The Archives nationales' mailboxes preservation workflow
Émeline Levasseur – emeline.levasseur@culture.gouv.fr
André Falut – andre.falut@culture.gouv.fr
Data migration and digital preservation managers

Département de l’administration des données – Department of Data Administration
Direction du numérique et de la conservation – Direction of Digital Support and Preservation
Archives nationales
1. Context
National Archives (France)

**Missions:** acquiring, arranging, preserving, restoring, valorizing and providing access to archives.

**Scope:** Historical public archives from the central administrations of State, archives of Paris notaries, private archives of national interest.
Organization of the archives network (national scale)

- **Appraisal**
  - Archives services of the ministries ("missions")
  - Archives services of the national operators, the high courts

- **Preparing transfer**
  - Interministerial Service of French Archives (SIAF)

- **Transferring**
  - Digital repository based on Vitam

- Creators (ministries, national operators, high courts)
  - Operational
  - IT

- Creating
- Collecting
- Records management

Preserving, giving access...
Technical and normative context

- The National Archives’ digital repository is based on Vitam (back-office) and interfaced with its own front-office (in-house archives managing software).
- In service since November, 2018.
- Goals: ensuring the ingestion, management and preservation of large amounts of digital records of administrative as well as archival value, while providing long-term access to these records.
- Data Exchange Standard for Archiving – SEDA (2006, 2022): models the data exchanges (electronic and paper) between archival agencies and their partners, translated as messages in XML.
2. The National Archives’ mailbox preservation workflow
Mailboxes in our archival holdings

• Digital archiving at the National Archives goes back to 1983...

• ... But the first mailbox entered our archival holdings in 2011. Our mailbox archiving strategy was developed in 2019 and applied to mailboxes collected from 2020 onwards.

• Reminder: appraisal, transfer preparation and ingestion are conducted by mission archivists, except for private archives

• Mailboxes are generally from ministries’ offices.

• In the digital repository: 41 PST files (representing 40 mailboxes) ingested.

• Outside the digital repository, pending processing and migration: 95 PST files, 7 mailboxes in Mbox.

• Majority of PST files, as Outlook is widely used by the central administrations of State.
Mailbox archiving workflow

Goals:

• Ensuring preservation of the messages themselves, their arrangement, their metadata and their attachments in their original format
• Making it possible to search and find messages in the digital repository
• Enabling access to the messages of a mailbox in a tree structure
• While leaving room for future options to represent mailboxes and their messages outside the digital repository (emulation?)
Mailbox archiving strategy

... In practice: 2 SIPS

1. The 1st contains the raw export (PST or Mbox) as received by the mission archivist

2. The 2nd contains:
   • The messages in EML format, rendered in their original tree structure
   • The body of each message in text format
   • The extracted attachments linked to their messages.
SIP preparation and mailbox extraction tool developed by the VITAM team.

Available on GitHub: [https://github.com/ProgrammeVitam/sedatools](https://github.com/ProgrammeVitam/sedatools)

To import a mailbox:
ReSIP: mailbox importation results

Extraction of the mailbox

A list of the messages
Fichier Contexte Traiter Import Export

Arbre du SIP (25663 archiveinit/25567 dog/40618 hdo/0 pdo)

Unité d'archive - ID13 ...

Niveau de description : Item

Titre : [Fwd: Re: Chantiers des Fonds - suite réunion du 19-10]

ID-SI prod : <472EDCA8.2040401@culture.gouv.fr>

Rédacteur : francoise.peyre, francoise.peyre@culture.gouv.fr

Destinataire : christine.noulet, christine.noulet@culture.gouv.fr

Date d'envoi : 5 novembre 2007 09:04

Date de réception : 5 novembre 2007 09:04

Recharger unité d'archive Sauve unité d'archive

Groupe d'objets - ID14 ...

Numérique - ID15 :

BinaryMaster_1, 472EDCA8.2040401@culture.gouv.fr, mes

Numérique - ID16 :

TextContent_1, 472EDCA8.2040401@culture.gouv.fr, txt, text
ReSIP: appraisal tools

Format statistics

Processing duplicates

Searching in the metadata (for example "confidential")
ReSIP: extraction settings

- **Edition des paramètres par défaut**

  **Import des messageries**
  - **Encodage par défaut des messageries**: windows-1252
  - **Extraction des fichiers textes des courriels**: des messages, des pièces jointes
  - **Extraction des métadonnées textuelles des courriels**: des messages, des pièces jointes

  **Import des hiérarchies sur disque**
  - **Fichiers exclus des imports**:
    - Thumbs.db
    - pagefile.sys
    - "\vcf
    - "\emz
    - "\dat
  - **Ignorer les liens symboliques et rassourcs**

  **Import/export des csv**
  - **Encodage des csv**: UTF-8
  - **Séparateur**: ;

- **Annuler** | **OK**
ReSIP: archival description and arrangement
Description de l'unité d'archive - ID10:

Titre:
Messagerie de Christine Noulet, archiviste aux Archives nacionales

Description:
Description of the mailbox

Date de début: 5 mars 2007
Date de fin: 26 novembre 2011
When processing is done... Exportation!

ReSIP: exportation
Two AIPs for a single mailbox:
View in the digital repository (search results)
Raw PST export AIP
ReSIP-processed AIP
ReSIP-processed AIP
3. Further prospects
Limitations

• Much of the process relies on the mission archivists and their IT service, who handle the exportation process from the client.

• Strategy relies on ReSIP… Importation errors occur often.

• Ingestion limited at 30 Gb. Require being careful for big mailboxes => several exports => several SIPs.

• Tree structure rendering is not always suitable.
Experiments regarding access

• First request for access on a mailbox! But...
• Legally, required an authorization, and therefore an investigation
• Technically, very specific case: not ingested yet, couldn't be imported in ReSIP...
• In order to make the investigation possible, experiments with Ratom and Epadd+ (with Emailchemy trial licence)
• Access was eventually denied
• But these experiments furthered our reflection regarding access
**Conclusion**

- Our strategy leaves space to consider potential future answers while preserving mailboxes and their messages.
- It can evolve...!
- More experiments needed to challenge this strategy (on access).
- ReSIP is also an option to answer request for access on processed and ingested mailboxes (DIP importation).
Thank you for your attention!