Intelligent data processing in healthcare & life sciences

nagarro

nagarro

Table of Content

Introduction

Did you know that in the 1960s, the vaccine for protection against mumps was the fastest vaccine to be developed - within four years? This record was broken only recently, when the medical fraternity developed and launched vaccines for Covid-19, in less than 12 months in 2021. This steep decline in vaccine launch time was mostly made possible because Covid-19 (or SARS-CoV-2) is a member of the Coronavirus family. The world had already seen two other outbreaks caused by viruses of the same family - SARS in 2002 and MERS in 2012. The research performed to study these viruses helped understand and fight Covid-19 more effectively.

Besides research, the innovations in biotechnology expedited the time taken for vaccine trials and rollouts. But biotechnology alone didn't win the battle against the pandemic. We must not forget emerging technologies (such as artificial intelligence (AI), machine learning (ML), Analytics, and cloud computing), which also played a crucial role in containing the outbreak by enabling data collection and analysis.

As infections spread worldwide, all governments and healthcare organizations had to jolt out of their inertia and were pushed to embrace technology to ensure operational speed and efficiency. On the flip side, digitization also brought along large amounts of complex data collected through prescriptions, medical records, reports, and images. Even if we say that's an ocean of data, it would be an understatement. The sector now needs intelligent solutions to sort, track and analyze all this data to deal with current and future outbreaks.

In this paper, we examine how an intelligent data processing system can sort, track, and analyze healthcare data to address these major challenges in the healthcare & life sciences industry.

Executive Summary

Data is at the heart of any digital transformation; unfortunately, 80% of all business data is embedded in unstructured formats like business documents, emails, images, and PDFs. The manual handling of document classification, key-value extraction, and metadata extraction in the healthcare industry is a tedious process, which requires enormous human effort. A complete end-to-end framework for document classification, key-value extraction, redaction of sensitive information, and converting unstructured data into standardized formats is imperative for the success of any healthcare or life sciences organization.

An Intelligent Data Processing (IDP) solution can transform unstructured and semi-structured information into usable data. IDP takes automation to the next level by capturing and processing various document formats with the help of technologies such as artificial intelligence (AI), natural language processing (NLP), computer vision, deep learning, and machine learning (ML) to classify, categorize, and extract relevant information and data. We bring you an IDP framework specifically designed to address the challenges faced by the healthcare and life sciences industry. The framework recognizes documents, extracts value, hides sensitive information, and converts unstructured documents into standardized formats, thereby increasing organizational efficiency and profits.

Challenges to data digitalization in the healthcare sector

Data Complexity and Volume: Most healthcare providers collect data through patient intake forms, consent forms, treatment evaluation, and health assessment forms. This data is either collected digitally or manually. It is then stored mainly in several different repositories, which creates silos and makes data utilization a complex process.

Multilingualism: In this era of globalization, where data comes from different parts of the world, multilingual support is essential for any product to capture the global audience. For example, if a drug is under trial in other countries, the clinical trial data will be in different languages for analysis, consolidation, and aggregation, causing a major challenge for the organizations in the sector.

Privacy and data security: There are various regulatory compliances that a healthcare organization must adhere to while dealing with digital data. Regulatory compliance requirements for privacy and data security like HIPAA (Health Insurance Portability and Accountability), GDPR, 21 CFR part 11 are just a few of such examples. Ensuring that subjects' data is stored securely can be tricky.

Multiple input channels: Data for various medical processes is collected through emails, social media, fax, scanners, online forms, and voice mails. Numerous data channels and file formats make data aggregation difficult. Every business requires this data in a uniform structure to aggregate it, report it to regulatory bodies, and analyze it.

Limitations of OCR (Optical Character Recognition) tools: Most of the OCR tools available for data recognition and extraction do not support the extraction of handwritten text. Since text recognition for scanned documents is quite poor, most OCRing solutions are unable to extract the text into meaningful data. The lack of interoperability restricts sharing of health information across institutions and software systems. The fractionalized nature of the healthcare system makes it difficult to track patients across different healthcare systems, resulting in incomplete data.

Making healthcare efficient with intelligent data processing

Organizations can apply IDP to most functions within their healthcare and life sciences teams. Here is a list of some applications of IDP in the sector:

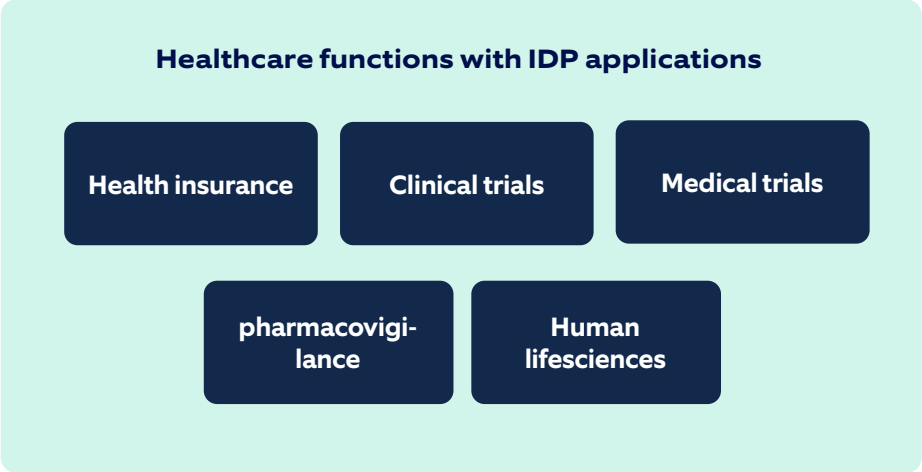

Figure 1

Health Insurance industry: Digitization of health insurance claims and data standardization can expedite the submission and processing of insurance claims while simultaneously bringing down costs. Organizations can use IDP in KYC (Know Your Customer) verification, policy issuance, fraud detection, and claims management.

- **KYC verification:** This verification extracts customers' data from identity cards and other documents, verifies the data, and stores it in digital format
- **Data Entry:** Extracts information from scanned customer application forms and inserts it into a CRM system.
- **Policy Issuance:** Processes large volumes of forms and documents to issue a policy instantly.
- **Fraud Detection:** Extracts claim information that is passed to trained machine models that work alongside adjusters.
- **Claims Management:** Intelligently identifies and classifies relevant data from documents such as claim intake, loss notification, and loss estimates (identification documents, bills, and health insurance forms). The IDP can identify the document type, treatment given, bills, and code the obtained data.

Medical records: Healthcare organizations can use IDP to process medical administrative records, claims data, and patient health records. The cost of processing medical records ranges from $0.25 to $2.00 per page. Implementing IDP can reduce this cost by extracting valuable information and securing sensitive data by redacting it.

Human data sciences: This is an emerging discipline that integrates human science and technology to advance our understanding of human health. In this amalgamation of science, data, and technology, the entire activity of collecting and collaborating with human health data is very crucial. Data sources such as patient health history, drug history, diagnosis, and treatment exist in various formats. But this data is often unstructured. Patient privacy concerns add to the challenges organizations face in the field. This is where IDP comes in as a great tool, which can address all these challenges.

Clinical Trials: The pandemic has highlighted the need for faster clinical trials, which would require automating data processing, to reduce the time for vaccine trials and launch. IDP can facilitate this by extracting data to generate reports that support the research. This would help while presenting the findings to the regulatory bodies. During the Covid-19 clinical trials, pharma companies chose subjects across the globe and offered them medicines at regular intervals. They had to observe the issues and note their parameters to see the medicine/vaccine efficacy. However, this data is usually collected and stored in different formats across different geographies. IDP can extract meaning from this data, create a database, and analyze it to gain useful information.

Pharmacovigilance: Adverse drug reactions are a severe concern for the healthcare industry, especially while testing and launching drugs. Pharmacology companies utilize numerous formats (forms) to report adverse effects reported for drugs, devices, and vaccines. Processing these forms to extract, standardize and report the data is highly cost-intensive. IDP can configure the documents to help extract the data automatically, redact sensitive information, and accelerate the regulatory report generation capability.

The framework

Nagarro is a digital product engineering company that enables organizations to utilize technology and data to address business challenges. Having worked with global players in the healthcare and life sciences sector, we understand the current landscape and realize that using data correctly could make or break an organization's future. To help organizations source, visualize, and apply data in the right manner, we have developed an IDP framework. This framework is designed specifically to solve the data processing-related problems faced by the sector.

Our framework automates any content-based business process to make sense of any unstructured data. Instead of typing in the data and routing it via paper or email, this framework allows you to capture and digitize the content from the very beginning. It automatically identifies the document type, extracts critical information, and delivers it directly to any repositories. The framework brings together API (Application Programming Interface) integration, big data, OCR, along with E-Text enabled data extraction, redaction, data storage, and transmission to process a large quantum of documents at higher speed and lower cost.

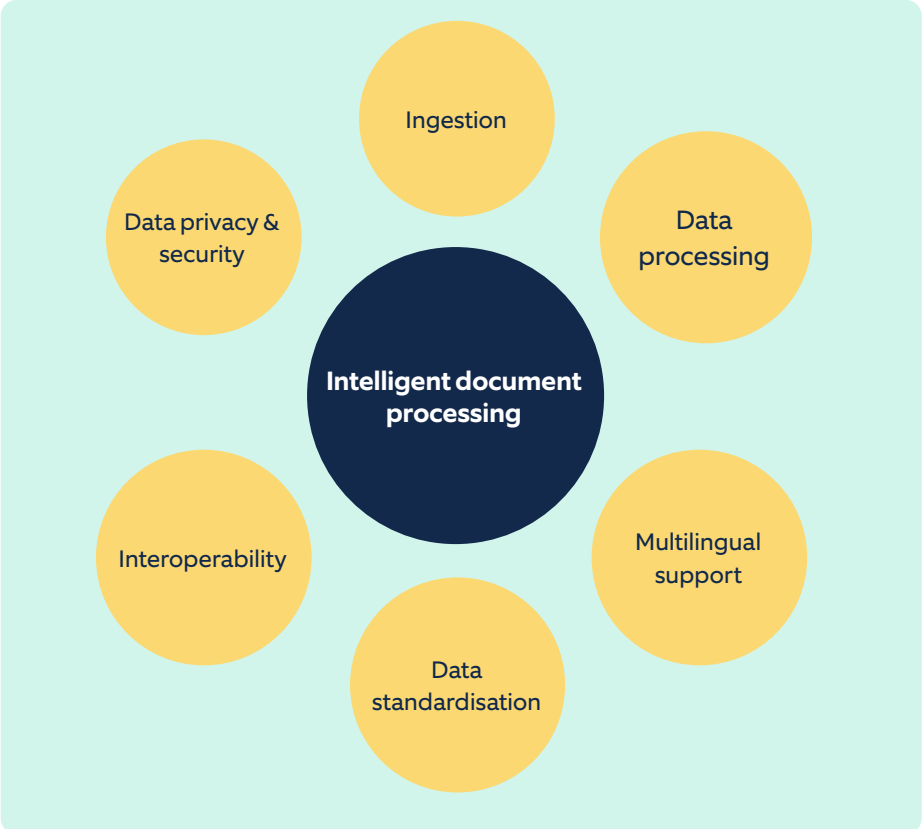

Figure 2: Visual representation of intelligent data processing framework

The framework - putting intelligence in IDP

The framework comes with multiple inbuilt capabilities such as:

Ingestion: It allows easy ingestion of documents from numerous platforms like mobile, email, fax, scanners, UNC folders, and MFPs. Automated scheduler to capture data from any instrument.

Data processing:

- **Image Processing:** Comprises document normalization through OCR, image processing, page processing, barcode processing, regular expression, and image comparison.
- **Classification:** Involves document type classification and automatic boundary identification.
- **Extraction:** Simplified data extraction that works autonomously with minimal or no user interventions. Automated extraction of configured data for defined fields uses a regular expression, barcodes, optical character recognition through business rules, database lookups, fuzzy logic, and NLP.
- **Validation:** This is designed to review extracted metadata fields. If the data requires manual intervention, there is a User Interface to review and validate data classification and extraction. If the fields/data have incorrect or incomplete information, it will ask users to complete or verify the captured data. The system can manually override data during validation.

Multilingual support: The platform supports multiple languages like English, Spanish, and French for text identification, classification, extraction, and validation.

Data standardization: The framework supports data extraction from structured and unstructured documents and organizes it into customized formats like XML and JSON (JavaScript Object Notation) consumed by the workflow processes.

Inter-operability: A system that supports API-based integration can create and exchange data across various leading industry technologies. As the first step of digital transformation, the framework can be bundled with new and existing products such as RPA (Robotic Process Automation) t ECM (Enterprise Content Management), and ERP (Enterprise Resource Planning). It allows customers to have best-of-breed solutions to meet their unique requirements and deliver value much faster.

Data Privacy and Security: We took exceptional care while designing the framework to ensure that the system extracts, sorts, and stores all the data files on the cloud. We have ensured that we follow all the associated regulatory compliance of data privacy and security.

- **Redaction:** This framework provides the capability to redact PII (Personally Identifying Information) and PHI (Personal Health Information) data, which protects against any exposure of patients/subjects-related personal information. The framework has a supervised learning feature for redaction and coordinate finding.

- **Storage:** This is a secure data storage medium to store original and extracted data files on the cloud. It offers a controlled environment with authorized access to data, audit, and version control.

Why should you consider this framework?

Increase Efficiency and Productivity: Intelligent data processing lets you process voluminous data quickly and eliminates paperwork to a great extent. Since your employees do not have to spend time on data entry, they can focus on more meaningful tasks, thus increasing efficiency and productivity.

Saves Time and Money: This framework saves all the time required in manual extraction of data from multiple sources, resulting in cost-saving.

Simplifies Data Privacy and Compliance: Data privacy is an understandably big concern for many organizations because any data breach can cause loss of trust and even monetary fines. Intelligent data processing allows you to redact and secure private and sensitive information. You can set it up to allow only authorized users to access selected information.

Supports Scalability: The rise of data, content, and document volume makes managing all your company's information impossible, especially when around 90% of this content is unstructured. With analysts predicting a tenfold growth rate of data over the next five years, the global data sphere will become even harder to manage. The intelligent data processing system is highly scalable, which allows you to file and organize as many paper and digital documents as required.

Reduce physical storage space cost: The rates of physical storage spaces have been rising worldwide. In the healthcare sector, the use of paper to capture data is still quite prevalent before digitization creeps deeper into the health delivery network. With IDP, we can integrate with cloud storage spaces which shall further reduce the necessity for physical storage spaces.

Outlook for the future

The pandemic may become endemic in due course of time but after having experienced the benefits of digitalization in healthcare, there is no turning back for organizations or patients. As technologies emerge and advance further, the data complexity in the sector will only intensify even more. Organizations need to think ahead and have tech-driven operational models that can evolve with time. Adopting an intelligent document processing model is the right step in that direction.

Having said that, it's also important to take that step properly. Moving towards a complete Intelligent Data Processing solution is a complex process with multiple steps. It requires an end-to-end view from a functional as well as technical perspective. The good news is that there are many accelerators within Nagarro which you can help and handhold you through this journey. Bringing our expertise and framework together, we can help your organization get its data strategy just right.

Interested? Reach us at lifesciences@nagarro.com

References

https://www.iqvia.com/-/media/iqvia/pdfs/library/white-papers/data-digitization-and-disruption.pdf

nagarro

About the Author

Sakshi Bishnoi

Sakshi Bishnoi is a consultant with Nagarro and has observed Healthcare and Life Sciences processes closely for over 12 years. She also carries qualifications in operation management and likes to make processes lean. She has tried to do the same for the healthcare & life sciences sector through this whitepaper. She lives in Gurgaon, India with her family that includes her husband, a kid and their cat, Cora

Editor
Shubhra Pant

Designer
Sri Harsha Dayanandachari

About Nagarro
In a changing and evolving world, challenges are ever more unique and complex. Nagarro helps to transform, adapt, and build new ways into the future through a forward thinking, agile and CARING mindset. We excel at digital product engineering and deliver on our promise of thinking breakthroughs. Today, we are 15,000 experts across 27 countries, forming a Nation of Nagarrians, ready to help our customers succeed.