

平成 27 年 3 月 18 日
山 形 大 学

平成 26 年度山形大学 JAIRO Cloud 移行実験レポート

1. はじめに

JAIRO Cloud は平成 24 年度から運用が開始された国立情報学研究所提供の SaaS 型の機関リポジトリサービスである。参加機関には、機関リポジトリシステム WEKO をインストール済みの環境が提供されるため、各大学の担当者は、ハードウェアおよび機関リポジトリシステムの管理から解放されるという大きなメリットがある。当初は自力構築が困難な機関が主な対象であったが、現在、既構築機関で、継続的な維持が困難な機関、JAIRO Cloud の先進機能の利用を望む機関へのサービス提供の検討を開始したところである。その際に問題となるのがデータ移行である。データ移行の可能性をさぐるために、平成 25 年度から実験は開始された。実証実験は DSpace、XooNIps、E-Repository 等の主要なリポジトリシステムごとに進められ、山形大学は NALIS-R の実験を担当することとなった。まず平成 25 年度に NALIS-R からのデータ抽出ツール開発に協力し、平成 26 年度に移行実験をおこなった。本報告では移行実験について報告する。なお実験にあたっては千葉大学の技術サポートを受けた。

2. 実験計画

2-1. 実験概要

移行実験のフローを図 1 に示す。移行時には既存の機関リポジトリシステムからのデータコンバートが必要である。主要な機関リポジトリシステムについては、移行実験を実施し、データコンバート用プログラムの開発・実証と、作業マニュアルの整備を行う。

移行実験の概要は以下の通りである。

平成 25 年度【準備段階】<本実証実験の対象外>

ベンダーとディスカッションの上で、データ抽出プログラム開発に協力

平成 26 年度【実証実験】

- ① サーバーへのデータ抽出プログラムのインストール
- ② データ抽出プログラムによるメタデータおよび本文コンテンツの抽出および Windows PC へのダウンロード
- ③ JAIRO Cloud 一括登録ツール (SCfW) の Window PC へのインストール
- ④ メタデータ変換フィルタの作成
- ⑤ JAIRO Cloud へのデータの一括アップロード

実験参加機関は、上記フローを各大学の環境で実行し課題点を抽出し、各機関リポジトリシステムから、JAIRO Cloud へのデータコンバート用のフィルタを作成する。作成したフィルタは、標準フィルタとして NII が無償で移行希望機関に配布する。そのフィルタは、多くの機関は、スムーズなデータのコンバートが可能となるよう設計する。機関によって

は、データのフォーマットに大きく手を加えている場合もあるであろうが、その場合でも、標準フィルタを修正する簡易なインターフェース、およびドキュメントを用意する。

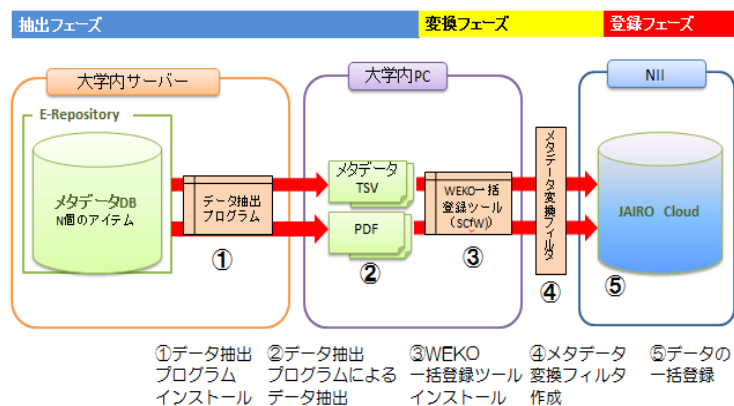


図1 データ移行作業のイメージ

2-2. 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	山形大学
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	山形大学 (千葉大学)
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	山形大学 (千葉大学)
大量データロード	大量データ (【件数】) を JAIRO Cloud 実験環境にロードする	山形大学 (千葉大学)
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 山形大学

2-3. スケジュール

移行実験の作業結果は以下のとおり。

表2 データロード実験スケジュール

		4	5	6	7	8	9	10	11	12	1	2	3
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

3-1. 実験の準備 (データ抽出、フィルタ作成・修正・サンプルデータロード)

実際は、フィルタ修正とサンプルデータロードは作業としてはセットであったため、サンプルデータロードも、実験の準備として整理する。データ抽出の手順は以下の通り

3-1-1. サーバーへの抽出プログラムインストール。

- ①IR サーバーの環境確認
- ②クライアントへの SFTP 対応プログラムのインストール (今回は FFFTP を用いた)
- ③NII 提供の抽出プログラムファイルの IR サーバーへのアップロード
- ④Oracle 接続用の Peal モジュールのインストール DBD-Oracle-1.17
- ⑤IR サーバー上での解凍、インストール
- ⑥Param. ini を vi で編集し、プログラム実行、動作確認。

初回作業 (2014.8.18) では、インストール手順書に従い、移行ツールのインストール作業を実施したが、Perl の ORACLE 用モジュールが未インストールであった。モジュールのソースをダウンロードし、インストールを試みたが失敗した。エラーメッセージより、環境変数の設定不足が原因ではないかと推定し、ORACLE_HOME のパスを設定、make までは完了した。が、make test では「共有オブジェクトファイルを開けません: そのようなファイルやディレクトリはありません」というエラーがでた。時間的制約もあり、初回の作業はこの時点で終了した。

二回目の作業 (2014.12.15) では以下の手順で ORACLE 用モジュールをインストール、設定した。ORACLE 用モジュールの最新版は DBD-Oracle-1.74 であるが、未対応のため、DBD-Oracle-1.17 をインストールしている。

インストール作業は以下のとおり。

- ① DBD-Oracle-1.17.tar.gz のダウンロード
- ② tar zxvf DBD-Oracle-1.17.tar.gz
- ③ cd DBD-Oracle-1.17
- ④ export ORACLE_HOME=/opt/oracle/product/10_2_0/db_1/
- ⑤ export LD_LIBRARY_PATH=/opt/oracle/product/10_2_0/db_1/lib
- ⑥ perl Makefile.PL
- ⑦ make
- ⑧ make test
- ⑨ make install

しかし、抽出ツールの output2weko.pl の挙動が不安定であり、何度か全件抽出を試行したが部分的にしかデータが抽出されないため、NII 担当者にエラー報告の上、作業を中止した。

三回目の作業(2015.03.05)では、バージョンアップ後のバグ修正済の抽出ツールをサーバーに再インストールし、全レコードの抽出に成功した。

3-1-2. フィルタ作成・修正・サンプルデータロード

初回作業時に山形大学の PC に SCfW を NII のサイトからダウンロードし、PC にインストールし、接続設定を行った。

第二回作業時には、不完全ではあるが、データを抽出できたのでサンプルアップロードテストを行った。データコンバートは筑波大フィルタを修正した。修正は XML ファイルをテキストエディタ TeraPad で直接編集した。

3-2. データロード (全件データロード)

3-2-1. データロード実験

全件データロードの実験は千葉大学で行った。PC に SCfW2-2-2 をインストールの上、全件アップロードテストを行ったが、SCfW2-2-2 は複数ファイルのアップロードに対応していないことが判明した。そのため SCfW2-2-1 にバージョンダウンの上、全件データロードを行った。

データ登録については、約 4500 件でメタデータ変換は 30 分ほどで終了した。データの JAIRO Cloud への登録は約 6 時間を費やした。

データのコンバートエラーが幾つか発生したが、以下の二つのパターンである。

- ・ファイル名が文字化けしている
- ・日付の入力形式が誤っている。

いずれも、個別に登録する場合は入力エラーとしてリジェクトされるであろうが、機関リポジトリシステム構築時には、おそらくダイレクトにデータベースにアップロードしたためメタデータ記述に誤りがあっても登録できたのではないかと思われる。今回は 30 件程度であったため、手作業で一件一件修正したが、エラー件数が多い場合は、なんらかの対処方法を検討を要する可能性がある。

4. 所感

移行ツールは、Oracle 用の Perl モジュールの事前インストールが前提であったが、NALIS-R については、少なくとも、山形大学では未インストールであることが判明した。かつ、インストールには UNIX 操作のスキルを要した。今回の移行実験作業の目的は、「高度なスキルを必要としない移行方法」の確立と考えているが、モジュール設定・インストール作業が困難な場合は、ベンダーである NTT データ九州への作業依頼も必要となる可能性がある。同様の問題は、NALIS-R 以外の DSpace でも生じる可能性がある。

WEKO が柔軟で拡張性の高いシステムであるということは今回の移行実験で理解したが、既存機関の JAIRO Cloud への移行を本格的に考えるのであれば

- 1) ドキュメントの整備
- 2) サポート体制(コミュニティ)の整備
- 3) 移行講習会の実施

が必要と考えている。

実際の移行作業では、現段階ではある程度の UNIX の操作スキルと、データコンバートのスキルが要求される。データ抽出にあたっては、UNIX サーバー上でのプログラムの実行を行わなければならないし、一般的な大学図書館員にとっては心理的なハードルが高いのではないかと。また抽出後のデータについても、データ修正が必要になる可能性が高く、TSV ファイルの処理のスキルも必要であろう。

5. 終わりに

地域共同リポジトリとして発足した「ゆうキャンパスリポジトリ」は、小白川図書館が中心となり、大学コンソーシアムやまがた参加図書館発行の紀要論文を主体にリポジトリ搭載作業を進めてきた。また、大学を取り巻く情勢が大きく変化することに伴い、各大学における情報発信機能強化の必要性はますます高まることとなった。さらに、平成 25 年の学位規則の一部改正に伴い、博士論文のインターネット利用による公表が原則とされるなど機関リポジトリが重要な手段として位置付けられるようになった。このような状況にもかかわらず、「ゆうキャンパスリポジトリ」を管理運用している小白川図書館においては、職員削減にともなう人材不足もあり、耐用年限を大幅に超過したリポジトリサーバーのシステム更新ができない状態が続いていた。

そのような折に、NII が「JAIRO Cloud」を構築し、平成 24 年度には運用を開始した。小白川図書館では「ゆうキャンパスリポジトリ」の JAIRO Cloud への移行の検討を進めていた。平成 25 年度には、NII から移行の試行についての提案をいただいた。「ゆうキャンパスリポジトリ」参加図書館へ「JAIRO Cloud」移行についての同意をとり付け、ようやく移行実験が開始される運びとなった。

平成 25 年度、本移行実験開始前の移行用ツール開発の協力については、当時、山形大学に所属していた三角氏（以下、敬称略）が担当した。準備作業を開始した時点では、三角は工学部の研究支援担当であり図書館から離れていたが、小白川図書館から工学部に要請し、三角が作業担当をすることとなった。平成 25 年 10 月に三角が、全学図書館の調整を担当する事務局渉外課に配置換となり、本格的に作業を開始した。作業内容は ORACLE

のデータベースからのデータのダンプおよび NALIS-R のバックアップであったが、すべて LINUX サーバー上の作業であり、UNIX および ORACLE の知識が無ければ困難なものであった。作業はすべて三角が行った。当時小白川図書館内にはデータ移行に関する専門的知識と技術を有する職員が不在であり、当時のシステム開発を委託した NTT データ九州に問合せたが、年月の経過とともに当時の事情を理解している担当者がいなかったため、移行実験を業者に依頼することもできず、専門的知識を有する三角に要請したものである。平成 25 年学位規則の改正により、博士論文のインターネット公表が原則とされてもそのような状況であった。平成 26 年度に正式に移行実験を開始し、山形大学は NALIS-R の移行実験を担当することとなった。しかし平成 26 年 4 月に三角は千葉大学へ異動した。山形大学としては三角に頼らざるを得ないことから出張依頼等で三角に移行実験を依頼したが、三角も千葉大学の業務もあり（千葉大学の JAIRO Cloud 移行実験も三角が担当していた）、作業時間をとることが困難であった。そのため当初の予定どおり移行実験を終了することができず、このたびようやく移行実験終了に至ったという状況である（実質的な作業日数は、山形大学で 3 日、千葉大学で 2 日の計 5 日程度である）。

このことは、自前のサーバーを所有しながら適切な運用を行う人材育成がなされていなかった事が原因であり、大いに反省している。今回の件を踏まえ、今後は学術情報の進展に対応できる人材の確保と育成に努めることとしたい。一方で、全国の多くの大学でも、同様にシステム管理人材が不足しているケースは少なくないのではないかと。移行実験で提供いただいたマニュアルおよびツール類では、ある程度のシステム管理のスキルがなければ、実行は困難であった。より初心者にもフレンドリーな環境整備を望みたい。

最後に、移行実験に当たって全面的にご協力いただいた NII の職員の皆様に心より感謝申し上げます。