

Semaine de webinaires pour
apprendre, échanger,
s'informer, questionner

Programme et inscriptions prochainement sur
<https://www.programmevitam.fr/>

Vitam en ligne

DU 06 AU 10 NOVEMBRE 2023

2E EDITION

JEUDI APRÈS-MIDI
09 NOVEMBRE 2023



**Archivage des courriers électroniques : théorie, retour
d'expérience sur la collecte et présentation
d'Archifiltre-Mails**

Avec Chloé Moser, mission Archives des ministères sociaux,
Nathalie Morin, mission Archives des services du Premier ministre et
Marion Ville, experte fonctionnelle du Programme Vitam



2013 – 2023 : retour sur 10 ans d'archivage des messageries

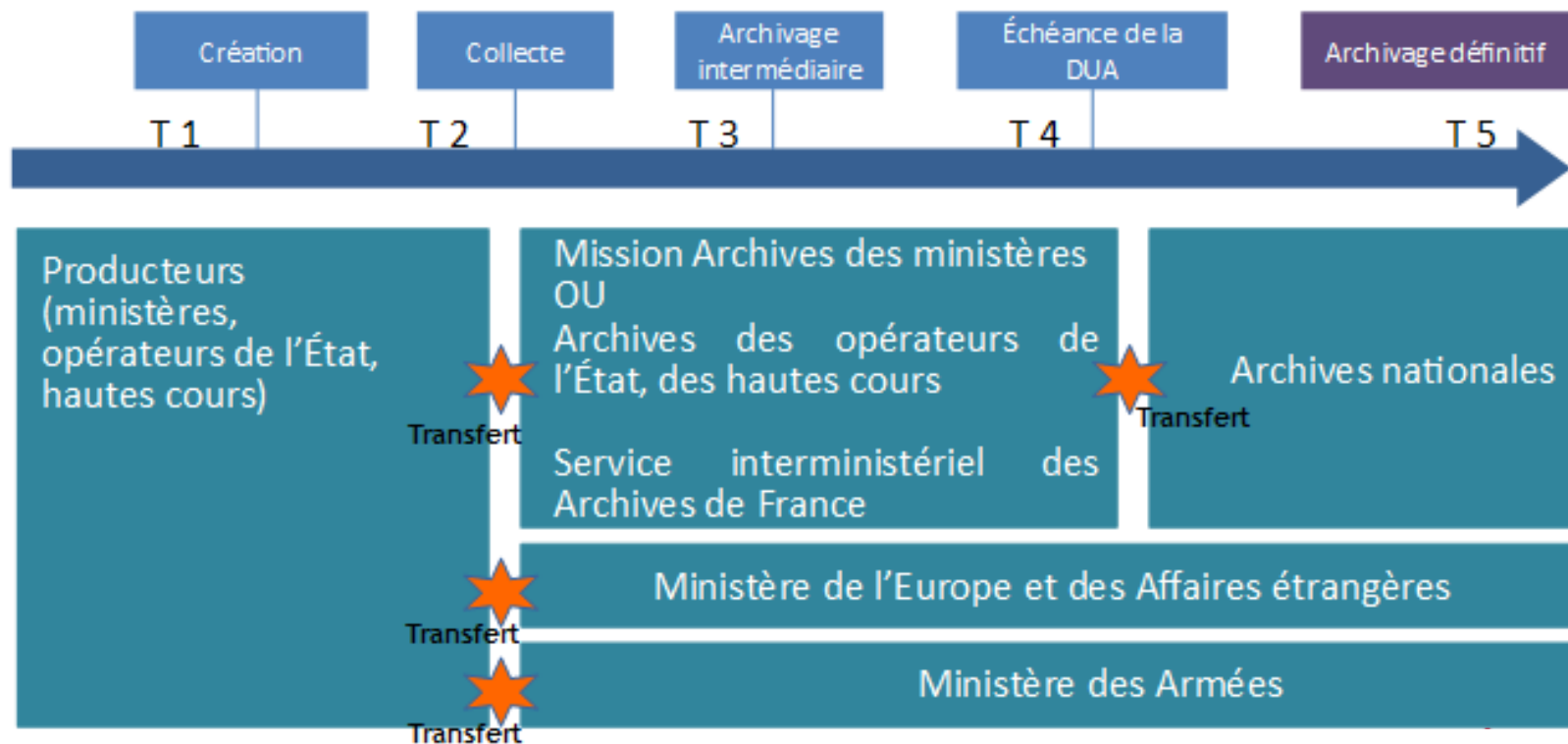
Marion Ville

9 novembre 2023

CONTEXTE ET PREUVE DE CONCEPT VITAM



Contexte : les archives de l'administration centrale



Objectifs et déroulement

4 partenaires :



MINISTÈRE
DE L'EUROPE
ET DES AFFAIRES
ÉTRANGÈRES

*Léonard
Agnès
François*



MINISTÈRE
DE LA CULTURE

*Léonard
Agnès
François*



MINISTÈRE
DES ARMÉES

*Léonard
Agnès
François*



3 objectifs :

Définir une stratégie pour la collecte, le traitement et la conservation des messageries électronique

Identifier les outils et les fonctionnalités techniques nécessaires au traitement technique des messageries électroniques

Définir un modèle de métadonnées pour les messageries conformément aux réflexions de l'équipe du projet Vitam

2 étapes :

entre mars et juin 2013

Revue de littérature

Expérimentation

1 livrable :

publié en octobre 2013



L'archivage des messageries électroniques

Preuve de concept VITAM

version 1.0 - 01 octobre 2013

Équipe projet VITAM
Hélène CHIFFOLEAU
Hélène CHIFFOLEAU
Lionel BARRON
Lionel BARRON
Nicolas FROST
Nicolas FROST
Lionel BARRON

Déroulement en détail

Première étape : revue de littérature

Objectif : réaliser une synthèse en français des travaux menés aux niveaux national et international

Revue de littérature française :

- instructions et législation concernant les emails
- Livres blancs émanant d'associations
- guides internes

Revue de littérature internationale avec 3 approches :

- une approche “pédagogique”
- une approche “collecte et conservation”
- une approche “diplomatique”

Deuxième étape: expérimentation

Objectif : tester la conversion de fichiers d'un format à un autre (.eml vers .mbox, .pdf; .mbox vers .csv, .eml, .pdf; .pst vers .eml, .mbox) et l'extraction des messages et de leurs pièces jointes.

Type de messageries : Thunderbird et Outlook

Outils :

- > 2 progiciels du marché (Aid4Mail, EmailChemy)
- > 2 outils développés par des services d'archives (CERP Email Parser and Xena)
- > Outils disponibles dans des bibliothèques open source (Apache POI, CLibPST, Java LibPST, Java Mail, Mime4J, Tika, DROID)

Conclusions et propositions

Propositions :

- **Propositions organisationnelles** : rappeler des règles simples relatives à l'utilisation des messageries.
- **Propositions juridiques** : rédiger des clauses-types dans les règlements intérieurs et les chartes informatiques, élaborer un protocole-type à faire signer par l'agent au moment de la collecte de sa messagerie.
- **Propositions techniques** : transmettre des instructions aux services informatiques, développer des plug-ins supplémentaires
- **Propositions archivistiques** : définir une stratégie d'acquisition du courrier électronique, développer un schéma XML pour la représentation des archives de courrier électronique.

Methodologie:

- **Collecte** : extraction des messageries aux formats .eml, .mbox et .pst. Pour ce dernier format, recommandation de réduire au maximum le temps de traitement entre l'extraction et le transfert pour archivage.
- **Constitution des SIP** : un fichier zippé comprenant un fichier au format XML décrivant la messagerie et respectant le Standard d'Echanges de Données pour l'Archivage (SEDA) et la structure hiérarchique des dossiers, avec, pour chaque message, le corps du message, les pièces jointes associées, et les messages dans leur format d'origine.

DE LA THEORIE A LA PRATIQUE



Identifier les messageries à collecter

2 stratégies de sélection :

- Identification et sélection des messageries jugées **les plus engageantes** (approche CAPSTONE)

Ministère de l'Europe et des Affaires étrangères :

Environ 500 comptes de messageries identifiés parmi les 15 000 messageries existantes (sous-directeurs, directeurs adjoints)

Mission Archives auprès des Services du Premier ministre :

Messageries ciblées : collaborateurs directs du Premier ministre, plus hauts fonctionnaires (directeur de l'Institut national de la fonction publique, secrétaire général du gouvernement, directeur du département numérique, etc.)

- Sélection de comptes de messagerie en usage pour mener à bien **une mission spécifique**

Ministère de l'Europe et des Affaires étrangères :

1.7772 comptes de messagerie en usage pour la COP21 (conférence de Paris sur le changement climatique en 2015)

Formats très variés (.pst, .msg, .msf, .eml)

Mission Archives auprès des Ministères sociaux :

143 messageries des agents employés en renfort au sein du centre de crise sanitaire

Département des archives du Ministère de la Justice :

15 comptes personnels et 2 comptes fonctionnels utilisés par la cellule dédiée à la crise sanitaire (janv. 20- déc. 22)

La collecte des messageries

Quand ?

- La plupart du temps, lorsque le compte de messagerie est fermé (ex. après les élections).

Comment ?

- Méthode 1 : collecter d'abord, traiter après
- Méthode 2 : évaluer et traiter avant de collecter

Mission Archives auprès des Ministères sociaux :

Collecte de 143 messageries des agents employés en renfort au sein du centre de crise sanitaire

*Aucun tri et aucune évaluation : quelques messageries ne contiennent que quelques messages.
Réévaluation à venir.*

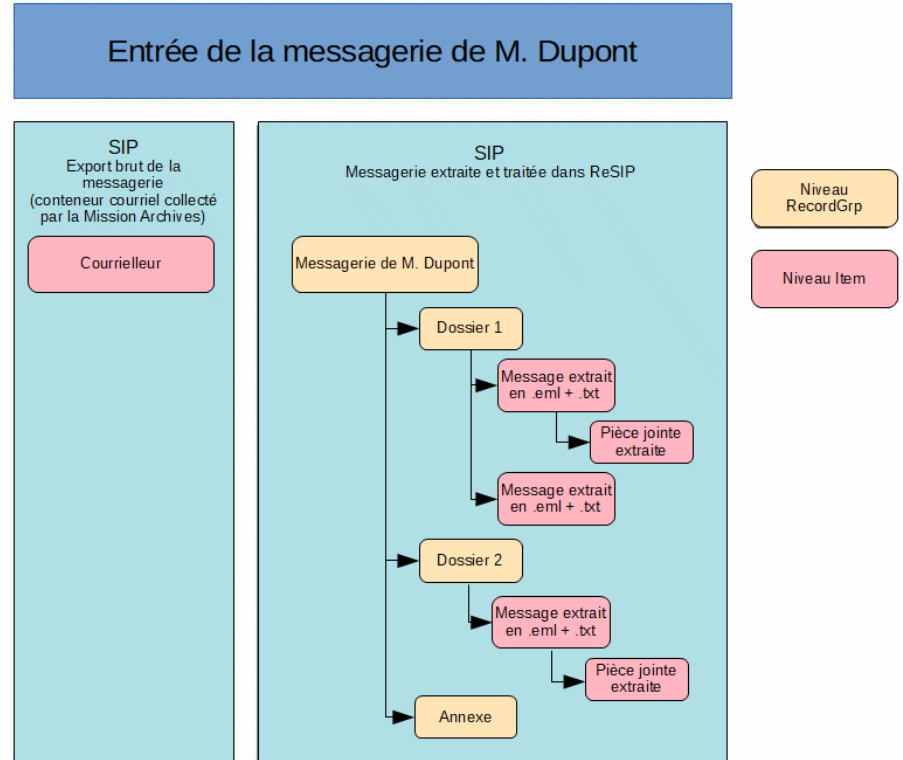
Département des archives du Ministère de la Justice :

- *Entretien avec le personnel de l'unité de crise sanitaire*
- *Evaluation : collecte de 15 messageries personnelles et de 3 messageries fonctionnelles (20 janv. - 22 déc.)*
- *Traitement : si le producteur n'a pas préparé de dossier sur le sujet, sélection des messages filtrés par une trentaine de mots-clés établis avec l'accord des agents.*

Méthode de collecte

2 stratégies de collecte :

- collecter uniquement les messageries dans leur format d'origine (Missions Archives)
- collecter l'export brut et une version extraite (Archives nationales, Ministère de l'Europe et des Affaires étrangères).



Proposition d'un modèle et d'un outil

ReSIP :

Le modèle et l'outil ont été conçus par le programme Vitam.

Objectif : import, traitement et export de messageries sous la forme d'un SIP

Fonctionnalités :

- extraction de messageries au format .pst, .mbox, .eml ou .msg et migration vers d'autres formats : .eml, .txt,
- extraction des agendas et des listes de contacts vers un fichier au format .csv,
- Extraction des métadonnées des comptes de messageries, et le contenu textuel du body des messages et des pièces jointes comme métadonnées encapsulées dans un fichier .xml conforme au SEDA ou dans un fichier au format .csv,
- traitements (ex. dézipper les pièces jointes, supprimer les dossiers sans messages, les dossiers contenant des messages considérés comme privés, les cartes de visite, les logos au format .emz, les fichiers au format .dat, etc.),
- ajout de métadonnées,
- export d'un SIP

Formalisation de stratégies d'archivage et construction réglementaire

- Formalisation des stratégies dans des **notes de procédure et des documents** :
 - instructions “officielles”,
 - notes et documents émanant d'institutions à usage interne (formes et approches variées)
- Encouragés par des **échanges** entre institutions au sein :
 - du programme Vitam,
 - de la Cellule nationale de veille des formats,
 - du réseau des archives publiques.

Une collecte croissante de messageries

Comparaison entre les différentes campagnes (Missions Archives) :

	2012 (mandat 2007-2012)	2017 (mandat 2012-2017)	2022 (pour la période 2020-2022)
Métrage papier	2557,4 ml	2000 ml	772 ml
Métrage numérique	615 Go	9 To	10,8 To
Nombre de messageries	7		577
	Pour 11 Missions Archives. 7 messageries pour 2 Missions Archives.	Pour 11 Missions Archives.	Pour 9 Missions Archives.

PROBLEMATIQUES ACTUELLES



Premier problème : une masse à traiter

- **Problème archivistique** : une difficulté à identifier rapidement les messages à supprimer, même si les archivistes savent quels types de messages électroniques doivent être supprimés (messages personnels, listes de diffusion, etc.). Certains archivistes recherchent des listes de mots-clés prédéfinis à l'aide de ReSIP ou d'Outlook.
- **Problème technique** : un traitement difficile ou impossible des messageries sans le client d'origine, en raison de plusieurs facteurs :
 - des problèmes de compatibilité entre le format de la messagerie et le logiciel utilisé pour effectuer le traitement (en particulier Outlook) ;
 - des postes de travail et des outils qui ne sont pas suffisamment puissants pour effectuer le traitement requis, en raison de la taille de l'archives ;
 - les erreurs contenues dans les messageries elles-mêmes.
- **Problème humain** : un manque de connaissances et une nécessité d'accroître les compétences sur les questions liées au traitement des messageries.

Premier problème : pistes de réponse

Archifiltre-Mails : développés par la Mission Archives auprès des Ministères sociaux.

Objectif : identifier rapidement les messages qui peuvent être supprimés.

Fonctionnalités :

- importer des messageries issues d'Outlook (sauf Office 365);
- explorer les messageries et afficher les messages par domaine mail, les contacts par domaine, les années par contact, puis les messages eux-mêmes ;
- exporter les messages au format .eml ;
- extraire les métadonnées et le contenu des messages sous forme de tableur aux formats .csv, .xlsx ou .json ;
- ajouter des tags « A supprimer » ou « A conserver » ;
- obtenir des informations statistiques.

Deuxième problème : une masse à préserver

- Les comptes de messagerie représentent une grande partie des archives numériques collectées ces dernières années par les services d'archives de l'administration centrale.

Ministère de l'Europe et des Affaires étrangères :

Les messageries constituent plus de 40 % du stock conservé dans le dépôt numérique SAPHIR.

Archives nationales :

Les messageries extraites depuis ReSIP représentent environ la moitié des métadonnées descriptives et techniques enregistrées dans la base de données pour moins de 0,5% des versements.

- Si le flux des versements se poursuit, combiné à une politique d'extraction des messageries depuis ReSIP, cela peut poser un problème de maintenabilité technique pour le système d'archivage.

Deuxième problème : pistes de réponse

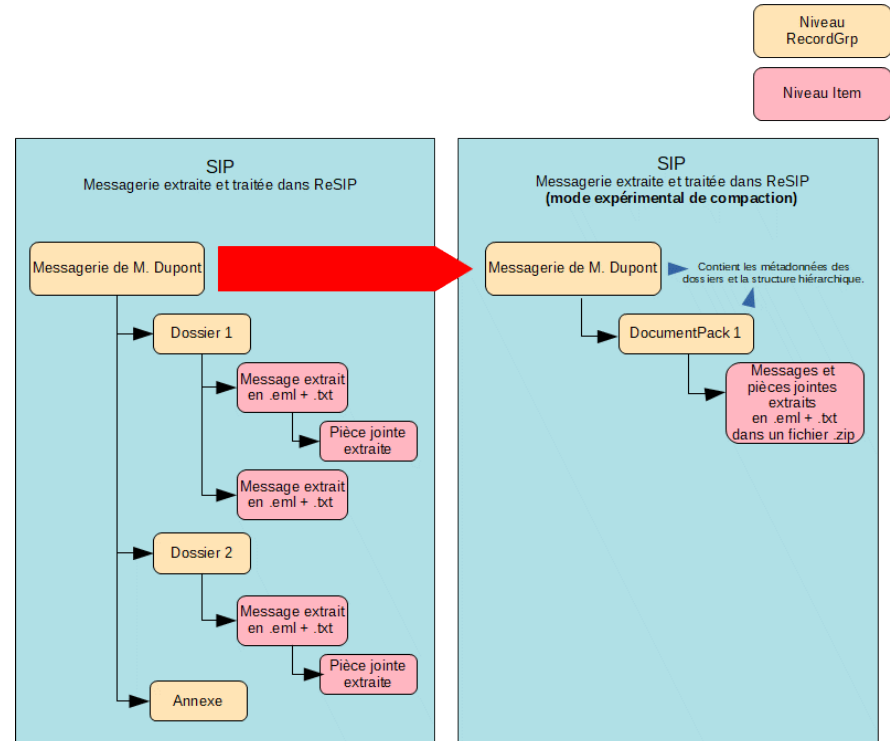
Nouvelle fonctionnalité de ReSIP :

Mode expérimental de compaction des messageries extraites (dossiers et messages convertis, pièces jointes) dans un fichier .zip.

Objectif : réduire le nombre de fichiers et de métadonnées extraits

Nouvelles questions :

- durabilité
- accès



Troisième problème : une masse à consulter

- De nouveaux types de demandes d'accès (loi CADA), portant sur des thèmes très larges ou sur l'occurrence d'un mot.
- Un accès difficile et manuel aux emails par import de la messagerie dans Outlook, ReSIP ou Archifiltre-Mails.
- Une difficulté à comprendre la logique organisationnelle de la messagerie, à évaluer les messages trouvés au regard du respect de la vie privée et des droits de communicabilité et à extraire ceux qui ont été identifiés comme communicables.

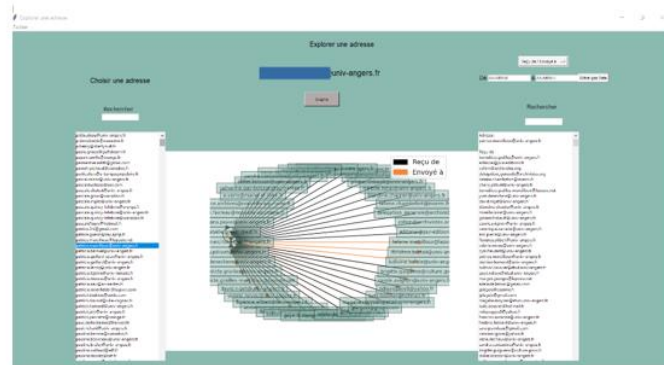
Troisième problème : pistes de réponse

- **Première réponse** : générer un fichier .csv à partir de ReSIP ou d'Archifiltre-Mails, qui inclurait les métadonnées des messages et serait archivé en même temps que la messagerie.
- **Deuxième réponse** : utiliser des outils existant tels que RATOM, epadd++, ou Pêle-Mél

Pêle-Mél:

Prototype d'exploration et de visualisation des comptes de messagerie acquis (format .eml uniquement) en utilisant des technologies d'intelligence artificielle et de machine learning.

Une équipe projet composée d'enseignants-chercheurs de l'Université d'Angers, et financée par le Ministère de la Culture.



Classification Interface

Mailboxes visualisation Interface

MERCI !



Contact / Nous suivre :

Directrice de la communication et des partenariats : alice.grippon@culture.gouv.fr

Experte fonctionnelle : marion.ville@culture.gouv.fr

47 rue de la Chapelle 75018 Paris, France

+33(0)186696025

programmevitam.fr

@ProgVitam

