Buyer Case Study

Agile Information Discovery at AstraZeneca

Carrie Solinger David Schubmehl

IDC OPINION

Agile information discovery is a challenge for many organizations with their current internal and external search and discovery methods. As cognitive technologies like natural language processing (NLP) and machine learning (ML) become more popular and integrated with many emerging technologies, so, too, does traditional enterprise search and discovery systems that utilize these capabilities. A large majority of information discovery systems face many criticisms and challenges. The systems are not agile, cost-effective, timely, or accurate enough to face the demands needed to handle the deluge of internal and external data sources in structured and unstructured formats. Thus many organizations are turning to vendors, such as Sinequa, that can offer these new cognitive technologies as part of their information discovery offerings. In addition:

- AstraZeneca has redefined the use of information discovery in terms of context, purpose, and solutions within its own company.
- AstraZeneca is using the concept of agile information discovery to develop task and role-based specific applications that solve very specific business problems.
- The company's use of Sinequa's platform provides an example of how organizations can take advantage of agile information discovery using cognitive systems technologies to solve real and specific business problems.

IN THIS BUYER CASE STUDY

This IDC Buyer Case Study examines how AstraZeneca has created an internal application store service for more agile information discovery across departments, not just where these solutions originated in research and development (R&D). As part of this Buyer Case Study, we look at how AstraZeneca is leveraging Sinequa's cognitively enabled search and analytics platform to create and use agile information discovery to improve and streamline R&D and business processes throughout its organization.

SITUATION OVERVIEW

Organization Overview

AstraZeneca is a biopharmaceutical company with R&D at its core. The company's business goals are to provide innovative, effective medicines that make a real difference to patients. AstraZeneca focuses on six specific areas within healthcare: cardiovascular, gastrointestinal, neuroscience, respiratory and inflammation, infection, and cancer. The organization invests over $4 billion annually in R&D, with over 15,000 professionals in 8 countries on 3 continents accessing and analyzing information as key

October 2016, IDC #US40797216

components of the company's strategy to improve the quality of its science, R&D productivity, and time to market.

Challenges and Solution

As a large biopharmaceutical company, AstraZeneca faced the challenge of making information discovery an agile process for scientists in the R&D department to find the right information in a timely manner. Accuracy, efficiency, and real-time results and insights were priorities. AstraZeneca understood the need for comprehensive information collection, aggregation, and discovery to accelerate research and innovation. To facilitate this information collection and discovery in a timely manner, AstraZeneca decided that it needed a new approach to how it researched and handled information. AstraZeneca's decision concurs with recent survey research by IDC that indicated there was a need to improve the information access capabilities of research staff. AstraZeneca focused on improving the internal R&D information access capabilities of research staff. In IDC's *Knowledge Worker Survey,* we found that 57% of all researchers stated that they needed access to four or more systems to perform their work and only 45% of all researchers rated their current process of finding information as being somewhat or very beneficial.

AstraZeneca found that other departments within the organization were similarly looking for targeted solutions to solve business problems specific to departmental needs. AstraZeneca's Data Analytics Lead, Rob Hernandez, understood that "from the beginning, we wanted to be more than [just traditional IT staff] and not just deliver traditional search but actually be able to build applications that were making use of the underlying information …. We wanted to bring under the same kind of roof every single document, and every single piece of information, from structured data to unstructured text that a scientist needs."

AstraZeneca's ambitious goal was to dramatically impact the time it took scientists to research, aggregate, and understand relevant information related to the company's various biopharmaceutical projects. AstraZeneca realized the need to provide a platform that allowed rapid and easy creation of business applications, in research as well as other departments throughout the organization. The overall requirement for this group was to work with all of the stakeholders to develop and implement the next generation of information discovery processes and systems at AstraZeneca and to do so in a timely manner while delivering ROI and value within the budgeted parameters provided to the group by senior management. These systems needed to handle a wide range of information and document types from a large number of sources such as SharePoint, Office 365, department file shares, EMC Documentum, eRoom, and AstraZeneca's own R&D wiki as well as several sources of large-value external data. "Once you have that … within a single system, you can then think of ways you might want to search and kind of re-delve into that information or query across that information or run analytics on that information to answer specific business questions," Hernandez explained. To provide a more agile and robust application, AstraZeneca also developed applications that made use of that underlying data to give new insights.

Sinequa is a cognitive search and analytics platform that provides relevant insights about information to users in their work environments. This platform discovers new information and hidden relationships and insights and analyzes user behavior and preferences to learn about their work context. Sinequa's platform extracts content from structured and unstructured data via natural language processing statistical and semantic analysis and a set of scalable machine learning libraries to continuously improve relevant information discovery and the relevance of information delivered to users. Structured data types include information from databases, and unstructured data types include information from

text, images, and video. Sinequa's search and analytics platform provides a flexible solution for data scientists, enterprise software, and end users. Sinequa integrates cognitive capabilities and machine learning to provide real-time, relevant results from unstructured and structured internal and external data. Sinequa's platform supports a wide range of machine learning algorithms and capabilities to improve findability and relevance such as:

- Filtering and providing recommendations collaboratively (ecommerce)
- Clustering topically related content
- Classifying by example but without classification rules
- Finding documents with similar content
- Computing missing values in a data frame, or predicting future values
- Identifying outliers

Results

In less than 12 months, AstraZeneca implemented Sinequa's real-time search engine for AstraZeneca's global R&D. This focused on all scientific information and core internal repositories and supported the creation of multiple search-based applications. Once that had been completed, AstraZeneca moved on to a new initiative to provide agilely developed role-based search applications for workers in R&D and throughout the company. To make these as widely available as possible, AstraZeneca decided to implement the concept of an app store for these tools where workers could easily download and start using the applications in their day-to-day work. AstraZeneca saw the need for similar agile enterprise environment solutions across other departments where accuracy, efficiency, and real-time results and insights were priorities too. AstraZeneca built over 10 applications in the first 3 months. The company now supports applications across all business functions such as:

- Find Scientific Partners
- Mobile productivity apps for employees such as Find People and Provide Approvals
- R&D Intelligence
- Find People at AstraZeneca
- R&D ChemSearch
- R&D News Alerts
- In-Video Search

AstraZeneca's app store has led to a better functional information discovery system and cross-departmental acceptance and use of better information discovery methods, which has improved its science, R&D productivity, and time to market. AstraZeneca's new global portal launched using Sinequa for enterprise search. Over the past several months, the app store has emerged from R&D to go global within the company and is available to over 60,000 employees. Targeted role-based search now enables AstraZeneca to socialize key findings from news and documents but also chatter, applications, people, and scientific tags, helping to connect people together. Already in preproduction, indexation of all of AstraZeneca's cloud repositories (Box, SharePoint, Veeva, etc.) makes even more data available. Today, AstraZeneca has over 180 million documents that are searchable in real time with key scientific vocabularies (SciBite) automatically tagged and findable.

AstraZeneca gained competitive advantage through its ability to innovatively use technology while also increasing the existing value of IT investment. The business-centric app store drives a disruptive culture across AstraZeneca departments and employees by accessing the latest technologies, creates

an environment to cultivate innovative ideas, and nurtures ideas that deliver immediate value to the business, which leads to step-change adoption. As a result, complex business processes can be delivered through intuitive search-based applications.

ESSENTIAL GUIDANCE

Information discovery is a challenge for many organizations. Their current internal and external search and discovery methods lack the necessary components to handle the influx of different data types. These systems lack agility, timeliness, accuracy, and cost-effectiveness. Agile information discovery systems address this problem through its all-encompassing methodology to treat the problem and provide a solution using a more holistic approach. As such, these newer systems meet the demands of the ever-increasing amount of internal and external data sources in structured and unstructured data types.

There are a number of lessons that other organizations can learn from AstraZeneca's experience:

- Role-based and tasked-based information discovery solutions can be very successful:
  - The goal is not to develop a perfect solution – rather, a solution that will successfully address 80% of employee needs.
  - This process allows the organization to move on to the next business problem.
- The Sinequa platform provides an agile platform for developers to build apps very quickly:
  - The search and analytics platform leverages a combination of natural language semantic analysis and scalable machine learning libraries.
  - Sinequa's cognitive-based services and solutions allow organizations to extract usable information from large amounts of structured and unstructured data types across a number of different data access points.
  - These robust applications assist organizations in uncovering new information, hidden relationships, and insights in real time.
- Enterprise organizations should look at adopting agile information discovery methods to help solve traditionally difficult business challenges:
  - Discovery and analysis of both structured and unstructured data types are more timely, accurate, and cost-effective, which in turn leads to a greater ROI.
  - This competitive advantage is derived from a more robust and holistic approach to information discovery.
  - A single unified search system allows a more comprehensive solution to enable organizations to solve difficult and specific business challenges.

LEARN MORE

Related Research

- *Market Analysis Perspective: Worldwide Cognitive Systems and Content Analytics Software, 2016* (IDC #US40797116, September 2016)
- *IDC's Worldwide Software Taxonomy, 2016* (IDC #US41572216, July 2016)
- *Worldwide Cognitive Systems, Content Analytics, and Discovery Software Forecast, 2016-2020* (IDC #US40305316, June 2016)

- *Requirements for Unified Information Access Systems* (IDC #244622, November 2013)
- *Unified Information Access 2013: Unifying Content Silos to Uncover Hidden Value* (IDC #240053, March 2013)

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com