PharmaSUG 2023 - Paper MM-206

Challenges with Metadata Repository System Implementation: A Sponsor's Perspective

Radhika Kale, Reema Baweja, Bristol Myers Squibb

ABSTRACT

In the modern world, Metadata Repository (MDR) systems are a critical element of metadata management commonly used in a variety of industries, such as biotechnology and pharma, finance, healthcare, manufacturing, to improve data governance and performance. MDR is a database created to gather, store, and distribute contextual information (i.e., metadata) about data. The most efficient MDR system helps to implement frustration free end to end implementation of standards while adhering to quality, automation, and flexibility. There are certain challenges that can arise when implementing and working with MDR systems that are commercially available or built in-house. This paper gives an overview of specific challenges from sponsor's standpoint and helps set expectations from an ideal MDR system. Challenges are described in two categories: Implementation Challenges including metadata strategy development, selection of standards and MDR system, building a metadata team, creating data catalog; and Operational Challenges pertaining to Data Quality, Data Governance, Integration with other systems, User Adoption, Maintenance and Support, and Scalability. Additionally, the paper also explores opportunities to mitigate specific challenges, thereby providing meaningful input for efficient product enhancements.

INTRODUCTION

A metadata repository (MDR) system is an important asset in the biotechnology and pharmaceutical industry, as it enables organizations to better manage and organize data and information related to their products and processes. An ideal clinical MDR system in conjunction with study automation platforms can help with expedited trial designs, offer early insights, minimize data rework, and ensure availability of validated metadata in a format that meets regulatory standards. However, some organizations may not realize full potential of such MDR system given the overall complexity, which can require significant resources and expertise to implement and maintain. Additionally, specific barriers such as lack of awareness of the benefits of metadata management; a resistance to changing established data management practices; lack of standardized metadata formats across different platforms can create challenges in data sharing and interoperability, and thus discourage companies from investing in MDR systems. There is an absolute need to leverage the use of MDR system in clinical space to ensure efficient management of large amounts of data generated during drug development, right from discovery to clinical trials.

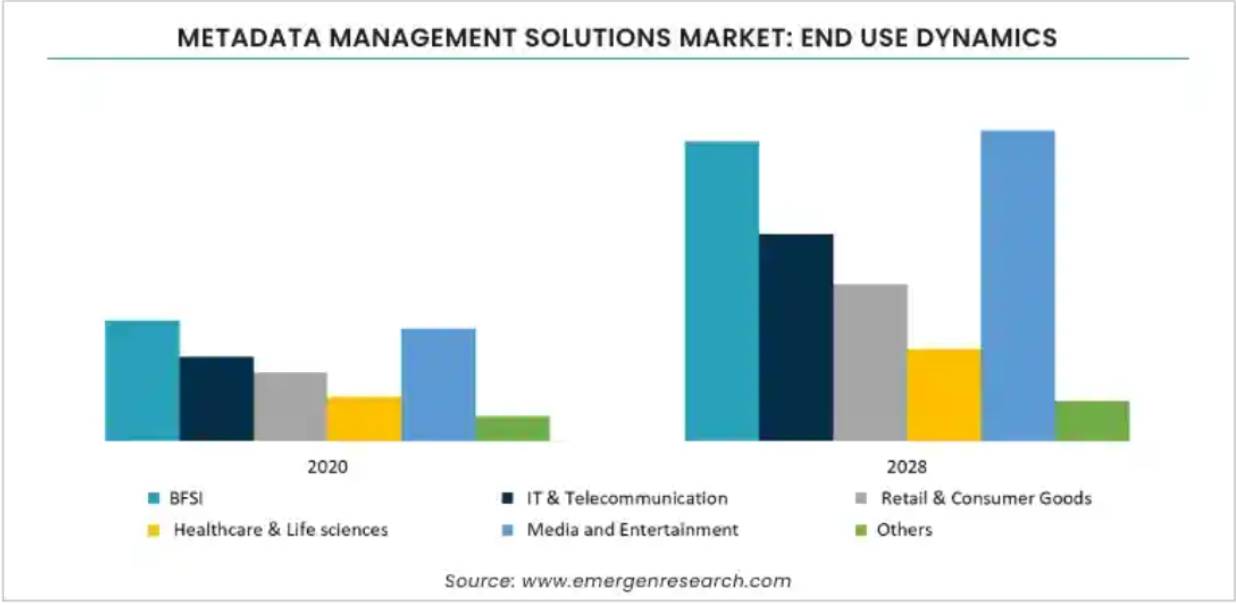

Figure 1. End Use Dynamics of Metadata Management tools across industry types

To ensure seamless MDR implementation, it is necessary for the sponsor companies to consider specific challenges that are commonly observed. These challenges are categorized as Implementation Challenges including metadata strategy development, selection of standards and MDR system, building a metadata team, creating data catalog; and Operational Challenges pertaining to Data Quality, Data Governance, Integration with other systems, User Adoption, Maintenance and Support, and Scalability.

MDR IMPLEMENTATION AND OPERATIONAL CHALLENGES

1. IMPLEMENTATION CHALLENGES:

a. Metadata Strategy Development & Selection of Standards:

Metadata strategy development in a Metadata repository system involves defining the goals and objectives of the metadata repository, identifying the stakeholders and their needs, determining the types of metadata to be included, and establishing policies and procedures for the creation, capture, storage, maintenance, dissemination, and use of metadata. The scope would be heavily dependent on the metadata types that are to be managed.

i. Nomenclature and Conventions:

One of the most significant components of metadata development strategy is the nomenclature conventions. To maintain standardization and consistency across collection metadata, it is imperative to use some naming conventions and rules. However, it is also important to define such conventions in a way that it does not interfere with the metadata repository's architecture and technical landscape.

Challenge Example: Common data points between CRF and external data – In different MDR systems, each data point is often considered as an object, an asset or a data element and has an internal relationship/linkage with other assets/elements. This is done to ensure a change when made to an object, is cascaded down automatically to all assets/objects linked with the parent; with no manual intervention and oversight. However, this relationship/linkage set up in MDR systems has seemed to intervene with how the metadata from different collection source systems is named. The two most common sources of collection are the EDC and vendor data, and most sponsors would

use the same name for a data point that is common between CRF and external vendor data. This has been a challenge when it comes to making any changes to the metadata object, originating from one source system only. As the object is named the same, but has its relationships established based on the origin, any update made to it leaves an impact on all its relationships. The change in this case is not localized and leaves the metadata curators no option other than defining a unique object name for the same data point, defeating the purpose and intention of standard naming conventions.

ii. Submission Standards:

Whether managing SDTM and/or ADaM metadata, it is necessary to have clear requirements around the components within each metadata type. For example, core requirements around SDTM metadata management would include Project/Study or Global level SDTM metadata, Domain and variable level metadata, Value-level metadata (VLM), Controlled Terminologies, RELREC specifications, sponsor's mapping conventions, general and domain derivations. Lack of clarity in any of these components would lead to incomplete scope for MDR, and cross functional challenges when the metadata is utilized by downstream consumers.

Challenge Example: Not Defining VLM model upfront – In different MDR systems, VLM model could be in the form of a specification table (like seen in Define.xml output) or in the form of relationship rules wherein the source variables with different Origin, Types, Length/Display Format are linked to a single target SDTM variable. It is critical to have a clear vision and requirement around such model to ensure complete SDTM model is managed within MDR without the need to manually handle any of its components outside of MDR (such as excel files), and complete set of specifications are extracted from MDR and delivered to downstream consumers. In essence, whether defining a global or project/study level SDTM metadata, the goal is to provide a standard framework to study programmers and minimize manual effort to create study level mapping specifications.

iii. Maintaining multiple SDTM standards in MDR:

An ideal MDR system is designed to harbor multiple SDTM standards including different versions of SDTM model/IG, NCI CT, TAUGs, Define/ODM configurations, etc. The design could be as simple as storing different versions in separate folders or stacks within MDR. Each stack can include the required metadata, such as variable definitions, data structures, and relationships, specific to that standard. For this purpose, it is important to ensure that the MDR is updated and maintained regularly to reflect any changes in the SDTM standards being used. This will help ensure that data collected across different studies is consistent and can be easily compared and analyzed.

Challenge example: While selecting MDR system, it is necessary to closely evaluate the MDR design to be able to maintain multiple versions of submission standards. Lack of such evaluation would result in a complicated/band-aid approach to implement specific submission standards at collection level and defeat the purpose of end to end standards implementation.

b. Building a Standards Team:

MDR system could be built in-house or vendor supplied. In either case, a centralized organization/functional group (such as Clinical Data Standards group) is a primary requirement that should include people who have extensively worked on functional standards. Individuals with expertise in collection standards (CRF, nonCRF/external) such as data managers, database developers; expertise in submission standards (SDTM, ADaM, TLFs) such as statistical programmers, data analysts/scientists, statisticians; technical SMEs such as information or solutions architect, IT program managers, etc. In general, Clinical Data Standards talent pool could ideally be a blend of experts who have extensive experience in development, implementation, and maintenance of end-to-end data standards for clinical trials. Defining a metadata model in MDR system requires appropriate SMEs' input, evaluation and assessment, and contributes to key business decisions.

Challenge Example: Lack of knowledge around the approved external data vendors and their standard data transfer specifications/methods could lead to a flawed nonCRF metadata model in MDR.

c. Creating Data Catalog:

A data catalog is an organized inventory of information that can be used by technical and non-technical users to find and access information quickly. Data catalogs help users go through all the context they need to know, at a glance with:

- Comprehensive business glossaries and descriptions
- In line annotations

Most MDR systems maintain a master data catalog that provides information about the technical landscape of the MDR and describes the attributes for each MDR asset/data element/object, like what a physical or online catalog would do.

Challenge Example: Complex Data Catalog structure and components – A complex master data catalog of MDR system attributes can result in significant learning gaps for metadata curators to understand the system specifics and use them to model the metadata attributes.

2. OPERATIONAL CHALLENGES:

a. Data Quality:

Several factors determine quality of metadata in a repository system. To name a few:

Accuracy: Metadata should be up-to-date, correct, and complete. To ensure accuracy, Data Standards team can develop metadata checks. For example, programmatic checks for operational metadata that align with EDC checks to ensure a validated framework is provided to cross-functional consumers. SDTM metadata can be validated either by passing the SDTM content from within a Define.xml to Pinnacle 21 validator or by developing programmatic checks aligned with Pinnacle 21 validation rules.

Consistency: Consistent metadata across the repository system. For example, the same terms and definitions, sponsor-defined metadata (ALS/SDTM) conventions should be used consistently throughout the system. Specific metadata compare reports could help ensure consistency of metadata across universal/global, TA subset, and study level libraries. For example, an ALS compare report could help identify the differences between a universal and study level metadata. Such inconsistencies could be identified and fixed at the time of initial study database build and downstream mapping or reporting issues could be avoided.

ProjectName	ProjectType	OID	DraftFormName
Universal MDR Library	Universal	UForm1	Study Drug Exposure - {Drug Name}
Study 1	Project	UForm1_S01	Study Drug Exposure- Drug X
Study 1	Project	UForm1_S01	Study Drug Exposure- Drug Y

Figure 2. Example – ALS Compare Report

Completeness: Metadata should provide a complete and comprehensive description of the data. It should include all relevant information, such as data source, format, and ownership.

Timeliness: Metadata should be updated in a timely manner to ensure availability of updated standards as released by CDISC, while adhering to the implementation timeline as defined by the sponsor.

Accessibility and Versioning: The metadata should be easily accessible to MDR users as well as to the downstream consumers for post-processing. It should be organized and presented in a way that is easy to navigate, understand, and extract. It is also helpful to provide a metadata difference report to the consumers to highlight the standards that were updated with reference to collection or submission metadata. An ideal MDR system can maintain multiple versions of an element or asset that is customized,

improved or updated. User should be able to easily identify and be confident they are using correct version of asset or standard.

Extract Version	domain	description	class	purpose	structure	mapping_instruction
SDTM_Extract_20220424 (Compare)	CM	Concomitant Medications	Interventions	Tabulation	One record per recorded medication occurrence or constant-dosing interval per subject.	CRF.FormOID1, CRF.FormOID2, CRF.FormOID3, CRF.FormOID4, nonCRF.OID1, nonCRF.OID2
SDTM_Extract_20220324 (Base)	CM	Concomitant Medications	Interventions	Tabulation	One record per recorded medication occurrence or constant-dosing interval	CRF.FormOID1, CRF.FormOID2, nonCRF.OID1

Figure 3. Example - Metadata Diff Report: SDTM Domain mapping attributes

Extract Version	domain	name	label	type	length	seq	cdisc_ type	cdisc_ origin	cdisc_role	cdisc_ core	source	variable_source	method_id	Status	Status_Date
SDTM_Extract_20220424 (Compare)	FA	FAREASND	Reason Not Performed	C	200	19	text	Derived	Record Qualifier	Perm	SDTMIG 3_2; SDTM 1_4	Src_Var1, Src_Var2	REASND	ACTIVE/APPROVED	2023-04-24
SDTM_Extract_20220324 (Base)	FA	FAREASND	Reason Not Performed	C	200	19	text	Derived	Record Qualifier	Perm	SHARE SDTMIG 3_2; SDTM	Src_Var1	REASND	DRAFT	2023-03-24

Figure 4. Example - Metadata Diff Report: SDTM Variable mapping attributes

Extract Version	FormOID	FieldOID	Ordinal	DraftFieldName	DraftFieldActive	VariableOID	DataFormat	DataDictionaryName	ControlType	IndentLevel	PreText
ALS_Extract_20230424	FormOID1	FieldOID1	12	FieldOID1	TRUE	FieldOID1	$200		LongText	0	Updated Pretext
ALS_Extract_20230323	FormOID1	FieldOID1	31	FieldOID1	TRUE	FieldOID1	$40	DD1	SearchList	0	Original Pretext

Figure 5. Example - Metadata Diff Report: ALS metadata

Impact Analysis: A high end MDR system should be able to analyze the impact of change. All upstream or downstream related standards and elements need to be analyzed when there is a change in any of the attributes of an element. Impact analysis help with informed decisions before any change is made in an established standard element.

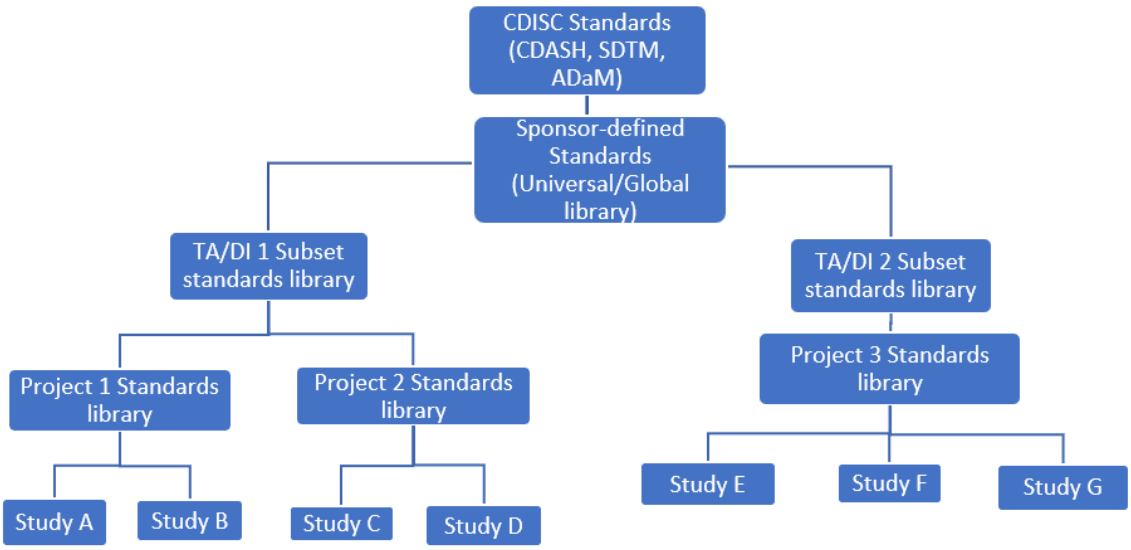

Figure 6. Downstream Impact on Metadata Standards Hierarchy

Ensuring high-quality metadata requires ongoing effort and attention around data governance policies and procedures, appropriate training to users, and using tools and technologies to automate metadata management tasks.

b. Data Governance:

A well-defined Data Governance model enables any organization from the get-go, to define the principles for aligning data strategy with business strategy, document the decisions, identify the owners and stakeholders and disseminate roles and responsibilities, establishing a solid roadmap for an effective metadata management, evolution and implementation for the organization. An ideal MDR system would enable any sponsor to customize and establish a seamless governance workflow to better manage and govern the metadata and its evolution over time.

Challenge Example: Disintegrated governance workflows – As much as it is important to have a governance workflow in place, it is important to keep in mind that the goal should be to:

- Initiate the governance workflow within the MDR
- Provide all stakeholders easy access to the MDR and governance workflow
- Provide access to the documents and specifications within the MDR, that support the metadata curation
- Minimize any major business process changes

If the only way to access the governance workflow and its components is via an email system or an external database outside the MDR, it makes it too tedious and cumbersome for the users to show adherence to the process. Needless to say, this approach meets with no success.

c. MDR Framework & Integration with other Systems:

Integration of a clinical MDR system with other systems requires establishing an API (Application Programming Interface) to allow communication between MDR and intended systems. In general, following series of steps can be followed to create a clinical MDR framework and API integration:

Define the metadata: Define the metadata per the selected standards. This involves identifying the data elements that are relevant to your clinical data, as well as the data definitions and models.

Design/Select the MDR: Based on the metadata definition, design or select the MDR system with thorough knowledge around database schema, data storage format, and user interface.

Develop the API: To integrate the MDR with other systems, create an API that allows for communication between the systems. This will involve defining the API endpoints, request and response formats (based on push and pull approach), authentication and authorization mechanisms.

Implement the API: Once the API is designed, implement it in the MDR and any other systems that will be integrating with it.

Test the integration: Before deploying the integration in a production environment, it is important to test it thoroughly in lower environment to ensure that it is functioning as expected. Testing strategy would differ in every sponsor company and could involve substantial effort from cross-functional teams. Necessary to factor in time and effort.

Maintain the MDR and API: As clinical data and metadata evolve, it is important to maintain the MDR and API to ensure that they continue to meet the needs of the organization.

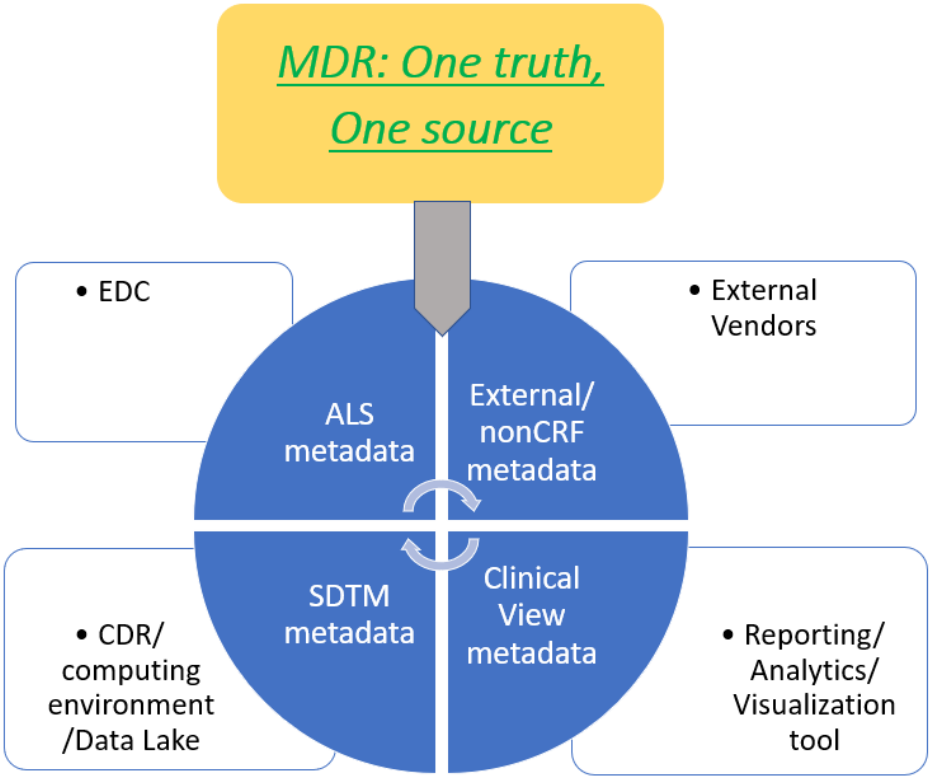

Figure 7. MDR API Integration with other systems

Creating a clinical MDR framework and API integration requires careful planning and execution, but it can provide significant benefits for consistently managing clinical data and metadata, as well as leverage automation thus eliminating the manual processes involved in transferring metadata from MDR to target system.

d. User Adoption:

Planned user adoption strategy can help organizations boost productivity, provide a better user experience, increase user engagement and make the most out of the new technology. One of the biggest barriers to a successful user adoption strategy is, users being unprepared to learn something new or

adopt something that is far too technical, non-intuitive and cumbersome. Even with the proper training material and lessons, dealing with an entirely new platform or software that is not intuitive or easy to use, can be daunting.

Challenge Example: Complicated User Interface and consumption of metadata by end users – Too often the MDR systems user interfaces are too graphical, technical and too far from real world view for end users, making it too system focused than end user focused. Most end users are known to maintain metadata as excel spreadsheets, a MDR system that has an interface more complex than what a simple excel spreadsheet provides, make the users apprehensive and demotivated to use the MDR. In addition to user disinterest, it also puts a lot of effort into developing training material to showcase the system and its capabilities. Also, if it takes complex programming and transformations to extract the metadata out of the system for downstream consumption and system integrations, in no time the system is considered a bottleneck and not user friendly.

e. Maintenance and Support:

Regular maintenance and support are essential to ensure reliability and functionality of MDR system. This involves specific tasks:

Data backups: Regular backups of the MDR system data are essential to ensure that the data is not lost in case of system failures, hardware malfunctions, or other disasters. Backups should be stored in a secure location and tested regularly to ensure data integrity.

Security management: Security aspect is critical as MDR systems contain sensitive clinical data. Regular security assessments and updates, including access control, data encryption, and auditing are necessary to protect the data against potential security vulnerabilities.

Performance monitoring: Monitoring the performance of the MDR system can help identify potential issues before they become critical. Regular performance tuning can help optimize the system's performance, improve user experience, and reduce downtime.

System updates and upgrades: Regular updates and upgrades to the MDR system software are necessary to ensure that it is up to date with the latest security patches and bug fixes. Upgrades can also bring new features and functionalities to the system.

User support: Users may require assistance with system usage, troubleshooting, or training. Providing timely and effective support to users can help ensure that the MDR system is being used effectively and efficiently. This includes responding to user inquiries and providing assistance with system issues.

f. Scalability:

Clinical metadata repository systems are designed to store and manage clinical data in centralized and organized manner. As the volume and complexity of clinical data continue to grow, it is critical that these systems are scalable to handle increasing amounts of data. Several factors are to be considered that can impact scalability of MDR system:

Data volume: As the amount of clinical data stored in the repository increases, the system needs to be able to handle the increased volume of data. This may require upgrades to hardware or software or using a cloud-based platform that can scale dynamically.

User concurrency: As the number of users accessing the system increases, the system needs to be able to handle multiple requests simultaneously. This may require upgrades to hardware or software, or the use of load balancing technologies.

Performance: The system needs to be able to deliver query results and reports in a timely manner, even as the volume of data and number of users increases. This may require optimizations to database schema, indexing, or query execution plans. Performance and scalability testing should be performed regularly to identify any bottlenecks and to optimize the system's performance.

Data security and privacy: As the repository grows, so does the importance of maintaining the confidentiality, integrity, and availability of the data. This may require upgrades to security protocols, access controls, and disaster recovery plans.

Challenge Example: Sudden, unexpected MDR downtime and subsequently the time taken by the development team to identify and implement resolution. Such situation heavily impacts scheduled metadata releases and deliverables that ultimately impact planned timelines around study milestones. One of the common causes being lack of scalability testing on a regular basis.

CONCLUSION

For a sponsor company to successfully implement MDR system, it is necessary to have a well-researched and complete set of requirements after a giving thoughtful consideration around all the challenges and possible mitigations. It is equally imperative to ensure that all the relevant stakeholders are onboard with the requirements as this journey is highly collaborative.

In addition to standards management, MDR usage can also be extrapolated to:

Fast and Effective Protocol Design – MDR can help ensure consistency and accuracy in protocol design. By storing all relevant information around protocol elements in one place, the system can help ensure that all stakeholders are working from the same set of specifications and guidelines.

Product Lifecycle Management - MDR can be used to manage the entire lifecycle of a product, from initial research and development through to manufacturing and post-market surveillance.

Product Documentation - MDR can be used to store and manage documentation related to products, such as product specifications, regulatory filings, and labeling information.

REFERENCES

Metadata Management Solutions Market, Dec 2021. Available at https://www.emergenresearch.com/industry-report/metadata-management-solutions-market

CDISC Foundational Standards: https://www.cdisc.org/standards/foundational

ACKNOWLEDGMENTS

The authors would like to thank the Clinical Data Standards Leadership team and BMS internal team for providing valuable guidance and feedback.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Radhika Kale
Bristol Myers Squibb
radhika.kale@bms.com
https://www.linkedin.com/in/radhika-kale-438ba318/

Reema Baweja
Bristol Myers Squibb
reema.baweja@bms.com
https://www.linkedin.com/in/reema-baweja-6a182715/