

Extrait du Le site de Beuz

<https://www.beuselinck.fr/article225.html>

Conservation des archives de messagerie

- Linux -



Date de mise en ligne : mercredi 10 octobre 2012

Copyright © Le site de Beuz - Tous droits réservés

L'usage de la messagerie institutionnelle aboutit à se poser la question de la conservation des archives de messagerie. Une fois la réponse obtenue, se pose la seconde question : oui, mais comment ? C'est à celle-là que je tente de répondre.

L'état des lieux

On considérera dans cet exemple que les messages à archiver sont stockés soit sur un serveur accessible par imap ou pop3 (peu importe que ce soit exchange, postfix ou exim...) , soit dans des fichiers PST , MBOX ou Maildir.

Les ingrédients

Nous utiliserons plusieurs outils :

- ▶ [Mozilla Thunderbird](#)
- ▶ le plugin "[Import Export Tools](#)" pour Thunderbird
- ▶ [Mhonarc](#)
- ▶ [readpst](#)

Leur installation sera abordé au point 4 de ce sujet.

Le principe

Nous allons stocker nos messages dans un fichier au format [MBOX](#) puis le transformer en pages HTML. Pourquoi ce choix ?

Les pages HTML répondront aux besoins de consultation d'archives courantes sans avoir besoin d'un client de messagerie. Il suffit d'un navigateur web. De cette manière, nous essayons d'atteindre universalité et pérennité d'accès à notre archive [\[1\]](#).

Le fichier MBOX assurera la conservation des informations probatoires (notamment les en-têtes de messages) avec la possibilité d'y appliquer un [hashage](#). Ici, nous essayons d'obtenir conservation et intégrité de l'archive [\[2\]](#).

Conserver sous forme de PST impose de continuer à utiliser Outlook dans les prochaines années et que ce format soit maintenu par son éditeur. Au vu de la progression de la qualité des webmail (OWA ou Zimbra par exemple), je ne parie pas là-dessus. MBOX tout seul nécessitera aussi un client lourd (même si on a une plus grande diversité d'outils). Les format Maildir, eml, msg posent très vite le problème de la quantité de fichiers (pour la technique de [hashage](#), ça ne va pas être simple) et celle du client lourd. Pour toutes ces raisons, le couple (pages HTML+MBOX hashé) m'a semblé répondre le mieux aux différents besoins. A vous de conserver l'un ou/et l'autre ou adapter la méthode à vos propres choix.

Installation des outils

Commençons par le plus simple : "Thunderbird" et son plugin "ImportExport Tool". Pour installer thunderbird, il suffit d'aller sur le site officiel : <http://www.mozilla.org/fr/thunderbird/>. Les utilisateurs de linux pourront faire confiance au paquet RPM ou DEB qui est proposé avec leur distribution.

Une fois Thunderbird installé, il faut encore le configurer. Au premier lancement, on vous propose de créer un compte ou d'ouvrir une boîte mail existante. On ouvrira bien évidemment la boîte email ou les boîtes emails que l'on veut archiver : `contact@votredomaine.com`, `sav@votredomaine.com` etc....

Lancez ensuite l'entrée "modules complémentaires" que vous trouverez dans le menu OUTILS, vous verrez plusieurs choses s'afficher mais ce qui nous intéresse est la zone de recherche en haut à droite qui va nous permettre de rechercher et installer "Import export tools".

Passons ensuite à Mhonarc. Vous irez sur le [site web Mhonarc](#) télécharger la dernière version (2.6.18 au moment où j'écris ces lignes).

MHonArc est écrit en langage perl, un langage non compilé (de type « script ») très répandu ce qui en facilite d'autant plus la lecture, notamment pour en vérifier l'inocuité.

Les utilisateurs devront donc préalablement disposer de Perl sur leur machine. Perl est inclus dans les systèmes linux et Mac. Il est installable assez facilement sous windows. Lire la [page de téléchargement de perl.org](#) ou sur [telecharger.com](#) si vous préférez.

Readpst est téléchargeable [sur le site de l'auteur](#) si vous utilisez Linux et il en existe un portage sous windows téléchargeable sur <http://www.marklyon.org/gmail/gmailapps.htm>

la mise en oeuvre

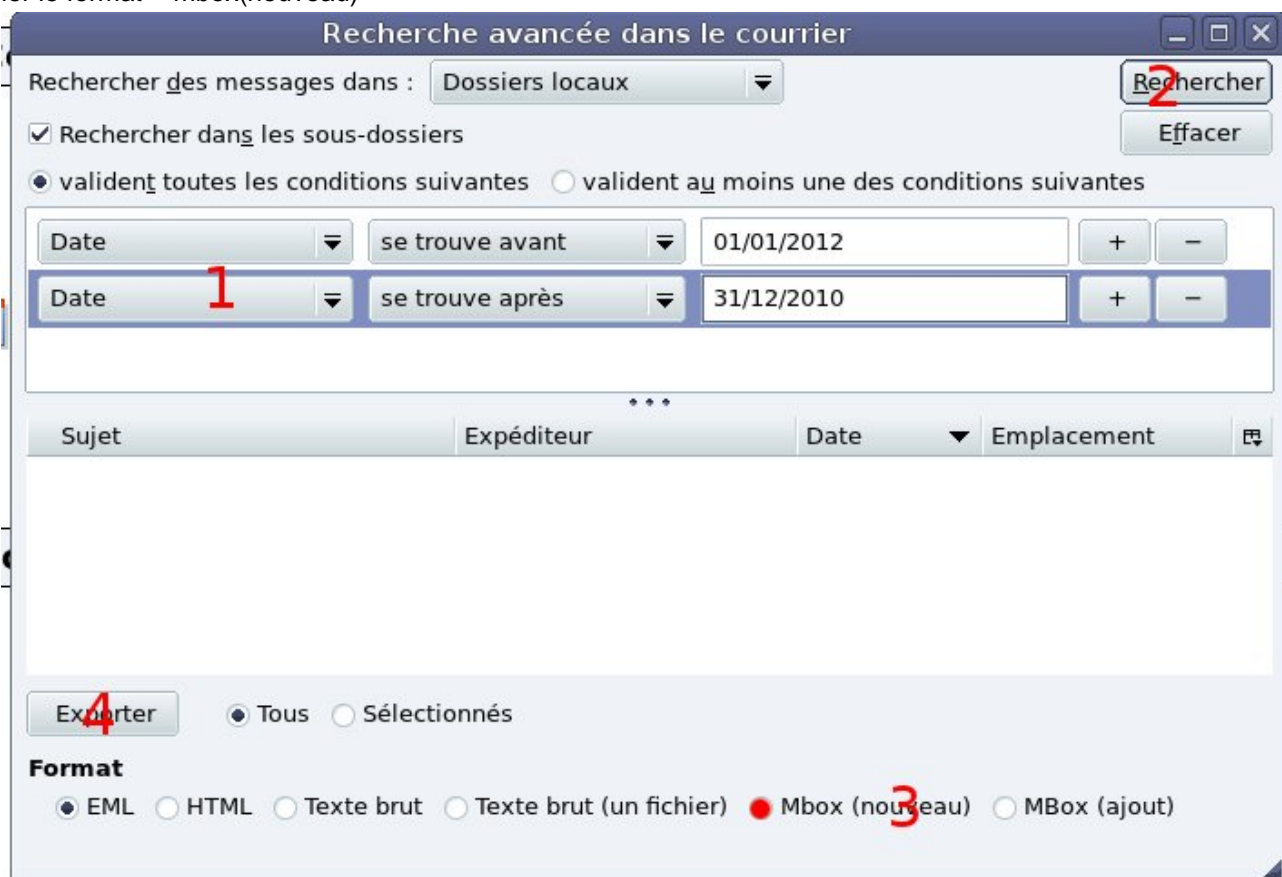
Nous allons commencer par les messages accessibles par Thunderbird. C'est surtout le cas des messageries qui utilisent le protocole [IMAP](#) ou le protocole [POP3](#) mais aussi les espaces de stockage au format Maildir. Depuis que vous avez installé le module IMPORTEXPORT TOOLS, il vous est possible en faisant un clic droit sur un de vos comptes de messagerie, de sélectionner « Import/Export »>Rechercher et Exporter »

Conservation des archives de messagerie



Dans la fenêtre suivante, l'export se fait en 4 étapes :

1. définir les conditions (dans la copie d'écran ci-dessous, j'ai choisi l'année 2011)
2. effectuer la recherche
3. sélectionner le format « mbox(nouveau) »
4. exporter



Pour les fichiers PST générés par Outlook, vous devrez utiliser la commande readpst de façon très simple :

```
readpst nomdufichier.pst
```

Cette commande exporte le fichier du dossier courant répondant au nom de *nomdufichier.pst*. Par défaut, le résultat est au format MBOX, se situe dans le dossier courant et le nom du fichier est le nom du dossier (par défaut outlook attribue « dossier personnel ») sans l'extension ".pst".

Vous avez maintenant toutes vos archives de messagerie au format mbox. C'est le bon moment pour effectuer une empreinte numérique par la technique md5, sha256 ou sha512. Voilà ce qu'on a :

```
ls -ogh *mbox
-rw-rw-r-- 1 363M févr. 16 22:26 2009_mbox
-rw-rw-r-- 1 1,3G févr. 16 22:27 2010_mbox
-rw-rw-r-- 1 1,9G févr. 16 22:27 2011_mbox
```

Il suffit alors de faire au choix :

```
md5sum *mbox >>mbox.md5
sha512 *mbox >>mbox.sha512
```

Cela crée un fichier mbox.md5 ou mbox.sha512 qui contient l'empreinte numérique :

```
cat mbox.sha512
a31ec4961167a2fc8b100f028dfc39952a9193e1236a29bb5bc00ca15563bffdde98c8f0a8e61cbf539349215841dfd3fae913bf5e0
dd8f241fa22273645ed1 2009_mbox
3cb8f9411c62875dd23ad9028810b1adb96aee2bd5312d93e15cbe0e27bdbe09d170dd6903f92639e2a453ab321e226af0a36b97e814
f8f301c741c231b75aec 2010_mbox
ea15b33bccfb745b6145a25b8bb40b2b538e8b6808a18f784e042c3866faa110fcc768471ee1418c33c492e22ec1d4d394823318ac2
828f61724062c468f7e0 2011_mbox
```

Enfin , nous utilisons mhonarc pour créer nos archives html :

```
mhonarc Nom_du_fichier_mbox -rcfile common_fr.mrc -outdir dossier_de_sortie
```

Dans la ligne ci-dessus, vous remplacez Nom_du_fichier_mbox par votre fichier mbox et le dossier_de_sortie par votre propre choix (c :\archives ou /home/votrepseudo/archives) pourvu que ce dossier existe.

L'option -rcfile vous permet d'utiliser le fichier common_fr.mrc (en téléchargement en bas de ce article) lequel va permettre que les fichiers HTML générés par Mhonarc soient en français. Si vous préférez l'anglais, ne mettez pas "-rcfile common_fr.mrc".

Attention, je vous déconseille de lancer mhonarc sans mettre l'option "-outdir" suivie du chemin. Le nombre de fichiers générés est très important donc remisez-les dans un dossier créé pour l'occasion (Sinon vos fichiers mbox seront noyés dans plein d'autres fichiers).

Dans ce nouveau dossier, vous trouverez une série de fichiers à l'extension .bin, .jpg, .pdf, .doc, .xls... Ce sont les pièces jointes que l'on retrouve dans les messages. Ne les détruisez pas.

Les fichiers à l'extension .html correspondent chacun à un message archivé sauf maillist.html et threads.html qui sont le catalogue complet de vos messages.

Pour lire vos messages archivés, il vous suffit maintenant de lancer maillits.html dans votre navigateur préféré. Vous pourrez ensuite cliquer sur un message, rechercher par date, par fil de discussion, par auteur ou titre. En cliquant sur un titre, vous verrez le contenu du message et les pièces jointes que vous pourrez également visualiser par un clic de souris.

Conclusion

Vous avez maintenant stocké vos archives de messagerie sous deux formats différents : mbox et html. Vous avez

Conservation des archives de messagerie

calculé les empreintes numériques de vos fichiers mbox. Il reste à stocker tout ça sur un support longue durée (DVD par exemple), lui-même rangé dans le coffre-fort. Les améliorations possibles sont :

- ▶ [la mise à disposition pour votre communauté de travail de ces archives au travers d'un serveur web *intranet* à accès sécurisé](#)
- ▶ [l'ajout d'un moteur de recherche interne](#)

[1] http://fr.wikipedia.org/wiki/NF_Z_42-013

[2] http://fr.wikipedia.org/wiki/NF_Z_42-013