Smart Data Management: Unlocking the Potential of Regulatory Data

WHITE PAPER

Abstract

With the Industrial Revolution 4.0, pharma regulatory affairs (RA) has embarked on a transformation journey. The industry is looking forward to bringing in modernization on four fronts, namely, data, processes, applications, and infrastructure. In particular, data transformation takes an underscored importance, owing to its impactful nature on processes, applications, and infrastructure.

As data becomes more prevalent than ever, it creates its own set of challenges, such as maintaining data integrity, breaking away from silos, and standardizing metadata. This white paper focuses on our vision of how regulatory data can be managed smartly and efficiently in the future, so that it is purpose-centric and readily available for consumption.

A Data-Driven Overhaul

In the recent decade, there has been an increase in data proliferation, value focus, and scientific and technological advancements in the pharma regulatory affairs (RA) domain. Alongside, the sector also had an unmet need for agile regulatory data, lean and intelligent processes, integrated and connected systems, and a scalable cloud-based infrastructure. In a way, the Industrial Revolution 4.0 set the wheels in motion for a four-pronged transformation – data, processes, applications, and infrastructure. This paradigm shift from tangible documents to digital data, has made data a critical element and strategic asset. The data deluge that followed this digitalization gave rise to various challenges including:

- **Data changes:** With a single change in data, hundreds of documents get updated, and this requires high implementation cost and time.
- **Data inconsistency:** Data redundancy and inconsistent formats pose challenges to finding a single source of truth, thereby hampering data integrity.
- **Unstructured data:** The identification and extraction of relevant data buried in varied unstructured sources can be laborious and time consuming.
- **Disconnected data sources:** Lack of contextualization of data stored in different sources can lead to an oversight on the right action for the right insight.
- **Tracking of data lifecycles:** Data stored across the product lifecycle undergoes many transactions and the identification of the right source of event can become difficult.
- **Lack of standardization:** The lack of standard terminologies can put companies at risk during audit findings etc.

The Need

Regulatory data comprises product licensing and other related information, such as clinical development, manufacturing, safety, labeling, supply, marketing etc. Throughout the product lifecycle, data undergoes various changes in these categories, and a single change can cascade into product information and terms getting updated. Today, product information is not connected enough across its lifecycle to get real-time visibility of updates that enable accurate intelligence and decision making for the early marketing and access of the product. The need is to have a holistic, connected, automated, and intelligent data management layer that manages the integrity and has the visibility of data exchange across the lifecycle.

The Proposed Solution

A data foundational layer can be a solution for managing the underlying data, its correctness, completeness, and accuracy. The proposed Regulatory Data Foundation Layer is based on scalable graph systems that represent regulatory data as nodes linked to one another by edges with a set of properties that delineate the relationship between the nodes. It can synthesize the knowledge of product lifecycles and simplify data capture and access, profile, connect, and harmonize data from varied sources. This layer can also lead to an increased flexibility in accommodating increasing data loads with minimum impact.

Solution Concept

The fundamental ingredients of the data fabric are:

- Reference domain model-driven Knowledge Graphs, leveraging ontologies and common vocabularies for common understanding and exchange
- Semantics for building data relationships and labels
- Metadata to act as the knowledge navigator for managing data flexibly with data lineage

Knowledge Graphs can synthesize and connect entities across systems and enterprises dynamically and generate a graph that can be used to derive intelligence for business value. It stores the data in an effective manner, thus helping with faster exploration, contextualization, and querying from different systems. Knowledge graphs can be of immense value as the data in disparate systems – regulatory, clinical, safety, manufacturing, supplies, labeling, packaging etc. – can be presented in a connected and holistic graphic view providing 360-degree product view.

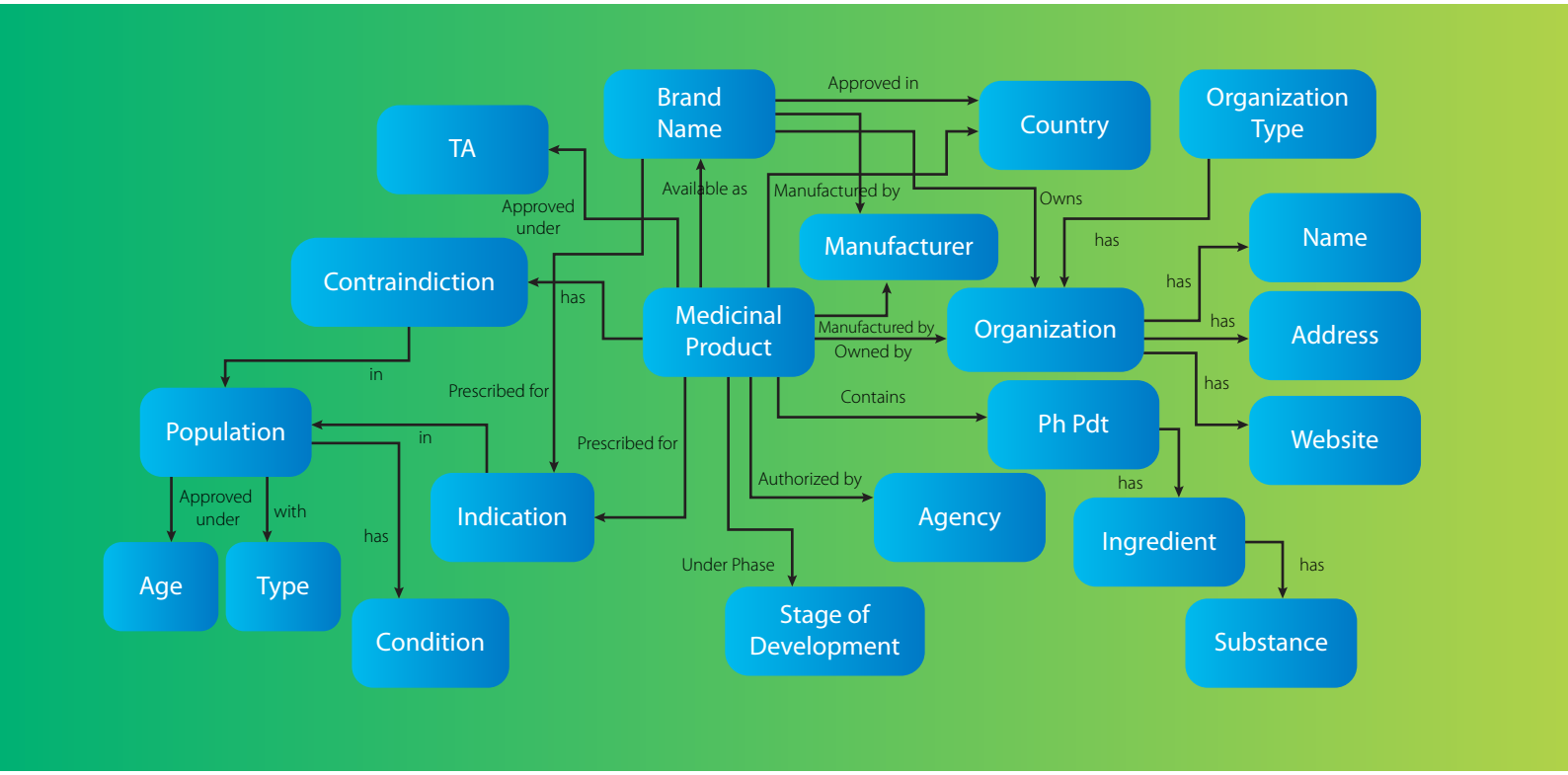

Figure 1: Connected Product Data in Knowledge Graph

Understanding Knowledge Graphs in the Regulatory Context

Building a regulatory knowledge graph is essentially extracting product data entities like medicinal product, brand, manufacturer, country, document, application, submission etc., their attributes and values like, name, type, identifier, address, country, document application submission, etc., and analyzing the linkages between them. The semantics on the top gives meaning, builds relations, and contextualizes the data in a complex environment.

With the adoption of standards like IDMP, FHIR, and PQ/CMC, and the usage of SPOR ontologies related to Regulatory Information Management (RIM), development of a baseline data/common information model can enrich and transform the regulatory data management process. Any data modification such as brand addition, safety profile, or country authorization enhances the organization of the data in a graph and defines the output data. An advantage of this is that it enables the granularity and independence of data while at the same time it can be assembled for a specific context or extracted without losing the context.

By linking various data domains and forming complex hyper Knowledge Graphs, like products linked to therapeutic area, disease, population, safety, conditions etc., the regulatory ecosystem can be aligned and understood in a better way. This also allows the content generators to work independently and yet collaborate for rapid share of data.

Emerging technologies like GraphdB/Neo4J with algorithms and visualization can facilitate connected data and create meaningful, purpose-driven insights. Such insights can also lead to the exploration of more opportunities, such as Regulatory Intelligence as it can synthesize the externally available knowledge by sensing and linking with the entities of interest.

Technologies like NLP, AI, and ML can also facilitate the use of ontologies, common vocabularies, and acronyms to auto-standardize and enhance contextualization for achieving accuracy, credibility, and faster intelligent decisions. ML algorithms can bring out faster inferences and increase efficiency, through various calculations run on data and their subsets, like CMC analytical data, which is otherwise human intensive.

Benefits

Knowledge Graph-driven data foundation layers can enable:

1. **Agility:** With connected product data and their build relationships, data is available, visible as graph, accessible, and ready to be consumed through faster query and search capabilities, hence, enabling faster submissions.
2. **Data integrity:** Any data change can show the impact on its connected data. Hence less rework and reduction in cycle time is guaranteed with ensured data quality and integrity.
3. **Audit and inspection readiness:** For audit inspection preparedness, data change and legacy of connected data can easily be captured with established and broken connections and metadata.
4. **Scalability:** Flexibility to add more information and linking it in the model has an advantage of absorbing and transforming data through the stages of onboarding, modeling, blending, and accessing to create a fully operational enterprise data fabric.
5. **Flexible and reusable:** Standardized data stored in knowledge graphs, can be reused at various points across the product lifecycle, thereby making data reusable, reducing cycle time, increasing quality and compliance.
6. **Visibility/traceability:** With 360-degree product view, any data can be found through better visualization to accelerate decision-making.

Conclusion

Standardized RIM and enterprise data fabric layers are the foundation for building trust and confidence about the quality and integrity of regulatory data across product lifecycles. Data-centric shifts underpinned by graphs, semantics, and metadata, can account for the much-needed agility in data management, thus opening up ventures for analytics to harness and unlock the potential of regulatory data.

About The Author

Rita Shah

Rita Shah is a Clinical Research professional with over 21 years of experience in academic research in the Life Sciences and Pharma domain. She is actively involved in research and development of innovative conceptual solutions leveraging automation technologies. Rita has a PhD in Organic Chemistry from Council of Scientific & Industrial Research (CSIR) and has more than a dozen research publications in peer reviewed journals. She has also worked as a Postdoctoral Fellow and Assistant Research Professor (Molecular Biology) at the Indiana University School of Medicine, USA.

Contact

Visit the Life Sciences & Healthcare page on www.tcs.com

Email: LifeSciences.Connect@tcs.com

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that has been partnering with many of the world's largest businesses in their transformation journeys for over 50 years. TCS offers a consulting-led, cognitive powered, integrated portfolio of business, technology and engineering services and solutions. This is delivered through its unique Location Independent Agile™ delivery model, recognized as a benchmark of excellence in software development.

A part of the Tata group, India's largest multinational business group, TCS has over 453,000 of the world's best-trained consultants in 46 countries. The company generated consolidated revenues of US $22 billion in the fiscal year ended March 31, 2020, and is listed on the BSE (formerly Bombay Stock Exchange) and the NSE (National Stock Exchange) in India. TCS' proactive stance on climate change and award-winning work with communities across the world have earned it a place in leading sustainability indices such as the Dow Jones Sustainability Index (DJSI), MSCI Global Sustainability Index and the FTSE4Good Emerging Index.

For more information, visit us at **www.tcs.com**

All content / information present here is the exclusive property of Tata Consultancy Services Limited (TCS). The content / information contained here is correct at the time of publishing. No material from here may be copied, modified, reproduced, republished, uploaded, transmitted, posted or distributed in any form without prior written permission from TCS. Unauthorized use of the content / information appearing here may violate copyright, trademark and other applicable laws, and could result in criminal or civil penalties.
Copyright © 2021 Tata Consultancy Services Limited

Corporate Marketing | Design Services | M | 03 | 21