Reimagine Pharma Regulatory Operations using Knowledge Graphs

Rita Shah

Abstract: *Capturing regulatory product data from various sources and storing in relation - built Knowledge graphs driven by standards ontologies and taxonomies enables data structured, standardized, and connected. It helps faster contextualization and consistent reuse of data in a compliant manner across product life cycle. It also facilitates faster analytics to derive insights and capture issues for audit inspection readiness.*

Compliance to enhance patient safety and patient centricity is of high priority for every pharmaceutical company. Managing compliance related to data and content across the product life cycle in the Pharma world is extremely critical, complex, and challenging. Companies want to:

- Reduce efforts spent to manage data quality.
- Maintain compliance and cope with ever - changing regulatory requirements for early regulatory approvals.
- Be audit and inspection - ready and reduce audit findings.
- Manage evolving data, content across the product lifecycle, and keep regulatory product - related data ready for submission.
- Reduce efforts spent on data and content migration as part of their mergers and acquisitions and modernization of application landscapes.

Digital transformation in the pharma industry paved the way for new age technologies like Natural Language Process (NLP), Neural Language Generation (NLG), and Knowledge Graph (KG) to solve data and content - related challenges. Standard - aligned data models use Knowledge graph (KG) that interconnects various data entities in context. Linking ontologies and providing semantics facilitates the use of AI/ML to further bring agility to operations. The scalable Knowledge Graphs are transforming the way of working and exploring new findings in pharma and healthcare.

Pharmaceutical Regulatory Affairs, a heavy content driven industry can take advantage of Knowledge graphs to bring operational efficiency as illustrated in the scenarios:

1) **Knowledge synthesis of large documents: Natural Language Process (NLP)** enables extraction and mining of loads of increasing information. **NER (Named Entity Recognition)** recognizes entities, and a knowledge graph is a network of entities in a graph with built - in relations. For example, in the SmPC Labeling document, attributes like the product, marketing status, country, dosage, form, package, clinical particulars, side effects, etc. are extracted and stored in a graph with a relationship between them. The integrated knowledge graphs (KGs) help to better understand and explore the interconnected nature of a product and its behavior to make faster decisions. e. g., capturing every knowledge or data about the product (like - its clinical use, varied patient types, offline use, safety parameters, recommendations etc.) storing in prebuilt related knowledge graph holistically represent the product knowledge when prompted or queried.

2) **Support standardization:** The adoption of Knowledge graphs in regulation provides an advantage by enabling storage of data in a structured and standard manner driven by data models and showing links between the data attributes in context. It uses ontologies, taxonomies, and controlled vocabularies to capture the data consistently as per the standard and create a rich source of scalable knowledgebase for discoveries.[1] For example - a growing knowledge of the drug, its condition, and target population, side effect etc. will get captured as per standards using MedRA and SPOR dictionaries for standardization.

3) **Better and faster contextualization:** Knowledge synthesis can reduce reviews by integrating different data attributes, correlating them, and contextualizing the knowledge faster, helping respond to queries faster. For example - the relation of "paracetamol to fever and ulceration can understood with established relations and provide intelligence.

4) **Data integration:** Clinical, safety preclinical, manufacturing, and regulatory data on medicinal products stored in different databases can get integrated in a knowledge graph to give a hypergraph of product providing 360 views representing all product relations. It can help to find patterns in complex information[2].

5) **Real - time data change:** A knowledge graph provides a network of various entities linked together with established relations. In the connected data ecosystem, if any data entity gets updated or changed, it is detected fast along with its metadata in a linked knowledge graph. The newer version of data gets contextualized faster and enables real - time analysis and faster implementation. For Example - new safety data linked to the product gets contextualized faster, enabling fast implementation or new label creation and stay compliant[3].

6) **Faster content generation:** With the help of a standard aligned connected data graph, the data entities mapped to prefilled XML - tagged templates get auto - populated by making use of RPA. It can reduce the overall time for content generation for submissions for example - Auto populating smart submission Forms like electronic application forms (eAFs), simultaneous submissions concept can take advantage of such KG.

7) **Content reusability:** Knowledge graph stores both structured and unstructured data with its metadata. The

stored content with its metadata can be repurposed by multiple templates with mapped ontologies. For example - Protocol, CSR, SAP, PSUR, PBRER, CCDS to SmPC to PIL, etc. Auto population of data and content in various templates through single data source ensures consistency and compliance across various linked documents.

8) **Audit or inspection readiness:** With the attached legacy information, source, and track of changed versions, metadata of data or content provides faster and seamless ways to conduct audits and inspections and find issues faster. Two drug products manufactured at one manufacturing site, showing two IDs for the same site can be a potential compliance risk. The knowledge graphs provide a view of identified risk.

9) **Data migration:** KG can behave as a digital tool to enable faster data acquisition from various sources and optimization and standardization of data before migrating to a mapped target system. It enables the migration of data with its lineage, metadata, and dependencies. For example – connected product detail data sets, metadata, and source information will get migrated along. It will help to assess the cost, impact, and risk of migration of selected data. It saves time of validation activity.

Benefits

Knowledge graphs are game - changing to keep the industry well - equipped to manage incoming data loads and ensure the high quality and integrity of data. It will increase data confidence among stakeholders, reduce review cycle time, improve compliance, and bring agility in operations by managing data and content needed for faster submission and drug approvals for patients in need.

Ability of KG to unite data from various sources and contextualize it faster, offer 360 views of data and its relations, and provide faster and more intelligent analytics needed for better decision - making and strategy building. Data Digital connectivity using KG can enable tracking data changes from source to consumption, critical for the highly regulated pharma industry.

Knowledge graphs in conjunction with visualization tools, digital assistants, LLMs Chatbots, automated workflows can significantly bring rigor in regulatory operations.

For the future of the data - driven industry, Knowledge graphs (KGs) will have a significant role in capturing, maintaining, and scaling the knowledge and making it readily available for reuse, analysis and improving overall efficiency to bring drugs early to the market for waiting patients.

References

[1] https: //enterprise - knowledge. com/knowledge - graph - accelerator - for - data - standardization - and - environmental - social - governance - esg/
[2] https: //datavid. com/blog/knowledge - graph - visualization
[3] https: //medium. com/neo4j/knowledge - graphs - llms - real - time - graph - analytics - 89b392eaaa95