

Docxonomy Case Study:
A Top 5 Pharmaceutical Company

Table of Contents

Business Need	3
The Challenge	3
The Solution	3
Results	4
About Docxonomy	5
Our Team	5

Business Need

The customer had hundreds of thousands of unclassified (or miss-classified) and untagged documents with incomplete metadata. Subsequently, their users struggled to find mission-critical documents during the course of their everyday work.

This customer also desired the ability to analyze and report on information within their documents (new drug applications, adverse event reports, etc.).

The Challenge

It was not feasible, incredibly expensive and error-prone to manually tag and classify the documents. Based on the volume, it was calculated to take over a year for a team of 50 people, fully dedicated to review and classify the content.

The content was spread among multiple repositories and file shares, largely disorganized and in some cases, with no metadata whatsoever.

Additionally, extracting and reporting on data points within unstructured content was also not humanly feasible, especially in large volumes.

The Solution

Using the Docxonomy platform, a machine learning model was trained using 3000 sample documents to recognize document types based on the context of the words within. Samples were separated into Training and Test sets with a 75/25 percent split. Training the model took approximately 1 hour to complete.

For this case study, Docxonomy's out of the box crawlers were used to traverse the repositories, analyze, classify and auto-tag the documents collected, outputting to an intelligent, searchable index.

During analysis step, analytics were extracted from the content, in the form of identified entities (drug names, compounds, side effects, etc.), noun phrases and proprietary, customer-specific terms.

Results

Thousands of resource hours were saved, with greater speed and accuracy than if this were attempted manually. Documents are now searchable from one location, fully classified and auto tagged.

Classification Results:

Auto Classification Results	
Overall Predication Confidence	97.5%

Document Classifications Identified in the Analyzed Repositories
Form 1571
Form 1572
Clinical Study Report
Cover Letter
CV
Investigational Brochure
Investigational Plan
Medwatch
Response
Summary of Clinical Pharmacology Information
Summary Safety Information

Analytical analysis output included frequency counts on particular instances where entities were found. For example, “Drug Product ABC” appeared 1850 times within the documents. Such analysis data points are outputted along with other auto-tagged items and customer-provided metadata, fit for direct ingestion into the customer’s business intelligence tool of choice.

The domain-specific data points that were extracted could then be aggregated and reported on: drug names, side effects, inactive ingredients, etc.

This led to actionable insight that was previously unknown or inaccessible. These insights could then enable data-driven decision making and improved organizational performance.

Entity Types Trained and Identified with Text:

Life Science Entities	General Entities
Pharmaceutical (Trade Name / Generic Name)	Person
Product Short Name	Place
Xxxx_Number (customer-specific identifier)	Number
Side Effect	Date/Time
Inactive Ingredient	Money

Compound Asset	Quantity
Compound ID	
Study Number	
Route of Administration	

143,000+ Entity instances were identified in all.

About Docxonomy

Docxonomy is an intelligent search and analytics platform for the enterprise that lets you crawl and analyze structured and unstructured data regardless of where it is stored. Docxonomy leverages artificial intelligence and machine learning to analyze text and rich media. Through this analysis, we can extract meaning and context, enabling the platform to classify files, identify entities, intelligently search, recognize file similarity, and answer questions. This process allows organizations to aggregate data that is otherwise inaccessible and draw insight from the valuable information they already own. Best of all, you can deploy Docxonomy on-premise or in your cloud of choice. Our proprietary crawler technology works with existing systems, so there's no need to migrate data to Docxonomy or spend months deploying a new platform.

Our Team



Bryan Reynolds
Founder & CEO

Mr. Reynolds has over 25 years of experience as a successful entrepreneur, senior executive and managing consultant with core competencies focused on enterprise content management, mobility, business process engineering, imaging, and records management. Currently, Mr. Reynolds is the Founder and CEO of Docxonomy. The breadth of his knowledge includes the architectural design and development as well as project management of numerous global, large-scale document/records management initiatives across multiple industries including pharmaceutical, biotechnology, medical devices, financial services, insurance, healthcare and public sector.



Thomas Kivlehan
Chief Data Officer

After graduating from Lehigh University with a Bachelor of Science degree in Computer Engineering, Mr. Kivlehan became immersed in consulting services and was focused on enterprise-wide content management systems. For over 10 years he was a self-employed entrepreneur. His work included customization of Documentum, Sharepoint, imaging systems, mobile development and more. He has worked with customers in the Life Science, Insurance, Financial, Packaging, Automotive, Petrochemical, Legal and Real Estate industries. Mr. Kivlehan is delighted to apply his 23 years of experience to help make Docxonomy a powerful, ground-breaking tool, essential to businesses of all sizes.



Stephen Maderak
Chief Information Officer

Mr. Maderak has over 15 years of experience in numerous technology roles including support, IT management, consulting and product development. He has experience working in various industries including payment processing, education, aerospace, life science, and more. He has primarily been focused for much of his career in unstructured data management and related processes and products. He was an early leader in search, where he oversaw the implementation of many solutions built on Documentum, Sharepoint, and many others. Mr. Maderak currently has responsibilities in the management, design, and implementation of Docxonomy's various products and infrastructure.