Recycle and Reuse of Medical Content: Intelligent Content Authoring



ANAND KIRAN
Co-founder, EVP of Global Operations and Medical Solutions at Indegene

Industry Overview/Trends

The life sciences industry today is facing a number of challenges, characterized by economic uncertainty, reform-driven pricing pressure, increasing demand for innovation, and an ever-tightening regulatory environment.

In such a global scenario, the velocity and volume of technology change are directly impacting the way pharma companies have been operating across the value chain. Recent research from MIT reports that around 90% of business leaders believe that their organizations are currently being disrupted by advances in technology and digitalization.[1]

Medical organizations are increasingly looking to drive efficiency and productivity in almost all areas of traditional workstreams.[2] Throughout the drug development process and life cycle management, digital content is "created" and "consumed" by various cross-functional departments like Regulatory, Safety, Medical, or Clinical.[3] The absence of a homogenous platform often leads to delay in bringing the right information to the right person at the right time.

Content Consumption

Content Management (Identification, Retrieval, and Archival)

Broadly speaking, a holistic content management solution for the life sciences industry would need to enable identification of relevant content from the complex mass of available data across several operational silos. This content comes in a variety of forms, ranging from unstructured to structured. In addition, a seamless content management model should enhance collaboration between the diverse stakeholders involved in the life cycle of the content and should provide a natural language-searchable catalog of content.

To summarize, the core objectives of such a model would be three-fold.

- **REDUCE** clutter in content
- **RECYCLE** core content efficiently for downstream usage
- **REUSE** core content effectively

Content Creation

Auto Generate and Repurpose Content

Traditional models, associated with content authoring and submission workstreams and processes, are human capital intensive and involve substantial expenditure of the most vital resource in the pharma environment—**time**. All content-related activities, spanning across R&D, Clinical, Safety, Regulatory, and even Commercial are a function of time.

An operational foundation based on artificial intelligence (AI) that extracts, interprets, stores, and connects unique content as components tagged on the basis of key messages/topics will enable repurposing of content to create documents across Clinical, Regulatory, Safety, and Medical. This would ensure accuracy and compliance besides expediting the time-to-agency and time-to-market.

Updates to Content/Document

Considering the nature of product and clinical development, and requirements of life cycle management, majority of Clinical, Regulatory, and Safety documents are dynamic in nature and tend to undergo regular or ad hoc content updates based on new information or changes to existing understanding. Managing such updates, primarily by ensuring that the content/information across documents is updated in a timely manner and is compliant with regulatory requirements, is currently a major challenge in the industry.

An AI-based system that builds content lineage would be instrumental in assessing the impact of an update besides accurately and effectively managing updates across the document chain to be compliant with regulatory requirements.

Potential for Disruption

There is a significant opportunity to disrupt the status quo by bringing together deep domain expertise and understanding from medical, scientific, clinical, and therapeutic perspectives and combining them with emerging modern technologies such as AI, deep learning, and natural language processing (NLP) for meaningful outcomes. Most incumbent solutions have focused only on parts of the complete content management process.

Early Generation Models

Substantial early focus, of what we call Generation 1 solutions, has been on enabling collaboration between the diverse stakeholders involved in the life cycle of these documents. Once these solutions matured and met the needs of life science organizations,

Generation 2 solutions made an attempt at reducing the cycle time for content creation and improving the quality of reviews using rigid business rules and templates. These solutions, however, were even more complex to implement as they were expected to function in a diverse set of contexts.

Today, thanks to the success of the early generation solutions, life science companies are demanding more as they look to continually improve their content management processes. The objective has now moved to extracting and identifying relevant content from unstructured sources. This process should ideally take into cognizance the different contexts that the content will be used in and should therefore be dynamic enough to adapt to the newer contexts of the future. At the same time, the focus is also on generating documents that are indistinguishable from expert authored documents.

Generation 3

The ability of a model to store content/documents from diverse sources with comprehensive metadata is the key enabler that defines Generation 3 solutions.

Optimizing content storage using machine-assisted classification, indexing, and tagging of documents and the content within these documents will ease the entire process of searching, authoring, and updating content.

The advancements in natural language understanding would allow for topic-wise deconstruction of document content not only through verbatim text analysis but also based on concept and context. The experience could be made even more seamless through a single user-interface linking databases across functions, providing ease of access to the consumers of content, without having to rummage through different databases.

Applications and Highlights

Clinical, Regulatory, and Safety documents draw in content/data from multiple sources, often reusing and repurposing content from preceding documents. It is an elaborate and time-consuming process that is prone to human errors and delays.

An intelligent content management tool should be able to reduce the time spent in creating these documents by facilitating reuse of core content and should assist in maintaining consistency across the documents for greater compliance.[4]

Indegene's ICA

The intelligent content authoring (ICA) model by Indegene is a Generation 3 solution, which stands out as an inclusive solution developed to encompass the contexts that define the industry. It has been built by medical experts for the industry to address a variety of use-cases. The model standardizes/ harmonizes content and more importantly drives reuse and repurpose of content. It lends the benefit of auto-generation of advance draft of a document that subsumes quality, compliance, and efficiency and expedites the time-to-regulator.

ICA addresses the gaps in efficiency using the latest developments in the field of technology.

The two key drivers in this process are the evolution of technology in NLP, natural language generation (NLG), machine learning (ML) classification models, and the evolution of data storage and processing technologies capable of managing highly connected dynamic semantic data structures that can process massive volumes of data at speed and store the data efficiently and at a reasonable cost.

While there have been technologies and models featuring content extraction, collaborative content management, and content archival, ICA goes beyond these traditional features. Indegene combines its deep medical expertise with these modern technologies to identify content and the contexts/key messages within the content. This content is then stored in a semantic structure that adapts to the nature of information being processed. AI and ML techniques make it possible for the solution to identify and include new semantic elements into the content structure. The solution uses NLG to generate downstream documents that are required by life science organizations. The solution enables the authoring of diverse sets of downstream documents by capitalizing on the built-in awareness of medical content document requirements, life cycle, and dependencies. It also provides a content catalog that supports natural language querying, making it easier for information seekers to access content.

The ICA Impact

The ICA model is able to store content from diverse and unstructured sources of data and is mutable according to the content context. The documents are, therefore, outcomes of a standardized authoring process and yet retain all the nuances of human authored documents. As a result, the workflow systematically becomes robust and efficient, increasing compliance and reducing the cycle time for document creation.

Indegene believes there are substantial gains in reducing time-to-agency and time-to-market while driving accuracy and efficiency improvements in the content authoring process. These gains are accomplished by leveraging modern technologies to intelligently curate, index, and automate reuse of content over the entire life cycle of medical content development. The applications of these technologies, with their impact on the content management life cycle, are invariably bound to change the way medical content is viewed and treated in the industry.

References

1. Bersin J. Digital leadership is not an optional part of being a CEO. *Harvard Bus Rev.* Dec 01, 2016. https://hbr.org/2016/12/digital-leadership-is-not-an-optional-part-of-being-a-ceo. Accessed July 08, 2018.
2. Innovating life sciences in the fourth industrial revolution: Embrace, build, grow. 2018 Global Life Sciences Outlook from Deloitte. Deloitte Web Site. https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-life-sciences-sector-outlook.html. Accessed July 08, 2018.
3. Hein T. Product lifecycle management for the pharmaceutical industry [An oracle white paper]. http://www.oracle.com/us/products/applications/agile/lifecycle-mgmt-pharmaceutical-bwp-070014.pdf. Accessed July 08, 2018.
4. Bernhardt SA. Improving document review practices in pharmaceutical companies. *J Bus Tech Comm.* 2003;17: 439. doi:10.1177/1050651903255345.