

平成 26 年 12 月 22 日
信州大学附属図書館

平成 26 年度 信州大学 JAIRO Cloud 移行実験レポート

1. はじめに

JAIRO Cloud とは、機関リポジトリを新たに構築する機関を当面の対象として、機関リポジトリのシステム環境を提供するサービスである。国立情報学研究所は、この JAIRO Cloud を平成 24 年度から運用している。JAIRO Cloud の今後の展開として、機関リポジトリの維持が困難な機関、JAIRO Cloud の先進機能の利用を望む機関に対しても、サービスの提供を検討している。その際に問題となるのが、既存のリポジトリシステムからの JAIRO Cloud へのデータ移行である。

本学は国立情報学研究所に協力し、データ移行の検証を行うべく、平成 26 年度より本実験に参加した。本レポートは、この実験における検証結果である。

2. 実験計画

(1) 実験概要

本移行実験は、機関リポジトリシステム DSpace1.6.1 から JAIRO Cloud に支障なくデータ移行ができるかどうかを確認するものである。国立情報学研究所が提供するデータ移行プログラム及びマッピング設定ファイル（フィルタ）、手順書、ワークシートをもとにデータ移行作業を行い、その評価及び問題点の指摘を行うことを主な目的としている。

データ移行においては、移行元システムに合わせたデータコンバートのモジュールを使用し、リポジトリのデータを実際にロードすることによるテストを行う。移行手順は、下図の通りである。

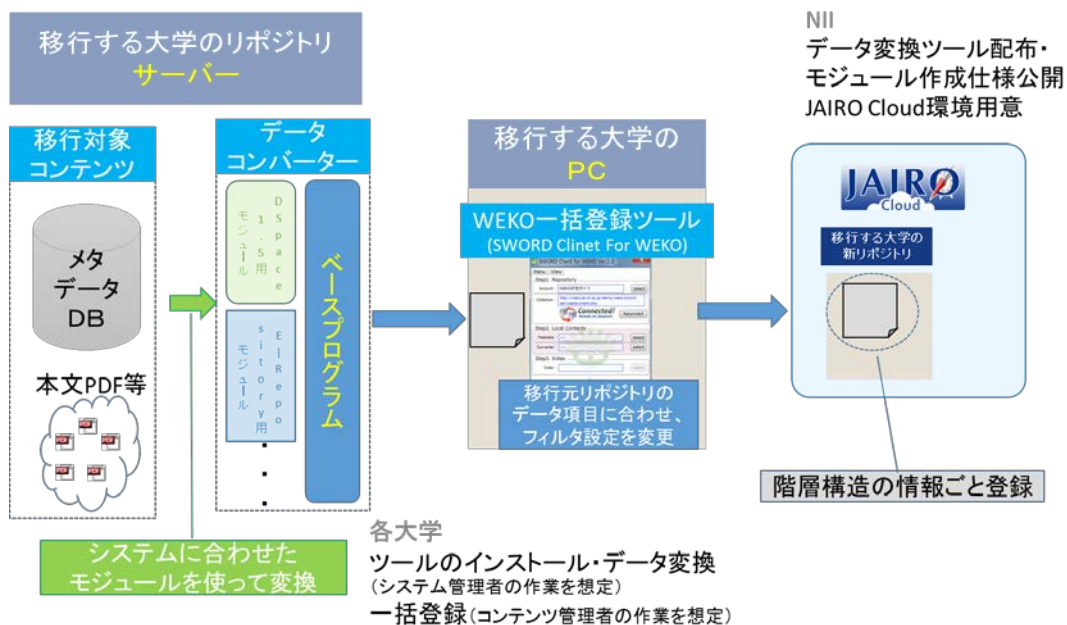


図1 データ移行作業のイメージ

(2) データロード実験作業

データロード実験計画における作業は次のとおり。

表1 データロード実験作業一覧

作業項目	作業内容	作業主体
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	信州大学
JAIRO Cloud 実験環境構築	データロード実験用の JAIRO Cloud 環境を構築	国立情報学研究所
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	信州大学
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	信州大学
大量データロード	大量データ (10,000 件程度) を JAIRO Cloud 実験環境にロードする	信州大学
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	国立情報学研究所・信州大学

(3) スケジュール

実験のスケジュールは次のとおり。

表2 データロード実験スケジュール

		4月	5月	6月	7月	8月	9月	10月	11月	12月
準備	移行元からのデータ抽出			→						
	JAIRO Cloud 実験環境構築				→					
	フィルタ作成・修正				→	→				
サンプルデータロード							→			
大量データロード								→		
登録結果確認								→	→	→

3. 実験の実施

(1) 実験の準備 (データ抽出、フィルタ作成・修正)

【データ抽出】

データ抽出は、抽出当時の全件 (15,070 件) を最上位のコミュニティ単位で 21 回データ抽出を行い、9 時間程度 (3 日間×3 時間)・35GB のデータ量になった。エラーは 2 種 (「ファイル名を含めたインデックス名が 200 バイトを超える」、「コミュニティまたはコレクション名に Windows で使用出来ない文字が含まれていたため全角文字に変換を行った」) 起り、いずれもマニュアルに記載されている想定内のものであった。「コミュニティ名が 200 バイトを超えた」については、ロードの際にエラーが起こる可能性があるとのことだったが、ドライブ直下に置き作業したところエラーが起きなかった。

本文ファイルの記述 (bitstream.description)、本文ファイルの閲覧権限がデータコンバータでは抽出できなかった。本文ファイルの記述 (bitstream.description) は、詳細の報告を行い次回のデータコンバータの改修が検討されている。本文ファイルの閲覧権限は、DSpace から別途抽出しデータロード後に個別に本文ファイルの公開日設定をすることで対応できた。

1 アイテムに複数の本文 PDF ファイルが登録されている場合、DSpace から「登録日時順」で出力した本文 PDF ファイルが必ずしも表示順と合っていない。本学は、2011 年に DSpace1.3.2 から DSpace1.6.1 にバージョンアップを行った。DSpace1.3.2 の間に登録したアイテムの本文 PDF ファイルが DSpace 上での表示順と逆順に出力され、DSpace1.6.1 移行後に登録したアイテムの本文 PDF ファイルは DSpace 上での表示順と正順に出力されることが分かった。両者の bitstream_id 順に違いがみ

られない。国立情報学研究所からの提案により抽出ツールでの対応策として、抽出ツールの本文 PDF ファイルの出力設定「ファイルの並び順を正順/逆順」によってコントロールすることで対応することとした。具体的には、下記の手順でデータ抽出・ロードを行う。

1. DSpace 1.3.2 のデータをロード

- (1) データコンバータで「ファイルの並び順を正順」に設定しデータを抽出。
- (2) 抽出メタデータから DSpace1.6.1 以降に登録したメタデータを手作業で削除。
- (3) このデータで一括ロード。

2. DSpace 1.6.1 のデータをロード

- (1) データコンバータで以下のとおり設定してデータをロード
 - ・ファイルの並び順を逆順に設定
 - ・データ抽出条件に、DSpace1.6.1 移行日を設定（差分更新用の機能を使用）
 - ・メタデータのみ抽出
- (2) 上記2-(1)のメタデータを、上記1の本文ファイルのあるフォルダにコピー（時間がかかるが、上記2-(1)で本文ごと抽出でもよい）
- (3) このデータを一括ロード。

【フィルタ作成・修正】

信州大学の DSpace で現在運用しているメタデータセットとクロスワークを再現するという方針でフィルタ作成を行った。

現在のコンバータでは、著者 ID が WEKO の著者 ID のフィールドに入らなかった。DSpace1.6.1 の標準機能 Authority Control の仕様を報告し、次回のデータコンバータの改修に含めてもらうこととなった。本学で使用している著者 ID は、Authority Control 機能によるものではなく、以前のバージョンからのカスタマイズにより運用されていることが分かった。

WEKO のメタデータから外部サイトへリンクを行うには、フィルタでリンク属性を使用することでリンクが可能である。リンク属性ではメタデータに「リンク名」、「リンク URL」を指定し、junii2 などのクロスワークには「リンク URL」の値が出力される仕様である。DOI をこのリンク属性にすると、junii2 への出力の際、フル URL (junii2 非推奨) が出力されることが懸念される。雑誌掲載論文をリポジトリ登録する際、出版者側の条件により出版社版の論文へのリンクが求められるケースがあり、DOI からのリンクによってこれを満たしている大学もあると考えられ、DOI からリンクを行いたいと希望する機関が多くあると予想される。

フィルタ作成の考え方、フィルタと WEKO アイテムタイプの関係を説明する資料があると取組みやすいと思う。フィルタの各設定の動作、仕様などについて説明資料があると良い。

また、WEKO のメタデータセットをフィルタで作成するにあたって、自大学の移

行前リポジトリのメタデータセットやクロスワークを確認しながら作業を行うとスムーズだが、移行を希望する機関がその確認方法に不安がある場合はサポートが必要になるかもしれない。先行機関や各種コミュニティからのサポートが適切だと思われる。

(2) データロード (サンプルデータロード、大量データロード)

計 10,000 件程度のデータロードを行った。1 回のデータロード量を 2,000 件程度に調整し 8 回に分けてデータロードを行ったところ、作業時間は合計 6 時間程度で順調に終了した。

データロード後に WEKO のアイテムタイプの変更を行うと、フィルタと WEKO のアイテムタイプの間で不整合を起し、その後データロードを行った際、変更を行った項目が登録されない事象が起こった。

フィルタで項目の属性タイプを「プルダウン」にした場合、ロード後、次の事象が起こった。「プルダウン」を選択した項目が NULL 値だった場合、そのアイテムを WEKO 上で空更新するとデータが崩れた。この現象は、最新の WEKO (ver. 2.1.7) では解消されており、最新の WEKO や SCfW ではエラーチェックされる。フィルタで属性タイプ「プルダウン」を選択する場合、メタデータに選択候補以外の値が入っていないか、複数の値が入っていないか事前に確認する必要がある。

4. まとめ

Dspace ver 1.6.1 からの JAIRO Cloud へのデータ移行については、データコンバータの改修を待つ点があるが、国立情報学研究所の若干のサポートがあれば可能であることがわかった。しかし、全国の機関リポジトリ構築機関から、JAIRO Cloud へのシステム移行を受け入れるにあたっては、データ移行マニュアルの整理等の対策が望まれる。