



特集1 機関リポジトリ — 病院図書館でもできるのか? —

機関リポジトリシステム導入とランニングコスト

福田 典雅

I. はじめに

今日は、システム的なお話ではなく、機関リポジトリを実際に構築して、それに対してどのくらいのコストがかかるか、そしてみなさんはどういうところに悩まれているところかについてお話ししたいと思います。

まずアグレックス (以下、AGREX) という会社をまったくご存知ないと思います。詳しくはホームページをまた調べてください。当社の特徴のひとつは、BPO (ビジネスプロセスアウトソーシング) という私どもが生み出した言葉ですけれども、システムだけを販売または構築といった業務を請け負うものとは違って、業務をすべてアウトソーシング化して、当社に取り込み、それをコストパフォーマンスを出して対応するという業務展開をしています。

今回の機関リポジトリで言うと、ただシステムを構築するだけではなく、運用のところ、たとえば許諾を聞いていくとか、OCRやメタデータの作成、一括登録などを行います。あとはシステムの保守・運用の保守、こういった形で業務全体をサポートしています。

今日は施設関係の方が多々いらっしゃると思います。当社はCROやSMOなどといった治験にかかわる業務も行っています。私どものメンバーがまた接触することもあるかと思いますが、その時はACメディカルというAGREXの100パーセント子会社になりますので、よろしくお願ひします。

今日発表させていただく内容ですが、まず導

入管理、導入ソフト運用パターンと導入実績、そして、共同リポジトリの実例、資料の電子化、著作権の処理、PDF、メタデータといった共同リポジトリの実例について、ご提案というか、こういう風に解決もできますよというひとつの投げかけをしようと思っています。

II. 機関リポジトリシステム導入関連

1. 導入ソフト

まず導入ソフトは、DSpaceとJAIRO Cloudの二大勢力になっています。『月刊DRF』2013年7月号によれば、DSpaceが96サイト運営されています。JAIRO Cloud、私たちはWEKOと呼んでいますが、こちらは71ユーザーの利用実績があり、申し込みは100ユーザー以上あります。

それら以外の導入例を表1にまとめました。XooNips (ズーニップス) は、慶応義塾大学、新潟大学、近畿大学や天理大学などが使っています。Eprintsは北海道大学が使っています。NARIS-Rは株式会社NTTデータがDSpaceをもとに作成した製品です。ここまではオープンソースといって誰でも使える自由なソフトです。

表1 ソフトウェアと導入機関例

ソフトウェア	導入機関例
XooNips (ズーニップス)	慶応大、新潟大学、近畿大学、天理大学など
ePrints	北海道大学など
NARIS-R	
eRepository	新潟大学、千葉大学、広島大学、CNS社など
infoLib-DBR	神戸大、山口大など
iLisSurf	同志社大学など

ふくだ つねまさ:アグレックス

あとは商業用のサービスというか商品です。eRepository は新潟大学、千葉大学、広島大学などが使っています。これは CNS 社が提供しています。infoLib-DB は、神戸大学や山口大学などが使っています。富士通の iLis Surf は同志社大学などが使っています。商業用は非常にコストも高いと私たちは聞いています。

2. 運用パターン

運用パターンは、機関内にサーバーを置く場合と機関外（外部）にサーバーを置く場合の2つがあります。まず機関外にサーバーを設置する共同リポジトリの場合（表2）は、奈良県立医科大学さんみたいに大学様が主体になり、ほかの病院がそこにのっていきます。だから、どこかにサーバーが必ず置かなければどうしようもないという状況です。

表2 機関外設置の場合の運用パターン

設置パターン	詳細
共同リポジトリ	どこかの機関が代表して施設内にサーバーを設置する
商用環境	1) Amazon などのクラウドサービスを使用して、ハードウェアを利用する 2) AGREX などの SaaS (ASP) サービスを使用して、ハードウェア、ソフトウェアを利用する
NII 環境	JAIRO Cloud を利用して、ハードウェア、ソフトウェアを利用する

商用管理のところは、当社でも使用している Amazon のクラウドサービスというハードウェアを利用し、ソフトを改良する方法です。Amazon のクラウドは東京都内にサーバーセンターやシステムセンターがあり、おおよその場所の特定はできます。しかしどこにあるのかわからないため、監査はできません。Amazon を信用するしかないのです。逆に言うと場所がわからないのでセキュリティ面では、いいかなとは思いますが。

当社では、サーバーの AGREX の SaaS 版である ASPaaS を提供しています。ASPaaS ではハードウェアとソフトウェアを使用します。データセンターは東京と大阪にあり、そちらを

利用します。

NII 環境は JAIRO Cloud を利用したソフトウェアです。

それぞれ利点はありますが、まず、Amazon はソフトウェアの用意や導入、運用管理が必要になります。2 番目の AGREX の ASPaaS は運用費用が必要になり、カスタマイズの自由度が若干低くなります。通常 DSpace では 100 アイテム以上、利用者の利便性を高めるためであるとか、管理者のためのツールというのが用意され、カスタマイズができます。

そういった意味では、NII が提供する JAIRO Cloud は、たとえば画面ひとつとってもカスタマイズに自由度がありません。

3. 導入実績

当社での導入ユーザーは、学内にサーバーを置いているユーザーが 41 あり、ASPaaS 使用は 17 ユーザー、保守利用が 48 ユーザーあります。

東京の大学の中には職員数の少ない施設もあります。また何年かごとに周期的に人が異動するため専任者が置けない状況です。そんな中で誰がリポジトリ導入にかかわる事務的な作業をするのが問題になっています。こういった問題を解決するために、当社では JAIRO Cloud 導入支援を提案しています。画面やメタデータを作る作業は、当社で行います。2014 年 2 月から 3 月にかけて 2 ユーザーの導入支援を予定しています。著作権の処理は、すでに終わっているところも現在進行形のところもありますが 18 ユーザーに対応しています。

DSpace のカスタマイズ例のひとつに PDF ファイルのカバーページ作成があります（図 1）。Google などの検索結果から直接 PDF を表示させる場合に備えて書誌事項を入れるのですが、データ量が大きいいため作成時はサーバー容量を考慮する必要があります。

ほかにはアクセス統計とかダウンロードの統計や、研究者データベースの著者連携などにも対応しています。最近ではスマートフォン向けのトップページを作成し、画面に収まる形で提



図1 PDF トップページ

供するサービスも展開しています。

4. 共同リポジトリの実例

いろいろな病院施設がまとまってリポジトリ公開する場合には、やはり共同リポジトリで展開しなければコストがかかりますし、事務局も大変です。そこで共同リポジトリの実例をお話します(表3)。奈良県立医科大学さんで行われている共同リポジトリは参加館募集中ですね。弘前大学、群馬大学、福井大学、ここまではAGREXが構築支援と保守を手がけています。規模の小さい大学は地域でまとまった共同リポジトリを運用しています。DSpace以外の実例では、eRepositoryがあります。これは広島県の共同リポジトリで使われています。

Ⅲ. 資料の電子化

資料の電子化に際してまずやらなければならないのは、リポジトリ公開するための運営指針の作成です。これらはインターネット上にありますので、それをコピーしてつくればあまり力はいらないと思います。

1. 著作権処理

著作権の問題は、やはり出版社の許諾が一番です。雑誌に記載された論文への場合は結構煩雑なところもあるため、当社では“グレーはあげない”方針で対応しています。共同で研究者発表される論文のトップの方には当然許諾はとりますが、たとえば4名いれば残りの3名の方にも必ず許諾を取らないといけません。あとあと問題になるということ当社も恐れていますので、基本は徹底して許諾処理をしております。また、出版社の許諾条件が著者の最終稿の場合は、当該の原稿の提出を先生に依頼します。ほとんどがこのような状態で行います。許諾については神経質にならなければあとで訴えられるという問題があります。

2. PDF化

運用で問題になるのがPDF化です。冊子体の場合はスキニングが必要です。PDFなどの電子媒体の場合は、PDFでの変換が必要です。また、全文検索できるような透明テキストをつけるためOCRをかけます。

今ではPDFにセキュリティを施して、ダウンロード後の改ざんを防ぐことも可能になりました。

表3 共同リポジトリの実例 (主にDSpace)

リポジトリ名、URL	参加館
奈良県立医科大学 http://ginmu.narmed-u.ac.jp/dspace	奈良県下の施設へ募集を行っている
弘前広域リポジトリ http://hrr.ul.hirosaki-u.ac.jp/	弘前大学、弘前学院大学、東北女子大、弘前医療福祉大学、東北女子短期大学、放送大学青森学習センター
群馬県地域共同リポジトリ https://gair.media.gunma-u.ac.jp/	群馬大学、育英短期大学、関東学園大学、関東短期大学、桐生大学・桐生大学短期大学、共愛学園前橋国際大学など
福井県地域共同リポジトリ http://crf.flib.u-fukui.ac.jp/dspace/	福井大学、福井県立大、福井工業大学、仁愛大学、仁愛女子短期大学、敦賀短期大学、福井医療短期大学
※ Dspace 以外 広島県大学共同リポジトリ http://harp.lib.hiroshima-u.ac.jp	

た。

3. メタデータ

メタデータとは、タイトルの項目など書誌の情報のことで、この作業が必要です。メーリングリストで事前に勉強されたそうですから、ここではメタデータの作成が必要なこと、実際に公開されるデータの入力、必ず書誌データを作るみたいな形で必要になることの説明にとどめます。

IV. 提案

1. 機器構成および費用

今までやってきて、コストがどのくらいかかるのかということをお話いたします（表4）。

表4 ソフト・ハードのメーカー別価格

	使用するメーカー	初期費用	月額
ハードウェア	AmazonAWS	10万円	25,000円
ソフトウェア	DSpaceを 機関内導入	80万円	35,000円～ 1人分の保 守込み
ソフトウェア	SaaS版導入 (ハードウェア不要)	20万円	35,000円～ 1人分の保 守込み

※ソフトウェアにはサーバ動作監視、環境バックアップを含む。
※保守費用は協議会から当社へお問い合わせ担当者数により増加します。

まず、Amazonを使用した場合の初期費用は、使用料初期10万、月額2万5千円が当社の標準価格です。容量は50GBです。ほとんどの大学機関では、この容量で対応できていますので、よほどのビッグユーザーでなければ対応できると思っています。あと、CPUの数やメモリ容量またはハードディスクの容量によって、月額単価が変わります。ハードウェアは以上の通りです。次にソフトウェアはDSpaceを使用する前提で学内、つまり機関内に導入する場合は、初期費用が80万円、月額3万5千円必要です。この月額費用は保守料です。1人分とありますが、ここはサービスできると思います。ただし、保守料はシステム保守だけではなく、運用保守もサポートしています。

誰が事務局などの担当者になっても、サポー

トは当社の長年のノウハウを生かして、ご指導または対応できますので、ご安心いただけたらと思っています。

SaaS版の導入の場合は、初期費用が15万、月額3万5千円という金額設定です。SaaS版はハードウェアを当社が提供しますので、お金はかかりません。3年ないし5年ごとに必ず行うサーバーのリプレース時の費用も無償です。機関内に導入された場合は加えてサーバー費が必要です。ハードディスクやバックアップのソフトなどもろもろがかかりますので、当社の利益なしでも100万強はサーバー代にかかるとなっています。

当然SaaS版には、サーバーの動作監視や環境バックアップも含まれたサービスですので、より安心してお使いになれると思っています。

2. 資料の電子化にかかる費用

資料を電子化する際の費用は、あくまでも参考金額です（表5）。特に著作権の許諾にはいろんな対応方法があるため価格に幅があります。調べたけれどダメでしたというのも1件とカウントします。成功報酬ではありません。だめでも1件につき1,000円かかります。

表5 電子化費用（一例）

	単位	価格
著作権の許諾※	1件	1,000円
冊子をPDF化（裁断可能）	1頁	10円
冊子をPDF化（裁断不可）		30円
OCR	1頁	50円
メタデータの付与	1論文	700円
メタデータ登録	500件まで	20,000円

※許諾が取れなかった場合にも必要

PDF化は、紙媒体を電子化したい場合は、裁断が可能かどうかで価格が異なります。つまり1冊しかない場合や貴重な冊子で裁断できない場合は、1ページあたり30円、裁断可能な場合は1ページあたり10円という価格設定です。1,000ページの1冊の本の場合PDF化に1万円かかります。データの場合、リポジトリ仕様に基づいたPDFであればPDF化の費用は不要で

す。ただし今までよくあるんですが、あまりにも解像度が低いPDFでは、OCRの作業でテキスト化できないことがあります。また、特に画像データなどのカラー画像は、やはりカラーで撮り、きれいな形で見せることを当社では考えているため、リポジトリ仕様に基づいていないPDFデータはお断りする場合があります。

OCR化は1ページ50円です。標準価格なので、高いとは思われないと思います。特にOCR化では、ページ内に英語と日本語がある場合、たとえばこれをAcrobatでとらえた場合、英語がOCRで読み取れず透明テキストができないこともあります。当社では1ページずつ見て、日本語と英語を分けてOCR化しています。

メタデータの作成は、ひとつひとつ論文を見て情報を書き込むため、当然機械化はできません。人が判断して入れるため、1論文あたり700円必要です。加えて登録代金は、1回500件までは2万円分と提示しています。

各仕様書および手順書を当社で作成します。事務局の方が交代しても、たとえば当社が手を引いた後にも、その運用を引き継いでいかなければ、精度を維持できません。そのために当社で作成した仕様書を必ず提供します。それをもとに引き継いで、精度を保っていく、そういう運用をお願いしています。

3. 構成と管理の方法

複数の病院で共同リポジトリを開く前提で提案します。図2は、相手が末端のところからの表記です。この場合は情報が分散される可能性は低いのですが、結構デメリットがあります。まず、管理主体が不明瞭なことです。どこの病院のなんという先生が書いたかも、一見してわかりにくいですし、病院単位へのハーベストに対応できません。加えて協議会全体の構成を検討する必要が出てくるため、非現実的かなと思います。

図3は病院単位で管理する場合です。施設の下層に科、たとえば脳外科や内科などのいろいろな科があり、所属の先生方が研究論文を記入し

ていく、そういう表記の仕方が考えられます。メリットは、管理主体がはっきりしているため、病院単位のひとつずつの構成を構築できます。またこれなら病院単位へのハーベストに対応できます。デメリットは、情報が分散化される可能性が高くなりますが、対策として、キーワードや分類などを入力すれば、その場で、その値でまとめて表示することもできます。

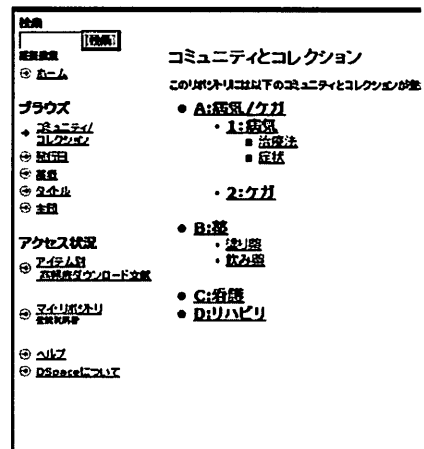


図2 リポジトリ構成例1

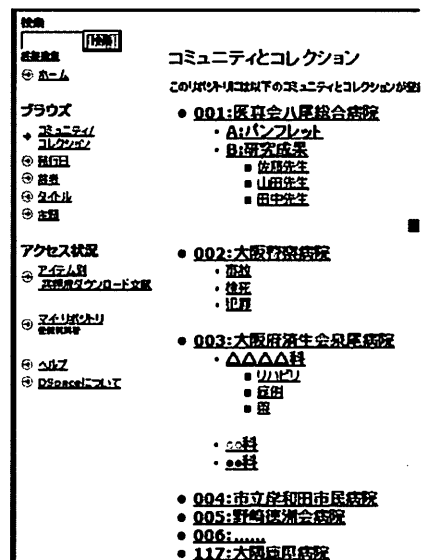


図3 リポジトリ構成例2

V. おわりに

当社は業者ですが、これまでもみなさんがリポジトリを立ち上げるにあたって、いろんなご支援を提供してきました。当社は今日・明日つぶれる会社ではありませんので、末永くご支援できることは、最大のメリットであると考え

ています。退職などで人は変わりますが、業務、サービスはお客様に提供する以上は低下はしません。担保されたサービスを提供できるのは最大のメリットかなと思っています。これで AGREX の提案は終わります。