

独立行政法人 科学技術振興機構

多様なデータソースを集約する 情報提供サイトで効率的な処理を実現

知の創出から研究成果の社会還元まで、我が国の科学技術基本計画の中核的な実施機関である独立行政法人科学技術振興機構(略称JST)。その事業の一環として、各種機関に蓄積された膨大な科学技術情報の活用するためのサービスを展開しています。JSTが進めるこうした文献や特許などの情報を集約・連携するサービス「J-GLOBAL」の構築にDataSpiderが採用されています。

Before

大量かつ多彩な情報を集約するため、効率的に前処理できるツールが必要だった



After

JST内でも、効率的にデータ処理スクリプトを開発することが可能になった

各分野バラバラだった 情報サービスのハブを作り上げる

我が国の科学技術基本計画の中核的な実施機関として「新技術の創出に資する研究」「新技術の企業化開発」「科学技術情報の流通促進」「研究開発に係る交流・支援」「科学技術理解増進」と、科学技術に関連した幅広い事業を手掛ける独立行政法人科学技術振興機構(略称JST)。

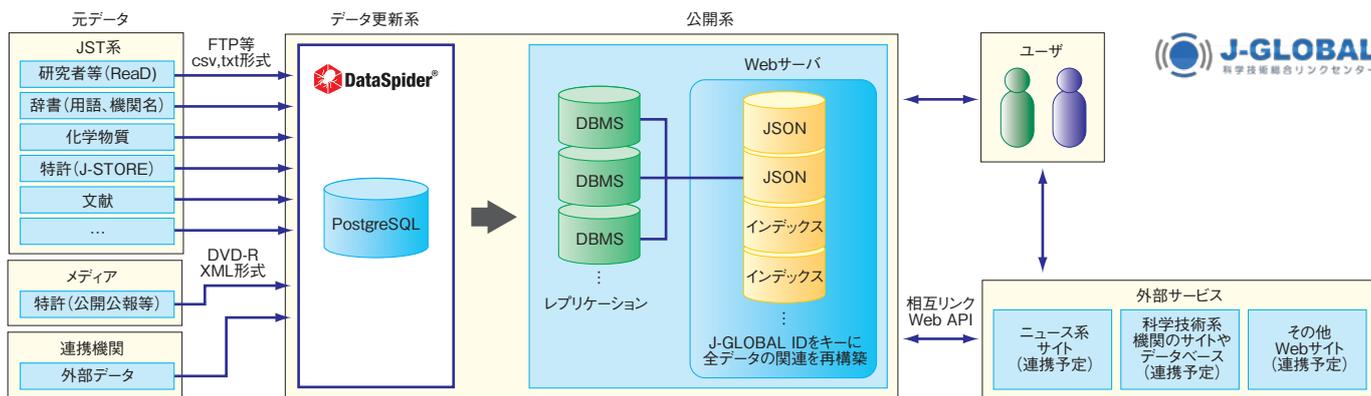
JSTは、前身である日本科学技術情報センターの時代から、文献情報、研究者情報、特許情報など、さまざまな科学技術情報を提供しています。これらのデータベースは科学技術や医学などの文献情報を提供する「JDreamII」、研究成果展開総合データベース「J-STORE」、研究開発支援総合ディレクトリ「ReaD」といった数々のサービスとして展開しています。

そして2006年、こうした多様なデータベースを連携させるべく、新たな開発がスタートし、2009年3月に科学技術総合リンクセンター「J-GLOBAL」としてサービスを開始しました。

「科学技術の幅が広がり、一つの方野だけで解決できない課題も増えてきました。JSTの目的の一つである「イノベーションを創出する」ためには、JSTをはじめ国内外の機関で各分野バラバラに提供されている情報サービスを結び、情報ハブとなるサービスを作り上げなければならないと考えたのです」と、イノベーション推進本部 研究基盤情報部 情報整備課 係長の黒沢努氏はその意義を説明しています。



独立行政法人科学技術振興機構
イノベーション推進本部
研究基盤情報部 情報整備課 係長
黒沢努氏



USER PROFILE



本 部 埼玉県川口市本町4-1-8 川口センタービル
代 表 者 理事長 北澤宏一
事業予算 1153億7600万円(2009年度)
お問い合わせ (JSTについて)
www-admin@tokyo.jst.go.jp
 (J-GLOBALについて)
<http://support-jglobal.jst.go.jp/>

PARTNER PROFILE



販 売 株式会社日本システムディベロップメント
本 社 東京都新宿区西新宿2-7-1
 新宿第一生命ビル16F
お問い合わせ 情報通信・公共システム本部
 03-3342-1200
E - M a i l DataSpider@nsd.co.jp
<http://www.nsd.co.jp>

PARTNER PROFILE



サ ポ ー ト 株式会社メディアフォース
本 社 東京都新宿区下落合2丁目3-18
 SKビル5F
お問い合わせ 事業統括本部 03-5988-2010
E - M a i l DataSpider@mforce.co.jp
<http://www.mforce.co.jp/>

異種かつ大容量のデータ処理を自分たちで行うためのツール



独立行政法人科学技術振興機構
イノベーション推進本部
研究基盤情報部 情報整備課
係長
関根基樹 氏

J-GLOBALのシステムを作り上げる上で課題となったのが、形式の異なるデータの統合です。既存の各データベースは、それぞれのサービスに最適化されているため、DBMSやデータベーススキーマが多岐にわたり、提供されるデータもtxtやcsvやXML、また、独自の形式など、使用する上ではさまざまな問題がありました。研究者名も、特許情報と科学論文では表記が異なり、名寄せの必要がありました。

また、J-GLOBALは扱うデータ量も膨大です。2009年の時点では、数の多いものだけでも、文献は約600万件、特許は約200万件、化学物質名称データベースが

約260万件となっています。より新鮮なデータを提供するためには、この大量のデータを迅速に処理する必要があります。

加えてもう一つ、JSTの事情がありました。独立行政法人として、外部委託を行う際には一般競争入札が義務づけられているのです。

「単純に外部に委託すれば、データ処理のノウハウが特定の委託先に集約してしまい、他に委託することが困難になってしまいます。それを避けるため、私たちの側で詳細な技術仕様を策定し、細部まで理解できるシステムが必須だったのです」(黒沢氏)

黒沢氏らは情報連携ツールの比較検討を重ねた末、DataSpiderの採用を決定しました。

「最終的には2つの製品を比較し、それぞれの営業担当から説明を受けました。アプレッソは回答が素早く、説明も分かりやすく、私たちが欲している製品だということが理解できました。DataSpiderはプログラムをスクラッチで開発する必要がなく、短期間でデータの変換も可能でした。データ構造の確認やテストも迅速で、イメージ図もどことなくJ-GLOBALに似たところがあって親近感を覚えました」

試行錯誤を繰り返した後にエキスパートの支援で公開へ



株式会社メディアフォース
事業推進室 室長
松本一 氏

こうして、J-GLOBALの開発がスタートしました。プロトタイプを作りながら、DataSpiderの活用法を探っていったそうです。

若干の苦労があったものの、黒沢氏らはプロトタイプで動作を確認しつつ、J-GLOBALの基本的な処理手順の開発を進めていきました。そして、基本的な構成ができあがってきたところで、新たな段階に入ります。今後のスクリプト開発を容易にし、技術を継承するためにスクリプトの標準化や性能向上を進めることにしたのです。

「将来に向けたマニュアル化に備えて、標準化の必要を感じていました。また、私たちが手探りで作ったスクリプトが本当に適切か不安もありました。そこで、アプレッソに相談し、DataSpiderのエキスパートを紹介してもらい、支援を受けることにしました」(黒沢氏)

支援に加わったのは、アプレッソの開発パートナーである株式会社メディアフォースの事業推進室長、松本一 氏でした。

「まずは、ディレクトリ構造や命名規則、処理順序などを標準化しました。それから、公開へ向けて処理性能を向上するためスクリプト改良に携わり、中には、2週間近くかかっていた処理を3～4時間と飛躍的に短縮する改善も実現しました。さらに現在、運用支援も行っています。システムインテグレータとしてのエッセンスを提供した、というところですね」(松本氏)

今後さらに改良を加えより充実したサービスを目指す



独立行政法人科学技術振興機構
イノベーション推進本部
研究基盤情報部 情報整備課
関本和美 氏

J-GLOBALは2009年3月、2度の調査とプロトタイプ開発を経て、β版としてリリースされました。

J-GLOBALは各種データベースから更新された情報を受け取り、バッチ処理で反映、毎週月曜に更新しています。更新処理を担当するイノベーション推進本部 研究基盤情報部 情報整備課の関本和美氏は次のように説明します。

「バッチ処理そのものは順調ですね。出勤してきて処理を開始したら他の作業をして、終わったところで確認するようにしています」(関本氏)

また、J-GLOBALで扱う情報の拡充も進められています。2009年10月からは、第二期として半年間の追加開発に取り組んでおり、同12月中旬には、当初の公開時点よりデータ量が2倍ほどになる予定です。

そして、使い勝手の向上も課題の一つです。これはJ-GLOBALに限ったことではなく、例えばReaDでは本人が自発的に更新する仕組みになっていることから、登録・更新作業をより容易に行えるようにして、データの鮮度や精度を高めることが求められています。

「例えば業績の登録をする際、その研究者の情報を文献や特許の情報ソースから抽出し、候補として示すような仕組みを考えていますが、DataSpiderを活用して進めていきたいですね」と、イノベーション推進本部 研究基盤情報部 情報整備課 係長の関根基樹氏は言います。J-GLOBALは単なる情報集約サイトでなく、逆にデータ提供元となっているデータベースへのフィードバックも考慮されているのです。

「情報を複合することで、新しいものが見えてきます。例えば、注目する技術や分野が今後伸びるかどうか、文献や特許をはじめとするさまざまなリソースを複合した時系列推移から、類推することが可能だと思っています。今後もDataSpiderを活用して、統合された科学技術情報の発信に努めていきたいですね」(黒沢氏)

APPRESSO、APPRESSO ロゴ、DataSpider、DataSpider マーク、Servista、Servista ロゴ、アプレッソ、サービスタ、データスパイダーは、株式会社アプレッソの商標または登録商標です。その他の会社名、製品名、サービス名等は、各社の商標または登録商標です。2009年10月の内容に基づいて作成しています。



開発・販売元
株式会社アプレッソ
〒112-0014
東京都文京区関口1-20-10
住友不動産江戸川橋駅前ビル2F
TEL : 03-4321-1111 (代表)
FAX : 03-4321-1112
<http://www.appresso.com/>
E-mail : info@appresso.com