

PRODUCT FLASH

Unified Access Case Study: MetaCarta Unlocks Geographic Content in a Global Energy Firm

Geoffrey Bock (contract writer)

Susan Feldman

IN THIS PRODUCT FLASH

This IDC Flash describes how one global energy company exploits the unique geospatial entity extraction capabilities that MetaCarta delivers to transform its corporate knowledge base into a strategic information resource for energy exploration and decision making. Geographic information is a vital component of energy exploration. Drilling in the wrong location can cost an energy company hundreds of millions of dollars — while drilling in the right place can lead to a multibillion-dollar gusher. It's critical to look at all the information that is available to make informed decisions — and that information is usually stored in many places, in many formats, and in many applications. Yet geography is often an important facet for unifying access to both unstructured and structured content. End users within this global energy company needed an effective way to retrieve information within their corporate knowledge base, and to rapidly find content that was locked in more than fifty information silos. With its geospatial information retrieval solution, MetaCarta provided unified access to these silos for the petroleum engineers, geologists, and other specialists who needed to conduct research, analyze problems, and make decisions.

SITUATION OVERVIEW

Oil and gas exploration is an information-intensive business. One global energy company that IDC interviewed maintains an extensive library of scientific and technical documents — containing millions of items and stored in both unstructured repositories and databases. They chronicle its exploration activities around the world. This corporate knowledge base covers years of the company's own exploration and development efforts, as well as those of firms that it has acquired.

End users within the company — petroleum engineers, geologists, and other specialists — rely on this knowledge base when deciding on where to prospect for new drilling sites. They need to assemble structured and unstructured information about predefined locations — including geophysical analyses, seismic mapping reports, leasing conditions and terms, and field-level production data. But accessing all of this content in a consistent manner, and finding complete and accurate information, has been difficult because of the differences in how the information is organized and the sites are described.

A variety of strategies for organizing, managing, and retrieving information to support exploration and development have yielded mixed results. For many years, the company had a formal records information system, modeled as a library catalog. End users worked with professional search specialists to query the online repositories and retrieve citations. When the company started to use the Web in 1995, it began to provide browser-based access to the electronic records. End users were pleased with self-service searching but were not confident about the precision and recall of their results.

In the new millennium, as the Web became increasingly popular and powerful, the energy company added browser-based access to many of its structured databases and full text retrieval of its unstructured document collections. While end users could query separate repositories, they had difficulty making sense of all the information they found. Each repository was a separate information silo, with its distinct user interface, search experience, and query language. There was no organizing metaphor, standardized taxonomy, or unifying capability for retrieving content among the disparate sources.

Since the energy exploration business depends on geography, end users spend a lot of time and effort searching documents and databases for references about geospatial locations, and then correlating the results among multiple repositories. Faced with the high cost of not finding relevant information in a timely fashion, the energy company sought to substantially improve the effectiveness of its corporate knowledge base.

In 2004, the energy company began working with MetaCarta, a geographic content search and discovery firm, to extend and unify the retrieval capabilities to its corporate knowledge base. MetaCarta adds geospatial expertise to content searching. It provides a geospatial, natural language processing (NLP) engine for an enterprise infrastructure. This engine analyzes, extracts, and geocodes both structured and unstructured content for names, places, and other kinds of geographical references. It can automatically differentiate between "Paris, France," "Paris, Texas," "Paris Hilton" (the person), and "Paris Hilton" (the hotel) as the terms appear in different contexts and disparate content collections. MetaCarta maintains an extensive and extensible gazetteer. It identifies the latitude and longitude of over 8 million places and also allows companies to add their own locations.

Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights. Visit us on the Web at www.idc.com. To view a list of IDC offices worldwide, visit www.idc.com/offices.

Copyright 2006 IDC. Reproduction is forbidden unless authorized. All rights reserved.

Filing Information: November 2006, IDC #204375, Volume: 1, Tab: Users

By enhancing its corporate knowledge base with MetaCarta's geographic searching capabilities, the energy company uses location as a powerful metaphor for aggregating content across diverse repositories. A single geospatial query retrieves information from multiple sources. Not surprisingly, the company has extended the basic MetaCarta gazetteer to include the company-specific geographic information — adding the names and locations of energy fields, well heads, pipelines, and pumping stations. An oil well might be called anything from "Mad Hatter" to "K2" to "ABS33345," but there is only one precise geospatial location for each well, calculated by its latitude and longitude.

Map-based filtering is a powerful way to filter the search results and return information in the specific geographic area of interest. But up-front investment always improves the ability of any language-based system to retrieve information. At the energy company, three subject matter experts have refined the retrieval terminology, devoting roughly six months to tune the geospatial search capabilities and ensure consistency of results across the fifty-plus repositories.

FUTURE OUTLOOK

This energy company considers its unified content access environment to be a strategic advantage. Its success, however, is the result of combining good technology with systematic information architecture. To ensure a consistent geospatial search experience, MetaCarta requires some effort at information integration: connecting MetaCarta to its various repositories, and then ensuring that MetaCarta accurately spiders each silo and returns accurate results that utilize the access security model of each individual repository. The energy company tailors the adapters and connectors to each of its repositories, and then tests and tunes the results. Along the way, it identifies and catalogues all of the company-specific places within the gazetteer.

For the energy company, the investment in MetaCarta yields substantial results. End users can rapidly find documents and data related to one or more oil fields. It's now also easy to spot related exploration information about adjacent energy fields. The search for documents and data, which used to take over a week of intense searching, now can be accomplished within a few seconds. The result is that the corporate knowledge base has become a strategic resource, easy to use to answer quick questions or for longer research projects. End users rely on it to return relevant and comprehensive results so that they can make timely decisions.

All in all, geospatial information retrieval is a technology whose time has come. Information mapping no longer requires explicit, a priori tagging and the predefined identification of places and locations. The energy company's experience demonstrates the practical capabilities of MetaCarta's entity extraction solution — to infer, identify, and tag locations by analyzing both unstructured and structured content stored in multiple repositories.

With a modest amount of effort, other enterprises and publishers can follow the energy company's lead and easily adopt MetaCarta to unify access among disparate content repositories. In short, IDC believes that geospatial information retrieval is now an effective tool for unifying information access.

Published Under Services: Search and Retrieval Technologies; Unified Access to Content and Data; Content Management and Retrieval Solutions