

Synthèse du travail mené dans le cadre du chantier « nettoyage du contenu du CFP » et proposition pour la suite

Version finale au 15/02/2022

GT "Contenu du CFP" Mathdoc-RNBM

Initié en 1995, le CFP a été constitué à partir des collections des bibliothèques de mathématiques, dont certaines étaient pluridisciplinaires (informatique, physique), et par l'import des données de la base ISSN. En plus des bibliothèques du RNBM, la BNF, l'INIST et l'INRIA avaient versé leurs fonds dans le CFP.

L'objectif initial de ce catalogue était de partager les informations de localisation entre les bibliothèques de mathématiques, et de localiser facilement une revue pour répondre à des demandes de fourniture d'articles.

Le CFP est mis en œuvre et maintenu techniquement par Mathdoc ; la mise à jour du contenu est effectuée en collaboration entre le RNBM et Mathdoc.

Dès les premiers travaux sur un Plan de conservation partagée des revues imprimées de mathématiques (PCMath), le CFP, déjà bien implanté dans la communauté, a été au cœur du projet et validé comme outil du plan. En parallèle, le CFP est envisagé comme outil utile pour d'autres projets comme la gestion des ressources électroniques.

Jusqu'en 2016, les mises à jour des données bibliographiques du CFP étaient sous la responsabilité de la documentaliste affectée à Mathdoc. De façon informelle, à la vacance du poste et avec le développement des besoins de mise à jour du PCMath, le comité technique (CT) du PCMath et le groupe Licence ont pris en charge les ajouts et modifications des notices dont ils avaient besoin.

Pourquoi un chantier de nettoyage ?

Du fait des conditions initiales de constitution du CFP, la couverture thématique des titres indexés s'étend bien au-delà des domaines des mathématiques. La base contient des notices bibliographiques de revues mais aussi de collections de monographies et de séminaires. Avec l'utilisation du CFP comme outil pour le PCMath ou la gestion des ressources électroniques, cette situation engendre un bruit important pouvant s'avérer gênant et alourdit la charge de mise à jour des données.

En 2014, Mathdoc et le RNBM lancent un travail d'analyse du contenu du CFP et de définition du périmètre du catalogue. Un *Etat des lieux du contenu du CFP* est dressé par Sophie Min-Picault, documentaliste à Mathdoc (voir document joint). Le CFP contient alors **19 380** notices, dont 5 206 sont classées dans les domaines mathématiques selon les classifications Dewey ou CDU. Ces notices sont associées à des notices d'exemplaires décrivant les états de collections imprimées des bibliothèques participantes : 3 694 notices ont des localisations dans au moins une des 38 bibliothèques participantes. 1 745 de ces titres étaient des unicas (i.e. avec une seule bibliothèque détentrice).

En 2015, un groupe de travail est mis en place, comprenant des représentants du PCMath, du Portail Math, et de l'actuel comité éditorial du site licence.rnbnm, pour faire le point sur le contenu du CFP et les fonctionnalités attendues pour chaque projet.

Le CFP est défini comme une base de connaissance sur les revues en mathématiques, pour servir de support à des services destinés à la communauté mathématique.

L'exploitation de la base du CFP permet :

- la sélection du corpus du PCMath ;
- la recherche et la localisation des périodiques ;
- la gestion des périodiques électroniques acquis par le biais d'une licence nationale ;

- la gestion des collections imprimées et électroniques dans les bibliothèques du réseau ;
- l'identification des ressources à numériser.

Le bon fonctionnement de ces services implique de pouvoir s'appuyer sur un corpus clairement défini et contrôlé.

Lors d'une réunion du groupe de travail le 16/06/2015, sur la base du document « Etat des lieux du CFP », il a été convenu que :

- les bibliothèques participantes au CFP sont celles qui sont parties prenantes des activités dont le CFP est le pivot. Cela a conduit à écarter la BNF, l'INIST et l'INRIA ;
- le **corpus** du CFP doit être revu, de façon à ne conserver que les titres constituant le cœur des revues mathématiques (avec les titres en lien, les titres définis dans les projets ou à la demande de la communauté) ;
- les données bibliographiques s'appuient prioritairement sur le Sudoc pour l'imprimé, sur la base ISSN pour l'électronique ;
- la mise à jour manuelle des données d'exemplaires par les bibliothèques est un travail redondant. Une solution d'export de données du Sudoc vers le CFP pour les données d'exemplaires (titres imprimés) est à étudier ;
- le constat d'une nécessité de nettoyer la base et de s'accorder au préalable sur des critères de contenu du CFP.

De 2015 à 2020, le CFP a été intégralement ré-écrit informatiquement pour permettre l'ajout de fonctionnalités liées au PCMath et à la gestion des ressources imprimées et électroniques (pour plus de détails, voir l'annexe p. 5). Cette ré-écriture était jugée nécessaire pour débiter le chantier de nettoyage.

Mise en œuvre du chantier

En 2020, les travaux de sélection du corpus du PCMath par les scientifiques, travaux préalables à la poursuite des phases de traitement du corpus du PCMath, rendent indispensable le chantier de nettoyage du CFP.

En mai 2020, le groupe de travail est ré-activé ; il comprend des représentants du PCMath (N. Granottier, D. Barrère), du groupe Licence (G. Thureau), du groupe Documentation électronique (F. Leinardi), du groupe Portail (A. Wojciechowska), ainsi que Mathdoc (O. Labbe, T. Bouche), le bureau du RNBM (E. Kneller, directrice) et C. Sabbah, chargé de mission au bureau du RNBM et membre du comité de pilotage du PCMath.

Entre juin et juillet 2020, le groupe s'est attelé à identifier et formaliser les critères de tri et à définir le contenu du CFP :

a – Critères permettant de mettre en place un tri automatique des 19 380 notices contenues dans le CFP

Sont à retirer du CFP : les titres sous lesquels seuls sont localisés : la BNF, l'Inria, l'INIST.

Font partie du corpus du CFP (et sont à garder dans le CFP) :

- le corpus du PCMath sélectionné à cette date, soit 1005 titres ;
- les titres qui ont au moins une bibliothèque localisée ;
- les titres du DOAJ Math ;
- les titres de la base zbMath ;
- les titres liés (historique des titres à garder).

Sur cette base, Mathdoc a pu mettre en place sur le CFP un filtre pour distinguer les titres du *Corpus CFP* et les titres *hors corpus CFP* (accès avec authentification administrateur).

À l'issue de cette première phase, le CFP contient 7 251 titres dans le CFP-corpus et 9 017 titres hors corpus-CFP.

b – Critères de contenu du CFP

Le groupe de travail convient que :

- Le CFP est un outil pour la documentation en mathématiques. Il est utilisé par les professionnels de la documentation spécialisés en maths et doit permettre d’avoir une vue d’ensemble des collections disponibles dans les bibliothèques affiliées. Il bénéficie aux scientifiques soit directement pour leurs recherches, soit indirectement par le biais du Portail Math par exemple.
- Le CFP est un outil de localisation des revues dans les bibliothèques desservant la communauté de recherche en mathématiques mais aussi un outil pour le Portail Math et pour l’application Geslic (gestion des accès aux revues électroniques).
- Le CFP a pour objectif de recenser les revues de recherche en mathématiques (localisation physique ou accès électronique), y compris les revues consacrées à l’histoire des mathématiques, l’enseignement des mathématiques et la vulgarisation des mathématiques.

- Font donc partie du corpus du CFP :
 - les titres du CFP localisés dans une bibliothèque de mathématiques ou ayant une composante mathématique ;
 - les titres du corpus PCMath ;
 - les titres liés par une relation historique à des titres présents dans le corpus dans la mesure où ils sont centrés sur les mathématiques ;
 - les titres ayant une période d’accès électronique, dont les titres électroniques DOAJ ;
 - les titres indexés dans la base zbMath ;
 - les titres indexés dans la base Mathscinet ;
 - les titres électroniques sans localisation papier mais avec une localisation « électronique », c’est-à-dire, soit un titre avec une licence active dans au moins une bibliothèque, soit un titre en libre accès sélectionné notamment en s'appuyant sur l'[AMS Digital Mathematics Registry 1](#) et le portail [Mathnet.ru](#) ;
 - les publications des IREMs et de l'APMEP ;
 - les titres mathématiques localisés dans le Sudoc qui seraient absents du CFP.

- Sont donc à exclure du CFP : les revues généralistes en science, même si elles contiennent des articles en mathématiques (exemples Nature, Science, etc).

Quelques critères ont été précisés par la suite :

- garder les titres en recherche opérationnelle / intelligence artificielle / histoire des sciences / logique ;
- garder les revues en lien avec les activités des IREM ;
- exclure les titres en astronomie / acoustique / optique / physique théorique ;
- garder uniquement la lignée « Mathématique » pour les titres historiques généralistes aux ramifications thématiques multiples ;
- garder pour les traiter à part ultérieurement les notices de collections de monographies et de séminaires.

Entre novembre 2020 et février 2022, le groupe a procédé à l’analyse et au balisage des titres en fonction des critères établis. 10 976 titres hors corpus CFP, sans bibliothèque localisée, ont été retirés du CFP.

Une prochaine étape prévue entre mars et juin 2022 va permettre de clore ce chantier en traitant les 521 titres qui devraient être retirés du CFP selon les critères de contenu mais qui ont au moins une localisation. Cette phase va nécessiter d’impliquer les bibliothèques concernées.

Un dispositif est mis en place dans le CFP pour permettre une concertation avec les bibliothèques avant de valider le retrait du CFP :

- chaque bibliothèque voit dans son tableau de bord (après authentification) les titres qu'elle détient parmi cette liste de 521 titres ;
- chaque bibliothèque envoie au groupe de travail un avis argumenté pour les titres qu'elle souhaite conserver dans le CFP ;
- après le 30 juin 2022, les titres non signalés seront retirés du CFP.

Ce dispositif sera présenté lors de la Journée des adhérents PCMath du 08/03/2022 et un message sera envoyé sur la liste du RNBM pour informer les bibliothèques hors PCMath.

Proposition pour la gestion du contenu du CFP

Des perspectives de service à la communauté, d'ores et déjà identifiées, nécessitent de s'appuyer sur un catalogue de qualité.

Pour garantir cette qualité et pour que ce travail de désherbage du CFP ne soit pas vain, il faut veiller à ce que dès à présent les critères de définition du corpus soient explicitement connus et que des procédures de mise à jour soient clairement établies.

Le groupe de travail propose donc la mise en place d'un comité éditorial du CFP au plus tôt.

Le comité éditorial aurait pour rôle de :

- garantir la qualité et la cohérence de la base ;
- définir les critères de contenu du CFP et les règles de mise à jour ;
- procéder à l'alimentation et au maintien du corpus du CFP ;
- accompagner les groupes dans leurs projets basés sur le CFP :
 - en identifiant les besoins, les améliorations possibles ;
 - en identifiant les nouvelles fonctionnalités utiles ;
 - en accompagnant les développements nécessaires.

Le comité éditorial serait composé de :

- un membre Mathdoc ;
- un représentant du comité stratégique du RNBM ;
- un membre de chaque groupe de travail concerné : CT du PCMath, groupe licences, groupe documentation électronique, etc. ;
- un représentant des bibliothèques utilisatrices ;
- un représentant de la communauté scientifique.

La mise en place du comité éditorial doit être effective rapidement et au plus tard avant la fin de l'opération de nettoyage (juin 2022).

Il nous semble nécessaire qu'**avant fin mars 2022**, le comité stratégique ait validé le dispositif proposé et lancé l'appel à candidature pour la composition du comité éditorial.

ANNEXE

Qu'est-ce que le catalogue fusionné des périodiques de mathématiques ?

Historique

Dès 1995, le CFP a été constitué à partir des collections des bibliothèques de mathématiques du RNBM ainsi que de celles de la BNF, l'INIST, l'Inria, et par l'import des données de la base ISSN

L'objectif initial était de partager les informations de localisation entre les bibliothèques de mathématiques, le Sudoc n'ayant été complètement déployé qu'en 2002. Le CFP permettait ainsi de localiser facilement un fascicule pour répondre à des demandes de fourniture d'articles, mais il nécessitait un travail important de mise à jour de la part des bibliothèques.

Cet outil est techniquement maintenu par Mathdoc qui l'a fait évoluer en coordination avec différents groupes de travail du RNBM en fonction des évolutions techniques et des besoins exprimés permettant des améliorations substantielles tant en termes de qualité des données que de fonctionnalités.

Dès les premiers travaux sur un Plan de conservation partagée des revues imprimées de mathématiques (PCMath), le CFP, bien implanté dans la communauté, a été au cœur du projet et validé comme outil du plan. En 2009-2010, une refonte du CFP a été menée par Mathdoc, notamment dans la perspective de la conservation partagée. Le nouveau CFP a été présenté lors des Journées RNBM en février 2011.

Dans le cadre du programme national CollEx, un soutien en CDD en 2016-2017 et 2017-2018 a été financé auprès de Mathdoc pour l'interopérabilité avec le catalogue du Sudoc, rapidement apparue comme essentielle pour éviter le doublonnage des actions et rendre visible au niveau national le travail de signalement du PCMath. Le catalogue est aujourd'hui l'outil de gestion du PCMath.

En parallèle, le CFP est apparu comme central pour faciliter la gestion des ressources électroniques et leur accès via le Portail Math. L'application GesLic a été conçue en 2017 pour centraliser les données sur les abonnements aux revues électroniques des bibliothèques du RNBM. Elle est adossée au site licence.rnbn.org pour l'extraction des données des contrats de licence des éditeurs.

Après une première version de l'application développée par Mathdoc et le RNBM entre 2017 et 2019, testée et alimentée par 5 bibliothèques pilotes (membres du GT Portail Math), l'application GesLic intègre en 2020 le logiciel du CFP pour faciliter sa maintenance. En 2021, une équipe d'administrateurs de GesLic a été constituée ; elle rassemble les membres du comité éditorial du site licence.rnbn.org et Simon Panay (Mathdoc). Cette équipe a permis d'améliorer les interfaces administrateurs et utilisateurs de GesLic.

Contenu

Le CFP contient des notices bibliographiques décrivant des revues mais aussi des collections et des séminaires. L'enrichissement des notices issues des catalogues des bibliothèques par des imports de la base ISSN a notamment permis l'intégration de titres liés historiquement.

En 2014, le CFP contenait 19 380 notices, dont 5 206 classées dans les domaines mathématiques selon les classifications Dewey ou CDU. Beaucoup de titres étaient interdisciplinaires et une bonne partie très généralistes, principalement dans le fonds de l'INIST.

Les notices initiales concernaient principalement des titres au format imprimé. La base a été enrichie avec des titres uniquement accessibles au format électronique.

Un lien vers la notice du catalogue national du Sudoc est automatiquement créé par l'interrogation d'un web services de l'ABES (le lien se fait au travers du numéro d'identification unique qu'est le PPN de la notice dans le Sudoc).

Les notices bibliographiques sont associées à des notices d'exemplaires décrivant les états de collection des versions imprimées des bibliothèques participantes. En 2014, 3 694 notices avaient des localisations dans au moins une des 38 bibliothèques participantes. 1 745 de ces titres étaient des unicas, i.e. avec une seule bibliothèque détentrice.

Depuis 2017, Mathdoc a développé une mise à jour automatique des états de collection du Sudoc. Les bibliothèques mettent à jour les états de collection et de lacunes dans le Sudoc et transfèrent ces données dans le CFP par un simple clic, simplifiant le travail qui devait être fait deux fois jusque-là. Inversement, le PCMath est visible au niveau national dans l'outil Périscope de l'Abes (<https://periscope.sudoc.fr/>).

Ces données d'exemplaires ont été enrichies pour les besoins du PCMath avec des informations de description technique de la collection (mètres linéaires, état de la reliure, état physique de conservation).

Le CFP a très vite évolué vers un catalogue signalant les versions imprimées et électroniques des titres de revue. Pour faciliter la gestion et éviter les doublons de notices, le CFP affiche sous une même notice pour un titre donné, les données sur les supports imprimés et/ou électroniques (données ISBD papier, ISSN, PPN électronique, lien vers la version électronique).

Les liens vers la version électronique des titres (périodes d'accès électronique) sous contrat de licence ont été importés de manière groupée à partir du site licence.rnbnm (<https://licence.rnbnm.org>)

L'application GesLic a été conçue pour signaler les abonnements électroniques de chaque bibliothèque du RNBM.

Depuis sa conception en 2017, elle s'est enrichie successivement des corpus de titres des licences nationales, des licences consortiales (type Couperin) puis des licences-type éditeur.

Les bibliothèques peuvent se rattacher aux licences souscrites afin que les données d'exemplaires (i.e. les périodes souscrites) s'affichent dans une frise sur chaque notice du CFP (comme pour les abonnements à la version imprimée).

En 2020, MathDoc et le RNBM ont souhaité valoriser les titres de revues mathématiques en Open Access dans le CFP. Les liens vers les revues en ligne de plusieurs DML (EuDML, Math.net.ru, GDZ, Scielo, etc.) et du DOAJ (dont les titres du Centre Mersenne et de Numdam). Un programme lié au DOAJ permet d'importer les liens des titres de revues du domaine des mathématiques. Ce programme peut être lancé à tout moment pour pouvoir actualiser ces liens.

<p>Fonctionnalités</p> <p>Le CFP est un catalogue public permettant une vision simplifiée et unifiée des périodiques en mathématiques quel que soit son format et sa localisation en France. Il est utilisé dans différents environnements et services proposés par le RNBM et Mathdoc.</p>	
<p>Signalement et gestion des collections imprimés des bibliothèques</p> <p>Le catalogue permet aux bibliothèques de :</p> <ul style="list-style-type: none"> • se localiser sous les titres qu'elles détiennent dans leurs fonds ; • afficher la liste des titres détenus ; • afficher pour chaque titre ses états de collection et de lacunes ; • décrire les états physiques de ses collections ; • identifier et afficher ses unicas. <p>Elles bénéficient de plusieurs fonctionnalités pour faciliter la gestion de leurs fonds et les mises à jour des autres catalogues qu'elles alimentent :</p> <ul style="list-style-type: none"> • import automatique des états de collection et états de lacunes à partir du Sudoc ; • hyperlien avec la notice bibliographique du périodique dans le Sudoc ; • export de ses titres pour des analyses et traitements locaux ; • affichage graphique des historiques des titres via un arbre graphique dynamique qui permet la navigation entre les titres et le repérage d'anomalies ; • import des collections d'une bibliothèque via un fichier de PPN. 	<p>Gestion du Plan de conservation partagée (PCMath)</p> <p>Le CFP est l'outil du plan de conservation partagée des périodiques de mathématiques (PCMath).</p> <p>Fonctionnalités spécifiques au PCMath proposées aux bibliothèques participantes (en plus des fonctionnalités de base du CFP présentées ci-contre) :</p> <ul style="list-style-type: none"> • affichage du corpus PCMath complet avec le statut de conservation de chaque bibliothèque participante selon deux vues (synthétique ou détaillée) pour des exports appropriés ; • à partir du tableau de bord personnel à chaque bibliothèque : affichage des collections PCMath propres à la bibliothèque, campagne de positionnement et de gestion des candidatures pôles de conservation et pôles associés, affichage ponctuel lié à un projet donné (exemple : phase finale du nettoyage du CFP chantier 2020-2022). <p>Un accès avec des droits étendus a été mis en place pour les membres du comité technique permettant :</p> <ul style="list-style-type: none"> • la correction des notices bibliographiques via l'import Sudoc et ISSN ; • la correction ou création de notice manuelle ; • la gestion des phases du PCMath ; • la saisie et la modification du positionnement des bibliothèques participantes lors des campagnes de candidatures : pôles de conservation et pôles associés ; • l'identification et la gestion des Colref ("Collections de référence") ; • l'affichage des autres plan de conservation nationaux et régionaux en lien avec le Sudoc. <p>Ces évolutions sont suivies de près par l'Abes qui souhaite s'en inspirer pour faire évoluer l'affichage des périodiques.</p>
<p>Signalement et gestion des abonnements électroniques</p> <p>À partir de l'application GesLic intégrée au CFP, les bibliothèques peuvent :</p> <ul style="list-style-type: none"> • signaler leurs propres abonnements électroniques ainsi que ceux de leurs tutelles et les abonnements nationaux ; • afficher sur chaque notice de périodique les périodes d'accès souscrites sur une frise de temps associée à une bulle de texte ; • à partir d'une liste AtoZ (ou via le service « Collections électroniques ») la bibliothèque peut visualiser un historique de ces abonnements et effectuer des recherches par titre, ISSN, PPN, éditeur. 	<p>Services web associés</p> <p>Numdam utilise un service web du CFP pour afficher l'historique des revues, à partir des données du CFP.</p>