

AIPを作成・利用する ツール紹介

中村覚(東京大学史料編纂所)

AIPを作成するために使用できる代表的なツール

オープンソースソフトウェア

- [Archivematica](#)
 - デジタルアーカイブに適合するためのAIPを生成するプロセスを自動化
- DSpace
 - アカデミック、非営利、商業組織向けのオープンソースのリポジトリソフトウェア
 - <https://wiki.lyrasis.org/display/DSDOC8x/DSpace+AIP+Format>

サービス

- Rosetta
 - Ex Librisによって開発されたデジタルアセット管理と保存のためのソリューション
 - https://knowledge.exlibrisgroup.com/Rosetta/Product_Documentation/Rosetta_AIP_Data_Model/01_Introduction_to_Rosetta_AIP_Data_Model
- AMLAD®
 - デジタルアーカイブソリューション | NTT DATA <https://www.kantei.go.jp/jp/singi/titeki2/forum/2017/siryous8.pdf#page=9>

関連発表およびプロジェクト

関連発表

- 中村覚, 金甫榮, 南山泰之, Archivematicaを活用したデジタルデータの長期保存を支援する簡易操作アプリケーションの開発, デジタルアーカイブ学会誌 7(S2) s107-s110 2023年10月, 10.24506/jsda.7.s2_s107
- 金甫榮, 中村覚, 渡邊英徳, [B2] 真正なデジタル化資料の長期保存と公開: ArchivematicaとOmeka Sを用いた事例, デジタルアーカイブ学会誌 6(s3) s147-s150 2022年11月, 10.24506/jsda.6.s3_s147

プロジェクト

- 中村覚, 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, 研究データエコシステム構築事業, 2023-2024年度

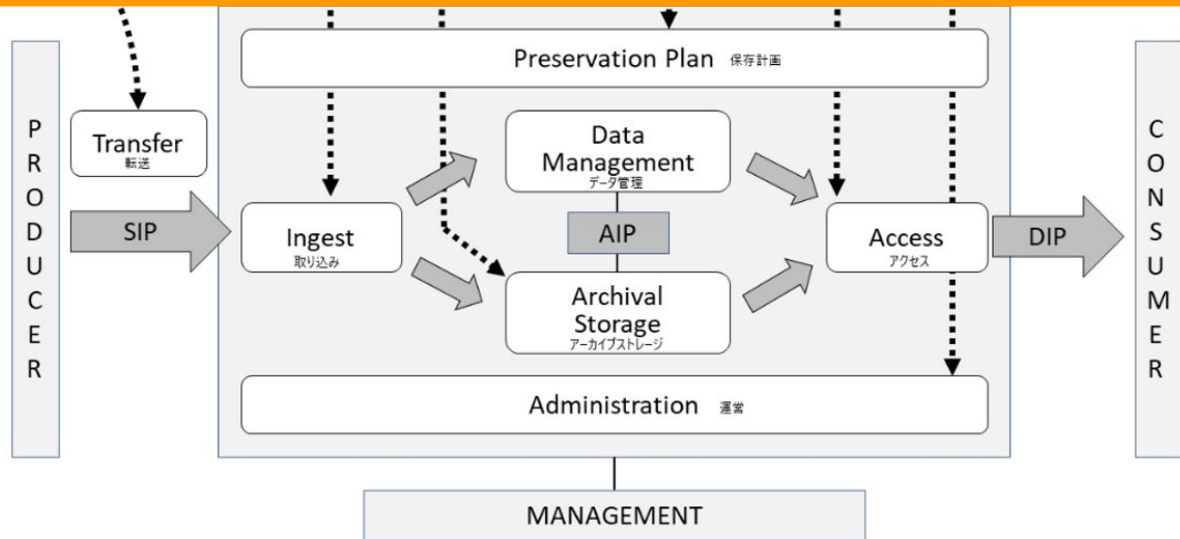
Archivematicaを活用したデジタルデータの長期保存 を支援する簡易操作アプリケーションの開発

※OAIS参照モデルとArchivematicaの対応関係

Archivematica
メニュー

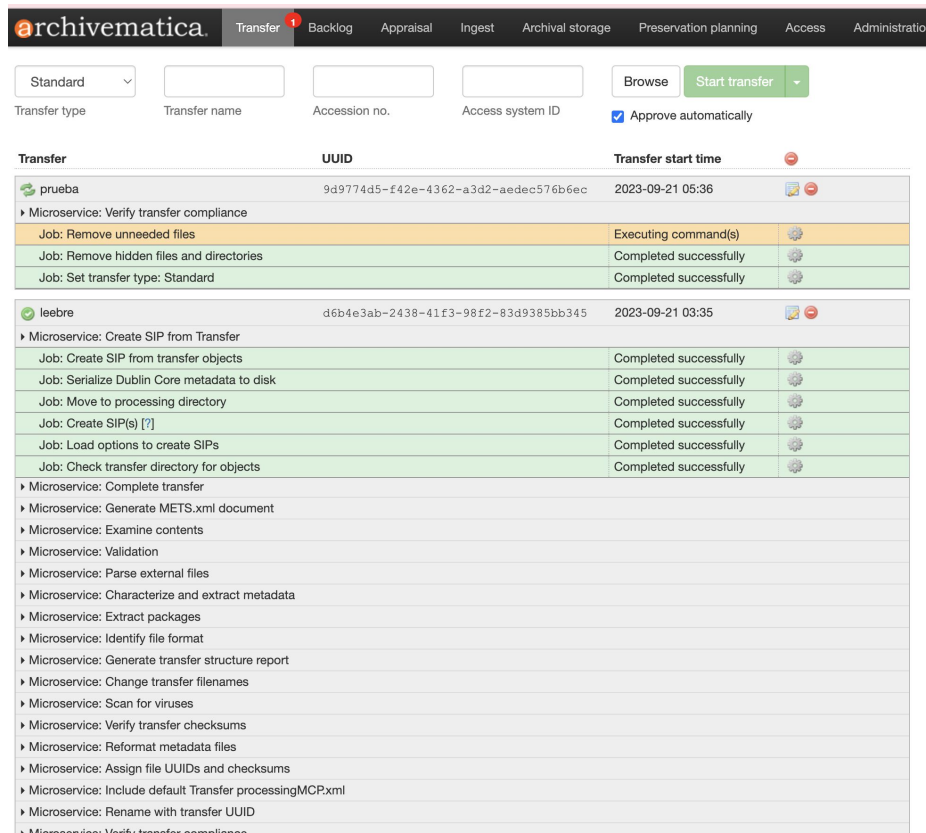


OAIS
参照モデル
















Archivematicaの課題

- 豊富な機能が提供されている => 操作方法になれるまで時間を要する
 - AIP (Archival Information Package) の出力内容を確認してみる、といった試行には不向き
- APIを介したシステム連携が進んでいるが、ArchivematicaのAPIに関する(特に日本語の)情報が不足している



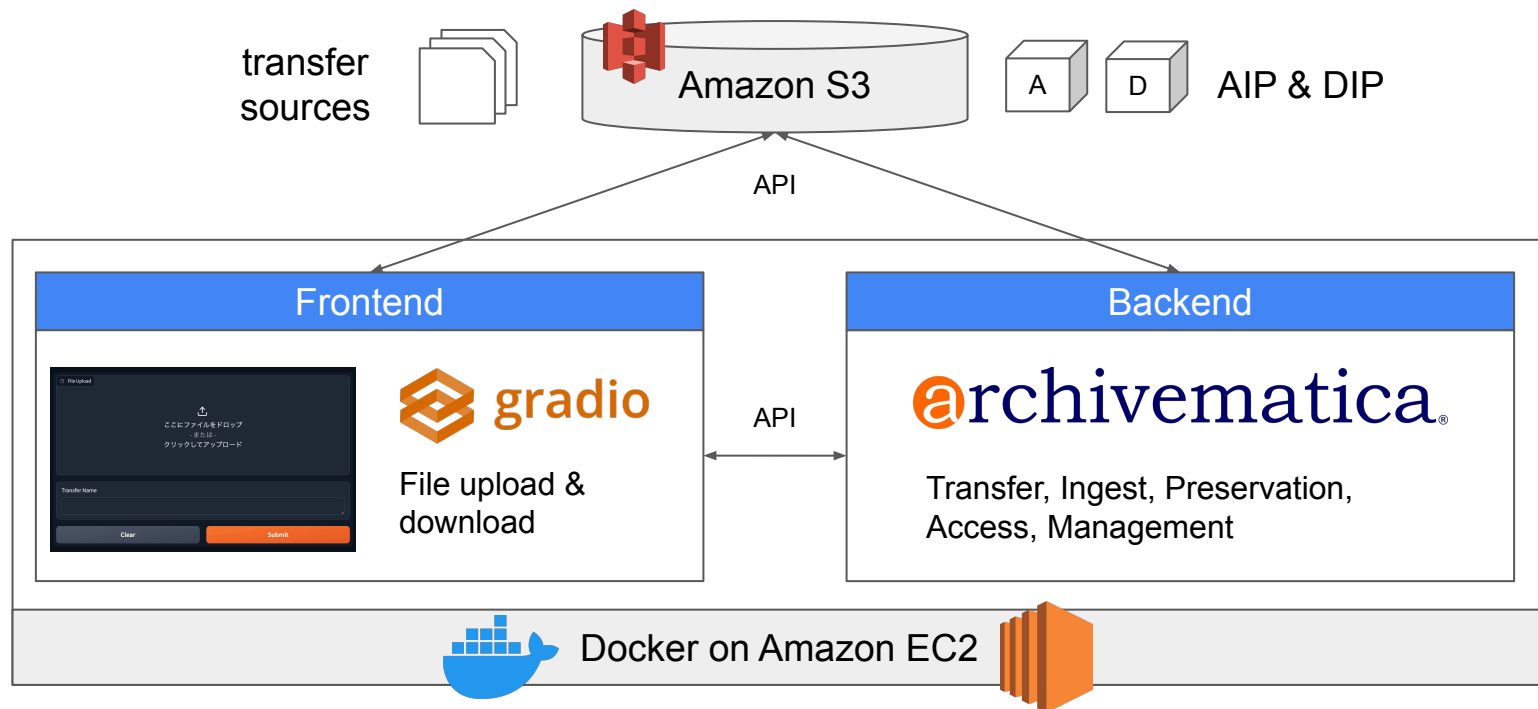
The screenshot displays the Archivematica web interface. At the top, there is a navigation bar with the Archivematica logo and several menu items: Transfer (with a red notification icon), Backlog, Appraisal, Ingest, Archival storage, Preservation planning, Access, and Administration. Below the navigation bar, there are input fields for Transfer type (set to Standard), Transfer name, Accession no., and Access system ID. There are also buttons for Browse, Start transfer, and a checkbox for Approve automatically (checked).

Transfer	UUID	Transfer start time	
prueba	9d9774d5-f42e-4362-a3d2-aedec576b6ec	2023-09-21 05:36	 
Microservice: Verify transfer compliance			
Job: Remove unneeded files		Executing command(s)	
Job: Remove hidden files and directories		Completed successfully	
Job: Set transfer type: Standard		Completed successfully	
leebre	d6b4e3ab-2438-41f3-98f2-83d9385bb345	2023-09-21 03:35	 
Microservice: Create SIP from Transfer			
Job: Create SIP from transfer objects		Completed successfully	
Job: Serialize Dublin Core metadata to disk		Completed successfully	
Job: Move to processing directory		Completed successfully	
Job: Create SIP(s) [?]		Completed successfully	
Job: Load options to create SIPs		Completed successfully	
Job: Check transfer directory for objects		Completed successfully	
Microservice: Complete transfer			
Microservice: Generate METS.xml document			
Microservice: Examine contents			
Microservice: Validation			
Microservice: Parse external files			
Microservice: Characterize and extract metadata			
Microservice: Extract packages			
Microservice: Identify file format			
Microservice: Generate transfer structure report			
Microservice: Change transfer filenames			
Microservice: Scan for viruses			
Microservice: Verify transfer checksums			
Microservice: Reformat metadata files			
Microservice: Assign file UUIDs and checksums			
Microservice: Include default Transfer processingMCP.xml			
Microservice: Rename with transfer UUID			
Microservice: Verify transfer compliance			

本研究の目的

- Archivemáticaの複雑な設定やカスタマイズの部分を自動的に補完するアプリケーションを開発する
 - Archivemática導入における専門的な知識の必要性、複雑な設定とカスタマイズといった障壁を軽減する
- 利用者からのフィードバックを通じて、本アプリケーションが長期保存に必要な作業の迅速化と作業者の負担軽減に寄与することを確認する
 - 上記の開発を通じて、Archivemática(特にそのAPI)に関する情報発信を行う

システム構成



Archivematica Gradio Demo

A simple demonstration of Archivematica and Gradio.

2ステップ

- ファイルのアップロード
- AIPファイルのダウンロード

File Upload ×

ja.zip 47.0 KB ↓

Transfer Name

Japanese Example

Clear Submit

☰ Examples

File Upload	Transfer Name
Images.zip	Images Example
ja.zip	Japanese Example

Satoru Nakamura, Boyoung Kim, Yasuyuki Minamiyama.

Development of a User-friendly Application to Support Long-term Digital Preservation Using Archivematica,

18th International Digital Curation Conference, 2024.

関連ツール

METSFlask <https://github.com/tw4l/METSFlask>

- ArchivematicaのMETSファイルを手間を省く方法で探索するためのウェブアプリケーション

参考記事

- <https://zenn.dev/nakamura196/articles/a5f7cd460195bd>

METSFlask

Home / METS.b341fe0f-25be-4ebf-b29b-f2dc50d17bc9.xml

METS File: METS.b341fe0f-25be-4ebf-b29b-f2dc50d17bc9.xml
AIP UUID: b341fe0f-25be-4ebf-b29b-f2dc50d17bc9

File format visualization

Original files (count: 1) [Hide visualization](#)

Show 10 entries

Filepath	Format	PUID	Size (bytes)	Last modified	Details
ユースケース公募提案書.docx	Microsoft Word (2007+)	fmt/412	51332	2024-02-26T03:34:17Z	View

Showing 1 to 1 of 1 entries

METSFlask [Upload](#)

Home / METS.b341fe0f-25be-4ebf-b29b-f2dc50d17bc9.xml / ユースケース公募提案書.docx

File: ユースケース公募提案書.docx

File details

Filepath: ユースケース公募提案書.docx
UUID: f48c30f3-c1d1-4ffc-a6d8-8fd98b152d66
Last modified: 2024-02-26T03:34:17
Size: 50 KB (51332 bytes)
File format: Microsoft Word
Version: 2007+
PRONOM ID (PUID): [fmt/412](#)
amdSec ID: amdSec_2
sha256 hash: 8e3c38e590e15f38b953011bcb3a379a857408520ed9547b3fac77fbb151ae2

PREMIS Events

Event UUID: 24741142-467a-45da-936e-78e43ab68a6c

Event type: ingestion
Event datatime: 2024-02-26T03:34:19.082563+00:00
Event detail: None
Event outcome: None
Event detail note: None

Event UUID: 52529f18-5818-4045-b505-00fe918ea50e
Event type: message_digest_calculation

AIPscan <https://github.com/artefactual-labs/AIPscan>

- 詳細なレポーティングソリューションを提供する
- ArchivematicaストレージサービスのAIPからMETSファイルをクローリングして、リポジトリ保有物についての表形式および視覚的レポートを生成する

AIPscan Archivermatica Storage Services AIPs Reports API *Crawl Archivematica AIPs to provide repository-wide reporting*

File: yuusukesukouboteiansho.docx

Filepath	objects/yuusukesukouboteiansho.docx
AIP	Japanese_Example-2698386a-ed98-4833-b409-ed198ffd5c4a
File type	original
UUID	c5635980-754d-4935-a504-9010a9592a37
Size	51.3 kB
Last modified date	2024-02-06
PUID	fmt/412
Format	Microsoft Word (2007+)
Checksum type	sha256
Checksum value	8e3c38e590e15f38b953011bcb3a379a857408520ed9547b3fac77fbb151aef2
PREMIS Object XML	<pre><?xml version="1.0" encoding="UTF-8" standalone="no" ?> <ns0:mdWrap xmlns:ns0="http://www.loc.gov/METS/" NDTYPE="PREMIS:OBJECT"> <ns0:xmlData> <premis:object xmlns:premis="http://www.loc.gov/premis/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="premis:file" xsi:schemaLocation="http://www.loc.gov/premis/v3/premis.xsd" > <premis:objectIdentifier> <premis:objectIdentifierType>UUID</premis:objectIdentifierType> <premis:objectIdentifierValue>c5635980-754d-4935-a504-9010a9592a37</premis:objectIdentifierValue> </premis:objectIdentifier> <premis:objectCharacteristics> <premis:compositionLevel>0</premis:compositionLevel> </premis:objectCharacteristics> </premis:object> </ns0:xmlData> </ns0:mdWrap></pre>

Report: file formats count
Storage Service: test
213 original files ingested between 2024-01-01 and 2024-02-24
11 different formats

Legend: JPEG, Transaction TGA Stream, Microsoft Word, HTML, Portable Network Graphics, XML, Plain Text, MPEG-10 Audio Layer 3, Exchangeable Image File Format (Compressed), Tagged Image File Format, Microsoft Rich Text

参考記事

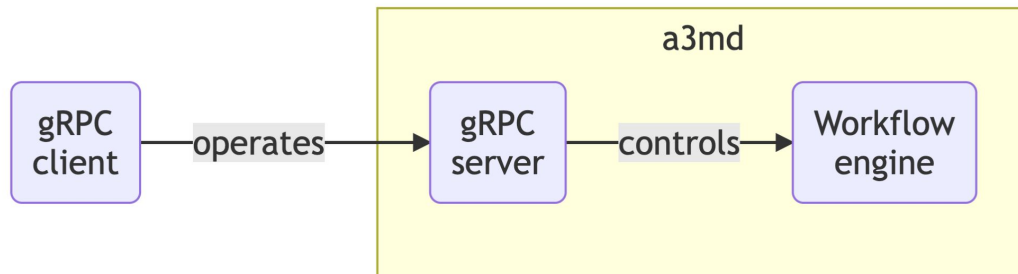
- <https://zenn.dev/nakamura196/articles/2917442dd6b9f5>

a3m <https://github.com/artefactual-labs/a3m>

- Archivematicaの軽量版で、AIPの作成に特化したツール
- 他のアプリケーションとの連携が容易となることが期待される

参考記事

- <https://zenn.dev/nakamura196/articles/7d0fd142f92a63>



日本語対応への取り組み

Archivemataのマニュアルの日本語化

- 試験公開版
 - <https://nakamura196.github.io/archivemata-docs/>
- 公式サイトに取り入れていただくため、開発元 (Artifactual社) と調整中

Archivemata 1.15.1 ドキュメント » Archivemata文書

Archivemata文書

はじめに

本マニュアルは、Archivemataをデジタル・ブリザベーション・ソリューションとして導入することを検討しているユーザーを対象としています。本マニュアルには、Archivemataプロジェクトに関する情報や、新規ユーザーが試用するためのクイックスタートガイドが含まれています。生産環境でのArchivemataのインストールや使用方法については、以下のユーザーマニュアルまたは管理者マニュアルを参照してください。

概要

概要では、Archivemataの定義、ベースとなっているOAISモデル、テクニカルアーキテクチャ、マイクロサービスのフレームワーク、Archivemataを導入するためのシステム要件について説明します。

- Archivemataとは？
- ウェブベースのダッシュボード
- 技術アーキテクチャ
- マイクロサービス
- 外部ツール
- システム要件

クイックスタートガイド

このセクションでは、Archivemataをテストするために、Archivemata Sandboxまたはローカルの仮想マシンを使用して、Archivemataをインストールする方法を説明します。完全なインストール手順については、Administratorマニュアルを参照してください。

- Archivemataクイックスタートガイド
- 設定

目次

- Archivemata文書
 - はじめに
 - 概要
 - クイックスタートガイド
 - トラブルシューティング
 - その他のリソース
 - ユーザーマニュアル
 - 転送
 - Backlog
 - 評価
 - インジェスト
 - Archival Storage
 - アクセス
 - 保存計画
 - 管理
 - メタデータ
 - 翻訳
 - 管理者マニュアル
 - インストールとセットアップ
 - メンテナンス
 - セキュリティ
 - 開発者マニュアル

前のトピックへ

Archivemata文書の内容

次のトピックへ

はじめに

We/The authors thank Crimson Interactive Japan – www.crimsonjapan.co.jp for their assistance in manuscript translation and editing.

日本語ファイル名の変換

課題

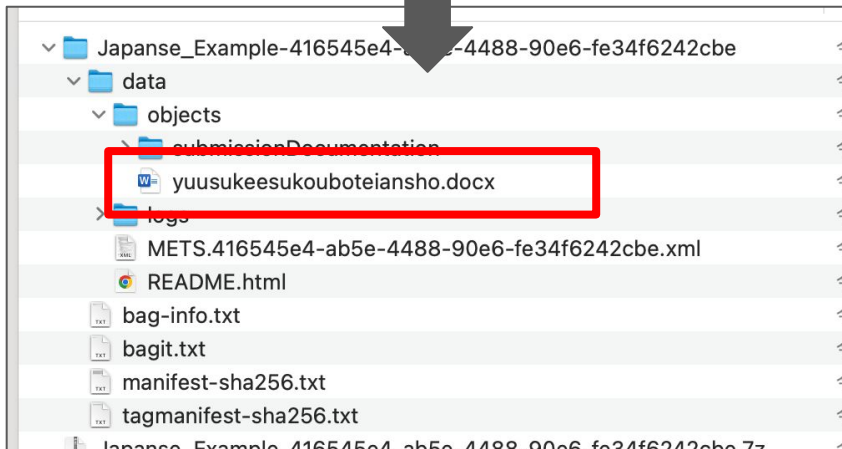
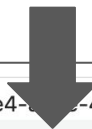
- 日本語のファイル名を持つファイルを入力すると、一見わかりにく名前に変換される
- 例: ユースケース公募提案書 .docx

取り組み

- この原因を調査し、対処方法を共有

参考

- <https://zenn.dev/nakamura196/articles/d2a2b323590f12>



ご清聴ありがとうございました。