



# ReSIP



29 septembre 2020



# SOMMAIRE



**01** Présentation de ReSIP

**02** Utilisation de ReSIP aux  
Archives nationales

**03** Utilisation de ReSIP en mission



# Présentation de ReSIP

# Un ensemble d'outils SEDA



La manipulation (création, édition, sauvegarde) des paquets SEDA est un problème partagé par tous les acteurs. Notamment la création de paquets est à faire dans de nombreuses applications, l'examen lors de la recette de certains SIP... Il a donc été envisagé d'avoir un ensemble d'outils pour simplifier ces tâches liées au SEDA.

On compte dans les sedatools :

- **Sedalib** : une bibliothèque permettant de manipuler à bas et haut niveaux des paquets SIP
- **Sedalib-samples** : des exemples d'utilisation de la sedalib comme pourrait le faire une application réalisant une extraction et construisant son SIP
- **Resip** : IHM de manipulation et édition des SIP et DIP

Intégration de **mailextract**, qui permet de faire de l'exportation de contenu de messagerie dans une forme normalisée eml avec les métadonnées SEDA (traite pst, thunderbird, mbox, msg, imap, gmail...).

# SedaLib

Elle permet de :

- Construire tout ArchiveUnitTransfer (fonctions bas niveaux sur les structures SEDA complètes)
- Construire, en quelques appels, un SIP en décrivant l'arborescence et les métadonnées essentielles (fonctions constructives de haut niveau)
- Construire un SIP depuis une simple arborescence disque, éventuellement porteuse de métadonnées dans des fichiers spécifiques
- Générer un SIP de classement à partir d'un CSV
- Analyser et parcourir un SIP existant ou un DIP  
...
- Construire un SIP à partir d'un CSV des métadonnées et une liste de fichiers.

A voir sur <https://github.com/ProgrammeVitam/sedatools/tree/master/sedalib>

**A noter** : il est utile de regarder en même temps **sedalib-samples** qui démontre l'usage de sedalib depuis une application construisant un SIP d'extraction

# ReSIP

L'application **ReSIP**, construite au dessus de la bibliothèque sedalib, permet de :

- construire et manipuler des structures arborescentes d'archives,
- d'en éditer les métadonnées, de les importer et exporter sous la forme de SIP, sous la forme de hiérarchie disque ou encore sous la forme CSV

Elle permet de :

- Importer SIP, DIP hiérarchie disque et messagerie
- Circuler dans l'arborescence et les objets
- Modifier les arborescences et les objets
- Modifier les métadonnées de manière assistée
- Exporter en SIP, hiérarchie disque...

The screenshot displays the ReSIP application interface. On the left, a tree view titled 'Arbre du SIP (50 archiveUnit/39 bd...)' shows a hierarchical structure of archive units. The selected unit is 'Brantes, Combes de Noïdes, torrents des Vabres et...'. The right pane shows the 'Métadonnées AU: de-2 - Brantes. Combes de Noïdes, t...' for the selected unit. The metadata is displayed in XML format, including fields like <AccessRule>, <StartDate>, <DescriptionLevel>, <Title>, <FilePlanPosition>, <ArchivalAgencyArchiveUnitIdentifier>, and <Description>. Below the metadata, there are buttons for 'Editer l'ArchiveUnit', 'Ajouter...', and a dropdown for '[A]ArchiveUnitProfile'. At the bottom, there are sections for 'Objets' and 'Détails'. The 'Objets' section shows 'ID00-de-2-bin:BinaryMaster'. The 'Détails' section provides specific information about the selected object, including 'DataObjectVersion: BinaryMaster', 'Nom: 5FI6/5FI6\_17', 'Taille: 7702 octets', 'SIP Id: ID00-de-2', 'URI: Content/1NUM\_2.JPG', 'MimeType: image/jpeg', 'PUID: fmt/43', 'Format: JPEG File Interchange Format', 'Digest: SHA-512 - 0e0cec05ald72ee5610eaa5afbc904c012d1900', and 'Path: C:\Users\jsl\Documents\Resip\FRAD'. Buttons for 'Ouvrir l'objet', 'Changer l'objet', and 'Editer le DataObject' are also visible.

# mailextract

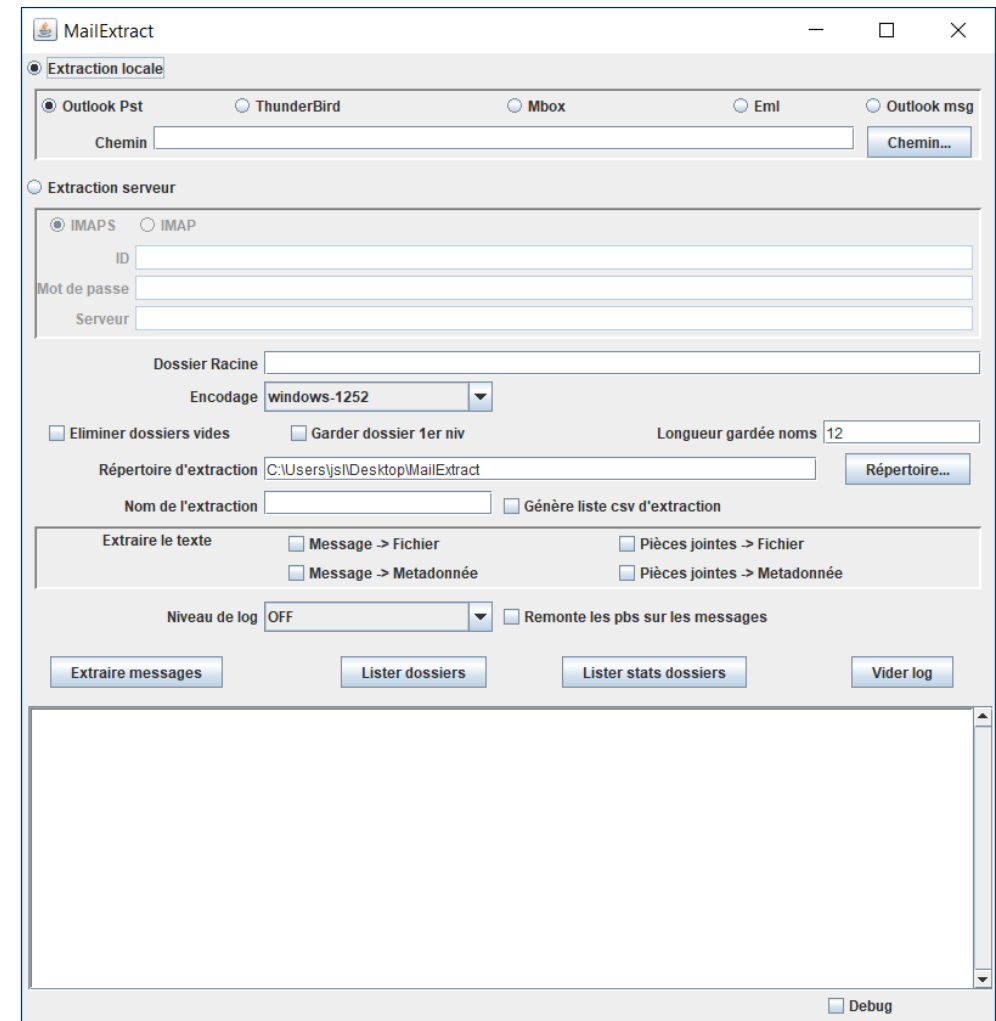
La **bibliothèque mailextract**, et l'interface associée, permettent d'extraire les messages, pièces jointes et, dans certains cas contacts et rendez-vous, de conteneurs de messagerie et informations associées.

Elle permet de :

- Extraire des conteneurs messagerie pst, thunderbird, mbox, msg, imap, gmail...
- Normaliser les messages en eml (RFC822)
- Extraire les pièces jointes à part
- Extraire les contenus textuels des messages et pièces
- Extraction des contacts des pst
- Mise sous forme SEDA sur hiérarchie disque

Est prévue l'extraction de tous les rendez-vous

**A noter** : la bibliothèque a été intégrée dans **ReSIP**





# Utilisation de ReSIP aux Archives nationales



# Utilisation de ReSip par les Archives nationales

## Contexte d'utilisation de l'outil

1. Reprise des versements réalisés avant la mise en service de la plateforme d'archivage (novembre 2018)

Ensembles bureautiques, messageries, fonds audiovisuelles (photo, vidéo, son)

2. Lors du traitement de fonds par les archivistes en mission :

Échanges sur la conception du SIP sur les aspects techniques

# Utilisation de ReSip par les Archives nationales

## 3. Vérifications réalisées sur les « manifests » avant versement des SIP :

- Contrôles sur les fichiers à partir des statistiques remontées par ReSip
- Contrôles de conformité XML, SEDA, profil minimum ADAMANT
- Conformité aux exigences archivistiques des AN

## 4. Outil de formation au traitement des archives numériques des archivistes (mission, AN) et étudiants en Master 2 (UP8 / UP13)

## 5. Outil permettant de visualiser et d'échanger sur des propositions d'organisation des données avec les partenaires de projets de versement complexes (Grand Débat National, Centre G. Pompidou, application de la Cour de cassation, Données comptables du SAE finances de La Poste etc)

## 6. Dans le cadre des développements réalisés pour la plate-forme d'archivage des AN (ADAMANT) ReSip est utilisé pour la conception de SIP test

# Traitement des ensembles bureautiques



- **Première étape :**  
contrôles techniques sur les fichiers

## **Vérifications réalisées à partir des statistiques :**

- détection et élimination des fichiers vides (0 octet)
- détection et élimination des fichiers indésirables (.\_TemporaryItems, desktop.ini...)
- détection et traitement des dossiers ou fichiers compressés (zip, tar, tar.gz...)
- détection et fusion des doublons

# Traitement des ensembles bureautiques



- **Deuxième étape : traitement métier**

- travail sur la cohérence de la structure du dossier
- détection d'éventuels dossiers personnels
- détection des caractères interdits xml
- travail de description
- renseignement des règles de gestion (communicabilité, diffusion, réutilisation)

*Ne pas oublier de sauver le projet en cours de traitement*

**En fin d'étape : régénération des ID continus, contrôle conformité SEDA puis export du SIP**

# Vérifications effectuées sur un manifeste avant versement

**Pré-requis : le manifeste doit avoir été enrichi via la plateforme d'archivage lors de la création de l'entrée**

## **Créer un « SIP de visualisation du dossier »**

1. créer un dossier « content » vide
2. zipper le dossier « content » et le manifest
3. importer le SIP dans ReSip

# Vérifications effectuées sur un manifeste avant versement

- **Vérifications d'ordre techniques (détection de fichiers indésirables, ou compressés)**
- **Cohérence de structure du SIP**
- **Conformité SEDA et au profil minimum ADAMANT**

# Traitement des messageries

- Formats rencontrés :
  - Thunderbird : Mbox + dossiers sbd
  - Outlook : Msg, Pst
  - Eml

Chaque message est converti en eml, le contenu textuel est extrait dans un fichier texte (version TextContent dans le GOT), les pièces jointes sont extraites.

Archivage du courrielleur dans un autre SIP.

Répond à des problématiques de préservation + empêche la recherche sur des métadonnées sensibles directement dans le manifest (pas d'extraction du texte dans le manifest)

# Traitement des messageries



- Extraction de métadonnées utiles : date d'envoi et de réception, expéditeur, destinataires, ...
- Organisation logique des UA
- Utilisation recherche sur les UA pour repérer les messages personnels, confidentiels, sans objet «perso», « maison », « confidentiel », « pas de sujet », « sans objet »



# Traitement des messageries



- Utilisation de l'export de la hiérarchie sur disque pour constituer un seul SIP à partir de deux conteneurs courriels (ex : un fichier pour la boîte de réception, un autre pour les messages envoyés)
- Utilisation de l'import de conteneur pour traiter des messages qui se trouvent dans des ensembles bureautiques (préservation + extraction des métadonnées)

# Points forts



- Seul outil à disposition pour traiter archivistiquement des messageries
- Le principe de l'extraction des messages compatible SEDA => organisation des UA et des GOT à l'extraction des messages et des p.j.
- Extraction des métadonnées
- Création des dates extrêmes au niveau des UA supérieures à partir des dates d'envoi et de réception des mails
- Option d'extraction du contenu textuel : préservation ++

# Points faibles



- Longueur de l'import des messageries lourdes
- Quelques imports KO, extraction avec log trop longue pour connaître la source de l'erreur d'import (mais pas eu le temps d'approfondir, valeur de test)
- Pas d'extraction à la demande des mails dans des ensembles bureautique (une fonctionnalité comme le dézippage serait la bienvenue) => demande prise en compte



# Utilisation de ReSIP aux Ministères sociaux

# Programme Vitam

47 rue de la Chapelle, 75018 Paris – France  
Tél. : +33 (0)1 86 69 60 03

[www.programmevitam.fr](http://www.programmevitam.fr)  
<https://twitter.com/@ProgVitam>  
<https://www.linkedin.com/grps/Programme-Vitam>

