuels la totalité de ce qui s'est publié depuis les *Éléments* d'Euclide. Elle serait étroitement liée avec les bases de données, et surtout aussi navigable et gratuite que possible.

L'exhaustivité n'est pas un rêve : c'est seulement une question d'argent et de volonté. La totalité du corpus ne couvre pas plus de 100 millions de pages, selon des estimations crédibles : c'est moins que ce qui s'est publié dans le seul domaine biomédical depuis l'an 2000, cela représente une masse d'information modeste qui tiendrait sur un seul disque dur !

La « navigabilité » est une des motivations fortes pour la communauté, du fait du caractère cumulatif de la littérature mathématique. Quelle facilité si la machine fournit le lien *ad hoc* chaque fois que l'on souhaite suivre une référence ! À comparer avec le nombre de sites qu'il faut connaître, de stratégies différentes qu'il faut y développer pour trouver quelque chose aujourd'hui.

Quant à la gratuité, elle intéresse bien peu les acteurs économiques. Il y aurait pourtant de grands bénéfices à en tirer, notamment parce qu'il serait bon que cette ressource basique soit disponible bien au-delà de l'environnement professionnel du chercheur en mathématiques. Une autre raison, plus pragmatique, est que la navigation est infructueuse si tous les liens pointent sur des portes closes. Or, du fait de la désorganisation des accès électroniques, aucun laboratoire dans le monde ne possède toutes les licences d'accès nécessaires.

En tant que projet *global*, la DML a peu avancé au cours des dix dernières années, malgré le soutien de l'Union mathématique internationale (IMU), de la Société mathématique européenne (EMS) et la tenue de nombreuses rencontres internationales sur le sujet.

Dans l'intervalle, si nous sommes toujours très loin d'avoir satisfait les ambitions du projet DML, le contenu mathématique existant au format numérique est devenu tout à fait considérable : il représente plus du tiers de la totalité du corpus, tous supports confondus. Le tiers des articles référencés dans les bases de données dispose d'un lien vers son texte intégral, bien que de très nombreuses ressources leur échappent, en général parce qu'elles sont produites et gérées par des non-professionnels. C'est au niveau des barrières d'accès que la situation est la plus décevante : conçue comme un service public dans un monde qui perd cette notion, la DML peine à trouver un modèle économique viable.

L'IMU a proposé depuis longtemps que les articles deviennent librement accessibles à l'issue d'un certain délai d'embargo pendant lequel l'éditeur récupère ses coûts de production (valeur généralement suggérée : 5 ans). Si cette recommandation a été entendue de nombreux acteurs académiques, elle n'a pas ébranlé les éditeurs commerciaux. Conscients de l'intérêt que les mathématiciens portent à leur patrimoine scientifique, les éditeurs commerciaux ont entrepris de numériser leur propre fonds qu'ils vendent désormais comme un

[3] Voir <http://www.mathunion.org/WDML/>, consulté le 2 juin 2012.

nouveau produit. Les projets américains sans but lucratifs, mais sans financement public récurrent, comme JSTOR ou Project Euclid[4], assurent leur pérennité en commercialisant l'accès à leurs services par les bibliothèques à un tarif qui peut nous paraître prohibitif. Au final, la quantité des projets apportant une partie des services attendus de la DML s'est considérablement accrue, et avec elle le nombre de conditions et de barrières d'accès !

Je décris dans ce qui suit deux projets de nature assez différente dans lesquels j'ai été impliqué au sein de la cellule Mathdoc : le projet de numérisation NUMDAM et son extension par des chaînes d'acquisition nativement numérique ; le projet EuDML qui a pour objet de porter au niveau européen une vision à la fois ambitieuse et pragmatique de la DML.

Je signale que l'histoire ne s'arrête pas là et qu'une rencontre organisée par l'IMU a eu lieu à Washington en juin 2012 pour relancer l'effort mondial.

3. Le programme NUMDAM

NUMDAM a été conçu comme un projet de numérisation de revues françaises de mathématiques (d'où le développement officiel de cet acronyme hybride en « Numérisation de documents anciens mathématiques ») dans une logique de soutien à la visibilité et à la diffusion la plus large possible des archives des revues académiques nationales. Le financement initial, de l'ordre de 450 000 €, a été débloqué par le ministère de la Recherche à une époque un peu particulière (la fameuse « cagnotte de Jospin »). Si les motivations étaient assez claires et convaincantes pour emporter cette décision, les modalités de mise en œuvre n'étaient pas finalisées du tout : les précédents directeurs de Mathdoc s'étaient inspirés du peu d'informations disponibles sur les rares projets de numérisation déjà actifs (JSTOR aux États-Unis, ERAM en Allemagne, Gallica en France, le projet européen DIEPER…) et avaient pour principal objectif de viser la meilleure qualité de numérisation. En ce qui concerne l'exploitation des données, rien n'était précisé : la diffusion des archives numérisées pourrait avoir lieu sur le site de chaque revue, sur la plateforme d'un sous-traitant, l'accès pourrait se faire moyennant une redevance modeste pour compenser ces coûts…

Les travaux préparatoires avaient commencé avant que l'argent ne soit versé, courant 2000, ils s'achevèrent par la rédaction d'un cahier des charges technique pour un appel d'offre de numérisation publié en juin 2001. Dans l'intervalle, des accords avaient été passés avec les responsables des 6 séries concernées par la première phase. Un recensement exhaustif du contenu et des particularités physiques des 6 collections avait été réalisé par des vacataires. Deux postes d'ingénieurs avaient été affectés par le CNRS, l'un sur un emploi permanent, l'autre en CDD, pérennisé un an plus tard.

La numérisation proprement dite débute en janvier 2002, chez le sous-traitant lauréat du marché. Une chaîne de contrôle qualité (manuel sur échantillons, et exhaustif pour tout ce qui pouvait être automatisé) est développée en interne ; il faut encore attendre six mois pour que la qualité du prestataire corresponde aux standards définis par le cahier des charges, et que la production prenne enfin sa vitesse de croisière. En parallèle, nous commençons à étudier les options pour la mise en ligne des articles et nous contactons les auteurs pour obtenir leur

[4] Voir <http://www.jstor.org/> et <http://projecteuclid.org/>, consultés le 2 juin 2012.

autorisation — l'un des éditeurs profitera de cette occasion pour faire signer des contrats d'édition à tous ses auteurs. Nous mettons en place des post-traitements pour les fichiers utilisateurs : ajout d'une page de garde, linéarisation et protection des PDF contre les modifications ; ajout de liens vers les bases de données pour les articles et ceux qu'ils citent en bibliographie… Un ingénieur de Mathdoc adapte un logiciel qu'il a développé dans un autre contexte (interface de recherche et de consultation du ZBMath) de façon à ce que nous disposions d'une maquette simple d'accès aux collections et commencions à réfléchir aux modalités de la mise en ligne. Cette maquette s'avère tellement fonctionnelle que nous décidons de l'adopter après quelques ajustements. Les deux premières revues apparaissent en ligne le 20 décembre 2002, les six revues de la première phase sont en ligne un an plus tard, soit 7 000 articles couvrant 200 000 pages dont la quasi-totalité sont en libre accès.

NUMDAM, en tant que projet pilote dont la première ambition était de numériser l'intégralité des archives de cinq ou six revues académiques de mathématiques publiées en France des origines jusqu'à l'an 2000, est alors dans une situation intéressante : moins de la moitié de l'argent alloué à cette opération a été consommée bien qu'elle ait duré 6 mois de plus que prévu. Les objectifs un peu vagues formulés lors du financement du projet sont largement atteints, mais le produit final ne ressemble pas du tout à ce qui avait pu être imaginé à l'époque, aussi bien sur le plan du modèle d'affaires (les articles sont en libre accès sur une plateforme entièrement gérée par la cellule Mathdoc) que de l'exploitation des données capturées. En fait, ce service sera le modèle explicite pour les projets qui naîtront en Europe à ce moment-là (DML-E en Espagne, DML-CZ en République Tchèque, BDIM en Italie), et ailleurs (*Canadian Journal of Mathematics*…).

Grâce au soutien constant de nos tutelles, la disponibilité du budget alloué en l'an 2000 ne sera pas remise en question pendant un peu plus de six ans : un deuxième marché pourra ainsi être financé et tripler la taille des collections. La troisième phase du programme NUMDAM s'est achevée en 2011, portant le nombre de séries numérisées à 66, dont 2 séries de mémoires et une collection de thèses. La couverture disciplinaire est désormais très large : de l'informatique théorique à la théorie des catégories, de la physique aux statistiques, de l'arithmétique aux sciences humaines. Des revues italiennes et hollandaises nous ont rejoints. Les collections débutent avec le premier article des *Annales de Gergonne*, première revue exclusivement consacrée aux mathématiques publiée dès 1810 à Nîmes, et comportent plus de 600 articles publiés depuis 2010, sur un total de 52 000.

Au-delà du seul programme de numérisation, le serveur NUMDAM abrite en effet la production des éditeurs qui veulent bien la lui confier. Cela concerne une vingtaine de titres publiés par une dizaine de structures, ce qui représente de l'ordre de 10 % des contenus diffusés. Nous avons à cette fin développé des chaînes d'acquisition numérique qui nous permettent de convertir les métadonnées des éditeurs à notre format interne.

Au final, le projet pilote de l'an 2000, prévu pour une durée de 3 ans, s'est discrètement transmué en un service pérenne de bibliothèque de référence archivant *a posteriori* une partie substantielle du corpus mathématique produit en France et alentour. Le coût consolidé sur ses dix premières années avoisine le million d'euros, soit à peu près un euro par page numérisée.

4. Le projet EuDML

4.1 Présentation

Depuis l'année 2003, la SME n'a cessé d'appuyer la constitution d'une

bibliothèque numérique de mathématiques, volet européen de la DML. Elle a à cette fin présenté plusieurs propositions à l'occasion de divers appels à projets (5e PCRD, eContent+, ESF, 7e PCRD…). C'est pour finir en 2009 que la première proposition fut retenue, dans le cadre du programme cadre pour la compétitivité et l'innovation, volet d'appui stratégique en matière de technologies de l'information et de la communication, thème : « bibliothèques numériques », objectif : « libre accès à l'information scientifique ».

Voici la description des projets attendus dans le cadre de cet appel[5] :

> « Des actions pilotes B seront financées pour mener des expériences concluantes de libre accès à des bibliothèques numériques de contenu scientifique/savant, y compris des expériences explorant de nouveaux paradigmes pour l'évaluation par les pairs, l'affichage, l'interrogation et la constitution de liens pour des contenus scientifiques/savants et (optionnellement) les ensembles de données sous-jacentes associés. »

Et voici l'impact espéré : « Accès libre à plus de contenu scientifique/savant et/ou le développement de nouvelles façons d'évaluer, afficher, interroger et lier du contenu scientifique/savant. »

Pour répondre à cet appel et à « l'amicale » pression de la SME, j'ai formé un consortium composé de 14 partenaires européens[6]. Une dizaine d'entre eux sont des institutions publiques qui apportent les principales collections de mathématiques numériques (en majorité numérisées) en Europe[7]. Un seul éditeur commercial (EDP Sciences) fait partie du consortium. Il participe avec cinq journaux français dont la série ESAIM, éditée sous les auspices de la Société de mathématiques appliquées et industrielles (SMAI) et les descendants de la revue *RAIRO* qui ont été numérisés par NUMDAM. De nombreux autres éditeurs sont associés au projet à travers leurs archives numérisées par un partenaire ou plus directement parce qu'ils utilisent pour leur édition électronique l'une des plates-formes partenaires du projet[8]. L'ensemble des partenaires représente une diversité notable de compétences techniques, depuis les bibliothèques numériques et les services d'édition électronique jusqu'au traitement automatisé du savoir mathématique.

Le consortium est renforcé par l'apport de ses deux partenaires associés, qui ne perçoivent aucun financement européen : la Société mathématique européenne (SME) est associée au projet comme autorité morale fixant les objectifs et évaluant l'utilité des résultats du projet. Elle préside un comité

---

5 *Competitiveness and Innovation Framework Programme* (CIP), *ICT Policy Support Programme*, *ICT PSP Work Programme 2009*, appel publié le 29 janvier 2009. [La traduction est de moi.]

6 La liste détaillée se trouve sur le site Web du projet : <http://www.eudml.org/>. La vision institutionnelle du projet par la Commission européenne est en ligne à l'adresse <http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=250503>, consulté le 2 juin 2012.

7 Projets « nationaux » déjà opérationnels : les tchèque DML-CZ, français NUMDAM, espagnol DML-E, polonais DML-PL, portugais PtDML, grec HDML, et des projets émergents comme le bulgare BulDML.

8 Ce qui signifie qu'environ 90 revues fournissent directement leurs articles récents : elles sont diffusées par EDP Sciences, le projet CEDRAM de la cellule Mathdoc, ou le service ELibM du FIZ Karlsruhe. Leurs éditeurs sont pour la plupart des petites structures, allant d'une équipe de volontaires à des sociétés savantes en passant par des structures académiques ou de petites sociétés privées. <http://publications.edpsciences.org>, <http://www.cedram.org/>, <http://www.emis.de/journals/>, consultés le 2 juin 2012.

scientifique consultatif. La bibliothèque universitaire de Göttingen apporte les journaux numérisés par les projets ERAM et RusDML, ainsi que la plus grande collection de livres numérisés de mathématiques.

Nous avons rédigé un argumentaire (insistant notamment sur le fait qu'en mathématiques le libre accès *à terme* est un objectif pertinent puisqu'il est primordial de disposer d'un corpus de référence validé et accessible sans barrières) et un plan de travail sur 3 ans découpé en 11 *work packages*, jalonné d'objectifs précis. La coordination du projet a été dévolue à un collègue portugais, seul membre de notre consortium doté de compétences managériales.

Le projet a été accepté sur la base d'une durée de trois ans (à partir du 1er février 2010), d'un volume de travail global de 487 personnes-mois, pour un financement maximal de 1,6 M€.

4.2 Objectifs

Par référence aux objectifs de la DML, le projet EuDML met l'accent sur quelques-uns des chantiers, espérant bénéficier d'une double masse critique : l'importance des contributions européennes en mathématiques, particulièrement au cours de la période très représentée des XIXe et XXe siècles, et la quantité de textes fondamentaux rendus disponibles numériquement par l'un de ses partenaires.

Un portail global. Le premier aspect, degré zéro de l'intégration, porte sur la mise en place d'un point d'accès unifié à l'ensemble des collections à travers un portail Web personnalisable permettant de feuilleter, fouiller, et naviguer dans les collections. Ce système central dispose de services pour transformer les références à ces articles en liens hypertextes, de façon automatisée.

De la sorte, on espère faciliter la fouille dans le corpus mathématique présent sur le portail, donner plus de visibilité à ce corpus précédemment éclaté, et mettre toutes les références hébergées à un clic de l'utilisateur.

Le défi est ici de définir un format d'échange suffisamment riche pour décrire de façon fine et précise la totalité des contenus (articles de revues, communications dans des séminaires ou des congrès, monographies, œuvres en plusieurs volumes…).

Une archive répartie. Une idée force pour la réalisation de la DML est non plus de partir d'un projet centralisé et pyramidal, mais de mettre en réseau les réservoirs de textes existants, et d'atteindre ainsi une masse critique. Le schéma proposé est que chaque éditeur identifie le partenaire du réseau le plus approprié pour y déposer ses archives (y compris récentes), et décide avec lui des modalités de cet archivage, à la façon d'un dépôt légal volontaire.

C'est ainsi que pourrait se constituer une archive pérenne de référence, indépendante des aléas économiques de l'édition, prise en charge par des institutions scientifiquement fiables et durables.

Cet aspect du projet n'a cependant pas convaincu grand monde au-delà des partenaires pour l'instant. D'autres bibliothèques numériques au profil similaire à celui des partenaires actuels se sont associées en cours de route : citons un projet italien (BDIM) et un russe (math-net.ru).

Libre accès à terme. La modalité de fonctionnement suggérée pour les éditeurs partenaires est la suivante :

1. lorsqu'un éditeur a complété sa production, une copie archivable (métadonnées et textes intégraux) est fournie à l'institution *ad hoc* ;
2. ces contenus sont validés et enregistrés ;
3. les nouveaux textes sont indexés et apparaissent dans les résultats de recherches ;
4. l'accès aux textes intégraux se fait sur le site de l'éditeur sous son contrôle ;
5. à l'issue du créneau mobile, la copie locale devient librement accessible.

De la sorte, les textes intégraux sont archivés par des tiers pérennes, la quantité de textes de référence en libre accès augmente, la mathématique établie, mère de toutes les sciences, est enfin à la portée de tous !

Sources	Nombre de textes
Allemagne ERAM/JFM, GDZ, ELibM	85 000
France Gallica-Math, NUMDAM, CEDRAM, EDPS	57 000
Rép. Tchèque DML-CZ	32 000
Russie RusDML	17 000
Pologne DML-PL	14 000
Grèce HDML	4 000
Espagne DML-E	6 400
Portugal SPM/BNP	2 000
Bulgarie BulDML	600
Italie BDIM	2 000

Innovation. Un aspect particulièrement intéressant et ambitieux de ce projet est d'aller bien au-delà de la simple intégration dans un « guichet unique », et d'essayer pour la première fois de déployer des techniques avancées de gestion des savoirs mathématiques, de telle sorte que les spécificités des textes mathématiques (haute formalisation, emploi de formules…) ne soient pas un handicap, mais un ingrédient essentiel pour la fouille et la navigation dans le corpus.

Pour ce faire, nous privilégions les métadonnées en MathML, que nous cherchons à produire à partir de toutes les sources disponibles : OCR pour les textes scannés (avec le concours du logiciel *Infty Reader*), conversions à partir de sources LaTeX, extraction et reconstruction à partir de PDF nativement numériques.

À partir de là, le projet vise à permettre la recherche de formules, la mise en relation des textes à partir de calculs de similarité de leurs contenus mathématiques. Un pan entier du travail est aussi dédié à rendre le corpus plus accessible aux utilisateurs handicapés par la dyslexie ou des problèmes de vue.

4.3 Résultats

Les collections prises en charge sont actuellement de 225 périodiques et séries, 235 000 textes, 2 600 000 pages.

Outre les articles de revues, le serveur actuel contient tous les types de documents pertinents pour les mathématiques : livres, actes de conférences ou de séminaires, œuvres en plusieurs volumes, œuvres complètes ou choisies de grands mathématiciens…

Au cours de ce projet, qui s'achèvera début 2013, un certain nombre de tâches ont été entreprises avec succès :

1. l'analyse des contenus disponibles et la définition d'un format d'échange. Ce format est basé sur le schéma XML "*Journal Archiving and Interchange Tag Set*" développé initialement pour PubMed Central, adopté par JSTOR ou Portico, et devenu récemment un standard NISO ;
2. une conférence a été organisée à Prague en octobre 2010 afin de discuter avec des partenaires potentiels (Springer, éditions de la *London Mathematical Society*…) de nos choix politiques et de leur capacité à attirer ou faire fuir de nouveaux partenaires, notamment parmi les éditeurs (libre accès à terme – créneau mobile, nouveautés fournies par les éditeurs et indexées rapidement, archivage partagé dans un réseau de bibliothèques numériques de référence). Nous n'avons pas entendu de *non* ferme et définitif, mais pas non plus de *oui* enthousiaste de la part des éditeurs recherchant le profit…
3. la constitution d'une base de données avec 235 000 références soit environ 1/5 du contenu numérique existant (1/3 de l'existant « DML ») ;
4. une série d'outils destinés à améliorer l'indexation, la navigation, l'accès et la visibilité pour ce corpus spécifiquement mathématique ;
5. un prototype désormais fonctionnel est en ligne et opérationnel depuis l'été 2012 sur <http://eudml.org/>. Le contenu disponible en ligne dépasse en quantité les prévisions initiales, quelques collections ayant continué à croître, et un projet extérieur s'étant associé (d'autres sont en cours de rapprochement). La diversité des types de contenus pris en charge (articles, livres, communications dans des conférences) couvre l'essentiel du corpus visé ;
6. la recherche fonctionne assez bien, on peut trier et raffiner les résultats par facettes. On dispose d'une recherche plein texte sur la moitié des textes et d'une recherche expérimentale sur les formules, dont la pertinence reste à évaluer. Un feuilletage est disponible par journal, classification mathématique ;
7. toutes les métadonnées disponibles ont été converties en MathML, si bien que les informations sur un texte sont parfois mieux présentées sur le site <eudml.org> que sur le serveur d'origine ;
8. la création systématique de liens est opérationnelle (liens vers les bases de données MathSciNet et ZBMath depuis les textes EuDML ou leurs références bibliographiques, mais aussi de nouveaux liens internes entre les textes EuDML), certaines fonctions Web 2.0 sont actives ;
9. les interfaces de service qui permettent des appels directs depuis des machines distantes de la base de données EuDML viennent d'être rendues publiques. Ceci ouvre notamment la possibilité de transformer de façon automatique toute référence à un texte en lien vers EuDML depuis toute page Web, donc réalise une des fonctions principales d'une bibliothèque numérique de référence : mettre ses collections à la disposition de ses utilisateurs, en un clic, où qu'ils soient. Il reste à faire adopter l'outil de la façon la plus large.

Au final, le projet EuDML, qui avait affiché de très hautes ambitions aura atteint une partie substantielle de ses objectifs. La cause de la DML a progressé grâce à lui, par la mise en œuvre de la première réalisation transfrontalière. Il reste un assez grand nombre de questions ouvertes et de défis à relever dans la nouvelle période qui s'ouvrira à la fin du financement européen. Il n'est en effet pas simple de concevoir et développer une infrastructure pérenne avec des moyens dépendant d'appels à projets.

Les directions à suivre pour faire évoluer le prototype vers une infrastructure mondiale et complète sont évidentes : il faut

- étendre la couverture en associant des partenaires de tous les continents et en intégrant des contenus de provenances variées ;
- consolider l'architecture technique pour disposer d'un service souple et facilement configurable pouvant s'adapter à de nouvelles collections et intégrer des services évolués.

L'avenir nous dira si nous aurons les moyens de poursuivre les voies qui ont été ouvertes par ce projet.