DATA GLOSSARY AS AN ASSET TO OPTIMIZE DATAOPS IN PHARMACEUTICAL INDUSTRY

Dissertation Submitted to

The IIHMR University, Jaipur

for the partial fulfillment for the award of the

degree of Master of Business Administration

(MBA)

in

Hospital and Health Management

by

Dr Pravie

Under the Supervision and Guidance of

Dr Dhirendra Kumar

2022

DATA GLOSSARY AS AN ASSET TO OPTIMIZE DATAOPS IN PHARMACEUTICAL INDUSTRY

Dissertation Submitted to

The IIHMR University, Jaipur

for the partial fulfillment for the award of the

degree of Master of Business Administration

(MBA)

in

Hospital and Health Management

by

Dr Pravie

Under the Supervision and Guidance of

Dr Dhirendra Kumar

2022

DECLARATION BY STUDENT

I hereby declare that this dissertation titled Data Glossary as an Asset to Optimize DataOps in Pharmaceutical Industry is the bonafide record of my original research work. I declare that it has not been submitted to any other university or institution for the award of any degree or diploma. Information derived from the published or unpublished work of others has been duly acknowledged in the text.

Dr Pravie

Date: 24.06.2022

CERTIFICATE

This is to certify that Dr Pravie in partial fulfillment of the requirements for the award of the degree of MBA (Hospital and Health Management) from the IIHMR University, Jaipur has successfully completed her/his internship at ZS Associates, New Delhi during February 07, 2022, to June 18, 2022.

Place:
Date:

Preceptor/Head of the Organization
Name of the organization

CERTIFICATE

This is to certify that dissertation titled Data Glossary as an Asset to Optimize DataOps in Pharmaceutical Industry is a record of the research work undertaken by Dr Pravie in partial fulfilment of the requirements for the award of the degree of MBA (Hospital and Health Management) from the IIHMR University, Jaipur under my guidance and supervision and her approach to the research study has been sincere, scientific, and analytical.

Dr Dhirendra Kumar	Dr (Col) Mahender Kumar
Faculty Guide	School Dean
IIHMR University, Jaipur	IIHMR University, Jaipur
Date	Date

ABSTRACT

In Pharmaceutical industry, the data is growing faster than ever before. The vast amount of data generated and collected by a multitude of stakeholders comes in so many different forms – labs, genomics, pharmaceutical R&D, clinical, wearables, SDOH, etc. Now-a-days Life Science companies are generating more data across many functions and are looking for ways to derive fact-based and informed decisions. With the increasing data, regulations across the world are also evolving. Hence, Life Science companies need to balance business opportunities and regulatory compliance. To cater these concerns, there is a need of data management as a service. Data Glossary can act as a crucial part of the data management framework. It can help in effective and efficient data stewardship and cataloguing. It helps to define industry specific concepts and helps in standardization and communication hence optimizing data operations within an organization. Data glossaries acts as one of the many key resources that data governance program produce to promote understanding and shared usage of language across an enterprise. A robust data glossary defines key terms and concepts based on a company-wide consensus-and establishes relationships between those terms and definitions that the whole organizations as well as the industry and regulatory bodies can comprehend, use in daily operations, and report on regulatory compliance. data glossary will mark as the starting point of any effective formalized data governance program resulting in efficient and effective data operations within the organization contributing to reduced time and cost of newer therapeutics to market.

ACKNOWLEDGEMENT

I would like to convey my sincere appreciation to ZS Associates, New Delhi for giving me this wonderful internship opportunity in their esteemed organization. I would like to thank Dr Mani Sharma, my Project Manager for assigning me on this project and providing her invaluable guidance throughout this internship. Thank you for trusting me with the responsibilities which were bestowed upon me during my internship tenure.

I would like to extend my gratitude to Mrs Shalvi Shethi, Consultant, ZS Associates, under whom I did this research. The completion of this project could not have been accomplished without her constant guidance and support.

I also wish to hold out my sincere gratitude to Dr Dhirendra Kumar, Professor, IIHMR University, Jaipur, my mentor, for his expert guidance, critical supervision, patience, motivation, and immense knowledge which helped me to come up with this report.

Last of all, I am thankful to all my colleagues for their continuous encouragement during the internship. I would like to thank all who in one way or other helped me in completion of my report.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Acronym/Abbreviation	Definition
DataOps	Data Operations
R&D	Research and Development
AI	Artificial Intelligence
ML	Machine Learning
FDA	Food and Drug Administration
CAGR	Compound Annual Growth Rate
PHI	Personal Health Information
DAMA	Data Management Association
DMBOK	Data Management Body of Knowledge
KDE	Key Data Element
CRO	Clinical Research Organization
FDA	Food and Drug Administration
PMDA	Pharmaceuticals and Medical Devices Agency
CDISC	Clinical Data Interchange Standards Consortium
EHR	Electronic Health Record
EVS	Enterprise Vocabulary Service
NCI	National Cancer Institute
NCIt	NCI Thesaurus
SDTM	Standard Data Tabulation Model
CDASH	Clinical Data Acquisition Standards Harmonization
SEND	Standard for Exchange of Nonclinical Data
PRM	Protocol Representation Model
ADaM	Analysis Data Model
WHO	World Health Organization
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
LOINC	Logical Observation Identifiers Names and Codes

Chapter 1: Introduction

1.1 Background

Drug development process is still an arduous, complicated, expensive, and time-taking even after years of efforts. Years of studies and research along with extensive data collection is needed for pharmaceutical research and development processes. It takes an average of 12-15 years for a drug to make its way through the development and review process at the FDA with a colossal amount of approximately USD 1 billion.

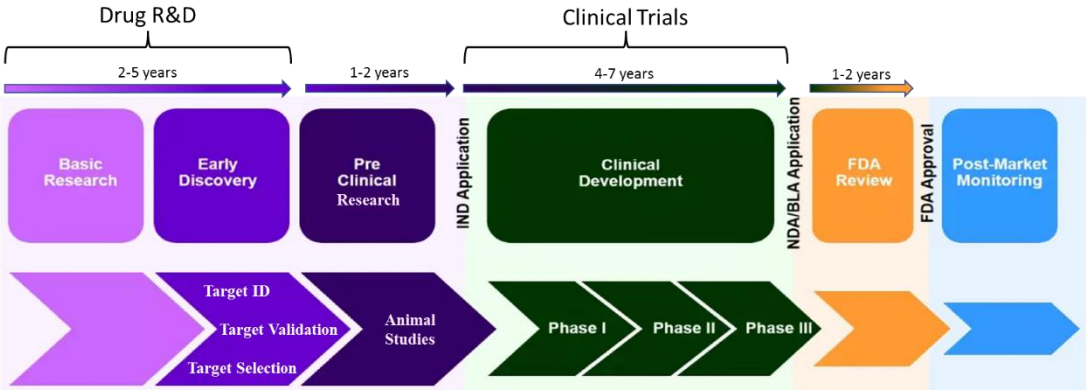

Figure 1.1: Timelines for Drug Development Phases

Even the pharmaceutical sector is not exempt from the issue of the growing amount of data that needs to be captured and monitored. Huge amounts of data are often accompanied by concerning data quality issues. The entire process necessitates continuous data quality monitoring because even a minor error may have serious ramifications, including risks to patients' health and lives as well as legal repercussions and reputational harm. It has been observed that only a small proportion of pharmaceutical corporations consistently track and attempt to gauge the financial effects of poor-quality data. The market study conducted by Gartner also supports this situation. According to Gartner's Data Quality Market Survey, January 18, nearly 60% of companies doesn't assess their annual financial cost of poor-quality data. Another survey conducted by World Class Regulatory Information Management (RIM) reflected that 62% of the 65 Life Sciences firms surveyed acknowledged that their data standards and data governance practices are inadequate.

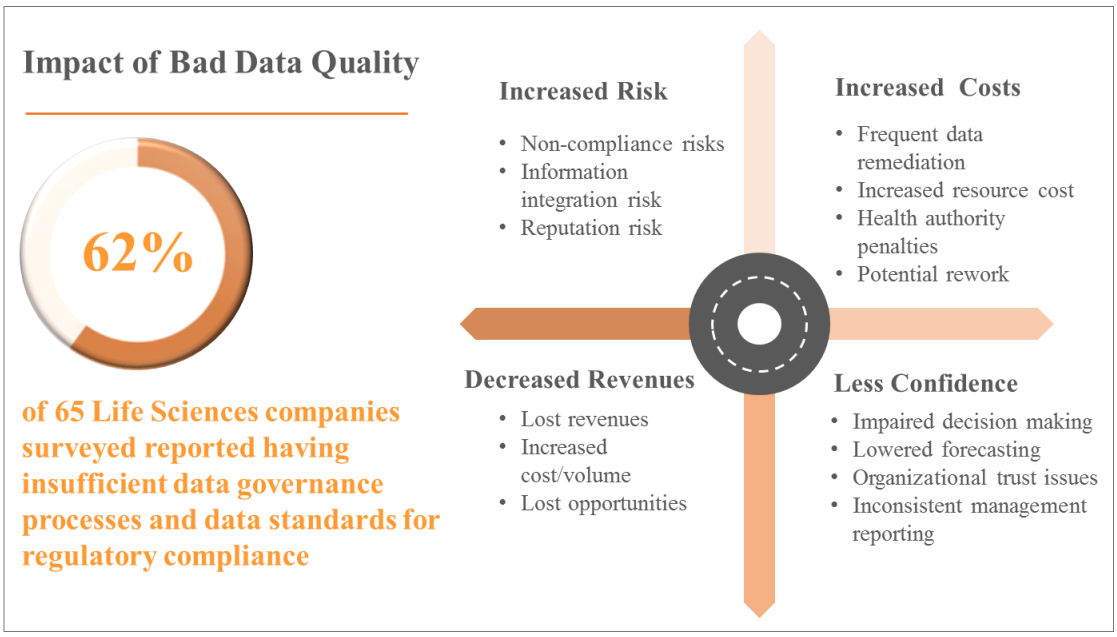

Figure 1.2: Impact of Poor Data Quality; Source: Pharmaceutical Commerce

DataOps can aid in addressing these challenges of poor-quality data and governance. There is a need of building and implementing intelligent assets to optimize DataOps.

1.2 Aim and Objective of the research

The purpose of conducting this research is to understand how we can build Data Glossaries as an asset that can help the pharmaceutical companies to optimize their data operations to rapidly bring drugs and therapeutics to market that too in reduced costs. To get answers to this question we first need to understand what data operations is, why there is a need of data operations, what are the data related challenges that pharmaceutical industry is facing, how can data operations cater to these challenges and most importantly how data glossaries can play their part to address these challenges resulting in optimized data operations.

1.3 Introduction to Data Operations

Data Operations, also known as DataOps, is an orchestration of people, processes and technology tools that allows consistent, automated, and secure data management. Data Operations can be defined as "an organization-wide data management practice that controls the flow of data from source to value, with the aim of speeding up the process of deriving value from data". The goal of Data Operations is to be a cross-functional approach to dealing with data regarding its acquisition/onboarding, storage, processing, quality monitoring, improvement, and delivery to the end users. The fundamental purpose of Data Operations is to make the teams across various verticals of the organization capable enough to oversee the critical business operations, analyse the value of each one

in order to eliminate data silos, and centralise those processes. The ultimate outcome of DataOps is ‘scalable, repetitive, and foreseeable’ data flows for various personas interacting with data on a day-to-day basis such as data engineers, data scientists, and business users. Data Operations across the entire enterprise results in:

- improved quality of data that is ready to be consumed by the end users
- rapid and secure migration of data to the cloud or other destinations
- reliable flow of data throughout data lifecycle
- supports integration of data science and machine learning for improved and fact-based business decision making
- establishment of standardized data that enables compliance and standardized data security policies.

1.4 Need for Data Operations

In the current environment, companies require a new approach to handle the complexity of ever-increasing volume, velocity, and variety of data, making it a must for organizations to manage an infinite number of data flows. The data is being produced in a variety of formats with diverse frequencies and there is no sign of data heterogeneity being slowed. These factors are fuelling the demand for quicker data movements. Data Operations helps to keep track of the data, secure and control data access, and ensure usability and discoverability throughout the data journey.

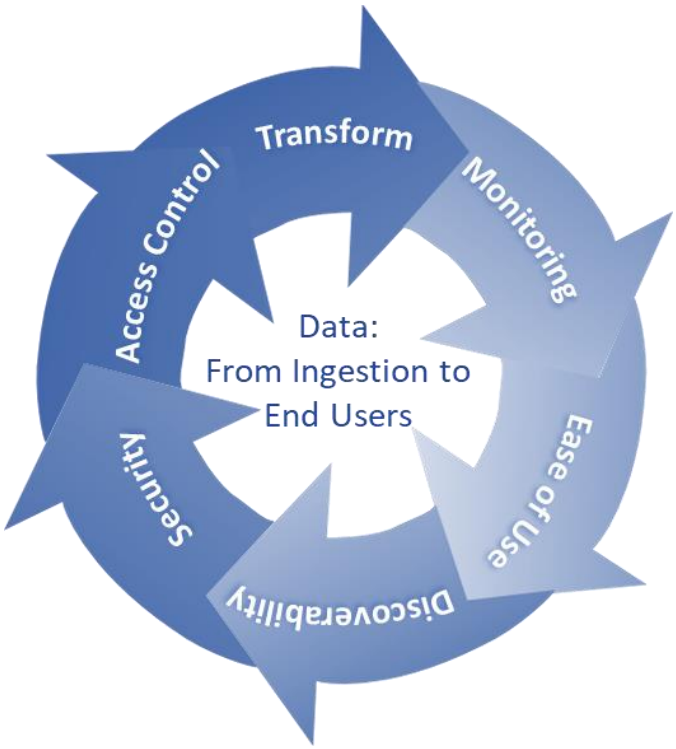

Figure 1.3: The Data Journey

- Monitoring – tracking the operational data flows and receiving prompt alerts whenever a change or anomaly in data is discovered.
- Security and Access Control – Access provisioning and monitoring is mandatory whenever the data is onboarded from third party data vendors along with access management at attribute-level leading to controlled data management.
- Ease of use – Data accessibility is crucial to all the personas of business to drive value from it.
- Discoverability – To make data democratized, discoverability and data mapping is important.

1.5 Data related challenges in Pharmaceutical Industry

One of the most growth-oriented industries globally is the pharmaceutical sector. By 2023, the global pharmaceutical market is anticipated to reach USD 1.5 trillion. The domestic market is anticipated to rise three time during the following ten years to come, according to the Indian Economic Survey, 2021. In light of this, the in-country pharmaceutical market in India is now estimated at USD 41 billion, and it is projected to increase to USD 65 billion by 2024 with a further rise of USD 120 to 130 billion by 2030. Over the next five years India's expenditure on medicine is expected to rise by 9% to 12%, placing it among the top 10 nations worldwide in terms of medicine spending.

Now-a-days, the pharma sector is dealing with several unique issues including

- pressure to lower costs to bring a new drug to market,
- speed up R&D processes to quickly bring new drugs and treatments to market,
- significant expenditure in R&D,
- reputational losses due to counterfeit drugs,
- stricter regulatory and compliance requirements and
- demand for improved and more effective data management and governance

Talking about the R&D, pharma businesses are vying for victory in today's dynamic and rapidly shifting competitive landscape by improving their performance without raising their overall operating costs. Pharmaceutical firms innovating quickly to maintain a competitive edge and taking advantage of market opportunities as a result of the growth of breakthrough technologies like AI-ML, automating processes and data analytics is the need of the hour.

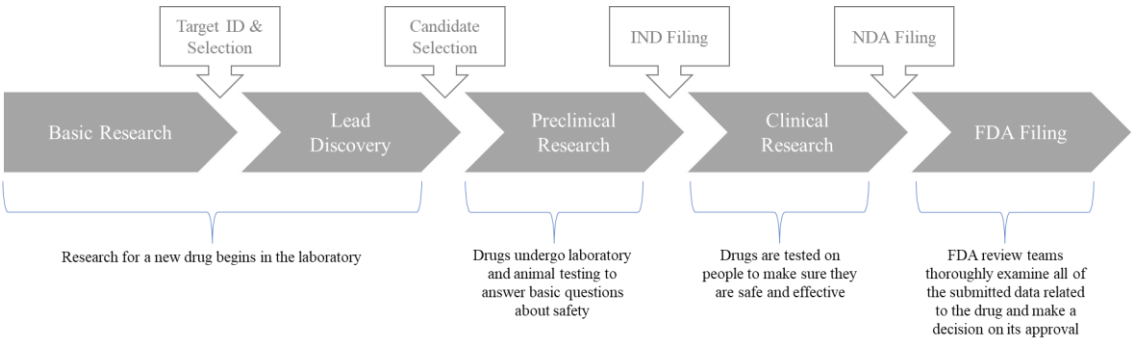

Figure 1.4 The Drug Development Process

A huge quantity of complex data is churned out during this whole process from basic research and discovery to clinical trials by the pharma or life sciences companies. This data is basically of drug components, diseases, patient information, test results, laboratory experiment findings, drug constitution and dose, FDA regulatory requirements, etc. The complexity of conducting clinical trials has increased as more vendors and service providers get involved. The following are some organizational challenges encountered by pharma/life sciences companies:

- siloed clinical data that too in diverse systems and formats,
- frequently changing ontology standards
- inability to utilize clinical data to gain insights and make informed decisions,
- lack of reproducible procedures that diminish the potential for efficiency gains across clinical trials.

It has become a challenge for the pharmaceutical businesses to make use of this large amounts of data that is being generated and stored as databases in various formats. It is crucial for pharma organizations to analyse and report the clinical data from multiple sources and make informed decisions based on more accurate, timely and quality data that also supports regulatory submissions.

To address these challenges, pharma companies are inclining towards digital technologies to integrate valuable insights from various data sources, elevating the productivity of clinical trials, and boosting the quantity and quality of data collected in trials. A report from GlobeNewswire predicts that the global life science analytics market is forecasted to elevate from $17.2 billion in 2021 to $27.59 billion in 2028 at a CAGR of 6.5% between 2022-2028. Using analytics to enhance the clinical trials efficiency by utilizing data would be a huge benefit for pharma companies.

However, with increasing technology comes stricter security scrutiny as data is being shared among multiple stakeholders. As the data sharing increases, access to sensitive data like patient's Personal Health Information (PHI) from clinical trials or past medical records pose a threat of confidential information being breached. Hence, pharmaceutical and life sciences companies must be compliant and responsive to regulatory guidelines.

Standardization of data to accelerate R&D efficiency while maintaining compliance to regulations with thorough security, audit trails and data mapping is the pressing need. Lack of data standards to oversee and regulate master data sharing among multiple stakeholders continues to be a persistent challenge for pharma companies.

1.6 DataOps catering to these challenges

Gartner defines DataOps as "a collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization." To address the aforementioned challenges an organization needs to have well established and concrete data management practices in place to ensure efficient and effective data operations. 'Data management' is an umbrella term that covers various functional areas of how an organization uses and manages its data. DAMA has developed a DMBOK Framework which displays a set of 11 data functional areas in a structured form.

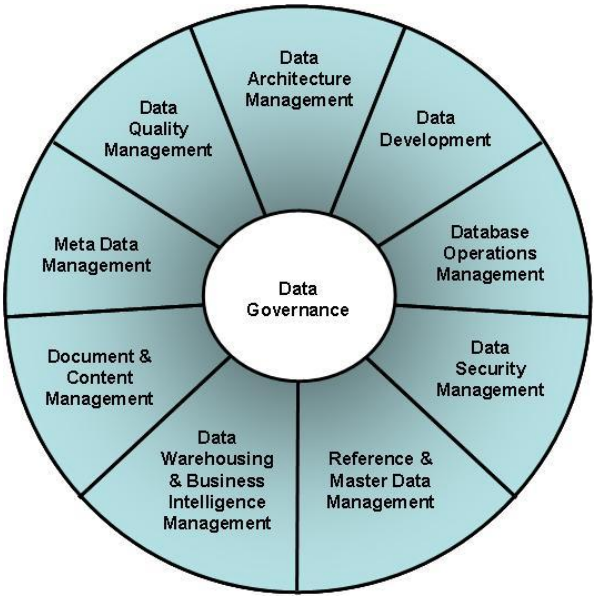

Figure. 1.5: DAMA-DMBOK Functional Framework

An organization requires various assets and accelerator to optimize these data management functional areas. The main aim of this research is to highlight how data glossaries can be built as an asset to increase the efficiency of some of the aspects of data

management ultimately leading to optimization of data operations. A data glossary is a fundamental feature of modern data management, crucial for defining a consistent technical vocabulary for an organization.

1.7 Understanding the Data Glossary

A data glossary is a very simple and straightforward yet a very crucial concept. A glossary is a collection of words with their definitions that provides context and aids in the organization of knowledge. A data glossary fulfils this exact function for all the data in an organization's data ecosystem. It is an easy-to-search compilation of all the terms that defines multiple data types and their key characteristics used within the organization as well as the industry so that everyone may refer and utilize the standard definitions and impart harmony around the system's guiding concepts. Well-maintained data glossaries make sure to improve the legibility of existing data and empower people both internal and external to better understand the data warehouse or database they are looking at. Additionally, it promotes consistency as everyone shares a similar understanding of the data terms and speaking the same language. A data glossary, in essence, serves as a sole source of veracity for all the data terminology associated with that specific area or domain such as clinical, biomarkers, genomics, proteomics data, etc.

1.8 Benefits of having a Data Glossary

Data glossaries can have a strong impact on business performance as it has several benefits that include:

- Single source of truth – standardized definitions of data terms within the organization as well as industry enables collaboration.
- Increased data visibility – Anybody that has access to these data glossaries is able to comprehend the data without consulting the data owners or stewards.
- Ease of understanding – As people, internal or external of the organization can easily understand the data with the help of data glossaries, it becomes easier for them to make their own fact-based and informed business inferences making the organization more data-driven.

Establishing and maintaining a data glossary within the organization can aid businesses to:

- Evaluate its KDEs – There is always room for interpretation in language. It is possible that a data term might mean different to different people if a standardized definition is not in place. This acts as a huge hurdle as using different definitions

can lead to having huge inconsistencies in the final analysis. For any business, accurately measing the key data elements and recognizing the growth drivers is the very first step to a successful data journey

- Make the analyses credible – Trust can easily fade if the employees of the organization are not aligned on a shared understanding of data terms, squandering all the efforts put into analysing the data to make crucial business decisions. A data glossary promotes enterprise data trust leading to productivity and efficiency of business.
- Endorse Data Governance – A solid foundation can be laid with the help of a well-maintained data glossary for an effective data governance program in an organization. The organization gains more understanding when precise standards for data terminologies and definitions are established. Data users can use the data glossary terms to design rules and govern access policies with the aid of this well-organized knowledge base.

1.9 Business need of Data Glossaries in Pharmaceutical Industry

The need of data glossaries for optimizing data operations in pharmaceutical industry can be understood if we investigate the challenges faced by these companies. These firms may integrate their data silos and gain insights more quickly that will allow them to identify, collaborate and derive value from data wherever it may be by using data glossaries. These challenges can be enumerated as follows:

i. Poor Quality of Data – To perform research and development, clinical trials, and run daily operations, pharma companies must analyse enormous amounts of data in multiple formats. The data to be ingested is unorganised, fragmented, unstructured, missing values, inconsistent with potential biases. Organizations are compelled to spend their valuable time onboarding, cleaning and arranging the data.

ii. Poor Performance – To obtain timely and relevant insights, pharma or life sciences firms must process vast quantities of data easily and quickly. It is very crucial for drug development to have efficient consolidation, authentication and mining of data generated from clinical trials.

iii. Lack of Collaboration – Decision-making is enhanced by having access to various sources of data. Masses of sensitive data are handled by pharma companies, necessitating frequent back-and-forth collaboration. Throughout the drug development process, volumes of patient records and laboratory results are shared

across several department and partners. Building diverse, legacy systems is therefore required to enable quick, simple and secure data flow.

iv. Poor Compliance – For the usage, storage and disposal of sensitive data, pharmaceutical companies must adhere to stringent regulatory body rules, government regulations as well as quality guidelines. This necessitates a wide range of security approvals and controls that allow for monitored and controlled access to all data.

1.10 Data Glossary as a Solution

In order to address these concerns, data glossaries can be beneficial in

- Standardized data onboarding and data transfers – standardized and structured data is easy to process and flow
- Data Safeguarding – Data glossaries can assist organizations in handling their data more efficiently and avoid legal and regulatory violations.
- Cost Reduction – By eliminating data silos, companies can better understand their data ecosystem, uncover redundancies, streamline processes resulting in decreased costs.
- Risk Mitigation - Data glossaries can help in identification of poor quality of data if the data is not at par with the defined standards hence highlighting the potential issues and preventing crisis.
- Regulate change – By ensuring operational consistency, minimizing risks and reducing costs, comprehensive data glossaries enable enterprises in their attempts to ease the pressure of regulations.

Chapter 2: Literature Review:

Sr. No.	Research Title	Authors	Year
01	Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey	Rhonda Facile; Erin Elizabeth Muhlbradt; Mengchun Gong; Qingna Li; Vaishali Popat; Frank Pétavy; Ronald Cornet; Yaoping Ruan; Daisuke Koide; Toshiki I Saito; Sam Hume; Frank Rockhold; Wenjun Bao; Sue Dubman; Barbara Jauregui Wurst	2021
02	Towards a Single Data Exchange Standard for Use in Healthcare and in Clinical Research	Jozef Aerts	2018
03	CDISC SHARE, a Global Cloud-based Resource of	Samuel Hume, Anthony Chow, Julie Evans, Frederik Malfait, Julie	2018

	Machine-Readable CDISC Standards for Clinical and Translational Research	Chason, J. Darcy Wold, Wayne Kubick, and Lauren B. Becnel	
04	Current applications and future directions for the CDISC Operational Data Model standard: A methodological review	Sam Hume, Jozef Aerts, Surendra Sarnikar, Vojtech Huser	2016
05	Clinical Data Acquisition Standards Harmonization Importance and Benefits in Clinical Data Management	Jagadeeswara Rao Gaddale	2015
06	Data Management in Clinical Research: An Overview	Binny Krishnankutty, Shantala Bellary, Naveen B.R. Kumar, and Latha S. Moodahadu	2012
07	Data Management Challenge in Pharmaceutical Industry	Muharman Lubis, Mira Kartiwi	2012

Chapter 3: Methodology

The main objective of this research is to determine how can we build a standardized data glossary that is specific to pharmaceutical industry to optimize the data operations. It is better to follow the existing industrial standards rather than building the data glossary from scratch or creating new ones. This will make the data glossaries standardized throughout the organization as well as industry no matter the data type or a new use case. We conducted exhaustive secondary research on a total of 43 companies that in one way or other constitutes pharmaceutical industry such as, top 15 big pharmaceutical companies (based on their R&D spending, 2021), top 10 CROs, 5 big consulting companies, 5 big Biotech companies, and few government organizations to identify the answers to following research questions:

i. Are there any type of standards that companies are either using or offering services that utilize these standards to standardize their data operations?
ii. If they are using standards, then what are those standards.

The outcome of the research is displayed in the following table:

Sr. No.	Name of the Company	Company Type	Any Standards for Data Operations	Type of Standards
1	AbbVie	Pharmaceutical	Yes	CDISC Standards

2	Accenture	Consulting	Yes	CDISC Standards
3	Amgen	Pharmaceutical	Yes	CDISC Standards
4	AstraZeneca	Pharmaceutical	Yes	CDISC Standards
5	Bayer Healthcare	Pharmaceutical	Yes	CDISC Standards
6	Biogen	BioTech	Yes	CDISC Standards
7	Boehringer Ingelheim Pharmaceutical	Pharmaceutical	Yes	CDISC Standards
8	Bristol-Myers Squibb	Pharmaceutical	Yes	CDISC Standards
9	Cognizant	Consulting	Yes	SNOMED CT, LOINC, RxNorm, ICD-10
10	Deloitte	Consulting	Yes	CDISC Standards
11	Eli Lily	Pharmaceutical	Yes	CDISC Standards
12	EY	Consulting	Yes	HL7 FHIR, SNOMED CT, LOINC
13	F. Hoffmann-La Roche	Pharmaceutical	Yes	CDISC Standards
14	Food & Drug Administration	Government	Yes	CDISC Standards
15	Genmab	BioTech	Yes	CDISC Standards
16	Gilead Sciences	Pharmaceutical	Yes	CDISC Standards

17	GlaxoSmithKline	Pharmaceutical	Yes	CDISC Standards
18	IBM (IBM Watson)	Consulting	Yes	CDISC Standards
19	ICON Clinical Research	CRO	Yes	CDISC Standards
20	IQVIA	CRO	Yes	CDISC Standards
21	Janssen (J&J)	Pharmaceutical	Yes	CDISC Standards
22	KCR	CRO	Yes	CDISC Standards
23	KPMG	Consulting	Yes	SNOMED CT
24	Labcorp (Covance)	CRO	Yes	CDISC Standards
25	Medpace Inc	CRO	Yes	CDISC Standards
26	Merck & Company	Pharmaceutical	Yes	CDISC Standards
27	National Cancer Institute	Government	Yes	CDISC Standards
28	National Institute of Allergy & Infectious Diseases	Government	Yes	CDISC Standards
29	National Library of Medicine	Government	Yes	CDISC Standards
30	Novartis Pharmaceutical Corporation	Pharmaceutical	Yes	CDISC Standards
31	Novo Nordisk	BioTech	Yes	CDISC Standards
32	Parexel	CRO	Yes	CDISC Standards

33	Pfizer	Pharmaceutical	Yes	CDISC Standards
34	PPD	CRO	Yes	CDISC Standards
35	Pharmaceuticals and Medical Devices Agency	Government	Yes	CDISC Standards
36	PSI CRO AG	CRO	Yes	CDISC Standards
37	Sanofi	Pharmaceutical	Yes	CDISC Standards
38	Syneos Health Inc	CRO	Yes	CDISC Standards
39	Takeda	Pharmaceutical	Yes	CDISC Standards
40	TCS Life Sciences ADD	Consulting	Yes	CDISC Standards
41	US Centres for Disease Control (CDC)	Government	Yes	CDISC Standards
42	Vertex Pharmaceuticals	BioTech	Yes	CDISC Standards
43	WuXi Clinical Development, Inc.	BioTech	Yes	CDISC Standards

Table 3.1: Standards followed by the Pharmaceutical Industry

Chapter 4: Results and Discussion

Out of 43 organizations, 40 of them are using CDISC standards while 3 are using other standards such as SNOMED CT, LOINC, etc. From the above outcomes it can be concluded that CDISC Standards are becoming the universal language for information exchange in the pharmaceutical industry R&D processes. Majority of the organizations are now adopting CDISC Standards to homogenize their clinical research data. Building data glossaries based on these standards will impart a significant impact on streamlining and optimizing the data operations.

CDISC elaborated as the Clinical Data Interchange Standards Consortium is an open, non-profit organization that creates and supports international standards for data to enhance the quality and interoperability of healthcare and medical research. CDISC Standards are extensively utilized for planning of clinical studies and collection, tabulation, analysis, and submission of data to various regulatory authorities globally such as FDA, PMDA, etc. These standards can aid in reinforcing the acquisition, sharing, submission, and archive of clinical as well as non-clinical research data and metadata. In order to harmonize clinical data and streamline research procedures from protocol to interpretation and documenting, and the usage of EHRs for speed up the acquisition of research data that is high in quality, CDISC fosters constructive collaboration among industry stakeholders. The CDISC standards can greatly speed up the production of safer and more effective therapeutics as well as a learning healthcare system while also reducing the time and expense of research. The vision of CDISC is to advance higher grade medical research to improve patient care and safety.

To create and foster controlled terminology for all CDISC core standards, CDISC collaborates with NCI Enterprise Vocabulary Services (EVS). The research and healthcare communities participate widely in the rigorous process of content development and public evaluation that goes into CDISC Terminology. As part of NCI Thesaurus (NCIt), EVS updates and disseminates CDISC Controlled Terminology.

CDISC has developed controlled terminology for its following foundational standards:

- Study Data Tabulation Model – a world-wide standard for clinical research data approved by the FDA as a standard electronic submission format. SDTM terminology consists of vocabulary for CDISC Questionnaires, Ratings and Scales.
- Clinical Data Acquisition Standards Harmonization – The CDASH Terminology is a subset of SDTM Terminology. In order to enable unambiguous traceability of submission data into the SDTM and increase visibility for regulators and those who execute data review, CDASH establishes a standard approach to gather data consistently across studies and sponsors.
- Analysis Data Model – ADaM controlled terminology defines standards supporting efficient generation, replication, review, and submission of analysis results from clinical trial data.
- Standard for the Exchange of Nonclinical Data – Controlled terminology for SEND defines the standards for collection and presentation of nonclinical data in

a consistent format. SEND is an implementation of the SDTM standard for nonclinical studies.

- Protocol Representation Model (PRM) – Protocol terminology elaborates the standards for planning and designing a research protocol with focus on study characteristics such as study design, eligibility criteria, and requirements from the government bodies such as ClinicalTrials.gov, WHO registries, and EudraCT registries.
- CDISC Glossary – This initiative of CDISC tends to harmonize definitions as well as acronyms, abbreviations, and initials used in CDISC standards for clinical research.

According to CDISC 2014 Business Case Highlights, 'the use of CDISC standards at project initiation can save 70-90% of time and resources spent prior to first patient enrolled and approximately 75% of the non-patient participation time during the study conduct and analysis stages.' Implementing CDISC standards in pharmaceutical R&D settings has a number of important advantages such as:

- Ready to plug and play – CDISC standards due to their maturity can be applicable to any type of clinical research studies. Standards are also offered for a wide range of therapeutic areas.
- Promoting Interoperability – CDISC standards can aid interoperability between clinical systems and EHRs, lowering the amount of data re-entry and improving data quality.
- Supporting Regulatory Submission – Data standardization facilitates effective and high-quality data review. In today's environment regulatory bodies have made it compulsory for the organizations to standardize their clinical trials data to facilitate timely reviews.
- Enabling Collaborations – Data sharing is greatly aided by CDISC standards, which improve R&D efficiencies both internally as well as with external partners.

To build a holistic and comprehensive data glossary terminologies associated with other clinical data standards can also be considered such as SNOMMED CT, LOINC, HL7-FHIR (Fast Healthcare Interoperability Resources), etc. but keeping in mind that there is no duplication of terms and their definitions. The idea is to make a data glossary that has terminology from all the possible sources but industry specific and each term in the glossary is unique.

Together, CDISC and HL7 collaborate to successfully connect healthcare and research while maintaining the integrity and quality of the data to support the use of HER data in clinical research. To speed up the transfer of data from Electronic Health Records to datasets that are ready for CDISC submission, the HL7 FHIR standard for electronically transferring healthcare information partners with the CDISC Standards CDASH, SDTM, and Lab Data Exchange.

But then the question arises here is why not just simply use SNOMED Clinical Terms instead of Controlled Terminologies for CDISC Standards. There are a few good reasons to this question:

i. The first reason is that CDISC is an open platform. Anybody can easily access CDISC terminologies. Whereas on other hand there is license requirement for use of SNOMED. Though governments of multiple countries have arranged licences for users but there are still some parts of the world where use of SNOMED is not free.
ii. The second reason is SNOMED terms lack definitions. Though the terms are very-well documented, they are not defined.

Chapter 5: Conclusion

In conclusion, though the idea of data glossary is very simple, but it plays a very important role in optimizing data operations and ultimately imparts a huge positive impact on business. There is no doubt that there are endless advantages of establishing a comprehensive and well-maintained data glossary such as – quicker time-to-market of drugs and therapeutics; streamlined processes, decreasing the error rate and time-saving; reduced costs and resources; increase in operational productivity; improved data and processes transparency and visibility across all the verticals of the organization; improved data access provisioning and monitoring, improved ease of data exchange and onboarding; enhanced data quality facilitating data governance, and many more. CDISC terminologies certainly aid in building standardized data glossaries. CDISC has proved to be a success in standardizing clinical research data for regulations worldwide. It aids in streamlining regulatory submission, review, and approval along with enabling data sharing and metadata analysis. For the purpose of creating data glossaries, it is strongly advised to take into account CDISC standards as they will offer advantages in terms of cost, time, and high data accuracy. Additionally, these guidelines serve as a catalyst to speed up the review process and assist businesses in releasing new medications onto the market. But to build a holistic and comprehensive Data Glossary all the terminology from

the existing standards should be reviewed and considered from all the possible sources while making sure that each data term is unique.

References

1. How the pharmaceutical industry can meet the challenges of data quality management [Internet]. Striped Giraffe. 2019 [cited 2022 Jul 1]. Available from: https://tinyurl.com/25vhdr3f
2. Balasubramanian V. Data as a strategic asset in the life sciences [Internet]. Pharmaceutical Commerce. 2020 [cited 2022 Jul 1]. Available from: https://tinyurl.com/2a29mexe
3. What is DataOps? [Internet]. Nexla. 2017 [cited 2022 Jun 29]. Available from: https://www.nexla.com/what-is-dataops/
4. Pharma R&D data [Internet]. Pharmafocusasia.com. [cited 2022 Jun 29]. Available from: https://www.pharmafocusasia.com/foreword/pharma-rd-data
5. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery: Principles of early drug discovery. Br J Pharmacol [Internet]. 2011;162(6):1239–49. Available from: http://dx.doi.org/10.1111/j.1476-5381.2010.01127.x
6. Economic Survey 2020-21 [Internet]. Indiabudget.gov.in. 2021 [cited 2022 Jul 1]. Available from: https://tinyurl.com/k8p428ch
7. Industry Research. Global Life Sciences Analytics market 2022 opportunity by upcoming trends, profit, industry dynamic, market contribution, size-share, top companies strategy, and consumption trend till 2028 [Internet]. Industry Research. 2022 [cited 2022 Jun 29]. Available from: https://tinyurl.com/25yvt7u5
8. Data standards and technology empower life sciences companies [Internet]. Forbes. 2014 [cited 2022 Jun 29]. Available from: https://tinyurl.com/2p997km8
9. Kuan B. Implementing DataOps in life sciences [Internet]. Tamr Inc. 2020 [cited 2022 Jun 29]. Available from: https://tinyurl.com/2ej2f4vn
10. Definition of DataOps - Gartner information technology glossary [Internet]. Gartner. [cited 2022 Jun 29]. Available from: https://www.gartner.com/en/information-technology/glossary/dataops
11. DAMA-DMBOK functional framework [Internet]. Damadach.org. [cited 2022 Jun 29]. Available from: https://tinyurl.com/2vnh7kkm
12. What is a data glossary? [Internet]. Castordoc.com. [cited 2022 Jun 29]. Available from: https://www.castordoc.com/blog/what-is-a-data-glossary
13. What is a data glossary, and how do you build one? [Internet]. Atlan. [cited 2022 Jun 29]. Available from: https://atlan.com/what-is-data-glossary/

14. What is a data glossary? [Internet]. Secoda.co. [cited 2022 Jun 29]. Available from: https://www.secoda.co/glossary/data-glossary
15. Prijic M. 5 data and Analytics challenges for life sciences companies [Internet]. IT Convergence. 2021 [cited 2022 Jun 29]. Available from: https://tinyurl.com/4daxjmyc
16. Buntz B. Pharma's top 20 R&D spenders in 2021 [Internet]. Drug Discovery and Development. WTWH Media; 2022 [cited 2022 Jun 29]. Available from: https://www.drugdiscoverytrends.com/pharmas-top-20-rd-spenders-in-2021/
17. Theodora S. Top 10 CROs to watch in 2022. [Internet]. Proclinical.com. [cited 2022 Jun 29]. Available from: https://tinyurl.com/4syufwpr
18. Philippidis A. Top 25 Biotech Companies of 2022. Genet eng biotechnol news [Internet]. 2022 [cited 2022 Jun 29];42(4):20–1. Available from: https://www.genengnews.com/a-lists/top-25-biotech-companies-of-2022/
19. CDISC terminology [Internet]. Cancer.gov. [cited 2022 Jun 29]. Available from: https://datascience.cancer.gov/resources/cancer-vocabulary/cdisc-terminology
20. CDISC 2014 Business Case highlights significant time and cost savings through use of CDISC standards in medical research studies [Internet]. PRWeb. 2014 [cited 2022 Jun 29]. Available from: https://tinyurl.com/4rvec67f
21. CDISC [Internet]. Cdisc.org. [cited 2022 Jun 29]. Available from: https://www.cdisc.org/
22. Why not just use SNOMED? [Internet]. Cdisc.org. [cited 2022 Jun 29]. Available from: https://www.cdisc.org/kb/articles/why-not-just-use-snomed
23. de Boer M. Data management in life sciences: overcome these 4 core challenges [Internet]. Scilife.io. Scilife; 2022 [cited 2022 Jun 29]. Available from: https://www.scilife.io/blog/data-management-in-life-sciences
24. Gaddale JR. Clinical Data Acquisition Standards Harmonization importance and benefits in clinical data management. Perspect Clin Res [Internet]. 2015 [cited 2022 Jun 29];6(4):179–83. Available from: http://dx.doi.org/10.4103/2229-3485.167101
25. Facile R, Muhlbradt EE, Gong M, Li Q, Popat V, Pétavy F, et al. Use of clinical Data Interchange Standards Consortium standards for real-world data: Expert perspectives from a qualitative Delphi survey. JMIR Med Inform [Internet]. 2022 [cited 2022 Jun 29];10(1):e30363. Available from: https://medinform.jmir.org/2022/1/e30363