

2023年2月6日開催

第6回東京大学学術資産アーカイブ化推進室主催セミナー

デジタル資料の保存管理の実際

# 東京大学での取り組み

## 学術資産アーカイブ化推進室

学術資産アーカイブ化推進室/情報システム部情報基盤課学術情報チーム（デジタル・ライブラリ担当）

田口 忠祐

# 東京大学デジタルアーカイブズ構築事業の概要

## 事業目的

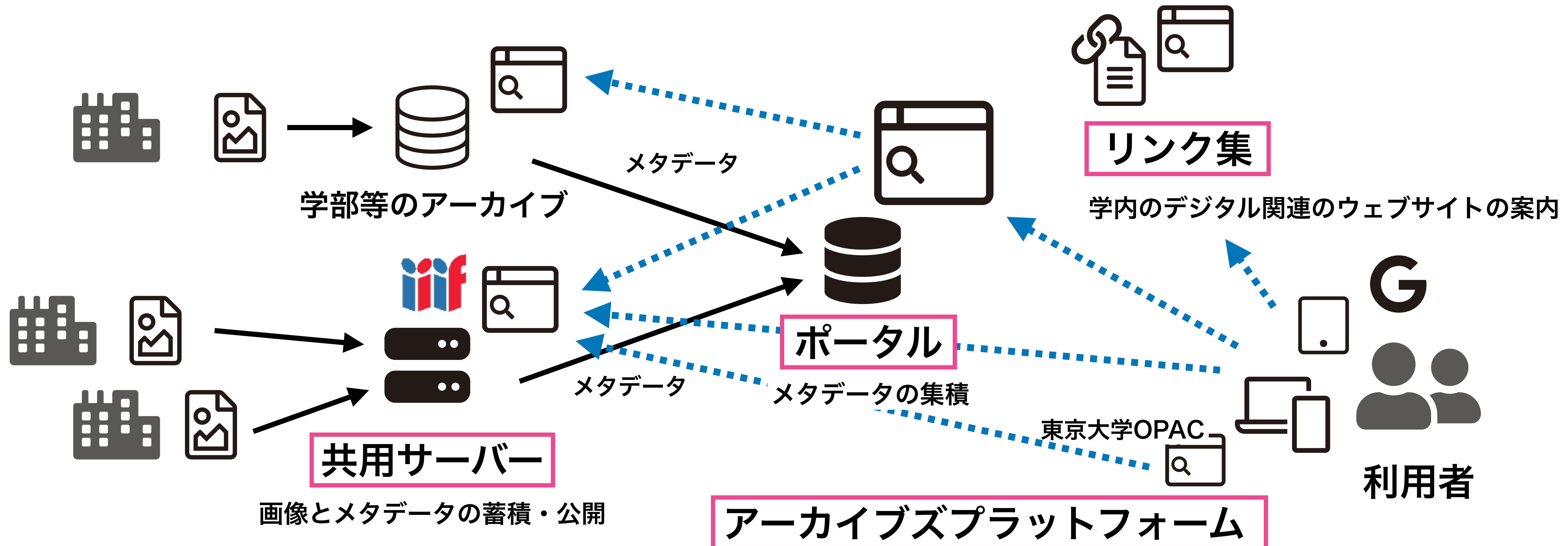
- 学内の多様な学術資産等のデジタルアーカイブ化を行い、かつ、国内外に向けて広く公開し、その活用を促進する。
- 本学の目指すべき理念や方向性をめぐる基本方針を定めた「[UTokyo Compass](#)」の目標の一つにもなっている。

## 経緯

- 2016年に東京大学ビジョン2020に基づき、学術資産等アーカイブズ委員会が設置
- 2017年に附属図書館に学術資産アーカイブ化推進室(教員+職員)が設置
  - デジタル・ライブラリ担当もメンバーとして公開システムの構築を支援
- 電子版貴重書サイトをアーカイブ化推進室の運営するシステムに順次移行
  - 統一的なシステムのため、利便性、効率性が向上

# 東京大学学術資産等アーカイブズプラットフォームとは

- ・ アーカイブズ構築事業の目的を実現するためのシステム
- ・ リンク集、共用サーバー、ポータルの総称



# バックアップの対象

- システムプログラム及び設定ファイル
- 共用サーバーの登録データ
- メタデータ
- 保存用高精度画像
- アクセスログ

# バックアップに利用しているストレージ等

## NAS (Network Attached Storage)

- Buffalo TeraStation 64TB x 1台 (RAID5 : 実効容量約50TB)
- Buffalo TeraStation 120TB x 1台 (RAID6 : 実効容量約84TB)

## テープドライブ

- HP StoreEver LT07 Ultrium 15000 (テープ容量15TB)

## オンプレミスサーバー

- HP DL160 Gen9 (CPU36コア、RAM128GB、HDD1TB x 4)

## 外部システム

- GitHub
- Internet Archive

# アーカイブズポータルへのバックアップ

アーカイブズポータルには画像ファイルがないため、メタデータのみをバックアップしている。

- バックアップ対象と方法
  - メタデータをCSV形式でオンプレミスサーバーに保存
  - cronによる自動処理
  - 実施頻度は月1回

# 共用サーバーのバックアップ

共用サーバーでは、デジタルアーカイブシステム（Omeka S）と  
IIIF画像サーバー（IIP Image Server）が稼働

- バックアップ対象と方法
  - 必要となるディレクトリをrsyncによりNASに2世代保存
  - データベースダンプ
  - メタデータ（Excel、CSV、RDF（JSON-LD））をGitHubの非公開レポジトリに保存
  - cronによる自動処理
  - 実施頻度は月1回

# Internet Archiveとの連携

共用サーバーに登録しているコレクションの一部をInternet Archiveへ提供

- バックアップ対象と方法
  - 総合図書館資料のうち自由利用可能なIIFで公開するコレクション
  - メタデータと画像データ
  - Internet Archive側がIIF manifest APIでアイテムのデータを取得

連携サイト：<https://archive.org/details/utokyo-general>



# 高精度画像のバックアップ

IIIF画像サーバーで公開するため各部局から預かった高精細画像をPyramid Tiled Tiff形式に変換している。原則、元の高精度画像は各部局で保存しているが、再度の作業が必要になる場合もあり、保存とバックアップを行っている。

- バックアップ対象と方法

- Tiffファイル（一部、JPEGファイルあり）約82万枚
- NASに保存
- コマンドによる手動処理
- 容量は約60TB

# バックアップの多層化

## 第1段階

- NASへ保存

## 第2段階

- テープカートリッジに保存（対象をカートリッジの容量に手動で合わせる）
- 同じデータを3セット作成

## 第3段階

- テープカートリッジを1セットを柏図書館に送付（2セットは本郷キャンパス）
- GitHubやInternet Archiveへの保存

# システムのマイグレーション

オンプレミスサーバーは保守が切れるタイミング（5年）でマイグレーション

- デジタルアーカイブシステム → 2022年に新サーバーに移行済
- IIF画像サーバー → 2022年に新サーバーに移行済
- アーカイブズポータル → クラウド環境に移行予定

NASは容量が不足しそうなタイミングでマイグレーション

- 48TB → 64TB → 120TB

# 課題

- バックアップにかかる時間
- ストレージを用意するための費用
- レプリケーション
- OAIS参照モデルに準拠したバックアップ
- バックアップの運用計画と適切な運用と定期的な見直し
- 担当の運用体制
- OSを含めたイメージバックアップ