



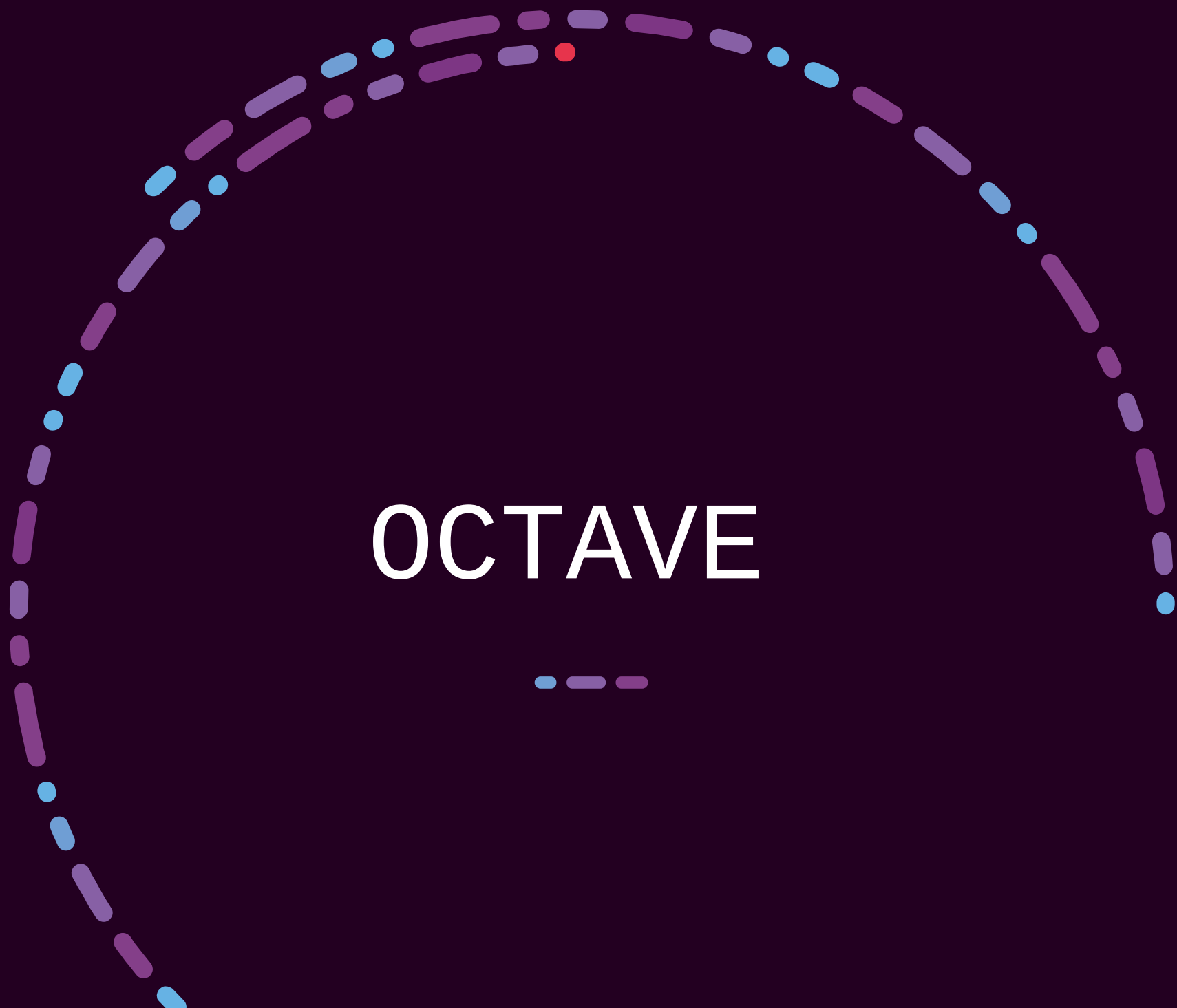
3^e partie

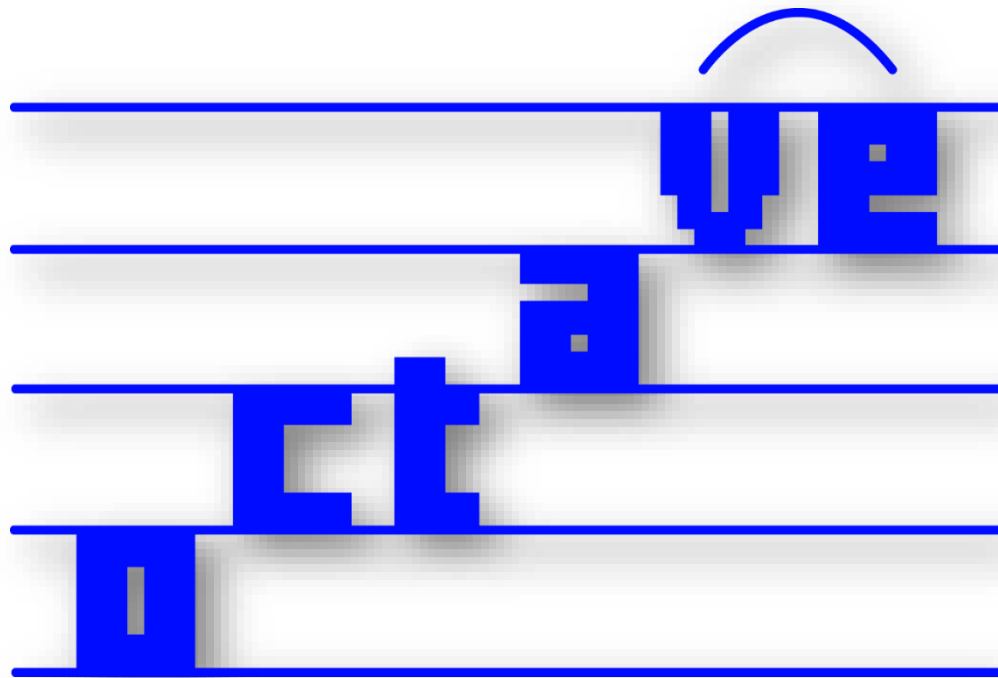
Les outils de Préparation des entrées



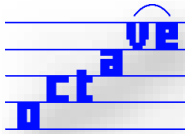
Archives nationales,
Site de Pierrefitte-sur-Seine

18/03/2019

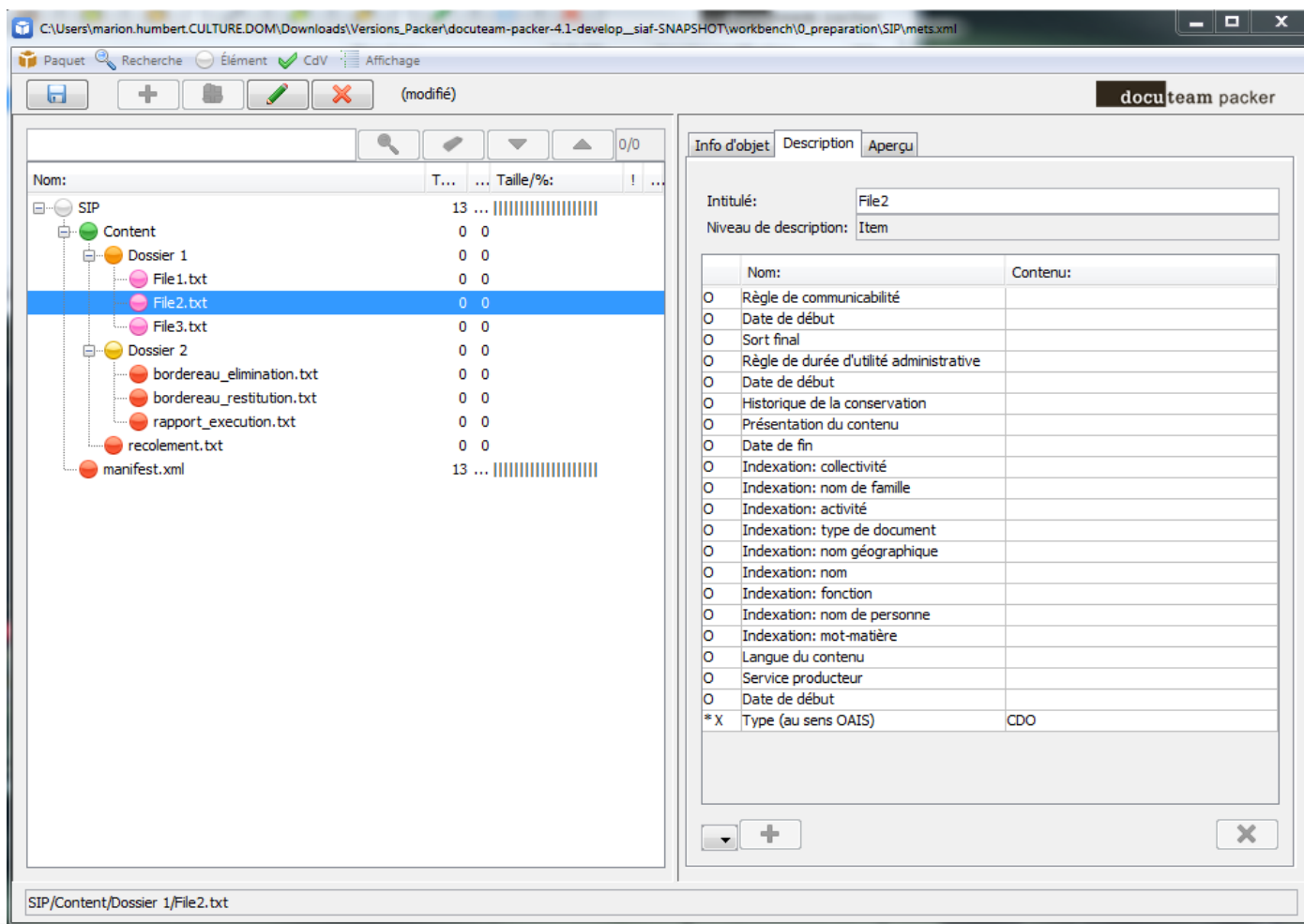




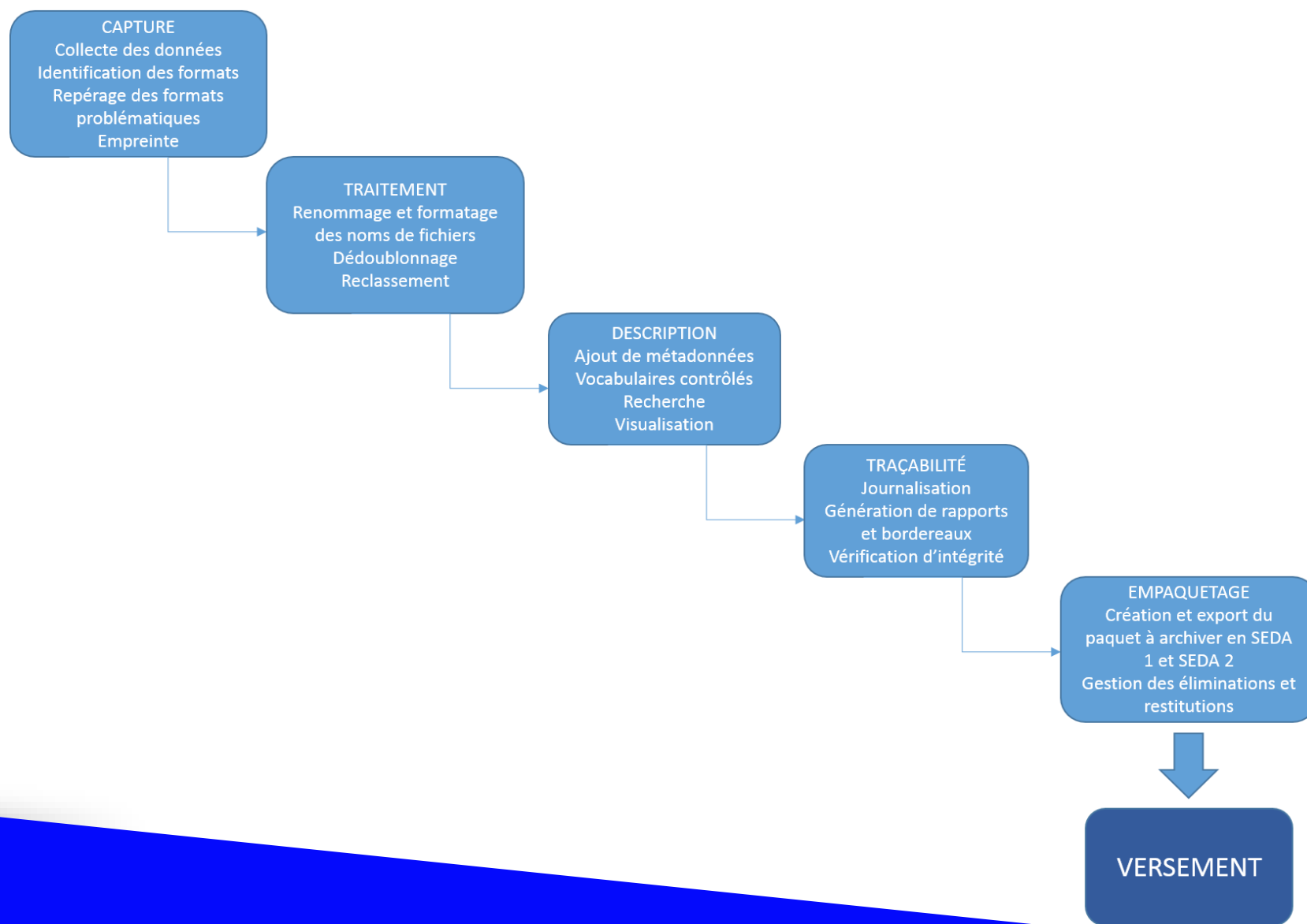
Outil de
Constitution et de
Traitement
Automatisé des
Vracs
Electroniques



- ❑ Animation en 2015 d'un groupe de travail et recueil de cas d'usage
- ❑ Tests des outils déjà existants (voir le carnet de recherche du SIAF)
- ❑ Cadrage du projet : étude de risques, chiffrage
- ❑ Recadrage du projet en 2017 au vu des fonctionnalités déjà existantes dans Docuteam Packer
- ❑ Développements depuis novembre 2017
- ❑ Réalisation en 2019 d'un audit de code en prévision du déploiement de la solution



- ❑ Logiciel libre
(licence GNU GPL) basé
sur l'outil Docuteam
Packer
- ❑ Client lourd JAVA,
utilitaire
- ❑ Deux distributions
disponibles : avec ou
sans JRE
- ❑ OCTAVE sera fourni
sous forme de dossier
zip



Pour en savoir plus:
marion.humbert@culture.gouv.fr



ARCHIFILTRE



Un outil pour l'audit des arborescences de fichiers bureautiques

archifiltre



par Chloé Moser

Adjointe à la cheffe de la Mission des archives de
France auprès des ministères sociaux

*Journée de présentation de la solution VITAM v2
18 mars 2019*



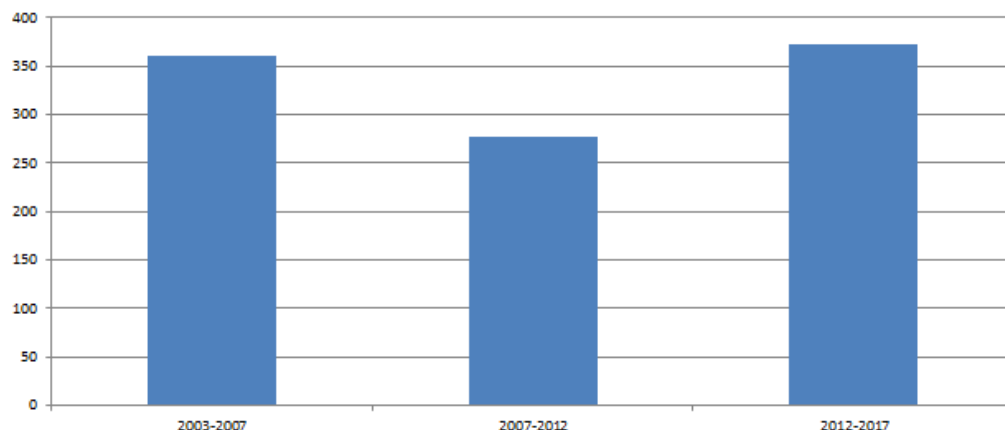
La mission des archives auprès des ministères sociaux

- 10 agents au sein du bureau dont 3 chargés du traitement
- 20 kml de préarchivage (papier)
- Moyenne des entrées papier:
 - 850 ml/an versés par les services des ministères à la Mission
 - 115 ml traités par la Mission et versés aux AN par an
- Moyenne des éliminations visées:
 - Pour les services des ministères: 300 ml/an
- Versements bureautiques:
 - 232 versements: de quelques Ko à 193 Go
 - 2,1 To de données



La collecte des archives de cabinet de 2017

Archives papier (ml)



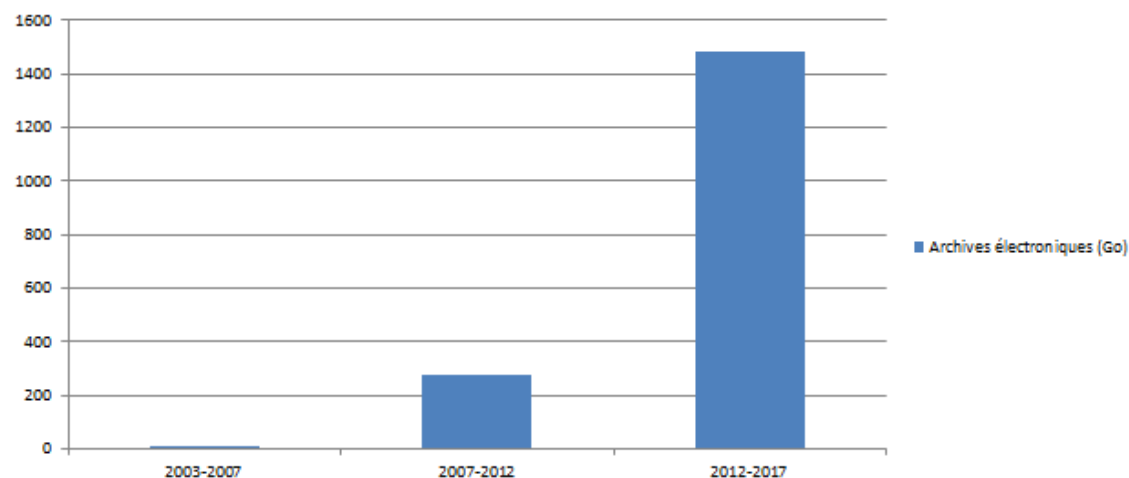
Une **collecte papier**:

- équivalente aux mandatures précédentes
- 136 versements
- plus de 250 ml

Une **collecte électronique** inédite:

- 95 versements
- 1,2 To de données
- 38 versements de bureautique
- 57 versements de messageries

Archives électroniques (Go)





Des expériences de traitement de bureautique

1 exemple de traitement :

- 6 Go de données
- Pas de détails sur le contenu
- 2 mois de traitement par une stagiaire archiviste
- 6% du fonds compris et décrit

Conclusions:

- Impossible de comprendre la totalité du fonds
- Manque de moyens humains
Les moyens humains ne vont pas augmenter

Les 2 problématiques de l'archiviste:

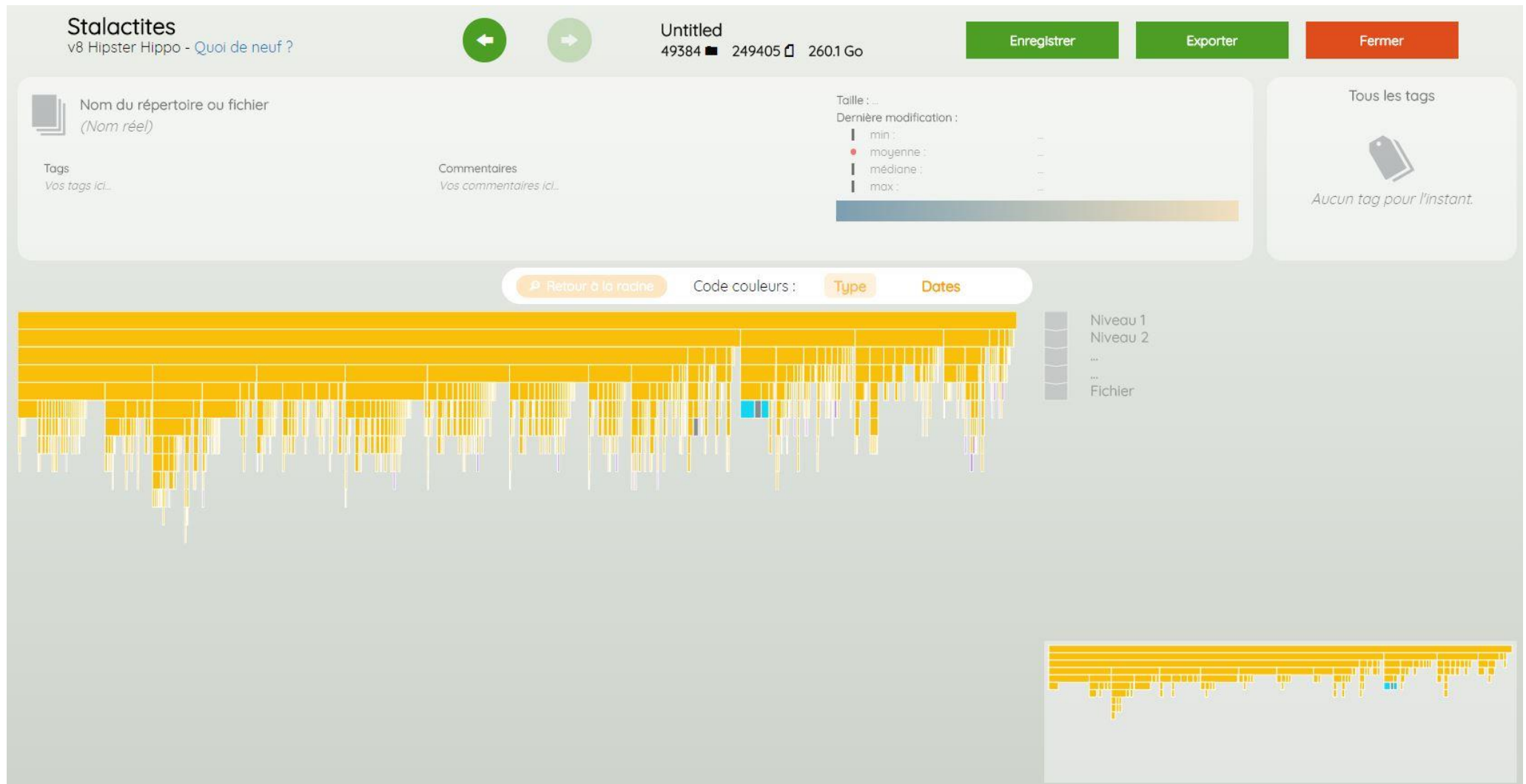
- L'appréhension, la compréhension des arborescences pour les décrire
- Le temps dont il dispose



Quel fonds vous sentez-vous prêt.e à attaquer en une ½ journée?

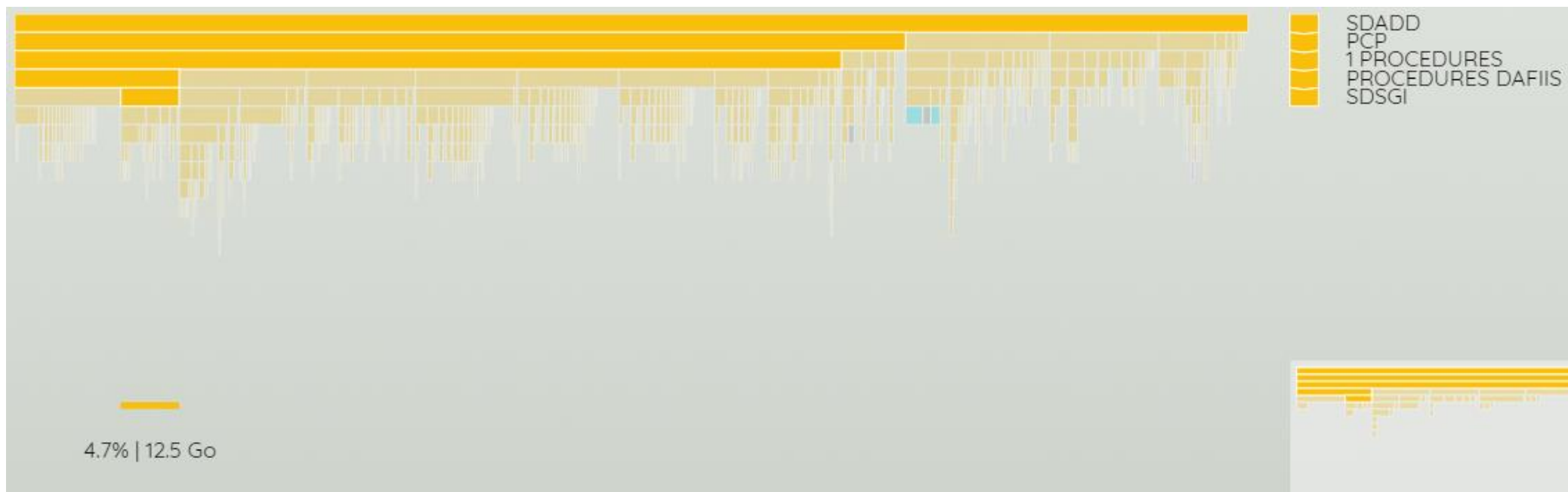


L'outil développé: la visualisation d'arborescences

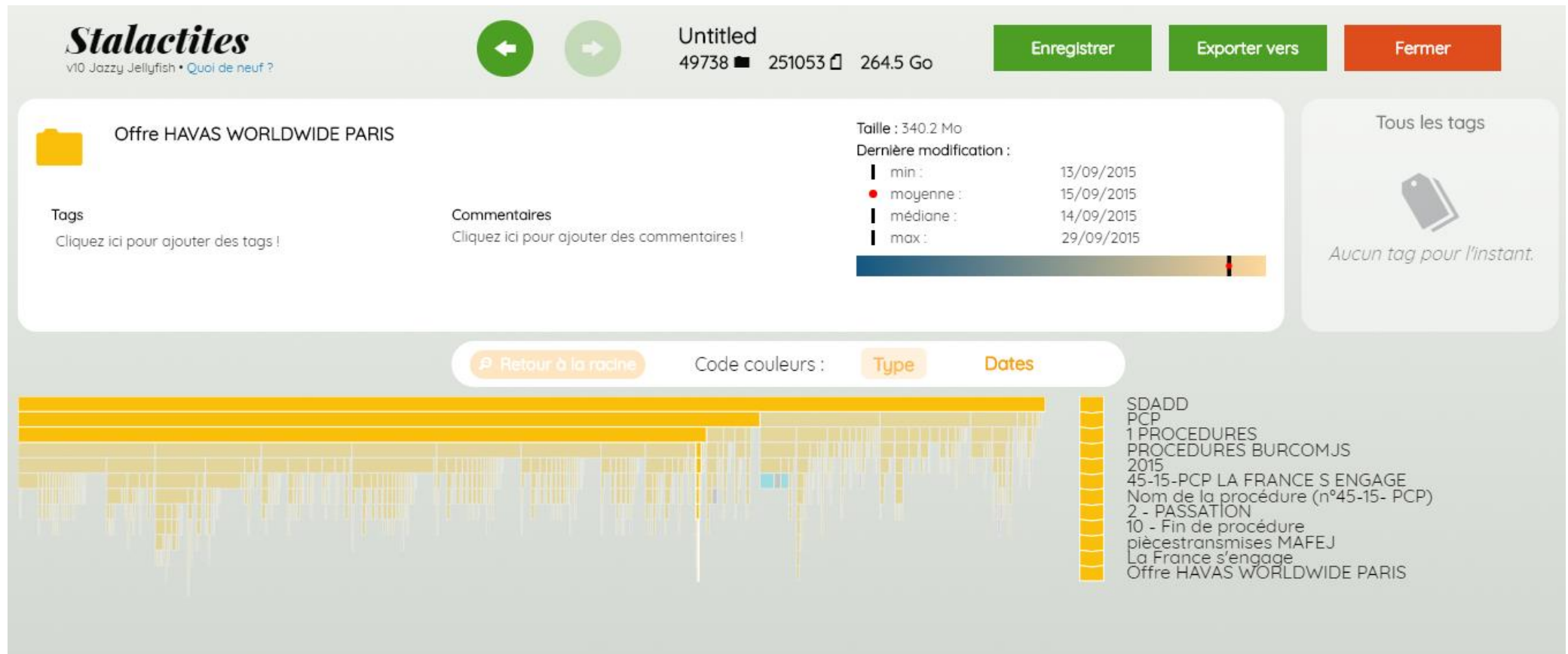




Une visualisation basée sur le volume des répertoires et fichiers



L'appréhension de la profondeur d'une arborescence



La navigation et le zoom

Stalactites
v10 Jazzy Jellyfish • [Quoi de neuf ?](#)


← →

Untitled
 49738 ■ 251053 □ 264.5 Go

Enregistrer

Exporter vers


Fermer


 Nom du répertoire ou fichier
 (Nom réel)

Tags
 Vos tags ici...

Commentaires
 Vos commentaires ici...

Taille : ...
 Dernière modification :
 | min : ...
 | moyenne : ...
 | médiane : ...
 | max : ...

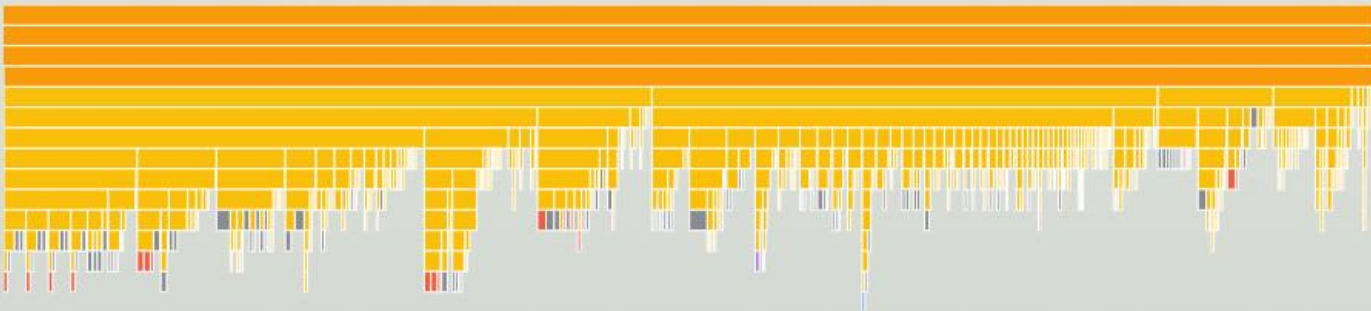
Tous les tags

 Aucun tag pour l'instant.

Retour à la racine

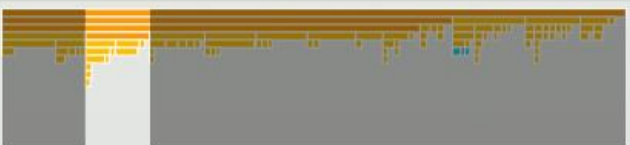
Code couleurs :

Type

Dates

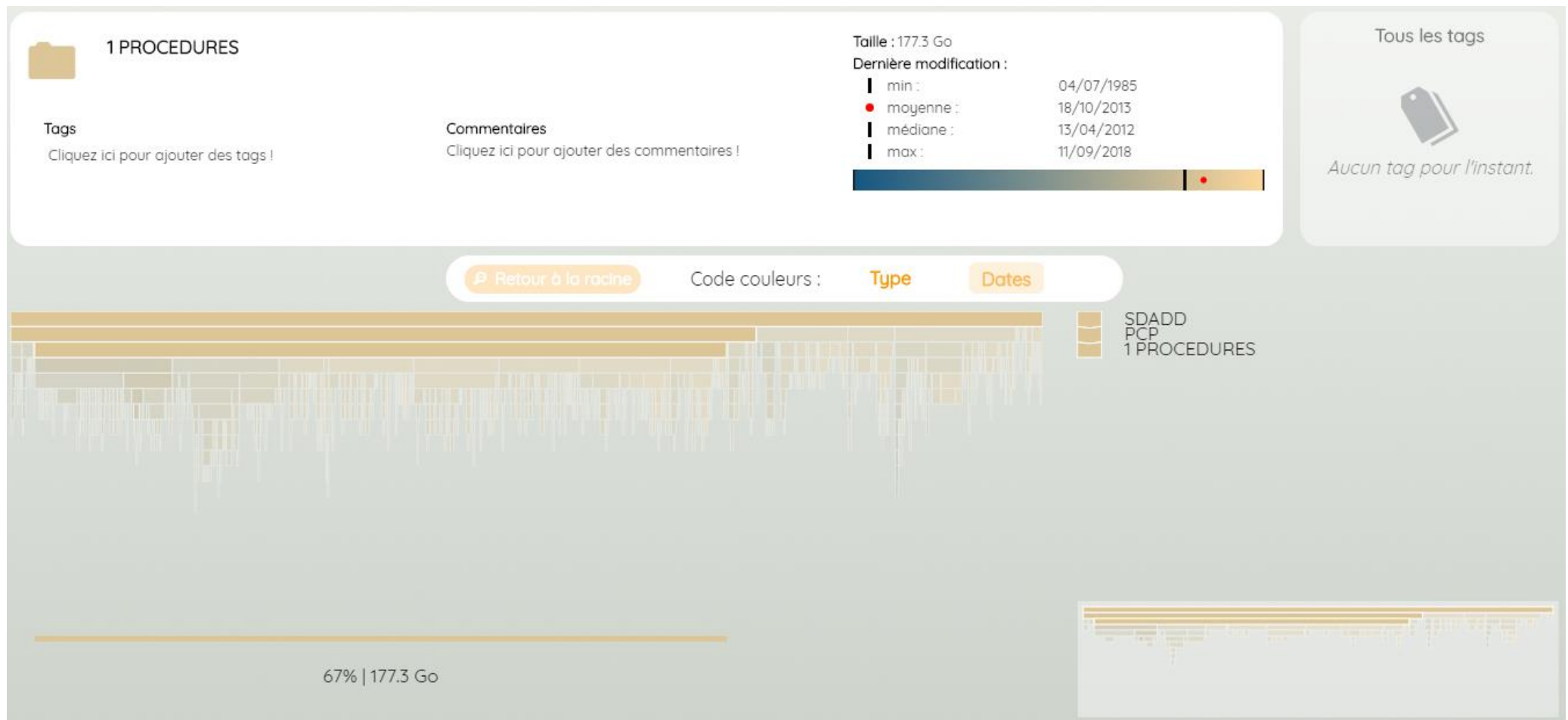


Niveau 1
 Niveau 2
 ...
 Fichier



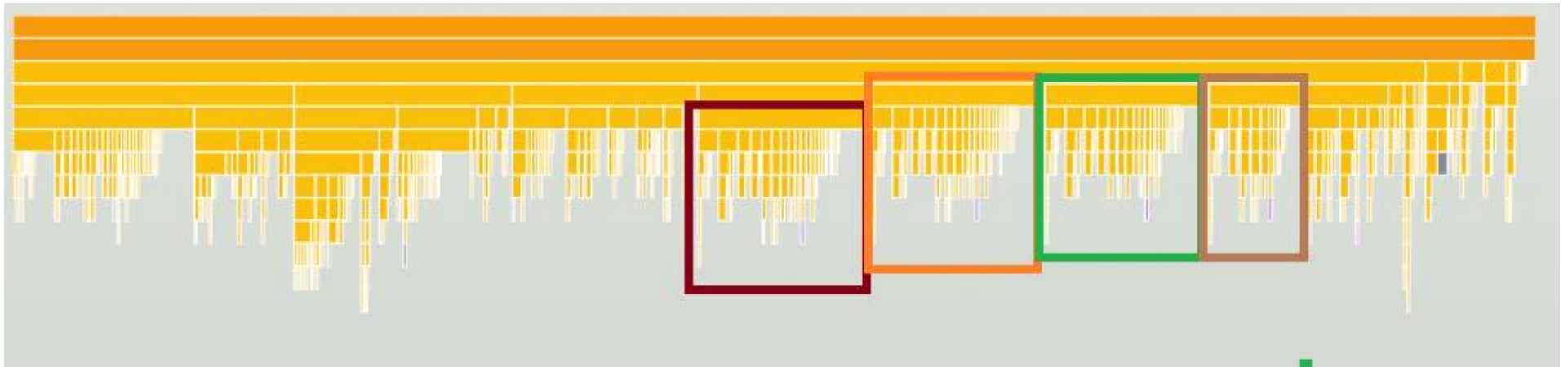


L'appréhension de l'ensemble des dates



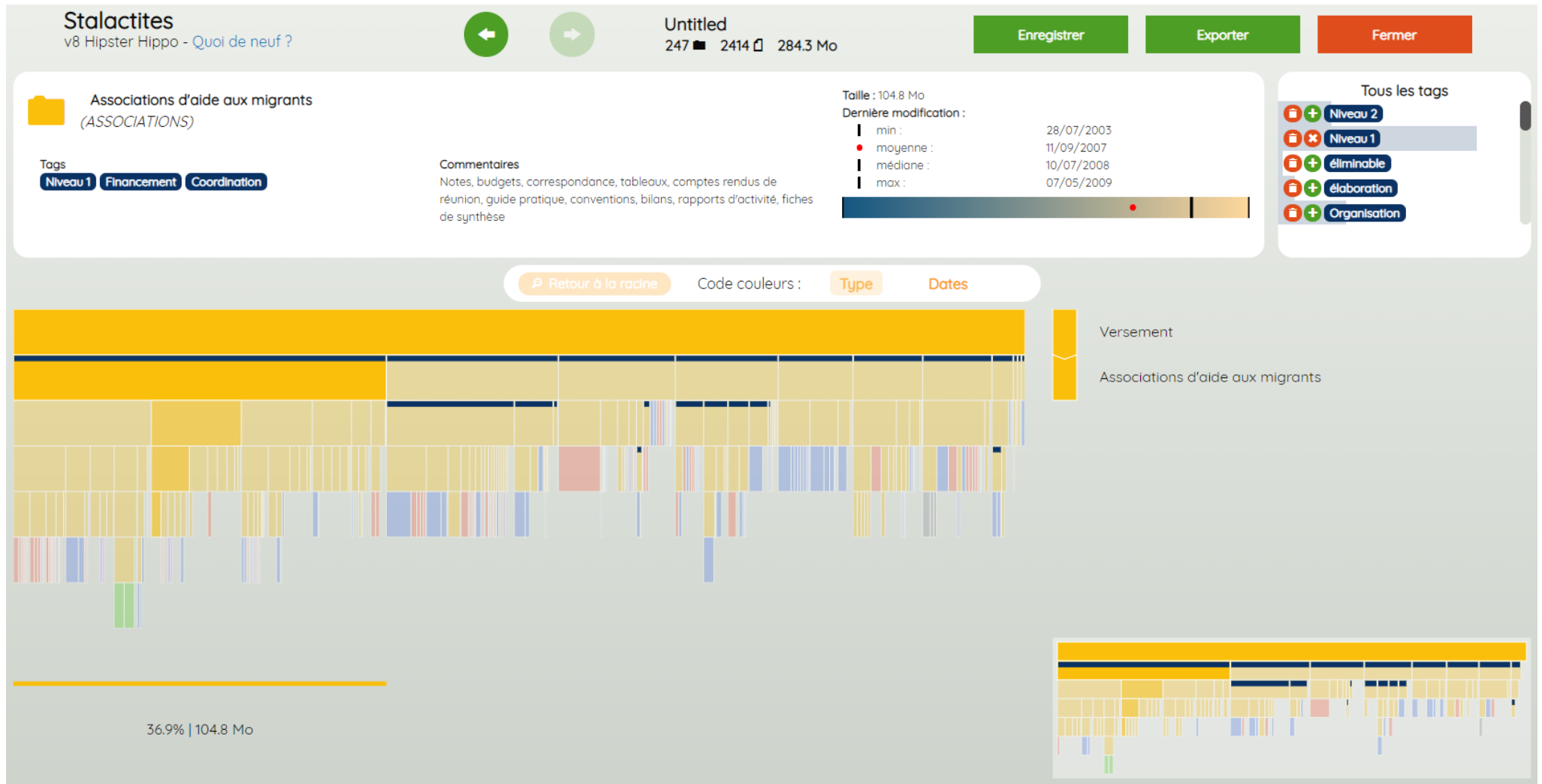


La détection de doublons d'arborescences ou d'arborescences approchantes





L'enrichissement des métadonnées





Formats de sortie

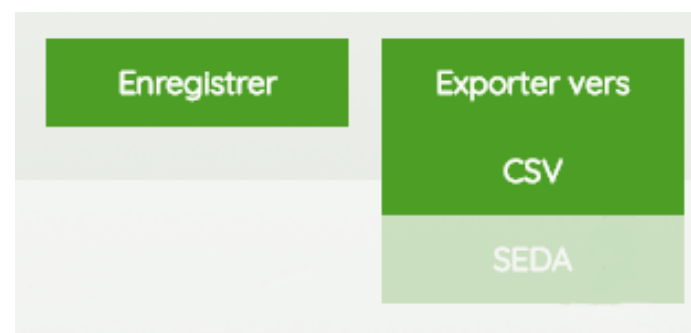
- Enregistrement au format .json

- Export possible:

- Au format .csv

- Au format SEDA 2.1

pour entrée dans VITAM



B	C	D	E	F
path	size (octet)	last_modified	alias	comments
SGMAS_2014_009/Tableau de bord/Médico-social/Réunion 13 décembre 2010/RE Réu	131413	Mon Jan 03 2011 15:37:24 GMT+0100 (heure nc		
SGMAS_2014_009/ETAT FINANCIER 2012 au 30 mars 2012/Maquette, annexe litt, Fichie	3810464	Mon Mar 05 2012 17:22:02 GMT+0100 (heure n		
SGMAS_2014_009/GESTION du RISQUE/Circulaire - plannig - indicateurs 2012/Indicate	17408	Mon Jan 16 2012 17:55:46 GMT+0100 (heure nc		
SGMAS_2014_009/GESTION du RISQUE/Handicap/Compte rendu CNAMTS agent compt	69632	Wed Jul 11 2012 15:07:50 GMT+0200 (heure d'		
SGMAS_2014_009/OGD/Rapport OGD v 24 midi.doc	190976	Wed Oct 24 2012 11:41:56 GMT+0200 (heure d'		
SGMAS_2014_009/GESTION du RISQUE/Imagerie/Questionnaire Imagerie 22 déc 2011/	41984	Thu Dec 22 2011 13:34:24 GMT+0100 (heure no		
SGMAS_2014_009/Docs stratégiques/COG RSI/cog_rsi_2007-11.pdf	1435072	Mon Apr 20 2009 17:03:48 GMT+0200 (heure d'		
SGMAS_2014_009/FIR/règlementation/arrêtés FIR 2013/dotations régionale FIR 2013 a	165785	Thu Mar 14 2013 16:12:12 GMT+0100 (heure nc		

Les cas d'usages d'Archifiltre

- **Audit** d'arborescences de gros volumes pour les services (jusqu'à 60 à 70% de documents à supprimer)
 - Outil supplémentaire pour l'offre de service de l'archiviste
- **Appréhension** d'arborescences à traiter et verser dans un SAE
- Définition du bon niveau de description et **Enrichissement** de l'arborescence à verser

Développements 2019

Passage d'ArchiFiltre en start-up d'Etat:

- Au sein de l'incubateur des ministères sociaux
- Poursuite des développements de 2018
- Basés sur les retours et les priorisations des utilisateurs
- Groupe utilisateurs en cours de constitution

→ *Venez nous rejoindre!*



Pour en savoir plus...

- Le site: <https://archifiltre.github.io>
- Le compte Twitter @ArchiFiltre
- Cheffes de projet
 - anne.lambert@sg.social.gouv.fr
 - chloe.moser@sg.social.gouv.fr





SEDATTOOLS



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

A ce jour existe aussi un outil traité à part, **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...). Il devrait sous peu entrer dans l'ensemble sedatools.

L'ensemble des sedatools est soumis à étude des porteurs et partenaires comme tous les sedatools pour déclaration d'intérêt.

A noter il conviendra en fin de release R10 de décider si ces différents outils rentrent dans ce qui est industrialisé et supporté au titre de la SL Vitam.

SedaLib

Elle permet de :

- Construire tout ArchiveUnitTransfert (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
- ...

Est aussi prévue à ce jour la construction d'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> (prochaine release 1.2.0 fin de semaine)

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

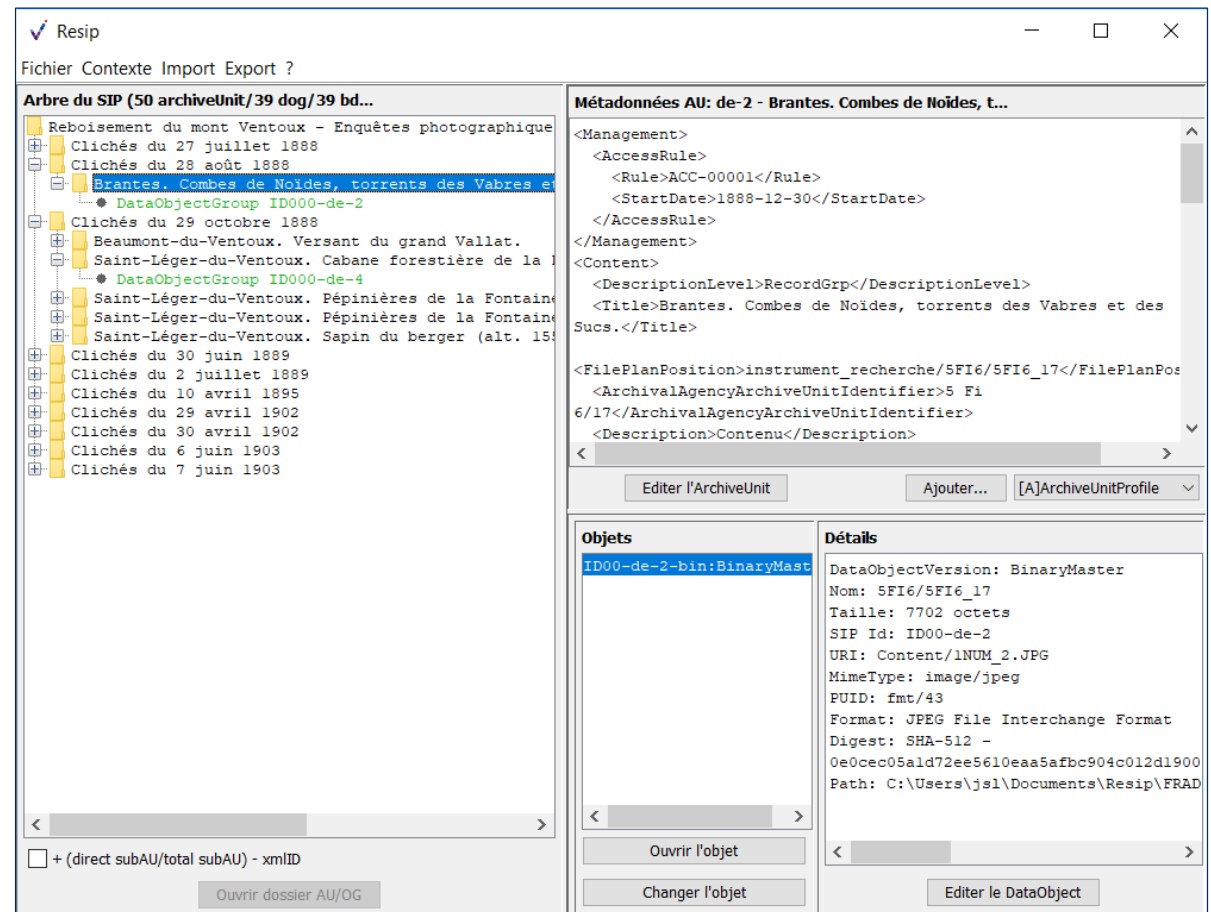
ReSIP

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 pour les plans de classement.

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...



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**

The screenshot shows the MailExtract application window. It has a title bar with the icon and name 'MailExtract'. The interface is divided into several sections. At the top, there's a tabbed interface with 'Extraction locale' selected. Under 'Extraction locale', there are radio buttons for 'Outlook Pst', 'ThunderBird', 'Mbox', 'Eml', and 'Outlook msg'. Below these is a 'Chemin' text field and a 'Chemin...' button. The 'Extraction serveur' section is currently unselected. It contains radio buttons for 'IMAPS' and 'IMAP'. Below these are fields for 'ID', 'Mot de passe', and 'Serveur'. There's a 'Dossier Racine' text field and an 'Encodage' dropdown menu set to 'windows-1252'. Below that are checkboxes for 'Eliminer dossiers vides' and 'Garder dossier 1er niv', and a 'Longueur gardée noms' text field set to '12'. There's a 'Répertoire d'extraction' text field with the path 'C:\Users\js\Desktop\MailExtract' and a 'Répertoire...' button. Below that is a 'Nom de l'extraction' text field and a checkbox for 'Génère liste csv d'extraction'. The 'Extraire le texte' section has checkboxes for 'Message -> Fichier', 'Message -> Metadonnée', 'Pièces jointes -> Fichier', and 'Pièces jointes -> Metadonnée'. There's a 'Niveau de log' dropdown menu set to 'OFF' and a checkbox for 'Remonte les pbs sur les messages'. At the bottom, there are four buttons: 'Extraire messages', 'Lister dossiers', 'Lister stats dossiers', and 'Vider log'. A large empty text area is at the bottom of the window. In the bottom right corner, there's a 'Debug' checkbox.

Programme Vitam

47 rue de la Chapelle, 75018 Paris – France

Tél. : +33 (0)1 86 69 60 25

www.programmevitam.fr

<https://twitter.com/@ProgVitam>

<https://www.linkedin.com/grps/Programme-Vitam>

