平成 27 年 3 月 26 日
国立情報学研究所
学術コンテンツ課

旭川医科大学 JAIRO Cloud 移行実験レポート

1. はじめに

国立情報学研究所の JAIRO Cloud (以下、「JC」という。) は、リポジトリを新規に構築する機関への提供を目的とした機関リポジトリの SaaS である。また今後は、すでに機関リポジトリを保持しているが、維持・管理を負担と感じている機関や JC の先進的な機能の利用を望む機関への提供も検討されている。

すでに運用されている機関リポジトリを JC へ移行するためには、データの移行が必要となる。平成 25 年度旭川医科大学と国立情報学研究所は、機関リポジトリシステム XooNIps にて運用されている旭川医科大学 学術成果リポジトリ(以下、「AMCoR」という。) のデータを JC に移行する実験を行った。また、本実験を通して XooNIps から JC へのデータ移行ツールの開発業者（以下、「開発業者」という。）による当該ツール実地検証も行った。

2. 実験の概要

本実験の目的は、実際にデータの移行を行うことで、想定されている移行の手順や新たに発見された課題の確認を行うとともに、AMCoR の JC への移行の見込を評価するものである。

実験は、国立情報学研究所より旭川医科大学にデータ移行ツール（以下、「データコンバータ」という。)、手順書を提供し実施した。なお、データコンバータを構成する XooNIps 用モジュールの運用は本実験が初であり、当該モジュールのレビューも併せて実施した。

データ移行手順のイメージを図 1 に示す。

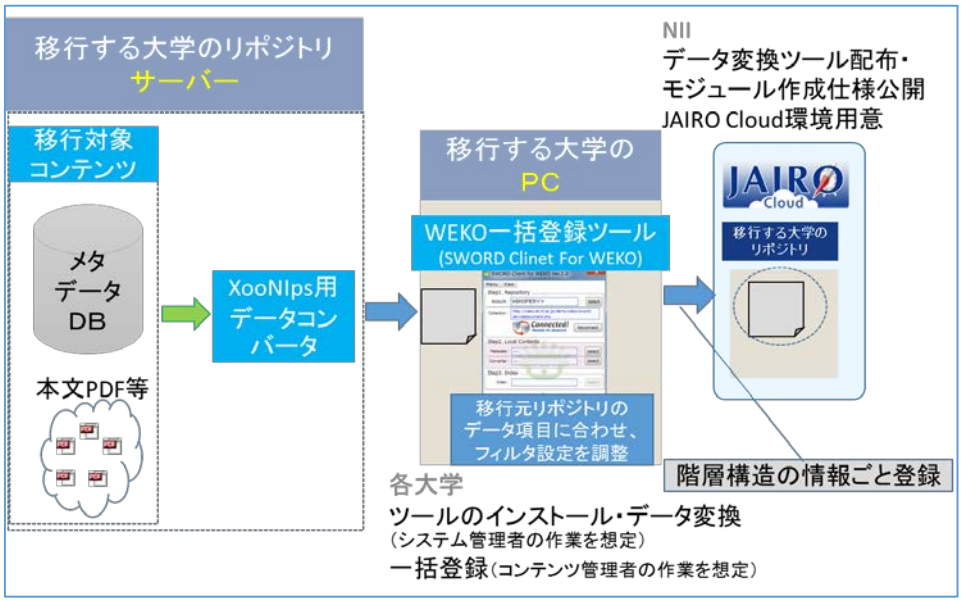

図 1 移行手順のイメージ

（１）実験を構成する作業

実験は表 1 に示す 6 つの作業から構成される。

表 1 作業一覧

作業名	作業内容	作業主体
1. 実験環境構築	抽出したデータのロード対象となる JC 環境の構築	国立情報学研究所
2.データ抽出	データコンバータのインストールとデータの抽出	旭川医科大学（開発業者の支援による）
3. フィルタ作成	XooNIps データ項目と JC のデータ項目の対応関係を設計(マッピング)とデータのロードに用いるフィルタの作成	旭川医科大学(開発業者の支援による)
4. サンプルデータロード	サンプルデータ（10 件程度）の実験環境へのロード	旭川医科大学(開発業者の支援による)
5. 大量データロード	大量データ（5,000 件程度）の実験環境へのロード	旭川医科大学(開発業者の支援による)
6. 登録結果確認	大量データロードの結果の確認	旭川医科大学・国立情報学研究所

（２）実験の実施期間

実験は平成 25 年 10 月から平成 26 年 5 月にかけて実施された。各作業の実施期間(実績)を図 2 に示す。

		10月	11月	12月	1月	2月	3月	4月	5月
実験準備	実験環境構築	→	→						
	データ抽出				→	→	→		
	フィルタ作成				→	→	→		
サンプルデータロード							→		
大量データロード							→		
登録結果確認								→	→

図 2 作業の実施期間(実績)

3. 実験の作業報告

（１）データ抽出

旭川医科大学において、AMCoR をホストする環境にデータコンバータのインストールを行い、平成 26 年 1 月 8 日にデータ抽出を行った。対象は XooNIps の Library Module で登録されたアイテムである。処理には 3 時間を要した。

さらに、データコンバータの修正を行い、平成 26 年 3 月 11 日に 2 回目のデータ抽出を行った。データ抽出の条件と対象となったデータは 1 回目のデータ抽出と同様である。

課題と対応

- 抽出したデータを Windows で解凍すると、日本語を含む本文ファイル名が文字化けする。

⇒　移行手順において、文字化けを起こさない Windows のファイル解凍ソフトの使用を推奨したが解消しなかったため、Linux 上で tar ファイルを展開し、一件ごとにダウンロードを実施した。

- 抽出したデータは、本文ファイル名とサムネイル画像ファイル名がデリミタを挟んで同じ項目で出力されている。
  ⇒　データコンバータの改修により、現在は別の項目で出力される仕様に変更されている。
- 現状の AMCoR からの抽出データは、JAIRO Cloud 登録前に正規化を進める必要がある。このため、著者名カナ表記の追加やデータフィールドの分割やマージなど、ツールで修正できないものについて、全件手作業で修正する作業が発生する。

（２）フィルタ作成

旭川医科大学おいて、XooNIps の Library Module のデータ項目と JAIRO Cloud のデータ項目の対応を設計し、「XooNIps・JAIRO Cloud データ項目対照表」にまとめた。さらに、当該設計を元にして Library Module から抽出したデータを実験環境にロードするためのフィルタを作成した。

なお、Library Module 以外の Module に対応するフィルタは、Library Module をテンプレートとして国立情報学研究所が作成した。

（３）サンプルデータロード

旭川医科大学において、平成 26 年 3 月 18 日にサンプルデータロードを行った。データは、平成 26 年 3 月 11 日に抽出したデータのうち 10 件を用いた。

（４）大量データロード

旭川医科大学において、平成 26 年 3 月 24 日～25 日にかけて大量データロード（Library Module の全件 ：4737 件）を行った。

課題と対応

●本文ファイルのサイズが 100MB を越える場合、アップロードがエラーとなる。
⇒ JC では、複数ファイルの一括データロードにおいて単一のファイルのアップロードに 10 分以上を要するとき、タイムアウトとなる仕様となっている。個別登録においてはタイムアウト設定が存在しないため、サイズが大きなファイルを低速

なネットワークからアップロードしようとするときは、個別登録の実施を推奨することとした。

●「上位タイトル_発行年月次」に全角スペースが含まれており、アップロードがエラーとなった。

⇒ アップロード前に、抽出データから全角スペースを除去することで正常なアップロードを行う運用を標準として定めた。

●本文ファイル名にハイフンやダッシュ等の機種依存文字が含まれていると、アップロードがエラーとなる。

⇒ 現状のデータコンバータの仕様では登録可能な形式に変換を行っている。

（５）登録結果確認

国立情報学研究所と旭川医科大学において、大量データロード完了後、登録されたアイテムのサンプル調査及びレスポンス等の検証を行い、大量データロードでエラーとなったもの以外に、共著者名が表示されない問題や、巻号表示に不具合があることを発見した。この問題は、データ全件をチェックと正規化を行えば解消することを確認した。

4. JAIRO Cloud の機能に対する要望

移行実験を通じた JC に対するオペレーションの経験をふまえ、旭川医科大学より以下の要望があった。

- ファイル公開日の設定を NULL としたまま、オープンアクセスに設定したい。
- XooNIps のアイテムの閲覧数とダウンロード数を WEKO のアイテム詳細画面の「利用統計を見る」に移行できる機能の実装。
- ストリーミングの動画配信機能の実装。
- プログラミング等の知識がなくてもできるフィルタの作成。

5. まとめ

本実験では、国立情報学研究所が提供するデータコンバータを用いて、XooNIps から JC へのデータ移行が可能であることを確認した。

ただし、データの正規化は、全件のデータチェックと不足するデータの追加が必要であり、移行にはかなりの工数を要することも判明した。また、フィルタの作成も相応の工数を要することとなったが、実際の移行では本実験で作成したフィルタをテンプレートとして用いることができるため、工数は削減できるものと考えられる。

なお、Library Module 以外のフィルタの作成は、国立情報学研究所にて行うことになった。