NLP text analytics for Regulatory Affairs

Reduce costs and speed compliance across labeling, regulatory changes and master data management

Introduction

The pharmaceutical industry is among the most heavily regulated in the world. AIML approaches like natural language processing (NLP) augments regulatory teams, and reduces the cost of regulatory compliance, as the huge volume of text-heavy documents means that manual efforts are often slow and expensive.

IQVIA provides advanced text analytics to scale extraction of key data from unstructured text, find and highlight key information within regulatory documents, code to regulatory standards for compliance, structure components for draft dossier regions, detect inconsistencies across documents and more. Recent and upcoming changes in regulation mean that companies require new tools and solutions to assist with regulatory review and compliance. In some cases, meeting the regulatory requirement is straightforward, while in other cases, accessing the necessary data can take a significant amount of time, money and effort, all of which increases costs but does not necessarily increase revenue. The case studies below demonstrate the value that text analytics can bring for regulatory affairs.

Natural Language Processing (NLP)-based text analytics can bring value in a wide range of regulatory use cases:

Regulatory labeling

NLP-powered search of reference Drug Labels (e.g. FDA, EMA, UK, France, Spain) empowers teams to find the right information rapidly, and compare side-by-side, for improved analysis of label information.

Regulatory intelligence

Application of NLP within integrated data flows to provide alerts and updates on regulatory guidance changes, both internal (e.g. CAPAs, RTQs) and external (e.g. Regulatory guidelines, FDA Letters).

Master data management and regulatory mappings

Extraction of data attributes from regulatory documents (SMPC, eCTDs, CMC documents) and mapping to standards (e.g. IDMP, xEVMPD) or master data management provides teams with access to metadata to improve compliance.

Regulatory document development

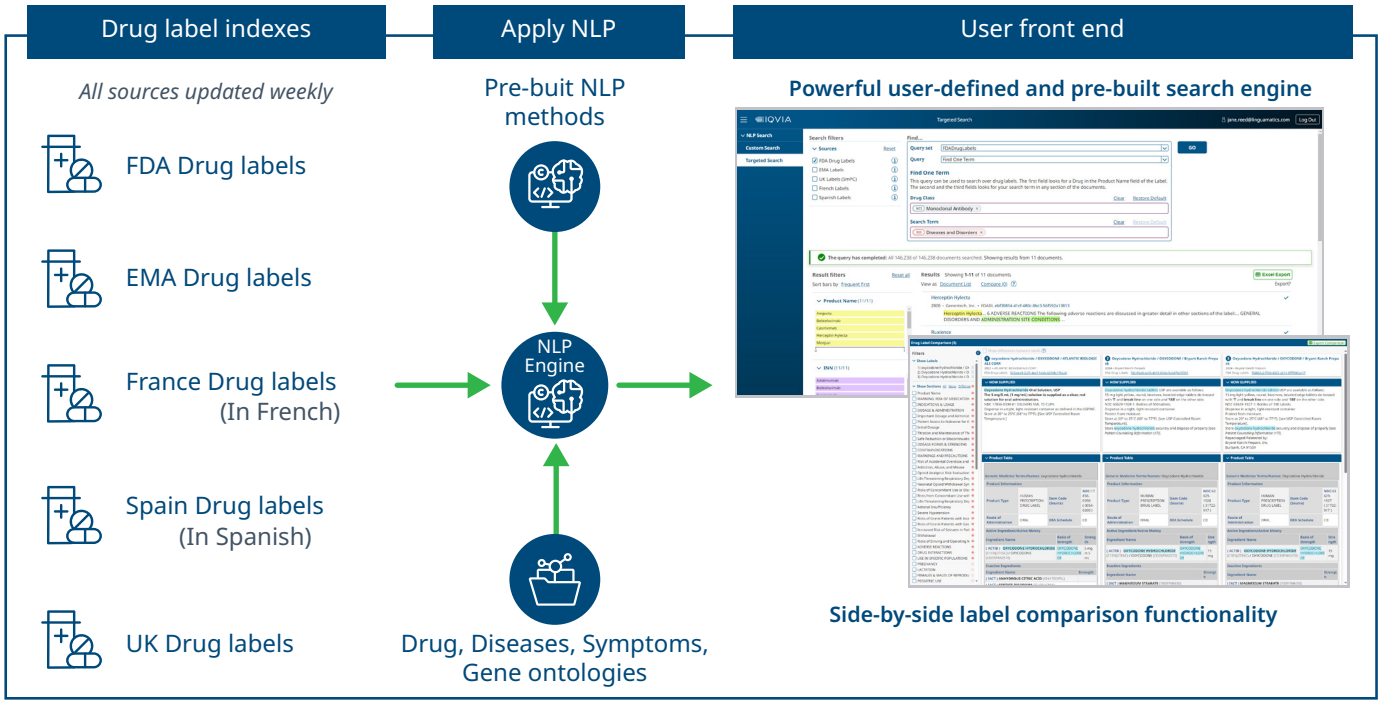
Combination of NLP technologies for component extraction and generative text for regulatory QA/QC and document digitalization, creating draft dossier and label sections, with human-in-the-loop review.

Regulatory labeling

Drug labels provide a vital conduit for relaying essential medication information to healthcare practitioners and patients. These labels encompass critical particulars like drug names, strengths, indications, contraindications, dosages, and administration instructions. Information within drug labels is dynamic and subject to continual updates across the drug's lifecycle. Innovative AI technologies such as NLP can augment the work of regulatory professionals to find, understand and manage label changes.

The dynamic nature of drug labeling, with hundreds of new or updated labels published weekly across regions, means that teams need to search through diverse data sources and languages. By leveraging NLP, labeling teams can save time and effort in finding, extracting and comparing drug-label information. NLP enhances analysis, enabling teams to extract maximum value from each label, expediting the process, and ensuring greater accuracy. The IQVIA Labeling Intelligence Hub (see figure 1) allows teams to build label searches or deploy pre-built searches to optimize label writing and analytics, culminating in more accurate information during the assessment stage and leading to well-informed decisions in response to proposed changes. By leveraging these technologies, organizations can free up valuable staff time, allowing them to focus on higher-value activities and strategic decision-making.

Figure 1: Schematic overview of IQVIA's Labeling Intelligence solution



Top 10 pharma reduces label comparison time by 95%

A top 10 pharmaceutical company has been using the IQVIA Labeling Intelligence Hub for two years, to explore drug label data efficiently. Various teams within the company, including global labeling, regulatory affairs, medical, and safety, use the Labeling Intelligence Hub to access key reference labels. The tool enables users to conduct customized searches, refine results, and export data for further analysis. Additionally, users can compare specific labels through an interactive view and access original documents directly. This solution streamlines the process of developing new labels, updating existing ones, and expediting regulatory approval, ultimately saving time for the teams involved. Assessment by the labeling team found using the Hub **reduces label comparison time by 95% and estimated an equivalent of 9 FTEs worth of manual work saved**, enabling experts to focus on deeper analytics and strategic outcomes.

Regulatory intelligence

Often, compliance teams depend on manual methods to monitor regulatory updates, such as having individual team members regularly perform checks of relevant agency websites or subscribe to industry emails, to stay up to date on recent guidelines, public consultations, and meeting conclusions. Although the process is important because it provides compliance teams with essential intelligence to identify key concerns, deadlines, events, and regulatory decisions for compounds of interest, it is generally costly in terms of resources and time. NLP can be applied within integrated data flows to provide alerts and updates on regulatory guidance changes, both internal (e.g. CAPAs, RTQs) and external (e.g. Regulatory guidelines, FDA Letters).

NLP for risk management provides insights for pipeline decision making

Risk management insights from internal and external feeds: A top 10 pharmaceutical company's product development and supply team needed a way to improve its understanding of internal and external risk management data to optimize the formulations, commercial supply, and postmarket regulatory compliance of its products. To fuel the initiative, the team developed a data lake to capture important internal and external feeds. Internal feeds included deviations, corrective and preventative actions (CAPAs), risks, and responses to questions (RTQs). External feeds included FDA warning letters, biological license applications (BLA) review reports, white papers, and industry benchmark repositories. The team employed NLP to structure and generate this intelligence data, extracting concepts, relationships, and sentiments embedded in the information. The data's value to the team was further enhanced by easy-to-understand visualizations, enabling end-users to drill down and navigate the information. These data pipelines and workflows were updated automatically to deliver sustainable and scalable reporting of the regulatory landscape, featuring key risks and recommendations to act upon. Using text mining over these large document saved the team significant time as NLP can search hundreds of pages in seconds, pulling back the key data that subject matter experts can then review effectively.

NLP surfaces key substance risk updates from regulatory guidelines

Intelligence for key substances from regulatory websites: One agrochemical company needed effective updates from specific EU agencies around particular substances and excipients. An NLP workflow was developed to semi-automate information acquisition and develop summaries. A key feature of the approach involved the integration of deterministic NLP technology with text summarization using Large Language Models (LLMs). Using this combination of NLP technologies, a regulatory intelligence assistant was created, which provided team members with user-friendly question-and-answer access to updated regulatory information and risk categorization for substances of interest. By employing this model, the team delivers dynamic insights into various regulatory fields, highlighting major areas of risk, by extracting, summarizing, and classifying information for user-specified substances.

Master data management and regulatory mappings

IQVIA NLP can save organizations time and money by rapidly finding, extracting, standardizing and structuring a broad range of regulatory data attributes from regulatory documents (SMPC, eCTDs, CMC documents) and mapping to standards (e.g. IDMP, xEVMPD) or internal master data management schema.

IDMP compliance provides a clear example of the benefits that automation within regulatory workflows can bring. IDMP (IDentification of Medicinal Products) is a set of international standards developed by ISO to help streamline and improve the safety of pharmaceutical operations across the entire drug development cycle. The standards were first published in 2012, have been through updates and iterations and are set to become mandatory in Europe in 2024; they are expected to be adopted by the FDA and globally over the next few years.

Capturing the hundreds of data attributes required per product, 70% of which lie in a variety of unstructured text sources, demands time, resources and investment. Structuring this valuable data can benefit the broader organization, enabling better data governance and master data management across discovery, development, clinical and manufacturing. IQVIA NLP can be deployed in workflows to find, extract, standardize and structure the required data elements from IDMP-relevant unstructured text documents. A text analytics approach brings multiple benefits over manual data extraction from unstructured text. Copying and pasting relevant data from documents into spreadsheets is labor-intensive, repetitive and tedious work, and is also prone to errors. Text mining uses business rules and standard vocabularies to systematically create a consistent, normalized set of product data, and can be used across tens, hundreds or thousands of documents. Business rules can be rapidly translated into search queries, and this flexibility is key as the IDMP framework evolves. This approach can also be used for other reporting frameworks such as xEVMPD, or to provide data suitable for enterprise master data management.

NLP saves manual resource with systematic extraction for IDMP compliance

A top 10 pharma consumer division used NLP to find and extract over 80 IDMP data attributes from a range of regulatory documents including Summary of Product Characteristics and regulatory dossiers (eCTD sections 3.2.S and 3.2.P). The challenges are familiar to anyone involved in IDMP compliance — varied document sets, in mixed formats (Doc, Docx, image and text PDFs), across different languages, integrating with SPOR vocabularies. The output needed to be mapped to the regulatory MDM schema for their IDMP submission and internal business use. NLP was highly effective for systematic attribute extraction across all document sets, saving the team significant time and resources.

Regulatory document development

Dossier submissions to regulatory authorities are labor-intensive document-driven processes and are very time-consuming for multiple staff across multiple departments. The source study data files tend to be text or image PDFs, which makes them hard to access and search effectively, hindering re-usability

and machine use. In addition, ensuring appropriate formatting, adherence to industry standard medical terminology, and accurate and consistent references to safety and clinical information can be challenging. NLP can be deployed within workflows for component extraction and generative text for document digitalization, draft dossier section creation, and discrepancy detection.

NLP for internal and external risk management and regulatory intelligence tracking

Augmenting the dossier creation process: NLP was deployed in a workflow for dossier digitalization. As source documents entered the system, they were passed to an OCR (optical character recognition) engine that transformed input PDFs to machine-readable standard files. NLP models were developed to extract relevant data attributes and data components from study documents, and these were provided for human-in-the-loop review and curation. This process output clean structured datasets, providing meaningful components for storage and downstream use by large language models for generation of draft dossier and label sections, again supported by human-in-the-loop review.

Detecting discrepancies in blinded data reviews: A top 10 pharma company deployed NLP in an automated process to improve quality control of regulatory document submission. The solution developed cross-checks for MedDRA codings, references from text to tables, decimal place errors, and discrepancies between the summary document and source documents (e.g. clinical listings). NLP powered the extraction of information from PDF tables as well as analysis of free text. Automated text analysis for quality assurance of submission documents can save countless hours or weeks of tedious manual checking, and potentially prevent a re-submission request. The key stakeholder reported "100% positive feedback from business teams" with the NLP solution.

ERROR CATEGORY	DESCRIPTION
Missing source tables	Identify references in the summary document to source tables not appearing in the same document bundle
MedDRA label errors	Identify source tables that should contain MedDRA terms and then check to ensure that accurate MedDRA terms appear
Incorrect formatting	Identify cells in tables that contain values with inconsistent formatting, such as doubled period, incorrect number of decimal places, addition of percentage sign
Incorrect calculation or threshold	Identify cells in tables where the numeric value for the particular cell is incorrect, or where the table title threshold is not met
Inconsistent units	Identify cells in tables where the units are not appropriate for the measurement reported, e.g. haematocrit, haemoglobin level, platelet count, etc.

Proven NLP addresses a range of regulatory challenges

IQVIA's NLP technology finds, highlights and extracts structured data elements from regulatory documents, which can provide rapid, systematic, repeatable analysis within regulatory applications. The platform can handle a broad variety of document formats and types, and provides agile querying and integration into enterprise workflows, enabling the flexibility needed to rapidly address critical business issues across regulatory affairs.

CONTACT US
+44 (0)1223 651 910 (U.K.) | +1 617 674 3256 (U.S.)
nlp@iqvia.com
iqvia.com/nlp