



# Drilling for Data

ABBYY, Agile Upstream and Alta Mesa automate lease management in the Oil & Gas Upstream market.

Sue Feldman, CEO, Synthexis

Synthexis

A SYNTHESIS WHITE PAPER

# Drilling for Data

Sue Feldman, CEO, Synthexis

ABBYY, Agile Upstream and Alta Mesa automate lease management in the Oil & Gas Upstream market.

## Executive Summary

The oil and gas industry runs on information. Quick access to the right information, and better, faster analysis of documents make all the difference in gaining a competitive edge. It makes sense, then, to streamline the data pipeline: from entering the data into a land system to analysis, alerting, and decision-making.

Land management is something of a labyrinth. Automating it requires extensive knowledge of the industry and in-depth understanding of acquisition and divestiture processes. There are significant hurdles to be leaped in the journey to automation. Buying and selling assets is a complex process, and contracts are rarely straightforward. Contract stipulations can be dense and unpredictable. Multiple documents, sometimes with conflicting rights, apply to a single property. They may go back decades or more. For that reason, oil and gas companies must make their way through a welter of what is usually aging paper. Little of it is digital. The analysts who manage this complexity daily must understand and retrieve not only the original contract but also the related documents and additions.

To create an effective well-designed application that fits the needs of a specific task like this one, technology vendors, industry experts and practitioners must work together. In this case study, ABBYY supplied the language-based technology foundation to power advanced information extraction. They constructed a specialized knowledge base for the oil and gas industry that their system used to accurately understand and retrieve information from leases. Agile Upstream was ABBYY's development partner, designing the user interaction, and supplying the process knowledge. They provided the big picture of the role that lease and document management plays in the oil and gas industry. Together, they developed a highly usable application. They also brought in Alta Mesa Holdings, their client: the right test case to help develop a practical working application.

## Contents

- 3 THE STORY
- 4 THE PLAYERS
- 7 THE CHALLENGE
- 8 THE TECHNOLOGY
- 8 MINING OIL AND GAS LEASES
- 12 THE DESIGN
- 15 CONCLUSION

**“ At the end of the day, the oil and gas business is an information business. You need to keep your house in order from a data standpoint to succeed. ”**

*--Alta Mesa*

## The Story

In 2015, ABBYY was seeking a significant project to demonstrate the value of their new semantic platform for information management. They realized that the oil and gas industry, overwhelmed by data, was ripe for this kind of highly integrated approach to managing information.

At the same time, Agile Upstream was looking for a technology design partner to help them co-develop a Land Insights system. Their customer, Alta Mesa, an upstream oil and gas company, wanted to streamline its land lease management process to tackle these problems.

# The Players



Alta Mesa Holdings, LP

**Alta Mesa Holdings, LP** is a mid-sized oil and gas company engaged in onshore oil and natural gas exploration and production. It has approximately 230 employees, and operations in Texas, Oklahoma and Louisiana. The Land Department, run by David Murrell, VP of Land and Business Development, acquires oil and gas leases to gain access to minerals. The process of deciding what to leases to buy and where to drill is highly sensitive and proprietary. Alta Mesa depends on 3D seismic and other geological and technical data, and tends to look for under-developed or overlooked areas to which they can apply new geological techniques in order to extract more oil and gas. In their quest to find these overlooked drilling locations, they are adept in analyzing various data sources and acquiring and analyzing oil and gas leases.

Lease analysis supports their acquisition strategy and varies according to level of activity, varying from a handful to upwards of several hundred leases in a given week, from which they expect to extract key facts like description, date of expiration, and provisions that trigger dollar compensation. They need to verify leases, to identify conflicts among related documents such as LPR's and plat's, and understand all the provisions. In the case of an acquisition, they can receive multiple physical boxes as well as virtual collections of leases and related documents. It takes a lot of personnel and time to verify and integrate a set of new files.

Alta Mesa recognized that it would be beneficial to improve its efficiencies and thereby get a faster handle on acquisition and divestiture documentation. In the event a timely conversion was not available, their analysts had to sort, classify and enter data into their land management system. The process takes time in addition to the day to day activities.

They wish to streamline the acquisition processing: to image and analyze their boxes of paper leases quickly, and to integrate that information into the system they already have in order to keep on top of expiring drilling rights and ownership specifications. Alta Mesa was interested in new approaches to integrating oil and gas lease information into its current system and data base.

## The Players



**Agile Upstream** is a technology integrator for the oil and gas industry. They design process-based software that integrates data and applications to streamline specific industry workflows such as capital budget management, accounts payable, well management, or joint interest billing. Agile creates portals and data rooms that allow buyers and sellers to search, collate and “slice and dice” documents quickly and accurately.

Land lease management processes depend today on stove-piped applications in which the same data is entered multiple times, sapping productivity and leading to errors and inconsistencies. Agile realized that too much information was getting lost in the cracks between the applications. This was apparent to them as they tackled the same problems in customer after customer. They realized that the need was industry-wide and that a common infrastructure and document standards would enable companies to exchange documents, perform due diligence, and speed up mergers and acquisitions.

Agile was looking for a new approach to knit the applications together and to minimize data entry. They needed a well-integrated technology platform that could ingest, manage and analyze paper and digital documents more intelligently. Their goal was to create a Land Insights system that bundled together features from content and records management, retention and governance, analytics, extraction monitoring and reporting. Because regulations vary by locality, the system had to be smart enough to apply the right regulations to each property. A Land Insights system combining cutting edge cognitive computing capabilities with domain specific knowledge in oil and gas could transform the way oil and gas leases are managed - significantly increasing operational efficiencies and driving massive value to acquisitions and divestitures.

## The Players

# ABBYY®

**ABBYY** is a technology and solutions company, founded in 1989. Their software tools run the gamut from OCR and recognition to document analysis, text analytics, classification and even inferencing. They needed a showcase for the new ABBYY InfoExtractor (AIE) SDK that can identify and extract entities, facts and relationships from complex documents. AIE is designed as a platform for information-intensive Smart Business Process Applications, to provide content-based intelligence to improve analysis and accelerate decision-making.

**ABBYY InfoExtractor incorporates Compreno, ABBYY's linguistic technology foundation for their text analytic capabilities at every software level. Compreno's advanced semantic analysis extracts concepts and terms, reconstructing facts, events and storylines, storing them in a graph database for pattern matching and reasoning. These insights enhance information discovery, search accuracy, and are the basis for visualizing collections of information effectively.**

The ideal of this highly integrated platform is to drive a process-driven document pipeline that can support information-dependent processes like the oil and gas industry's appetite for acquisitions and divestitures.

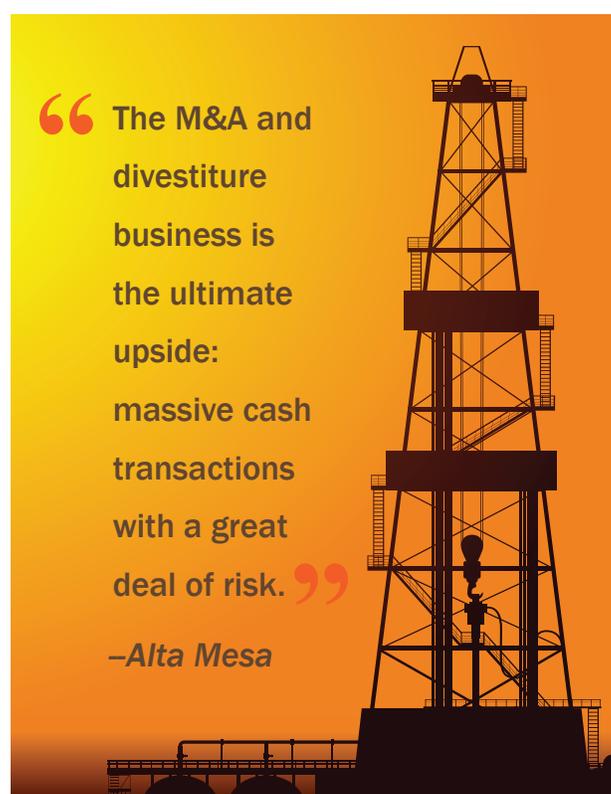
The implications of this approach are myriad. Certainly by deploying this kind of information platform, an oil company can streamline its acquisitions and divestiture process as well as increase accuracy and productivity. However, there is also potential upside for the entire industry. Its impact grows as other players standardize on this approach. Using a shared domain-specific terminology, or ontology, to extract facts, events and storylines from complex leases will provide much greater information transparency for all parties involved in M&A transactions, accelerating deal making while controlling risk and optimizing results across the board.

## The Challenge

To develop an application that will support property acquisition and divestiture in the oil and gas industry. The system must create a single accessible collection from both paper and digital documents, then uniformly extract facts such as lease terms, locations, and drilling rights no matter how they are phrased. It must find relationships among leases and related documents.

Furthermore, the system must:

- Support ad hoc querying and analysis without relying on schemas or specialized query languages
- Connect to standard land management and document management systems to update databases and keep all systems in sync
- Provide a process-based visual interface to support all land lease management processes
- Automate acquisition and divestiture processes, but with manual override as necessary
- Be easy to use for non-technical users
- Create an integrated pipeline from paper document to acquisition/divestiture analysis
- Adapt and evolve as terminology and the business change



Property acquisition and divestiture are two sides of the same coin. In both cases, all documents relating to a property must be located, and then scrutinized to determine the value of the property to the current business. The information must be accurate, displayed so that it is quickly understood, and show changes in ownership over time. The ability to trace the system’s interpretation of a term back to the original wording is a requirement. The system must be able to adapt to and integrate new information and terminology. It must evolve as situations change. The machine and the human must interact smoothly and complement each other, with the system managing predictable and repetitive tasks, and the human overseeing the process and making corrections that can improve the machine learning process.

# The Technology

Oil and gas companies depend on structured document management systems (and at times file shares) to track their leases and locate terms and conditions. The primary purpose of any document management system is to create a reliable, trusted archive. Entering data into the system is manual and data-intensive. Accuracy is critical because the data is the foundation for its business. Its value extends far beyond that of a static repository. Information-dependent companies like Alta Mesa realize that these systems, properly tapped, contain both their history and their future. They can disclose patterns of ownership, acquisitions, competition, and even the possible location of mineral deposits.

In order to tap this potential, however, the facts that the documents contain must be identified and input as metadata to the system. Simple keyword indexes, although a good first step, are not sufficient: they don't retrieve documents that say the same thing with different words. 42° 06' 12"N, 76° 15' 43"W is hard to recognize as a synonym for Owego, NY.

For decades all of this data has been entered manually. The results are intelligent, but prone to data entry errors. Automating the process with a high degree of accuracy and consistency is a challenge, however, particularly when source documents use non-standard terminology, are blurry, or in unpredictable formats. To succeed, a good document management access platform must combine intelligent, semantically based data capture with a specialized knowledge base and an underlying ability to track and uncover relationships among entities and documents, no matter how they are phrased.

## Mining Oil and Gas Leases

Like any contract, oil and gas leases are long and convoluted. No two are alike in wording or format. Extracting the salient names, locations, terms and conditions for entry into lease management systems has been a manual operation that is time consuming and error prone. Streamlining this process will save

time and money. More, it may make a significant difference in the profits or losses of the company.

We have the technology today to build a software solution that automatically extracts names, lease terms, time, and locations, as Agile Upstream has

done with Agile Land Insights and the underlying ABBYY Compeno technology. The color-coded example below is taken from a lease that is 3,400 words long. It shows some of the features that Alta Mesa needed to identify and store in its Land Lease Management system:

- Time
- Location
- Lessor name
- Lessee name
- Obligations of lessee
- Rights granted by lessor
- Monetary considerations
- Other numeric terms (e.g., acreage)

### Contract Extract

THIS AGREEMENT made this 22nd day of October 2014, between GEORGE COSTANZA and wife, SUSAN ROSS, lessor, whose address is 1590 SHOTWELL LANE, ROUND TOWN, TX 12345 and VANDELAY ENERGY RESOURCES, LLC, 12345 KATY FREEWAY, SUITE 100, HOUSTON, TX 77777, lessee...

1. Lessor, in consideration of Ten dollars ...does hereby grant, lease and let unto lessee the land ...with the exclusive right of exploring, drilling, mining and operating for, producing and owning oil, gas (including carbon dioxide), sulphur and all other minerals, ...right to make surveys ..., lay pipe lines, establish and utilize facilities for surface or subsurface disposal of salt water, construct roads and bridges, dig canals, build tanks, power stations, power lines, telephone lines, employee houses and other structures on said land, The land ...is located in the County of CLARKE, State of ALABAMA, and is described as follows.

#### TOWNSHIP 1 NORTH, RANGE 2 EAST SECTION 3:

... 400 feet South and 500 feet West of the NE corner of the SE 1/4 of the NE 1/4 of Section 3; thence South 600 feet; thence West 700 feet; thence North 800 feet; thence North 90 degrees East 900 feet; thence East 100 feet to the point of beginning, less and except the 80 foot right-of-way of the Kramer & David Road and being a part of the SE 1/4 of the NE 1/4 of Section 3.

Lessor does hereby let and lease to Lessee all their mineral interest owned or claimed in Section 24 Township 6 North, Range 2 East, ... said land shall be deemed to contain 2.80 acres,

Lessee and agrees: (a) To deliver to the credit of lessor... one-eighth part of all oil produced and saved by lessee... to bear one-eighth of the cost of treating oil to render it marketable pipe line oil; (b) To pay lessor on gas and casinghead gas produced ... one-eighth of the amount realized by lessee... (c) To pay lessor on all other minerals mined and marketed or utilized by lessee from said land, one-tenth either in kind or value at the well or mine at lessee's election, except that on sulphur mined and marketed the royalty shall be one dollar (\$1.00) per long ton....this lease shall continue in force as though operations were being conducted .... after the expiration of the primary term, ... [If] there are no operations on said land, then at or before the expiration of said ninety day period, lessee shall pay ... one dollar (\$1.00) for each acre of land. ....

not more than 80 surface acres plus 10% acreage tolerance; ... may be enlarged to contain not more than 640 acres plus 10% acreage tolerance...

...6. This is a PAID-UP LEASE. ...Lessee shall not be obligated...to commence or continue any operations during the primary term. No well shall be drilled nearer than 200 feet to the house or barn ...Lessee shall pay for damages caused by its operations to growing crops and timber on said land.

### Extracted Entities

Time	Location	Lessor Name	Lessee Name
22nd day of October 2014	1590 SHOTWELL LANE, ROUND TOWN, TX 12345	GEORGE COSTANZA and wife, SUSAN ROSS	VANDELAY ENERGY RESOURCES, LLC
this lease shall continue in force as though operations were being conducted	12345 KATY FREEWAY, SUITE 100, HOUSTON, TX 77777	lessor	lessee
after the expiration of the primary term	County of CLARKE, State of ALABAMA TOWNSHIP 1 NORTH, RANGE 2 EAST SECTION 3		

Rights Granted by Lessor	Monetary Considerations	Other Numeric Terms
right of exploring, drilling, mining and operating for, producing and owning oil, gas (including carbon dioxide), sulphur and all other minerals	Ten dollars To deliver to the credit of lessor... one-eighth part of all oil produced and saved by lessee... to bear one-eighth of the cost of treating oil to render it marketable pipe line oil; (b) To pay lessor on gas and casinghead gas produced ... one-eighth of the amount realized by lessee... (c) To pay lessor on all other minerals mined and	80 surface acres plus 10% acreage tolerance not more than 640 acres plus 10% acreage tolerance 2.80 acres
Lessor does hereby let and lease to Lessee all their mineral interest owned or claimed in		

**Obligations of Lessee**

[If] there are no operations on said land, then at or before the expiration of said ninety day period lessee shall pay

No well shall be drilled nearer than 200 feet to the house or barn ...Lessee shall pay for damages caused by its operations to growing crops and timber on said land

## The Technology *cont.*

ABBY InfoExtractor is “taught” to identify each of these elements of meaning with a set of examples. The system must be able to identify multiple expressions of the same idea, as a human would. Any Land man can tell you that a “Pugh clause” and a “freestone rider” are synonyms, but this is not easy for a machine to do. However, given the right technology, it’s possible to “teach” a computer how to understand text.

Machines learn to use a variety of clues to help them understand what a word means, just as people do. They rely on what they already know about a topic, as well as how the word is used in order to triangulate on what a word means within a given sentence or paragraph. For high accuracy, good information access systems today build a dictionary of specialized terms and their synonyms—a “knowledge base.” More advanced knowledge bases take advantage of ontologies that define classes of entities, their attributes, and relationships within a specific industry use case. These knowledge bases, along with sample documents, are used to train the system to understand industry concepts.

In order to meet the requirements for land lease management, the underlying technology has to understand language on multiple levels. Specifically, it needs to:



- Find concepts, no matter how they are worded. This includes recognizing all the variations in phrasing for specialized terms and concepts.
- Merge similar concepts across structured databases and text documents, no matter how they are expressed.
- Recognize documents that are related in some way, even if the similarities are approximate rather than exact. For instance, a piece of property might be described as an address or a set of coordinates. Or the property may have changed ownership.
- Adapt to changes in how leases are worded and add new terms as they arise. A completely rule-based system is too rigid to support work in a changing and dynamic industry.
- Understand the natural language that a user might choose in querying a system so that users don’t have to learn a specialized query language.
- Scale to ingest paper and digital documents when they arrive in large batches, as well as singly.

## The Technology cont.

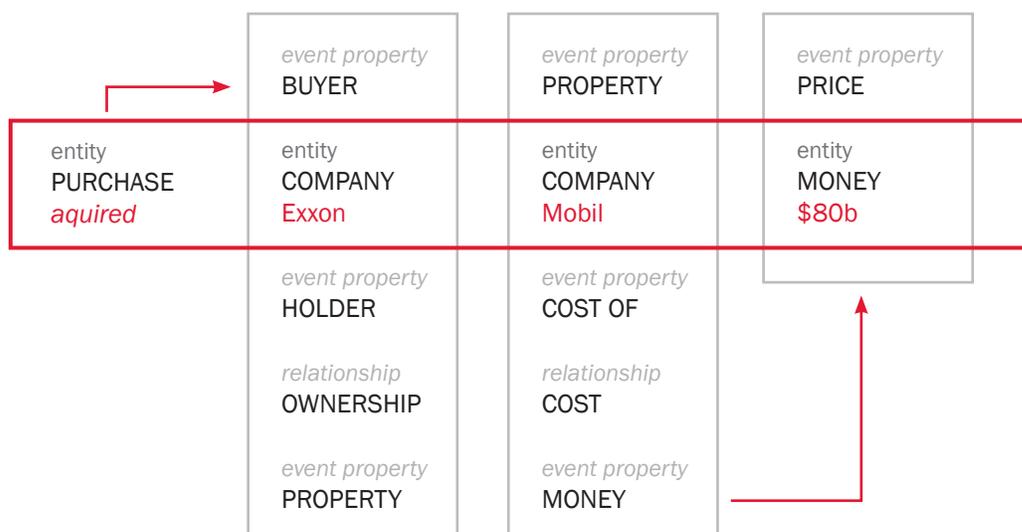
Systems use text analytics to identify the important elements in a text. For instance, they may:

- Extract names—of places, people, events, objects or organizations
- Extract expressions of time (e.g., dates for deadlines but also lengths of time such as “two weeks”) and also construct timelines for a sequence of documents or to determine which provisions are the most current
- Extract geographic locations in a variety of formats and relate them all to a location on a map
- Extract details such as amounts, costs, or contingencies
- Extract properties for each class of entity (e.g., places have locations and addresses; events have speakers, who have names and affiliations)
- Extract the type of relationship that exists between all the entities and events (e.g., acquired by, owned by, defaulted, sold, etc.)
- Extract the entire fact: a computable “who did what to whom” and store it for analysis

ex

### Fact of Acquisition

“Exxon acquired Mobil for \$80 billion.”



ABBYY Compeno Solutions

*ABBYY InfoExtractor examines the relationship of one word to another. It extracts and stores “triples” (or “facts”) of subject (Bill), verb (hit) and object (Bob). Unlike a traditional search engine, a semantic system must be able to differentiate that a company was acquired from the company that acquired it.*

ABBY's use of linguistic technology throughout its stack, from data capture to fact extraction, makes it possible to find these meaning linkages across documents and decades. The metadata that is extracted helps users explore a large collection to find hidden or unknown relationships. Strong workflow and visual interfaces hide the underlying complexity and invite simple navigation that steps the user through a task. With a strong linguistic foundation, it's possible to ask complex questions like, "Which Wolfcamp prospects do I have to drill in the next 6 months to retain the lease?"

This new, adaptive, interactive type of application brings us squarely into the realm of cognitive computing. It is meaning-based, and designed to solve complex problems.

## The Technology *cont.*

## The Design



The goal of any software application is to solve a problem by making it easy to accomplish a specific task. Any successful application is rooted in good technology. But surfacing the features of that technology platform so that they are apparent to the user, and so that they fit conveniently within the task's workflow is both an art and a technical challenge. Agile melded their knowledge of the land lease management process with the features of ABBY's Compreno technology to design an interface that will boost productivity and accuracy.

# The Design cont.

In the screen shot below from Agile Land Insights, you can see that they have taken the elements identified by ABBY InfoExtractor—names of the lessor and lessee, date, locations, contract terms—and used them to help analysts navigate a contract. The expert is also able to validate the accuracy of the system by comparing the OCR text and data extractions with an image of the original lease.

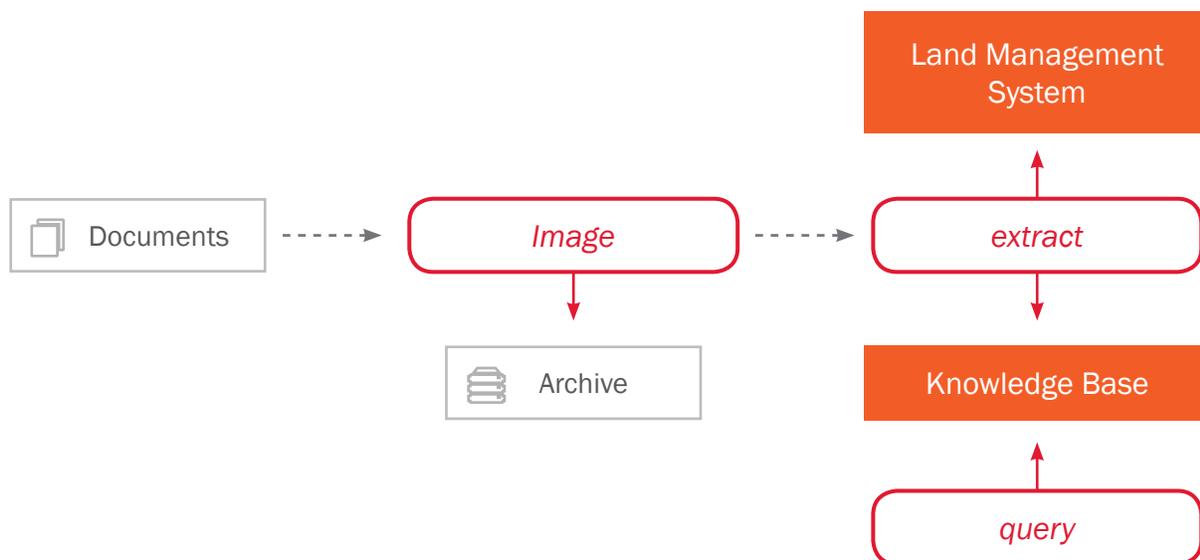
The screenshot displays the Agile Land Insights interface. On the left, a document viewer shows an 'OIL, GAS AND MINERAL LEASE' document. The central sidebar contains navigation options: Dashboard, Upload, Search, Processing (18 items), Needs Review (10 items), Validated, Error (6 items), and Admin. The right-hand panel, titled 'Lease Costanza | Pending Validation', shows the document's status as 'Pending validation since 12/22/15 9:26 AM'. Below this, the 'Parties' section lists the Lessor Name as 'GEORGE COSTANZA and wife, SUSAN ROSS' and the Lessee Name as 'VANDELAY ENERGY RESOURCES, LLC'. The Lessee Address is '12345 KATY FREEWAY, SUITE 100, HOUSTON, TX 77777'. The document text in the background includes details about the lease location in Township 4 North, Range 2 East, Section 24, and the lease terms.

This design fits the normal processes the user follows in the many tasks that the underlying information will be used to support, from finding the terms of a specific lease to understanding the company's holdings in a larger geographic area. It supports answering specific questions with high precision and the larger overviews that are necessary for the company to get a analytic and/or visual understanding of its holdings.

## The Design *cont.*

Once the elements needed to navigate the information were established, the next step for Agile was to analyze and codify the processes that are germane to land lease management. For example, as shown in the diagram below, the document ingestion and access process digitizes a document, uses semantic data capture to extract the text, stores the extracted terms and relationships in both the land management system of record, and the linguistic knowledge base.

### *Agile Land Insights*



Users can then access the system to answer questions such as:

- Find all documents related to Property X
- What leases will expire in July?
- What do we already own that is near Property Y, and what properties might we acquire that are available?

Working with ABBYY, they connected the information sources that will feed information to the system and normalized across data types. They also created standardized outputs to feed the data into Alta Mesa's Land Management system, replacing much of the manual data entry now required.

Finally, Agile designed and tested the user interface, based on their analysis of business processes such as acquisition and divestiture. The resulting application encapsulates process knowledge and documents in a set of tools that support land lease management, taking boxes of paper documents, and turning them into a digital information foundation for the business.

## Conclusion

Far from the frenzy of Internet IPO's and tech bubbles, companies like ABBYY, Agile Upstream, and Alta Mesa are solving practical problems that increase revenue, streamline processes and improve productivity. Measurably.

Agile Land Insights, powered by ABBYY InfoExtractor, gives oil and gas Upstream companies like Alta Mesa a significant information advantage over their competitors because it enables them to analyze more leases at greater depth and speed. A quicker understanding of their current holdings will give them greater agility to react to the fluctuations of oil prices in the marketplace. They should also be able to optimize production for each property by being alerted to lease terms—drilling depths, ownership, or expirations—in a timely manner.

Both Alta Mesa and Agile Upstream believe that

### **systems like Agile Land Insights will produce significant return on investment for the companies**

Automating lease management will reduce the data entry error rate. Even more importantly a system such as Agile Land insights, based on a specific oil and gas ontology developed using ABBYY InfoExtractor, will increase the information transparency of transactions. Transaction parties will be able to work more efficiently together and optimize their respective interests, faster, streamlining the all-important merger, acquisition, and divestiture processes that lie at the heart of the oil and gas industry. That should be a win across the entire industry.

We believe that Agile Land Insights represents a significant trend in the technology industry. As advanced semantic platforms are embedded within a process-oriented task-based application, the technology is hidden, and the user is able to simply work, without having to become a technical wizard first. These early cognitive systems adapt to new information and will help, not hinder users in getting their work done better and faster.



# About Synthesis

Synthesis is a research focused investigative and consulting firm working in the area of emerging software technologies. Synthesis provides business advisory services to vendors and buyers of cognitive computing, search and text analytics technologies. We maintain active research programs in these areas, write about these topics, speak at industry conferences, craft marketing messages, and analyze strategy and positioning for our clients. Our analysts represent over 50 years in the IT and online industries. Synthesis and its affiliates advise vendors on market trends, potential alliances and competitors. We guide their customers in setting information strategies, measuring ROI, and understanding how the impact of current and future market trends and demand patterns will shape the implementation of tomorrow's systems. The Synthesis Blog can be found at [www.synthesis.com](http://www.synthesis.com).

## About the Author

**Sue Feldman, CEO, Synthesis**



Sue Feldman is founder and CEO of Synthesis. She speaks frequently at industry events on topics such as trends in computing, conversational systems, big data technologies, and the hidden costs of information work. She wrote the chapter on search engines for the Encyclopedia of Library and Information Science, and was the first editor of the IEEE Computer Society's Digital Library News. In her book, *The Answer Machine* (Morgan & Claypool, 2012), Sue discusses the technologies behind information seeking and analysis, and their central role in the future of computing.

Before founding Synthesis, Sue was Vice President for Search and Discovery Technologies at IDC (International Data Corporation), where she directed research on the technologies and markets for search, text analytics, categorization, translation, mobile and rich media search. Prior to coming to IDC, Ms. Feldman was founder and president of Datasearch, an independent technology advisory firm, where she consulted on usability and on information retrieval technologies. She is a founder and former president of the Association of Independent Information Professionals, a member of the Association for Computing Machinery, and has won numerous research and writing awards. Ms. Feldman holds degrees from Cornell University in linguistics and from the University of Michigan in information science.

Synthesis white papers place a product or service within the context of our independent market and technology research. All opinions are our own.