

オープンアクセス・サミット2014

第2部：未来への飛翔～機関リポジトリの更なる発展を目指して

セッション2：JAIRO Cloudの新展開

JAIRO Cloud移行実験 事例紹介 (3)

千葉大学附属図書館

三 角 太 郎

ShaRe

Image the
Colourful IR in
one system

ShaRe (Shared Repository -シェア-)

国立情報学研究所学術機関リポジトリ構築連携支援事業
平成20-21年度委託事業領域2プロジェクト 「共同リポジトリ:モデルの構築と普及」

目的

2008年度計画

連携機関

関連資料

関連リンク

連絡先



NII学術情報リポジトリ
構築連携支援事業

●目的

現在、日本で運用されている機関リポジトリは予算や人的資源の豊富な比較的大・中規模機関が数多くを占めていますが、今後、日本で生産される学術情報を網羅的に発信するためには、中小規模の学術機関でも機関リポジトリを構築し、運用することでオープンアクセスの裾野を広げることが最も大きな課題であると考えられます。中小規模の機関では、費用や労力をシェアできる共同リポジトリが機関リポジトリの構築・運用にとって有効な手段の1つであることは、先行例からも明らかです。本プロジェクトは、共同リポジトリのシステムと運用モデルの改善・構築、および担当者育成を行うことで、複合的な視点から共同リポジトリの構築・運用を支援し、共同リポジトリを全国規模で普及させることで、日本国内でのオープンアクセスの推進に寄与することを目的とします。

●2008年度計画

(1) EPrintsによる独立型システムモデルの構築・運用

(2) 担当者育成

(予定)

- ・ 9月29日(月) 情報交換会(埼玉県大学・短期大学図書館協議会)
- ・ 10月29日(水)~30日(木) DRF/Share地域ワークショップ(広島大学図書館)
- ・ 11月 長崎地区システム構築支援・研修会(長崎県大学図書館協議会)
研修会(新潟県大学図書館協議会)
- ・ 12月 ワークショップ DRF/Share-Yamagata(山形大学附属図書館)

●連携機関

広島大学図書館
岡山大学附属図書館
山形大学附属図書館
埼玉大学図書館
広島工業大学附属図書館
長崎国際大学附属図書館
北海道大学附属図書館(DRF)

<http://www.lib.hiroshima-u.ac.jp/share/share.html>

共同リポジトリプロジェクト報告書 -国内の地域共同リポジトリの分析-

<http://ir.lib.hiroshima-u.ac.jp/metadb/up/ZZT00002/ShareReport.pdf>

4年前の報告書で、すっかり忘れていたが
ネタ探しに昔のファイルを見ていたら発見。

みすみは

第5章共同リポジトリのコスト分析
を書いた。

共同リポジトリプロジェクト報告書 -国内の地域共同リポジトリの分析-



2010 年 3 月

共同リポジトリプロジェクト : ShaRe

山形大学・新潟大学・埼玉大学・文教大学
福井大学・岡山大学・広島大学・広島工業大学
山口大学・長崎国際大学・琉球大学
北海道大学・千葉大学・金沢大学・大阪大学

項目としては、コンテンツ作成コストばかりでなく、コンテンツ保存に要するコストや、広報・研修等の円滑な運用サポートのためのコストも包括的に組み込まれている。LIFEやBCLは英国のシステムをベースにしたものであり、そのまま日本の機関リポジトリに適用できるかどうかは検討が必要である。しかし、同様の包括的なコストモデルが無ければ、客観的な評価や検証は困難であり、**日本のシステムにあわせたモデルの構築**が望まれる。共同リポジトリについては、特有の問題もあり、共同リポジトリ向けのコストモデルの構築が必要となるであろう。

システムについては、幾つかのベンダーでクラウド型サービスの提供を開始しており、また**NIIもシステム基盤（ハードウェア・ソフトウェア）提供を告知**している。システム構築の際には、それらのサービスとの**コストの比較検討**も必要であろう。

当時は本気でコスト計算しようなんて思ってなかった

しかし、今こそコスト計算が必要なのではないか？

JAIRO Cloudに乗り換えるか？ 乗り換ええないか？

流されない！

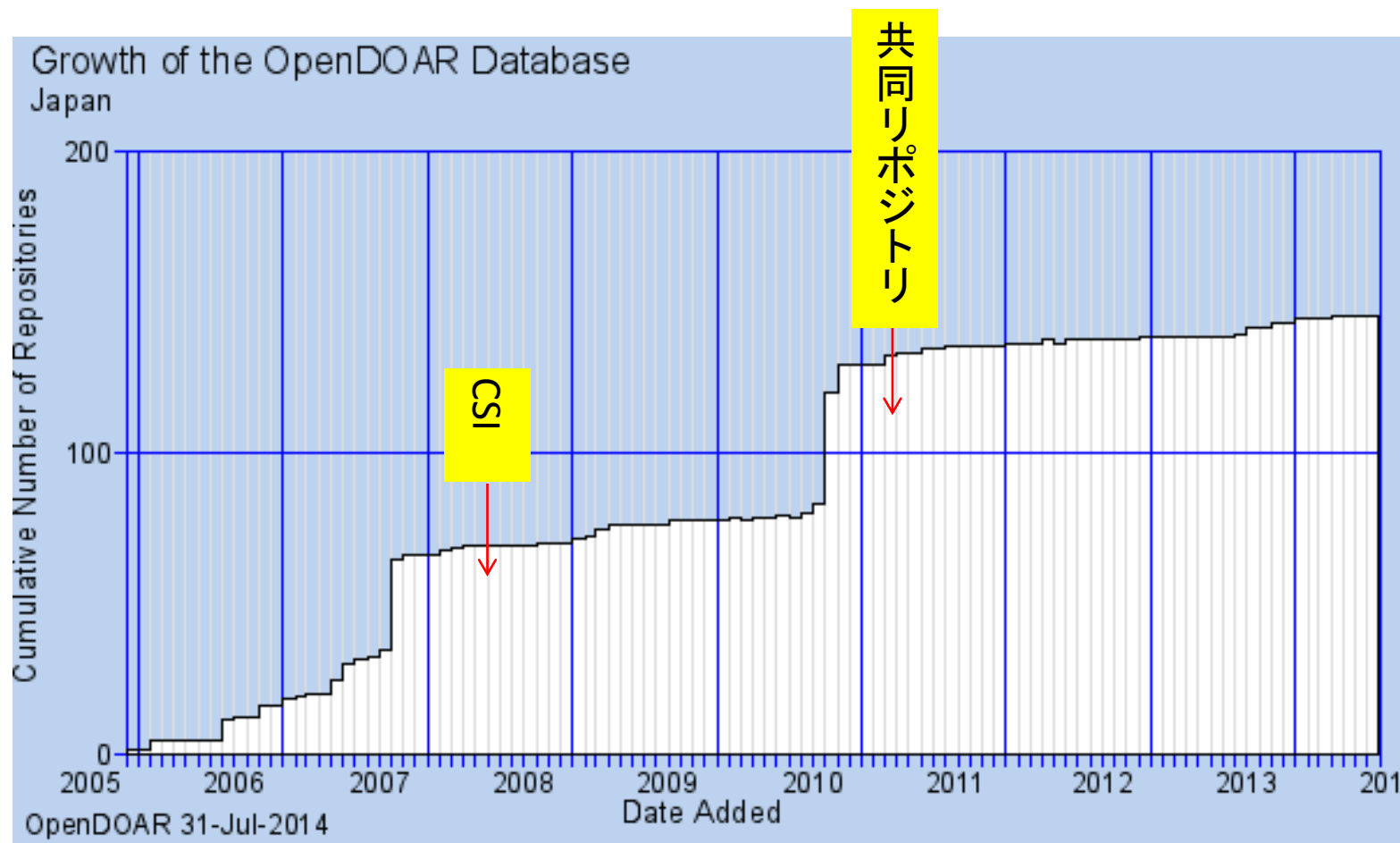
JAIRO Cloudとは

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。

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。

機関リポジトリの裾野

OpenDOARにおけるリポジトリの累積グラフ: 日本

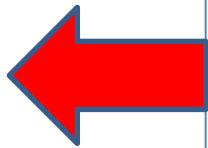


※OpenDOAR <http://www.opendoar.org/> 「Repository Statistics」より

機関リポジトリシステムの運営に必要なコストは大きく二つ

1) メタデータ・コンテンツのマネージメント

2) システムのマネージメント



「システムのマネージメントはサービスの本質ではない、人的リソースを削減できるところは削減し、浮いたリソースは他のサービスにまわすべきである」

千葉大学の機関リポジトリとは

千葉大学の機関リポジトリ「千葉大学学術成果リポジトリ（**CURATOR**: Chiba University 's Repository for Access To Outcomes from Research）」は、千葉大学内で生産された電子的な知的生産物（学術論文、学位論文、プレプリント、統計・実験データなどの学術情報）を蓄積、保存し、学内外に公開することを目的とし、**平成15年3月に試験公開を開始し、平成17年2月に正式公開をおこなった。日本国内において最初に構築された機関リポジトリである。**

日本最初の機関リポジトリ！

国立情報学研究所の学術機関リポジトリ構築支援事業にも積極的に参画した。

主担当をつとめたプロジェクトテーマとしては、

- ・機関リポジトリの評価システム
- ・研究コミュニティ創出支援
- ・e-Science 基盤構築のための
データ・キュレーション機能拡充

がある。

CURATOR
10周年
イベント

千葉大学CURATOR10周年
リニューアルオープン



学術情報のオープン化が導く大学の将来
土屋俊氏(大学評価・学位授与機構教授)

20世紀におけるソフトウェアのオープン化に端を発し、21世紀には研究成果のオープン化、学習資源のオープン化、高等教育のオープン化として観察される知識のオープン化の動向は、学術、知識の役割の再考を促す要素となっている。大学という脈絡においてこの動向は、大学図書館の役割を超え、大学の社会的機能、役割の変容の原因あるいは結果として理解することができる。

とくに近年の関心のひとつは、誰もが負担なしに利用できる便宜の質の維持である。オープン化した環境において、たとえば査読による質の維持、管理、大学システムの観点からの評価による質の保証という概念とそれに基づく方法の実効性を信じていることができるのかという問題を論じる。

2014年度 第3回 千葉大学アカデミック・リンク・セミナー
オープン化する研究情報流通と
学習との接点

日時：2014年10月21日(水) 14:30~16:30

会場：アカデミック・リンク・センター
1 棟1 階コンテンツスタジオ
終了後、リニューアルオープンした1 棟の見学会を開催します。

千葉大学CURATORの10年
オープンアクセスの10年

尾城孝一氏

(国立情報学研究所学術基盤推進部次長)

CURATORの開設から10年このかたの、オープンアクセス思潮をとりまく動向を概観し、現在の大学及び大学図書館が直面している学術情報流通の諸課題の今後について展望する。



<お申込み方法>

Academic Link

氏名・所属・職名・連絡先 (E-mail) をご記入の上、前日までにメールでお送りください。
宛先: alc-info@chiba-u.jp
件名: セミナー申込み

問い合わせ先：千葉大学アカデミック・リンク・センター
http://alc.chiba-u.jp/ Tel:043-290-2243, FAX:043-290-2255

千葉大学のポリシー

Principle of Principled Promiscuity (原則的無節操の原則)

大学のアクティビティの成果とみなせるものは、何でも受け入れる。

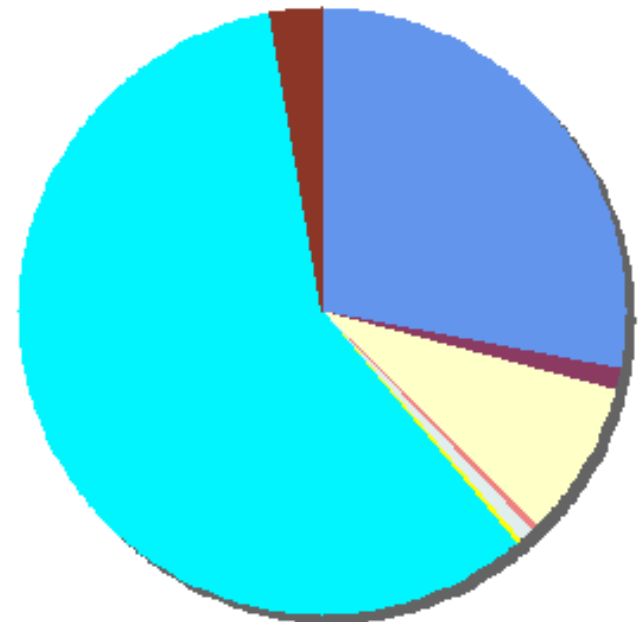
そもそも発信の主体者は、コンテンツの作成者でありセルフアーカイビングが原則であって、図書館ではない。図書館の役割は発信のサポートである。

日本語でも英語でも、あまり上品な表現ではないが・・・

コンテンツ構成

資源タイプ別コンテンツ数内訳(割合)

Journal Article (学術雑誌論文)	25,169 (28.2%)
Thesis or Dissertation (学位論文)	887 (1.0%)
Departmental Bulletin Paper (紀要論文)	7,439 (8.3%)
Conference Paper (会議発表論文)	30 (0.0%)
Presentation (会議発表用資料)	115 (0.1%)
Book (図書)	35 (0.0%)
Technical Report (テクニカルレポート)	35 (0.0%)
Research Paper (研究報告書)	803 (0.9%)
Article (一般雑誌記事)	0 (0.0%)
Preprint (プレプリント)	12 (0.0%)
Learning Material (教材)	210 (0.2%)
Data or Dataset (データ・データベース)	52,223 (58.4%)
Software (ソフトウェア)	0 (0.0%)
Others (その他)	2,436 (2.7%)
合計	89,394



Data or Database
が圧倒的に多いのが、千葉大の特徴

E-repository

- 千葉大CURATOR構築のためにCMSが開発

<http://www.cmssc.co.jp/modules/office/index.php?id=1>

- つまり、日本最初の機関リポジトリ管理システム
- 比較的、構造はシンプル

E-repositoryを一番最初に導入し、一番実績のある千葉大が、一番最初に移行するのは自然と言えば自然な流れ？

実験の流れ

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の仕様作成に協力

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

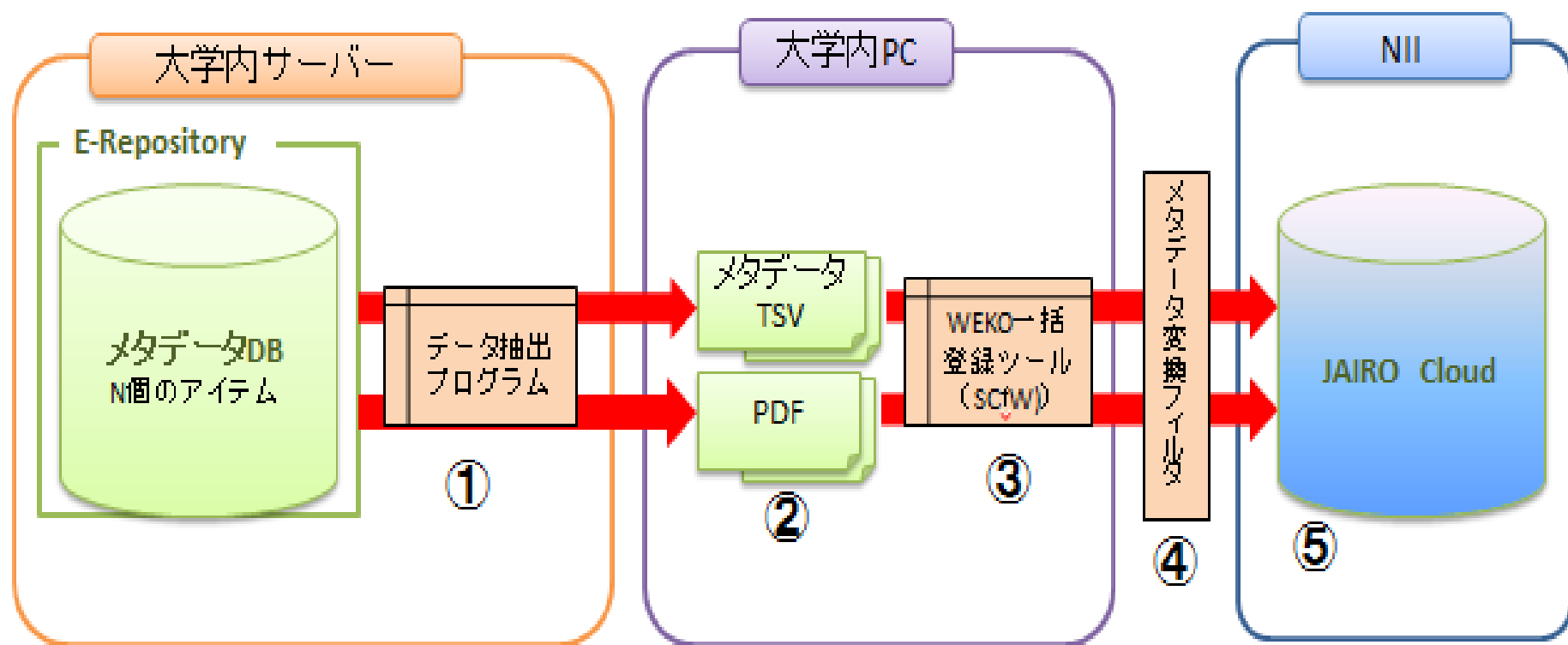
実験参加機関は、各大学の環境で実行し課題点を抽出、各機関リポジトリシステムから、JAIRO Cloudへのデータコンバート用のフィルタを作成。

作業フロー

抽出フェーズ

変換フェーズ

登録フェーズ



①データ抽出
プログラム
インストール

②データ抽出
プログラムによる
データ抽出

③WEKO
一括登録ツール
インストール

④メタデータ
変換フィルタ
作成

⑤データの
一括登録

作業結果

→ 計画
→ 実際

		4月	5月	6月	7月	8月	9月	10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							→	→

- フィルター最終修正作業が少し残っている
- 最終報告書は提出、NIIの確認待ち。

作業にかかった、おおよその日数

サーバへの抽出ツールインストール	半日
サーバからのデータ抽出テスト	2日
フィルタ作成～サンプルロードテスト	4週間
大量ロードテスト	3日
データ検証	3週間

← 実験なので、いろいろ余計な事も試しているし（いろいろな失敗？）、他の業務の片手間だったので、時間がかかってしまった（実は、みすみはリポジットの実務担当者ではなくて、ブランクもあったのでてまどった・・・）

現役の担当者ならもっとスムーズにできると思う。

集中的に作業した場合の日数

多分、こんなもんだろう

データ件数1万件であれば

- | | |
|------------|-----|
| 1) データ抽出 | 2日 |
| 2) サンプルテスト | 1週間 |
| 3) 全データロード | 3日 |

本日、一番重要なスライド

← 過去に凝った事はしていないし、今回もしない、という前提。

変換フィルターをいじりはじめると、もっとかかる。
実験では一から作ったので時間がかかったし、修正もけっこう手間がかかる。

移行実験の実際

- 当初計画の終了予定であった、9月中に終了させるべく少々強引に進めた。検討に時間をかけずに、トライアンドエラーで進めたため、失敗しまくっている。詳細は10分では無理。
- 作業の概要は、最終報告書にまとめてNIIに提出済みだが、まだ未確定。
- そもそも、失敗だらけの実験の実際を話しても、参考にならないと思うので、モデル的な進め方(移行の手引き)を作成で、それに沿って説明する。

JAIRO Cloud移行の手引き

【1】検討

【2】計画

【3】実行

【4】移行終了後

【1】検討

1. JAIRO Cloudの特徴

2. 移行前に調べておかなければならないこと

1) 機関リポジトリのシステム

○ 運用システムの種類

- ・ Dspace、XooNips、E-Repository、NALIS-RはNII側でデータ抽出ツールと変換フィルタを準備済み

- ・ その他のシステムの場合は、所定のフォーマットのメタデータと本文ファイル、メタデータ変換フィルタを準備する必要あり。

- ・ NIIが抽出ツール準備済みでも、Peal、データベース接続用のPealモジュールがインストールされていることが前提なので要確認。

追加インストールが必要になる可能性あり

○ ディスク容量の確認

圧縮作業で、作業領域が必要となる

【1】検討

2) コンテンツの中身の確認

- ・移行コンテンツ数の確認（移行期間に影響）
- ・HTMLファイルの場合、本文中の画像をJAIR0 Cloud上で開くことができない（パスが自動では書き換わらない）。
- ・学内限定公開コンテンツは、JAIR0 Cloudでは対応できない（認証が必要になる）。

【2】計画

- ・移行コンテンツの確定
- ・移行後の利用設定の確認

クリエイティブコモンズ対応で、柔軟な利用許諾方法が

選択可能

- ・データコンバートフィルタの確認
- ・作業担当者の決定
- ・NIIへの申請
- ・移行スケジュールの確認

【3】実行

＜移行ツール対応のシステムの場合＞

1. 事前準備

- 移行ツールのセットアップ
- FTPの設定
- scfwのセットアップ
 - i) インストール
 - ii) 標準フィルタの設定および各機関にあわせて修正
 - iii) JAIR0 Cloudへの接続設定および接続確認
- JAIR0 Cloudのユーザー設定
 - # 管理者設定および作業用アカウント登録

【3】実行

2. 移行作業

※ いろいろ方法はあるでしょうが、2000件から3000件程度のロット毎にデータ抽出～データ登録作業を行うことをおすすめします。

1. データ抽出ツールのparm.iniを設定

2. データ抽出実行

3. データをFTPでローカルPCへ

4. scfwで一括登録

5. JAIR0 Cloudでアップロード確認確認

データが確認できたら、上記の1に戻り、登録を繰り返す

※慣れてきたら、1～3と4は平行作業も可です

すべてのデータ登録が終了したらインデックス、アイテムタイプの確認を行ってください。特にアイテムタイプは、重複でできてしまっているので整理してください。

【3】実行

＜移行ツールが未対応のシステムの場合＞

- ・ 何らかの方法でメタデータと本文ファイルを抽出し、ローカルPCへ
- ・ scfw のルールにあわせて、メタデータファイルを加工
- ・ コンバートフィルタを作成
- ・ 一括登録

基本的に個別対応となりますので、NII側と十分に相談してください

【4】移行終了後

1. データ確認の上、NIIに報告
2. 公開日の決定
3. 旧システムの停止日の決定
4. 広報作業
5. 移行！

E-repository標準フィルタについて

○アイテムの区分をどうするか？

#標準フィルタでは、NII typeで区分したが
E-repository独自の区分とはズレて
いる可能性がある

○移行時にURIが変わってしまう

#標準フィルタでも、旧URLとして移行する
設定にはしたが・・・

○未対応の項目 (selfDOI、isbn、NAID、ichushi)
については、今回は設定していない

本移行にむけて

○ 実験を通して、JAIRO Cloud (WEKO)の仕様として、ユーザー側が、メタデータ仕様を柔軟に設定できるシステムであることが分かった。

← メタデータ項目を自由に拡張できる！！

○ 研究データや、研究助成情報とかを格納できるように拡張することも簡単にできそう。

← ハーベスティングは可能か？

← とりあえずは手動エクスポートでの対応か？

○ 研究データのメタデータのスキーマの検討

チョット
宣伝

大学の知の発信システムの構築に向けて—機関リポジトリの
新たな可能性を探る研究データへのアプローチ:エジンバラ大学図
書館の実践事例を中心に

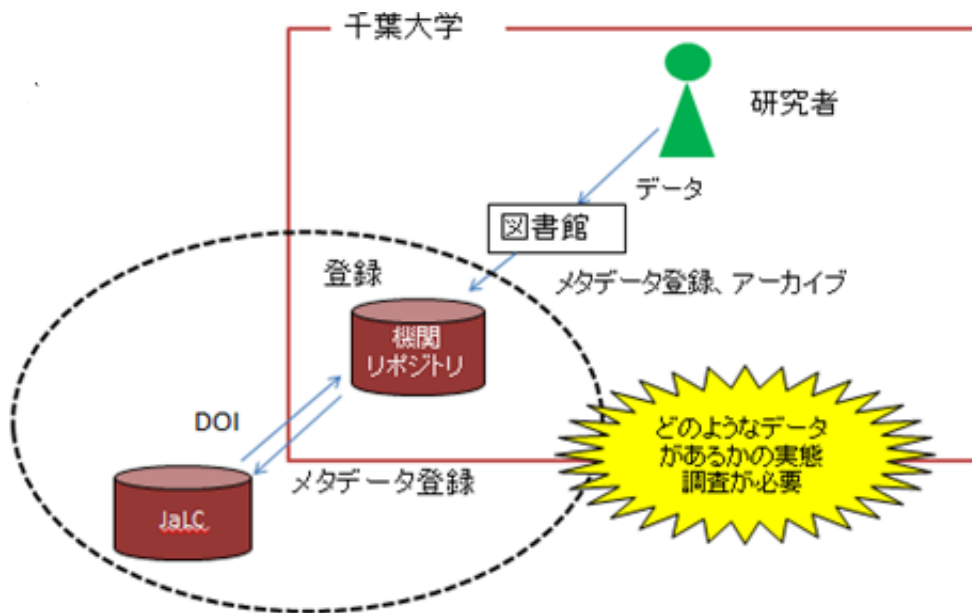
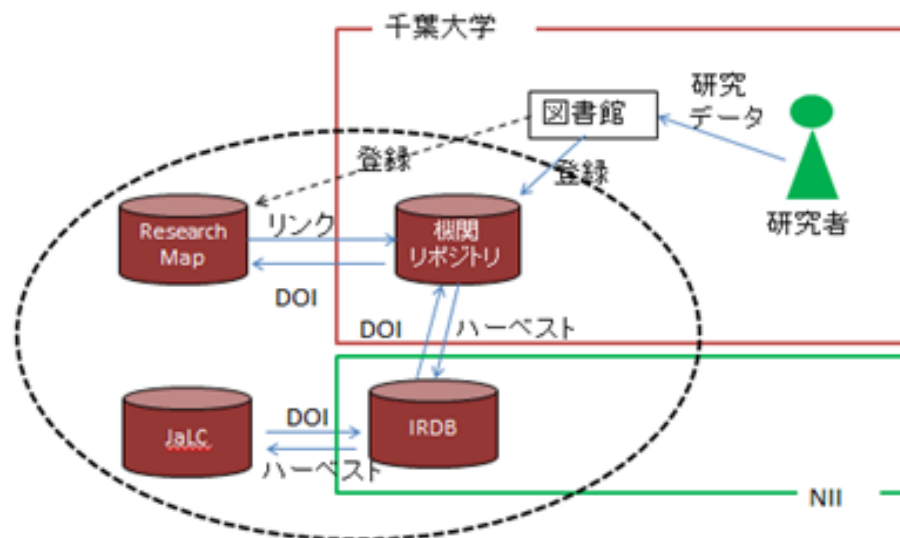
「図書館総合展 11.6 13:00-15:00

<http://2014.libraryfair.jp/node/2113>

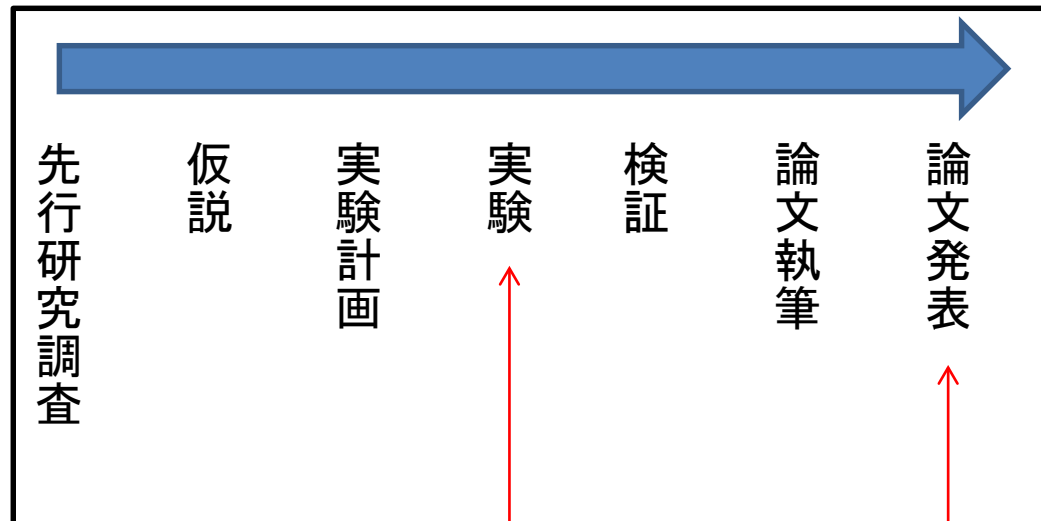
個人的には
こういうことが
やってみたい！

学内の調整も検討も
まだ何もやっていない
が...

一般的な研究データ登録フロー



研究データのオープンアクセス化



一般的な研究サイクル

科学において、論文は研究のエッセンスではあるが、あくまでエッセンスに過ぎず論文中のデータは一部でしかなく、データが無ければ、その結果を他の研究者が継承・発展させるのは困難

ここで産み出される
膨大なデータも
オープンアクセスの
ターゲットに！

今までの
オープンアクセスの
ターゲット

ファウンダー側としては、発展性
が無い研究への出資は望ましくない

しかし、分野によって文化が大きく異なるので、一律に進めるのは無理だが、データのオープン化の方向で進むのは間違いない

政策レベルでは、リサーチデータのオープン化が進められている

メタデータの拡張

○Junii2にはないが、**著者ID**は必要。

← 無いとResearchMapにデータを流せないなので
ResearchMapにあわせる

○助成情報(助成機関、プログラム名、年度、プロジェクト番号)も必要

← 助成機関はCrossRefで良いと思うが、そのまま表記すると、利用者から見ても何だかわからないので、日本語表記項目をつけたほうが良いかもしれない。プログラム名のIDがあるかどうかかわからないし、表記が正規化されているかわからないが、助成機関と調整の上、入力にブレがないようにする必要がある。

← ブレると統計が狂う

最後に

裾野は、CSI事業に牽引された単独の機関リポジトリから、2010年頃には地域共同リポジトリに移った。そして今、共有リポジトリに移りつつあるのは確かである。しかし、地域共同リポジトリにしても、ハードを用意すれば進むというものではなかった。地方の場合は、地方国立大学がホスト校と運用の中心となっているケースがほとんどであった。

そして、その地域共同リポジトリが活性化するかどうかは、ホスト大学の「本気度」にかかっていたし、それはキーマン（ウーマン）の本気度でもあった。業務を粛々とこなせばその延長上で立ち上がる、というものではなく、**ホスト大学にキーマン**

（ウーマン）がデンと構えていて、「A大学に聞けばいい」ではなく「A大のBさんに聞けばいい」という信頼関係を確立できて、はじめて動いている、というように見えた。

共有リポジトリでも、結局は「人」ではなかろうか。ハコを用意すれば良いというものではないし、ルールを作れば良いというものでもない。ホスト、つまりNII側が「顔の見える」「頼りになる」キーマン（ウーマン）を立てられるかどうか、JAIRO Cloudの発展の鍵となるのではないかと考えている

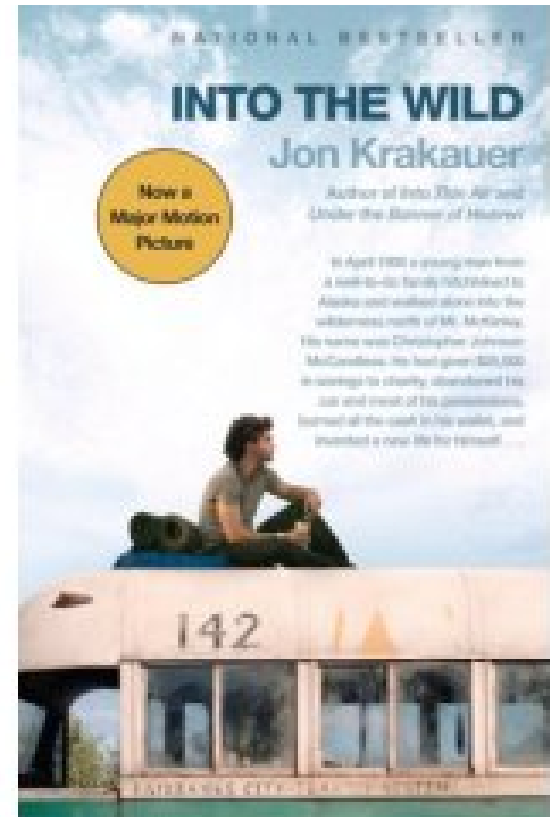
リポジトリ、人、ネットワーク
をいかに育てていくか？

← 2009年のDRF-ShaRe-Hiroshimaのプレゼンと同じ結論だが・・・

Happiness Only Real When Shared

from “Into the Wild

“Jon Krakauer



Thank You!



Ashizuri Kaiyokan Aquarium, Kouchi

misumit@chiba-u.jp