

放医研機関リポジトリの再構築

○下村岳夫^{A)}，大竹淳^{A)}

^{A)}独立行政法人放射線医学総合研究所 研究基盤センター情報基盤部 科学情報課

1. はじめに

独立行政法人放射線医学総合研究所(以下放医研)では2002年以降、所の研究成果に代表される業務実績にかかる電子情報を蓄積・公開するシステムを独自開発し、運用してきた。収集された実績情報は個人・部門・組織の評価にも利用される。しかしその格納対象はテキストデータに限定される事、入力操作の煩雑さ、所外から隔離されガラパゴス化してしまっている事による弊害など、問題点を多く抱えていた。特に学術情報を取り巻く世の環境はここ数年で大きく変化し、世界標準的なデータベースインフラが広く整えられつつある[1]。これらに準拠した所外の文献データベースや検索システムと、所内の研究成果データベースの連携が全く出来ない状況は、研究所の重要課題である国際研究開発競争力向上において大きな阻害要因にすらなりつつあり、研究・事務両部門からシステム刷新を求める声が上がっていた。

2. 目的

旧システムが抱えていた問題点の抜本的対策を講じる事で、「所の業務実績を効率良く収集・蓄積・公開する」従来機能のパワーアップ版を提供する。これの運用により、所の情報発信力強化とユーザ利便性の向上を図る事を目的とする。

3. 方法

3-1) 学術情報流通にかかる世の状況把握と技術調査

- ー平成 22 年度学術ポータル担当者研修参加:「機関リポジトリとは何か」「その技術や課題は何か」を把握。
- ー国際カンファレンス eSciDoc days (2010 年 11 月:コペンハーゲン) への参加 with 放医研将来構想発表。
- ー所内外の図書担当者・他研究所リポジトリ担当者・ミドルウェア提供者へのヒアリング。
- ー文献・Web 情報の収集、解析。

3-2) 所の業務分析&部門間調整及び旧システムの問題点分析

- ー企画・研究推進・総務・人事・知財など各事務部門を巻き込んで打合せを重ね、情報収集と調整を実施。必要に応じて業務フローや運用ルールを変更してもらう場面も。
- ー旧システムを徹底的に調べ上げ、問題点の列挙→原因・解決策の抽出→工数・効果・リスクを勘案した優先順位付けにより、開発方針を模索。

3-3) ミドルウェア選定におけるトライ&エラー

外部サービス連携性を高めるべく、メタデータを OAI-PMH などの世界標準規格に準拠させ、所外検索サービスからのハーベスティングに対応させる為には、専用ミドルウェアの利用が欠かせない[2] [3]。放医研でも開発に先立ち、幾つかの候補をテスト利用してみた。

ーeSciDoc: 独マックスプランク研究所・カールスルーエー大学

日本の研究所における採用もある[4]が国内の認知度は低い。環境構築後、協力者と連携して試用してみるも、当時日本語対応が不十分であった事、ユーザインターフェースのカスタマイズが困難で、放医研の業務に馴染まない事などから採用を断念。

－DSpace: 米マサチューセッツ工科大学・ヒューレット・パッカード社[5]

国内の大学・研究機関でリポジトリ構築に広く利用されており、日本におけるデファクトスタンダード的地位を確立している。環境構築後、技術情報収集を試みるも意外と公開されている情報は少ない。導入に際しては業者にカスタマイズを任せているケースが多いらしく、自由度の高い開発を狙う放医研のニーズに沿わない事から採用を断念。

－WEKO: 国立情報学研究所 (NII)

同研究所提供の CMS "NetCommons" のリポジトリ構築専用モジュール[6]。DSpace の機能を網羅しており、かつ日本語対応もしっかり行われている。カスタマイズに対する柔軟性が高く、現在これを所外公開サーバに採用する方向で進めている。

3-4) 情報発信力強化／利便性向上のための主な方策

- －所内外他システムの Web-API を利用したシステム連携(データインポート機能)
- －自然な流れに沿った入力手順と、入力分類の統廃合(ユーザが迷わないインターフェース)
- －本文 pdf ファイル等の登録と公開設定(ファイルアップロード機能と、公開／非公開制御)
- －入力文字列の正確さとマスタメンテナンスにかかる労力のトレードオフ(マスタ情報管理の一部廃止)
- －相互チェックによる情報の正確性・正当性担保(共著者間コミュニケーション機能)

3-5) 所謂『機関リポジトリ』としての公開に向けて

旧システムも業績情報を蓄積・所外公開しているという点で、広い意味で「機関リポジトリ」と言えたかも知れないが、例えば原著論文はアブストラクトを含む書誌情報までと、その公開範囲は十分な物では無かった[7]。新システムは本文ファイルアップロード&公開機能を擁し、今後導入予定の所外向け外部サーバとのリアルタイム連携を実現させる方向で準備を進めている。運用面の課題として問われる著作権問題は、登録時の著者相互チェックに図書室が加わる事で解決する[8]。

4. 結果

職員の実績評価時期に合わせ 2013 年 11 月、旧システムから 10 余年分の業績データ約 38,000 件を移行し、ユーザ説明会を経て所内向けサーバを新システムに切替えた。以来半年間、今の所大きな問題無く運用を継続できている。職員や組織の評価に際しシステムを直接利用できる様になった事、所の研究活動を俯瞰的に把握できる充実した集計ができる様になった事などは、複数の事務部門担当者だけでなく上層部からも好意的に受け止められている。一方でユーザによるマスタメンテナンスの手間を省き、分かり易い自然な流れで入力できる様改善した点、また所の研究者が多く利用する PubMed(医学・生物学分野の学術文献サービス)や幾つかの所内業務 Web システムからのデータインポート機能を実装する事で、データ入力にかかる労力を大幅に軽減させた点は、利用者から高く評価されている。機関リポジトリとしての完成形は『新』所外向け公開サーバとの連携が完了する事であり、現在本年(2014 年)夏の公開を目指し準備を進めている。

5. 考察

2011 年 4 月より調査を開始し 3 年余りの時間を要した当開発も、大詰めを迎えている。旧放医研リポジトリは、所内向けサーバが「業務実績登録システム」、所外向けサーバが「発表論文等データベース」と呼ばれ、これら 2 つのシステムを夜間バッチで部分的に一方向連携(所内→所外)させた物であった。新システムもこのサーバ構成は守りつつ、リアルタイム反映、pdf などの添付ファイル登録及び公開、世界標準フォーマット準拠による検索サービスのハーベスティング対応、所外 Web-API サービス連携によるデータインポート(現状 PubMed のみ、将来的には CrossRef・KAKEN・researchmap・SHERPA/RoMEO 等との連携も視野に入れている)、などが実現できる見通しであり、デー

タの蓄積・公開両側面について大幅な改良が行われる事となる。これにより①データ入力／管理作業にかかる工数の大幅削減、②業績を更に広い分野に向け、かつ詳細情報を付与して PR できる、③論文被引用機会の新規創出、④個人・部門・組織毎の業績一括管理がより容易になる、など放医研にとってのメリットが多く得られるばかりでなく、所の研究成果情報をより広く具体的に発信する事で、より多くの方に活用して貰える、といった社会貢献にも大きく寄与できると考える。

所内向けの業務実績登録システムに蓄積される情報は所の全活動と、それに関連する報告が網羅される為、非常に多岐に渡る。旧システムではこの分類が非常に複雑で 1 件データを登録する為に作業者がユーザーマニュアルを参照しなければならないケースも多々あった。一方でその分類は集計上ほとんど利用されていないといった状況がある事も、調査を通じて浮き彫りとなった。このような無駄を一つ一つ削り落とす作業には非常に多くの時間と労力を要したが、結果として所の業務フローに最適化された、柔軟性の高いシステムを構築できたと自負している。

6. 参考文献

- [1]筑木一郎, 学術情報流通と大学図書館の学術情報サービス, カレントアウェアネス, No293, P21-30, 2007
- [2]尾城孝一, OAI-PMH をめぐる動向, カレントアウェアネス, No278, P12-13, 2003
- [3]吉田暁史 横谷弘美, 学術情報流通における OAI-PMH の役割, 大手前大学論集, 第 10 号, P237-257, 2009
- [4]高久雅生 谷藤幹子, 材料系研究所における機関リポジトリ NIMS eSciDoc の開発から応用まで 研究者総覧 SAMURAI と研究ライブラリコレクション, 情報管理, 55(1), P29-41, 2012
- [5]MacKenzie smith, DSpace An Open Source Dynamic Digital Repository, D-Lib Magazine, Vol9(1), 2003
- [6]山地一禎, 学術資源共有基盤 WEKO の開発, デジタル図書館, 36, P51-61, 2009
- [7]倉田敬子, 機関リポジトリとは何か, MediaNet, No13, P14-17, 2006
- [8]谷山秀幸 杉本若葉, 機関リポジトリと著作権. 概論, 平成 21 年度学術ポータル担当者研修資料, P1-19, 2003