The Evolution of Clinical Data Management into Clinical Data Science

An SCDM Position Paper on ***how to create a Clinical Data Science Organization***

"Your path toward Clinical Data Science"

12 SEP 2022

Society for Clinical Data Management

© Society for Clinical Data Management

Our Vision

"*Leading innovative clinical data science to advance global health research and development"*

Our Mission

"Connect and inspire professionals managing global health data with global education, certification and advocacy"

Copyright © 2022 Society for Clinical Data Management. All rights reserved

- TABLE OF CONTENTS -

Contents

Contents **2**
1. Foreword **3**
2. Abstract **3**
3. Methodology **4**
4. Acknowledgements **4**
5. The path to Clinical Data Science **5**
5.1 Background 5
Understanding our CDM roots (i.e., our starting points) 5
Driving forces behind the evolution of CDM into CDS 5
What is Clinical Data Science? 7
5.2 Transformation to "create" a CDS organization: 7
A – The assessment Phase (i.e., Know where you start from) 8
B – The envisioning Phase (i.e., Know where you go) 9
Foundational Framework 9
Future proofing capabilities 10
Required operating model to enable the target state capabilities 13
C – The Design Phase (i.e., The roadmap from current to end state) 13
The roadmap definition 14
The organizational design 15
D – The transition phase (i.e., how to implement the roadmap) 16
Planning the transition phase 17
Transition into the steady state 18
Conclusion **19**
Main abbreviations **20**
References **21**

1. Foreword

Clinical Data Management (CDM) evolved over the last two decades from managing data entered on paper Case Report Forms (CRFs) to managing data transcribed into Electronic Data Capture (EDC) systems. The Society for Clinical Data Management (SCDM) strongly contributed to this first significant CDM evolution through its Good Clinical Data Management Practice[1] (GCDMP©) Chapters first published in 2000 and its certification program for Clinical Data Managers[2] subsequently released in 2004. Now, CDM must evolve even more dramatically to a patient-driven (i.e., executing clinical trials decentralized at patient's will) and to risk-based clinical trial and data management environments. SCDM wants to continue its mission and support CDM professionals adapting to this second ongoing evolution by guiding them into Clinical Data Science (CDS).

As such, after publishing three reflection papers on the evolution of CDM into CDS, the SCDM Innovation Committee and the SCDM Board want to formally establish SCDM's position on CDS and how SCDM's members and practitioners of CDM may evolve into it. This position paper will not rehash the content of the previous reflection papers but rather emphasize on key CDS concepts and focus specifically in providing insights on how CDM professionals can efficiently set their path forward toward CDS.

2. Abstract

The evolution of drug development and the decentralization of clinical trials accelerated by the COVID-19 pandemic combined with recent geopolitical events, has required us to act decisively. As a result, to remain effective and meet these rising needs, some CDM organizations have started to seize the opportunities offered by technology and regulatory changes to augment their CDM capabilities and become resilient to unexpected emergencies. Others still need to initiate the move to support complex study designs in a fit-for-purpose and patient-driven data framework. To support those in need for a direction, the SCDM Innovation Committee seeks to set a position on possible pathways to evolve into CDS by building upon the content of the three previous SCDM reflection papers on the "evolution of Clinical Data Management toward Clinical Data Science"[3,4,5].

This position paper is not intended to be used as an exhaustive change management guide but provides a set of CDM specific recommendations that are essential to consider while building, evolving, or transforming a CDM organization. Additionally, it is important to acknowledge that each organization is unique and that a one-size-fits-all strategy would lead to more failures than successes. Therefore, it is key for each organization to tailor their own evolution strategy seeking to address the four Ws (Why, What, Who, and When) considering their current CDM scope, roles & responsibilities, talents, partners, technologies, and operating models.

Lastly, there is a natural human *tendency* of minimizing change, making it incremental, keeping the organization structure as is, and focusing on technology as the answer. However, a true and effective transformation needs an achievable vision which should be unconstrained by the current state. CDM Leadership teams will therefore need to be comfortable with being uncomfortable and foresee the need for enough investment and clear communications in their organization. This is the core role of change leadership, a critical component to the change management journey needed to build consensus and confidence with their peers and their respective functions.

3. Methodology

The SCDM Innovation Committee seeks to provide Thought-Leadership to our industry and support the SCDM vision of "leading innovative clinical data science to advance global health research and development". To that end, the SCDM Innovation Committee strives to demystify CDS and support the development of all CDM professionals, from subject matter experts (SMEs) working on clinical studies to CDM leaders setting the direction of their organizations. Understandingly, due to the recent emergence of the CDS discipline and the absence of a comprehensive literature base regarding CDS, this content was gathered from industry leaders through a consensus-based methodology. As CDS matures, and technologies evolve it is anticipated that literature on CDS will blossom.

4. Acknowledgements

Disclaimer: The views expressed in this reflection paper do not necessarily reflect those of the companies or entities the authors are employed by or affiliated with.

Lead authors:

- Patrick Nadolny, Global Head, CDM, Sanofi, SCDM Innovation Committee Chair
- Demetris Zambas, Global Head, Data Monitoring & Management, Pfizer, Former SCDM Board Chair

Content contributors and reviewers:

- Mayank Anand, VP Global Data Strategy, Monitoring & Management, GSK, SCDM Board Chair (2022)
- Sanjay Bhardwaj, Head, Clinical Technology Strategy & Operations, Abbvie, SCDM Board Chair (2021)
- Ward Lemaire, Global Head, Data Management, Janssen
- Arshad, Mohammed, Head, Global Data Management, GSK, SCDM Vice Chair (2022)
- Ian Shafer, Principal, PwC

SCDM Innovation Committee authors who contributed to the 3 Reflection Papers and 6 Topic Briefs used as a foundation to this position paper.

- Sanjay Bhardwaj, Head, Clinical Technology Strategy & Operations, Abbvie, SCDM Board Chair (2021)
- Catherine Celingant, Sr Director, Data Monitoring & Management – Oncology, Pfizer
- Lynne Cesario, Sr. Director, Global Risk Based Monitoring Program Lead, Pfizer
- Steve Chartier, Sr. Director of Engineering, CALYX
- Catherine Hall, Vice President, Product Strategy, endpoint
- Patrick Nadolny, Global Head, CDM, Sanofi, SCDM Innovation Committee Chair
- Prasanna Rao, Sr. Director, Artificial Intelligence and Machine Learning, Pfizer
- Inder Sachdeva, Portfolio Delivery Leader, Clinical Data Sciences & Safety, Cognizant
- Ian Shafer, Principal, PwC
- François Torche, CEO, CluePoints
- Joshua Wilson, Executive Director, Biometrics, Strategic Technology Advancement, Syneos Health
- Richard Young, Vice President, Vault CDMS Strategy, Veeva
- Demetris Zambas, Global Head, Data Monitoring & Management, Pfizer, Former SCDM Board Chair

SCDM would also like to acknowledge the many volunteers who have participated in the SCDM Innovation Committee and have contributed to forming the thoughts expressed in this position paper.

5. The path to Clinical Data Science

5.1 Background

Understanding our CDM roots (i.e., our starting points)

An organization must understand its journey to current state (i.e., Where they are coming from) in order to plot a course to its future (i.e., Where they want to go). CDM has evolved from humble beginnings focusing on designing data collection forms, executing double data entry, reviewing data listings to identify erroneous data and issuing Data Correction Forms (DCFs) to investigational sites. Roughly two decades ago the industry began to differentiate between what it saw as "core" clinical development capabilities vs. transactional and repetitive supporting processes. For the latter, organizations promptly focused on commoditization mainly through outsourcing and offshoring.

The commoditization of CDM in large drove the generation of detailed output metrics to monitor efficiencies and relative quality of deliverables from the "data management production lines". This unintentionally resulted in shifting the focus entirely on measured outputs (i.e., volume and cycle time) with little focus on meaningful outcomes (i.e., was the data fit for purpose?). As an example of "side effect" of the CDM commoditization, some companies began to use the # of queries generated by an individual in a day to measure efficiency. On the surface it may appear as a good means to measure resource utilization and throughput, however such measures provide no insights into whether the resulting clinical database is fit for purpose. Moreover, the metric could incentivize "bad behaviors", such as issuing unnecessary queries to sites when used as an assessment of an individual's productivity. To reinforce the point, a recent article[6] demonstrated that queries are only leading to 1.7% of eCRF changes after data has been entered by the site. This indicates that the outcome does not proportionally reward the significant effort that many CDM organizations place on querying data through traditional data review and data validation strategies especially for non-critical data.

Additionally, depending on the level of outsourcing and/or offshoring, this trend surfaced the criticality of the CDM knowledge base. Whether an organization elects to insource or outsource CDM activities, maintaining a critical knowledge base is key to success. Perhaps more importantly, a rich knowledge base is the foundation to gaining control of an organization's data and to facilitate innovation.

All organizations are capable of innovation. An organization's focus (e.g., outputs vs. developing new therapies to meet unmet medical needs) will result in different types of innovation. Regardless, CDM must reinvent itself to prepare to handle the rising complexity of Drug Development, the 5Vs of Clinical Data and the virtualization of many clinical trial activities.

Driving forces behind the evolution of CDM into CDS

Since the release of the first reflection paper[3] in June 2019, the core drivers of the CDM evolution toward CDS remained consistent and centered around four main themes.

1. The **complexification of clinical trials designs** (e.g., Adaptive, Master Protocols, Synthetic Control Arms) where studies regularly include different patient populations and endpoints with evolving and flexible study characteristics.
2. The **decentralization of clinical trials (DCT)** accelerated by the COVID-19 pandemic. This is leading to a wider adoption of patient centric solutions such as telemedicine, home nursing and electronic consenting. A DCT framework **allows patients to decide what suits them the best (i.e., being Patient-Driven)**. However, the volume of direct data capture (i.e., eSource) clearly challenges the centricity of traditional EDC technology and processes relying on site source's transcription, SDV and queries.
3. The **adoption of risk-based CDM approaches**[7] fostering Quality by Design (QbD) and the focus on what matters the most (i.e., Critical to Quality (CtQ) Factors, critical data and processes) according to the latest ICH E6 (R2)[8] and ICH E8 (R1)[9] guidelines.
4. The **automation of CDM activities**[10] leveraging maturing intelligent technologies acting as virtual assistants to the Clinical Data Scientist. Some technologies consist in automating repetitive and reproduceable activities using Robotic or Intelligent Process Automations (RPA & IPA) such as data transformation and reconciliation. Other advance capabilities such as Machine Learning (ML) or Natural Language Processing and Generation (NLP & NLG) can automate more complex activities such as query detection and query writing as well as predicting data issues.

Additionally, as we evolve, it is critical to remind ourselves to go beyond data integrity (i.e., achieving data quality). While ensuring data integrity is essential, reaching data quality to ensure the reliability of the trial results is at the core of our evolution and is fundamentally changing the way we manage data by focusing more on its meaning and its value.

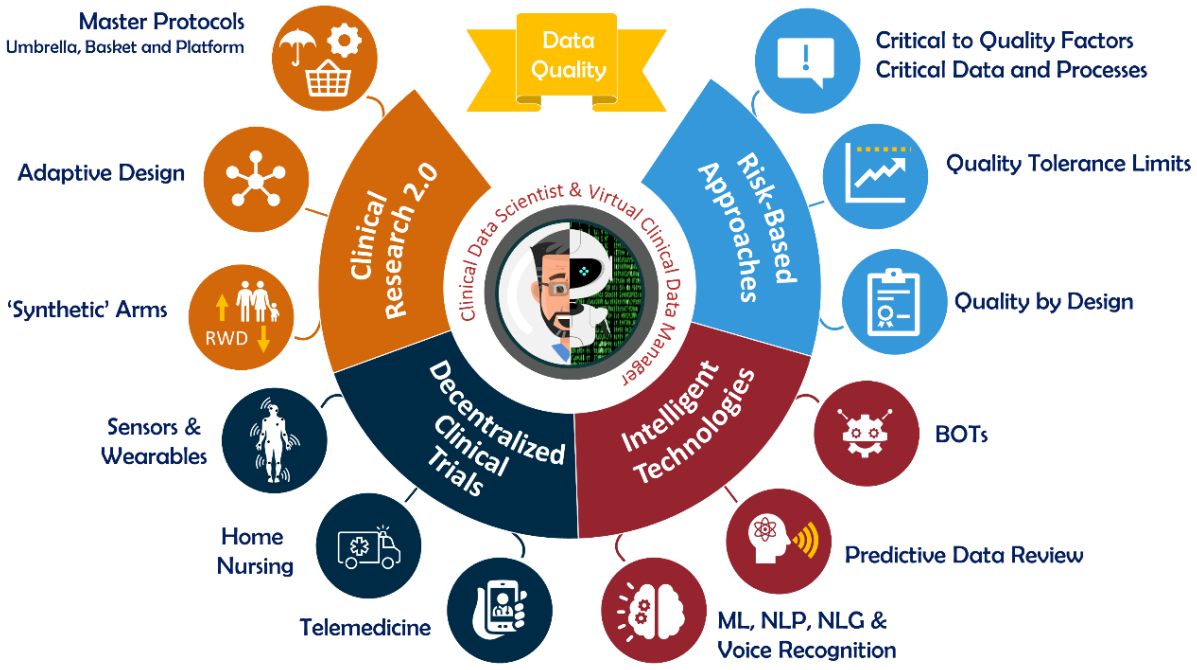


Fig 1: Driving forces behind the evolution toward CDS

What is Clinical Data Science?

First, Clinical Data Science is not to be confused with the general discipline of Data Science which applies across multiple industries. From an SCDM point of view and as expressed in the third reflection paper[5], Clinical Data Science is an evolution of Clinical Data Management. Clinical Data Science encompasses domain, process, and technology expertise as well as data analytics skills and Good Clinical Data Management Practices essential to prompt decision making throughout the life cycle of Clinical Research. Clinical Data Science can be defined as the strategic discipline enabling the execution of complex protocol designs in a patient centric, data driven and risk-based approach ensuring subject protection as well as "the reliability and credibility of trial results"[8].

In contrast, Clinical Data Management is responsible for the life cycle of clinical data from collection to their delivery for statistical analysis in support of regulatory activities. Clinical Data Management is primarily focusing on dataflows and **data integrity** (i.e., data is managed the right way). Clinical Data Science broadens this focus by adding the data risk, data meaning and value dimensions for achieving **data quality** (i.e., data is credible and reliable). Clinical Data Science also expands the scope of Clinical Data Management beyond the study construct by requiring the ability to generate knowledge and insights from clinical data to support other clinical research activities which requires different expertise, approaches, and technologies.

5.2 Transformation to "create" a CDS organization:

An effective transformation plan supports an organizational strategy, and a strategy should support an organization's mission and purpose. There is no cookie-cutter formula to establishing strong CDS capabilities. The authors and contributors of this position paper have established or restructured multiple top 10 pharma and smaller Clinical Data Management organizations from different starting points and successfully reached impactful end states tailored for their specific companies. Each case and starting point were as unique as the ingredients one may find in different households' kitchen pantries. One is able however to define a series of assessments and decision points to guide a team or organization toward a successful outcome. Much in the same way a creative chef in an unfamiliar kitchen is still able to create a palatable meal pleasing their actual guests.

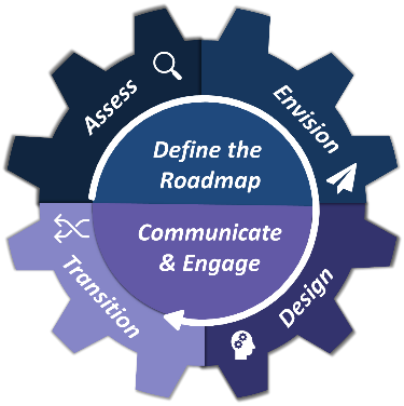


Fig 2: The 4 key Evolution and/or Transformation Steps

Basic geometry has taught us that the shortest distance between two points is a straight line. Inversely, a line is defined by Euclid as an interval between two points. Besides the obvious expectations required for any successful effort such as leadership support, stakeholder's engagement and funding, the most significant initial focus should be **assessing** and understanding an organization's existing capabilities as the first point and **envisioning** a target state which would be the second point. Hence a line or roadmap can be **designed**. Only then, an organization will be able to initiate its **transition** from one point to another.

Additionally, too many transformation initiatives focus on "People, Process and Technology" and forget to consider the internal and external Partnership dimension which is crucial as no one can succeed alone. All of those four dimensions need to be considered throughout the evolution of any CDM organization.

A – The assessment Phase (i.e., Know where you start from)

Assessing your current CDM capabilities objectively takes time but is critical to succeed. The assessment needs to be complete and unbiased enough to prevent unnecessary rework and most importantly avoid underestimating the journey. To start the transformation, there is a need to identify the sponsors, owner roles, functions, and partners in scope but also key talents and leaders to lead the change. Change leadership is critical and must encourage teams to be self-aware and be critical of their own current state.

The most obvious place to start the assessment is to perform a **S**trengths, **W**eaknesses, **O**pportunities and **T**hreats **(SWOT)** Analysis of the team, technology, capabilities, processes, and partners. Additionally, some critical aspects of the assessment phase include but are not limited to the following:

- **Technology maturity and reliance**. Some organizations may have traditional EDC centric processes with limited capabilities to manage complex external or eSource data and will therefore need to focus on technology enablers toward Data Aggregation, Data Exploration, DCT and complex protocols. On the flip side, some companies may have little internal technology reliance and/or internal system integration challenges and therefore may have more flexibility to explore new ways of collecting and managing data.
- **Internal and external dependencies** on other transformation initiatives, partner function's inputs and deliverables (e.g., digital, clinical/scientific operations), technology or CRO partnerships must be understood as those elements intersect and need to evolve with the CDM organizations.
- **Current size and operating models**: As an example, larger companies tend to have a bigger internal CDM footprint than a smaller company usually relying on CROs. Thus, requiring different change management considerations such as upskilling of internal staff vs. implementing or acquiring new CRO capabilities.
- **Understand CDM liabilities** (e.g., Current regulatory or Corrective Actions and Preventative Actions (CAPA) commitments, existing contract terms, ongoing study portfolio) as some of those will heavily influence the transition strategy from gradually evolving to radically changing capabilities.
- Understanding **the potential points of resistance** to change as they will require specific attention to key stakeholders and roles in the roadmap and transition phase.
- **Portfolio Strategy:** The portfolio diversity in term of Therapeutic Areas and study phases has an impact on the overall CDS priorities and focus. As an example, CDS organizations with focus on early development studies (e.g., Phase I / IIa) could anticipate faster pace with time sensitive Go/No Go decisions, higher number of start-ups and locks, wider adoption of eSource data and more complex master protocols and adaptive designs. So, speed and agility would be essential in that context. In contrast, CDS organizations with focus on later phase studies (e.g., Phase III) could anticipate an increase in clinical trial decentralization and a need to adopt risk-based CDM approaches in line with new regulations. So, technology and process evolution may be more deeply rooted to meet the future needs.
- **Merger & Acquisition (M&A) Strategies**: In the current industry landscape, larger companies tend to grow their development portfolio through varied M&A strategies. On the opposite side, some start-ups deliver on a proof-of-concept development of novel therapy with the anticipation of joining or collaborating with a larger company to further its development. In such cases, CDM organizations will need to tailor the transition considering ongoing studies and portfolio activities impacted by the M&A which were governed under different systems, processes, and partners. Such a transition as and when it occurs could be greatly simplified if anticipated during the design of both organizations.

- **CDM fundamental capabilities**: What capabilities exist in the organization and what state of evolution are they in? It is possible that certain capabilities are dispersed across an organization or are completely non-existent. It is critical to assess these capabilities as well as their health relative to the aspects listed above and to evolving industry benchmarks. This will also enable an assessment of possible gaps in the event that the current operating models are not adequately supported by existing CDM capabilities. These are detailed in the following section as "foundational capabilities."
- **Avoidable complexity (What can be simplified):** Some current process and technology complexities may stem from historical constraints no longer applicable. For example, it could be valuable to question why one organization may still need to collect the same information across multiple systems when data integration capabilities have matured, what is the value of having the same level of edit checks for critical and non-critical data, what is the cost of not decommissioning some legacy systems used on only few studies, etc. Those could end-up being low hanging fruits leading to meaningful improvements with low investment.
- **Unavoidable complexities (What needs to be managed):** The evolution of drug development is inevitably driving unavoidable changes. For example, CDM Organizations must be able to adapt the data collection tools and other dependent systems (e.g., data review analytics) rapidly if required during the conduct of an adaptive design study. It is no longer acceptable to take weeks to perform system change controls when the system is expected to change constantly. CDS functions must anticipate those unavoidable complexities and redesign strategies enabling drug development as opposed to constraining it.

While those are a few examples, they illustrate the need to deeply understand the current CDM state before envisioning a CDS strategy. This could range anywhere from creating a function virtually from scratch within an organization, to planning an evolution from well-established foundations. The latter does not necessarily mean an easier course since entrenched roles and processes may likely pose a difficult change management journey. Other more general (i.e., non CDM Specific) and traditional areas to assess during a CDS transformation include company culture, financial constraints, performance metric, geographic footprint, change barriers, and talent pool.

B – The envisioning Phase (i.e., Know where you go)

The target state vision should focus on the "Why" and the "What", not the "Who", "How" and "When". While the requirements articulated in the previous reflection papers[3,4,5] on the evolution of CDM into CDS to help answer "Why" and the "What", it is important to emphasize that the target state needs to support each company's unique mission and must be realistic, innovative, and achievable. This is why the assessment phase is so crucial (i.e., what is meaningful and achievable considering a company current state, based on company financial constraints, culture, strategy, etc.). CDM leaders need to connect the dots in a meaningful way based on what is critical to their own company success without assuming that everything mentioned in this paper is applicable to their own CDM organization. It is up to the leadership to envision a future state that not only solves their current gaps but also defines the state an organization will need to reach in the longer horizon. If the target is too focused in addressing the current challenges, the new organizational deliverables will likely arrive just in time to find a new set of obstacles to address.

Foundational Framework

The target state in of itself needs to include the fundamental capabilities to execute best CDM Practices as defined in GCDMP's[1] leveraging the expected competencies from certified Clinical Data Managers[2].

In conjunction with these core CDM capabilities, there are also a series of enabling capabilities necessary for any well-functioning organization such as

- Project Management
- Software Development Life Cycle (SDLC) / User Acceptance Testing (UAT) management
- Training and Upskilling
- Key Performance and Quality Indicators (KRIs & KPIs) and supporting operational metrics
- Quality Management System (QMS)

Future proofing capabilities

As stated, there is no one size fits all strategy and contextualizing the strategy and capabilities around the particular organization is key. That said, the mission of most CDS organizations would be expected to **operationalize complex clinical trials designs in a risk-based and patient-driven decentralized framework leveraging advanced technologies such as automations capabilities**.

Once the foundational elements have been defined, the end state should be complemented by the capabilities articulated in figure 1 and any recent or emerging regulatory expectations (e.g., Audit Trail Review[11], QTLs, focus on critical data, critical processes and Critical to Quality factors). Those capabilities need to be 1) relevant or required to support your company goals and strategy, 2) challenging yet achievable based on the outcome of the assessment phase and 3) bringing meaningful and quantifiable value **to your own company**. Such multi-layer strategic alignment approach of the capabilities to the overall value proposition to the organization can help assure support from the sr. leadership, partners, and the staff.

In the context of the evolution of CDM into CDS, the comparison between the current CDM state and a desired CDS end state were summarized as follow:

Typical CURRENT STATE focus	Potential TARGET STATE focus
Logical thinking (Output)	Critical thinking (Outcome)
Data integrity	Data quality
Quality Controls (QC)	Quality by Design (QbD)
Randomized controlled trials	Adaptive and master protocols
Focused on site generated data (Traditional EDC centric)	Focused on eSource data (i.e. DCTs and Patient Centric sources)
Standard processes across studies (one size-fits-all)	Risk-based processes tailored for each study (focus on what matters)
Low volume of data and sources	High volume of data and sources
Simple data flows	Complex data flows
Vendor management	Vendor oversight
Data validation	Data curation, validation, review, tagging, filtering, and exclusion
Patient data review	Patient data and audit trail reviews
Reviews of data after their collection	Risk-based and predictive data monitoring
Project Management	Cross-functional leadership
Clinical research standard	Clinical research and healthcare standards
Clinical research data	Clinical research and healthcare data
Traditional programming (SQL, C#, SAS, etc.)	AI/ML based automations (R, Python, etc.)

To support the evolution of the CDM Role[12], the SCDM innovation Committee has identified 4 core emerging best practices and published corresponding Topic Briefs that could be used as a reference:

The 5 Vs (Value, Veracity, Velocity, Variety and Volume) of clinical Data[13] supporting advanced clinical research and technologies

- Enabling new protocol designs such as adaptive and master protocol
- Enabling the decentralization of clinical trials
- Generating secondary data assets (e.g., synthetic control arms)
- Integrating various data structures through fit for purpose technologies

The evolution of Clinical Data Reviews[14] to ensure data quality (i.e., data reliability)

- Adapting to the decline of traditional EDC centricity and increase in data variety
- Focusing on the credibility and reliability of trial results
- Reviewing increasingly complex and study specific data

The automation of CDM Driven Activities[10] to drive process efficiencies and innovations

- Leveraging all components of Artificial Intelligence (AI)
- Maximizing machine to human interactions
- Enabling new capabilities such as identifying data anomalies without explicit validation checks programming

The adoption of Risk based CDM approaches[7] (i.e., risk prevention) aligned with recent regulations including

- Quality by Design (QbD), Quality Tolerance Limits (QTLs), Critical to Quality (CtQ) factors and Critical data and processes
- Risks lifecycle management including assessment, root cause analysis, etc.
- Ensuring operational feasibility

The list above could serve as a strawman, a framework to which organization can add their specific evolution requirements. Some organization specific examples may additionally include:

- Evolving the resourcing model (e.g., focus internal expertise toward future proof and/or partner facing activities and externalizing non-core activities),
- Evolving technology footprint (e.g., internal hosting vs. cloud, retiring/externalizing legacy footprint slowing innovation),
- De-risking the current state,
- Simplify, or optimize the current state to free-up capacity or address existing gaps (e.g., What capability gaps do we have? What capabilities do we need to augment?),
- Ensuring resiliency to emergency crises,
- Decentralized/hybrid trial execution capabilities,
- Addressing any current weaknesses such as readiness by First Patient First Visit (FPFV),
- Implementing committed CAPA(s),
- Meeting latest regulatory requirements.

In summary, the table below describes the required foundational framework according to GCDMP Chapters and CCDM Program as well as the Future Proofing Framework expanding the foundation according to the CDS Reflection papers.

Holistic CDS Framework	
Foundational Framework **CURRENT STATE**	**Future Proofing Framework** **ADDITIONS TO REACH THE TARGET STATE**
Best Practices	
Data Management Plan Paper CRF and EDC Lifecycle Vendor Selection and Management External Data Management Assuring and Measuring Quality Patient-Reported Outcomes Medical Coding	Risk-based CDM approaches *(Incl. Operational feasibility, QbD and CtQs)* Advanced Clinical Data Reviews *(incl. Story telling visualizations, systematic errors detection, Data Tagging, Exclusion and Curation)* Adapting to the 5Vs of Clinical Data *(Incl. Secondary data assets such a synthetic control arms and complex data Integration)* Automation of CDM Activities *(Incl. Robotic and Intelligent Process Automation)*
Soft Skills	
Logical thinking Adaptability Communication Ability to work with cross-functional teams Ability to troubleshoot complex data trends	Critical thinking Influential leadership Pragmatism Ability to manage ambiguities Ability to make and own decisions
Core Competencies Areas	
Design Programming Data Processing Project Management	Risk Management Vendor Oversight Patient Driven Technologies Deployment Process Management RPA and ML Based SDLC Advanced data exploration and interrogation *(e.g., audit trail, non-SQL, R, Python)*
Foundational Knowledge	
Clinical Development Methodology Regulations Software Development Life Cycle (SDLC) Audit Methodologies Standard Models and Terminologies Workflow Management	New protocol designs *(e.g., Master, Adaptive)* Decentralized Clinical Trials approaches Risk-based methodologies and regulations Health Care Data (RWD/RWE), Standard Models and Terminologies Emerging Data Structures *(e.g., Sensors)* Automation and Artificial Intelligence concepts

Required operating model to enable the target state capabilities

Additionally, when envisioning the target state, there are fundamental operational elements required to function in a healthy fashion and deliver the expected CDS capabilities. These may or may not currently exist but need to be considered within the definition of desired end state.

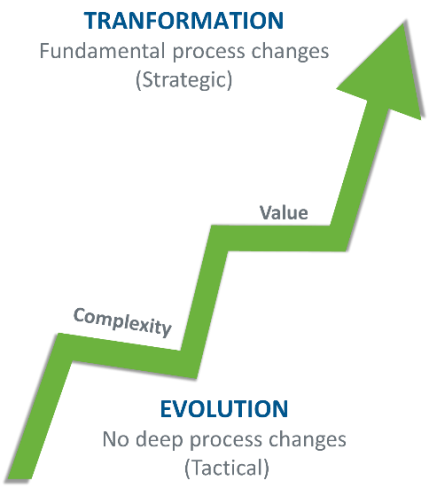
They include but are not limited to:

- A high-level **operating model** (i.e., how is work structured, disseminated, executed, and delivered).
- **Organization alignment** (e.g., by Therapeutic Area, Functional capabilities, geography, hybrid).
- A **resourcing model** (e.g., internal staff vs third party, on-shore vs off-shore resource proportions).
- Contingency, scalability, and resiliency strategy.
- Expected **key transformation outcomes** (e.g., Submissions without unanticipated data queries from regulators?)
- What are the most critical outputs that can be measured as leading metrics to directionally predict the probability of your intended outcomes?

Lastly, once the vision has been defined, it must be socialized and vetted with management and impacted organizations before initiating the concrete design stage. The vision should also describe what is the impetus for creating or evolving CDM into a CDS organization and the value proposition for the wider organization and CDM Stakeholders (i.e., There is a need for a business case with ROI).

C – The Design Phase (i.e., The roadmap from current to end state)

Only when the current CDM landscape is well understood and the CDS future state vision is clear, can an organization design its core transformation deliverable: **a tangible roadmap** to detail 'what', 'how' and 'when' to get to the target state. The vision serves as a blueprint for the leadership to design the roadmap which needs to cater for the current organization characteristics and consider the effort required to meet your future goals while resisting the tendency to solve all challenges with technology alone.



The extent of the roadmap will be proportionate to the distance between the current state and the desired target state. For some organizations, the roadmap will simply be a tactical **evolution** whereas others will require a much deeper and longer strategic **transformation**. The value proposition but also the complexity of the transition into the target state needs to be carefully assessed based on the organization ability to deliver the roadmap.

Additionally, one should **involve internal and external partners** in the design phase anticipating their contributions and dependencies during the implementation phase (i.e., dependencies related but not limited to Process, Systems, Portfolio, Services, and Budget). Reaching buy in at the design stage is critical to prevent later roadblocks.

The roadmap definition

When it comes to the roadmap itself, an effective way is to segment it by key deliverable themes, such as core characteristics of the future state of the CDS organization and develop the solution for each addressing the needs of **people, process, technology, and partnership**. Then organize each of them by "Delivery packages" with actionable phases, measurable goals, objectives, and discrete milestones that can be evaluated during implementation.

The CDM to CDS delivery themes in the roadmap may include but not limited to:

- **Foundational gaps** identified during the assessment phase that must be prioritized.
- **Organizational design:** leadership structure, onboarding plan, training, upskilling, and mentoring.
- **Operating model** including the evolution of technology and CDS services partnerships.
- **New CDS capabilities** that will become the new CDS Best Practices.
  - Next Generation Data Capture such as moving from traditional EDC centric to Patient Driven data capture (e.g., DCT Capabilities, eSource, Sensors, and Wearables),
  - Advanced data review capability (e.g., Storytelling visualizations, ML Based data issue detection and statistically driven pattern/outlier detection),
  - End to end risk based CDS processes.
- **Resourcing algorithms** aligned with the envisioned target state. This can be a complex undertaking to account for presumed efficiencies to be gained but critical to demonstrate financial/efficiency benefits (in addition to quality and speed) from current state and gain organizational support.

The next step is to define the delivery sequence considering all dependencies. If foundational elements are lacking, they must be prioritized. One may also require establishing stop gap measures in their current ecosystem first. If the organizational foundations are strong across the key dimensions defined in the holistic CDS Framework, then the approach can directly focus on innovation and optimization. Understanding your path is key to defining the **duration of the transformation** (i.e., defining how long it will take to move from the current state to the target state). The duration of the roadmap must be aligned to the complexity and readiness for change of the CDM organization. This is a key element which will heavily influence the success and failure of the change. If too long, it may not get the right management priority nor engagement from all stakeholders by not **generating some sense of urgency**. If too short, it may not be transformational enough or achievable (i.e., attempting to implement too many changes at once or not building foundational capabilities first). A typical transformation roadmap lasts between 3 and 5 years. To define the target duration, the following could be considered by design and any relevant gaps should have been identified during the assessment phase:

- What structural changes to the CDM organization are needed to align accountability and responsibilities to build the future-state CDS organization?
- Are there any higher management (e.g., increase portfolio by X% with Y% R&D budget evolution), dependent drug development (e.g., reach Z% of Clinical Trial Decentralization) or digital (e.g., reduce legacy system footprint) goals to align and/or to contribute to? Many organizations initiate regular transformation projects. Aligning to their timelines and key objectives would create momentum, ensure synergy and increase buy-in. This can also apply to critical study deliverables which may either need to be ring-fenced (i.e., excluded from transformation) or intentionally targeted (i.e., prioritized for transformation) depending on current CDM capability strengths and weaknesses.
- Are there timebound constraints such as budget cycle, system deployment targets and, CAPA commitments to consider?

- Are there any critical existing technology projects that would directly or indirectly impact your capabilities timeline?
- Is this a one-off change or an ongoing organization strategy? Change is a constant, as a good practice, some leaders maintain rolling roadmap to ensure ongoing organization alignment, anticipate the next evolution step and communicate a clear sense of direction to their teams and stakeholders.

Additionally, the path should address:

- The major steps and building blocks from one another to drive the necessary change (i.e., focused on building the correct base and fundamentals first),
- If critical gaps have been identified, assure they are prioritized appropriately,
- Identify low hanging fruits to show early results, build momentum and credibility,
- Key enablers (i.e., What capabilities are needed to support the operating model and to deliver),
- Delineate capabilities between existing, needing changes vs non-existing requiring creation,
- Dependencies (e.g., technology deployments, hiring, deliveries of other initiatives),
- Roles required to support the target operating model.

As an example, one of the CDM Leader contributing to this paper has been using a home grown 3 steps roadmap strategy called "SISTER": **S**implify, **I**mprove, **S**tabilize, **T**ransform, **E**xecute, and **R**efine:

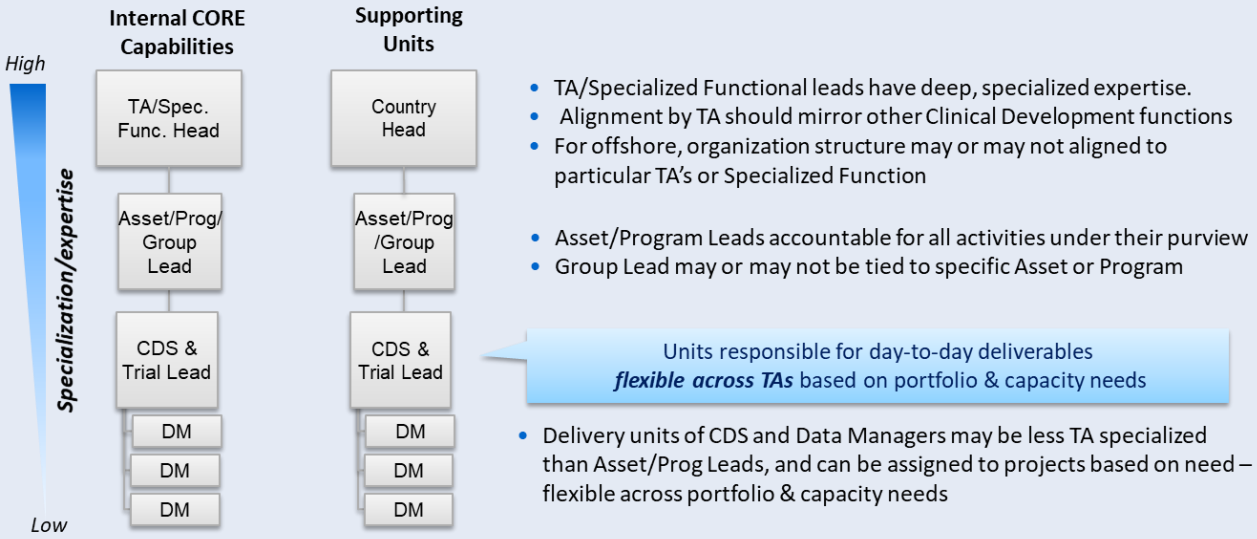
Step 1 (**S**implify and **I**mprove)	In this early stage the goal is to address the impactful low hanging fruit identified during the assessment phase and build momentum and credibility. In parallel designing and establishing the building blocks required for a deeper transformation.
Step 2 (**S**tabilize and **T**ransform)	This is a critical stage where the first improvements become a norm and where the transformational elements are piloted prior to ramp-up.
Step 3 (**E**xecute and **R**efine)	This is the stage where the process is refined based on the early live pilots and vetted before ramp-up. This is also where the implementation strategy is adjusted, if necessary, prior to the large-scale adoption phase.

The organizational design

Last, define fundamental organizational structure to deliver the roadmap. However this has to be done in conjunction with the operating model as one cannot design an efficient organizational structure without considering the end-to-end delivery model. Within the organization there are multiple activities required which could be executed by a varying number of individuals depending on the required scale of the organization. Smaller organizations will need to take a more generalist role approach while larger ones will tend to evolve towards specialist roles for economy of scale. Typical segmentation dimensions to consider include:

- Therapeutic Area (TA) alignment
- Function (e.g., CDM, Data Technology)
- Role (e.g., Clinical Data Scientist, EDC Builder, Clinical Programmer)
- Study team facing vs. support role
- Regional needs (e.g., Global vs. Local)

A sample large pharmaceutical company organizational framework is as follows:



Key Considerations:

- Despite flexibility, TA/Specialized Functional alignment intended to be sticky – colleagues more likely to move to new projects within the same TA to build team cohesion and specialized expertise
- Training and mentorship required for colleagues switching to a different TA/Specialized Function
- The more senior the colleague's role, the greater the stickiness to TA/Specialized Function

Regardless of the roadmap duration, structure, and content, it needs to be clearly articulated, easily communicated, time bound, and deliver a combination of quick wins and long-term benefits. So, once the roadmap has been set, it is advisable to create and socialize a high-level roadmap before venturing into too many details. To be credible and achievable, the path needs to be balanced. If the correct path is planned, one can avoid under or overbuilding.

D – The transition phase (i.e., how to implement the roadmap)

Once the previous phases have been completed, one can focus on the core aspects of transitioning into the future state solutions. The transition phase includes the conversion of the roadmap into a detailed and actionable implementation planning, change management, the transformation program close and the transition into the steady state. Change leadership plays an important role within the transition phase of the transformation, as real change occurs at middle management, ultimately being accountable within the line functions for the target state. Change leadership must be effective models for their teams, embrace and reinforce the vision, and help to navigate uncertainty and overcome obstacles or roadblocks.

Additionally, the best transformational programs are the ones that complete and end. Programs need to establish clear criteria for the "when" is the organization ready to transition from the implementation to steady state and when we have all required components to sustain the change.

Planning the transition phase

Depending on the complexity and extent of the transition, it may be advisable to **establish first, the leadership framework** which will be accountable for the new CDS organization when reaching the steady state. There is no substitute to accountability and ownership. Therefore, large transformations are typically more successful when the ultimate owners of the outcome are empowered and held accountable as Change Champions to drive the implementation of the roadmap and if possible, the design of the roadmap itself. Without empowerment there cannot be true accountability. Therefore functional leaders must empower operational managers to drive and monitor change. The Change Leadership team must be supported by a strong **Program Governance Model** (e.g., Steering Committee) with clear communication pathways for issues and escalations and a dedicated **Program Management Office** to assess progress and variance to the overall implementation and drive risk mitigation and issue resolution. The planning phase should be organized around the delivery themes defined in the roadmap starting by establishing or acquiring the **foundational building blocks** including:

- Establishment of the Program Governance Model and dedicated Program Management Office
- Defining accountabilities and ownership for implementing the transformation
- Early establishment of culture as a critical element
- Human resources strategies: Job classification, career ladders, talent acquisition, compensation, onboarding, training, upskilling, mentoring and evaluation
- Orchestrating the changes to the technology, processes, and operating models with all internal and external partners
- Defining a cross functional training plan for new processes and tools
- Defining the monitoring strategy and validate the success criteria for the transformation (i.e., what outcomes define success?)

Once the planning of the core elements is articulated, define the **transition strategy for clinical studies** considering the portfolio and book of work, with an emphasis on de-risking any pivotal trials or critical assets. Consideration for scope of applicability may include:

- Need for early adopter studies to vet changes first
- Implementation time point for new studies
- Need to keep and maintain two sets of procedural documents (e.g., Standard Operating Procedures, Work Instructions, Job Aids, …) for current and future state until completion of the transition
- Benefit risk for ongoing studies. Not all changes being equal, it is important to carefully assess the impact and applicability of the new practices for ongoing studies. For example, introducing new data collection methodologies such as DCT would present complexity in a study using traditional EDC processes. However, implementing ML based anomaly detection or RPA on top of legacy process would represent a low risk while bringing additional efficiency and increase quality
- Considerations for long duration study (e.g., Studies locking more than X years post CDS framework implementation) as the risks of grandfathering them may outweigh the risks of transitioning into the CDS framework
- Is there an additional value for ongoing studies in a specific TA and/or phase such as automating SAE collection and reconciliation in a large phase III study in a sick patient population?
- The transition/implementation plan needs to carefully consider external and internal drivers (from assessment) that could influence the plan. Everything from EDC licensing contracts to planned reorganizations that can impact the transition approach

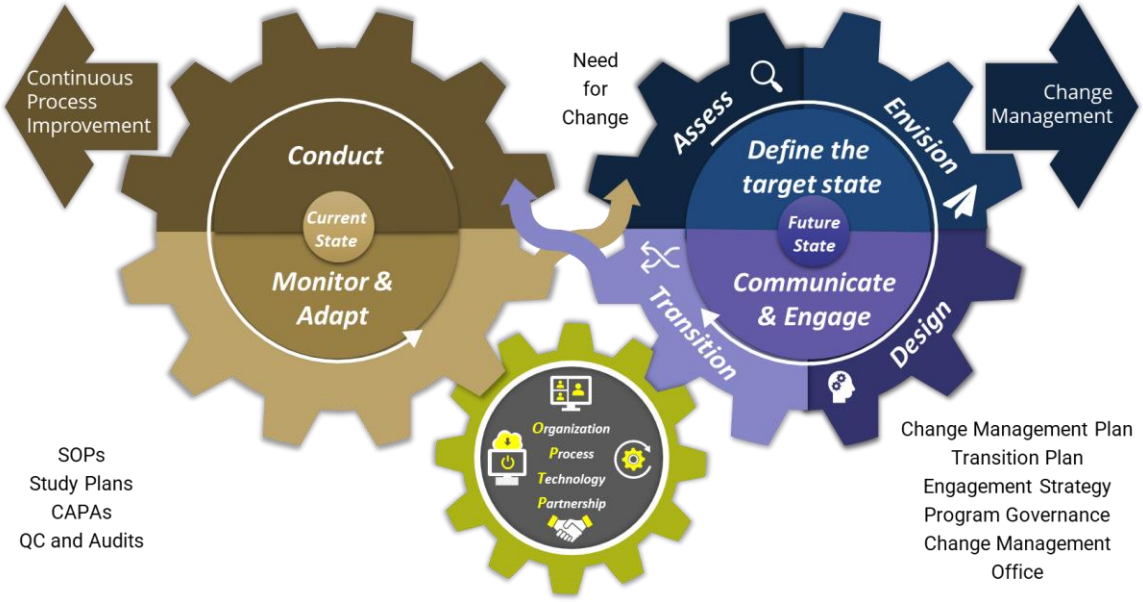
Additionally, there is a need to provide a clear and complete study specific transition guide to support the transition of ongoing studies. Some changes will lead to the update of existing documents such as the study Data Management Plan (DMP). Others such as assessing the risk benefit of applying the change may be better documented in a study specific transition plan. Regardless, it is wise to anticipate future audits and inspection questions by documenting all decisions and changes unambiguously.

Transition into the steady state

The transition phase will start when the roadmap has been converted into an actionable implementation strategy considering both the organization and study levels transitions.

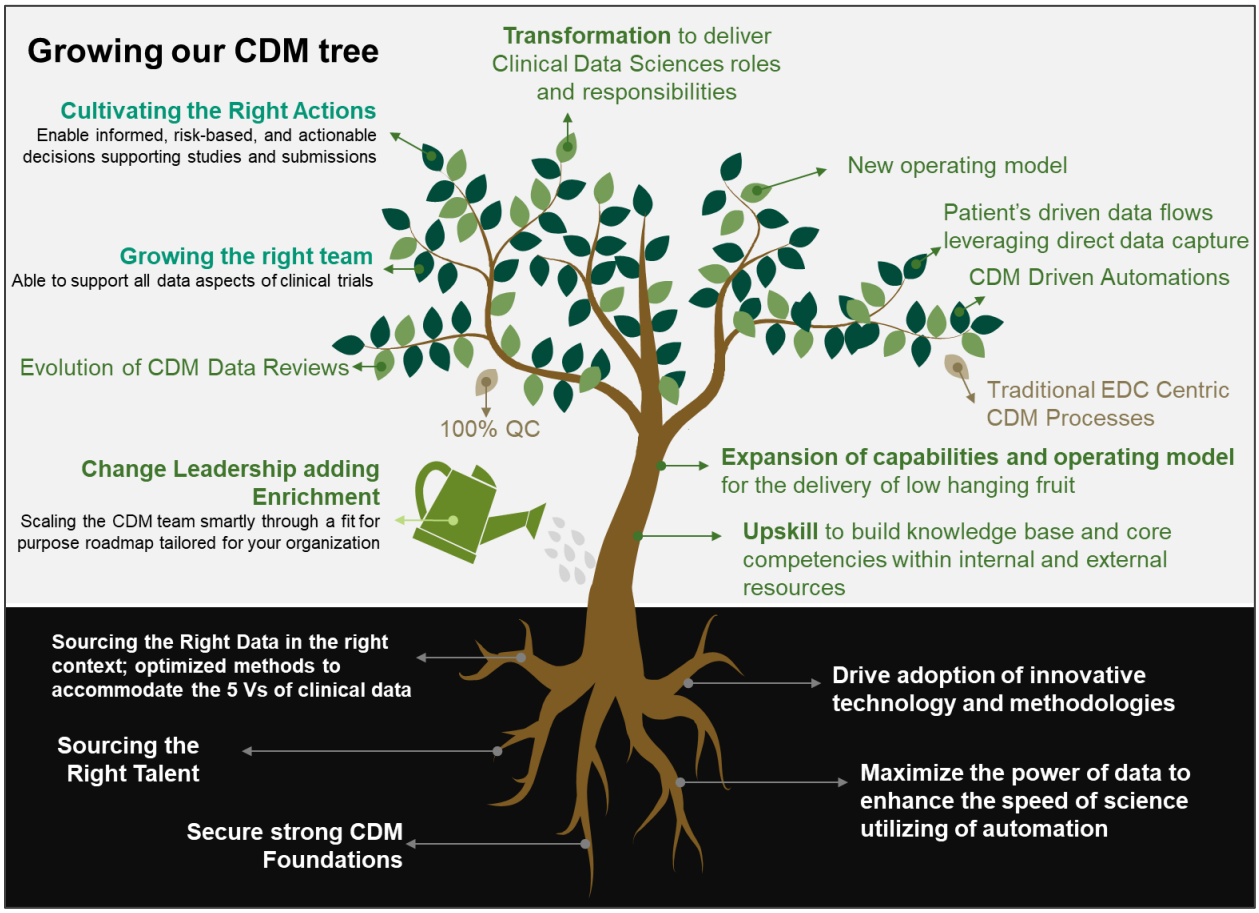
Transition steps may include the:

- Launch of the targeted training and upskilling programs for internal and external partners (e.g., CROs and FSP Partners) covering
  - Processes
  - Technologies
  - Regulations
  - Behavioral / Soft skills
- Launch and Support pilot studies if applicable
- Expand/scale-up from pilots and reinforce for broad deployment
  - Apply CDS practices to new studies
  - Fully or partially apply CDS capabilities to ongoing studies
- Demonstrate process "control"
  - Random checks of outputs or Risk-based, targeted quality checks
  - Continue leadership check-ins
  - Monitor for improvement opportunities
  - Monitor desired performance metrics
- Upon ramp-up, launch the continuous process improvement mechanisms
  - Learn and Adapt (Lessons Learned)
  - Monitor for industry trends and technology advancements
- Update your roadmap yearly and keep an eye on the next few years



Conclusion

The evolution into a CDS organization requires to grow our CDM tree through strong and decisive Change Leadership. First, we must ensure that our CDM foundations are deeply rooted before venturing into transformational CDS capabilities. Along the way, some old leaves of our CDM tree will have to fall (e.g., reliance on 100% QC) to allow new ones to grow (e.g., risk-based Clinical Data Reviews). Regardless of your journey and travel speed, there is no shortcut to a successful and sustainable transformation and all main phases must happen starting from the assessment of the current state to the implementation phase through the envisioning and design phase.



Main abbreviations

AI	Artificial Intelligence
CAPA	Corrective Actions and Preventative Actions
CCDM	Certified Clinical Data Manager
CDM	Clinical Data Management
CDS	Clinical Data Science
CRF	Case Report Form
CRO	Clinical Research Organization
CtQ	Critical to Quality
DCF	Data Correction Form
DCT	Decentralized Clinical Trials
DMP	Data Management Plan
EDC	Electronic Data Capture
FPFV	First Patient First Visit
FSP	Functional Service Provider
GCDMP	Good Clinical Data Management Practices
IPA	Intelligent Process Automation
KPI	Key Performance Indicator
KRI	Key Risk Indicator
M&A	Mergers & Acquisitions
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
QbD	Quality by Design
QC	Quality Control
QMS	Quality Management System
QTL	Quality Tolerance Limit
ROI	Return On Investment
RPA	Robotic Process Automation
RWD	Real World Data
RWE	Real World Evidence
SCDM	Society for Clinical Data Management
SDLC	Software Development Life Cycle
SISTER	Simplify, Improve, Stabilize, Transform, Execute, and Refine
SME	Subject Matter Expert
SWOT	Strengths, Weaknesses, Opportunities and Threats
TA	Therapeutic Area
UAT	Use Acceptance Testing

References

1 SCDM Good Clinical Data Management Practice Guidelines, Available at https://scdm.org/gcdmp/

2 SCDM Certification program for Clinical Data Managers, Available at https://scdm.org/certification/

3 SCDM, June 2019, The Evolution of Clinical Data Management to Clinical Data Science – A Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management. Available at https://cdn.fs.pathlms.com/a1AnYc9xSiu8GbVj4bHi?cache=true

4 SCDM, March 2020, The Evolution of Clinical Data Management to Clinical Data Science (Part 2: The technology enablers) – A Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management. Available at https://scdm.org/wp-content/uploads/2020/03/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science-Part-2.pdf

5 SCDM, August 2020, The Evolution of CDM to Clinical Data Science (Part 3: Evolution of the CDM role) – A Reflection Paper on the evolution of CDM skillsets and competencies. Available at https://scdm.org/wp-content/uploads/2020/08/SCDM-Reflection-Paper-CDM-Role-Evolution-Part-3-.pdf

6 Stokman and al, 2021, Risk-based Quality Management in CDM An inquiry into the value of generalized query-based data cleaning, Journal of the Society for Clinical Data Management, available at a https://www.jscdm.org/article/20/galley/40/download/

7 SCDM Topic Brief, July 2022, Risk-Based Clinical data Management, Available at https://scdm.org/clinical-data-science/

8 ICH, Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2). Available at https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf

9 ICH, ICH E8, General Considerations for Clinical Trials Draft Guidelines, Available at https://database.ich.org/sites/default/files/E8-R1_EWG_Draft_Guideline.pdf

10 SCDM Topic Brief, June 2022, The Automation of Clinical Data Management Activities, Available at https://scdm.org/clinical-data-science/

11 SCDM and eCF, Audit Trail Review, An Industry Position Paper on the use of Audit Trail Review as a key tool to ensure data integrity Available at https://scdm.org/wp-content/uploads/2021/04/2021-eCF_SCDM-ATR-Industry-Position-Paper-Version-PR1-2.pdf

12 SCDM Topic Brief, July 2022, The Evolution of the CDM Role, Available at https://scdm.org/clinical-data-science/

13 SCDM Topic Brief, March 2022, The 5Vs of clinical Data, Available at https://scdm.org/clinical-data-science/

14 SCDM Topic Brief, April 2022, The Evolution of Clinical Data Reviews, Available at https://scdm.org/clinical-data-science/