

JOUR 1 et 2: Module Technique

Lundi 12 et mardi 13 décembre formation technique

Journées assurées par Didier BELLAMY

- Formateur
- Migrateur pour PMB services

-Horaires : 9h30/12h30 - 13h30/17h30 (*je vais demander 10h-18h pour la première journée*)

- Pré-requis : connaissances Apache, Mysql, base de données

- Contenu : formation à l'installation et la MAJ de PMB, outils d'aide à la migration de données

- Contenu détaillé (jour 1 et 2) :

- Présentation rapide / prise en main du logiciel.
- Installation générique de PMB
 - pré-requis
 - démo sur Debian
 - questions / réponses.
- Mise en place d'un système de moissonnage de bases externes avec OAI-PMH
 - Protocole
 - Client / Serveur dans PMB
- Principes d'import dans PMB
 - Format pivot d'import Unimarc
 - Utilitaires de conversion vers Unimarc intégrés à PMB.
- Réalisation d'un import spécifique depuis un fichier au format .csv
- Réalisation d'un import spécifique depuis un fichier au format xml.

Tour de table, bagage informatique de chacun?

- O.Gavotto: développeur,bdd, à installé PMB sur son serveur.
- G.Debarros: géomaticien, bdd
- Amandine Montaigne: développement web, a une ancienne version de PMB, quelques développement de code
- B.Faucheux: développeur, a ses données dans Access 1997. A tenté un import dans PMB avec des scripts unimarc et modules d'import. Quelques difficultés. A réussi à importer des notices mais quelques problèmes sur les imports.
- E.Hamdi: responsable pôle informatique.

Historique, présentation logiciel

- Logiciel créé par un bibliothécaire
- Langage php, bdd mysql.
- Logiciel entièrement web utilisable à travers un navigateur
- 2 parties
 - gestion backoffice (documentalistes)
 - OPAC pour la consultation et éventuellement de la réservation (parfois il y a du dépôt de document à travers un catalogue public) ex: *institut agronomique méditerranée de*

Montpellier.

- Possibilité d'ajout d'un portail=CMS avec du contenu éditorial (rubriques, articles) : *Un exemple le site www.sigb.net*

Les onglets du backoffice:

- onglets « circulation » gestion des prêts,
- onglet « catalogue » (documents bibliographique) possibilité de créer des notices et périodiques,
- onglet « autorités » possibilité de modifier le nom des auteurs pour toutes les notices à partir d'une liste.
- onglet « éditions »: imprimer des listes (prêts en court, retard, état de collection, état de circulation). Les états sont personnalisables ou paramétrables: requêtes SQL réalisées dans l'interface d'administration.
- onglet « DSI »:requête préparées "exemple orchidées val de Loire" et ensuite mailing liste pour que les gens reçoivent les nouveautés dans ce domaine.
- onglet «administration »:paramétrage de l'interface

Vocabulaire

- Auteurs: conception ouvrage
- Editeur (celui qui publie) des collections et suites
- Séries: regroupement d'une suite d'ouvrages (ex: bande dessinée)
- Index décimal: un indice décimal qui permet de retrouver un ouvrage (ex: géographie).
- Catégorie: système d'indexation
- Monographie= document autonome
- Périodique= exemple le canard enchaîné, un bulletin toutes les semaines, il comporte des articles= dépouillements
- Exemplaires physiques dans les deux cas (monographie ou bulletin).
- **ISBN** identifiant de monographie (livre=booknumber)
- **ISSN** identifiant de bulletin (serialnumber)

Onglet administration:

- exemplaire= document physique
 - supports= durée de prêt, localisation (local par exemple espace jeunesse dans une bibliothèque) + section (rayonnages par thème ex: romans, livres enfant).
 - Statut= confidentialité des documents (exclu du prêt), code statistique (savoir si c'est pour des enfants par exemple), informations personnalisables (champ à ajouter dans le logiciel)
- notice biblio= pas d'existence physique

Les statuts des documents permettent de savoir si les notices, les exemplaires, les documents numériques sont visualisables par l'utilisateur.

Les origines d'un document permettent de savoir d'où vient la notice, si elle est interne, ou appartient au catalogue de la BNF.

Attention il faut qu'entre les notices et les exemplaires, les champs personnalisables soient cohérents

entre eux.

Par ailleurs on peut préciser si chaque champ est exportable (si un champ est vraiment propre à un catalogue on ne permet pas de l'exporter).

Création d'une nouvelle notice

Dans l'onglet document /nouvelle notice

La notice a plusieurs champs standards dans le logiciel: besoin d'un titre, d'un type de document... La liste proposée est issue de la norme Unimarc (en dur dans les fichier). Par exemple pour le type de documents numériques = texte imprimé, texte manuscrit.... On peut ajouter des types de documents mais cela ne sera pas forcément interopérable.

Chaque notice à besoin à minima (pour la MONOGRAPHIE)

- d'un titre propre = titre principal de l'ouvrage
- un auteur principal
- autres auteurs
- des auteurs secondaires (exemple illustrateur d'ouvrage, photographe)
- éditeur
- collections
- sous-collection
- année de publication
- ISBN
- collation= informations sur les caractéristiques physiques des exemplaires (nombre de pages, d'éléments, autres caractéristiques, format, prix, matériel d'accompagnement)
- notes (générales, de contenu, résumé)
- indexation (catégories d'un thésaurus, nidxation décimale issue d'un plan de classement, langue de la publication)
- liens: URL associés
- Champs personnalisés
- notices liées (exemple monographie en français et même exemplaire en anglais)
- informations de gestion (langue de la notice)
- statut (visible dans opac ou non?)
- commentaire
- vignette de notice.
- droit accès emprunteur

Pour les périodiques:

- On crée une notice pour le périodique et pour chaque publication possibilité d'ajouter un numéro/bulletin (avec numéro, date publication, éventuellement un titre propre). Il est aussi possible d'ajouter un dépouillement (un article spécifique).

Pour les tirés à part (proposition de D. Bellamy) :

- création d'un type de notice "tiré à part". On crée une notice. Ensuite on fait un lien dans l'article avec un type de relation "contenu dans" par exemple.

Gestion des types de documents

Ajouter un fichier doctype_subst.xml

Catalogage

Il est nécessaire avant toute chose de se définir un cadre:

- façon de cataloguer
- ce que l'on saisit (exemple façon d'écrire les auteurs)
- types de liens entre les différentes notices

Moissonnage

Le moissonnage n'est pas une recherche en direct. Il y a import des informations dans une base locale et ensuite on cherche en local.

A l'INRA il existe un catalogue PMB centralisé (<https://belinra.inra.fr>): possibilité de chercher dans l'ensemble des centres de documentation qui existent et 2 centres à part traités en tant que ressource externe (moissonné).

*Question : * si je moissonne différents notice il y a-t-il un identifiant propre dans le catalogue FBCN ou est-ce le permalink qui sert d'identifiant ?*

- comment on organise la recherche fédérée? OAI qui ne supporte pas les choses qui ne sont pas dans la norme.

Il existe une maison d'édition de littérature dans les sciences humaines et sociales : possibilité moissonner notices distantes puis et synchroniser l'import des notices dans une bdd locale.

Installation du logiciel

- Il existe des notices pour centos, ubuntu, debian, windows.
- Documentation pré-requis techniques du serveur : [prerequis_techniques_linux_20160203.pdf](#).
- Possible de séparer les composants sur des serveurs différents: base de donnée de l'application et interface PMB/OPAC.
- Pour le serveur mysql plus de RAM nécessaire que pour l'OPAC
- Pour env. 100 k notices :
 - Config. mini : 4 CPU + 8 Go de RAM
 - Espace disque : 2 Go + 10 Go BDD + Espace pour les docs numériques
- Une version majeure par an qui suit les versions de Php (PMB 5, PMB 7).
- 2 URL à prévoir 1 public (= OPAC) + 1 gestion
 - doc.fcbn.fr ⇒ OPAC

- doc.fcbn.fr/opac_css⇒ Admin PMB
- Configurer les virtualhosts d'Apache

Librairies

- **php_yaz** est nécessaire pour interroger les serveurs de notice en z39-50 indispensable.
- **php-ldap** et **php-cas** pour l'authentification, de base le système d'authentification est en base de données.

Fichier de conf

- **max_input_var** indispensable sinon les notices ne seront pas postées totalement
- Installation en **utf8** de base. Si on a des install en latin1 PMB peut fournir un script de conversion.

Attention à bien sauvegarder les répertoires qui contiennent les documents physiques.

Ouverture vers l'extérieure

Le protocole pour communiquer avec les serveurs externes: z39-50 ne supporte pas les proxy. Il faut faire des ouvertures spéciales.

Remarque: Il est possible de récupérer des images des notices Amazon à partir des numéros ISBN.

Chiffrage des mots de passe

La fondation Mozilla fourni (lets_encrypt, les certificats sont valable 3 mois et renouvelles automatiquement) a fait une autorité de certification de chiffrage des mots de passes qui est gratuite. Document signé certifié avec une clef publique de cryptage délivrée par une autorité. Cette autorité a validé le fait qu'on possède une clef privée et qu'on chiffre le flux. D'un côté j'ai un mot de passe qui permet de crypter les données et qui est privé et un mot de passe qui permet de décrypter ces mots de passe et qui est public.

Installation de la base de données

- Installation de Mariadb puis lancer la requête suivante pour s'assurer que la base a été bien installée:

```
mysql show global variable like '%engin%'
```

Pour le moment le serveur de stockage est innoDB mais PMB passe à myISAM (système d'indexation full text qui n'existait pas pour innodb et qui a été ajouté dans la dernière version).

- Pour modifier les paramètres mysql il peut être intéressant de rajouter un fichier **pmb_my.cnf** (facile sous debian, mais sous ubuntu 2 fichiers).

→ demander le fichier exemple au formateur

- Si latin1 il faut mettre latin_swedish_si
- Configuration du fichier php: attention à l'upload max des fichiers qui est à 64 Mb pour le moment (s'il y a des photos à télécharger ou des ouvrages faisant plus de 64 megas.

Mise à jour

- Ecraser les fichiers par la nouvelle version - Lancer l'interface d'admin

Droits sur les fichiers

```
chmod -R root www-data html/ chmod 750 -R html/
```

Droits en écriture

```
chmod g+ws pmb_documents_numeriques chmod g+ws pmb/temp/ chmod g+ws  
pmb_documents_numeriques chmod g+ws pmb/opac_css/temp/ chmod g+ws  
pmb/admin/backup/backup/ chmod g+ws pmb/admin/import/
```

Paramètres

gestion > admin > Outils > Parametres > Parametres généraux [ESC + x]

Format d'export des notices

- Les formats d'export sont le csv, l'xml, le text
- Le format pmb xml unimarc est un format pivot entre le format csv, xml, text avec le
- Format marc ou unimarc qui est le format d'import dans pmb qui est un fichier assez illisible.
- On peut importer soit uniquement les notices, soit les notices avec exemplaire.

OAI/PMH

- protocole très simple utilisé pour la récupération d'article, voir <https://www.openarchives.org/pmh/>

Dans les connecteurs sortant → choisir le connecteur qui exporte des données OAI

- xml labellisé à peu près équivalent à l'unimarc en plus lisible
- dublin core (on peut ajouter une feuille de transformation au Dublin core)

Faire un set de notices

pour les connecteurs sortant: on peut rajouter un statut de notice "notice pour export OAI" et choisir que seules les notices avec ce statut sont exportables dans les "sets pour connecteurs sortant". Il existe une gestion des doublons sur l'identifiant de notice et l'ISBN (le premier entrant est choisi).

http://multimedia.bnf.fr/unimarc_b_trad/Bsection5-Label_notice-6-2010.pdf

Unimarc

Dans `\pmb\includes\marc_tables\fr_FR` :

- 'doctype.xml' ⇒ Types de documents
- 'country.xml' ⇒ liste des codes pays en UNIMARC

Dossier `doc_conversion`:

- Document sur les label de notice de la norme unimarc.

Avec l' unimarc, la notion de bulletin n'existe pas.

Le label de notice est un code de 24 caractères avec un type de document, un niveau bibliographique et un niveau hiérarchique

Ensuite il y a des informations réparties en champs (zone) et en sous champs. Le champs est une zone numérique sur 3 chiffres et le sous-champ est '\$' + une lettre ou un chiffre.

exemple:

- 700 auteur principal; 701 auteur autres;702 auteurs secondaires
- 200\$a pour le titre propre
- 200\$d pour le titre parallèle
- 700\$a pour le nom
- 700\$b pour le prénom
- 101\$a pour la langue originale
- 101\$b pour la langue de l'exemplaire (ou de la notice, à vérifier)

Dans PMB tout n'est pas géré car la norme comprend 900 champs et beaucoup de sous-champs.

Le dossier `xampp\xampp\htdocs\pmb\doc` contient le fichier `import_unimarc.txt` qui décrit tous les champs de la norme.

Dans le même dossier le fichier `import_unimarc_lien.txt` permet de voir les liens pour l'import ou export des notices.

Il existe un outil pour faire de la transformation de csv à xml. Cet outil est intégré à PMB. Il sera présenté en 2ème partie de formation. Il se trouve dans le répertoire `convert_import` dans PMB

Les fichiers subst

Ils permettent de modifier les configurations de PMB sans modifier les fichiers sources.

Cela concerne les types de documents, les messages dans l'interface, les relations entre notices, le catalogue des fonctions de conversion.

Exemple:

- `pmb/include/messages`
- le fichier `doctype.xml` qui permet de modifier les types de documents
- `relation_typedown.xml` "tiré à part"
- `relation_typeup.xml` "a pour tiré à part"

Attention ! Il faut avoir les mêmes documents côté OPAC et PMB pour être cohérent.

Il faut donc que les fichiers subst soient recopiés côté OPAC sinon si par exemple j'ai créé un type "rapport" côté PMB et qu'il n'existe pas côté OPAC, le document n'apparaît pas dans la recherche: c'est bloquant.

Conversion de fichier

Dossier /var/www/html/pmb_cbn_corse/pmb/opac_css/admin/convert/imports:

Dans le fichier Catalog.xml, on voit les formats qu'il est possible d'importer dans PMB (les type avec "yes")

1. On fait un catalog_subst.xml on ne met que les lignes qui nous intéressent.
2. On copie le dossier pour la conversion **text2pmbxml** en **text2pmbxml_cbn**
3. Aller dans PMB Gestion :Admin / ConversionsExports / Conversion de fichier externe. La liste 'Type de conversion' doit afficher le contenu du fichier XML 'catalog_subst.xml'
4. Dans le fichier params.xml, on décrit le type de fichier qu'il va prendre en entrée (ex: text pour les csv) avec le séparateur pipe, le header etc...
5. Puis on attribue à chaque colonne une correspondance dans les label xml.
6. Première conversion fichier
7. CSV CBN→pmb-XML MARC
8. Puis récupération du xml produit
9. Deuxième conversions pmb-XML MARC→MARC ISO2709

Remarque :

- les "zones" (label des balises) du PMB unimarc sont décrits dans C:\Users\anais\Desktop\PMB\xampp\xampp\htdocs\pmb\doc\import_unimarc.txt et import_unimarc_lien.txt
- on peut ajouter un doctype

<dt>b</dt> -pour texte manuscrit

La description des doctype.xml se trouve dans ce dossier :

C:\Users\anais\Desktop\PMB\xampp\xampp\htdocs\pmb\includes\marc_tables\fr_FR

- Attention si on décrit la langue du document zone 101 c il faut que cette langue soit définie dans le fichier lang.xml (c'est un code langue)
- on peut rajouter un niveau bibliographique
 - <bl>s</bl> c'est serial (=revue) pas de soucis si pas d'enfant
 - <bl>b</bl> c'est bulletin pose soucis si pas de parent
 - <bl>m</bl> c'est la monographie
 - <bl>a</bl> c'est un article pose soucis si pas de parent (la revue)
- on peut ajouter un niveau hiérarchique
 - <hl>0</hl>

Pareil il faut être attentif à l'existence d'un parent pour les articles, les bulletins

Si on veut faire le text to iso2709 on va dans le fichier params.xml du dossier pmbxml2marciso et on copie colle la balise <step>

Voir dans PMB_explication_conversion c'est expliqué.

reprendre le CSV et le params.xml de l'exemple du fichier de Corse.
Didier Bellamy doit aussi nous fournir les types de champs perso.

Exemple d'import

Définir le mapping entre les champs du CSV et les champs de PMB

Pour les champs les plus communs

- * Auteur :
 - * 700 = Si 1 Auteur unique
 - * 701 = Si plusieurs auteurs
 - * 702 = Auteurs secondaires (par ex traducteur ...)
 - * \$a = nom
 - * \$b = prénom
- * Date publication :
 - * 210 \$d = Date de pluli
- * Titre :
 - * 200 \$a = Titre
 - * 200 \$i = Titre de série (si monographie)
- * Numéro :
 - * 200 \$h = Numéro dans la série (si monographie)
- * Page :
 - * 215 \$a = Num de page
- * Mots clefs (utilisation d'un thésaurus) :
 - * 610 \$a = Mots clefs libre dans la notice
 - * 676 = Si plan de classement
 - * 606 \$a = Si associés à un thésaurus
- * Site de la zone d'étude :
 - * 610 \$a = Mots clefs libre dans la notice
 - * 676 = Si plan de classement
 - * 606 \$a = Si associés à un thésaurus
- * Numéro ISBN :
 - * 010 \$a = code
- * ID bd source (= nouveau champ non par défaut dans PMB)
- * Champs perso de PMB :
 - * 900 \$a = <valeur>
 - * 900 \$n = <nom>
 - * 900 \$l = <libelle>
- * Cote & licalisation dbiblio :
 - * 995 \$a = localisation
 - * 995 \$f = Num de code barre (ou d'exemplaire) (*)
 - * 995 \$k = cote (*)
 - * 995 \$q = public visé
 - * 995 \$r = Type de document (support) (*)
 - * 995 \$u = Commentaire non bloquant

(*) = champs obligatoires

Thésaurus

- Partager les thésaurus au sein du réseau a travers le wiki (ou alors mettre en place un ftp pour la doc avec les différents documents)
- Pour refaire le lien entre les documents physique et la notice il faut créer une fonction qui permet de faire le lien entre le xml et la table de PMB qui stocke les répertoires d'upload (avec une partie de l'identifiant)→upload_repertoire

Conversion XSLT

aix2pmbxml permet de traiter des sorties de logiciel superdoc. A l'origine il s'agit d'une sortie xml de Superdoc qui va avoir des tag xml superdoc (particulier).

Type d'entrée=xml

rotelement: inmrecordset

Etapes de transformation: des étapes qui s'enchainent.

La feuille xsl permet de compléter la transformation si certaines choses ne sont pas passées dans le "mapping" traditionnel. C'est du xml qui permet en plus des transformations avec des fonctions. Cela récupère des balises dans le fichier d'origine et les transforme en autre choses. Cela permet de mettre des conditions pour le mapping (cela prend en compte les valeurs contenues dans les champs).

Par exemple pour mes mots clefs, si les séparateurs sont du ; ou de la , on expose le champs en fonction soit de la virgule soit du point virgule...

From:

<https://wiki.fcbn.fr/> - Wiki - FCBN

Permanent link:

https://wiki.fcbn.fr/doku.php?id=gt:formation_pmb_2016_12_12&rev=1500472900

Last update: **2022/03/07 12:00**

