

Journée des adhérents au PCMath 2022 – le 8 mars 2022 à l’Institut Henri Poincaré

Matinée :

09h00-09h30 -Accueil et tour de table

09h30-10h00 - Faits marquants

10h00-10h20 - Appel d’offres Soutien aux bibliothèques : bilan

10h20-10h35 - Colref

10h50-11h05 - CFP : nouvelles fonctionnalités

11h05-11h25 - CFP : nettoyage et fonctionnement

11h25-11h35 - Fichier Qualité Abes

11h35-12h25 - Echanges autour de points techniques

12h25-12h30 - Conclusion : on se retrouve les manches !

Après-midi :

14h00-14h20 - archivage pérenne, une introduction

14h20-15h30 - La BNF et l'archivage pérenne

16h00-17h00 -Transferthon

Faits marquants 2021

Voir diapos 4-7

Transferts : Le bilan montre un nombre de transferts important bien que les conditions de fonctionnement aient été perturbées par la COVID et un changement de marché au CNRS.

La procédure avec un formulaire en ligne et un bordereau à envoyer par mail était un peu complexe. Une nouvelle procédure est en place à partir de ce jour.

Un seul formulaire regroupe maintenant les 2 anciens formulaires (demande de transfert et bordereau d'envoi). Ce formulaire doit être rempli que l’on demande un financement ou pas. Les règles de financement par lePCMathrestent identiques : transferts financés pour les titres du corpus vers 1 bib PC (colref ou non). Un formulaire par bibliothèque destinataire, rassemblant l’ensemble des titres et des colis destinés à une bibliothèque. Ce formulaire doit être complété au moment de la mise en place du colis. En cas de financement par le PCMath, l’enlèvement des cartons est géré par le CT : il faut être précis pour les informations utiles au transporteur (horaires, jours d’ouverture, conditions d’accès aux locaux), et pour indiquer le poids et les dimensions des colis. Le traitement de la demande sera refusé par le CT si elle ne correspond pas aux critères **OU si les informations sont incomplètes dans le formulaire. Précision : en cas d’envoi de plus de 5 colis ou de plus de 10 titres vers une même bibliothèque, il faudra remplir deux formulaires.**

[L’appel d’offre pour les transferts est ouvert en permanence.](#)

Budget :

Voir diapos 8-9

Le CTLES précise que le mode de financement va changer à partir de 2023. Une information sera adressée sous peu aux animateurs de plan pour les informer des nouvelles dispositions.

Retro planning 2022 : un appel d'offre reliure devrait être ouvert cet été. Une phase 5 devrait être déclenchée également cet été. Le comité technique a adressé une liste de 1776 titres aux chercheurs (suite au travail de nettoyage du CFP) pour qu'ils poursuivent la sélection du corpus avec une phase 5 qui pourra être déclinée en plusieurs sous-phases si nécessaire.

Le comité de pilotage a validé le projet de marquer les dix ans du PCMath au CIRM en 2023 lors des journées du RNBM. Les dates ne sont pas encore fixées et dépendront des disponibilités du CIRM.

Appel d'offres Soutien aux bibliothèques : bilan

Voir diapos 12-13

- Denis Roura (Polytechnique - Palaiseau) indique les difficultés pour recruter sur une très courte durée. Il n'est pas non plus facile d'estimer précisément la durée nécessaire. Il reste 94 titres à corriger dans les catalogues.
- Isabelle Famechon (Lyon) : Le CDD a été chargé uniquement du récolement physique. Le recrutement a été simplifié car il y a eu un accord avec le SCD qui avait recruté un CDD pour un autre plan. Cela a permis de prolonger ce CDD : 2 mois en Bu Sciences + 1 mois en Bib Maths. Cependant les sources de financement étant différentes entre le CNRS et l'université cela a posé un problème. Il a été nécessaire de faire 2 contrats CDD : un université, un CNRS. C'est le laboratoire qui a servi de support pour le transfert financier.
- Marianne Sermain (ENS Paris) : elle a été recrutée sur un CDD PCMath et a travaillé pour ENS/Poitiers/Villetaneuse. Elle a été confrontée à de nombreuses questions sur la façon de saisir les données dans les catalogues. Elle a regretté de ne pas avoir pu se déplacer à Poitiers pour voir la bibliothèque et rencontrer sur place les bibliothécaires.
- Myriam André (Poitiers) : Le travail a été fait à partir d'un fichier, à distance par Marianne Sermain, mais cela a posé problème car les fichiers étaient incomplets et cela a donc nécessité des vérifications à faire en rayon. Il faut prendre en compte que cela représente une charge de travail pour la bibliothèque.

Conclusion : Pour le prochain appel d'offre une hypothèse serait de prévoir un CDD long avec des déplacements sur sites, ce qui impliquerait des frais d'hébergement et de déplacement. Il faut prévoir des CDD suffisamment longs pour garantir de pouvoir finaliser le travail.

Il est suggéré de créer une sorte de boîte à outil à l'intention des CDD avec toutes les informations nécessaires pour le traitement des collections dans le cadre du PCMath. En attendant vous pouvez vous référer au site de l'ABES (mémo sur les états de collection dans le cadre des PCP) et à la dernière newsletter pour les spécificités PCMath. Attention aux différences entre la saisie directe dans WinIBW et la saisie dans Colodus.

Un formulaire va être envoyé dans la semaine prochaine pour recensement des besoins :
[date limite le 8 avril](#)

Label Colref

Voir diapos 14-20

Attention engagement : quand on est pôle de conservation et surtout colref, il n'est pas obligatoire mais fortement conseillé d'essayer de maintenir l'abonnement papier.

Discussion sur la notion de Pôle associé :

Pour la charte du PCMath et pour les PCP gérés par le CTLes : Pôle Associé = membre du plan, ce qui signifie que la collection est incluse dans le PCP et que si on veut s'en défaire, l'engagement est pris de la proposer aux PC avant tout désherbage. Cela ne signifie pas obligatoirement qu'on est prêt à donner des fascicules pour combler les lacunes des PC. Les collections non positionnées (PC ou PA) sont hors du PCP.

Dans les faits, plusieurs interprétations ont cours au sein des établissements participant au PCMath. Certains s'attachent à la définition de la charte et du CTLES rappelée ci-dessus. D'autres n'ont mis en PA que les titres qu'ils accepteraient de donner. D'autres bibliothèques n'ont tout simplement pas eu le temps de positionner les titres en PA.

En conclusion, comme l'interprétation du positionnement PA est hétérogène, quand on cherche à combler une lacune, il est préférable de contacter toutes les bibliothèques conservant le titre et pas seulement les pôles associés (grâce au nouveau bouton « Contacter les correspondant.e.s » du CFP).

Précision : Pour tout changement de positionnement, veuillez adresser un mail au comité technique en précisant pour chaque titre le changement à effectuer.

Les nouvelles fonctionnalités du CFP

Voir diapos 21-22

Bouton magique pour contacter les correspondants : Pour permettre le bon fonctionnement de ce bouton, il est indispensable que chaque bibliothèque tienne Mathdoc informé de tout changement qui pourrait survenir. Pour l'instant c'est Mathdoc qui est en charge de maintenir l'annuaire mais cela est amené à changer. Pensez peut-être à créer une adresse générique pour ce type de demande qui pourra être maintenue même en cas de changement de personnel.

CFP : nettoyage et fonctionnement

Voir diapos spécifiques

- Un nettoyage de la base de données du CFP a été effectué. Nous sommes passés de 19380 notices à 5335 notices au 25/02. Parmi ces notices 531 ont été notées hors corpus math mais des bibliothèques sont localisées sur ces titres. Chaque bibliothèque a accès à la liste des titres qui la concerne dans son tableau de bord. Si vous souhaitez qu'un de ces titres soit maintenu dans le CFP vous devez adresser une demande motivée au comité technique qui fera suivre au comité éditorial du CFP.

- Une proposition a été adressée au comité stratégique de créer un comité éditorial pour veiller au contenu du CFP.

Sélection du corpus Phase 5

Voir diapo 24

1^{ère} liste cet été pour des candidatures jusqu'en octobre

Fichier Qualité Abes

Voir diapos 25-27

Conclusion :on se retrouve les manches !

Voir diapo 29

Le comité technique et le comité de pilotage vont travailler sur les séminaires et séries.
Attention : merci pour le moment de **suspendre les opérations de désherbage** sur ce type de document.

Résumé :

- **Appel d'offre reliure 1** (devis le 15 mars – réalisation avril)
- **Appel d'offre reliure 2** (probablement cet été)
- **Recensement des besoins de soutien aux bibliothèques** (date limite 8 avril)
- **Fichier qualité sudoc (Juin 2022)** et **toujours en cours fichier qualité CFP**
- **Transferts toujours en cours –nouvelle procédure**
- **Vérification des localisations de vos bibliothèques sous les titres susceptibles d'être retirés du CFP (30 juin)**

Articulation Papier / électronique

Voir diapos 31-35

Un constat : accélération de l'arrêt de la production et/ou de la diffusion des versions imprimées.

Une question : comment dans ce contexte continuer à garantir une conservation pérenne de l'information. Il faut distinguer l'accès pérenne de la conservation pérenne. Il est nécessaire de distinguer stockage, sauvegarde et archivage pérenne.

Il apparaît indispensable de décrypter les licences sur ce point pour déterminer les garanties apportées par les éditeurs. Mais il faut prendre en considération l'histoire complexe des périodiques dans le passé et dans le futur.

Il faut déterminer si une version électronique existe, pour quelles périodes, sous quels formats. Il est nécessaire de déterminer pour toute l'histoire du périodique, pour chaque séquence :

- Qui a les droits de propriété
- Qui a les droits de diffusion
- Quelles sont les procédures mises en place pour la conservation pérenne et l'accès pérenne.
- Quels sont les formats, normes et plateformes utilisées.
- Qu'en serait-il si un accident survenait (dépôt de bilan, crash, revente...)

Pour chaque séquence de l'histoire d'une revue, quelles sont les conditions d'accès post-abonnement, d'accès pérenne ou de récupération de l'archive pérenne.

Le PCMath doit identifier les versions numériques des titres du corpus et les lier aux informations de conservation et d'accès pérenne. L'objectif n'est pas d'analyser toutes les licences, mais plutôt d'identifier des bases qui rassemblent déjà ces informations. Nous allons entamer ce travail en collaboration avec le groupe licence. En parallèle le comité stratégique a lancé un travail sur la conservation pérenne électronique de manière plus large, notamment concernant la conservation des contenus créés et publiés en France. Accessoirement, le travail d'identification des versions électroniques des titres du corpus pourra permettre d'identifier des lacunes et de proposer une politique de numérisation.

Intervention BNF – Archivage et préservation numérique

Intervenant : Thomas Ledoux, informaticien à la DSI de la BNF

Informations générales

BNF sur 7 sites en France, 2200 ETP (site Richelieu va rouvrir bientôt)

Le Dépôt légal des imprimés existe depuis 1537 ; en 2006 le dépôt légal des sites web est mis en place et depuis décembre 2021 la BNF implémente le dépôt légal des documents numériques (actuellement uniquement la musique, mais ensuite les ebooks et l'audiovisuel). La conservation est au cœur des missions de la BNF et cela s'étend désormais au numérique avec des chantiers de numérisation et plus récemment aux documents né-numériques. A la différence des archivistes qui font une conservation sélective la BNF doit tout conserver.

Projet SPAR

Utilisation de la norme OAIS (conceptualisée par agences spatiales, les premiers à avoir perdu des données numériques) => le système est basé sur le modèle conceptuel OAIS (langage commun entre informaticiens et bibliothécaires) et utilise la syntaxe XML.

Le système est autonome et placé dans le système informatique global de la bibliothèque. Ce système est déconnecté d'internet ce qui assure la protection des données qui ne peuvent être modifiées ou détruites par une intrusion extérieure. Le système de sauvegarde pérenne de la BNF est exploité depuis mai 2010 et répliqué sur 2 sites (actuellement une recherche est en cours pour trouver un 3^{ème} site).

Actuellement le triplestore compte 7 milliards de triples (métadonnées élémentaires). Des migrations technologiques complètes doivent être faites de façon régulière (2 déjà réalisées)

Tous les paquets qui sont entrés suivent le même processus défini via différents modules : pré-versement – transformation versements – gestion des données, stockage et en parallèle

abstraction du stockage (indexation pour une préservation à long terme) – accès (ce n'est jamais l'archive qui est fournie mais une copie traitée/transférée).

Le système nécessite une grosse infrastructure matérielle : 2 sites identiques pour la préservation + 2 copies pour l'exploitation.

Modèle de paquets

Que préserve-t-on ?

La description, fichiers de différents formats (image, texte...). Le document complet est préservé et pas seulement une page pour garantir de conserver la structure physique.

Un paquet d'archive (AIP) = une archive TAR contenant :

- le manifeste des métadonnées au format METS : avec plusieurs schémas et grammaires xml pour bien pouvoir décrire tout le contenu
- l'ensemble des fichiers à préserver, avec une décomposition par niveaux : set, group, object, file, le set étant récursif

Élaboration politique de préservation à long terme

Mise en place de filières différenciées pour traiter l'hétérogénéité des données. Pour les documents du dépôt légal, la BNF a la maîtrise complète donc elle sélectionne les formats à conserver. La BNF gère aussi des contenus pour des tiers (tiers archivé), dans ce cas la BNF ne maîtrise pas les contenus qui ne lui appartiennent pas, les dépôts peuvent être très hétérogènes. Il est alors nécessaire de « contractualiser la filière » pour chaque type de document et pour chaque tiers.

"Contractualiser la filière" : contrat établi entre les parties qui permet de définir les étapes de traitement : définition des actions à chaque étape (même si le processus général est le même, le traitement sera différent selon le type de documents).

- filière de référence : description formats, filières et outils (toutes les grammaires de référence sont incluses dans le système lui-même, ceci afin que le système soit autonome)

Dépôt légal des né-numériques

Exemple : le son (jusqu'à très récemment laBNF numérisait les sources de différentes natures mais maintenant elle accepte le né-numérique).

Le système de préservation est utilisé comme un véritable magasin avec en parallèle un signalement. Comme pour les autres documents ce n'est jamais l'archive source qui est fournie mais un artefact de l'objet qui est constitué au moment de la restitution.

Coopération sur les bibliothèques numériques

La « *Marque blanche* » : c'est une collaboration avec d'autres bibliothèques, notamment patrimoniales : partage de numérisation, intégration de fonds numérisés, préservation des données numériques, diffusion par Gallica (diffusion via Gallica d'autres bibliothèques numériques). La BNF est co-proprétaire des données et les conserve. La BNF met à disposition une instance personnalisée de Gallica : actuellement 16 conventions de ce type. La BNF souhaite se limiter à 4 conventions de « *Marque blanche* » par an.

Tiers archivé : BNF n'est pas propriétaire des données ; elle est prestataire de services.

- ex : bibliothèque numérique de la statistique publique : les documents versés sont nativement électroniques : l'INSEE n'a pas voulu mettre les données au format csv, les éléments d'accompagnement sont restés au format propriétaire Excel et ne sont pas donc pas pérennes, alors que l'article en pdf est pérenne.

Questions

- Sur Gallica : pour certains titres (exemple : CRAS) il y a des manques : peut-on compléter les manquants ? Oui. Si pas déjà numérisé : contacter le département collaboration de la BNF pour lancer une numérisation.
- Quelles sont les spécifications de la numérisation ? Numérisation en couleurs, jpeg2000, 400 dpi.
- Protocole IIF : pour images, élargi à d'autres formats : permet d'avoir une interface unique pour consulter des morceaux de collections sur plusieurs bibliothèques.
- Quelle collaboration avec le CINES ? Il y a bien une collaboration notamment au niveau d'un "réseau" PAIN. Nous avons le même type d'architecture, de technologie, de périmètres.
- Quand débutera le dépôt légal né-numérique des périodiques ? 2024-2025. Cela concernera les périodiques français. L'accès est limité en interne dans les locaux de la BNF sauf accord de l'éditeur mais cela semble complexe.
- Quel lien avec les projets d'archivage pérenne comme Clockss, Portico ? des échanges et un dialogue. Le risque principal pour la préservation numérique est celui des moyens financiers nécessaires sur le long terme.

Transferthon

Voir diapo 37