

パネルディスカッション

人文系データがもたらす 分野融合の可能性

菊池 信彦 (国文学研究資料館)

中村 覚 (東京大学)

橋本 健広 (中央大学)

大向 一輝 (東京大学)

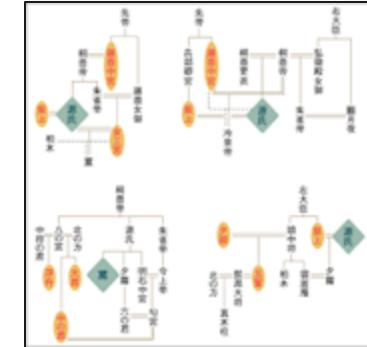
パネルディスカッションの流れ

- 趣旨
 - ユースケース創出課題採択一覧
- 自己紹介
 - 4名×3分
- ディスカッション
 - 5つのキーワード
 - ご質問（Slidoより）

人文学と情報技術



いづれの御時にか 女御更衣あまたさぶらひたまひけるなかに いとやむごとなき際にはあらぬがすぐれて時めきたまふ ありけり…



資料・史料



画像



テキスト



知識



スキャン・撮影



OCR



構造化

メタデータ

画像処理

コンピュータ

テキストアナリティクス

インターネット

クラウドソーシング

大規模化

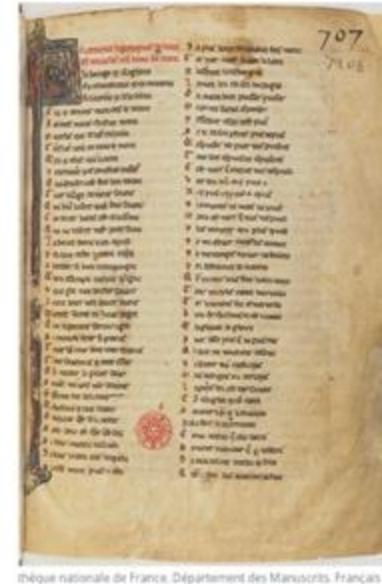
共有と協働

透明性



人工知能

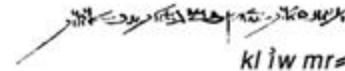
多言語・多文化



M17-G43-I9-G17-Q3-X1-N1
i-w-f-m-p-t-pt

5

- 1 13世紀・アラビア語で書かれたギリシア医学に関する写本
- 2 17世紀・ヘブライ語で書かれたユダヤ法解釈に関する印刷資料
- 3 13世紀・古フランス語で書かれた宗教説話集



kl i w mrs=f tgy mst=f p3 nt wnm=f 6

- 4 13世紀・古ノルド語で書かれたアイスランドにおける法律に関する写本
- 5 ヒエログリフ(エジプト語)とその文字番号・転写
- 6 デモティック(エジプト語)とそのアルファベット翻字

パネルディスカッションの流れ

- 趣旨
 - ユースケース創出課題採択一覧
- 自己紹介
 - 4名×3分
- ディスカッション
 - 5つのキーワード
 - ご質問（Slidoより）

古典籍テキストデータを活用したデータ駆動型人文学のための 研究資源構築プロジェクト

菊池信彦(研究代表／国文学研究資料館)

1. プロジェクトについて

背景①:30万点の古典籍画像データベース「国書データベース」の構築

背景②:「データ駆動による課題解決型人文学の創成～データ基盤の構築・活用による次世代型人文学研究の開拓～」
(FY2024～2033予定)

目的①:古典籍画像データからテキストデータの作成、TEI/XMLデータへと変換

目的②:データ作成ノウハウも蓄積、公開

意義:日本の古典籍資料を、国際的に共有可能な形で提供し、世界的視野でも分析可能な研究資源に

⇒ データ駆動型人文学の研究資源の確立を目指す

2. GakuNin RDMの利用

メンバーそれぞれのTEI/XML作成実践
データとマークアップ方針の作成・共有で利用

3. 成果

- ①(数点)と②に加え、
③万葉集データの可視化ビューワの開発
- ④マークアップ効率化のため歌人人物データ作成
- ⑤国書データベースの書誌データ
TEI/XML化方針のための検討材料

① プロジェクト概要

- ・ 文学研究向けの大規模言語モデルを使用した研究資源の構築・分析とオープン・プラットフォームの構築
- ・ AI支援による文学研究の新たな方法論の確立
- ・ 古典文学から現代文学まで幅広いジャンルを対象
- ・ 国内外のデータベースを接続する学際的アプローチ

▣ 人文研究向けオープン・プラットフォームの特徴

- ・ 人文学研究者の誰にでも開放
- ・ データセットを作成可能
- ・ 大規模言語モデルの影響分析・各種計量分析ツール
- ・ 外部データベースとの連携を検討中（Japan Search・nihuBridge・国書データベースなど）
- ・ データ保存にGakunin RDMなどを使用

■ 具体的な取り組み

- ・ 人文研究者向けオープン・プラットフォームの作成
- ・ 専門家による評価基準の作成
- ・ 英米文学、日本文学、その他諸外国文学全般の文学テキストの分析

■ 構成

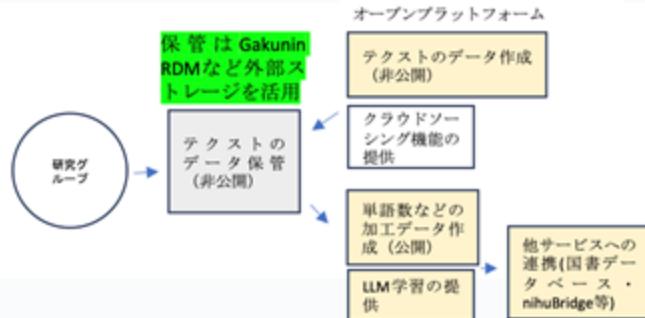
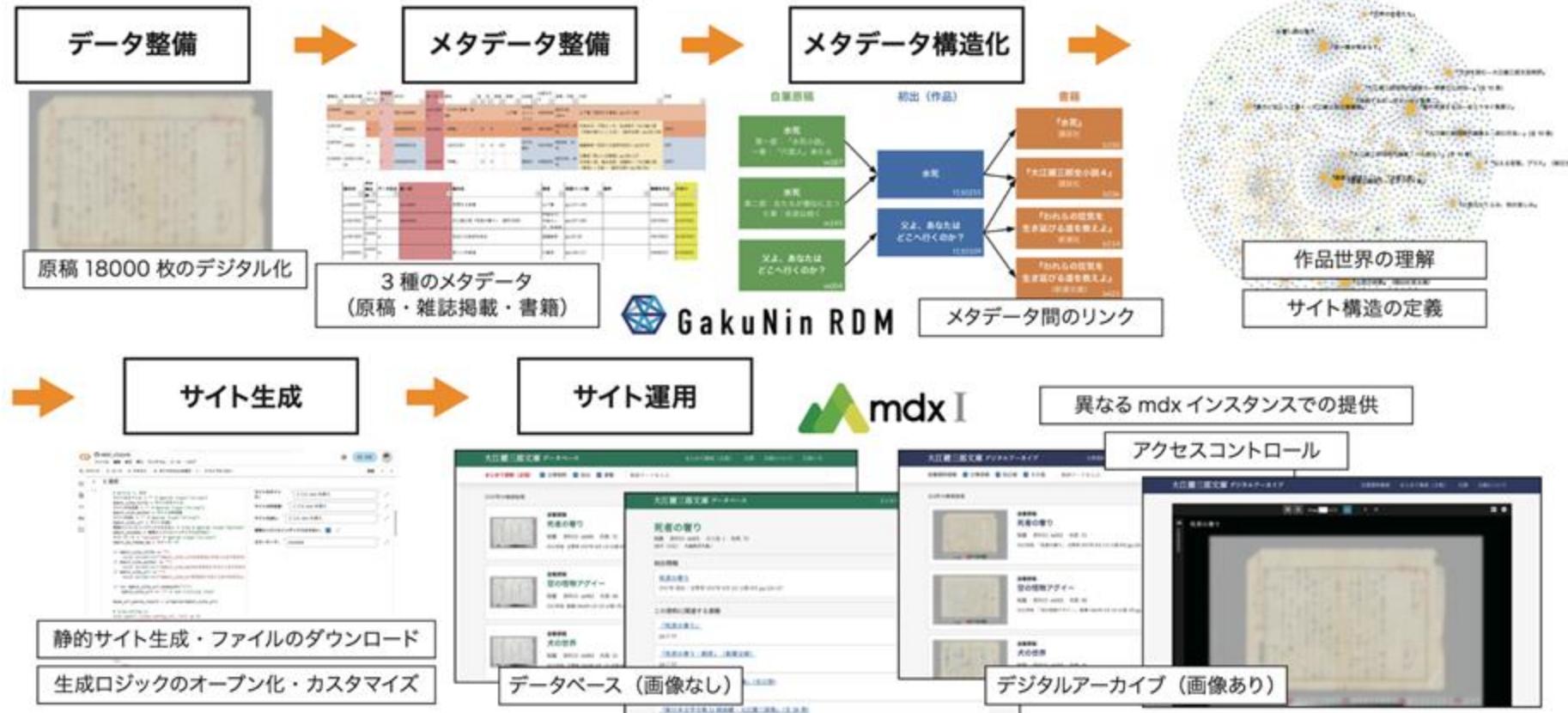


図2:オープンプラットフォームの構成

人文学研究における「読み」を共有するためのデジタルアーカイブ構築・AI活用ワークフローの確立

大江健三郎文庫自筆原稿デジタルアーカイブ・書誌情報データベース <https://oe.l.u-tokyo.ac.jp>

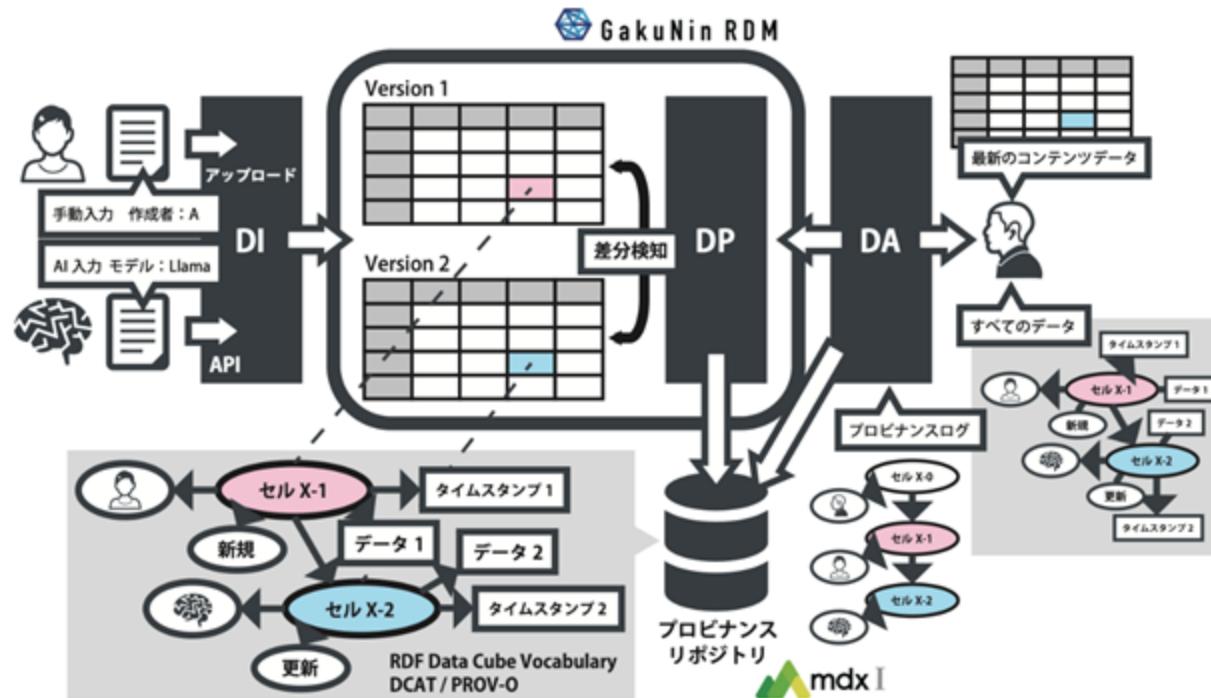


専門家・市民・AIの協働によるデジタルアーカイブ構築に向けたデータプロビナンス基盤のデザイン

データプロビナンス基盤

更新され続ける DA の信頼を担保する

- ・「いつ」「誰が」「どこに」「何を」を記録
 - ・主体の多様性（人・AI）
 - ・再現性から透明性へ
- ・イベント駆動データモデルの導入
 - ・Data Cube / DCAT / PROV-O
 - ・スナップショットへのアクセス
 - ・ログへのアクセス
- ・GRDMとの連携
 - ・GRDM上でのイベント・差分検知
 - ・プロビナンスリポジトリ



キーワード



時

間



技術
術



制度



信頼



期待

期待