

05.03 - récupération de l'historique

Base de données des indicateurs esup : récupération de l'existant

- [Base de données des indicateurs esup : récupération de l'existant](#)
- [Généralités](#)
- [Initialisation](#)
- [Etat des lieux](#)
- [Script final d'initialisation et de récupération des archives](#)

Généralités

De nombreux établissements utilisateurs d'esup-portail utilisent la chaîne de traitement des indicateurs distribuée avec le package.

Ce traitement produit 2 fichiers xml mensuels qui agrègent les informations statistiques esup-portail du mois concerné.

Le problème qui peut se poser : il est probable que les fichiers de plusieurs mois précédents ne soient plus en conformité exacte avec les fichiers de paramètres qui ont permis de les générer : types.map et services.map .

Il a pu y avoir entre temps de nouveaux services ou types de population, ou modification de libellé (2ème champ de ces 2 fichiers de mapping).

Est décrite ici la méthode utilisée dans notre université pour permettre un import facile à partir des anciens fichiers xml.

Initialisation

La base est initialisée avec la dernière version des fichiers types.map et services.map.

Etat des lieux

L'utilitaire DBmajIndics.pl peut être utilisé pour recharger des fichiers xml de sessions et services de n'importe quel mois antérieur : avant d'enrichir la base, il supprime les informations statistiques du mois concerné.

Il contrôle la cohérence des types de population et de services entre la base d'une part, et les fichiers xml du mois d'autre part.

Il suffit donc de tenter l'importation des fichiers antérieurs, et de déterminer par sondage à quelle période il y a eu un changement.

On est remonté à l'envers, pour déterminer les différents changements par rapport à la situation actuelle.

Dans notre cas :

- Avant mai 2007, pour le type de population ' type.map ', la description des populations était différente, mais pas le ciblage. On a donc créé un fichier types0505.map qui correspondait à la situation antérieure. On a ensuite continué à remonter le temps, toujours par sondage
- A partir de janvier 2006, plusieurs services ont été ajoutés aux traitements statistiques. On a créé un fichier services0601.map correspondant à la situation antérieure.
- Ce fichier convenait au premier mois d'archive de stats, soit mai 2005.

Script final d'initialisation et de récupération des archives

Suite à cette investigation, nous avons pu monter un script d'initialisation de la base de données, et de récupération des archives.

Le voici :

```
#!/bin/sh ./DBinit.pl -etablissement "universite nancy 2" -codetab "n2" -all

MOIS="2005-05 2005-06 2005-07 2005-08 2005-09 2005-10 2005-11 2005-12"
for i in $MOIS do
    echo "traitement du mois $i"
    ./DBmajIndics.pl -codetab "n2" -ficmaptypes types0505.map -ficmapservices services0601.map -ficsessions xml
/stats${i}.sess.xml -ficservices xml/stats${i}.serv.xml -nocontrol_service
done

MOIS="2006-01 2006-02 2006-03 2006-04 2006-05 2006-06 2006-07 2006-08 2006-09 2006-10 2006-11 2006-12 2007-01
2007-02 2007-03 2007-04"
for i in $MOIS do
    echo "traitement du mois $i"
    ./DBmajIndics.pl -codetab "n2" -ficmaptypes types0505.map -ficsessions xml/stats${i}.sess.xml -ficservices xml
/stats${i}.serv.xml
done

MOIS="2007-05 2007-06 2007-07 2007-08 2007-09 2007-10 2007-11 2007-12 2008-01"
for i in $MOIS do
    echo "traitement du mois $i"
    ./DBmajIndics.pl -codetab "n2" -ficsessions xml/stats${i}.sess.xml -ficservices xml/stats${i}.serv.xml
done
```

Durée d'exécution, sur une petite machine, pour la récupération des infos statistiques sur près de 3 ans : moins de 5 mn, sans aucune fatigue du serveur.

Quelques remarques :

- Ceci est un exemple. Il y a des cas particuliers dans chaque établissement, mais ça montre une manière de procéder.
- Sur la première boucle : l'option "-nocontrol_service" : sans cette option, l'utilitaire refuse de s'exécuter lors du contrôle de cohérence : tous les services déclarés dans la base ne sont pas présents dans la base de données