



# Archivage à Mathdoc

ANF RNBM - 21/11/2023

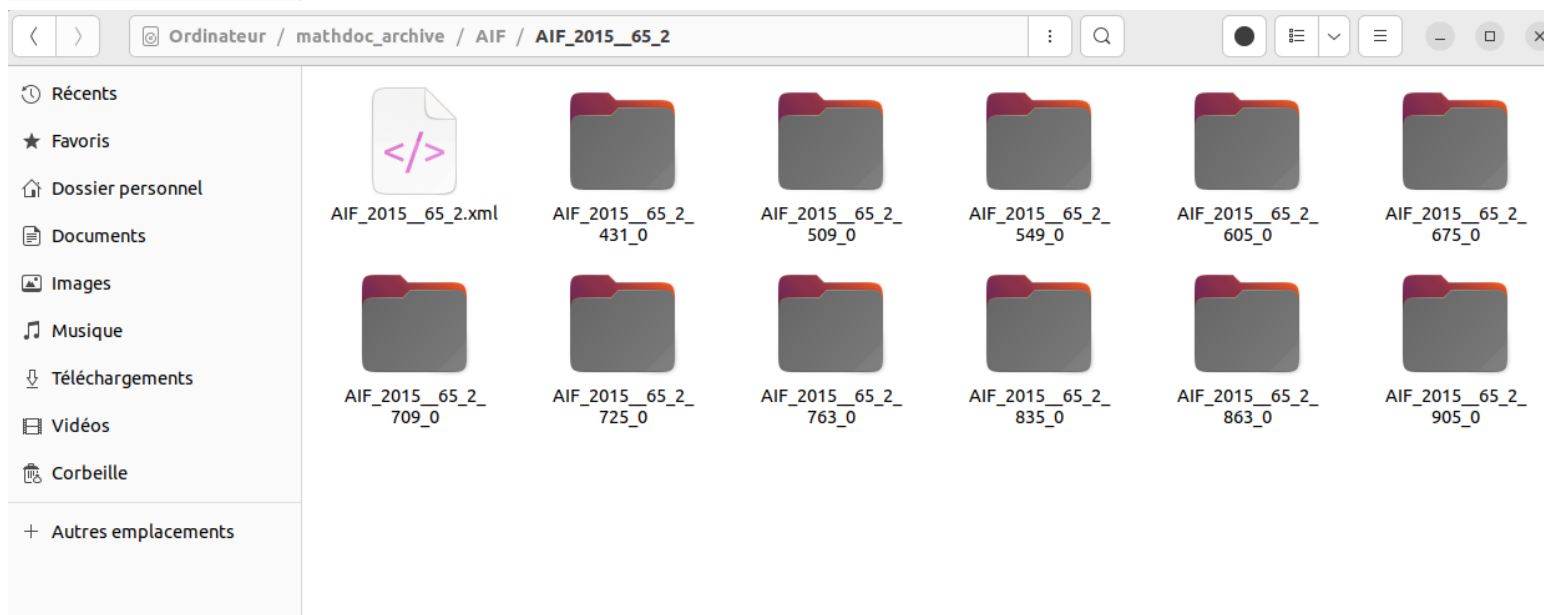
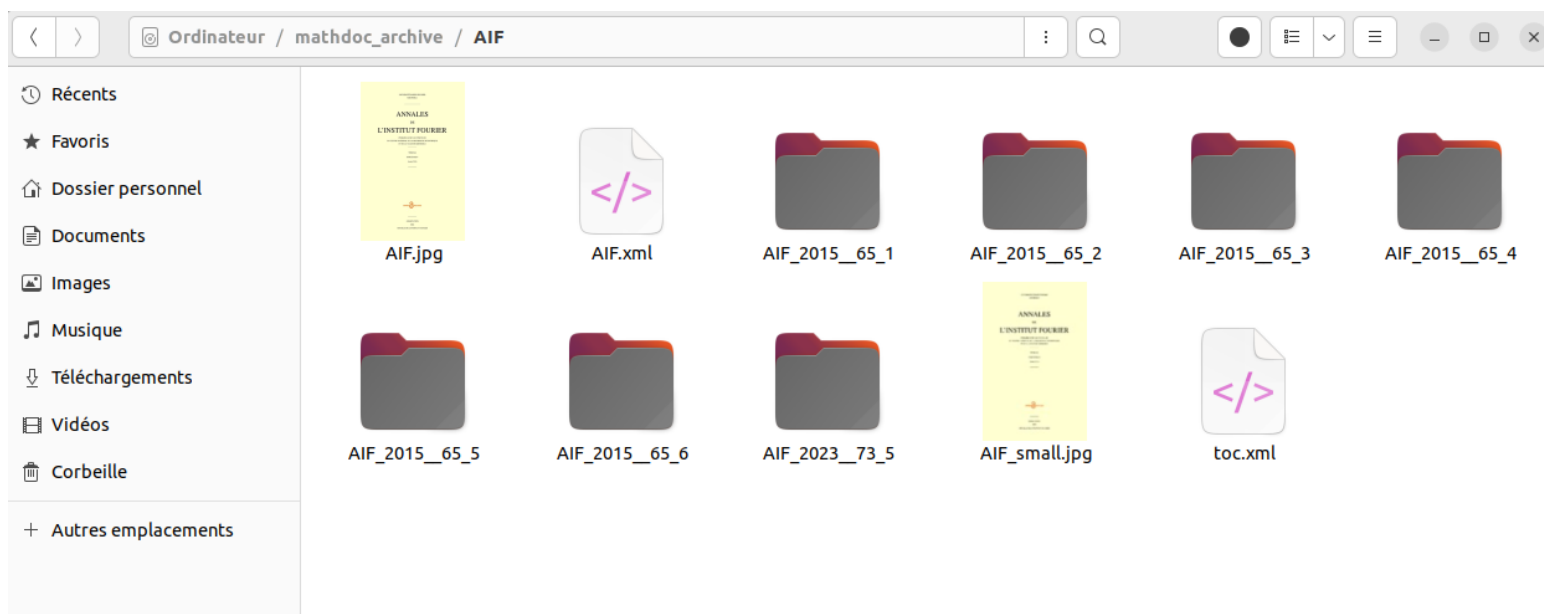
# Contenu sélectionné

- Livres et articles de Numdam et du centre Mersenne
- Données. Pour chaque article/livre :
  - PDF, PDF/A et DjVu
  - Matériel supplémentaire fournis par les auteurs
  - Images d'illustration
- Métadonnées. Pour chaque numéro/volume et livre :
  - XML au format JATS/BITS

# Contenu sélectionné

- Sources des documents du centre Mersenne :
  - TeX (fichier TeX, bib, ...)
  - Images utilisées dans le PDF
- Sources des documents numérisés :
  - Images scannées (tif, jpg)
  - XML du prestataire
- Sources des documents nativement numériques acquis par Numdam :
  - XML fournis par l'éditeur

# Contenu rangé



# Contenu rangé

AFST/

toc.xml

AFST.xml

AFST.jpg

AFST\_2022\_6\_31\_3/

AFST\_2022\_6\_31\_3.xml

AFST\_2022\_6\_31\_3\_861\_0/

AFST\_2022\_6\_31\_3\_861\_0.pdf

AFST\_2022\_6\_31\_3\_861\_0.djvu

src/tex/

# Contenu mis à jour régulièrement

On repart de zéro à chaque itération de Numdam

numdam en un après leur parution.



Cours de Jean-Pierre Serre



Compositio Mathematica



Publications Mathématiques de l'IHÉS



Comptes Rendus Mathématique

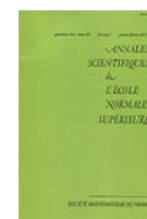


62138 articles dans 111 périodiques, 691 livres dans 5 c

parution.



Bulletin de la Société Mathématique de France



Annales scientifiques de l'École Normale Supérieure



Cours de Jean-Pierre Serre



Annales de mathématiques pures et appliquées



60930 articles dans 81 périodiques, 682 livres dans 5 c

numdam-pre



numdam.org

# Sauvegarde de l'archive

- Sur les serveurs de Numdam, du centre Mersenne
- Sur les serveurs de Mathrice
  - Grenoble
  - Angers
  - Lille
- Sur des disques durs externes
- Offre de stockage sur bande à l'étude

# Création de l'archive

- Automatiquement, de façon contrôlé, par une application
- Outil de vérifications, écrit avec du code différent de celui qui sert à créer l'archive
  - Vérification de la présence des fichiers
  - Cohérence entre le contenu des fichiers XML et les autres fichiers sur disque





# Utilisation de l'archive

- Par les développeurs pour recréer des sites web avec du contenu
  - Environ 1 jour pour un créer un site local (grâce à notre plateforme PTF de sites web)

# OAIS



- Modèle abstrait
- Spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage

<https://www.cines.fr/archivage/un-concept-des-problematiques/le-modele-de-reference-loais/>

# OAIS



« une information de structure qui explique la façon dont d'autres informations sont organisées (ex. : les tables de correspondance entre les noms de fichiers et les numéros de page pour un ouvrage numérisé) »

```
<article>
  <idart>AIF_1949__1__1_0</idart>
  <cphys>page0011.tif page0012.tif
page0013.tif page0014.tif</cphys>
</article>
```

# LOCKSS

- v1 : début des années 2000
  - v2.0-alpha: 04/2019
  - v2.0-alpha5 : 12/2021
  - v2.0-alpha6 : 01/2023
  - v2.0-alpha7 : 08/2023
- => Pas de version stable avant 2-3 ans ?
- « It is not yet possible » de convertir la v1
  - un prototype pour tester la migration, à des fins de test

# LOCKSS

## Moissonner les métadonnées

- Écriture de code XML et Java pour chaque éditeur
- CLOCKSS pour Mersenne : serveur FTP, puis site web local, moissonné par une instance LOCKSS

=> Même travail que Geodesic



# LOCKSS

Comment accéder aux données ?

CLOCKSS : « steps are required to extract the content from the directories or archive files that were copied from the CLOCKSS box repository, and generate from them a web site »

[http://documents.clockss.org/index.php/CLOCKSS:\\_Extracting\\_Triggered\\_Content](http://documents.clockss.org/index.php/CLOCKSS:_Extracting_Triggered_Content)



# Archivage pérenne pour collections RNBM

Geodesic étendu à des collections sous licences

- Gestion des droits d'accès (déjà fait par le Cedram)
- Stockage des PDFs