EDM Council Pharma / Life Sciences Global Data Management Benchmark Report

2021

EDMCouncil

OSTHUS

strategy&

Part of the PwC network

Photo by Alex Wong on Unsplash

Executive Summary

The EDM Council is proud to present the results of our 2021 Pharma / Life Sciences Global Data Management Benchmark Study. More than 20 firms participated in this first-of-its-kind global data management survey, including 7 of the 10 largest global pharmaceutical companies in the world.

Since 2015, the EDM Council has conducted Global Data Management Benchmark studies, measuring how organizations are handling their Data Management Programs. As with previous benchmark studies, the EDM Council conducted this year's benchmark study utilizing the DCAM—Data Management Capability Assessment Model, which provides the key dimensions to build and assess a firm's Data Management Program. New to the DCAM in 2020 was the addition of a Data Analytics component, which guides and assesses how organizations need to support their data analytics, AI, and ML programs—a growing concern as the scope and breadth of data, and the power of analytic tools, continues to grow in size and importance.

The role of the CDO in Pharma / Life Sciences

Out of all the Pharma / Life Sciences companies surveyed, roughly 75% now have a Chief Data Officer (CDO) function established. This indicates that the majority of organizations in the sector are already deriving value from data, trusting it to power their businesses. However, it is important to keep in mind that most CDOs have been established recently. For a number of companies, the function is less than one year old, which means it will take time for this role to stabilize.

This year's survey highlighted a disparity in reporting structures for these new CDOs. The survey revealed a mix of reporting lines—to the Chief Information Officer (CIO), the Chief Operating Officer (COO), and in some cases, directly to the Line of Business (LOB). Although CDOs in other industries are increasingly assuming responsibilities for their firms' analytics agendas, 43% of Pharma / Life Sciences companies surveyed said responsibility for analytics falls to LOB rather than the CDO (see page 36). Similarly, as more CDOs in other industries are assuming responsibility for Data Ethics, again, more than half of the companies surveyed assign Data Ethics responsibility to an office separate from the Office of Data Management.

Managing data in a highly regulated industry

In the increasingly regulated and supply chain-driven environment of Pharma / Life Sciences, it is crucial to build a foundation that supports the leveraging of data through analytics, while protecting sensitive data. Pharma companies are recognizing the need to establish data protection and security standards in alignment with regional regulations to leverage the insights from data and analytics, which has the power to not only improve operational efficiency, but also facilitate significant knowledge transfer and improve patient treatment and outcomes.

Industry Call To Action

Overall, the survey reveals that the Pharma / Life Sciences industry is stepping up its Data Management game by taking the following measures:

- **ESTABLISH A CDO FUNCTION TO HELP DERIVE BUSINESS VALUE FROM DATA**
- **EMPLOY ANALYTICS TO EXECUTE THEIR BUSINESS STRATEGY**
- **INVEST IN EXPANDING THEIR ABILITY TO LEVERAGE THEIR DATA**

The value of Data Management and Data Analytics is clear

This year's benchmark survey revealed that the components of a well-defined Data Management Program, e.g., data strategy, Data Governance, and Data Architecture, are well established in the Pharma / Life Sciences industry and are bringing concrete value their businesses.

The survey also revealed that the majority of the analytics programs of the companies surveyed were well aligned with the goals and objectives of their data programs and with their businesses. A sizable portion of respondents—45%—anticipate data and analytics will have the greatest opportunity to add value to Research and Development (see page 37).

According to a recent study performed by PwC, the global data and analytics market is forecast to grow to $90 billion by 2025, which is more than triple of its size in 2015 ($23 billion). Whereas previous growth in this area can be attributed to software development, future growth will be driven by cutting edge techniques in AI/ML, the internet of things (IoT) and Cloud computing, all critically dependent on accurate, timely and trusted data. According to the PwC study, Pharma—and, more broadly, Healthcare—are highly affected by these technologies.

More than 20 firms participated, including 7 of the 10 largest global pharmaceutical companies in the world.

The future looks bright!

Pharma / Life Sciences companies report exciting plans for the years ahead. Establishing and growing the role of Data Management and the responsibilities of the CDO encourages companies to take active responsibility for their data and to improve their organization's data literacy for all employees. Comparing Pharma / Life Sciences Data Management 'state of the state' to the Finance industry, the survey revealed that Pharma / Life Sciences companies' level of Data Management maturity compares favorably, and, in fact, is ahead of other companies in non-financial sectors. These developments, which have largely taken place within the last three to four years, are remarkable, and they show the fast pace at which Data Management is evolving within the Pharma / Life Sciences industry. We are excited to see the next steps in this journey.

I would like to thank our partners, OSTHUS and Strategy& (a part of the PwC network), for their assistance with the survey and the production of this report. And I would like to thank all those who participated in this survey, without which, we would not have been able to see and appreciate the progress and identify the areas for improvement that will continue to advance the Pharma / Life Sciences industry.

Sincerely,

John A. Bottega
President, EDM Council
July 2021

Executive Summary

Table of Contents

Executive Summary **2**
Table of Contents **4**
Survey Questions **6**
DCAM®: The Data Capability Assessment Model **15**
About DCAM 15
Scoring Scale 16
Demographics & Profile of Participants **18**
Data Management Program **21**
Analytics Program **34**
Data Management Drivers & Priorities **39**
Methodology 42
Survey Results & Observations **43**
1.0 The Data Management Strategy and Business Case 43
2.0 The Data Management Program & Funding 47
3.0 Business and Data Architecture 55
4.0 Data & Technology Architecture 60
5.0 Data Quality Management 65
6.0 Data Governance 70
7.0 Data Control Environment 77
8.0 Analytics Management 81
DCAM® Composite Scores **90**
2021 Benchmark Advisory Team **94**

Note: The 2021 EDM Council Pharma / Life Sciences Global Data Management Benchmark Report represents the opinion of the EDM Council, not individual members nor organizations. This analysis is intended for information purposes only.

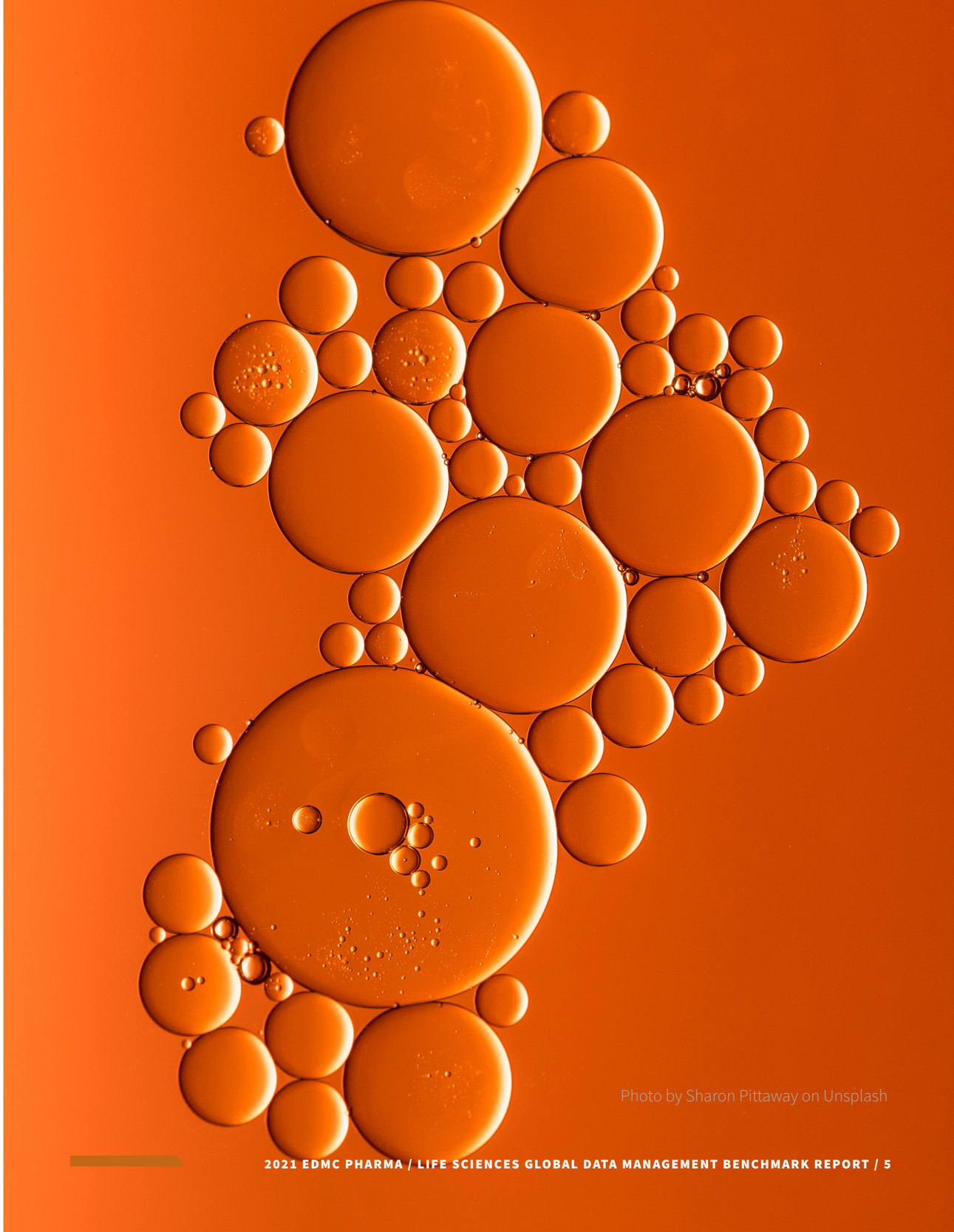


Photo by Sharon Pittaway on Unsplash

Survey Questions

SURVEY CAPABILITIES	DESCRIPTION		DCAM CROSS-REFERENCE	COMPONENT
1. A formal Data Management Strategy and approach has been developed and communicated to organizational stakeholders.	A Data Management Strategy formally exists and defines how the organization will approach the management of data content in a way that is meaningful to business stakeholders.		DCAM 1.1 and 1.3	The Data Management Strategy and Business Case The Data Management Strategy & Business Case determines how Data Management is defined, organized, funded, governed, and embedded into the operations of the organization. The strategy defines why the initiative is needed, how it will be implemented, and provides the rationale for the investment in the Data Management initiative.
2. A Data Management Business Case (including requirements and prioritizations) has been developed and communicated to organizational stakeholders.	The Data Management Business Case, describing the rationale for the data program investment, has been articulated and communicated to key organizational stakeholders.		DCAM 1.2	
3. Your organization has a formally established, structured and funded Data Management Program.	Data management program is formally established, funded and is operational.		DCAM 2.1, 2.2, 2.3	The Data Management Program and Funding Model The Data Management Program is a formally established organizational function created to support the proper curation and use of a firm's information assets. The Data Management Funding Model ensures that appropriate levels of funding and resources are applied to ensure program sustainability.
4. Formal plans, roadmaps and deliverables have been defined and articulated to program stakeholders.	Once established, the data program has defined and effectively communicated the program's objectives, deliverables and roadmaps.		DCAM 2.4	
5. Data Management Process Excellence is formally established.	Data Management Process Excellence is defined through the implementation of standardized, organization-wide processes that are repeatable, sustainable, measurable and auditable, and adhere to industry best practices.		DCAM 2.5	
6. Data Management Stakeholder Engagement is established and confirmed.	A broad set of stakeholders is required to effectively manage the data, including support from the business, technology, operations and cross-organizational control functions. A successful Data Management Program requires committed participation and resources across the organization.		DCAM 2.6	
7. Data management communications and training programs have been developed and implemented across the organization.	Internal communications and formal training are needed to affect the required organization-wide behavior and cultural change. Both are needed to drive awareness and adherence to the Data Management Program.		DCAM 2.7	
8. Metrics have been defined and are being captured to determine the success and continuous improvement of the Data Management Program.	Program metrics such as KPIs (Key Performance Indicators) and KRIs (Key Risk Indicators) are being captured and used to measure the success of the Data Management Program, as well as ensure continuous improvement.		DCAM 2.8	

Survey Questions

Survey Questions

SURVEY CAPABILITIES	DESCRIPTION		DCAM CROSS-REFERENCE	COMPONENT
9. Your organization has a formally established and supported Data Architecture Program (focus on the definition and modeling of data content).	A formally established Data Architecture function exists, focused on the design and definition of data content (its structure, definition, relationships, etc.).		DCAM 3.1	Business, Technology and Data Architecture Business Architecture is the strategy and design of efficient processes to support business objectives. Data Architecture defines the data. Technology Architecture defines the physical infrastructure. Together, all three must work in concert to have a successful Data Management Program.
10. Business Architecture is established and integrated into your Data Management Program.	Business Architecture determines the scope and requirements of data (what data is needed), considers all data restrictions and protections (e.g. use of PII – Personally Identifiable Information), and ensures that the right data is being used for the right business objective (appropriate use of data).		DCAM 3.2	
11. Logical data domains, models and metadata have been identified and documented and usage is supported by policy.	Data is defined through the identification and documentation of logical data domains. Logical data domains represent categories of data that are needed to run various business functions. For each domain, logical data models and definitions of data attributes are required, as well as a fully populated metadata (data describing data). Use and definition of data are then governed by policy and procedure.		DCAM 3.3 and 3.4	
12. Technology vision and strategy, in support of the Data Management Program, have been developed and documented.	The role of the Technology function is to define, design and implement the physical architecture needed to support business and data requirements. Technology defines the database strategies, analytics platforms, middleware solutions, storage and retention technologies, information security considerations, and all other aspects of the holistic technology infrastructure needed to support the Data Management goals and objectives.		DCAM 4.1	
13. Tools required to support the Data Management Program have been identified and implemented.	Data management tools (e.g. Data Quality tools, Data Governance platforms, metadata repositories, etc.), must be determined based on business and Data Management requirements. Technology is then responsible for designing the roadmaps, implementing the required tools, and ensuring appropriate governance is applied to the use of these tools.		DCAM 4.2	
14. Data Management Operational Risk planning is in place.	The Data Management governance structure must be in alignment with operational risk governance and engaged in the contingency planning and testing for data access and maintenance in the event of an operational disruption.		DCAM 4.3	

Survey Questions

Survey Questions

SURVEY CAPABILITIES	DESCRIPTION		DCAM CROSS-REFERENCE	COMPONENT
15. A Data Quality Management program is formalized and established.	A formal approach to Data Quality must be established within an organization. The goals, objectives and approach must be defined and communicated to all stakeholders. Dedicated time and resources must be committed to ensure data is fit for purpose.		DCAM 5.1	Data Quality Management Data Quality Management defines the goals, approaches and plans of action that ensure data content is of sufficient quality to support defined business and strategic objectives of the organization.
16. Data is profiled, measured, monitored and maintained.	Understanding the current state of your data is critically important to a successful Data Management Program. Profiling and measuring creates the Data Quality benchmark. Monitoring and maintaining the data ensures quality control points are implemented, DQ metrics are captured and continuous monitoring is established.		DCAM 5.2, 5.4	
17. Data Quality root cause analysis is routinely performed.	Data remediation plans must be developed and executed to resolve the most pressing DQ issues. The remediation must include both correcting the existing data and performing root-cause-fix to eliminate future data defects.		DCAM 5.3	
18. Data Governance function is formally established and operational.	Data Governance strategy and approach has been defined, communicated and reviewed and approved by program stakeholders. Roles and responsibilities have been assigned, project management, program funding and issue management are all operational.		DCAM 6.1, 6.3	Data Governance The Data Governance function is the backbone of a successful Data Management (DM) initiative. Data Governance is the process of setting standards, defining rules, establishing policy and implementing oversight. It is these steps that ensure adherence to DM best practices. Governance formalizes and empowers the DM initiative to ensure propagation and sustainability throughout the organization.
19. Policy and Standards have been written, approved and implemented.	DM Policy and Standards must be established for the organization and approved by stakeholders and executive governing bodies. The Policy and Standards must align with cross-control function Policy and Standards and be auditable.		DCAM 6.2	
20. Governance and maintenance of authorized data domains, data structures, data models, data definitions and data glossaries, is established and operational.	Governance of all related data structures (data models; data definitions; data glossaries), and the identification and governance of authorized data domains, has been established and is operational.		DCAM 6.4	
21. The access and use of data, driven by access controls, data-sharing agreements and contractual use of [market] data, is governed.	Governance is appropriately implemented in controlling the access and use of data, enforcing the contractual restrictions of third-party data, and establishing and monitoring adherence to the Data Sharing Agreements.		DCAM 6.5	
22. The ethical access, use and outcomes of data are considered, reviewed and governed.	Governing the Data Ethics includes establishing a formal Data Ethics oversight function, adhering to the ethical access and appropriate use of data, and monitoring whether the outcomes of data use adhere to established ethical standards.		DCAM 6.6	

Survey Questions

Survey Questions

SURVEY CAPABILITIES	DESCRIPTION		DCAM CROSS-REFERENCE	COMPONENT
23. A Data Control Environment, (established controls of data across the data lifecycle) and collaboration with cross- organizational control groups, is evident and operational.	Evidence of the Data Control Environment is the result of effectively integrating the capabilities of data strategy, program, architecture, Data Quality and Data Governance across the organization. Active engagement by stakeholders and cross-organizational control functions is required to ensure the success of the data program.		DCAM 7.1, 7.2	Data Control Environment The Data Control Environment refers to the state of operation in which the data assets of an organization are managed holistically throughout the organization. The purpose of the Data Control Environment is to coordinate the people, process and technology to ensure Data Management is part of the operating culture of the organization.
24. Risks associated with the access and use of data are being tracked, prioritized and mitigated.	The formal process of identifying data risk must be integrated into the Data Management initiative. Risks must be tracked, prioritized, mitigated and integrated into the overall risk management framework of the organization (e.g.: three lines of defense; risk; audit).		DCAM 7.3	
25. The Analytics function is established.	The Analytics function is formally established, with documented and approved strategy, operating model, and funding model, and with effective governance.		DCAM 8.1	Analytics Management The purpose of Analytics Management (AM) is to formalize how the Analytics activities of an organization are structured, executed, and managed and to ensure they are aligned with the DM activities. The degree to which Analytics teams are either centralized or distributed in an organization will depend on the structure and culture of the organization.
26. Analytics is aligned with business and Data Management Strategy.	The Analytics and DM functions must be aligned and together must support the business goals. Analytics must be prioritized to meet the needs of business strategy and drive business value.		DCAM 8.2	
27. Analytics is aligned with Data Architecture.	Analytics uses approved business glossaries and follows the organization's standards for data sourcing, identification, classification, and metadata. These standards are reflected in a data preparation standard adopted by Analytics.		DCAM 8.3	
28. Analytics is aligned with Data Quality.	Data Quality measures produced by Data Quality Management to ensure that data used by Analytics is fit-for-purpose. Data Quality issues discovered by Analytics are managed using the Data Quality Management framework.		DCAM 8.4	
29. The Analytics platform is designed and operational.	The Analytics platform meets the needs of the Analytics operating model, addressing the different requirements for innovation and production. It implements a version control regime and supports defined strategies for data obfuscation. Scalability requirements are understood and supported.		DCAM 8.5	
30. Model operationalization is established.	Model testing, approval, release, and regular review processes are in place and are aligned with Data Ethics and privacy governance. Model bias is managed and requirements for model explainability are understood and supported.		DCAM 8.6	
31. The Analytics culture and education needs are managed.	The desired behaviors for an Analytics culture are understood and initiatives are in place to address culture gaps. Education initiatives exist to address gaps in the skills required by Analytics practitioners.		DCAM 8.7	

Survey Questions

Photo by Bill Oxford on Unsplash

DCAM®: The Data Management Capability Assessment Model

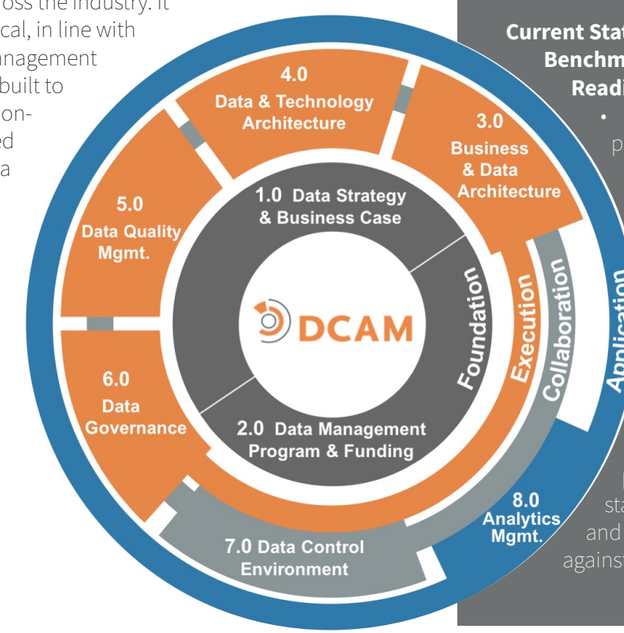
About DCAM

DCAM® (the Data Management Capability Assessment Model) is the industry-standard, best practice framework designed to assist today's information professional in developing and sustaining a comprehensive Data Management & Analytics Program. DCAM is organized into eight major components.

Each component is defined by a series of required Data Management & Analytics capabilities with each capability supported by a series of sub-capabilities, objectives, implementation advice and sample artifacts of evidence needed for verification of achievement.

DCAM, first published in 2014, was developed and continues to evolve based on the collective experiences of hundreds of Data Management & Analytics professionals.

DCAM is the synthesis of research and analysis of Data Management & Analytics practitioners across the industry. It was designed to be practical, in line with core principles of Data Management and Analytics. DCAM was built to be easily understood by non-specialists, while organized and structured for the data professional, to enable the implementation and continued improvement of a successful and sustainable data management program.



How Is DCAM Applied?

Firms are using DCAM in a variety of ways throughout the Data Management & Analytics Program lifecycle, including these use cases:

Program Initiation

- DCAM serves as a guideline to help launch new Data Management & Analytics Programs as well as align existing programs to best practice.
- DCAM provides an established set of criteria for sustainable Data Management & Analytics implementation and growth.

Current State Assessment, Benchmarking, & Regulatory Readiness

- DCAM assessments provide demonstrable and auditable evidence to market authorities on the adoption of Data Management & Analytics best practice.
- DCAM directly maps to data privacy regulations, such as GDPR.
- Firms are using DCAM to evaluate their programs, educate their stakeholders, identify gaps, and compare their progress against industry peers.

To learn more about DCAM, visit **edmcouncil.org**.

How Is DCAM Scored?

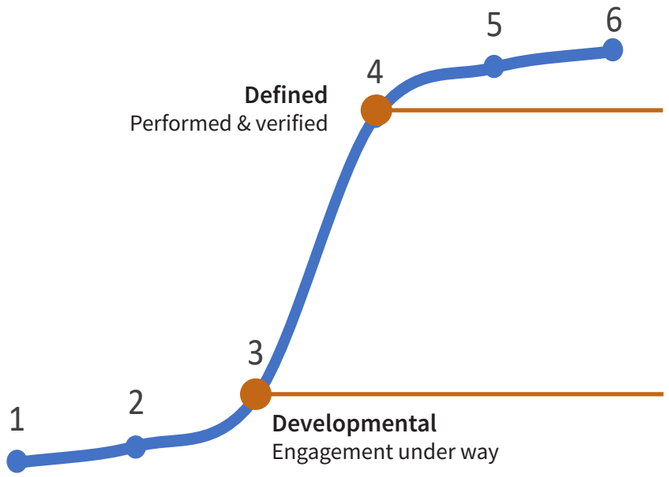
DCAM is structured to define and measure Data Management & Analytics capability. The singular goal of DCAM is the achievement of the requirements necessary to develop, implement and sustain an effective Data Management & Analytics Program.

Each requirement clearly defines the rationale for the specified capability and the dependencies that link the components of Data Management & Analytics into a cohesive program. DCAM is linked to a scoring matrix developed by the EDM Council to evaluate achievement of capability from three critical dimensions:

Engagement to ensure that the right people with the appropriate levels of authority are participating in the Data Management Program

Process to measure the degree to which Data Management processes are established, structured and repeatable

Evidence to provide the business artifacts that are necessary to audit against each capability statement



Scoring Scale

1	Not Initiated	**Ad hoc Data Management** Performed by heroes
2	Conceptual	**Initial planning activities** White board sessions
3	Developmental	**Engagement under way** Stakeholders being recruited and initial discussions about roles, responsibilities, standards, and processes
4	Defined	**Data management capabilities established and verified by stakeholders** Roles and responsibilities structured, policy ad standards implemented, gloassaries and identifiers established, sustainable funding
5	Achieved	**Data management capabilities adopted, and compliance enforced** Sanctioned by executive management, activity coordinated, adherence audited, strategic funding
6	Enhanced	**Capabilities are integrated fully into the operating culture of the organization**

DCAM®

EDMCouncil

SOLVE DATA MANAGEMENT & ANALYTICS CHALLENGES, TODAY & TOMORROW

There is no shortage of opportunities for data management & analytics professionals striving to advance their firms through automation, big data, and advanced analytics. At the same time, firms must address emerging regulatory and data ethics challenges. Through best practices and data standards, the EDM Council collaborates with its members to advance cross-industry Data Management & Analytics Programs and increase the business value of data assets.

Become an EDM Council member and gain access to several benefits:

DCAM CDMC EDMCouncil Cloud Data Management Capabilities

Implement Data Management & Analytics Best Practices & Data Standards with DCAM & CDMC (Cloud Data Management Capabilities).

eLEARNING EDMWebinar

Access E-learning, training, webinars, and certification for staff and executive management.

EDMConnect

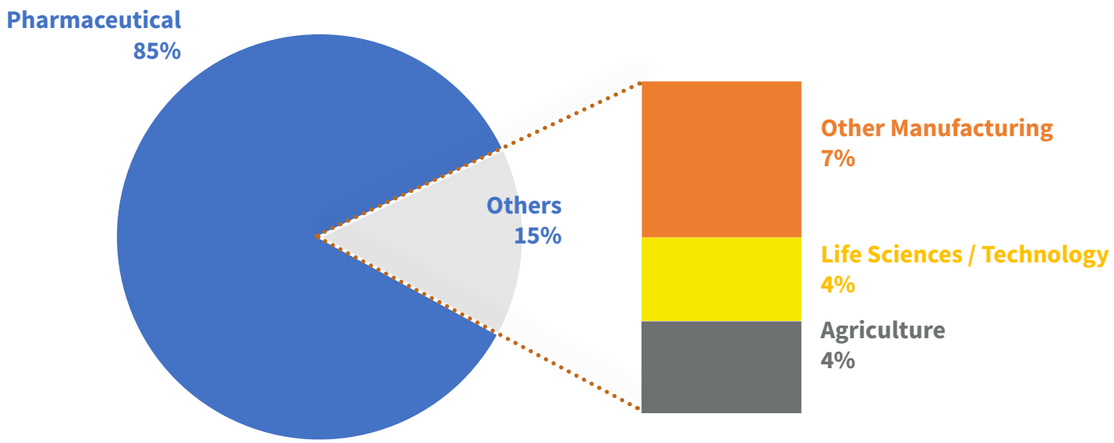
Network with 250+ global member firms and 10,000+ data management professionals via EDMConnect, our online membership community.

Data ROI EDMCouncil ESG EDMCouncil DataVision DATA ETHICS & RESPONSIBLE AI

Engage with interactive working groups, such as Data ROI, ESG Data, Women Data Professionals; DataVision conferences, and thought leadership opportunities.

Photo by NASA on Unsplash

Demographics & Profile of Participants

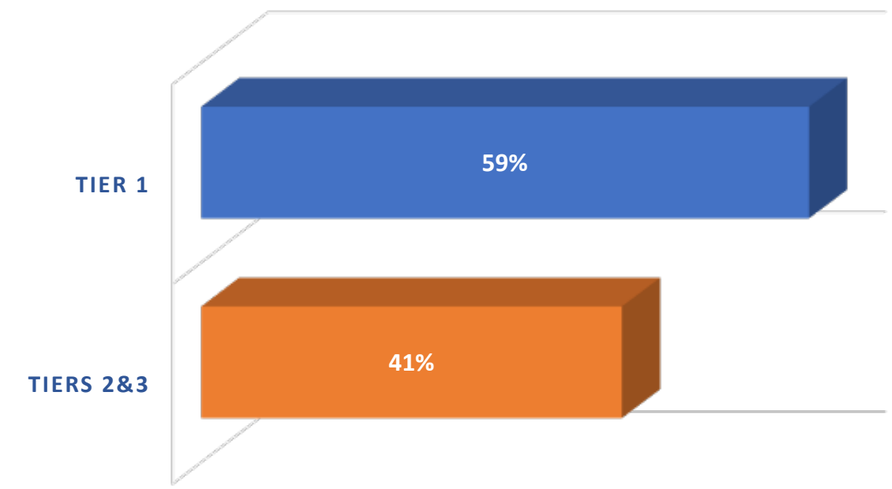


INDUSTRY PARTICIPATION

The 2021 EDMC Pharma / Life Sciences Data Management Global Benchmark Survey comprises responses from pharmaceutical (85%) and other life sciences (15%) organizations, all of which have a global scope of operation.

ORGANIZATION TYPE

Data Management Programs are present in both Tier 1 and Tier 2 & 3 firms.



TIER 1: REVENUE > $10B

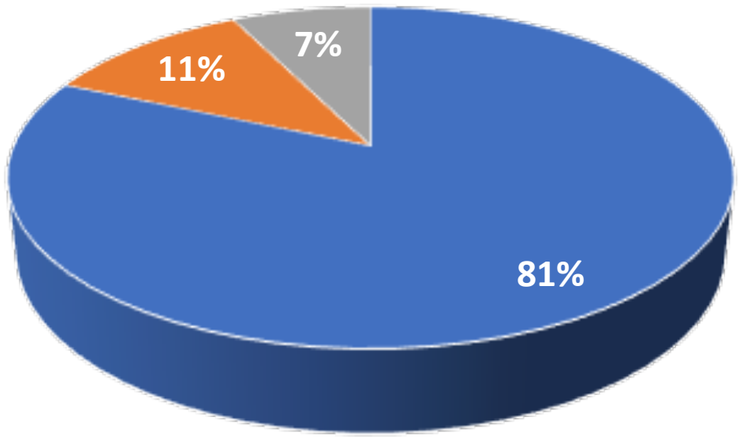
TIER 2: $100M < REVENUE < $10B

TIER 3: REVENUE < $100M

Demographics & Profile of Participants

Photo by National Cancer Institute on Unsplash

Data Management Program



- Global / Group (Enterprise-wide)
- Line of Business
- Regional / Country

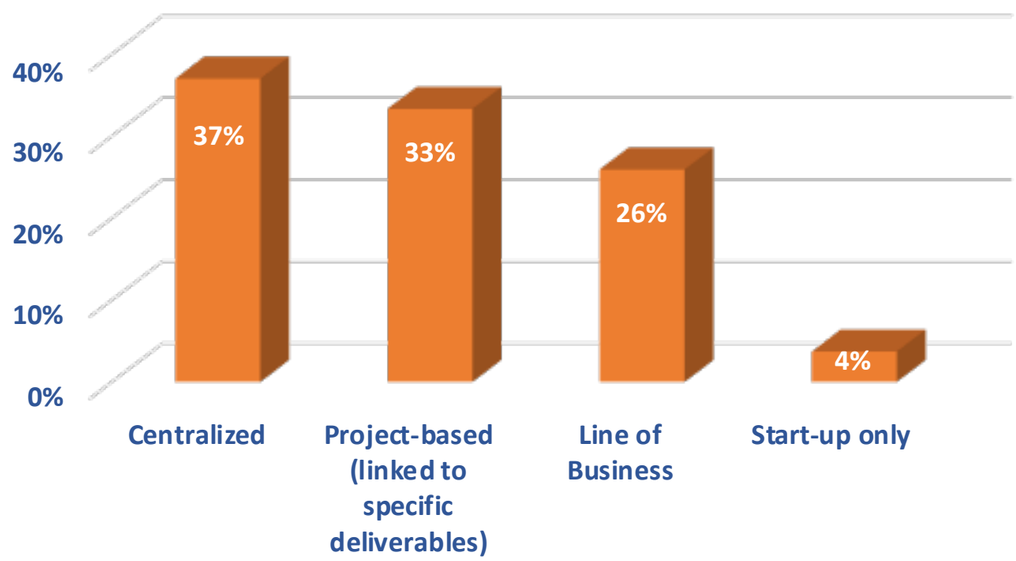
PROGRAM SCOPE

Eighty-one percent of Data Management Programs operate on an enterprise-wide basis.

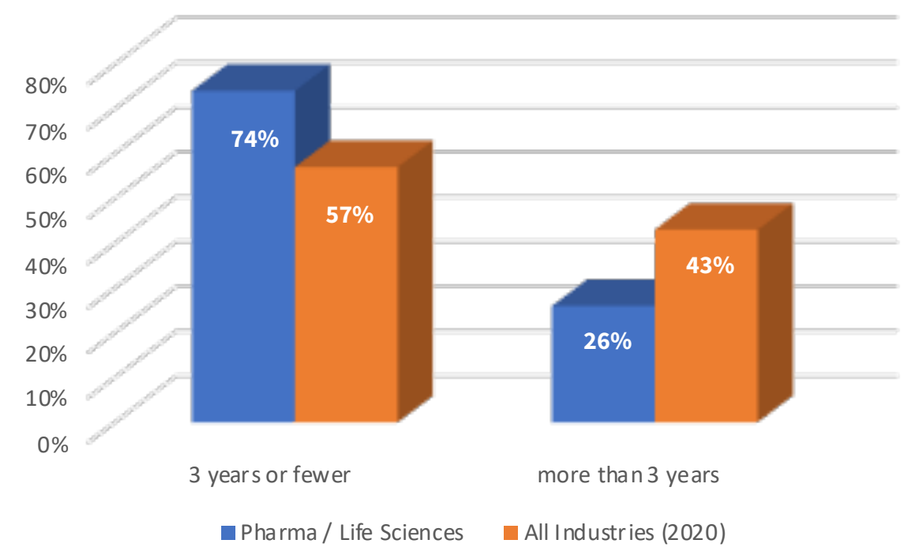
Data Management Program

FUNDING

There is no dominant approach to funding Data Management Programs. Thirty-seven percent of the survey respondents say their programs are centrally funded versus LOB / Project or Start-up.



THE PHARMA / LIFE SCIENCES INDUSTRY STEPS UP ITS DATA MANAGEMENT GAME



PROGRAM AGE

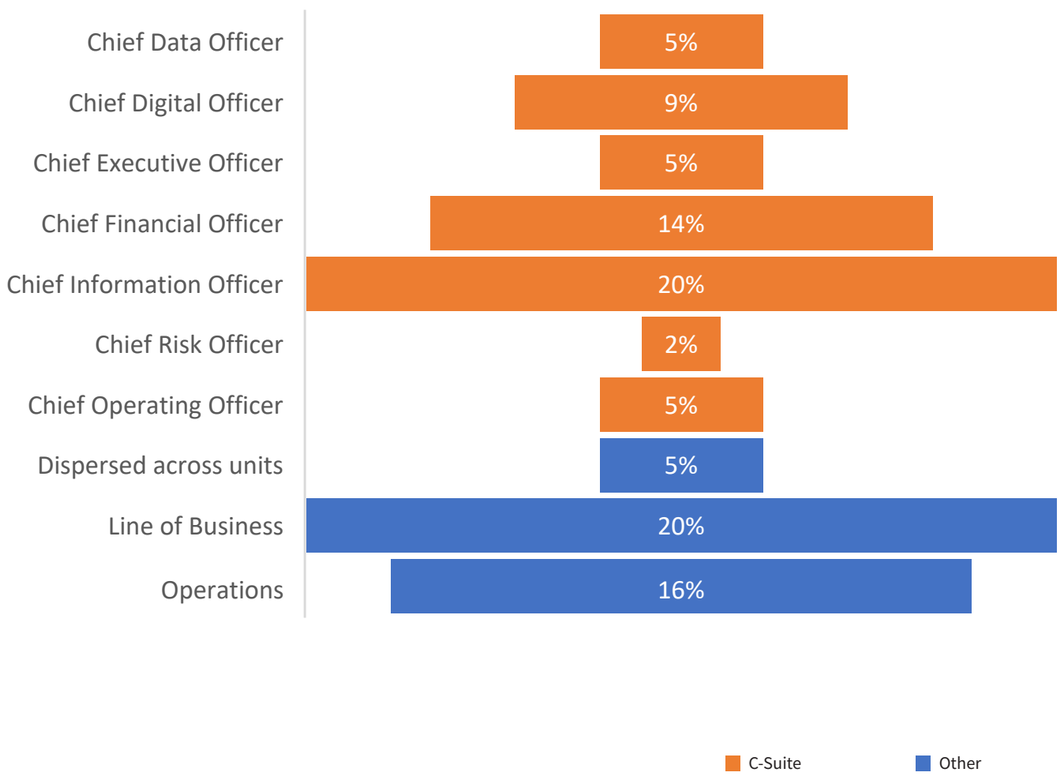
A majority of Data Management Programs were kick-started in the past three years. In the Pharma / Life Sciences industry, 74% of programs were initiated in the past 3 years. Only 26% of programs have been in place for 3 years or more in the Pharma / Life Sciences industry, which is less than the 43% observed for all industries in the 2020 survey.

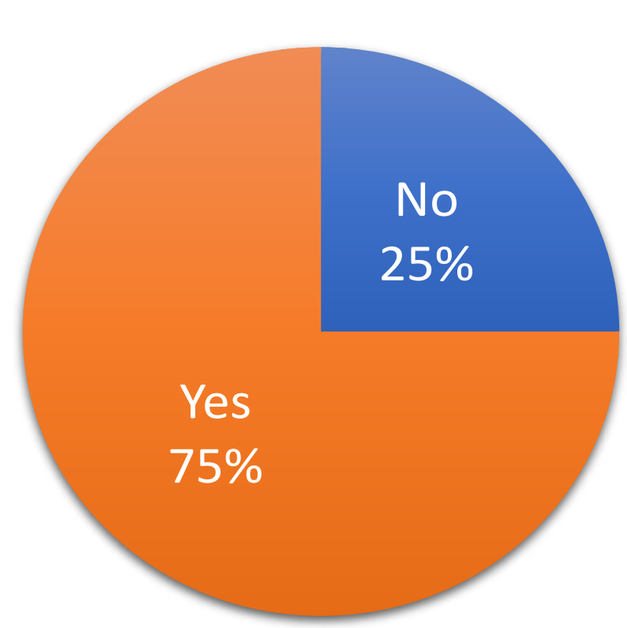
Data Management Program

Data Management Program

REPORTING

A majority of Data Management Programs report to the C-Suite. Fifty-nine percent of Data Management Programs report to the C-Suite in the Pharma / Life Sciences industry. In all industries, C-Suite reporting was 76% in 2020.





CHIEF DATA OFFICER

Seventy-five percent of those surveyed have a Chief Data Officer.

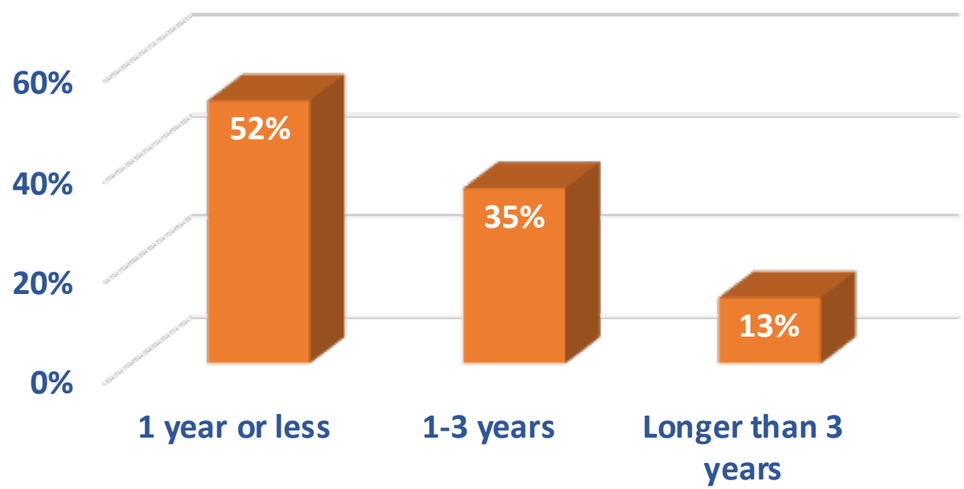
Data Management Program

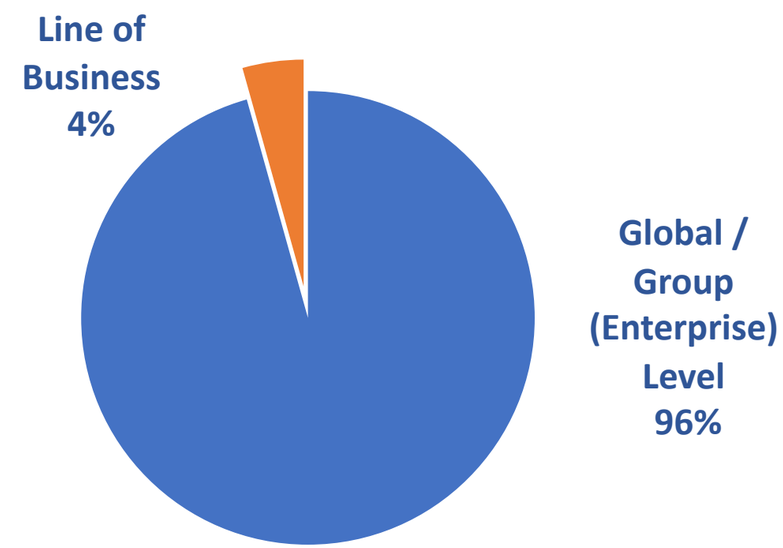
Data Management Program

TENURE

Many Pharma / Life Sciences companies are quite late to the game, with only 52% having hired a Chief Data Officer in the past year.

Forty-eight percent of Pharma / Life Sciences companies have had a Chief Data Officer for longer than a year, in contrast to 88% for the Finance industry and 65% for Other industries observed in the 2020 report.





SCOPE OF RESPONSIBILITY

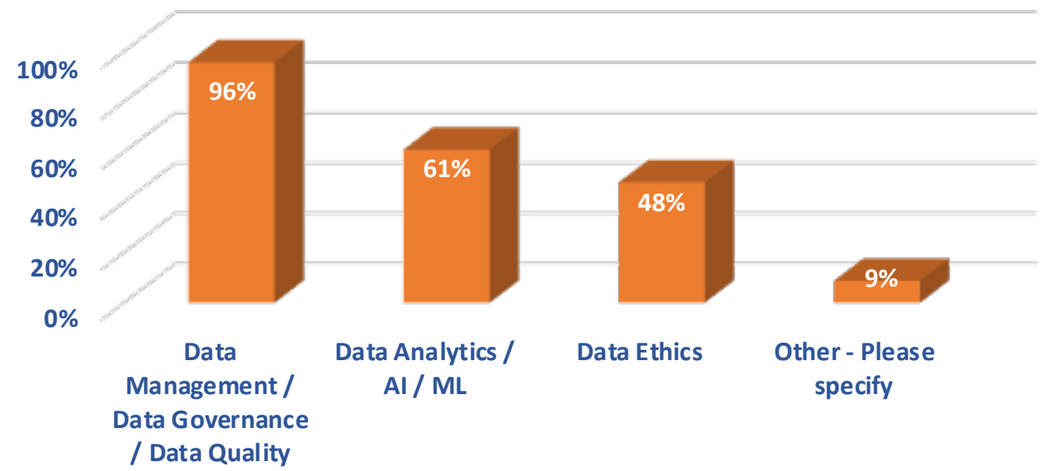
More Chief Data Officers are assuming global or enterprise roles. Ninety-six percent of organizations with a Chief Data Officer classify their roles as 'global.'

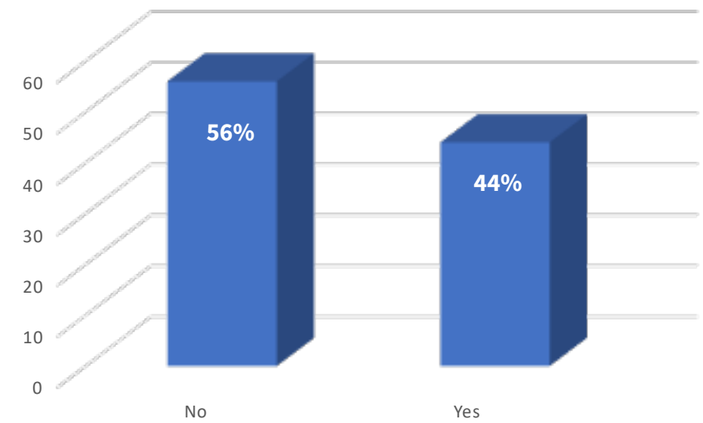
Data Management Program

Data Management Program

EMERGING RESPONSIBILITIES

Similar to the findings of the 2020 survey, 96% of Chief Data Officers are responsible for Data Management (Data Governance and Quality). Data Analytics and Data Ethics are again considered a crucial part of their role (compared to 52% for Data Analytics and 42% for Data Ethics in the 2020 survey).





DATA MATURITY MODEL USE

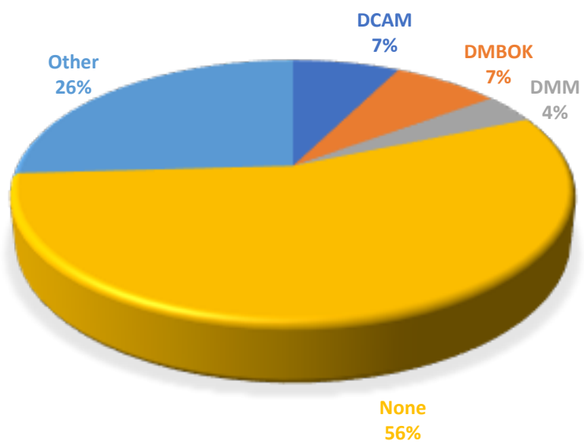
Forty-four percent of respondents state they are using a data maturity model for their programs.

Data Management Program

Data Management Program

MODEL TYPE

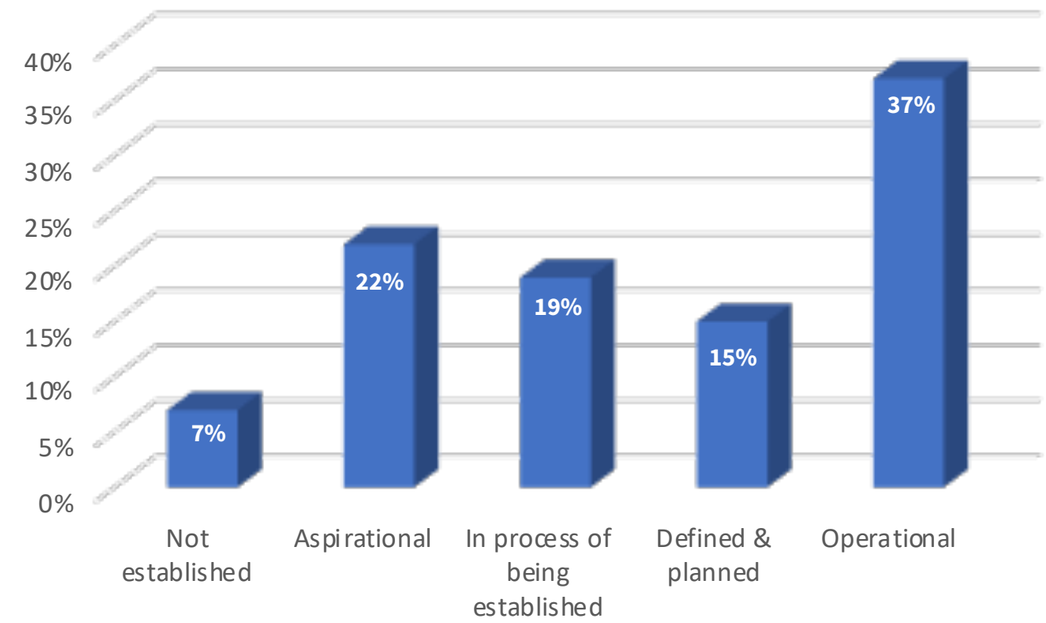
Forty-four percent of organizations surveyed report using a data maturity model, but only 18% of respondents use standard maturity models (DCAM, DMBOK, CMMI). Twenty-six percent use hybrid models developed in house and/or based on standard models.



DMBOK: Data Management Book of Knowledge—DAMA

DCAM: Data Management Capability Assessment Model—EDM Council

DMM: Data Maturity Model—Carnegie Mellon



FORMAL ANALYTICS & BIG DATA PROGRAMS

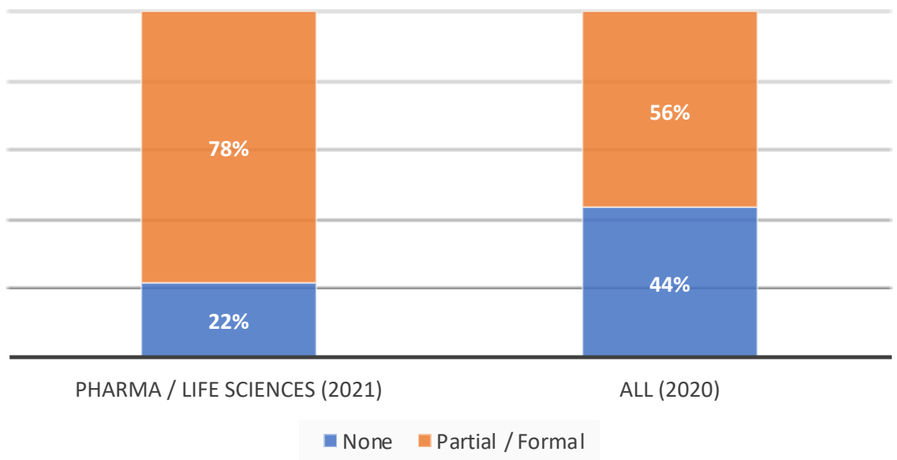
The 2021 Pharma / Life Sciences survey shows that 52% of respondents already have an Operational / Defined analytics and big data program, compared to only 41% in the 2020 survey. This indicates the importance of analytics for scientific endeavors in the Pharma / Life Sciences industry.

Data Management Program

Data Management Program

DATA MANAGEMENT OVERSIGHT FOR ANALYTICS

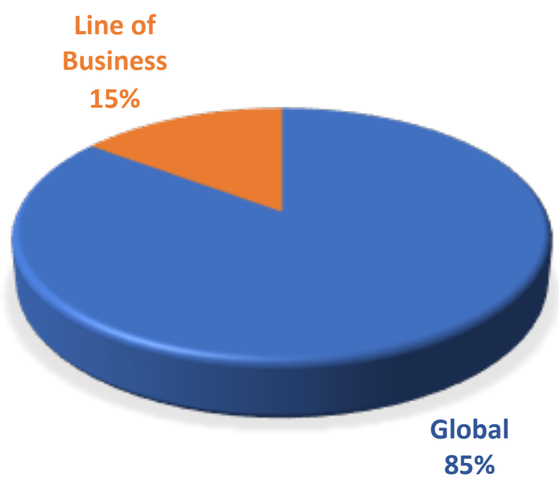
For 78% of companies in this survey, Data Management functions have either partial or full oversight over their analytics programs. This is considerably higher than the 56% observed for all industries in the 2020 survey.



Data Management Program

Photo by Jaron Nix on Unsplash

Analytics Program

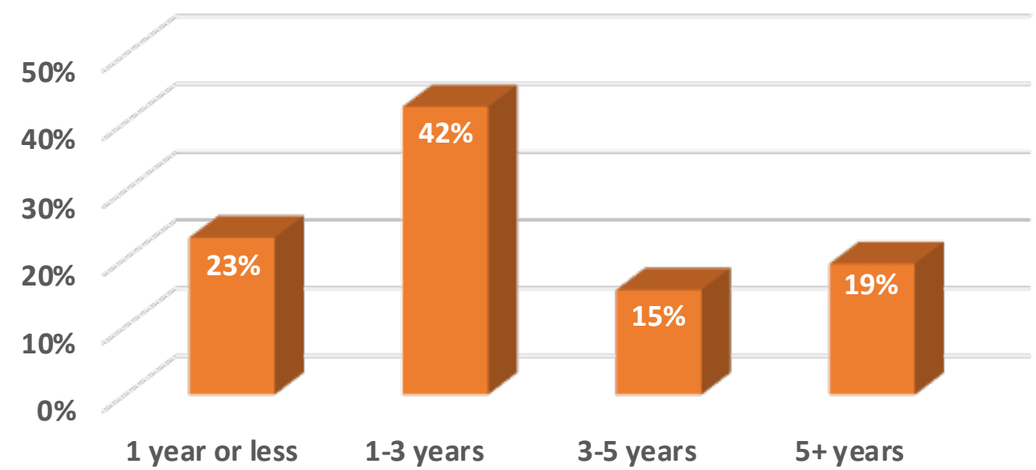


SCOPE OF ANALYTICS PROGRAM

The majority of respondents—85%—have an analytics program that is global in scope.

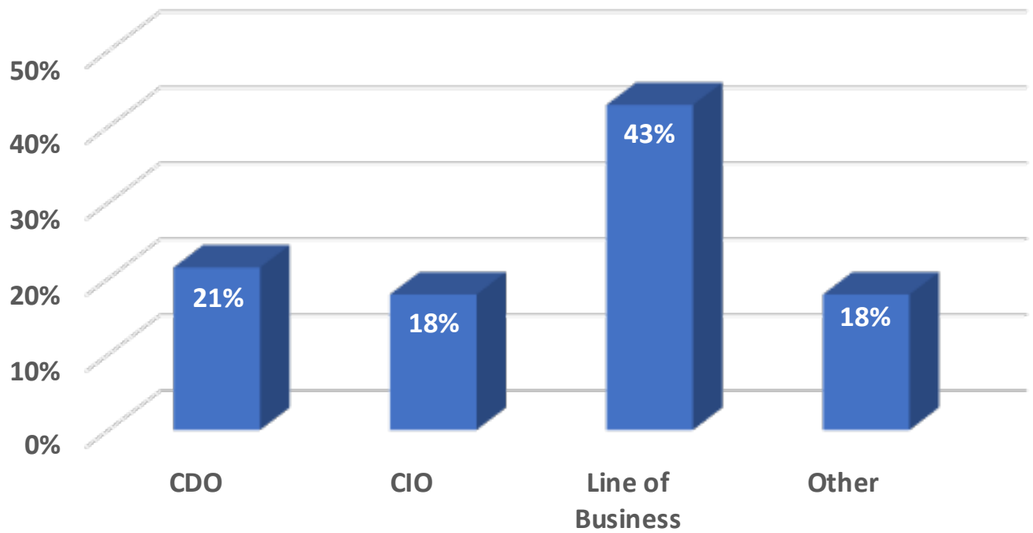
ANALYTICS PROGRAM YEARS IN OPERATION

The need to structure Analytics activities into a formal program has been recognized only recently by the majority of companies. Sixty-five percent of programs have started within the last three years.



Analytics Program

Analytics Program

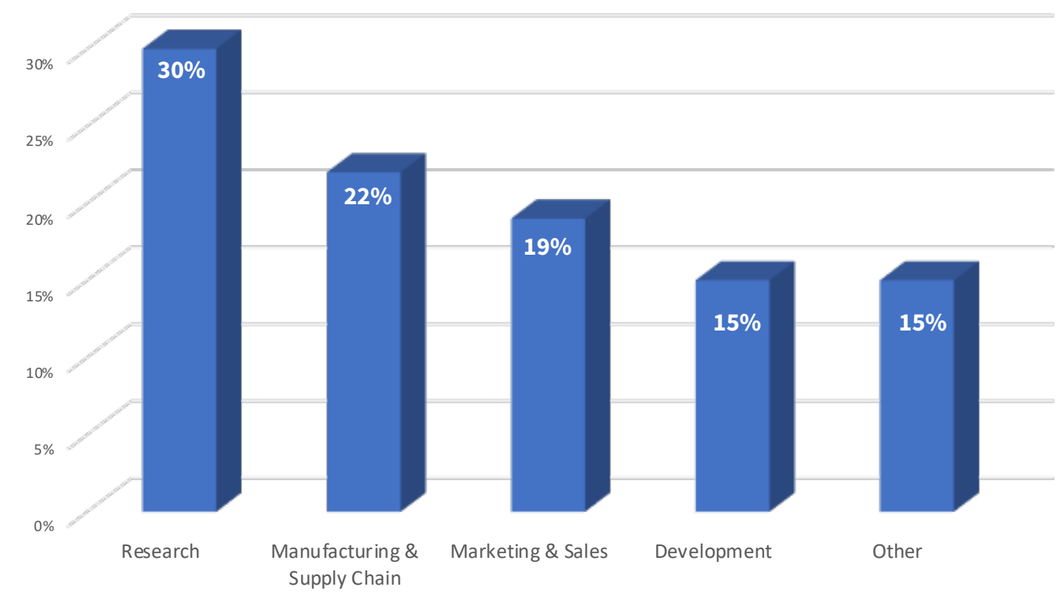


ANALYTICS PROGRAM RESPONSIBILITY & REPORTING

For 43% of the survey participants, the Analytics program reports into the Line of Business (LOB). Another 21% of respondents say Analytics reports to the CDO, and 18% to the CIO.

ANALYTICS PROGRAM OPPORTUNITY

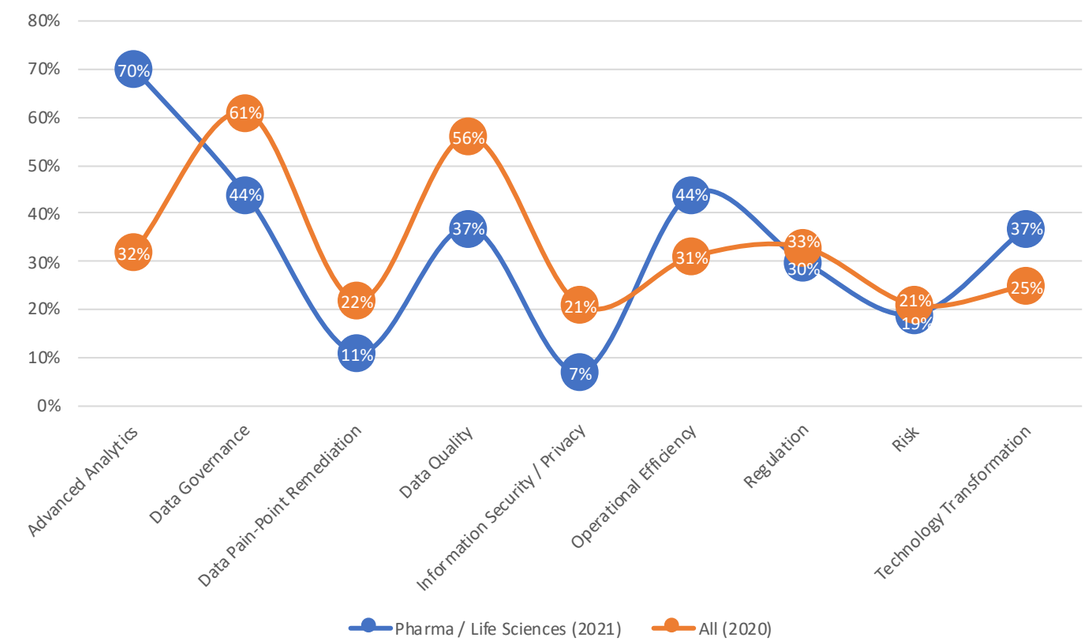
Survey respondents were asked which areas of activity presented the greatest opportunity for Data & Analytics in the next five years. They reported that the greatest opportunity arises from Research (30%) and Manufacturing & Supply Chain (22%).



Analytics Program

Photo by Obi Onyeador on Unsplash

Data Management Drivers & Priorities

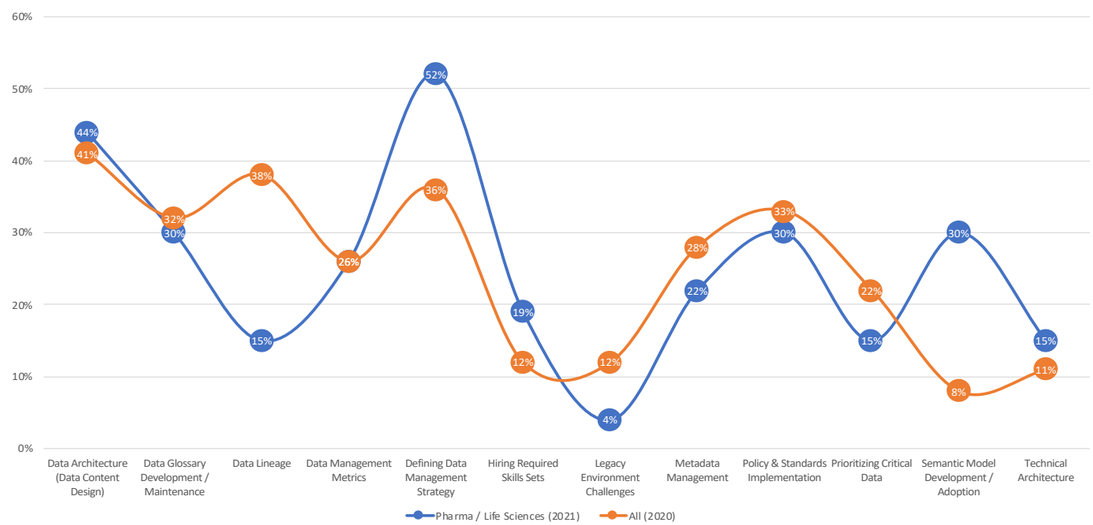


TOP BUSINESS DRIVERS

Respondents provided their top three drivers of (or, motivations for) Data Management. Advanced Analytics was named by 70% of participants in the 2021 survey. In comparison, Advanced Analytics was listed as a top business driver by only 32% of respondents in the 2020 survey. Similar to the 2020 survey, Data Governance continues to rank as a top driver in the 2021 survey. Rounding out the top three drivers in the 2021 survey is Operational Efficiency at 44%.

TOP DATA MANAGEMENT PRIORITIES

As for respondents' top three Data Management priorities, last year's leader—Data Architecture—has moved down a spot to make way for Defining Data Management Strategy, the top priority for the 2021 survey. Despite the drop in rank, Data Architecture was named by a slightly greater proportion of respondents—44%—in 2021, compared to 41% in 2020. Policy and Standards Implementation also ranks highly among the top three priorities in both years. In contrast to the findings of the 2020 survey, in which Data Lineage and Data Glossary round out the top five priorities, the 2021 survey respondents gave higher priority to Semantic Model Development / Adoption and Data Glossary Development / Maintenance.



Data Management Drivers & Priorities

Methodology

In this survey, the EDM Council focused on ensuring the participation of the top Pharma / Life Sciences companies by market capitalization. Twenty-one companies responded to the survey, including 7 of the world's top 10 Pharma / Life Sciences organizations by revenue.

In some cases, one-on-one interviews were conducted with Heads of Data Management & IT, Enterprise Architects, and Chief Data Officers, among other leadership roles. This provided a diverse range of perspectives.

Multiple responses were received from four companies. In these cases, the company's score for each question was calculated as the mean of the responses.

The next section of this report contains tables and charts describing the 'Average DCAM Score' and '% Achieved' for each of the survey questions. The individual DCAM scores provided by the respondents were combined into an Average DCAM Score. The percentage of responses with scores of 5 or 6 for each question determined the % Achieved. These aggregate scores and individual score distributions are used to compare the 2021 Pharma / Life Sciences Industry scores with Finance and Other Industry scores as reported in the 2020 Global Benchmark Report.

When referring to an organization's maturity, we combined the responses of 'Not Initiated' and 'Planning' and the responses of 'Defined' and 'Achieved' to create a clear visualization of the comparisons across industries, with the data presented for every response included below the chart.

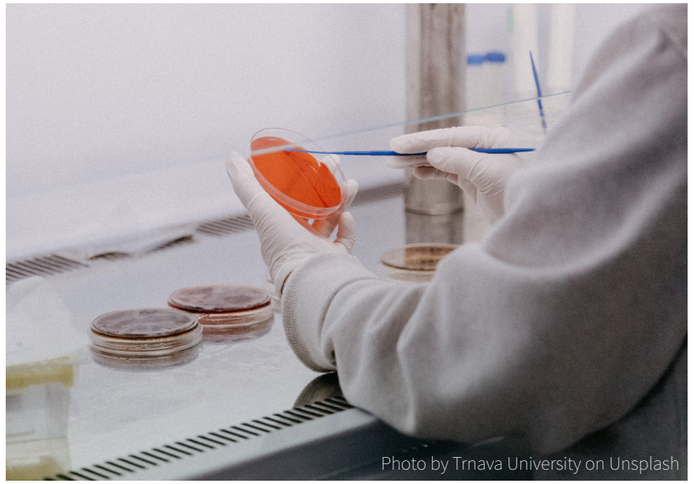
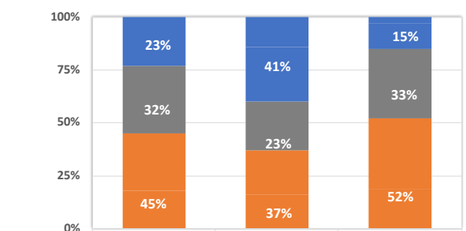


Photo by Trnava University on Unsplash

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.47	18%
Other (2020)	3.00	16%
Finance (2020)	3.52	19%



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	15%	3%
Planning	23%	26%	12%
Under Way	32%	23%	33%
Defined	27%	21%	33%
Achieved	18%	16%	19%

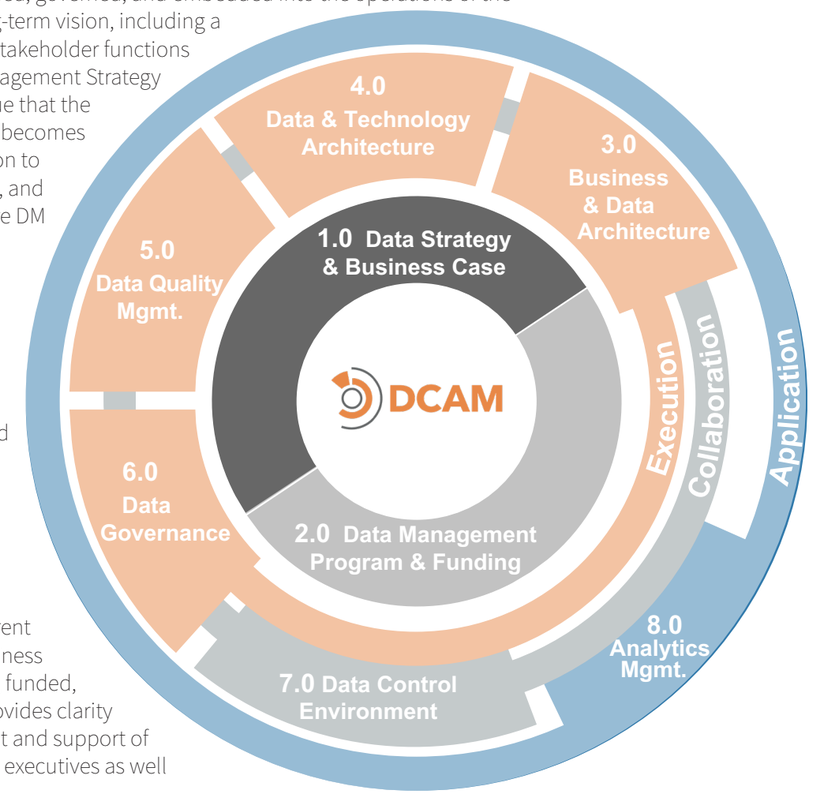
Survey Results & Observations

Survey Results & Observations

1.0 The Data Management Strategy and Business Case

The Data Management Strategy & Business Case determines how Data Management (DM) is defined, organized, funded, governed, and embedded into the operations of the organization. It defines the long-term vision, including a description of stakeholders or stakeholder functions that must be aligned. Data Management Strategy demonstrates the business value that the program will seek to achieve. It becomes the blueprint for the organization to evaluate, define, plan, measure, and execute a successful and mature DM initiative.

The purpose of developing a DM strategy and business case is to articulate the rationale for the DM initiative. The strategy defines why the initiative is needed, as well as the goals and expected benefits. The strategy also describes how to mobilize the organization to implement a successful DM initiative. The DM business case provides the rationale for the investment in the DM initiative. DM is no different than any other established business process. It needs to be justified, funded, measured, and evaluated. It provides clarity of purpose, enabling agreement and support of initiative objectives from senior executives as well as program stakeholders.



Component 1.0
Defines the vision and the purpose of the Data Management Program. Why is Data Management important?!

Survey Results & Observations

Question 1: A formal Data Management Strategy and approach has been developed and communicated to organizational stakeholders

The Data Management Strategy determines how Data Management is defined, organized, funded, governed, and embedded into the organization's operations. It defines the long-term vision and becomes the blueprint for the organization to evaluate, define, plan, measure, and execute a successful, mature data initiative.

Analysis

- Unsurprisingly, the highly regulated Pharma / Life Sciences and Finance industries scored similarly, and higher, than Other industries.
- Overall, low scores of less than 20% Achieved across all industries demonstrates the ever-evolving nature of Data Management.

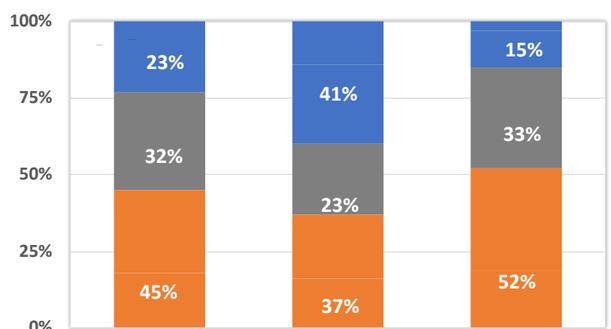
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.47	18%
Other (2020)	3.00	16%
Finance (2020)	3.52	19%

As increasing amounts of data are made available in today's world, developing a strategy to manage it becomes increasingly important. Awareness of data as an asset is needed to realize its value through actionable insights. This remains a challenge, with low success rates, particularly in big data implementations.

A formal Data Management Strategy is needed to identify short- and long-term goals, and to formulate the necessary steps to achieve them. As an ever-evolving field, continuous re-evaluation is necessary to ensure that an organization's Data Management Strategy is aligned with business goals, including regulatory compliance and Data Ethics. Additionally, keeping an eye on new and evolving data types (unstructured, social media, sensor, signal) will aid in this periodic re-evalution.

Organizations in the highly regulated Finance and Pharma / Life Sciences industries generally have an established Data Management Strategy.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	15%	3%
Planning	23%	26%	12%
Under Way	32%	23%	33%
Defined	27%	21%	33%
Achieved	18%	16%	19%

Question 2: A Data Management Business Case (including requirements and prioritization) has been developed and communicated

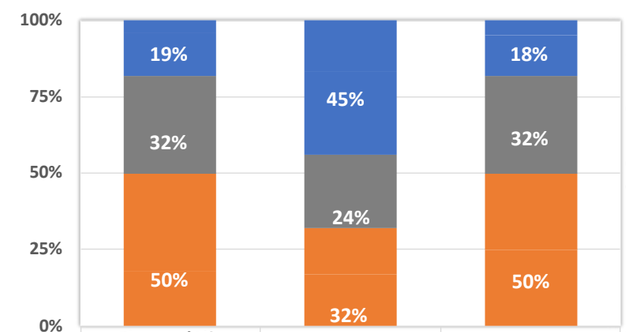
The purpose of developing a Data Management Business Case is to articulate the rationale for the data initiative. The Business Case is the justification for the initiative. It articulates the data-related issues facing an organization and describes the expected outcomes and benefits that may be achieved through the successful implementation of a DM program.

Analysis

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.52	18%
Other (2020)	2.99	17%
Finance (2020)	3.52	25%

- The highly regulated Finance and Pharma / Life Sciences industries lead Other industries substantially in terms of Average DCAM Score.
- Half of the organizations surveyed in the Finance and Pharma / Life Sciences industries report the Data Management Business Case has been either Defined or Achieved.
- While Finance respondents report an even split between the Defined and Achieved stages, 32% of Pharma / Life Sciences respondents indicate that their companies' Data Management Business Case is in the Defined stage.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	18%	6%
Planning	14%	27%	13%
Under Way	32%	24%	32%
Defined	32%	15%	25%
Achieved	18%	17%	25%

Not only is an organization's capital allocation important for maximizing profit, but it is also an investment in efficiency. An effective Data Management Business Case shines a spotlight on the escalating complexity of the data landscape to help senior executives realize how critical it is to start tackling these challenges with formal, sustained investment.

Industries differ in their approaches to Data Management. While earlier efforts focused mainly on data defense, increasingly, companies are turning to the offensive Data Business Case to harness and leverage the potential of data.

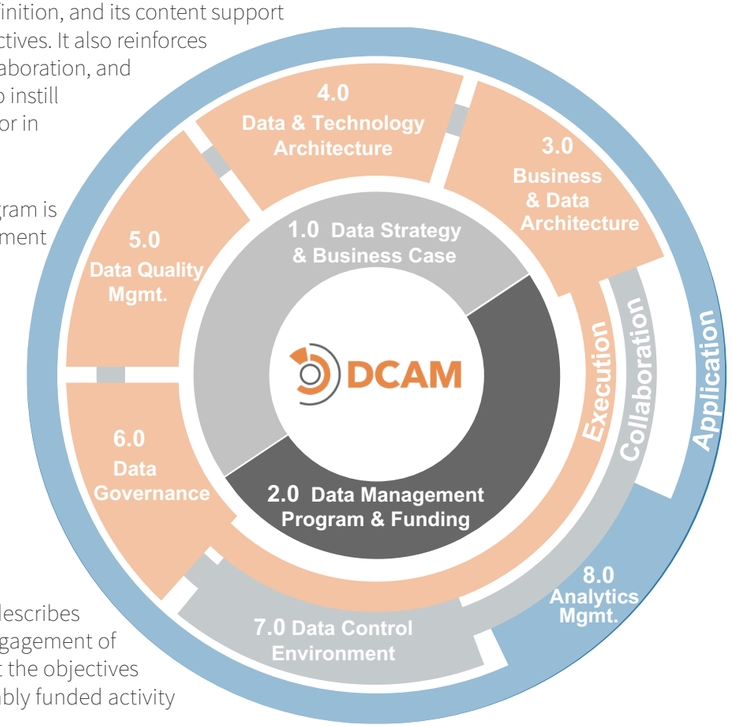
Survey Results & Observations

Photo by Andreas M on Unsplash

Survey Results & Observations

2.0 The Data Management Program & Funding

The Data Management Program is an organizational function dedicated to the management of data as an asset throughout an organization. It illustrates how the management of Data Quality (DQ), its definition, and its content support strategic, business, and operational objectives. It also reinforces the necessity of orchestration, active collaboration, and alignment among diverse stakeholders to instill confidence in data as a trusted input factor in business and operational processes.



The purpose of a Data Management Program is to organize and embed the Data Management (DM) concepts into the operational framework of an organization on a sustainable basis. The creation and implementation of the DM program elevates the importance of DM and integrates it as a core aspect of organizational operations. It establishes DM as a sustainable activity by ensuring sustainable funding. It reinforces the importance of managing data across the organization via education, training, and communication.

The Data Management Funding Model describes the overall framework and high-level engagement of senior management used to ensure that the objectives and processes of DM become a sustainably funded activity within the organization.

Component 2.0
Describes the organizational structure and funding model of the Data Management Program.

Survey Results & Observations

Question 3: Your organization has a formally established, structured, and funded Data Management Program.

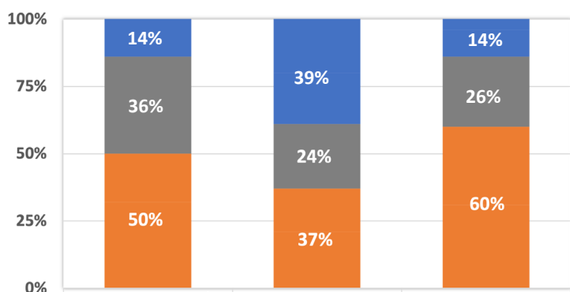
A Data Management Program is a formally established, independent, and sustainable part of the organization. It defines the lines of responsibility and accountability. It ensures access to appropriate staff resources and functional capabilities, has strong support of executive management, and is granted appropriate governance authority to ensure the implementation of a successful Data Management initiative.

Analysis

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.73	32%
Other (2020)	3.02	21%
Finance (2020)	3.77	31%

- Fifty percent of survey respondents in the Pharma / Life Sciences industry indicated that their Data Management Program is either Defined or Achieved in the 2020 survey, compared to 60% in the Finance industry.
- Both Pharma / Life Sciences and Finance industries show high Average DCAM Scores of 3.73 and 3.77, respectively.
- Other industries lag behind with only 37% Defined /Achieved and the lowest Average DCAM Score of 3.02.



	Pharmaceutical / Life Science (2021)	Other Industry (2020)	Finance Industry (2020)
Not Initiated	0%	20%	4%
Planning	14%	19%	10%
Under Way	36%	24%	26%
Defined	18%	16%	29%
Achieved	32%	21%	31%

A Data Management Program organizes and embeds Data Management concepts into an organization's operational framework in a sustainable manner. Whereas earlier Data Management efforts tended to consist of uncoordinated attempts by isolated individuals leading to disparate outcomes, a formal, funded Data Management Program facilitates the implementation of sustainable Data Management tools, which aid in maintaining consistency in Data Quality and processes, and engendering trust in data.

In this survey, a great majority of organizations in the Pharma / Life Sciences and Finance industries are at the Under Way, Defined, or Achieved level in establishing, structuring, and funding a Data Management Program, showing that senior management formally recognizes and acknowledges how important data and its proper management are to organizational success.

Question 4: Formal plans, roadmaps, and deliverables have been defined and articulated to program stakeholders

Program roadmaps must describe the steps required to attain the DM initiative's target state for its organizational structure and function. Stakeholders must be engaged and validate the roadmaps. Once roadmaps are approved, they must be translated into detailed project plans.

Analysis

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.72	32%
Other (2020)	2.93	19%
Finance (2020)	3.58	23%

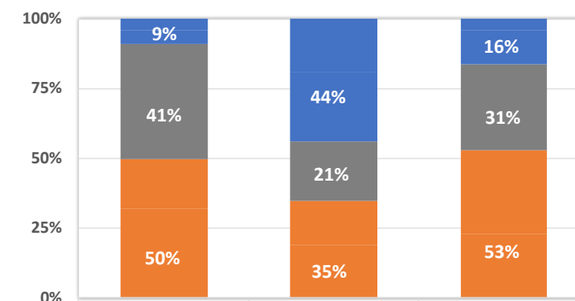
- Among the highest Average DCAM Scores in the survey, the Pharma / Life Sciences industry scored 3.72.
- Again, high percentages of the Pharma / Life Sciences and Finance organizations report that their formal plans, roadmaps, and deliverables are at the Defined / Achieved level.

Once a Data Management Strategy has been established and funding secured for the program, clear-cut objectives, plans, and deliverables should be defined and communicated to all stakeholders.

The Pharma / Life Sciences industry has the highest score on this dimension, likely due to strict regulatory requirements.

The Finance industry, another highly regulated industry, also fares relatively well, achieving an Average DCAM Score only slightly lower than that of the Pharma / Life Sciences industry.

Meanwhile, companies in Other industries lag behind, highlighting the role that regulatory pressure may play in driving the establishment of program roadmaps.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	4%	19%	4%
Planning	5%	25%	12%
Under Way	41%	21%	31%
Defined	18%	16%	30%
Achieved	32%	19%	23%

Survey Results & Observations

Question 5: Data Management Process Excellence is established formally

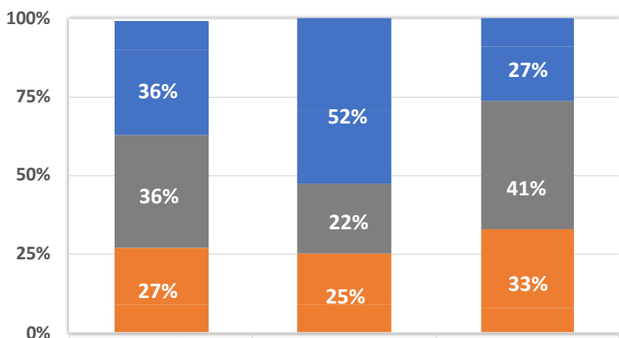
A successful Data Management initiative depends on the replicable execution of a standard, sustainable, and measurable process. In collaboration with Line of Business (LOB), the Data Management Process Excellence approach is designed to help establish processes that improve the scope and quality of enterprise data.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	2.98	9%
Other (2020)	2.54	9%
Finance (2020)	3.04	8%

Analysis

- More than half of organizations surveyed in Other industries are still at the Not Initiated / Planning stage, though the percentage of respondents who reported that Data Management Process Excellence has been Achieved is on par with both the Finance and Pharma / Life Sciences industries.
- Overall, all industries reported low percentages of Achieved scores.
- The Finance industry has a slightly better Average DCAM Score for established Data Management Process Excellence compared to the Pharma / Life Sciences industry.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	9%	29%	10%
Planning	27%	23%	17%
Under Way	36%	22%	41%
Defined	18%	16%	25%
Achieved	9%	9%	8%

A sustainable Data Management Program requires establishment of standard Data Management processes, preferably across the whole Data Management framework, as well as a feedback loop to address emerging problems for long-term continuous improvement.

Question 6: Data Management Stakeholder Engagement is established and confirmed

Data Management Stakeholder Engagement requires participation and cooperation across the enterprise, and among stakeholders outside the organization. Stakeholders must be engaged in and held accountable for delivery of timely, high-quality data. To strengthen that commitment, performance in support of the Data Management Program should be reflected in stakeholder reviews and compensation.

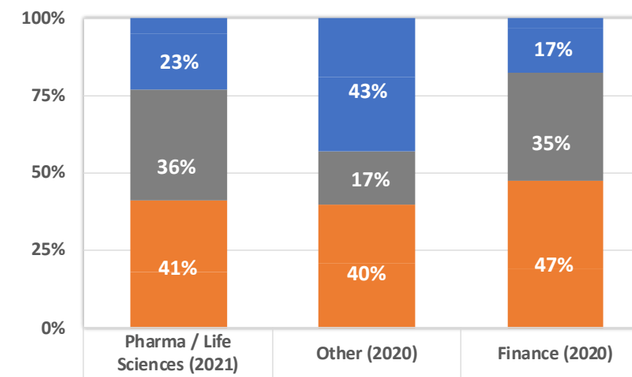
Analysis

- The Pharma / Life Sciences and Finance industries report comparable Average DCAM Scores (3.36 and 3.48, respectively) and % Achieved (19% and 18%, respectively) for Data Managment Stakeholder Engagement. Other industries report a slightly higher percentage of companies at the Achieved level (20%).
- However, Other industries have a much greater proportion of organizations (43%) at the Not Initiated or Planning stages, and a much smaller pipeline of activity at the Under Way level (17%) than the Finance and Pharma / Life Sciences industries.

A successful Data Management Program requires the formalized, accountable engagement of stakeholders across an organization, as well as with stakeholders outside the organization.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.36	18%
Other (2020)	3.03	21%
Finance (2020)	3.48	19%



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	19%	3%
Planning	18%	24%	14%
Under Way	36%	17%	35%
Defined	23%	19%	28%
Achieved	18%	21%	19%

Survey Results & Observations

Question 7: Data Management communications and training programs have been developed and implemented across the organization

Internal communications and formal training are needed to effect change in organizational culture. Plans are needed for internal communications to drive awareness and adherence to the Data Management Program. The communications strategy should be designed and implemented according to the organizational culture, and incorporated into the Data Management Strategy.

DCAM

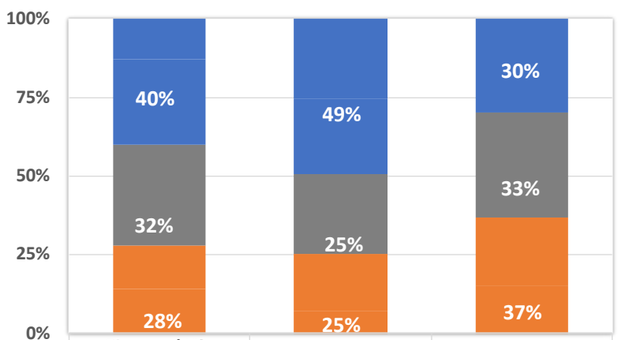
Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	2.92	14%
Other (2020)	2.58	7%
Finance (2020)	3.11	15%

Analysis

- With relatively low Average DCAM Scores across all industries for communications and training programs, the Finance and Pharma / Life Sciences industries have more organizations with initiatives Under Way than Other industries.
- Other industries scored 2.58 with around half the organizations surveyed still at the Not Initiated / Planning stage.

Communications and training programs are critical to the success of a Data Management Program. Many organizations lack a clear communications structure/matrix, leaving some stakeholders out of the conversation. Such exclusion may result in a range of detrimental consequences, so companies should invest in engaging stakeholders during the development of the processes necessary for successful Data Management.

This stakeholder engagement includes training that enhances the transparency of the Data Management Program and assists both internal and external stakeholders in their professional development. With technology advancing at such a rapid pace, and with data being using by increasing numbers of stakeholders within the data supply chain, it is inevitable that stakeholders will require periodic training to up-skill and maintain knowledge of Data Management processes.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	13%	25%	12%
Planning	27%	24%	18%
Under Way	32%	25%	33%
Defined	14%	18%	22%
Achieved	14%	7%	15%

Question 8: Metrics have been defined and are being captured to determine the success—and support the continuous improvement—of the Data Management Program

Metrics are essential to validating the benefits of your Data Management Program. Having a program in place to identify key performance indicators (KPIs), track program implementations, define and document outcomes, and capture and report financial metrics is a required capability in a mature Data Management Program..

Analysis

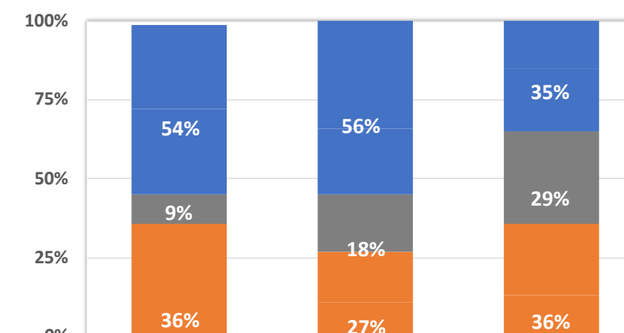
- None of the Pharma / Life Sciences organizations surveyed reached the Achieved stage.
- More than half the organizations surveyed in the Pharma / Life Sciences and Other industries remain at the Not initiated / Planning stage.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	2.59	0%
Other (2020)	2.46	11%
Finance (2020)	2.99	13%

Performance metrics are needed not only to measure the effectiveness of Data Management Programs, but also to determine areas for improvement, justify current funding, and procure future funding.

More than 50% of the survey respondents stated that they are not capturing the metrics needed to determine the return on investment (ROI) of their data programs. Thus, there is a need to develop standards to measure and track ROI in data programs.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	27%	35%	15%
Planning	27%	21%	20%
Under Way	9%	18%	29%
Defined	36%	16%	23%
Achieved	0%	11%	13%

Survey Results & Observations

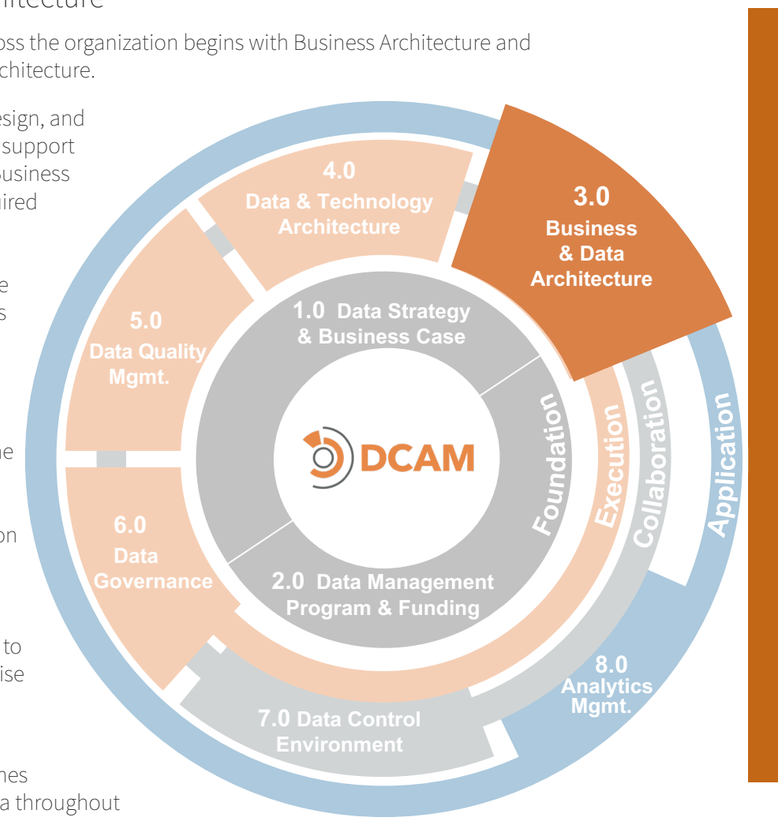
Photo by Joel Filipe on Unsplash

Survey Results & Observations

3.0 Business and Data Architecture

The path to integrated architecture across the organization begins with Business Architecture and how it defines requirements for Data Architecture.

Business Architecture is the strategy, design, and execution of the capabilities needed to support the organization's business functions. Business Architecture defines the processes required to meet the objectives of the business. The processes have requirements for data and Data Management (DM). These requirements must be defined as inputs and outputs of the business process.



Data Architecture is the strategy and execution of how data is designed (identified and described) to support the business objectives. Data Architecture speaks to the design, definition, management, and control of information content. Data Architecture identifies data domains, documents metadata, defines critical data elements, and establishes taxonomies and ontologies to ensure that the meaning of data is precise and unambiguous, and that data use is consistent and transparent.

The Data Architecture function establishes consistency in definition and use of data throughout an organization. Adhering to a prescribed Data Architecture forces business and technology resources to take the necessary steps to define and document data meaning, define the appropriate use of the data, and to ensure that proper governance is in place to manage data as meaning on a sustainable basis.

Component 3.0

Data Architecture is the design of information content, driven by Business Architecture (requirements), aligned to real-world objects and entities.

Survey Results & Observations

Question 9: Your organization has a formally established and supported Data Architecture Program

Data Architecture speaks to the design, definition, management, and control of information content. It identifies data domains, documents metadata, defines critical data elements, and establishes taxonomies and ontologies. Data Architecture is critical to ensuring that the meaning of data is precise and unambiguous, and that the use of data is transparent.

Analysis

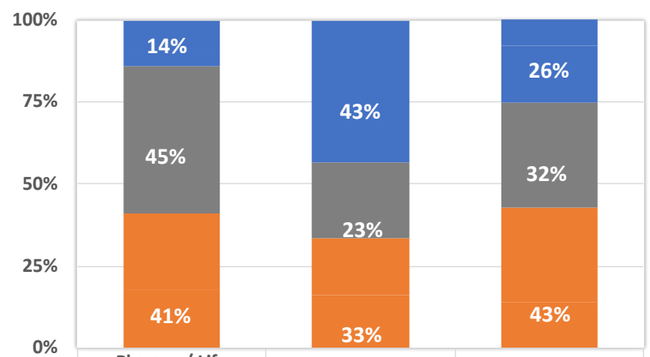
- The Pharma / Life Sciences industry is ahead of the pack when it comes to Data Architecture, with 86% of Data Architecture programs either Under Way, Defined, or Achieved.
- Similar Defined / Achieved percentages for the Pharma / Life Sciences and Financial industries (41% and 43%, respectively), stand in contrast to a disparity in percentages at the Planning stage (5% and 17%, respectively). For reasons yet to be explored, Pharma /Life Sciences companies seem to experience fewer bottlenecks at the Planning stage than those in the Finance industry.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.43	18%
Other (2020)	2.90	16%
Finance (2020)	3.24	14%

Data Architecture provides unambiguous language for Data Management. It establishes how stakeholders define the meaning of data, and codifies this in data domains, definitions of critical data elements, metadata schema, taxonomies, and ontologies. Data Architecture is critical to ensuring that the meaning of data is precise and unambiguous, and that the use of data is transparent.

A slightly higher percentage of Pharma / Life Sciences and Finance organizations attained a Defined or Achieved level, compared to Other industries. Overall, the Pharma / Life Sciences industry is more mature than Other industries.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	9%	18%	9%
Planning	5%	25%	17%
Under Way	45%	23%	32%
Defined	23%	17%	29%
Achieved	18%	16%	14%

Question 10: Business Architecture is established and integrated into your Data Management Program.

Business Architecture is the strategy, design, and execution of the capabilities needed to support organizational business functions. Business Architecture defines the processes required to meet the objectives of the business, which, in turn, define the data requirements—the inputs and outputs of the business process.

DCAM

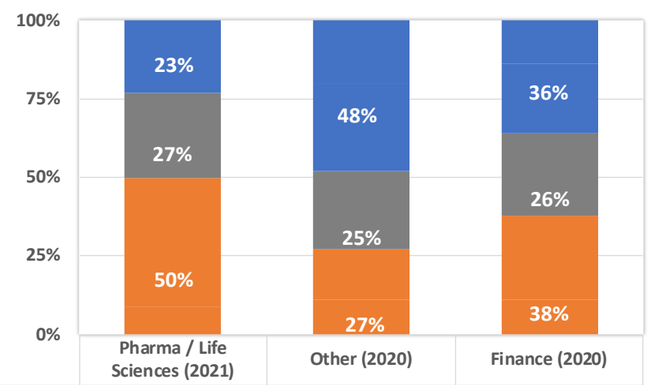
Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.45	9%
Other (2020)	2.71	11%
Finance (2020)	3.00	11%

Analysis

- Fifty percent of Pharma / Life Sciences respondents have integrated Business Architecture into their data programs.
- The Pharma / Life Sciences industry Average DCAM Score of 3.45 is significantly higher than those of the Finance and Other industries.

Business Architecture defines the requirements for Data Architecture, both of which should be reflected in the Data Management Strategy. Although data has traditionally been considered an IT domain, things are changing fast, with organizations recognizing that Data Architecture needs to be aligned with clear, directional business goals to have a positive impact. Business Architecture is a bridge between the requirements for data of the business process and the physical execution of that data in technology infrastructure.

Fifty percent of the Pharma / Life Sciences industry demonstrated alignment of their Business Architecture to their Data Management Programs, indicating a recognition of the impact of business processes on the ability to leverage data.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	20%	14%
Planning	23%	28%	22%
Under Way	27%	25%	26%
Defined	41%	16%	27%
Achieved	9%	11%	11%

Survey Results & Observations

Question 11: Logical Data Domains, models, and metadata have been identified and documented, and usage is supported by policy

Logical Data Domains are the categories of data needed to satisfy specific business functions. Identification of domains must be driven by the business, from the perspective of what data is needed to perform the required business functions.

Analysis

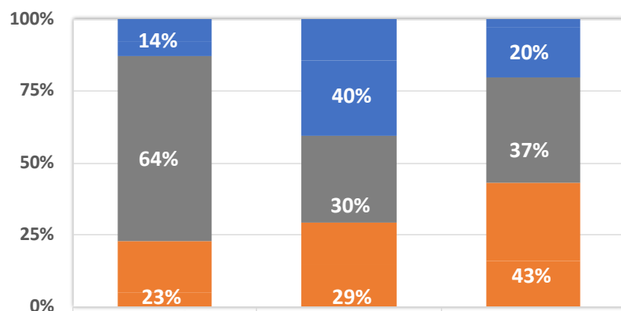
- Finance industry respondents scored higher than those in Pharma / Life Sciences in identifying and using Logical Data Domains.
- Significantly fewer Pharma / Life Sciences firms report that their Logical Data Domains are at the Achieved level than the other industries.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.11	5%
Other (2020)	2.91	14%
Finance (2020)	3.37	16%

As organizations' focus undergoes a paradigm shift from 'databases' to 'data,' increasing numbers of companies are applying Logical Data Domains, models, and metadata to manage better large volumes of data.

Given that Pharma / Life Sciences has a greater proportion of new programs, it is not surprising to see a significantly higher number of data domain activities in the Under Way stage. From a forward-looking perspective, this is a very promising score.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	9%	14%	3%
Planning	5%	26%	17%
Under Way	64%	30%	37%
Defined	18%	15%	27%
Achieved	5%	14%	16%

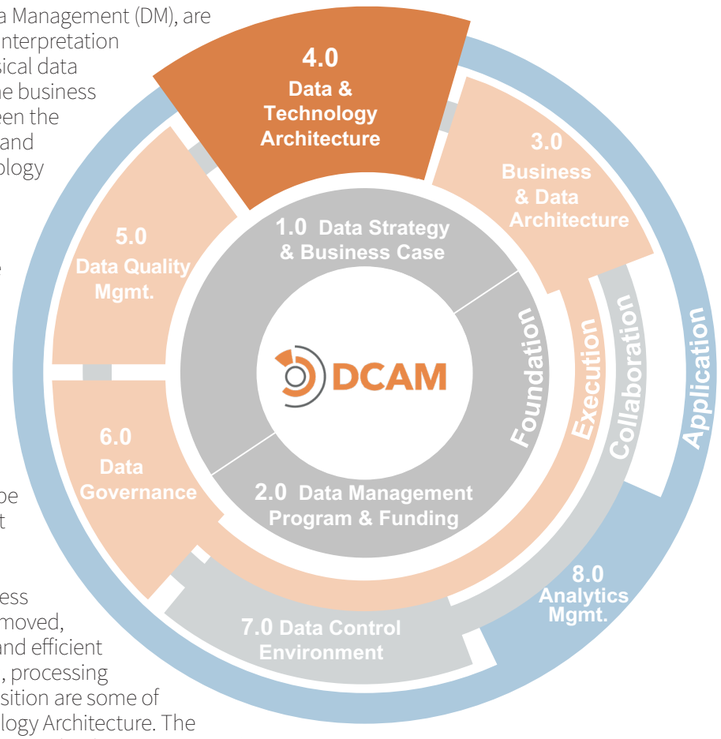
Survey Results & Observations

Photo by National Cancer Institute on Unsplash

Survey Results & Observations

4.0 Data & Technology Architecture

The requirements for data as well as for Data Management (DM), are interpreted through Data Architecture. This interpretation defines the requirements to design the physical data consumed, produced and provisioned by the business process. Data Architecture is a bridge between the data requirements fof the business process and the physical execution of that data in technology infrastructure.



Technology Architecture is the strategy and execution of how the physical infrastructure is designed to support the business and data needs of the organization.

Technology Architecture refers to the strategy, design, and implementation of the technology infrastructure, which supports the defined business and Data Architecture. Technology Architecture defines the platforms and tools, and how they need to be designed for maximum efficiency to support the requirements of the business process, data, and DM. The purpose of Technology Architecture is to support the business process and define how data is physically acquired, moved, persisted, and distributed in a streamlined and efficient manner. Physical data proximity, bandwidth, processing time, backup, recovery, archiving, and disposition are some of the important elements of a mature Technology Architecture. The efficient and effective movement of data is critical to business operations. Technology Architecture determines how data, tools and platforms operate in collaboration to satisfy business requirements. The proper alignment of these components dictates application efficiency and system processing speed. This enables organizations to control costs and achieve infrastructure scalability and elasticity, which are characteristic of an organization that is designed for long-term implementation success.

Component 4.0
Technology Architecture addresses the physical implementation of Data Management (platforms, DBs, tools), in collaboration with the business and the Data Management Office.

Question 12: Technology Architecture in support of the Data Management Program has been developed and documented

The role of the Technology Architecture is to define and design the infrastructure needed to satisfy the data requirements of the organization. Working in collaboration with the Data Management Program, Technology Architecture defines the database strategies, analytic platforms, and all other aspects of an holistic technology infrastructure.

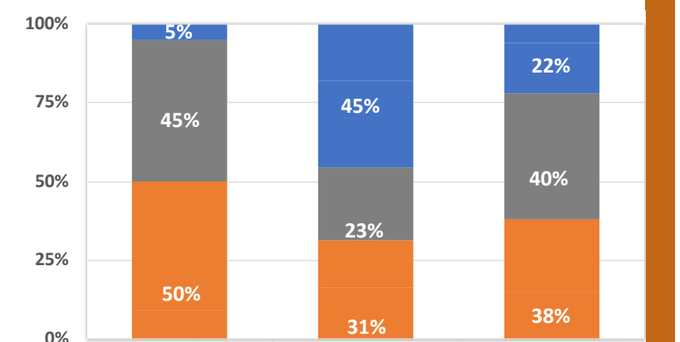
Analysis

- Ninety-five percent of surveyed organizations in the Pharma / Life Sciences industry report alignment between their technology vision and data programs is either Defined or Under Way.
- Meanwhile, almost half of organizations surveyed in Other industries report their Technology Architecture is in the Not Initiated / Planning stage.

The execution of a Data Management Program requires a technology infrastructure. The Pharma / Life Sciences industry leads all other industries in Technology Architecture. This technological maturity in the Pharma / Life Sciences industry may reflect the exceptional need for precision in the industry.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.57	9%
Other (2020)	2.84	16%
Finance (2020)	3.27	15%



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	18%	6%
Planning	5%	27%	16%
Under Way	45%	23%	40%
Defined	41%	15%	23%
Achieved	9%	16%	15%

Survey Results & Observations

Survey Results & Observations

Question 13: Tools required to support the Data Management Program have been identified and implemented

Many tools available in the marketplace may be used to support the Data Management Program. The selection of such tools must first be driven by business requirements, then reviewed, implemented, and governed by the Technology Architecture.

DCAM

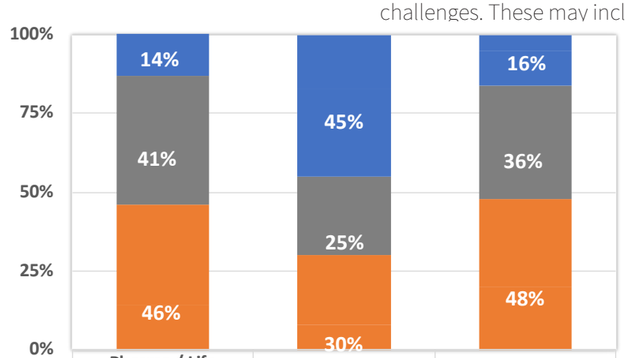
Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.58	14%
Other (2020)	2.75	8%
Finance (2020)	3.46	20%

Analysis

- The Pharma / Life Sciences and Finance industries show similar Average DCAM Score distributions.
- The Finance industry reports the highest Achieved percentage of 20%.
- Other industries lag behind with 45% at the Not Initiated / Planning stage.

The advancement of Data Management Programs requires a corresponding upgrade to tools and platforms to deal with new challenges. These may include governance tools, metadata, modeling, extract-transform-load (ETL), Data Quality audit, and Analytics. However, individual selection of fit-for-purpose tools is essential to maximize utility while minimizing costs.

Many organizations in both the Pharma / Life Sciences and Finance industries report being adept at incorporating suitable tools into their Data Management Programs with a high percentage of scores at the Under Way, Defined, and Achieved levels.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	17%	5%
Planning	14%	28%	11%
Under Way	41%	25%	36%
Defined	32%	22%	28%
Achieved	14%	8%	20%

Question 14: Data management operational risk planning is in place

A mature Data Management Program addresses operational risk, business continuity, and disaster recovery strategies. Data Management must be in alignment with operational risk governance and engaged in contingency planning and testing for data access and maintenance in the event of an operational disruption.

Analysis

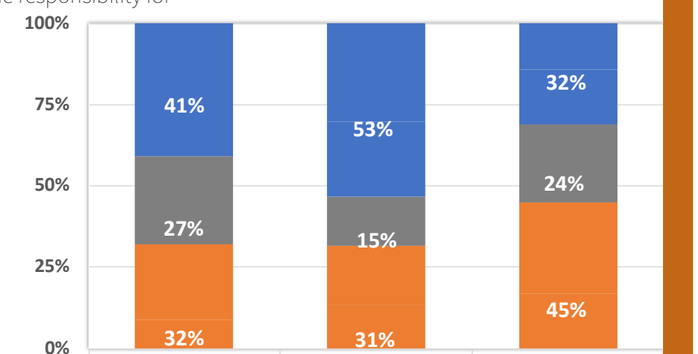
- Forty-five percent of Finance industry respondents report their Data Management operational risk planning is at the Defined / Achieved level.
- Pharma / Life Sciences organizations trail behind, with only 32% reporting their Data Management operational risk planning has reached the Defined / Achieved stage, while Other industries reported 31% at the Defined / Achieved level.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.01	9%
Other (2020)	2.83	13%
Finance (2020)	3.16	17%

An operational risk plan is an essential part of any Data Management Program, and should not only contain corrective steps for potential risks, but also preventive steps to mitigate risk proactively. This is especially important for critical business and operations functions. The responsibility for operational risk planning lies not only with Data Management, but also must be coordinated with business functions to ensure sustainable operations.

The Finance industry fares marginally better than the Pharma / Life Sciences industry and significantly better than Other industries on this dimension. Though the Pharma / Life Sciences industry is not more advanced in this area, it has the lowest percentage of companies that are yet to initiate Data Management operational risk planning.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	30%	15%
Planning	36%	23%	17%
Under Way	27%	15%	24%
Defined	23%	18%	28%
Achieved	9%	13%	17%

Survey Results & Observations

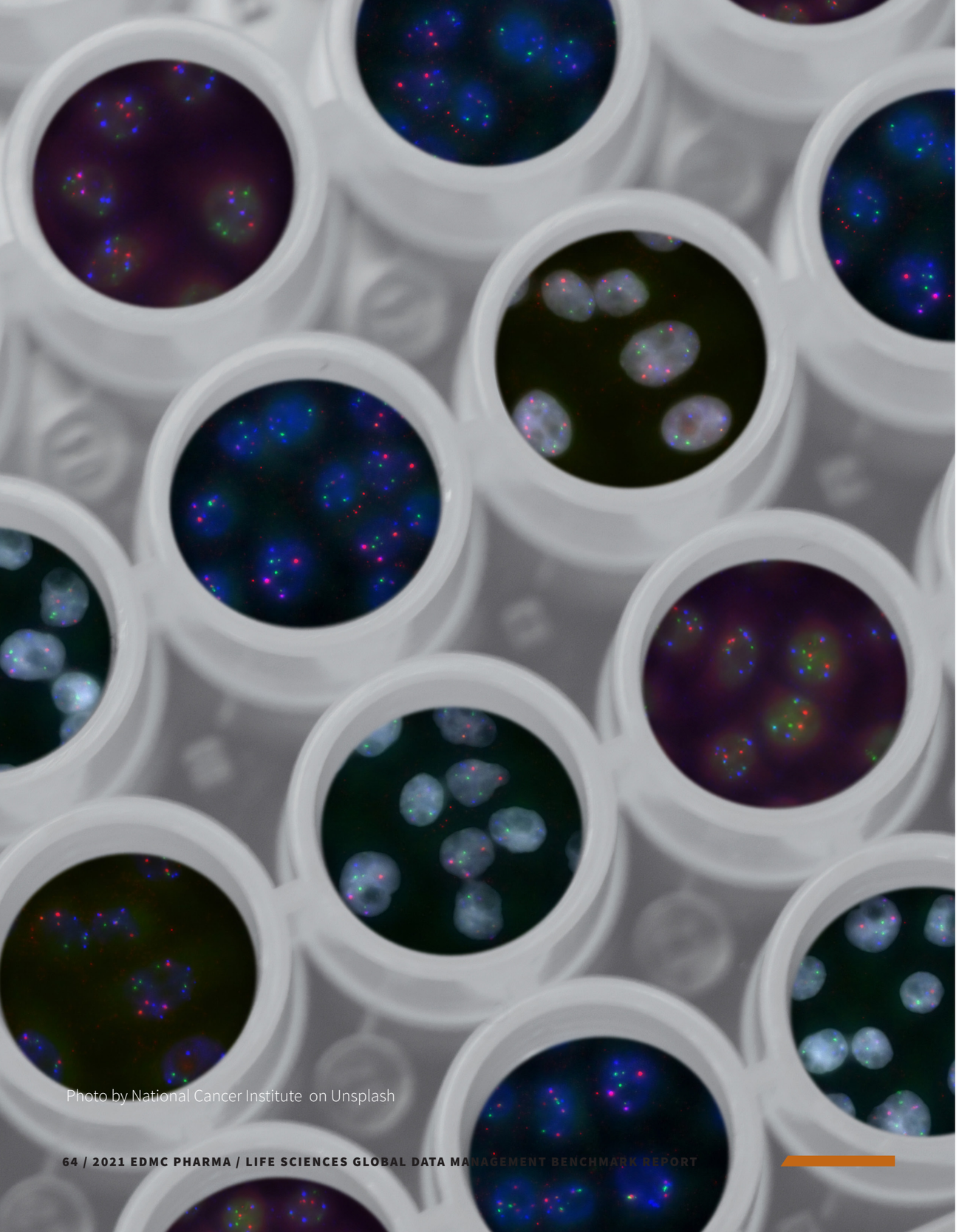


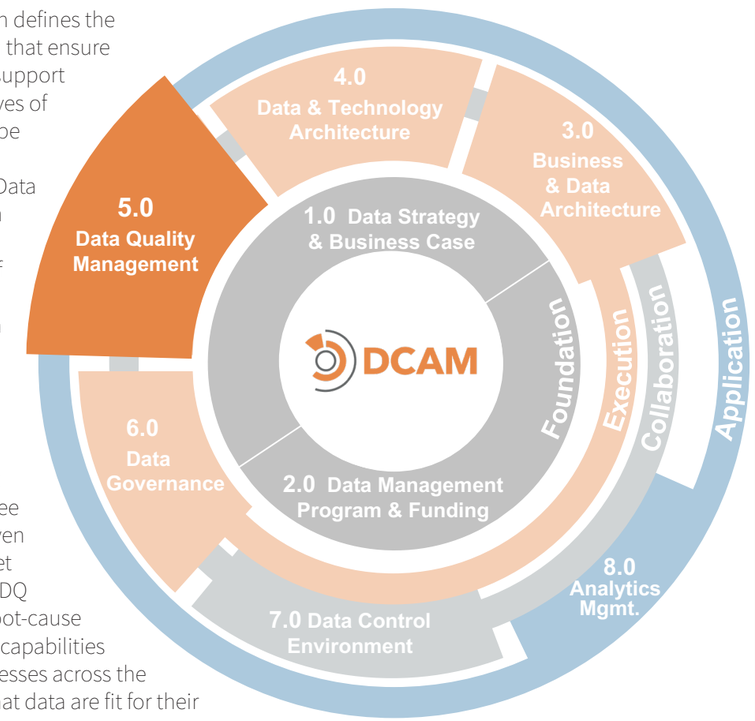
Photo by National Cancer Institute on Unsplash

Survey Results & Observations

5.0 Data Quality Management

The Data Quality Management function defines the goals, approaches, and plans of action that ensure data content is of sufficient quality to support defined business and strategic objectives of the organization. The function should be developed in alignment with business objectives, measured against defined Data Quality (DQ) dimensions and based on an analysis of the current state of DQ. Data Quality Management is a series of processes across the full data supply chain designed to ensure that the data provisioned meets the needs of its intended consumers.

DQ requires an understanding of how data is sourced, defined, transformed, provisioned, and consumed. DQ is not a process itself, but describes the degree to which data is fit for purpose for a given business process or operation. It is a set of capabilities to define data profiling, DQ measurement, defect management, root-cause analysis, and data remediation. These capabilities allow the organization to execute processes across the Data Control Environment, ensuring that data are fit for their intended purposes.



Component 5.0
Deliver data that are trusted and fit for purpose, where users have confidence that the data are what they need, without reconciliation.

Survey Results & Observations

Question 15: A Data Quality Management Program is established and formalized

The Data Quality Management Program must be defined and approved by stakeholders. Roles and responsibilities across the stakeholders must be established with operational, auditable processes in place. Once established, the Data Quality Management Program must be empowered formally by senior management, and its role must be communicated to all stakeholders.

Analysis

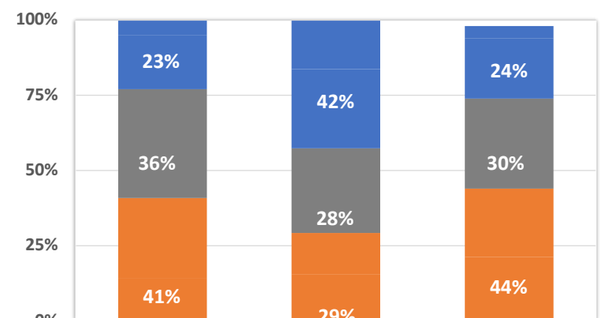
- The Pharma / Life Sciences and Financial industries display similar score distributions and Average DCAM Scores.
- Other industries lag behind with only 29% at the Defined / Achieved level and 42% at the Not Initiated / Planning stage.

Data Quality ensures that data content is of sufficient quality to support defined business and strategic objectives of the organization, meaning that data is 'fit for purpose.' It is vital to establish a Data Quality Management program that defines the goals, approaches, and plans of action that ensure sufficient Data Quality. Ensuring Data Quality should be a common responsibility of all stakeholders who come in contact with the data.

In terms of Data Quality, the Finance and Pharma / Life Sciences industries report a similar distribution of scores, while Other industries are far behind.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.31	14%
Other (2020)	2.88	15%
Finance (2020)	3.39	21%



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	16%	4%
Planning	18%	26%	20%
Under Way	36%	28%	30%
Defined	27%	14%	23%
Achieved	14%	15%	21%

Question 16: Data are profiled, measured, monitored, and maintained

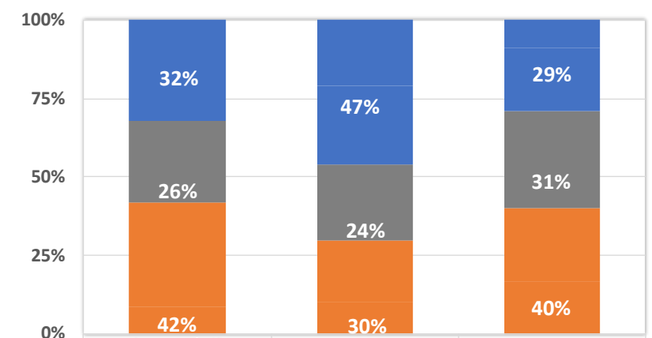
Profiling and measuring the data includes prioritizing the data in scope based on criticality and materiality, defining and testing Data Quality rules based on business rules, and verifying that the data are fit for purpose. Monitoring and maintaining the data includes implementing DQ control points, capturing DQ metrics, and performing continuous monitoring.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.11	9%
Other (2020)	2.71	10%
Finance (2020)	3.21	17%

Analysis

- The Pharma / Life Sciences industry leads the Finance industry on data profiling, measurement, monitoring, and maintenance, with 42% and 40%, respectively, at the Defined / Achieved level.
- Notably, 0% of surveyed Pharma / Life Sciences organizations at the Not Initiated stage. In other words, all Pharma / Life Sciences respondents have started planning, at a minimum.
- Other industries lag behind with 47% at the Not Initiated / Planning stage and the lowest Average DCAM Score of 2.71



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	0%	22%	9%
Planning	32%	25%	20%
Under Way	26%	24%	31%
Defined	33%	20%	23%
Achieved	9%	10%	17%

As the volume of data increases, there is a growing need for automated data profiling. Automated processing to better understand the structure, content, and interrelationships among data provides an initial overview of the suitability of data for future use, and the ability to identify and address issues early.

Across all industries, only about one-third of organizations attained a score corresponding to the Defined and Achieved levels, with the Finance industry slightly ahead.

Survey Results & Observations

Question 17: Data Quality root-cause analysis is performed routinely

Data remediation includes correcting defective data and determining the root cause of the DQ deterioration to avoid recurrence of defective data in the future. Data defects may stem from a people, process, data, or technical source. Having the right subject matter expertise from each of these areas is instrumental to the analysis of the root cause.

Analysis

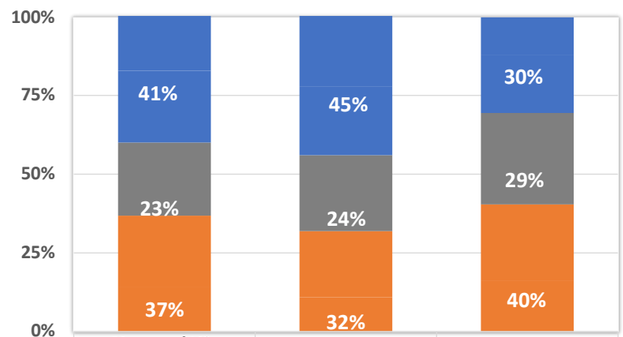
- The Finance industry leads with the highest Average DCAM Score (3.15) with 40% reporting that root-cause analysis is Defined / Achieved.
- Forty-one percent of Pharma / Life Sciences organizations report that their root-cause analysis capability is at the Not Initiated / Planning stage.
- Forty-five percent of organizations in Other industries report that their root-cause analysis capability is at the Not Initiated / Planning stage.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	2.98	14%
Other (2020)	2.76	11%
Finance (2020)	3.15	16%

Routine root-cause analysis across the entire data supply chain is necessary to identify, analyze, and remediate defective data at its source. This facilitates the identification of the cause—rather than just the symptoms—of the problem.

The score distribution for this question is similar to those of questions 15 and 16, which makes sense because these functions are inter-dependent and inter-related. The Pharma / Life Sciences industry has an unexpectedly high percentage of scores at the Not Initiated and Planning stages. However, given the nature of the work in this industry, it is possible that Data Quality root-cause analysis is performed on an ad hoc basis, rather than on a pre-determined schedule according to a Data Quality program.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	18%	23%	12%
Planning	23%	22%	18%
Under Way	23%	24%	29%
Defined	23%	21%	24%
Achieved	14%	11%	16%

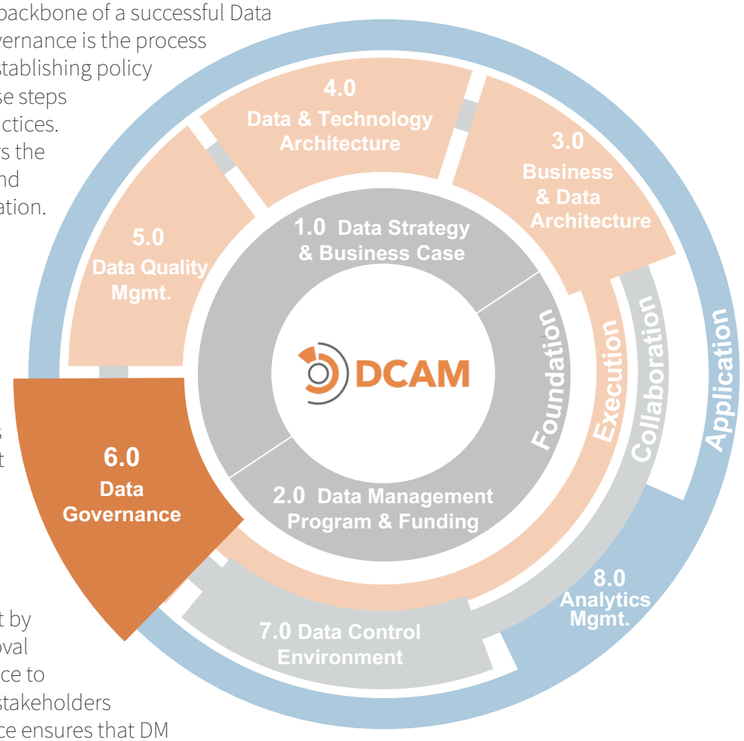
Survey Results & Observations

Photo by Ahmed Yaanui on Unsplash

Survey Results & Observations

6.0 Data Governance

The Data Governance function is the backbone of a successful Data Management (DM) initiative. Data Governance is the process of setting standards, defining rules, establishing policy and implementing oversight. It is these steps that ensure adherence to DM best practices. Governance formalizes and empowers the DM initiative to ensure propagation and sustainability throughout the organization.



The purpose of Data Governance is to formalize DM as an established business function. Data Governance establishes the rules of engagement, drives the prioritization of funding and enforces compliance. Data Governance delineates the guidelines for data movement. These movement guidelines prescribe how data will be acquired, persisted, distributed, appropriately used, archived, and/or defensibly destroyed.

Data Governance formalizes oversight by establishing control guidelines, approval processes, and evaluation of adherence to policies and procedures. It identifies stakeholders and empowers them. Data Governance ensures that DM principles are fully detailed, and adoption is achieved. Business, data, and technology functions are held responsible for the maintenance, quality, and proper use of data throughout the organization as part of the Data Governance function.

Component 6.0
The rules of engagement for Data Management, focused on the implementation of policies, standards, and operational procedures necessary to ensure that stakeholders 'behave.'

Question 18: Data Governance function is established and operational

Data Governance strategy and approach must be defined and reflect the related vision and objectives of the Data Management Strategy. Once established, it must be empowered formally by senior management and its role must be communicated to all stakeholders.

Analysis

- The Finance industry leads with a high Average DCAM Score of 3.87 with 64% of respondents reporting their Data Governance function ranks at the Defined / Achieved level.
- The Pharma / Life Sciences industry slightly lags behind Other industries in Average DCAM Score, though its respondents report a slightly greater proportion at the Achieved level for Data Governance function.

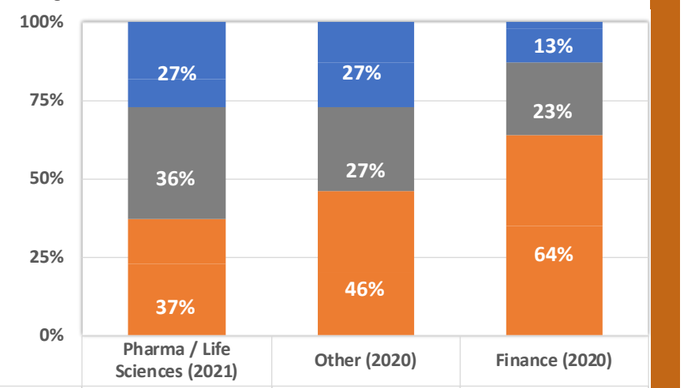
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.21	23%
Other (2020)	3.26	20%
Finance (2020)	3.87	35%

A Data Governance function empowers an organization's ability to make an authoritative decision about data and Data Management. In order for this to occur, a well-defined data strategy has to be communicated to relevant stakeholders to implement processes that allows the organization to achieve its objectives.

The Finance industry is well ahead of the Pharma / Life Sciences and Other industries with almost two thirds of the companies surveyed attaining a Defined / Achieved score. This may be due to heightened regulatory measures introduced after the 2008 financial crisis to eliminate ambiguity and unify Data Management for better outcomes.

Surprisingly, while Pharma / Life Sciences organizations appear to be ahead in other aspects of Data Management, they are lagging behind in establishing the Data Governance function, for reasons that are yet to be explored.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	18%	13%	2%
Planning	9%	14%	11%
Under Way	36%	27%	23%
Defined	14%	26%	29%
Achieved	23%	20%	35%

Survey Results & Observations

Survey Results & Observations

Question 19: Policy and Standards have been written, approved, and implemented

Policy and Standards must be established and approved by stakeholders and executive governing bodies and must reflect the basic principles of how business, technology, and operations functions manage and control data. Policy and Standards must be shared and reviewed by stakeholders and must be supported by audit.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.24	9%
Other (2020)	3.02	14%
Finance (2020)	3.89	34%

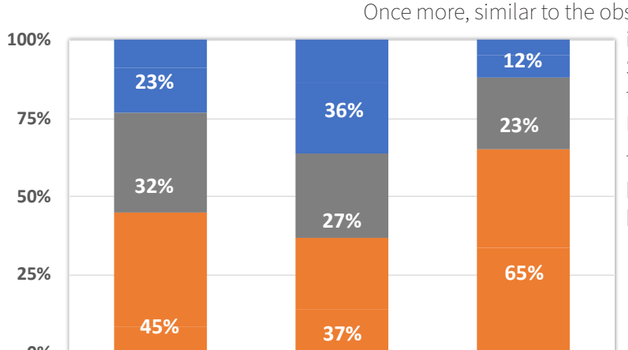
Analysis

- Similar to question 18, the Finance industry leads with a high Average DCAM Score of 3.89 and 65% of respondents reporting their Policy and Standards are at the Defined / Achieved stage.
- The Pharma / Life Sciences industry, 3.24 manages a higher Average DCAM Score than Other industries, 3.02

A Data Governance policy provides the guidelines for an effective Data Management Program, addressing how data is acquired, managed, maintained, and delivered throughout an organization. The Policy and Standards also serve as a foundation to transform the Data Management initiative from conceptual to functional.

Once more, similar to the observations for question 18, the Finance industry is well ahead of the Pharma / Life Sciences and Other industries with almost two thirds of the companies surveyed attaining a Defined / Achieved score.

The Pharma / Life Sciences industry lags behind the finance industry, but fares slightly better than Other industries.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	9%	14%	5%
Planning	14%	22%	7%
Under Way	32%	27%	23%
Defined	36%	23%	31%
Achieved	9%	14%	34%

Question 20: Governance and maintenance of authorized data domains, structures, models, definitions, and glossaries are established and operational

Once established, authorized domains, models, and definitions are the resources used in application development. Governance must be in place to ensure the proper use of these resources, and to ensure data consistency across the organization.

Analysis

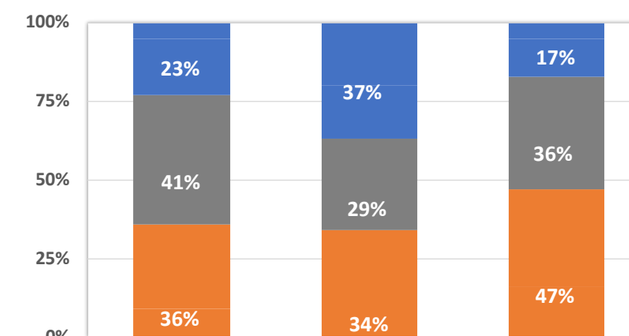
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.21	9%
Other (2020)	2.88	10%
Finance (2020)	3.41	16%

- The Finance industry leads with the highest Average DCAM Score (3.41) and 47% reporting that governance and maintenance of authorized data domains, structures, models, definitions, and glossaries is at the Defined / Achieved level.
- While the Pharma / Life Sciences industry's Average DCAM Score (3.21) is much higher than that of Other industries (2.88), it has a slightly smaller percentage at the Achieved level (9%) compared to Other industries (10%).

Clearly defined data domains, structures, models, definitions, and glossaries, which give structure and meaning to data within an organization, are becoming increasingly important. Through governance and policy implementation in this arena, the structures facilitate the uniform and unambiguous use of data for business needs.

Around half of the Finance organizations surveyed reached the Defined/ Achieved level, which is higher than that observed for the Pharma / Life Sciences and Other industries, showing that many Finance organizations recognized early on the importance of these Data Architecture elements for efficient operational processes.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	20%	5%
Planning	18%	17%	12%
Under Way	41%	29%	36%
Defined	27%	24%	31%
Achieved	9%	10%	16%

Survey Results & Observations

Survey Results & Observations

Question 21: The access and use of data—driven by access controls, data-sharing agreements, and contractual use of data—is governed

Governance processes must be established to ensure control over the identification, definition, and use of data. Governance over data includes policies related to controlling access and use; enforcing contractual restrictions of third-party data, and establishing and monitoring adherence to Data Sharing Agreements.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.77	27%
Other (2020)	2.92	23%
Finance (2020)	3.36	19%

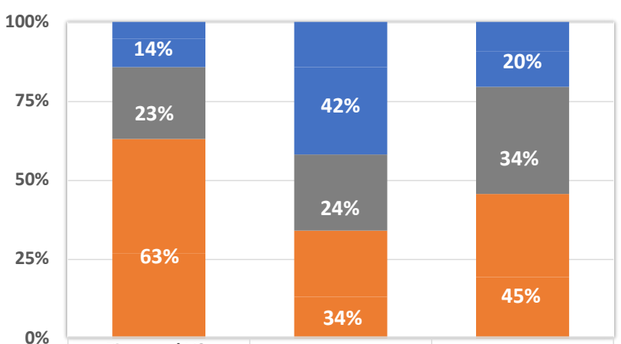
Analysis

- The Pharma / Life Sciences industry leads substantially with the highest Average DCAM Score of 3.77 and 63% of respondents reporting their governance of data access and use is at the Defined / Achieved stage.
- The Finance industry trails behind with a 3.36 Average DCAM Score and 45% of data access and use governance programs at the Defined / Achieved level.
- Other industries report a low Average DCAM Score of 2.92 and 34% of data access and use governance programs at the Defined / Average stage.

Emerging legislation requires organizations to ensure appropriate access and use of data, which is one of the foundational concepts of a well-implemented Data Governance program.

The Pharma / Life Sciences industry attained a high percentage of scores at the Defined/ Achieved level. This may reflect the strict privacy regulations that have been in place for a long time.

The establishment of authoritative data domains facilitates the control of data access and use. The discrepancy between the low Average DCAM Score for the Pharma / Life Sciences industry on Question 20 and its high score on Question 21 bears further analysis. What other dynamics are in the mix?



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	14%	9%
Planning	9%	28%	11%
Under Way	23%	24%	34%
Defined	36%	21%	26%
Achieved	27%	13%	19%

Question 22: The ethical access, use, and outcomes of data are considered, reviewed, and governed

Governing Data Ethics includes establishing a formal Data Ethics oversight function.

Analysis

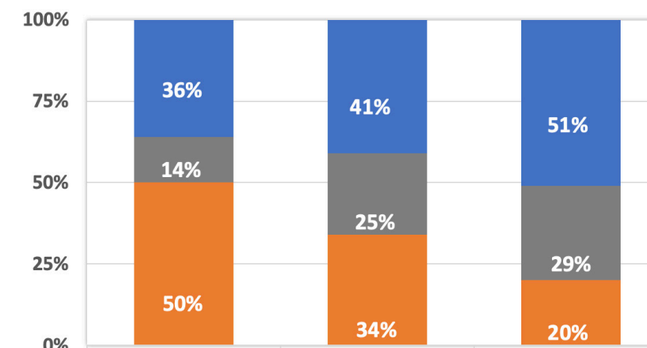
- Pharma / Life Sciences industry leads with a 3.39 Average DCAM Score and 50% Defined / Achieved
- Finance industry has the lowest score of 2.49 and only 20% Defined / Achieved
- In a rare instance, Other industries outperform the Finance industry with respect to both Average DCAM Score and proportion of Data Ethics governance initiatives at the Achieved level.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.39	27%
Other (2020)	2.80	12%
Finance (2020)	2.49	5%

Organizations acquire and analyze data to inform business decisions. As the volume and types of data increase in scale, questions arise concerning what should and should not be done with the data, which creates the need for governing the Data Ethics. For example, Data Ethics controls are needed to detect and mitigate bias in Artificial Intelligence / Machine Learning models and the datasets used to train them, as well as continuous monitoring and review to ensure equitable outcomes.

The Pharma / Life Sciences industry has the highest percentage of Data Ethics efforts at the Defined / Achieved level. As in the previous question, this result may reflect strict, long-ago-established privacy regulations. However, it is concerning that about a third of the Pharma / Life Sciences organizations surveyed have yet to initiate or are just in the planning stages of their Data Ethics governance.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	9%	26%	26%
Planning	27%	15%	25%
Under Way	14%	25%	29%
Defined	23%	22%	15%
Achieved	27%	12%	5%

Though the Finance industry, too, is highly regulated, more than half of the Finance organizations surveyed are at the Not Initiated or Planning stages.

Survey Results & Observations

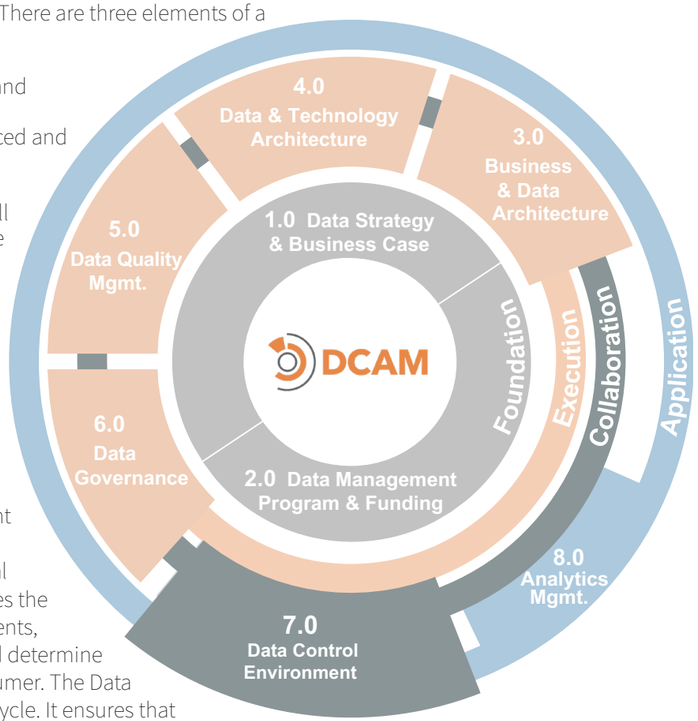
Photo by Sharon McCutcheon on Unsplash

Survey Results & Observations

7.0 Data Control Environment

The Data Control Environment refers to the state of operation in which the data assets of an organization are managed holistically. There are three elements of a successful Data Control Environment.

1. The Data Management (DM) objectives and capabilities described within this document have been embraced and adopted throughout the organization.
2. The data lifecycle is supported fully by all stakeholders. These stakeholders ensure understanding, awareness, and control of data throughout the data supply chain—from source to consumption to disposition.
3. DM is part of the organization's data ecosystem. It is integrated and coordinated with all other control functions organization-wide.



The purpose of the Data Control Environment is to coordinate the people, process and technology of DM into a cohesive operational model. The Data Control Environment defines the mechanisms used to capture data requirements, unravel data flows and linked processes, and determine how data is to be delivered to the data consumer. The Data Control Environment supports the data lifecycle. It ensures that proper resources and controls are in place as data moves throughout its journey..
Also, the Data Control Environment ensures collaboration and alignment to cross-organizational control functions. Areas such as Information Security, Data Privacy, and Change Management must synchronize with DM to ensure data are properly managed across all business functions.

To the extent that the Data Control Environment is not achieved, it presents a potential data risk. Data risk should be managed in alignment with the overall risk management framework of the organization. Data risk scope includes areas such as Data Architecture risks, metadata risks, Data Quality risks, Data Governance risks, and Master Data risks.

Component 7.0

The Data Control Environment ensures data operations, cross-control function alignment, and collaborative Technology Architecture operate cohesively across the data supply chain.

Survey Results & Observations

Question 23: A Data Control Environment and collaboration with cross-organizational control groups is evident and operational

Evidence of the Data Control Environment is the result of effectively integrating the capabilities of data strategy, program, architecture, Data Quality, and Data Governance across the organization. Active stakeholder engagement and cross-organizational control functions are required to ensure the success of the Data Management Program.

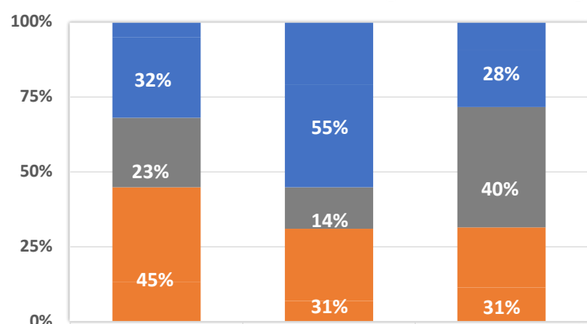
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.23	13%
Other (2020)	2.62	7%
Finance (2020)	3.05	11%

Analysis

- The Pharma / Life Sciences industry leads with a 3.23 Average DCAM Score and 45% report their Data Control Environment is at the Defined / Achieved level, compared to 31% for both the Finance and Other industries.
- More than half (55%) of the respondents in Other industries report their Data Control Environment ranks at the Not Initiated / Planning level.

An ideal Data Control Environment, in which data are curated, governed, and stewarded throughout the data lifecycle—from creation and/or acquisition through consumption and eventual defensible destruction—is a tall order for organizations in the digital era. Collaboration with cross-organizational control groups, as well as other relevant stakeholders, is part of the significant continuous improvement effort required for an effective Data Control Environment.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	5%	21%	9%
Planning	27%	34%	19%
Under Way	23%	14%	40%
Defined	32%	24%	20%
Achieved	13%	7%	11%

Almost half of the Pharma / Life Sciences organizations reached a Defined or Achieved level, which is considerably higher than the percentages for the Finance and Other industries. This might reflect a lower tolerance for avoidable mistakes in the Pharma / Life Sciences industry, along with the implementation of innovative practices to ensure integration and coordination of data processes. The Finance industry has a higher percentage of organizations at the Under Way level, indicating that organizations in this industry are beginning to devote significant resources to achieve effective Data Control Environments.

Question 24: Risks associated with the access and use of data are being tracked, prioritized, and mitigated

The formal process to identiofy data risk must be integrated into the Data Management initiative. Risks must be tracked, prioritized, mitigated, and integrated into the overall risk management framework of the organization (e.g., three lines of defense; risk; audit).

Analysis

- Forty-five percent of survey respondents in the Pharma / Life Sciences industry report that their efforts to track, prioritize, and mitigate data risks ranks at the Defined / Achieved level, with a 3.29 Average DCAM Score.
- Other industries have the lowest Average DCAM Score of 2.81 and the highest percentage, 46%, at the Not Initiated / Planning stage.

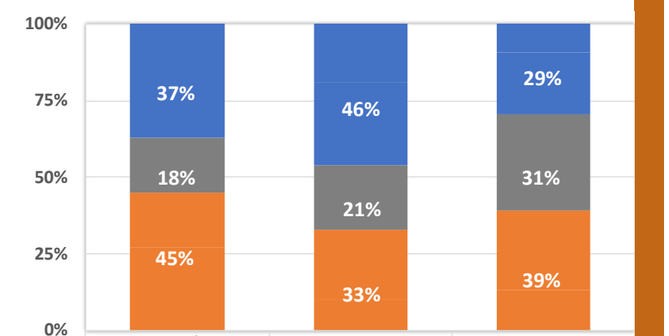
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.29	27%
Other (2020)	2.81	10%
Finance (2020)	3.17	13%

The increasing data volume and rapid adoption of advanced data tools and technologies has made data risk a prevalent concern for organizations. While data breaches targeting large organizations tend to grab the public's attention, often small businesses are targeted in ransomware attacks.

The risks associated with data are interconnected with business and technology functions. A formal process to identify data risks helps organizations address issues from a holistic perspective.

Almost half the Pharma / Life Sciences organizations surveyed reached the Defined/Achieved level, compared to slightly lower percentages for the Finance and Other industries.



	Pharma / Life Sciences (2021)	Other (2020)	Finance (2020)
Not Initiated	14%	19%	9%
Planning	23%	27%	20%
Under Way	18%	21%	31%
Defined	18%	23%	26%
Achieved	27%	10%	13%

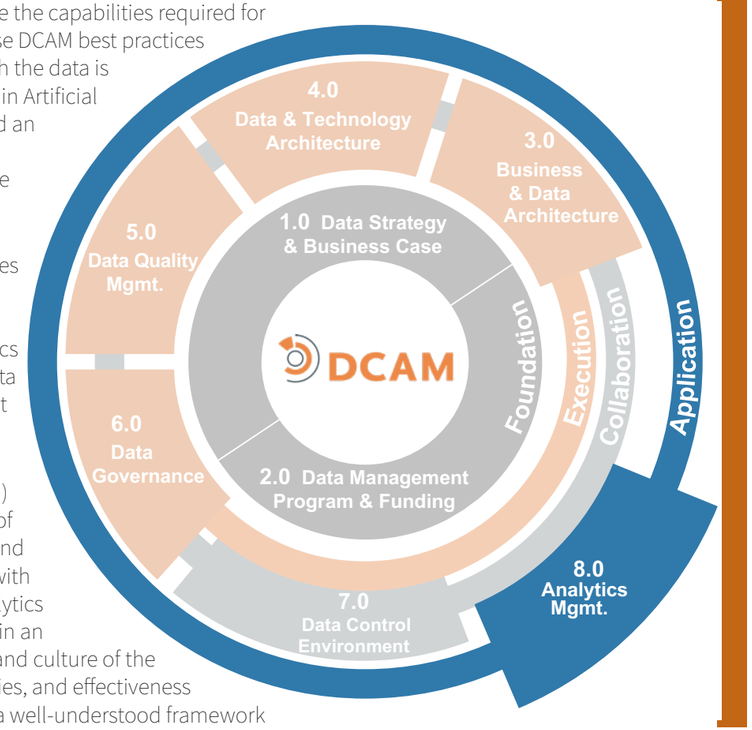
Survey Results & Observations

Photo by Pawel Czerwinski on Unsplash

Survey Results & Observations

8.0 Analytics Management

The first seven components of DCAM define the capabilities required for best practice Data Management (DM). These DCAM best practices guide us regardless of the purpose to which the data is subsequently applied. The issues inherent in Artificial Intelligence / Machine Learning (AI/ML) and an organization's Code of Data Ethics in these components are particularly relevant where Analytics consumes the data.



Consumption of data for analytical purposes is increasingly important for many organizations. Analytics is dependent on high-quality, well-understood data. Analytics functions are, in general, dependent on data produced by areas outside Analytics or that are sourced from outside the organization.

The purpose of Analytics Management (AM) is to formalize how the Analytics activities of an organization are structured, executed, and managed, and to ensure they are aligned with the DM activities. The degree to which Analytics teams are either centralized or distributed in an organization will depend on the structure and culture of the organization. However, synergies, efficiencies, and effectiveness will be maximized if teams operate within a well-understood framework as part of a coherent Analytics strategy.

Component 8.0 was added to DCAM in October 2020. The questions in this section were not included in the 2020 survey performed earlier that year, hence no comparisons may be drawn.

Component 8.0

Analytics Management formalizes how the analytics activities of an organization are structured and performed, ensuring alignment with business priorities and Data Management capabilities.

Question 25: The Analytics function is established

The Analytics function is established formally and governed effectively with documented and approved strategy, operating model, and funding model.

Analysis

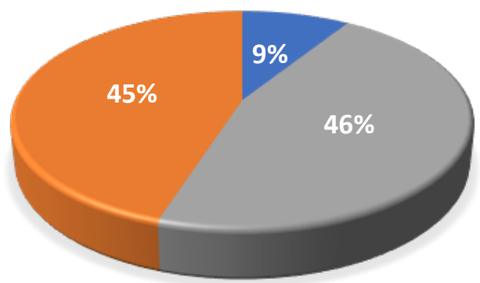
- As with data profiling, 0% of surveyed Pharma / Life Sciences organizations report having Analytics at the Not Initiated stage. In other words, all of the respondents to the 2021 survey report having an Analytics function at least in the Planning stage, with 91% at the Under Way, Defined, or Achieved level.

Analytics involves deriving new insights from data to enhance decision making. Current Analytics trends advance at a rapid pace, and technological innovation arising out of necessity during the COVID-19 pandemic has hastened adoption.

Eighteen percent of respondents report that Analytics are applied to many data processes in the Pharma / Life Sciences industry already. Further, 27% of these organizations are at the Defined stage, and 46% report that Analytics efforts are Under Way. What's more, all organizations surveyed have initiated the Analytics function, at the very least. In addition to deploying Advanced Analytics to tackle complex scientific questions, the popularity of Analytics also is driven by the need to lower healthcare spending and improve patient outcomes.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.64	18%



Pharma / Life Sciences (2021)

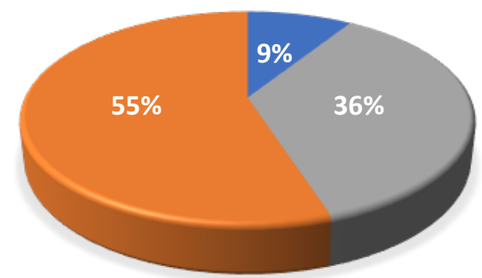
Not Initiated	0%
Planning	9%
Under Way	46%
Defined	27%
Achieved	18%

Question 26: Analytics is aligned with business and Data Management Strategy

The Analytics and DM functions must be aligned and, together, must support the business goals. Analytics must be prioritized to meet the needs of business strategy and drive business value.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.70	23%



Pharma / Life Sciences (2021)

Not Initiated	0%
Planning	9%
Under Way	36%
Defined	32%
Achieved	23%

Analysis

- Again, a small proportion of all Pharma / Life Sciences organizations surveyed report their Analytics efforts are in the Planning stage at a minimum, while the vast majority—91%—place their alignment of Analytics to the business and Data Management Strategy at the Under Way, Defined, or Achieved level.
- Respondents notably report greater maturity in aligning Analytics to the business and Data Management Strategy than in establishing the Analytics function.

  The challenging market landscape in the Pharma / Life Sciences industry compels organizations to consider carefully their business strategy, which should be supported by Analytics and Data Management, to drive business value and growth.

  Responses to this question and its relatively high Average DCAM Score show not only that an Analytics function is prioritized in the Pharma / Life Sciences industry, but also that the Analytics function is perceived to be well aligned with the aims and goals of the business and supported by the Data Management Program.

Survey Results & Observations

Question 27: Analytics is aligned with Data Architecture

Analytics uses approved business glossaries and follows the organization's standards for data sourcing, identification, classification, and metadata. These standards are reflected in a data preparation standard adopted by Analytics.

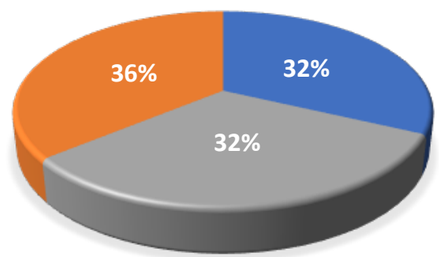
Analysis

- Unlike questions 25 and 26, about a third of respondents (32%) indicate their aligment between Analytics and Data Architecture is Not Initiated.
- Unsurprisingly, the Average DCAM Score for the Pharma / Life Sciences industry is lower (3.18) in comparison to questions 25 and 26 (3.64 and 3.70, respectively).

Data Architecture refers to all the strategies, processes, and methodologies that address how data is designed, including data domains and glossary, logical and physical data requirements, metadata, models, ontologies, and taxonomies, among others. Useful Analytics must be conducted in accordance with approved business definitions and taking into consideration the organization's identification and classification standards. This is crucial for effective interpretation and use of Analytics results.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.18	18%



Pharma / Life Sciences (2021)

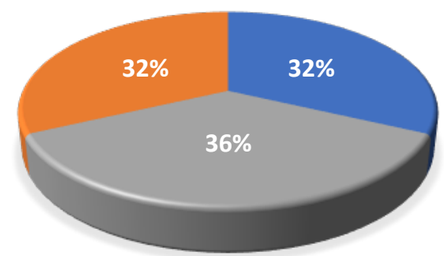
Not Initiated	5%
Planning	27%
Under Way	32%
Defined	18%
Achieved	18%

Question 28: Analytics is aligned with Data Quality

Data Quality measures are produced by Data Quality Management to ensure that data used by Analytics is fit for purpose. Data Quality issues discovered by Analytics are managed using the Data Quality Management framework.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	2.93	5%



Pharma / Life Sciences (2021)

Not Initiated	14%
Planning	18%
Under Way	36%
Defined	27%
Achieved	5%

Analysis

- Similar score distributions to that observed for question 27
- Only 5% of Pharma / Life Sciences organizations report their Analytics function is aligned with Data Quality at the Achieved level.
- Fourteen percent of organizations indicate their Analytics function is Not Initiated.

Data Quality measures are used to determine whether data is fit for its intended purpose. At the very least, a minimum Data Quality standard must be defined and regularly checked. This is important because data of poor quality my lead to skewed interpretations, which may have unanticipated negative consequences.

Question 29: The Analytics platform is designed and operational

The Analytics platform meets the needs of the Analytics operating model, addressing the different requirements for innovation and production. It implements a version control regime and supports defined strategies for data obfuscation. Scalability requirements are understood and supported.

Analysis

- A relatively high proportion (41%) of Pharma / Life Sciences organizations report the design and operation of Analytics platforms at the Defined /Achieved level.
- Again, 0% of Pharma / Life Sciences organizations categorize the design and operation of the Analytics platform as Not Initiated.

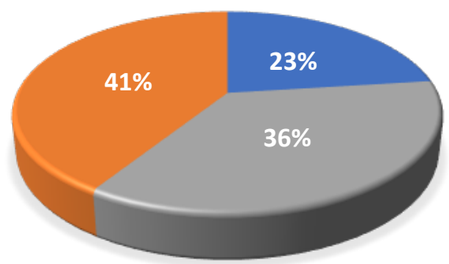
DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.54	23%

The Analytics platform comprises the tools, applications, and infrastructure that enable the functioning of the analytics team to generate information required by the business and stakeholders to make decisions. It is important that the platform aligns with the Analytics operating model approved by all relevant stakeholders.

It is reassuring to see that all organizations surveyed have initiated the design and operation of an Analytics platform. However, the slightly lower percentages of organizations at the Defined / Achieved level, as compared to questions 25 and 26, may indicate that more time is needed for the Analytics platforms to mature, or that operationalizing the Analytics platform design occurs later in the process.

.



Pharma / Life Sciences (2021)

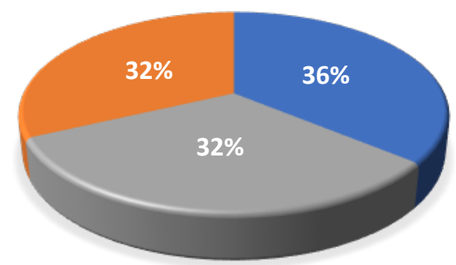
Not Initiated	0%
Planning	23%
Under Way	36%
Defined	18%
Achieved	23%

Question 30: Model operationalization is established

Model testing, approval, release, and regular review processes are in place and are aligned with Data Ethics and privacy governance. Model bias is managed and requirements for model explainability are understood and supported.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.08	9%



Pharma / Life Sciences (2021)

Not Initiated	9%
Planning	27%
Under Way	32%
Defined	23%
Achieved	9%

Analysis

- Again, about one third of respondents indicate that model operationalization is in each of the Not Initiated / Planning, the Under Way, and the Defined / Achieved stages.

Analytics starts with the identification of a business problem, followed by the iterative loop of data processing and analysis, with the aim of developing a deployable model. Analytics activities also may be exploratory or one-off events.

Models must be deployed in a controlled and governed manner. There should be a regular, standardized plan to test, approve, release, and review model deployment.

Model approval should be aligned with Data Ethics and privacy governance. Additionally, special attention should be given to potential biases in the model, including those that arise from programming assumptions and training data. Ongoing monitoring of outcome equity should translate back into model refinement when necessary.

Finally, model deployment should be communicated in a precise and repeatable manner.

Survey Results & Observations

Question 31: The Analytics culture and education needs are managed

The desired behaviors for an Analytics culture are understood and initiatives are in place to address culture gaps. Education initiatives exist to address gaps in the skills required by Analytics practitioners.

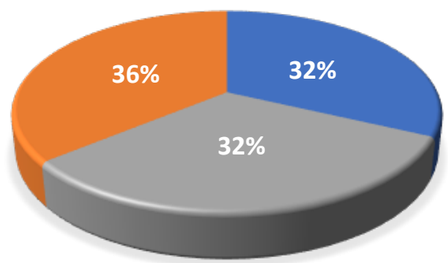
Analysis

- Again, about a third of respondents report that their Analytics culture, and the education initiative needed to support that culture, fall into the Not Initiated / Planning, Under Way, and Defined / Achieved categories, each.

Rapid advancement in technology creates an ever-increasing need for Analytics skills. Organizations take a variety of steps to establish an Analytics culture and education program, including determining the behavioral needs for an Analytics culture, addressing any cultural gaps, and providing educational programs to close these gaps.

DCAM

Industry	Average DCAM Score	% Achieved
Pharma / Life Sciences (2021)	3.64	18%



Pharma / Life Sciences (2021)

Not Initiated	9%
Planning	23%
Under Way	32%
Defined	18%
Achieved	18%

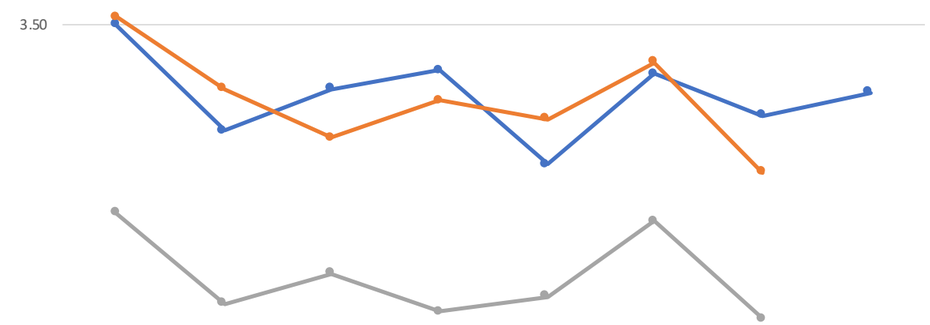
Survey Results & Observations

DCAM®
Composite Scores

DCAM Component	Capabilities	Pharma/ Life Sciences (2021)	Other (2020)	Finance (2020)
Data Strategy & Business Case	Question 1: DM Strategy has been developed and communicated Question 2: Business Case is developed and communicated	3.50	3.00	3.52
Data Management Program & Funding	Question 3: The Data Program is formally established Question 4: Formal Roadmaps are defined Question 5: Process Excellence is established Question 6: Stakeholder Engagement is confirmed Question 7: Communication & Training Programs have been implemented Question 8: Metrics are defined, captured and applied	3.22	2.76	3.33
Business & Data Architecture	Question 9: Data Architecture Program is formally established Question 10: Business Architecture is integrated into the data program Question 11: Logical data domains are identified, documented	3.33	2.84	3.20
Technology Architecture	Question 12: Technology vision and strategy is developed Question 13: Required tools are identified and implemented Question 14: Data Management Operational Risk is in place	3.38	2.74	3.30
Data Quality Management	Question 15: A Data Quality Program is formally established Question 16: Data is being profiled, measured, monitored and maintained Question 17: Root-cause analysis is being performed	3.13	2.78	3.25
Data Governance	Question 18: Data Governance program is established formally Question 19: Policy & Standards are written and approved Question 20: Authoried data domains, models, and glossaries are governed Question 21: Data Access Controls and Contract governance in place Question 22: Ethical access, use and outcomes of data are governed	3.37	2.98	3.40
Data Control Environment	Question 23: Data Control Environment is established Question 24: Data risks are being tracked, prioritized and remediated	3.26	2.72	3.11
Analytics Management	Question 25: The Analytics function is established Question 26: Analytics is aligned with business and Data Management Strategy Question 27: Analytics is aligned with Data Architecture Question 28. Analytics is aligned with Data Quality Question 29. The Analytics platform is designed and operational Question 30. Model operationalization is established Question 31. The Analytics culture and education needs are managed	3.32	–	–

How do industries compare?

Pharma / Life Sciences, Finance, and Other industries scored similarly across each DCAM component, though Other industries scores skew about half a point lower. Given the regulatory scrutiny facing organizations in the Pharma / Life Sciences and Finance industries, this slight deviation may represent the compliance-focused nature of these two industries.



	Strategy & Business Case	Program & Funding	Business & Data Architecture	Data & Technology Architecture	Data Quality Management	Data Governance	Data Control Environment	Analytics Management
Pharma / Life Sciences (2021)	3.50	3.22	3.33	3.38	3.13	3.37	3.26	3.32
Finance (2020)	3.52	3.33	3.20	3.30	3.25	3.40	3.11	
Other (2020)	3.00	2.76	2.84	2.74	2.78	2.98	2.72	

Pharma / Life Sciences (2021) Finance (2020) Other (2020)

DCAM®
Composite Scores

DCAM®
Composite Scores

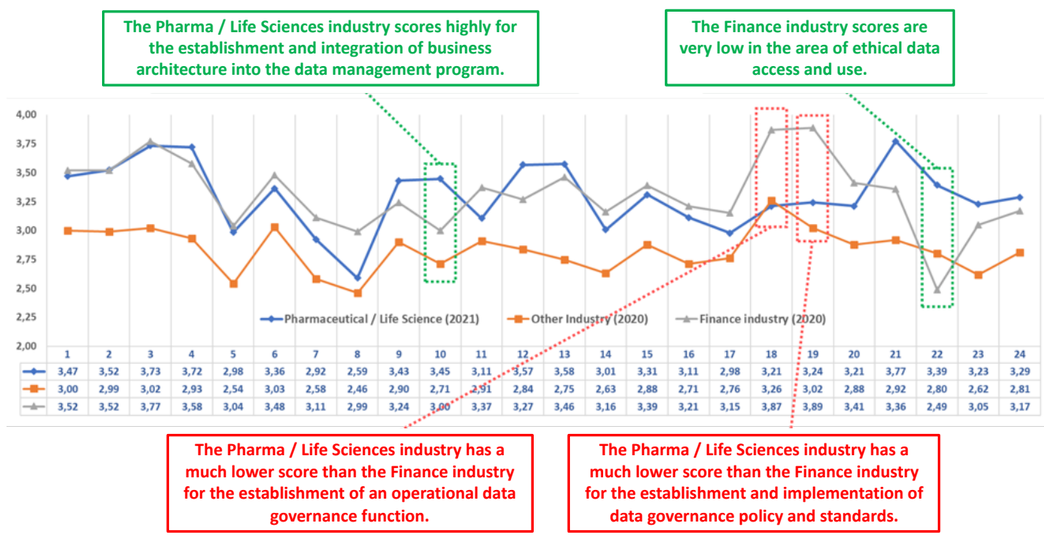
In general, as with the component scores, the Pharma / Life Sciences and Finance industries have similar tendencies in their scores. Noticeable inconsistencies have been highlighted in the figure.

The Pharma / Life Sciences industry scored considerably lower than the Finance industry for the establishment of an operational Data Governance function, having a score similar to Other industries.

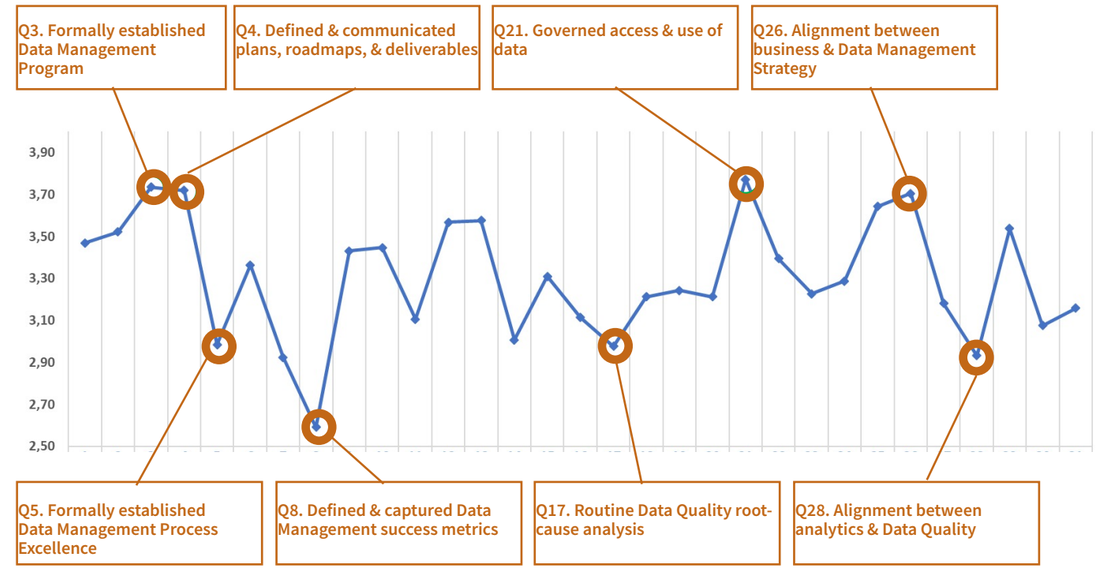
The Pharma / Life Sciences industry had the highest percentage of organizations at the Not Initiated stage, higher than even that of Other industries. Not surprisingly, similar patterns and scores also are observed for the the establishment and implementation of Data Governance Policy and Standards.

For the ethical access to and use of data, the Pharma / Life Sciences industry scored considerably higher than the Finance industry, which ranked low in this area.

In addition, the Pharma / Life Sciences industry scored considerably higher than the Finance industry in the establishment and integration of Business Architecture into the Data Management Program.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Pharmaceutical / Life Science (2021)	3,47	3,52	3,73	3,72	2,98	3,36	2,92	2,59	3,43	3,45	3,11	3,57	3,58	3,01	3,31	3,11	2,98	3,21	3,24	3,21	3,77	3,39	3,23	3,29
Other Industry (2020)	3,00	2,99	3,02	2,93	2,54	3,03	2,58	2,46	2,90	2,71	2,91	2,84	2,75	2,63	2,88	2,71	2,76	3,26	3,02	2,88	2,92	2,80	2,62	2,81
Finance industry (2020)	3,52	3,52	3,77	3,58	3,04	3,48	3,11	2,99	3,24	3,00	3,37	3,27	3,46	3,16	3,39	3,21	3,15	3,87	3,89	3,41	3,36	2,49	3,05	3,17



How do industries compare?

In general, the Pharma / Life Sciences industry has demonstrated Data Management capability at the developmental level. This means that engagement is under way; stakeholders are being recruited and initial discussions about roles, responsibilities, standards, and processes are continuing. However, there is room for improvement.

The highest score achieved was for the governed access and use of data (Q21), driven by access controls, data-sharing agreements, and contractual use of data, which likely is motivated by strict legislation. This is followed by the establishment of a formal Data Management Program (Q3) and the corresponding plans, roadmaps, and deliverables (Q4), which must be communicated to stakeholders. The alignment of Analytics with Business and Data Management Strategy (Q26) also ranks highly, reflecting the growing emphasis on deriving return on investment from data and analytics.

Meanwhile, the Pharma / Life Sciences industry obtained the lowest score for the definition and capture of Data Management success metrics (Q8), which might represent the difficulties faced in measuring the benefits of successfully implementing a Data Management Program.

Alignment between analytics and Data Quality (Q28) and routine Data Quality root-cause analysis (Q17) also drew low scores, highlighting the challenges inherent in developing and adopting standard measures for Data Quality.

Finally, the Pharma / Life Sciences industry also derived low scores on the questions addressing formal establishment of Data Management Process Excellence (Q5).

DCAM®
Composite Scores

2021 Benchmark Advisory Team

EDMCouncil

JOHN A. BOTTEGA

President

John Bottega began working with the EDM Council as an industry contributor in 2005 and served as Chairman from 2007 to 2014. He joined the Council's executive team as a Senior Advisor in 2014, took over as Executive Director of the EDM Council in 2017 and was promoted to President in 2020.

John is a senior strategy and Data Management executive with more than 30 years of experience in the industry. Over his career, John has served as Chief Data Officer in both the private and public sectors, serving as CDO for Citi and Bank of America, as well as for the Federal Reserve Bank of New York.

EDMCouncil

MICHAEL MERITON

Co-Founder & COO

Mike Meriton is a Co-Founder of the EDM Council and served as the first Chairman and active Board member since inception in 2005. Mike joined in 2015 as a Senior Advisor, promoted to COO in 2020, to lead Industry Engagement strategy, new member services and Council Operations.

Previously, Mike was the CEO of GoldenSource and held key executive roles at CheckFree (Fiserv), D&B and Oracle.

EDMCouncil

COLIN GIBSON

Project Manager

Colin Gibson is a senior executive and consultant with 30 years' experience in financial services. He specialises in architecture, Data Governance and data management, most recently as head of data architecture for the investment banking divisions of RBS and HSBC and as head of enterprise architecture for Willis Towers Watson.

Colin combines an enthusiasm for all things data with experience in software development, technology operations and management of major change programmes.

EDMC Partner Affiliates

OSTHUS

ABHISHEK PRASAD

Senior Consultant
Data Governance & Strategy
DataGovernance@osthus.com

Abhishek Prasad is a senior consultant with over a decade plus of experience consulting top tier R&D Life Science, Asset Management, and Retail firms across EU/UK/US. He is an expert in data strategy and governance and advises CDOs on the same.

He has the proven ability to lead both large teams and large-scale projects from the design phase all the way through to the end-user implementation. He's Also DCAM accredited, TOGAF Certified. He's also implemented various Data Catalog solutions for clients coming from a spectrum of industry domains.

ACCURIDS

HEINER OBERKAMPF

Co-Founder & CEO
heiner.oberkampf@osthus.com

Heiner Oberkampf is the co-founder and CEO of ACCURIDS and Head of Data Governance at OSTHUS. He is advising large organization regarding their information and Data Governance strategy.

In 2019, Heiner co-founded ACCURIDS to create a software product that allows to register data and lookup information in a data-centric and application agnostic way and thus providing a long term stable data foundation for the digital era..

strategy&
Part of the PwC network

HOLGER SCHMIDT

Partner
holger.schmidt@pwc.com

Holger Schmidt is a Partner within Strategy& Germany and member of the global Health Practices. He has more than 18 years of experience advising executives in the health industry particularly with pharmaceutical, biotech, and medical device companies and providers. He is an expert for R&D and commercial strategy, as well as digital and data strategy.

strategy&
Part of the PwC network

RALF SCHOENFELD

Director
ralf.schoenfeld@pwc.com

Ralf Schoenfeld is a Director within Strategy& Germany and member of the European Pharma Life Science Practice. He has more than 8 years of experience advising clients across the healthcare and life sciences industry globally. He specialized in (digital) commercial excellence and data strategy.

EDMCouncil

OSTHUS

strategy&

Part of the PwC network

edmcouncil.org

info@edmcouncil.org