



HAL
open science

Démarrage du projet EuDML - La bibliothèque numérique européenne de mathématiques

Thierry Bouche

► **To cite this version:**

Thierry Bouche. Démarrage du projet EuDML - La bibliothèque numérique européenne de mathématiques. Gazette des Mathématiciens, 2010, 124, pp.65-74. hal-00559055

HAL Id: hal-00559055

<https://hal.science/hal-00559055>

Submitted on 28 Feb 2022

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.

INFORMATIONS

Démarrage du projet EuDML

La bibliothèque numérique européenne de mathématiques

Thierry Bouche¹

Les mathématiciens et les utilisateurs de résultats mathématiques dépendent de façon critique d'un accès aux textes originaux dont le contenu a été scientifiquement validé. Le corpus constitué par l'ensemble de ces textes interdépendants (articles, monographies, mémoires ou synthèses) forme un réseau complexe, chaque texte fournissant les fondations de travaux ultérieurs, tout en reposant sur un ensemble de textes et de connaissances auquel il se réfère, parfois implicitement.

Les mathématiciens se sont organisés au fil des siècles pour disposer à travers le monde de bibliothèques de référence leur permettant de travailler. Ils se sont également investis très tôt dans les outils leur permettant de découvrir les articles utiles pour leurs recherches (journaux recensant les parutions, classifications par sujet). La littérature imprimée étant en cours de disparition rapide, ils se sont pris à rêver d'une bibliothèque numérique mondiale et exhaustive, dans laquelle tout article un peu ancien serait accessible d'un clic.

Au niveau national, particulièrement en Europe, les programmes de numérisation ont consacré au cours de la dernière décennie des efforts considérables en vue de construire des archives numériques de la littérature mathématique éditée dans le monde entier. Il existe en outre de nombreux projets pluridisciplinaires qui comportent quantités de textes mathématiques importants. Cependant, ce corpus n'est pas encore aussi accessible et utilisable qu'il devrait l'être. Ceci est dû principalement à un manque de coordination parmi des parties prenantes, si bien que les archives existantes ne sont pas interopérables. Les références d'un article dans une archive fournissent rarement le lien vers leur cible dans une archive différente. Les outils de recherche des différents services ont des capacités variables et beaucoup de textes ne disposent pas des métadonnées essentielles pour pouvoir être correctement indexés et faciles à trouver. Un plan d'action pour une conservation concertée du corpus mathématique numérique sur le long terme est également nécessaire.

Contexte

La mathématique, science exacte par excellence, dépend entièrement de sa littérature. La mathématique étant la mère de toutes les sciences, lesquelles ont besoin de bases fiables, les résultats mathématiques publiés doivent être soigneusement vérifiés, et les versions vérifiées doivent être conservées indéfiniment, sans

¹ Institut Fourier & Cellule MathDoc, CNRS/Université de Grenoble I.

modification. Le stockage doit être soigneusement organisé, avec un catalogue propre et détaillé, de sorte qu'il sera possible à tout moment de faire référence sans ambiguïté à chacun de ces textes. Le graphe des références devrait être construit et conservé aussi, de sorte que l'on puisse faire confiance aux nouvelles avancées reposant en partie sur des travaux antérieurs.

Puisque les utilisateurs des résultats mathématiques ne se servent pas uniquement de la production mathématique récente, c'est tout le corpus qui doit être facilement accessible sur de longues périodes.

Les modes évoluent : les critères de sélection de telles archives ne devraient pas être la popularité d'un auteur ou d'un sujet à une époque, mais la conformité à des normes rigoureuses de production et validation. Chaque résultat nouveau doté d'une preuve originale qui a été soigneusement vérifiée par des experts indépendants peut devenir une référence cruciale pour des développements inattendus, et trouver des applications spectaculaires dans d'autres domaines scientifiques ou technologiques.

Ces considérations expliquent pourquoi les mathématiciens ont toujours pris grand soin de leurs bibliothèques, qui sont l'infrastructure centrale de tous les laboratoires de mathématiques dans le monde. La bibliothèque idéale devrait être exhaustive, acquérir les nouvelles publications en temps réel, et être largement ouverte (dans le temps et à tous les visiteurs). Grâce à l'obstination de la communauté mathématique, les bibliothèques (papier) approchant de cette situation idéale ne sont pas rares, et bien distribuées dans les pays développés. Cependant, chaque laboratoire ayant ses sujets de prédilection, et un budget limité, aucune de ces bibliothèques de laboratoire ne conserve la totalité du corpus mathématique. Le prêt interbibliothèque assemble cependant ces bibliothèques dispersées en une ressource globale virtuelle (certes un peu lourde) qui remplit à peu près la fonction attendue. Mais le papier en tant que format d'archivage est en perte de vitesse rapide (il existe de plus en plus de ressources uniquement numériques, mais il faut aussi savoir que le tirage papier d'un original numérique peut avoir une durée de vie extrêmement courte en fonction des technologies et des supports utilisés).

Un point à souligner est que la valeur de ce système de bibliothèque de référence ne peut être réduite à la possibilité pour les chercheurs d'un accès rapide aux ressources les plus demandées. Pour parer immédiatement un de nos travers contemporains : ça n'est pas le *nombre* d'accès à un texte qui détermine son importance ou l'utilité de le conserver. Au contraire : la littérature mathématique est d'abord difficile pour les non-spécialistes, elle a donc une toute petite audience et est peu consultée, mais il serait tout simplement impossible de faire de la science sans les bases fiables fournies par le corpus mathématique dans son ensemble.

La naissance de communication électronique à la fin du 20^e siècle, qui est devenue un moyen omniprésent, presque exclusif de diffuser les connaissances de nos jours, n'a pas changé radicalement les besoins de la science. Elle a créé de nouvelles opportunités pour une diffusion plus facile et plus rapide, et des outils de fouille parmi les résultats scientifiques plus puissants. Malheureusement, elle a également stimulé un tel niveau de concurrence et de désorganisation parmi les fournisseurs de contenu numérique que beaucoup de scientifiques font maintenant face à une difficulté croissante pour accéder aux références publiées nécessaires à leur travail :

– La compression des budgets des bibliothèques et la multiplication des *big deals* (vente forcée de bouquets électroniques) font que la part du corpus papier disponible physiquement à la bibliothèque locale est en forte décroissance.

– Les bibliothèques locales conservent des versions papier éphémères (tirages laser, p. ex.) et des abonnements à des services en ligne dont les contenus originaux (numériques) ne sont pas archivés de façon sérieuse : le risque d'avoir perdu des pans entiers du corpus récent d'ici quelques années est réel.

– L'offre électronique a pris rapidement un tel essor que pratiquement tous les journaux vivants ont une édition électronique, une partie substantielle des journaux anciens a été rétronumérisée. Les livres, les thèses, les actes de séminaires ou de congrès, ainsi que d'autres composants utiles d'une bibliothèque de recherche en mathématiques sont de plus en plus souvent disponibles numériquement. Mais tout cela est dispersé parmi une myriade de fournisseurs, chacun ayant une politique spécifique sur les conditions d'accès à son contenu numérique. En outre, ces fournisseurs, leurs services, leurs serveurs sont très volatils : des collections entières se déplacent, ou disparaissent, quand des compagnies d'édition sont vendues, fusionnées, ou font faillite.

– Quelques ressources importantes sont bien souvent inaccessibles parce qu'elles ne sont pas référencées dans l'un des services professionnels de fouille dans la littérature, ou parce que l'URL fournie par ces services ne fournit pas le type d'accès qui a pourtant été payé.

Notre vision

Considérant les besoins des mathématiciens, de la science dans son ensemble, et que la bibliothèque papier se transforme peu à peu en une archive morte, nous concluons qu'il faut une nouvelle infrastructure fournissant le service attendu de la bibliothèque mathématique de référence dans le paradigme numérique. Comme un travail considérable a déjà été effectué pour convertir les textes mathématiques (anciens et actuels) au format numérique, nous estimons que l'effort devrait porter maintenant sur l'intégration de ce contenu dispersé dans une bibliothèque numérique distribuée de mathématiques.

Les résultats principaux du service de bibliothèque envisagé seraient de mettre en place un réseau d'institutions où les textes numériques seraient physiquement archivés. Chaque institution *locale* se chargerait de la sélection, de l'acquisition, du développement, de la maintenance, du catalogage et de l'indexation, ainsi que de la préservation de ses propres collections selon des politiques clairement établies : elle recevrait une sorte de « dépôt légal » pour une partie bien définie du corpus mathématique.

Le réseau formé par l'ensemble de ces institutions constituerait *une bibliothèque virtuelle globale* disposant d'un point d'accès au contenu distribué, au travers d'interfaces faciles à utiliser. En outre, le recours à des standards éprouvés permettrait à cette bibliothèque virtuelle de servir de couche d'infrastructure interopérable avec n'importe quelle composante de l'environnement de travail des scientifiques, permettant de transformer une référence intellectuelle à un résultat en un lien effectif vers sa rédaction.

Mon article [2] expose cette vision en détail, et aborde quelques-uns des défis qui restent à relever pour la réaliser. Une de ses conclusions est qu'on ne peut

pas espérer voir chaque acteur se conformer spontanément à des standards d'interopérabilité exigeants et coûteux, que la seule voie réaliste vers l'intégration passera donc par des solutions automatisées pour produire des métadonnées peut-être approximatives, mais permettant à un tel système de fonctionner.

Le projet EuDML

À l'échelle mondiale, il n'a pas été possible d'atteindre un consensus parmi les acteurs de la documentation mathématique (les mathématiciens en tant qu'auteurs, en tant qu'éditeurs ou en tant qu'utilisateurs, mais aussi d'autres scientifiques, les éditeurs de toute sorte, les documentalistes, les agrégateurs d'information scientifique et technique, les responsables d'organisations scientifiques...). Le projet pilote EuDML implémentera cette vision à l'échelle européenne, avec un nombre de partenaires réduit mais représentant une masse critique en termes de contenus et de diversité structurelle. La stratégie est d'élargir progressivement le groupe initial pour approcher ainsi par itérations successives du but visé.



Il faut saluer le soutien de l'Union européenne, qui permet la première avancée réelle dans ce domaine depuis plus de dix ans! Ce projet a en effet obtenu l'appui de la Commission européenne dans le cadre du programme « Compétitivité & Innovation » (CIP ICT PSP, « libre accès à l'information scientifique », projet n° 250503). Il a formellement démarré le

1^{er} février 2010 pour une durée de 36 mois avec un budget global de plus de 3 M€, pour un financement européen maximal de 1,6 M€. La réunion de lancement a eu lieu à Lisbonne les 4 et 5 février.

Le consortium EuDML

Le consortium EuDML se compose de 14 partenaires européens². Une dizaine d'entre eux sont des institutions publiques qui contribuent les principales collections de mathématiques numériques (en majorité numérisées) en Europe³. Un seul éditeur commercial (EDP Sciences) fait partie du consortium. Il contribue cinq journaux français dont la série ESAIM, éditée sous les auspices de la SMAI et les descendants de la revue RAIRO qui ont été numérisés par NUMDAM et sont en cours de mise en ligne. De nombreux autres éditeurs sont associés au projet, à travers leurs archives numérisées ou plus directement parce qu'ils utilisent pour leur édition électronique l'une des plates-formes partenaires du projet⁴. L'ensemble des

² La liste détaillée se trouve sur le site web du projet : www.eudml.eu. La Cellule MathDoc a une multiplicité 2 du fait de son statut d'unité mixte.

³ 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 BuDML.

⁴ Ce qui signifie qu'environ 90 revues fourniront directement leurs articles récents : elles sont diffusées par EDP Sciences, le projet CEDRAM de la Cellule MathDoc, ou le service ELiBM de la SME et du FIZ. 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.

partenaires représente une diversité impressionnante des 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 recevront pas de 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 présidera un comité scientifique consultatif. La bibliothèque universitaire de Göttingen contribuera les journaux numérisés par les projets ERAM et RusDML, ainsi que la plus grande collection de livres numérisés de mathématiques.

Le coordonnateur général du projet (gestion administrative, financière et technique) est José Borbinha, de l'Instituto Técnico superior (Lisbonne, Portugal), qui a été associé à la numérisation du journal *Portugaliae Mathematica* par la Bibliothèque nationale du Portugal, et a une grande expérience dans le secteur des bibliothèques numériques. J'en suis le coordonnateur scientifique (ce qui signifie surtout que je vais essayer de faire en sorte que les activités qui démarrent ne produisent pas un prototype éphémère, mais un service utile).

La situation actuelle

Vers l'an 2000 naissait en Amérique du Nord le concept DML (*Digital mathematics library* ou bibliothèque numérique de mathématiques, John Ewing a écrit l'un des textes fondateurs [7] à la demande de Philippe Tondeur, qui était alors à la tête de la division mathématique de la NSF). La bibliothèque de l'université de Cornell décrochait en 2002 une bourse de la NSF pour « planifier » la mise en œuvre du concept, qui s'est décliné par la suite dans le monde entier (WDML alias World DML de l'Union mathématique internationale (UMI), EMANI de Springer et quelques bibliothèques, DML nationales, etc.), sans jamais prendre corps.

Rétrospectivement, le principal bénéfice de ces initiatives aura été de motiver des projets nationaux, entre lesquels une certaine interaction a été maintenue par la participation à quelques conférences. Voir la table 1 pour une image très approximative des collections existantes⁵. On peut aussi recommander deux sources d'information sur l'état des collections numériques [8, 6].

J'ai publié un panorama de la situation en France et dans le monde il y a quelques années [1] (texte publié en 2008, mais rédigé pour l'essentiel en 2005). J'y constate que la situation mondiale est au point mort, essentiellement à cause de conflits d'intérêts⁶, tandis que le microcosme français, très en pointe dans ce domaine, préfigure toutes les composantes prévues de la DML et leurs interactions⁷.

⁵ Les textes considérés sont des textes originaux de mathématiques, plutôt niveau recherche, comme des livres, des articles de revues ou des thèses. En fonction des sources mentionnées ci-dessous, il n'est pas toujours possible de déterminer une estimation, même grossière. Une borne inférieure pour le nombre de textes numériques actuellement disponibles est suggérée par les 1,2 million de liens directs enregistrés dans les bases de données *Math. Reviews* et *Zentralblatt*.

⁶ Pour donner une idée de ces conflits, posons-nous quelques questions en apparence naïves... À qui appartient la littérature mathématique ? À qui doit-elle profiter ? À qui appartient son catalogue ?

⁷ NUMDAM comme bibliothèque des journaux et séminaires français (60 séries depuis 1810, 40 000 articles sur 1 million de pages en 2010) alimentée par numérisation et par un nombre croissant d'éditeurs (à ce jour : Elsevier/IHP et ÉNS, Springer/IHÉS, CEDRAM/laboratoires

TAB. 1. Estimation du contenu DML existant.

<p>Amériques : JSTOR (235 000 textes), project Euclid (100 000), Canadian Math. Society (4 000)</p> <p>Asie : DML-JP (30 000 textes)</p> <p>Europe : partenaires EuDML, <i>associés</i> et envisagés (205 000 textes)</p> <p> Allemagne : ELibM, <i>Mathematica</i>, <i>ERAM/JFM</i> (85 000 textes)</p> <p> Bulgarie : BuDML (2 500 textes)</p> <p> Espagne : DML-E (5 000 textes)</p> <p> France : <i>Gallica-Math</i>, <i>TEL</i>, NUMDAM, CEDRAM (50 000 textes)</p> <p> Grèce : HDML (15 000 textes)</p> <p> Pologne : DML-PL (13 000 textes)</p> <p> Portugal : SPM/BNP (2 000 textes)</p> <p> Rép. Tchèque : DML-CZ (26 000 textes)</p> <p> Russie : <i>RusDML</i> (13 000 textes)</p> <p> Serbie : bibliothèque informelle (3 700 textes)</p> <p> Suisse : <i>SwissDML</i> (5 000 textes)</p> <p>Commercial : 700 000 textes ?</p> <p> Springer : 14 journaux chez <i>GDZ</i>, 1 chez NUMDAM, 120 dans <i>Online Archives</i>, 179 vivants (300 000 textes)</p> <p> Elsevier : 4 journaux chez NUMDAM, 63 dans <i>Backfiles</i>, 100 vivants (320 000 textes)</p> <p> Autres : Cambridge University Press : 20 journaux, Oxford University Press : 30, Hindawi : 18, Walter de Gruyter : 13, Wiley : 42, Taylor & Francis : 58. . .</p>

Plusieurs conclusions du projet NSF de Cornell [9] sont toujours pertinentes. Tandis qu'un certain nombre d'entre elles mène à une impasse parce qu'aucun accord n'a été conclu à travers le monde sur des questions importantes comme la sélection des textes, le droit d'auteur ou le modèle économique, d'autres sont consensuelles et ont de fait été approuvées par l'UMI en 2002 (navigation et métadonnées libres, libre accès à terme [3]) et 2006 (meilleures pratiques pour la rétronumérisation [5], une vision pour la DML [4]).

Cependant, de nombreuses activités ont eu lieu dans l'intervalle, ce qui modifie profondément le paysage, et pose des problèmes différents. Le contenu mathématique numérique disponible est maintenant considérable, particulièrement en Europe où les projets nationaux ont réuni une partie significative du corpus au format numérique (voir la table 2). Les éditeurs commerciaux, qui étaient peu disposés à investir dans la numérisation et espéraient le déblocage de fonds publics, ont maintenant sauté le pas, et proposent une version privatisée de la fonction bibliothèque qui soulève quelques interrogations sur les critères de sélection des collections ou la pérennité d'une telle entreprise.

français, EDP Sciences/SMIAI, IMS/Euclid/IHP, SFdS). Rayons spéciaux de la DML française non encore intégrés à NUMDAM : Gallica-Math, archives Bourbaki, publications d'Orsay... CEDRAM, plate-forme d'édition exemplaire en terme d'ergonomie et d'interopérabilité. LiNum et Mini-DML, bases de donnée basiques permettant de chercher dans plusieurs collections à travers le monde (y compris certaines qui ne sont pas cherchables sur leur propre site!), supposées démontrer la faisabilité de la DML telle que je la conçois, mais illustrant aussi les difficultés car des dizaines de sources en principe favorables au projet ne nous fournissent pas de données exploitables.

TAB. 2. Services intégrés ou intégrables dans EuDML.

<p>Partenaires EuDML : bibliothèques numériques</p> <p>DML-CZ : DML tchèque : http://dml.cz/</p> <p>NUMDAM : DML journaux français (Cellule MathDoc) : http://www.numdam.org/</p> <p>HDML : DML grecque : http://dSPACE.eap.gr/dSPACE/handle/123456789/46</p> <p>DML-PL : DML polonaise : http://matwbn.icm.edu.pl/</p> <p>SPM/BNP : <i>Portugaliae Mathematica</i> numérisée : http://purl.pt/index/pmath/PT/index.html</p> <p>DML-E : DML espagnole : http://dml.e.cindoc.csic.es/en/portada_en.php</p> <p>Partenaires EuDML : édition électronique</p> <p>CEDRAM : Centre de diffusion de revues académiques mathématiques (Cellule MathDoc) : http://www.cedram.org/?lang=en</p> <p>ElibM : Bibliothèque de journaux sur le serveur EMIS : http://www.emis.de/journals/</p> <p>EDP Sciences : 5 revues mathématiques : http://www.edpsciences.org/</p> <p>Partenaires & collections associés à EuDML</p> <p>Gallica-Math : Contenu mathématique de Gallica indexé par la Cellule MathDoc : http://math-doc.ujf-grenoble.fr/GALLICA/</p> <p>TEL : Thèses électroniques françaises : http://tel.archives-ouvertes.fr/</p> <p>GDZ : Collections Mathematica et RusDML de la bibliothèque de Göttingen : http://gdz.sub.uni-goettingen.de/ et DigiZeitschriften : http://www.digiZeitschriften.de/</p> <p>Futurs associés d'EuDML ?</p> <p>Serbie: eLibrary of Mathematical Institute of the Serbian Academy of Sciences and Arts : http://elib.mi.sanu.ac.rs/</p> <p>Suisse : SwissDML : http://retro.seals.ch/</p> <p>Italie : DIGIMAT (projet de l'Unione Matematica Italiana)</p> <p>EuDML : http://www.eudml.eu/</p>
--

Tandis que des bibliothèques numériques locales étaient créées, une communauté de recherche active, basée principalement en Europe, a émergé sous l'intitulé MKM (*Mathematics knowledge management*, gestion automatisée des savoirs mathématiques). Elle vise à développer des outils pour gérer les savoirs mathématiques au format numérique, faisant un pont entre les mathématiques formalisées et les outils de démonstration automatique et... les textes relativement peu structurés que les humains écrivent. Il existe déjà des logiciels de reconnaissance optique de formules mathématiques, d'autres pour l'extraction ou l'inférence automatique de métadonnées. D'autres logiciels permettent de mettre en relation une citation écrite et les entrées correspondantes dans plusieurs bases de données. Ces technologies ne sont pas toutes directement exploitables en production, mais elles permettent d'envisager des chaînes de traitement largement automatisées pour extraire des métadonnées riches à partir desquelles on peut imaginer changer de paradigme pour la recherche de textes mathématiques.

La conservation à long terme du corpus mathématique est une question importante. Il est certain que de mauvaises décisions à cet égard ont déjà affecté des articles édités au début de l'ère électronique, et ce probablement jusqu'à très récemment. Nous espérons qu'il est encore possible d'identifier rapidement les textes en danger et de les sauver avant leur obsolescence.

EuDML : stratégie, actions et calendrier

Il nous semble que c'est le moment opportun pour inviter les parties concernées à coordonner leurs efforts, et créer une infrastructure utile et efficace. Le contexte européen est assez diversifié, tout en restant de dimension maniable, pour y concevoir et réaliser une version réduite mais pleinement fonctionnelle de la DML. Nous comptons sur l'effet catalytique du financement européen pour réaliser rapidement un premier prototype du service envisagé, ouvrant d'emblée un accès libre et facilité à un ensemble important de textes fondamentaux hébergés par nos partenaires. Au cours de la seconde phase du projet, tandis que les partenaires technologiques s'efforceront de rendre le système plus performant, nous tenterons de convaincre les utilisateurs et les contributeurs potentiels des bénéfices qu'ils pourront tirer de ce service. De la sorte, nous allons définir non seulement des normes techniques d'interopérabilité, mais aussi des standards non contraignants de coopération entre les différents acteurs. L'extension à de nouveaux partenaires devrait alors se faire naturellement.

La première étape des travaux, qui vient de commencer, consiste à agréger un dépôt central de métadonnées et doter chaque texte d'un identifiant permanent. C'est le degré zéro de l'intégration : la création d'une base de données unique de tous les textes contribués par les partenaires du projet. Cette tâche sera réalisée au cours de la première année, tandis que le système central et les services associés seront conçus et réalisés. Nous comptons avoir un site web fonctionnel en service autour de l'été 2011, et un système beaucoup plus sophistiqué vers la fin 2012. Nous en appellerons à la communauté mathématique pour tester le système quand ces jalons importants seront atteints.

Tous les services à valeur ajoutée seront développés à partir de cette plate-forme de base.

– Les humains pourront interroger la base EuDML par l'intermédiaire d'une interface web, les programmes disposeront de services web. Pour une meilleure implication des utilisateurs et une grande interactivité, nous ajouterons des dispositifs Web 2.0 comme la possibilité de créer un environnement personnalisé de travail. Il sera par exemple possible d'enregistrer des annotations personnelles sur des textes de la bibliothèque. Elles pourront être privées, partagées avec une communauté d'intérêt, ou finalement intégrées dans la base. Songez que les deux articles qui démontrent le théorème des nombres premiers ne l'appellent jamais ainsi !

– Puisque les métadonnées existantes sont très hétérogènes, souvent incomplètes et monolingues, nous voulons les améliorer en ayant recours à toutes les technologies dont nous pourrions disposer. Ce sera un processus itératif et continu utilisant toutes les relations entre objets indexés pour en déduire des métadonnées plausibles. Parmi ces techniques, citons

- la reconnaissance optique de caractère : textuelle, structurée, mathématique, avec pour but ultime une version XML/MathML exploitable des textes ;
- l'exploitation des formules mathématique comme métadonnée ;
- l'identification des citations ;
- la génération de liens internes tous azimuts.

– Pour améliorer la fouille dans notre corpus, nous voulons profiter du fait que le contenu de ces collections est par nature fortement mathématique (au lieu de le subir comme une punition !). Nous aurons recours à des techniques MKM pour surmonter les barrières linguistiques, pour mettre en relation des textes sur la base de leur contenu mathématique comme par exemple la similarité des formules employées.

– Nous nous efforcerons aussi de rendre notre contenu plus accessible, notamment aux utilisateurs malvoyants ou dyslexiques, en le rendant disponible aux formats *ad hoc*.

– Pour montrer l'utilité de notre entreprise, nous escomptons faire apparaître des références à des textes mathématiques fondateurs sous la forme de liens dans de nombreux sites, en fournissant à leurs auteurs les outils pour ce faire. Un versement du contenu EuDML dans la bibliothèque numérique européenne (www.europeana.eu) est prévu, et des expositions temporaires pourront y être organisées. Certains collègues vont jusqu'à prétendre que cela fera remonter la cote des mathématiques européennes parmi les citoyens de l'Union, et suscitera donc plus de vocations au sein de la jeunesse !

En parallèle, nous inviterons tous les acteurs qui le souhaitent (les utilisateurs, les éditeurs, les documentalistes, les organismes de recherche, etc.) à discuter des politiques à long terme en vue de la pérennisation du service au-delà de la durée de vie du projet. L'objectif ultime étant de définir une politique raisonnée d'archivage pérenne et de libre accès à terme pour le corpus mathématique.

Le versant français

Les partenaires français du projet EuDML sont : EDP Sciences, l'université Joseph-Fourier et le CNRS, ces deux derniers apparaissant en fait comme tutelles de la Cellule MathDoc (UMS 5638). EDP Sciences contribue 5 journaux, une expertise dans la production, la gestion et la conversion de métadonnées, et le point de vue d'un éditeur privé proche des communautés scientifiques. La Cellule MathDoc contribue ses 65 séries (en ligne ou en cours de traitement), ses outils de création de liens, son expérience sur tous les fronts du chantier DML depuis plus de dix ans, et l'auteur de ces lignes, promoteur infatigable de la DML, mandaté par la SME pour réunir un consortium et définir le projet qui vient donc de démarrer.

NUMDAM abrite la quasi-totalité des revues vivantes de mathématiques éditées en France, et sert plus ou moins de tête de pont pour le réseau français des centres de numérisation de textes mathématiques (notamment à travers ses collaborations avec la BNF, le RNBm, les bibliothèques d'Orsay, de Polytechnique, de Jussieu, etc.). Nous comptons servir d'intermédiaire pour contribuer toutes les collections mathématiques françaises qui le souhaiteront.

Toutes les revues de la SMAI sont donc concernées par ce projet, des archives numérisées par NUMDAM aux articles nouveaux publiés soit par EDP Sciences, soit par le CEDRAM. Le *Bulletin* et les *Mémoires* de la SMF, numérisés par NUMDAM jusqu'en l'an 2000, sont également de la partie, mais la chaîne d'acquisition pour la production post-numérisation n'est pas encore en place : espérons que ce projet soit l'occasion de faire avancer ce dossier !

Outre le rôle d'animation et de communication qui échoit au coordinateur scientifique, la Cellule MathDoc supervise la définition des standards de métadonnées

et le moissonnage des collections partenaires. Elle adaptera aussi les outils qui font de NUMDAM et CEDRAM des services salués par les mathématiciens du monde entier (notamment les liens vers les bases de données de référence et vers les articles cités, les métadonnées duales TeX/MathML). Elle espère apprendre de ses partenaires des techniques qui lui font aujourd'hui défaut et enrichir ses services. Finalement, elle espère contribuer à la réussite du projet en relevant quelques défis (qui ne sont pas tous technologiques), et voir enfin ses projets de bibliothèque numérique portés à l'échelon européen.

La prochaine réunion du consortium est prévue en juillet à Paris, à l'occasion de la tenue au CNAM de la série de conférences CICM (Conferences on Intelligent Computer Mathematics), dont l'atelier DML où devraient être présentés les premiers résultats du projet.

Références

- [1] T. BOUCHE – « Toward a digital mathematics library? », in *Communicating mathematics in the digital era* (J. Borwein, E. Rocha & J. Rodrigues, eds.), AK Peters Ltd, 2008, p. 47–73.
- [2] _____, « Digital Mathematics Libraries : The Good, the Bad, the Ugly », *Mathematics in Computer Science* 3 (2010), special issue on Authoring, Digitalization and Management of Mathematical Knowledge (Serge Autexier, Petr Sojka, and Masakazu Suzuki eds.), sous presse.
- [3] COMMITTEE ON ELECTRONIC INFORMATION COMMUNICATION OF THE INTERNATIONAL MATHEMATICAL UNION – « Best Current Practices : Recommendations on Electronic Information Communication », *Notices of the AMS* 49 (2002), no. 8, p. 922–925.
- [4] _____, « Digital Mathematics Library : A Vision for the Future », http://www.ceic.math.ca/Publications/dml_vision.pdf, August 2006.
- [5] _____, « Some Best Practices for Retrodigitization », http://www.ceic.math.ca/Publications/retro_bestpractices.pdf, August 2006.
- [6] AMS Digital Mathematics Registry : <http://www.ams.org/dmr/JournalList.html>.
- [7] J. EWING – « Twenty Centuries of Mathematics : Digitizing and Disseminating the Past Mathematical Literature », *Notices of the AMS* 49 (2002), no. 7, p. 771–777.
- [8] U. Rehmann – *Retrodigitized Mathematics Journals and Monographs* : http://www.mathematik.uni-bielefeld.de/~rehmann/DML/dml_links.html
- [9] S. E. THOMAS, R. K. DENNIS & J. POLAND (eds.) – *Digital Mathematics Library. A one-year (2002-2003) planning project coordinated by Cornell University Library and funded by the U.S. National Science Foundation (NSF) toward the establishment of a comprehensive, international, distributed collection of digital information and published knowledge in mathematics. Final Report*, Cornell University Library, October 2004, http://www.library.cornell.edu/dmlib/DMLreport_final.pdf.