

JAIRO Cloud移行実験 事例紹介

既構築機関DSpaceの移行実験

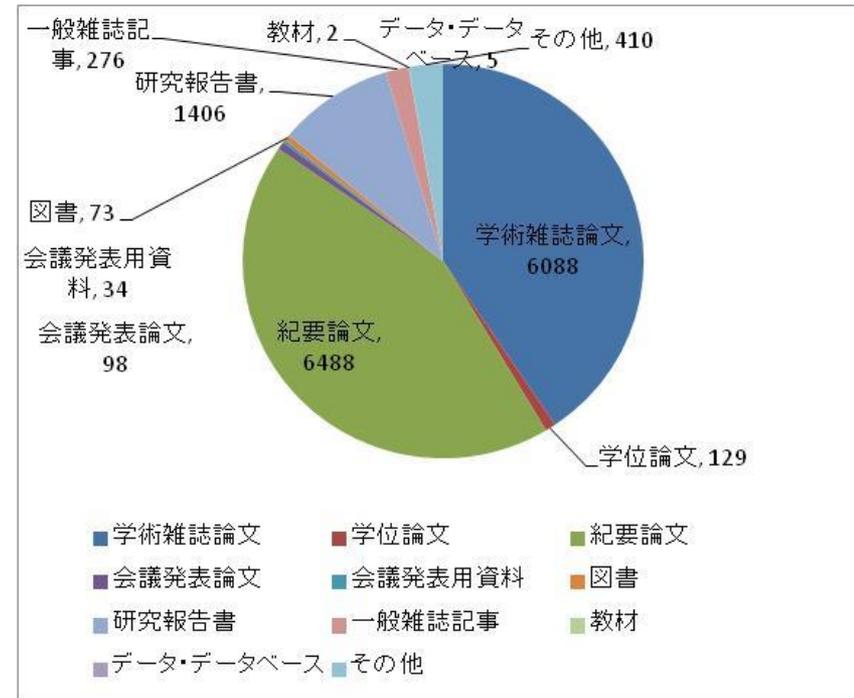
オープンアクセス・サミット2014 2014年10月22日（水）
セッション2 : JAIRO Cloudの新展開
信州大学附属図書館
徳永 澄子

移行実験 参加の経緯

- 本学リポジトリ(SOAR-IR)について
- 参加の経緯
- 移行実験

SOAR-IRについて

- 試験運用開始 H19.1
 - DSpace1.3.2
- 本運用開始 H19.8
- サーバの更新 H23.11
 - DSpace1.6
 - 学内の仮想サーバ
 - Webデザインカスタマイズ
- コンテンツ数 15,070件
H26.7.16時点



移行実験 参加の経緯

- 信州共同リポジトリをJAIRO Cloudで運用
- DSpace1.6でも移行ツール等が動作するか検証
- 著者 I D (本学独自)を使用していること

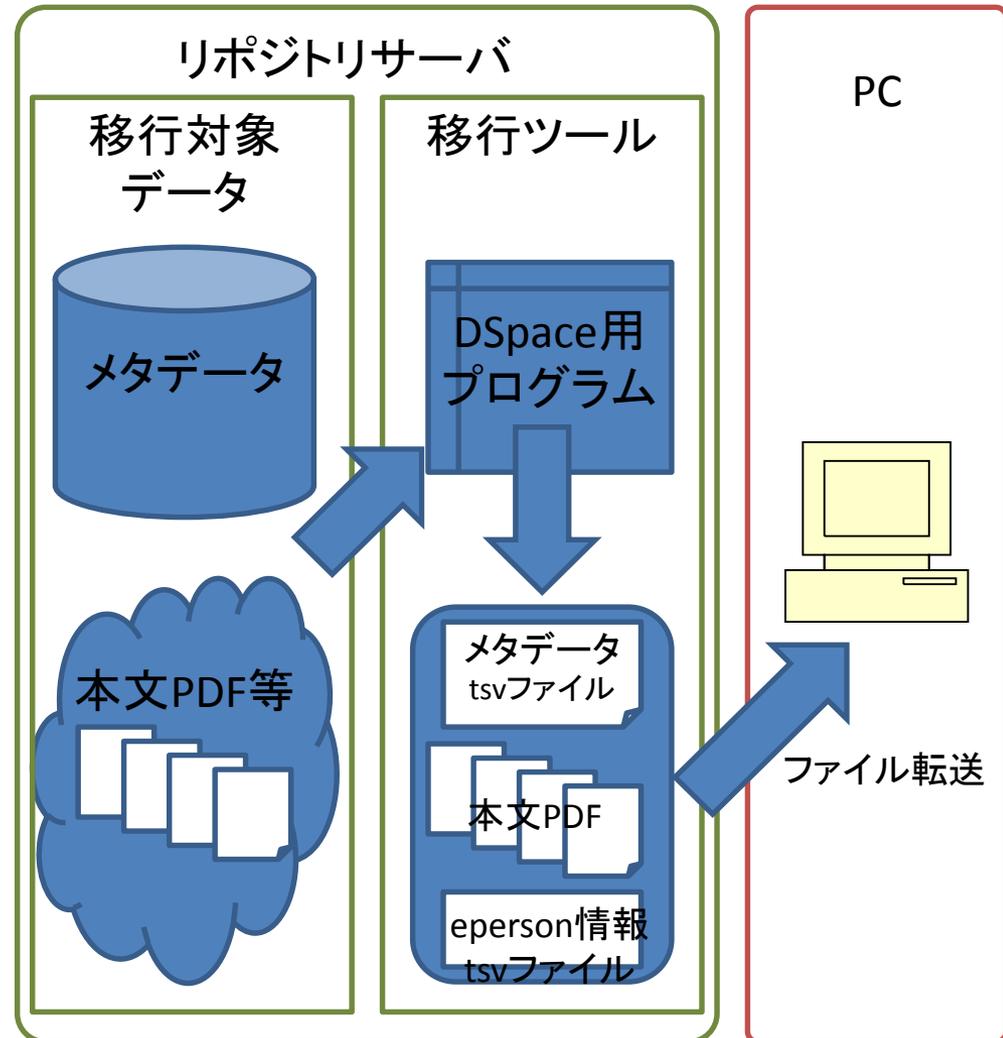
移行実験のスケジュール

	作業	作業内容	担当	時期
①	移行元リポジトリからのデータ抽出	移行元リポジトリからのデータ抽出プログラムをインストールし、データ抽出	信州大	6月中旬～7月中旬
	JAIROCloud実験環境構築	データロード実験用のJAIRO Cloud環境を構築	N I I	7月中旬
②	フィルタ作成・修正	移行元システムのデータ項目とJAIRO Cloudのデータ項目とのマッピング設定	信州大	7月中旬～9月上旬
	サンプルデータロード	サンプルデータ（100件程度）をJAIRO Cloud実験環境にロードし問題がなくデータがロードされたかどうかを検証	信州大	9月中旬
③	大量データロード	大量データをJAIRO Cloud実験環境にロードする	信州大	10月上旬
	登録結果確認	大量データロードについて、問題がなくデータがロードされたかどうかを検証	N I I・信州大	10月上旬

① データ抽出

作業の流れ

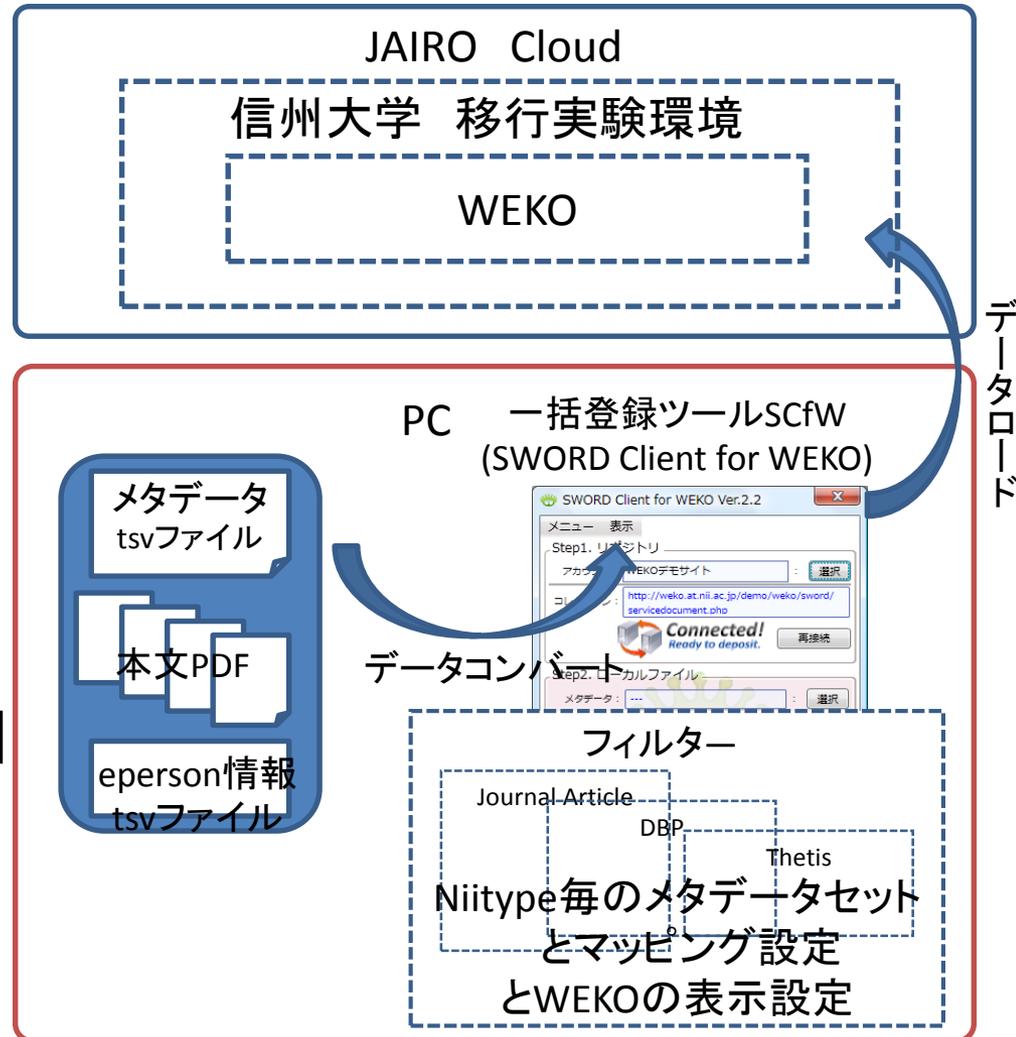
- 移行ツールインストール
- データ抽出
 - ・ 設定ファイルで条件指定
- PCへファイル転送



- 最上位のコミュニティ単位
- 21回に分けて 3日x3H
- データ量 約35GB
- 15,070件を抽出

② フィルター修正、サンプルロード

- 筑波大フィルター提供
- DSpaceの各項目データ数をSQLで調査
- 信大DSpaceのメタデータセットとマッピング設定を再現
- 次の点が課題
 - DSpaceデフォルト項目の移行後の要否
 - 使用頻度の低い項目の移行後の要否



③大量データロード、登録結果確認

大量データロード

- データロード数 10,000件程度
- 実作業 8回のデータロード 全6時間程度
 - 1回のデータロード数の上限2,000件程度。

登録結果確認

- 100件程度

今回分かった点

抽出ツール

- 本文ファイル記述事項(bitstream.description)が抽出されない
- 本文ファイル閲覧権限が抽出されない
- リンクが変換生成できない 例.)DOI,PMID

著者ID

- 著者IDがWEKO「著者ID」に入らない
- 本学著者IDはDSpace1.6標準機能ではない…

データの不整合(フィルター設定とDSpaceデータ)

良かった点

- 15,000件程のデータ移行が2ステップ、1週間位の実作業で行える
- DSpaceのコミュニティ/コレクション構造がWEKOのインデックスに再現
- メタデータと本文ファイルそれぞれに公開日設定が可能

本学の課題

SOARの全体像

- 機関リポジトリ : SOAR-IR
- 研究者総覧 : SOAR-RD
- 視認度評価分析システム : SOAR-RVAS

特徴

- 研究者総覧と機関リポジトリの相互リンク

The screenshot displays the SOAR website interface. At the top, there is a navigation bar with 'HOME', 'SOARとは', '研究者総覧', '機関リポジトリ', '視認度評価分析システム', and 'お問い合わせ'. Below this is a banner for '未来創造のきっかけはここにある。' (The starting point for future creation is here.) featuring a magnifying glass icon. To the right of the banner is an 'お知らせ' (Notice) section with a list of recent updates and contact information for the SOAR team.

The main content area is divided into two columns. The left column is titled '研究者総覧 - 研究者情報を調べる-' (Researcher List - Search for researcher information-) and contains a search box and a list of departments: 人文学部, 教育学部, 経済学部, 理学部, 医学部, 工学部, 農学部, 繊維学部, and 全学教育機構. Below these are links for '大学院から選ぶ' (Select by Graduate School).

The right column is titled '機関リポジトリ - 研究成果を調べる-' (Repository - Search for research results-) and contains a search box and a list of departments: 人文学部, 大学院人文科学研究科, 教育学部, 大学院教育学研究科, 経済学部, 大学院経済・社会政策科学研究科, 大学院法曹法務研究科, 理学部, 医学部, 大学院医学系研究科, 工学部, 農学部, 大学院農学研究科, 工学部, 法学部, 大学院法務研究科, and 工学部.

A large orange double-headed arrow labeled '相互リンク' (Mutual Link) points between the two columns, indicating the cross-linking between the researcher list and the repository.

本学の課題

大学の研究者総覧との相互リンクの維持

- 著者名フィールドからリンクは困難
- 研究者総覧へのリンクフィールドを用意？
メタデータtsvファイルの編集が必要

視認度評価分析システムの維持