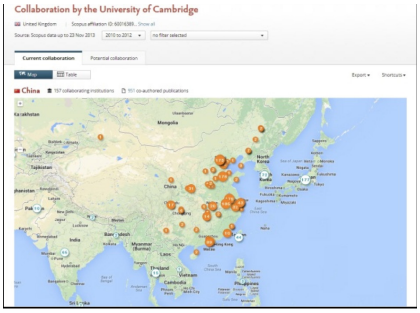
Case Study:

Elsevier's SciVal Story



Research continues to grow, with the number of researchers, currently approximately 7 million, increasing at 1% a year, and scholarly output growing at around 3% annually. Added to this the traditional research superpowers, such as the United Kingdom, United States, Japan, and Germany are being joined by the new breed of rapidly growing research nations China, India, and Brazil.

Overview

Scientific policymakers, research funders and research institutions, increasingly need objective information to use alongside expert opinion and peer review when setting policies and goals, charting progress, and making budgetary decisions about where research dollars should be focused. Research institutions have found it extremely time-consuming and difficult to locate suitable data and to generate the information in the format that they want.

Elsevier has been publishing scholarly research for over 130 years, and is the single biggest science, technology and medicine publisher in the world. This unique vantage point means that it is in a perfect position to address these needs, and over the last decade the company has transformed itself into a digital research information provider, making significant investments in data indexing, structuring and storage, to allow users to find the answers they need quickly.

Elsevier's SciVal enables users at research organizations across the world to benchmark flexibly to understand the position of their research relative to their peers, as well as relative to global and domestic standards, by allowing them to configure, visualize and export information according to their personal needs and preferences within just a few seconds. SciVal builds on Scopus, the largest abstracting and indexing data set of peer-reviewed literature, to provide insights that can be used to help answer the enormously diverse questions that are increasingly faced in the management of research, such as identifying research strengths, uncovering current collaborations and finding potential new partners, and scenario modeling. The results are also a powerful communication tool to showcase the distinctive strengths of research institutions to students, academics and funders.

SciVal brings together the 4 key elements that underpin any evidence-based research tool: data, technology, metrics and visualizations.

- **Data:** transparent, comprehensive, and up-to-date.
  - Scopus data, comprehensively covering 21,000 titles from 5,000 publishers is available in SciVal, and updated weekly.
  - SciVal is a ready-to-use solution with 4,600 research institutions and 220 countries pre-defined. Universities do not need to invest any resources in setting up SciVal and can benefit immediately.

- **Technology:** pre-defined standard report views, with flexible, real-time benchmarking.
  - Reed Elsevier's High Performance Computing Cluster (HPCC) Systems, underpins SciVal and is one of the most advanced, fast-performing Big Data technologies available today. It is used to solve large-scale, complex data and analytics challenges.
  - Users can query around 75 trillion metric values, refreshed every week, such as those for entire countries and institutions.
  - HPCC Systems easily accommodates unlimited numbers of datasets created by users. For example, if a single user creates only 20 datasets then HPCC Systems may generate an additional 100 million values per user per day.
- **Metrics:** broad range, clearly defined, and can be selected by the user.
  - 15 metrics cover productivity, disciplinarity, citation impact, and collaboration, both international and with industry.
- **Visualizations:** multiple, adaptable, easy to understand, and exportable
  - Every user can configure the visualizations to their personal preferences so that they are relevant to their challenges, such as:
  - "I want to benchmark my institute against my peers, and my departments want to benchmark themselves as well. How can we all do this effectively to suit our different realities?"
  - "My Vice Chancellor is going to China, who do our academics collaborate with there?"
  - "How can we demonstrate excellence in a way that best shows our unique strengths to secure funding and attract students?"
  - "I want to explore the various scenarios I'm considering to set up a center of excellence. How can the data provide me with insights?"

Elsevier will continue to invest in SciVal to extend its capabilities and address broader parts of the research landscape, such as funding, commercialization, and societal impact, to cater to the dynamic customer needs within the changing research environment.

The Challenge

Elsevier decided to enhance the analytical capabilities of their first generation SciVal solutions to better serve research institutions in generating hard facts to support their decision making. They not only integrated the existing functionality, but increased the level of sophistication by adding new options of slicing the data, new metrics, and new visuals.

The research community was critical in ensuring that this functionality effectively addressed their needs, and that the information was presented in a simple and intuitive interface. "The new generation SciVal has been

"Outsell's view of the new SciVal demonstrates its extraordinary utility: The system calculates complex benchmarking of more than 4,600 research institutions' outputs on highly specific scientific topics in seconds. The product is packaged in an extremely simple user interface. It elegantly allows users to create and share insightful charts, graphs, and diagrams in a way that prior versions of SciVal did not"

—from "Why 2014 Will Shape Up as a Battleground Year for Research Analytics Titans", by Rich Kreisman, Outsell Insights, 7 January 2014.

LexisNexis®

developed following years of close co-operation with leading research institutions globally," said Dr. Nick Fowler, Managing Director of Academic and Government Institutions at Elsevier. "We hope that research leaders will view the result as a significant step forward in helping them to monitor and manage their institution's research strategies."

In addition, it was essential that Elsevier could address the need of their users that the new tool should provide information quickly. The problem was made even tougher because users emphasized that they need to be able to create data slices and visualizations in real-time.

The Solution

The solution to providing information quickly and flexibly is a 40 node HPCC Systems cluster. This enables SciVal to process around 40 terabytes of linked data on a weekly basis.

A specific example is the creation of complex visualizations showing areas of an institution's global strengths. Prior to starting to use HPCC Systems, it took 1 to 2 days to generate this visualization for one institution, and now it takes about half an hour.

The Results

The use of HPCC Systems by SciVal has brought two enormous benefits to its users.

Firstly, users can customize their own visualizations, which are generated in just a few seconds from almost 40 terabytes of data currently underpinning the tool.

And secondly, HPCC Systems works in 2 modes: offline crunching of huge, pre-defined requests, and smaller calculations in real-time on customer-generated data slices. This enables users to benefit from both structured information available immediately upon implementation, as well as having the flexibility to tailor-make their own views of the world.

About LexisNexis and HPCC Systems

LexisNexis® currently has four petabytes of information from 10,000 different sources. This information is structured, semi-structured and unstructured data, and is leveraged in various information and analytics products and services to help customers verify identity, detect fraud, uncover hidden relationships and locate people, businesses and assets. Customers include financial institutions, insurance carriers, health care organizations, law enforcement, legal firms, and academia.

The HPCC Systems platform was built specifically to analyze large volumes of data in minutes to solve complex problems. The platform can deliver

"The new generation SciVal has been developed following years of close co-operation with leading research institutions globally,"

– Dr. Nick Fowler, Managing Director of Academic and Government Institutions at Elsevier.

LexisNexis®

the data to more people and provide fast reporting. Since only small development teams are needed—this reduces investment in large teams and keeps processes agile.

The HPCC Systems platform has three distinguishing factors that make it an effective choice for big data analytics and processing:

- The HPCC Systems Data Refinery engine (Thor) helps clean, link, transform and analyze Big Data. Thor supports ETL (Extraction, Transformation and Loading) functions like ingesting unstructured/structured data out, data profiling, data hygiene, and data linking out of the box. In addition, Thor supports flexible record oriented data structures.
- The HPCC Systems Data Delivery engine (Roxie) provides highly concurrent and low latency real time query capability. The Thor processed data can be accessed by large number of users concurrently in real time fashion using the Roxie. The Roxie queries are typically complex and could include embedded rules logic.
- The programming language called Enterprise Control Language (ECL) is used to program both the data processing jobs on Thor and the queries on Roxie. ECL is a declarative, implicitly parallel and data flow oriented programming language that abstracts complex data processing tasks by providing a simple programming interface. HPCC Systems helps organizations in mission-critical 24/7 environments gain competitive advantages by leveraging all the data to help scale for innovation and growth. The streamlined platform needs less resources to operate and eliminates expensive legacy technology to lower the Total Cost of Ownership (TCO) of Big Data solutions.