

CASE STUDY

Key federal agency cuts costs, boosts effectiveness with open source Business Intelligence, Data Management and Data Integration

The **National Institute of Food and Agriculture (NIFA)**, an agency within the **U.S. Department of Agriculture**, advances knowledge for agriculture, the environment, human health and well-being, and communities—all by supporting research, education, and extension programs in the U.S. Federal Land-Grant University System and other partner organizations. To this end, NIFA helps identify and fund important research, education, and extension programs at the state and local level, and provides program leadership in these areas.

Business/Technical Challenge

To facilitate its dual roles of oversight and leadership in government-sponsored research related to agriculture, health and environment, NIFA pursues (among others) three key strategies. First, NIFA requires—and collects—regular reports of progress and findings from the research efforts it helps to fund. These reports are prepared by research organizations at Land Grant universities across the U.S.

Second, it's NIFA's job to cull important information and insight from these reports and pass them along to the funding entity—Congress. NIFA and its partner research organizations use a pair of custom-developed applications—Statement of Work (SOW) and Current Research Information System (CRIS)—to capture and store these highly textual reports. Each of these systems focuses on specific types of research that are sufficiently different to warrant the use of both.

Finally, NIFA also acts as a key, on-demand information source to Congress and various government agencies, who request information about (or conclusions from) all of these disciplines and research efforts. To this end, NIFA staff must digest, analyze and synthesize data from among the thousands of reports it collects. NIFA also works to make these reports available to the public through various means. Finally, in its leadership role, NIFA has an internal research staff charged with their own investigative and analytic endeavors.

NIFA faces a number challenges in conducting this work, but the greatest of these is efficiently and accurately extracting “impact and outcome” meaning from a huge—and growing—number of unstructured, primarily textual documents, then making it available both for analysis and public access.



INDUSTRY Government

ORGANIZATION

National Institute of Food and Agriculture, agency of the USDA

SOLUTION

Provide information on government-sponsored research to the public

RESULTS

Using open source solutions from Jaspersoft, Talend and Infobright, NIFA implemented a new reporting portal at much less cost and with greater flexibility.

“To achieve the kind of structure, reporting and query support we required, we would need to upgrade to a new Oracle solution—priced at over \$200,000.

So we started looking for new solutions—especially those in the open source arena.”

Joe Barbano
NIFA

CASE STUDY

“The researching entities provide the information to us through SOW and CRIS,” explains Joe Barbano, at NIFA. “But that’s the easy part. Pulling out meaning, identifying trends, getting a complete view of impacts and outcomes—that’s the challenge. We have thousands of these reports. How do we make it meaningful for the public, and for our in-house scientists who have their own information needs?”

There’s also a requirement to provide the reports, formatted for readability, back to the reporting entities. Many of these reports, accessed regularly by hundreds of end users at partner universities, may exceed 100 pages in length.

To address these needs, NIFA designed a data warehouse—implemented on the Oracle database platform—to classify the incoming data along a number of key dimensions (the identification of which has proven a challenge in itself). Research entities provide keywords and tags as part of their reporting, which aids in this classification. NIFA has also experimented with a number of search technologies in an attempt to parse and classify incoming documents.

“We still haven’t been able to meet our precise requirements for search,” reports Barbano. “But we continue to try.”

In a sense, NIFA then faces a huge, metadata-driven desktop publishing effort. To provide reporting access to the normalized data in the warehouse, NIFA at first used the Crystal Reports solution—but there were challenges. Specialized staff was required to define reports.

Additionally, costs became a larger and larger issue—notably those associated with the Oracle platform. *“To achieve the kind of structure, reporting and query support we required, we would need to upgrade to a new Oracle solution—priced at over \$200,000,” says Barbano. “So we started looking for new solutions—especially those in the open source arena.”*

Using an Open Source Business Intelligence Stack

NIFA began looking at other technologies in an effort to cut costs and more effectively meet its unique information management needs. Finding the perfect-fit search solution has remained a challenge, but the team has had great success in building the business intelligence platform. NIFA chose Jaspersoft for reporting and dashboards, Infobright for the analytic database and Talend for data integration.

Reporting and Dashboards. NIFA worked with analysts from Jaspersoft to prototype some of the reports it desired, based on its CRIS database, using Jaspersoft’s iReport graphical report designer. The team found that it worked best to create a Java program for transferring the CRIS data into an XML format that included key metadata such as information tags. iReport could then be used to define the final output.

Given the success of the prototyping effort, the significantly lower costs involved, and the security of choosing a popular, widely used solution, NIFA decided to move forward with Jaspersoft.

CASE STUDY

Jaspersoft reporting capabilities are used to format a large variety of HTML-based reports, many 60 to 70 pages in length; these are hosted online for access by internal and partner research teams. Some of these reports are accessed by drilling through from dashboards, which also present helpful graphic representations of the information, created with Jaspersoft's JasperServer report server.

Data Integration. Textual data—together with metadata tags entered by the researching entities—are imported from the CRIS and SOW systems into a MySQL database, using ETL technologies from Talend. Where sufficient tag information is not available, search technologies and automated thesauri are used to derive the metadata. For example, input documents are examined for the presence of keywords and indexed according to matches found. This approach helps classify documents along key dimensions for storage in the MySQL database.

"We love using Talend's ETL solution because it lets us effectively share metadata among the researchers and developers—and because it's affordable."

Database. To reduce costs, speed development of new capabilities and add flexibility, NIFA also decided to migrate its Leadership Management Dashboard, developed years earlier on the Oracle Portal platform, to an open source stack including Linux, JBoss, and Infobright, the MySQL-based analytic database. Previously the dashboard infrastructure was based on a tightly integrated stack including the Oracle database, Oracle Portal, and PL/SQL stored in the database. NIFA wanted to be able to separate the logical tier from the database itself, and to leverage tools with greater flexibility and lower cost. Concluding that they had achieved as much as was possible using Oracle Portal, the team wanted to move to a more flexible environment—and doing so with Oracle would have required the licensing of Oracle Fusion at a cost of nearly half a million dollars.

NIFA decided to use the Ruby on Rails development environment with the new platform: *"With Ruby on Rails, we can deliver some very nice ad hoc queries that access the data through the Infobright technology. This gives us fast, highly efficient performance thanks to Infobright's columnar architecture,"* says Barbano.

Benefits of the New Solution

The new open source approach gives NIFA the flexibility it sought, and more readily provides internal users and the public with the information they really need.

"There's no longer a need to examine potentially hundreds of documents to locate information needed by a researcher or a Congressional office," says Barbano. *"Once workable queries are defined by staff,"* he adds, *"they can easily be shared with other researchers."*

Given the challenges of automating thesaurus-aided search of textual data, NIFA's quest for an ideal search solution is ongoing. *"But now, we have a much more flexible stack,"* Barbano points out. *"That means it's easier for us to try various tools and algorithms, and swap them out when we experience improved results."*

CASE STUDY

Now, with a fully open source stack, including business intelligence from Jaspersoft, ETL from Talend, high performance analytic queries enabled by Infobright, and efficient, low-cost MySQL data management, NIFA is better positioned to fulfill its mission—while keeping costs low.

About Jaspersoft

Jaspersoft provides the most flexible, cost effective and widely deployed Business Intelligence suite in the world, enabling better decision making through highly interactive, web-based reports, dashboards and analysis. Leveraging a commercial open source business model, Jaspersoft provides end-to-end BI capabilities at a fraction of the cost of other vendors. The BI suite includes pixel-perfect enterprise reporting, ad hoc query, dashboards, OLAP and in-memory analysis, and data integration. Jaspersoft is the only BI vendor that enables companies to adapt to the new, virtualized world by providing a complete spectrum of on-premise, multi-tenant SaaS and cloud-based deployment options for both embedded and standalone business intelligence.

For more information visit: <http://www.jaspersoft.com> and <http://www.jasperforge.org>.

About Infobright

Infobright's high-performance database is the preferred choice for applications and data marts that analyze large volumes of "machine-generated data" such as web data, network logs, telecom records, stock tick data and sensor data. Easy to implement and with unmatched data compression, operational simplicity and low cost, Infobright is being used by enterprises, SaaS and software companies in online businesses, telecommunications, financial services and other industries to provide rapid access to critical business data.

For more information visit www.infobright.com or join our open source community at www.infobright.org.

About Talend

Talend is the recognized market leader in open source data management & application integration. Talend revolutionized the world of data integration when it released the first version of Talend Open Studio in 2006. Talend's data management solution portfolio now includes operational data integration, ETL, data quality, and master data management. Through the acquisition of Sopera in 2010, Talend became also a key player in application integration.

Unlike proprietary, closed solutions, which can only be afforded by the largest and wealthiest organizations, Talend makes middleware solutions available to organizations of all sizes, for all integration needs.

Visit us at www.talend.com.

CONTACT Jaspersoft

Jaspersoft Headquarters
539 Bryant Street, Suite 100
San Francisco, CA 94107
1-888-399-2199
Phone: 415.348.2380
Fax: 415.281.1987
Email: sales@jaspersoft.com

CONTACT Infobright

Infobright Headquarters
47 Colborne Street, Suite 403
Toronto, Ontario M5E1P8 Canada
Tel. 416 596 2483
Toll Free 877 596 2483
Fax 416 352 5674
Email: info@infobright.com

CONTACT Talend

Talend Headquarters
9 rue Pagès
92150 Suresnes France
Tel: +33 1 46 25 06 00

Talend US:

5150 El Camino Real, Suite C-31
Los Altos, CA 94022
Tel: +1 (650) 396 7738