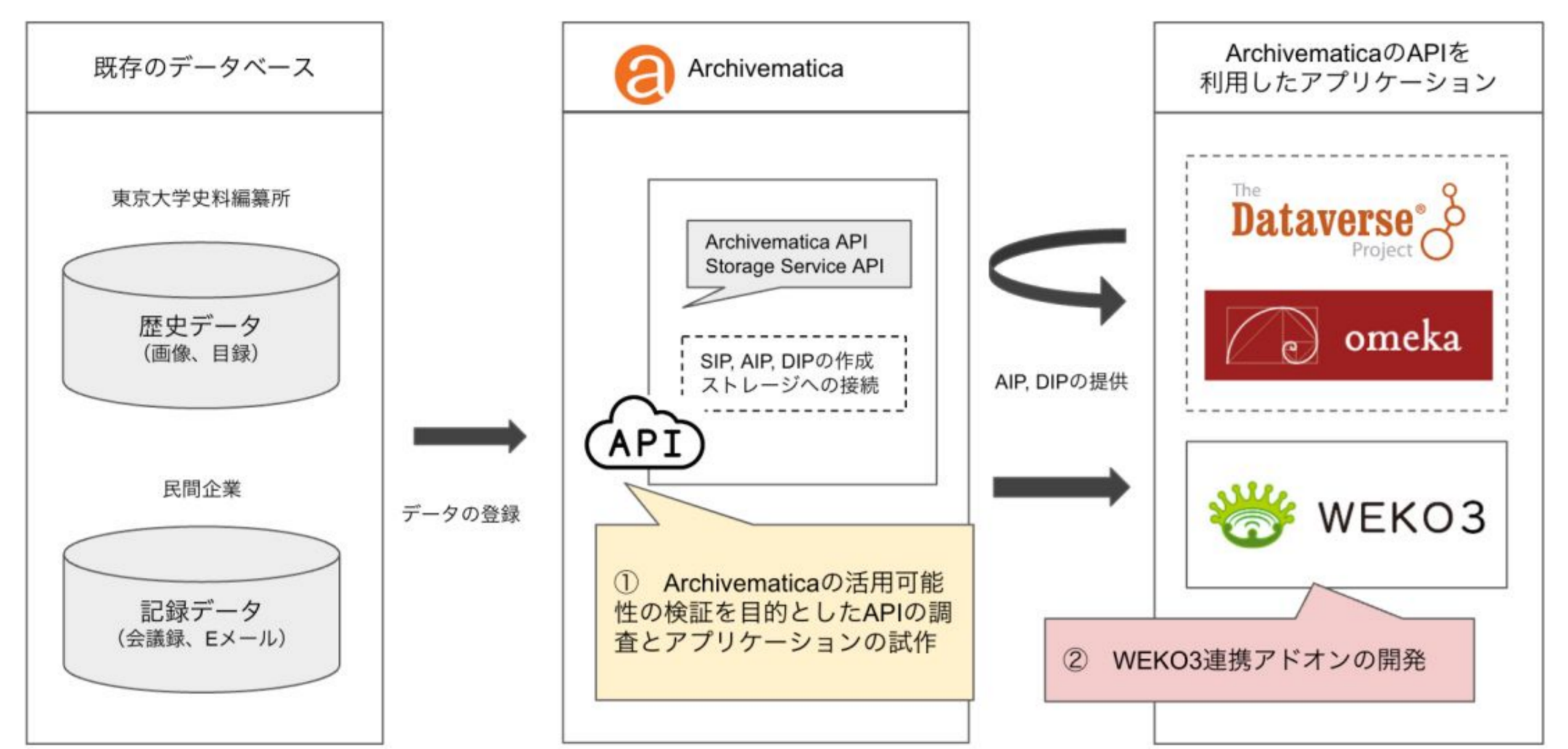


# Archivematica を用いた人文学データの キュレーションプロセスの自動化に関する試み

中村 覚 1), 金甫 榮 1)2), 南山 泰之 3), 山田 太造 1)  
1 東京大学 2 渋沢栄一記念財団 3 国立情報学研究所

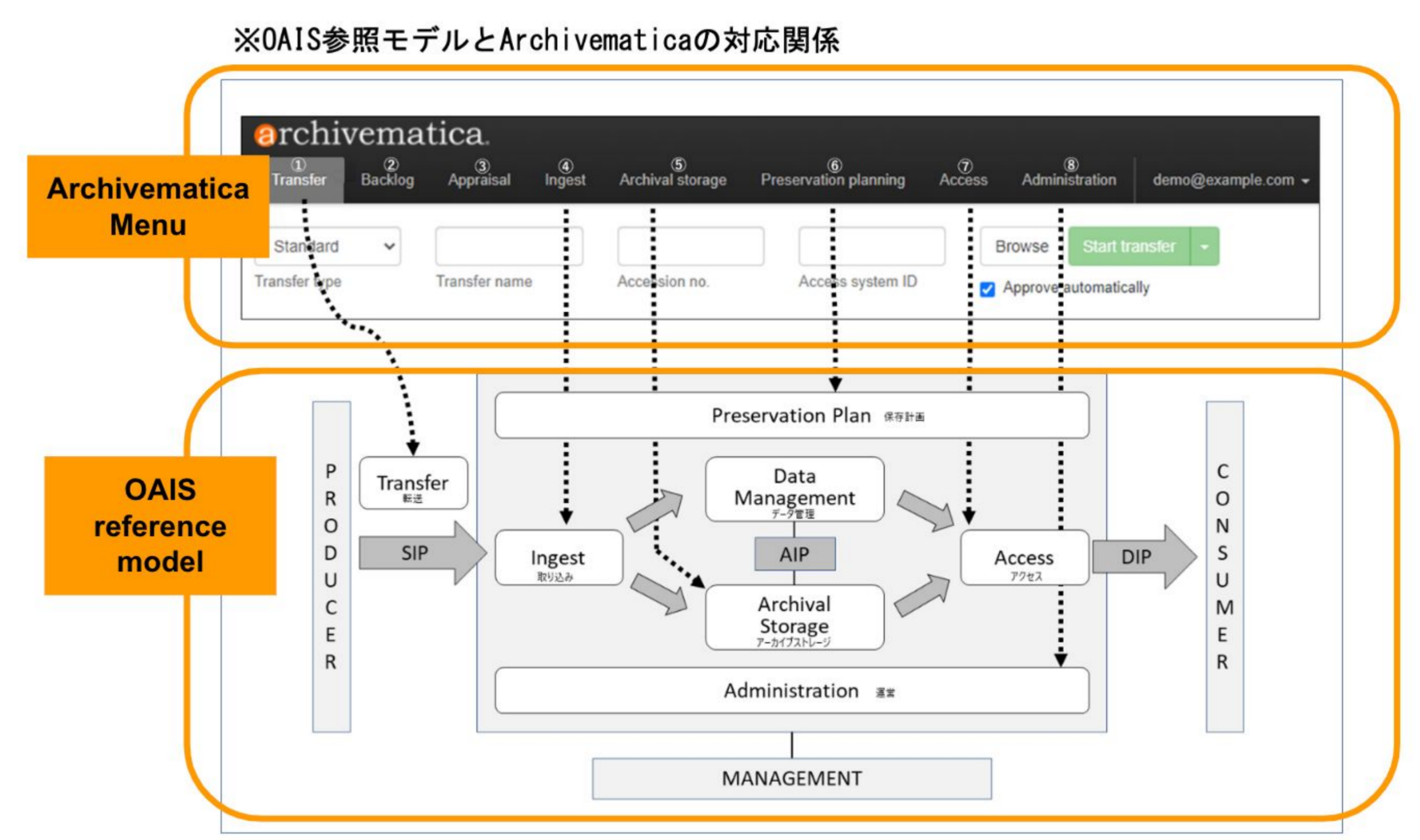
## 研究目的

- 研究データの長期的な利活用支援を目的とし、長期保存を支援する Archivematica を用いた人文学データのキュレーションプロセスの自動化に取り組む。
- 具体的には、WEKO3 との連携プロトタイプを開発し、人文学系機関などが有するデータで実験する。
- 人文学データの長期的な利活用に必要な具体的な技術要件を明らかにしつつ、NII-RDC で開発予定のデータキュレーション支援機能に対して、人文学データによる実証事例を提供する。



## Archivematica

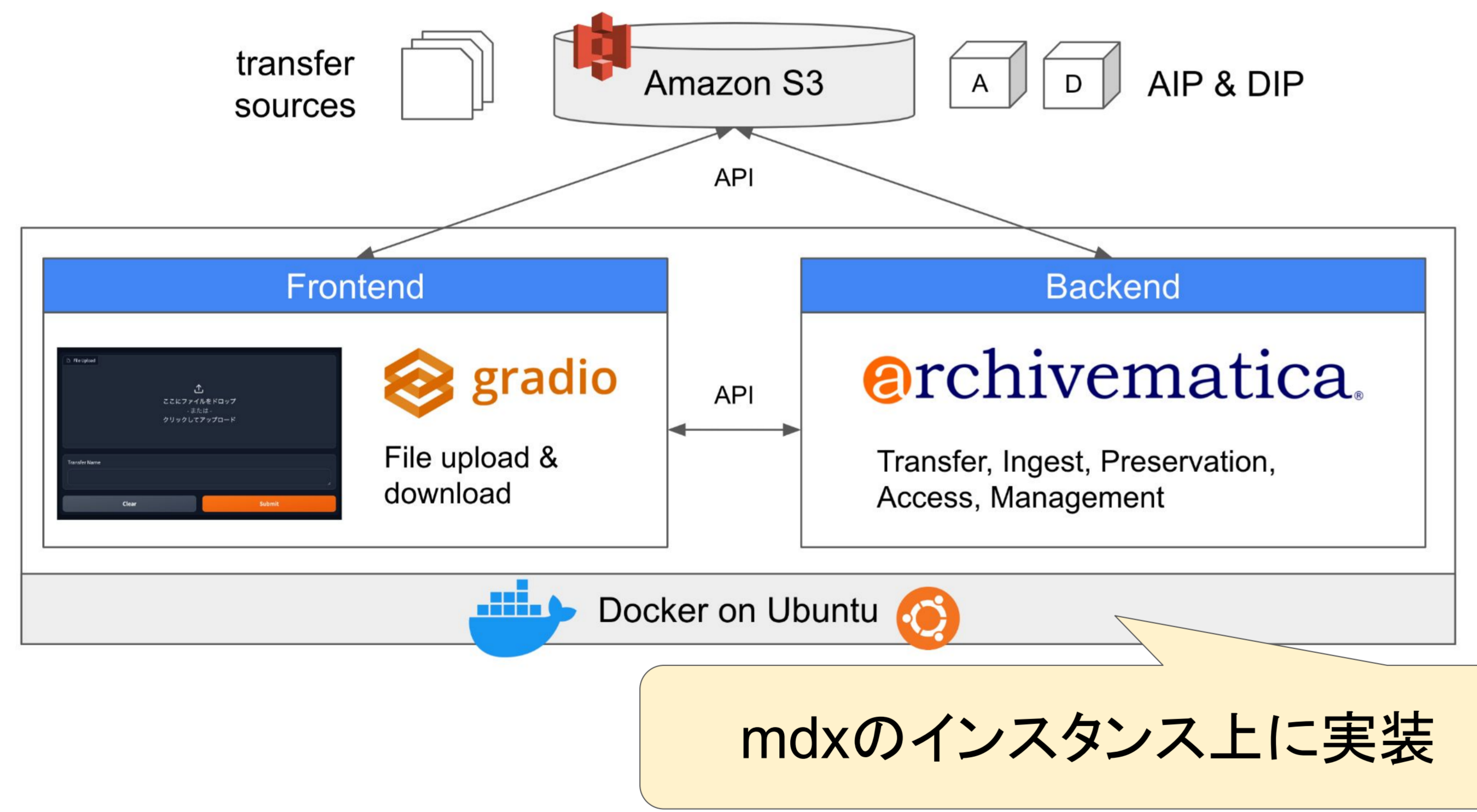
Archivematicaは、デジタルデータの長期保存に特化したオープンソースであり(OAIS参照モデル準拠)、多様なデータフォーマットの識別と変換、階層構造を持つデータのパッケージ化のほか、各種リポジトリソフトウェア(DSpace、Dataverseなど)との連携といった豊富な機能を持つ。



## 取組1: 簡易アプリケーションの開発

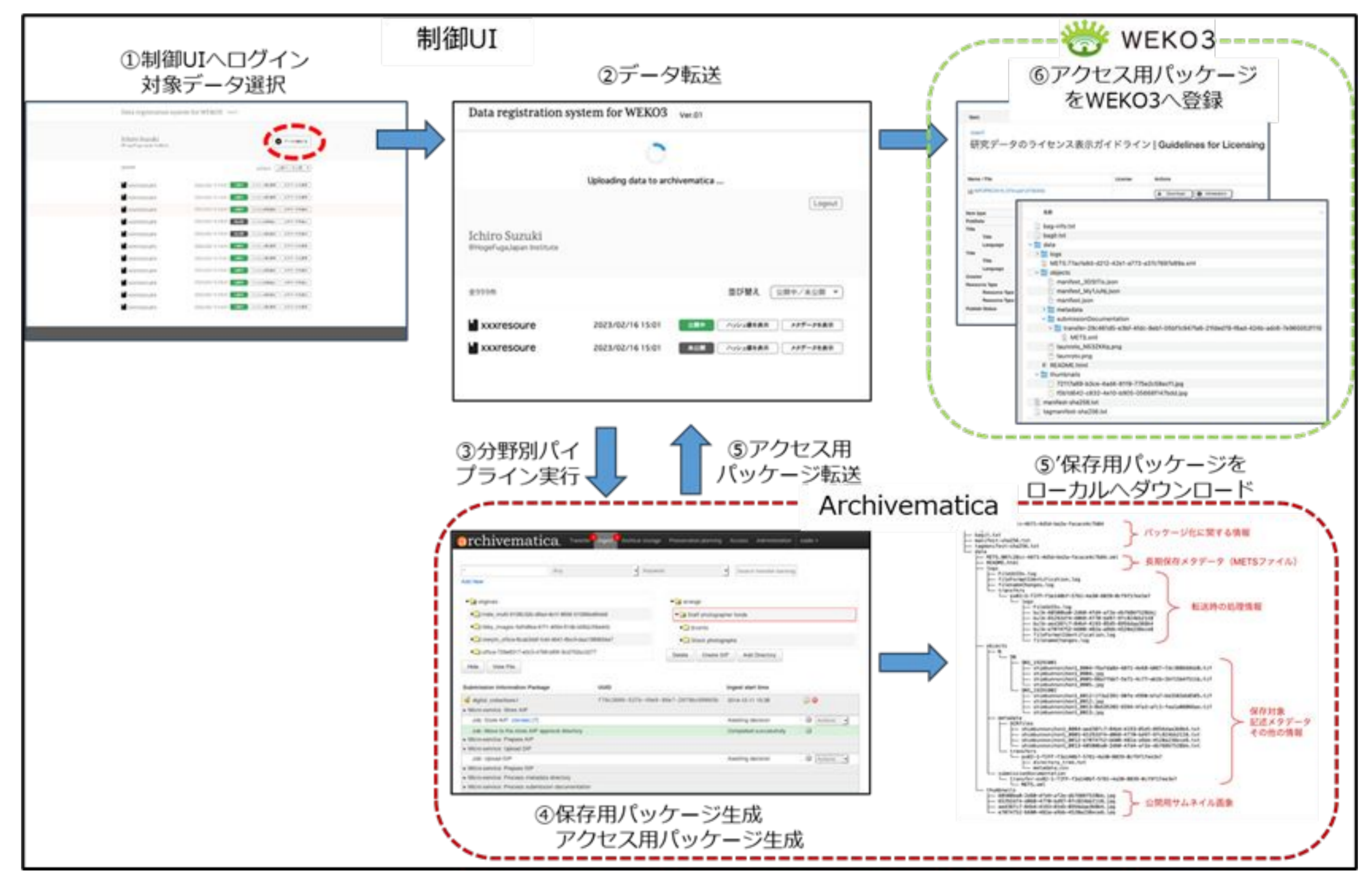
Archivematicaを用いたデータキュレーションプロセス自動化のためのプロトタイプシステム

ArchivematicaのAPIを利用して、Archivematicaの複雑な設定やカスタマイズを自動的に補完するアプリケーションを開発した。



## 取組2: WEKO3との連携

- 取組1での経験をもとに、WEKO3への登録フローを開発
- 長期保存に必要なパッケージとアクセス用パッケージを自動生成し、研究者による作業を軽減



## 取組3: 東京大学史料編纂所における実験

← Archivematicaが作成したAIPの例

```
final-e38b64d7-1386-4032-97a9-faacf945495b
bag-info.txt
bagit.txt
data
  METS_e38b64d7-1386-4032-97a9-faacf945495b.xml
  README.html
  logs
    FileUIDs.log
    fileFormatIdentification.log
    filenameChanges.log
  transfers
    final-cb0e1e2d-4619-4bf5-94ad-0e7ef2324eb5
      logs
        FileUIDs.log
        fileFormatIdentification.log
        filenameChanges.log
  objects
    BD2023-009700
    00000004-1b26d652-cab1-4db5-b779-c4692737a629.tif
    00000004.jpg
    00000005-3d68e011-1c7b-4a98-97b3-15617184ccf7.tif
    00000005.jpg
  metadata
    transfers
      final-cb0e1e2d-4619-4bf5-94ad-0e7ef2324eb5
        directory_tree.txt
        metadata.csv
    submissionDocumentation
      transfer-final-cb0e1e2d-4619-4bf5-94ad-0e7ef2324eb5
        20210721_tomishiroshi_obogaki.pdf
        2023nen7gatsu_tomishiroshimazuteikyokasho.pdf
        METS.xml
        _besshi_satsueikiboushiyou_tomishiroshimazutei2023nen7gatsu_.xlsx
  thumbnails
    a394bd3f-1fa4-4d7f-a81f-3d2276dafa3f.jpg
    b171db09-3f0a-4e5f-ae38-115ecbb0a1a2.jpg
  manifest-sha256.txt
  tagmanifest-sha256.txt
```

- 東京大学史料編纂所が保有するデータの例
- Archivematicaが出力するAIPの内容分析実施し、保存用メタデータとして有効性を検討 (じんもんこん2024に投稿中)

## 取組4: 日本での普及に向けて

- Archivematicaのファイル名変換において、漢字や平仮名などが日本語として認識されない問題の解決
- Archivematicaの英語マニュアルの日本語訳の作成

