Journal of Innovation
February 2025 | 26th Edition

Object Management Group.

Unlocking The Full Potential of Enterprise Data

Managing Valuable Data Assets Through Their Lifecycle

Author:

Bassam Zarkout
IGnPower Inc.
bzarkout@IGnPower.com

CONTENTS

FIGURES

TABLES

This article is about unlocking the value of enterprise data[1] in industries[2] and aims at prompting discussions within the C-suite about the criticality of this important issue. The article takes a data-centric perspective and focuses on the necessity of managing enterprise data assets through:

- The different stages of Digital Transformation,
- The different lifecycle tracks: transactional, operational, business, compliance, legal, and archival,
- The migration of these data assets across dataspaces and systems.

The article posits that enterprise data assets can hold substantial tangible and intangible value, which can be realized in the present or the future. It also introduces the concept of Digital Enterprises—organizations that have achieved advanced levels of digitization and digitalization. With their digital DNA, these enterprises are better equipped to innovate, disrupt markets, and accelerate revenue growth. They generate vast amounts of rich and granular data in diverse modalities and from sources such as digital twins, AI, IoT systems, business systems, messaging and collaboration systems, and dataspaces.

This article also highlights the importance of harvesting, protecting and fostering the value of enterprise data assets, and proposes a path to innovative strategies, methods, and programs for managing the lifecycle of these assets and safeguarding their value.

1 Digital Transformation

The article begins by examining at a high level the topic of Digital Transformation (DX) in industry. This subject is not new as it has been debated extensively in the public domain, including publications from the Digital Twin Consortium[3] and Industry Internet Consortium[4], but it is worth describing at a high level.

Definition: Digital transformation is the innovative, principled and strategic application of digital and connected technologies, coupled with organizational and process restructuring, to generate new value for the organization and its stakeholders.

There are many drivers for digital transformation, including:

- Evolving customer expectations
- Rising competitive threats, especially from digital-native competitors

[1] The terms “data” and “information” are used interchangeably in this article (in reality “information” is “data with context”)

[2] Example: manufacturing, cities, power grids, healthcare, and many more.

[3] DTC *www.digitaltwinconsortium.org*

[4] IIC *www.iiconsortium.org*

- Relentless need for more operational efficiencies and cost savings across value chains and ecosystems
- Advanced AI and analytics tools that can unlock further value from big data
- Expansion of digital ecosystems and the opportunity to use them
- Growing market pressures to meet sustainability goals

The disruption brought about by digital transformation can take many forms, such as:

- Disrupt markets and avoid being disrupted,
- Leapfrog the competition,
- Improve the ROI,
- Ensure long-term viability and success.

Over the past decade, industries have witnessed a surge of innovative and transformative solutions, driven by a vast array of emerging and emergent digital technologies, including digital twins, IoT, AI, distributed ledger, XR, and more.

Figure 1-1 illustrates that digital transformation is not a project. Rather, it is a journey with multiple stages, which in some cases can slightly overlap.

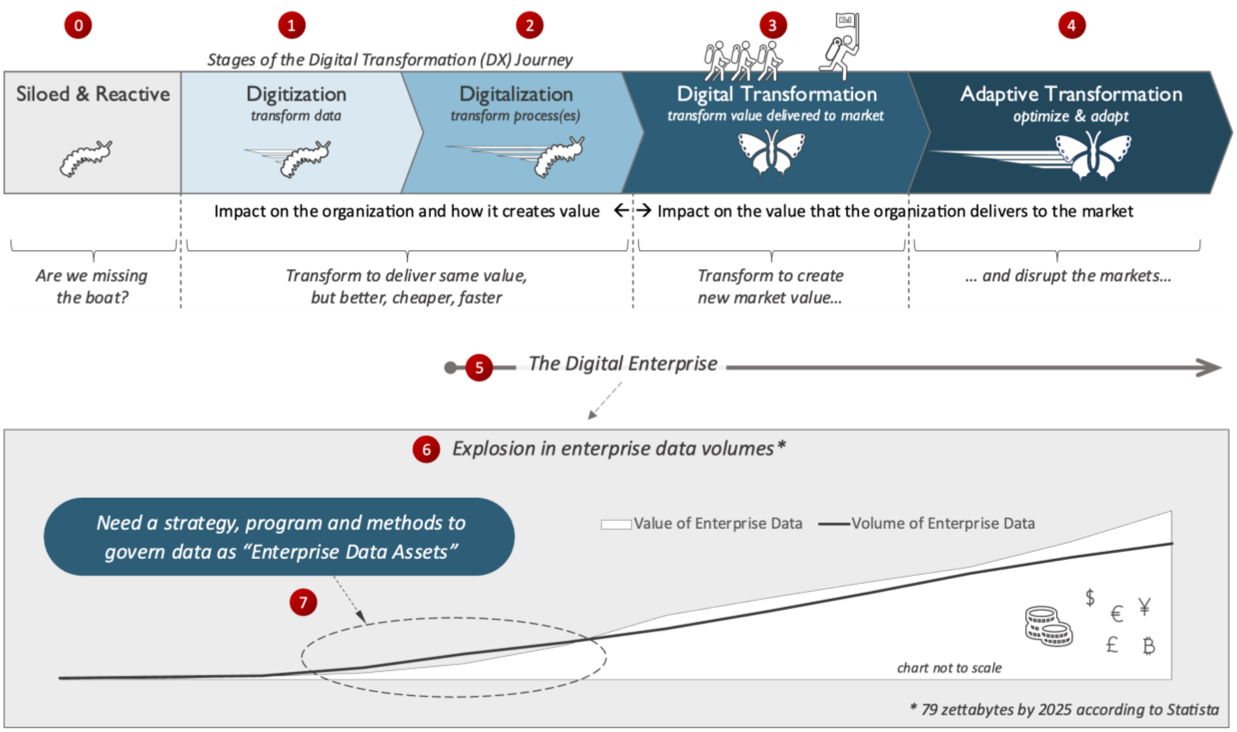

Figure 1-1: Digital transformation and the digital enterprise. (Source: IGnPower)

Note: Explanations ⓿ - ❼ below refer to Figure 1-1.

1.1 DIGITAL TRANSFORMATION IS A JOURNEY

Before embarking on the DX journey ⓿, organizations often face numerous disjointed processes and data silos, many of which still rely on analog data. Such organizations are typically slow to react to challenges, lack agility, and are vulnerable to market changes and threats from innovative digital-native competitors.

Digitization ❶*:* This stage focuses on converting analog data of core processes into digital form and integrating them into processes. Eliminating reliance on analog data addresses a major source of inefficiencies in white-collar and blue-collar work environments.

Digitalization ❷*:* In this stage, the focus shifts to incorporating digital data into core processes, optimizing these processes, and on harmonizing and integrating them. This applies to both internal and cross-ecosystem processes. The main benefits here include streamlined operations, data-driven decision making, and cost reduction.

> Note: With digitization and digitalization, the organization continues to produce the same value for its market constituents but does so more efficiently and more cost-effectively. Some of these benefits can be passed on to customers through lower prices and more reliable product availability.

Digital Transformation ❸*:* This stage leverages the work done during the previous stages and focuses on transforming the value delivered to customers and other stakeholders in the ecosystem. Outcomes may include new business models and new types of product offerings, for example product-as-a-service. The objective is to innovate, become customer-centric, leapfrog the competition, enter new markets, and potentially create blue ocean markets.

The DX journey continues beyond this point ❹, where the focus shifts to optimizing and scaling of the transformed business to grow new revenue streams, accelerate profits, leverage blue ocean markets, and ultimately future-proof the organization.

1.2 DIGITAL ENTERPRISES

During the digital transformation (DX) journey, the DNA of the enterprise evolves, leading to the emergence of the digital enterprise ❺. This transition occurs when the organization achieves significant levels of digitization and digitalization, alongside organizational and cultural changes, and the seamless optimization and integration of physical and digital spaces. Becoming a digital enterprise is a relatively well-charted and highly rewarding endeavor, that results in a more valuable and agile organization[5]. It enhances the organization's ability to pivot, disrupt business and markets, and integrate new technologies more effectively and efficiently.

[5] A major global bank (top 5) involved in M&A activities has stated that becoming a digital enterprise can add 10% to the market valuation of an organization.

Moreover, becoming a digital enterprise is a major steppingstone towards achieving true market-facing digital transformation, which provides the greatest value and competitive advantage. Digital enterprises are at the heart of the concepts of digital engineering and digital threads.

1.3 Big Data

As an enterprise becomes a digital enterprise, the volume of data produced by a proliferating number of digital systems (digital twins, AI, IoT and more) and consumed by the digitalized processes will skyrocket ❻. Examples include IoT sensors that generate massive amounts of operational data, supply chain partners that exchange significant volumes of data related to inventory and product and component information. This data can be fed into digital twins and analytics engines to generate insights, make decisions, and create even more data which in turn may be exchanged again with ecosystem partners.

The resulting volume of data exhibits many typical "V" characteristics associated with *big data*:

- *Volume:* The volume of this data (operations, engineering, etc.) is massive[6] and potentially dwarfs the volume of corporate information produced by business systems.
- *Variety:* Data used by digital twins and IoT systems is quite different from data used in traditional business systems in both[7] format and modalities. There are also differences in architecture and data sources (edge) and storage locations within the infrastructure. As a result, traditional IT best practices developed for business systems need to be adapted or even redefined.
- *Velocity:* Given the highly distributed nature and demanding performance[8] requirements of IoT systems (sensors/edge/cloud), data moves through the enterprise architecture and beyond at high velocity and in all directions. This requires the organization to address the unique performance, latency, scalability and security characteristics of industry data, in all their states: at-rest, in-motion and in-use.
- *Value:* For many organizations in industry sectors, industry data represent valuable Enterprise Information Assets which must be governed. More about this in Chapter 2.

Other V-characteristics of big data apply here, for example veracity and variability.

1.4 Relationship between Industry Digital Transformation and Data

The interplay between digital transformation and data volume is symbiotic. DX initiatives generate vast amounts of rich and granular data that informs decision-making, optimizes processes, and personalizes user experiences. In turn, this data accelerates transformation by providing deeper insights, enabling process automation, and fostering innovation and agility. It

[6] More than 70 zettabytes of IoT data in 2025 according to Statista.
[7] Time series, spatial data, indexed video and images, IR video and images, etc.
[8] Example: low latency.

also facilitates the scaling of knowledge across the enterprise and its ecosystem[9]. In this feedback loop dynamic, increased availability of rich and granular enhances the speed and quality of transformation, which in turn generates even more valuable data.

2 Enterprise Data Assets

The adage that "data is the new oil" remains prescient and valid and should apply to both IT-based business data and OT-based operational data[10] in industry. In digital enterprises, the ability to capture, produce, consume, process, store, and protect data, extract insight from it, exchange it with ecosystem partners, migrate it to other systems and other custodians, and eventually dispose of it when it is no longer needed[11] is critical.

These capabilities are essential for the operation and well-being of enterprises and their ability to create and deliver new value to the market and stakeholders. Therefore, enterprise data should be managed as a critical and valuable resource and asset ❼.

2.1 Enterprise Data as Valuable Assets

Section 1.2 discussed the notion that data is a critical resource and a major byproduct of digital transformation and digital enterprises, driving processes and the overall business of the enterprise. In a growing number of organizations, data is collected, cleansed, integrated and repackaged to be sold to partners in the ecosystem[12].

The Conceptual Framework for Financial Reporting publication[13] offers the following definition of an asset: *It is a present economic resource controlled by the entity as a result of past events, and where the economic resource is a right that has the potential to produce economic benefits.* With the rise of internet-enabled Operational Technology (OT) systems and the growing role these systems play in enabling operational efficiency, competitiveness, and value delivery, there is no reason to question the notion that data produced by these systems can also be enterprise data assets. For example, many modern OEMs regularly receive operational data from the IoT-enabled equipment they sell to customers and use this data to improve product design, performance, reliability, durability, serviceability and circularity.

Figure 2-1 breaks down the data value into four main areas.

---

9 *https://hbr.org/2021/11/the-essential-components-of-digital-transformation*

10 Example: digital twin and IoT data.

11 End of Life (EoL)

12 Smart Cities: Connectivity and Data Trend Report 2023 *https://www.smartcitiesworld.net/trend-reports/connectivity-and-data-trend-report-2023-v1*

13 *https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/conceptual-framework-for-financial-reporting.pdf*

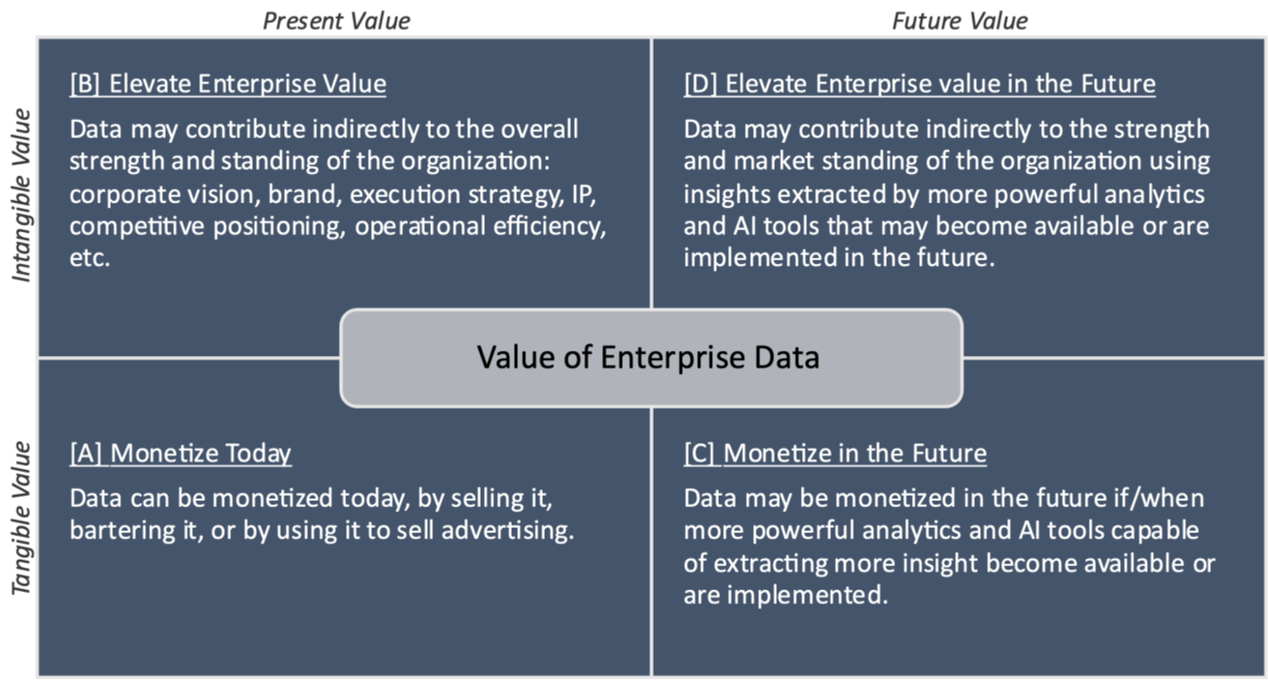

Figure 2-1: Value of data, tangible vs intangible, present vs. future. (Source: IGnPower)

Quadrant [A]: This is data, whether raw or processed, that has been captured, cleansed, and packaged for resale and monetization within the organization's ecosystem. The value of this data is tangible and usually quantifiable. For example, in some smart cities, some of the data collected from sensors is processed, packaged and sold to businesses[14], such as retail stores seeking insights about foot traffic patterns. Entities involved in the sales of data have well established frameworks, methodologies and best practices for quantifying this value.

Quadrant [B]: This data, whether raw or processed, may not be intended for resale but it may contribute significantly to the organization's overall strength and standing. This value is intangible and difficult to quantify as it may be a mix of the following:

- Shaping and refining the corporate vision and optimizing its execution strategy,
- Improving brand image,
- Enriching corporate intellectual property,
- Enhancing competitive positioning,
- Strengthening legal defensibility,
- Boosting operational efficiency,
- Improving product design and recyclability.

Quadrants [C] and [D] are analogous to quadrants [A] and [B], but they represent values that may become accessible in the future with the deployment of advanced analytics and AI tools (which may not exist yet). For instance, AI is increasingly proficient at analyzing X-ray images to detect

[14] *https://bloombergcities.jhu.edu/news/5-data-practices-demonstrate-whats-next-cities-everywhere*

complex lung cancer cases or identifying patterns in patient biodata that can help with early disease diagnosis. The value of this data, whether tangible or intangible, is a future value.

> Note: A data set can belong to more than one quadrant. For example, the data set may be monetizable today and at the same time it may represent intangible value that can elevate the enterprise value.

The value of enterprise data assets may also have multiple dimensions, namely:

- Transactional value
- Operational Value related to the physical assets, customer or partner that the data is associated with
- Business value
- Legal and compliance value
- Archival value

These value dimensions and their lifecycle tracks are discussed in detail in section 3.2.

2.2 Quantifying Data Asset Value

The methodologies and best practices for quantifying the intangible value of data assets are still in their infancy. For example, how do you assess the value of product usage data when it is used to enhance product design in terms of reliability or durability.

At the MIT Information Quality Industry Symposium in 2011, Doug Laney (Deloitte) made a presentation titled "*Infonomics: The Economics of Information and Principles of Information Asset Management*"[15]. Laney continued his infonomics work at Gartner where the term was later defined as "*the emerging discipline of managing and accounting for information with the same or similar rigor and formality as other traditional assets and liabilities)*." The characteristics of "valuable data" include utility, scarcity, accuracy, relevance, actionability, and many more.

Since that period, several organizations[16] have conducted research on the subject, with the ultimate goal to develop frameworks and methodologies for quantifying the value of information assets. The focus then was on data related to IT-based business processes. The subject of infonomics for industry data is virtually nonexistent.

2.3 Not All Data Have Lasting Value

The value of enterprise data is highly contextual and directly related to the role this data plays in processes and business functions. Value also varies at different stages of the data lifecycle.

---

[15] *http://mitiq.mit.edu/IQIS/Documents/CDOIQS_201177/Papers/05_01_7A-1_Laney.pdf*

[16] IBM Switzerland, Haute École de Gestion Genève , MIT, Gartner (Doug Laney), RSD Geneva (Stéphane Croisier), etc.

Consider the temperature readings inside a pressure vessel taken every few seconds (time-series). These readings are crucial for detecting changes in the vessel's conditions. One could argue that not every reading needs to be retained indefinitely; only the changes and their occurrences hold significant value, representing the critical information[17].

In other words, aside from change data, individual temperature readings quickly lose their value. But it can be more complicated. Some jurisdictions may require operators to retain all such data. Other jurisdictions may require the retention of this data as long as the pressure vessel itself is in use. Finally, insurance companies may request specific data retention requirements as a condition of insurance coverage.

Organizations should survey the types of data they produce and consume and assess the value attributes of this data across the full lifecycle. Chapter 3 provides a detailed discussion about the stages in the lifecycle of enterprise data assets.

2.4 What is Included in the Enterprise Data Assets?

Organizations must have a clear definition about the constituent elements of an enterprise data asset, for example:

- Actual data item,
- Creation date/time and versioning data,
- Unique data item identifiers,
- Ownership, right-to-use and custodianship,
- Provenance and lineage,
- Data protection requirements: security, confidentiality, privacy and residency,
- Lifecycle-related data.

2.5 Impact of Data Authenticity on Value

Establishing and maintaining the authenticity of enterprise data assets throughout their lifecycle is foundational to establishing and maintaining the value of these assets. In certain use cases and depending on the industry, the requirements for data authenticity are particularly stringent, especially when it comes to the legal admissibility and defensibility of OT data in the event of accidents, litigations, and investigations.

2.6 Data Ownership and Custody

A critical consideration in assessing the value of OT data involves the access, ownership, and custody of the data generated by systems or equipment. As OEMs increasingly shift their

[17] In information theory, this is referred to as information entropy - *https://en.wikipedia.org/wiki/Entropy_(information_theory)*

offerings to outcome-as-a-service[18] models, some of the operational data produced by the equipment may be transmitted directly to the OEM, in a manner that bypasses the operator's IT infrastructure. This creates challenges for the operators who may lack visibility into data raising significant questions about data ownership[19]. Organizations need a clear understanding of the data produced and consumed by their equipment, including data that bypasses their IT infrastructure. Furthermore, they must be aware of their rights and responsibilities towards this data, given the profound economic, legal, and technical implications.

3 Lifecycle of Enterprise Data Assets

We begin the discussion about the lifecycle of data by outlining several important principles:

- Data (including Enterprise Data Assets) has a lifecycle,
- Data in industry may outlive the systems that created them or are currently storing them,
- The lifecycle of data may be dependent on that of the physical assets they relate to, the customers they relate to, supply chains, procurement, and more,
- Data may need to migrate to other systems during the lifecycle.

The lifecycle starts when the data is created or captured (sensor reading, capture from an external system, etc.) and ends when this data is "no longer needed" (End of Life or EoL). At this point, the data is "disposed of": deleted (in most cases) or in some cases, mainly in government, it is transferred to another entity for long term archival. One should distinguish between the eligibility for disposition and the moment when the actual action of disposition is executed at a technical level. This is because data may be aggregated with other data that may have different lifecycle rules. De-aggregating expired data to dispose of them separately can be very complex and impractical, let alone cost justifiable. Risk management, Legal and IT should collectively discuss these situations and agree on a pragmatic approach to address them.

In industry, the duration of the lifecycle can range from a few seconds or minutes to years and decades, depending on the data type and a variety of other considerations. The lifecycle characteristics of the data type must be expressed in policies that are formally approved and authorized by the enterprise's top management.

3.1 Information Lifecycle Management

The modern principles of information lifecycle management can be traced back to the post-World War II era, when the US Federal Government recognized the need to develop a lifecycle management methodology to manage the vast quantities of paper records they had to:

- Identify records that are "no longer needed" and earmark them for disposal,

[18] Example: Rolls Royce TotalCare® aka propulsion-as-a-service. *https://www.rolls-royce.com/media/our-stories/discover/2017/totalcare.aspx*

[19] *https://www.tandfonline.com/doi/pdf/10.1080/2573234X.2021.1945961*

- Develop classifications for record categories,
- Establish lifecycle retention policies for these categories,
- Categorize remaining assets,
- Enforce the appropriate retention policies on these assets.

The 1990s marked the onset of digitizing and digitalizing front-office and back-office functions, resulting in the generation of massive volumes of electronic records such as documents, emails, transaction reports, and more recently social media data. The lifecycle methodology initially developed for paper records was adapted to manage these new formats. This methodology must be adapted again to suit the new modality of OT data.

3.2 Lifecycle Tracks

The value of data assets, whether tangible or intangible, present or future, encompasses multiple dimensions: transactional value, operational value, business value, legal and compliance value, and archival value. There may be more dimensions to the data value.

Figure 3-1 shows the different lifecycle tracks that are associated with these value dimensions. It also shows a liability track which represents the notion that data assets can become a liability during their lifecycle. Refer to the discussion in section 3.4 for details. The lower part of the diagram indicates that data assets may migrate across multiple systems during their lifecycle. Consequently, the lifecycle policy assigned to the asset must persist through these migrations. Refer to section 3.6 for details.

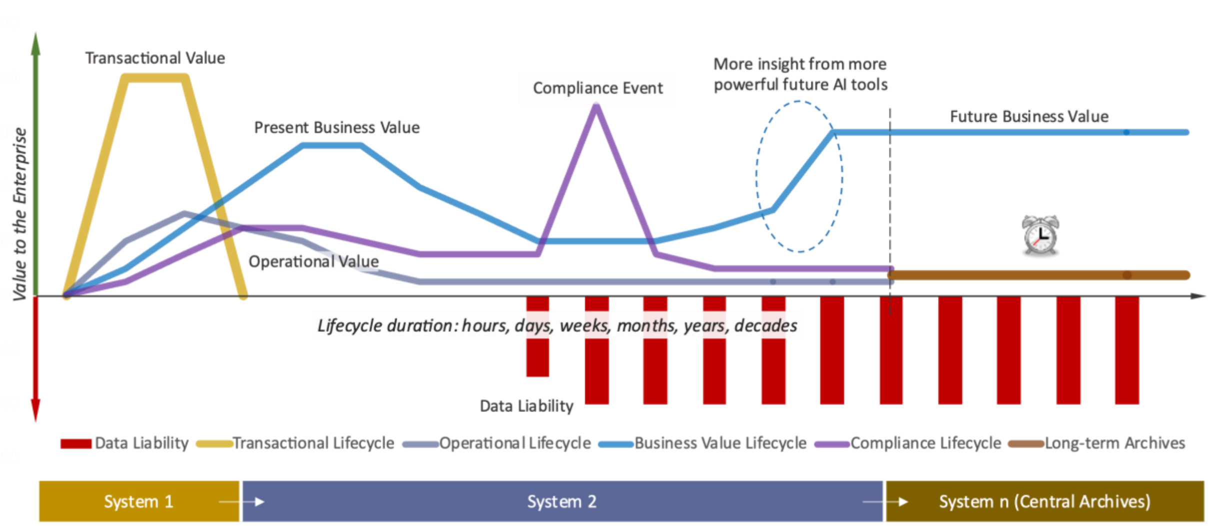

Figure 3-1: Lifecycle tracks of enterprise data assets. (Source: IGnPower)

The lifecycle tracks are outlined in Table 3-1.

Lifecycle Track Description	Data Location	Stakeholders
■ <u>Transactional Lifecycle Track</u> All data in industry have a defined transactional lifecycle, which begins the moment the data is generated or captured and ends when the data has fulfilled its specific transactional purpose and thus is no longer. The duration of this lifecycle track is typically short, and if no additional lifecycle tracks are required, the data is deemed to have reached its EoL point and can be "disposed of."	• Sensors • Edge devices • Digital twins • Analytics/AI models	• Design Engineers • Equipment Operators • Operational Engineers • Safety Engineers • Risk Managers
■ <u>Operational Lifecycle Track</u> Some data may need to be retained for the full lifecycle of the physical asset to which it is associated, or it may be needed to support a particular ecosystem partner or customer relationship. For example, some operational data about a pump may need to be retained for as long as the pump is in use, which may be many years. This means that the organization must implement methods and procedures to associate the lifecycle of these operational data to the lifecycle of these assets.	Depends on the organization	• Design Engineers • Equipment Operators • Operational Engineers • Safety Engineers • Risk Managers
■ <u>Business Value Lifecycle Track</u> Some data items may have business value on top of their transactional and operational lifecycle. This lifecycle track lasts as long as this data assets has business value. The data may be part of an organization's value proposition to its customers, and as described in section 2.1, the value may be tangible or intangible and present or future.	Depends on the organization	• Business Managers • Product Managers • Analytics and AI Engineers
■ <u>Compliance Lifecycle Track</u> Some data assets must be retained as evidence of compliance with regulations, laws, and insurance requirements. Such data must follow a specific lifecycle track, determined by the cumulative requirements of compliance obligations. Developing a lifecycle policy for compliance requires domain-specific expertise: business, technical, legal, and regulatory. The lifecycle policy must incorporate some elements of risk management because some regulations may have conflicting requirements. The policy must also address multi-jurisdictional environments to ensure compliance with laws and regulations across the different jurisdictions. Additionally, organizations should be aware that parties conducting investigations or litigation may be authorized to override existing lifecycle policies and impose their own rules	Depends on the organization	• Compliance Managers • Risk Managers • Legal Counsel • Product Managers • CxOs • CISO • Analytics and AI Engineers

Lifecycle Track Description	Data Location	Stakeholders
on the data and information related to the case (evidentiary and discovery data)[20]. The legal costs involved in the analysis and production of "evidentiary" data can be significant, especially if the data is disorganized, incomplete, orphaned, stored without context, unaccounted for, or if the expired data is not culled.		
■ Archival Lifecycle Track Some data assets (mainly in government) may have long term "archival value" beyond the other lifecycle tracks. Such data assets should be transferred to an "archival authority" where they can be stored for extended periods (decades). The archival process may also encompass a change of custody and ownership of the data. The strategy must also cater for the long-term retrievability of data and its context.	Depends on the organization	• Legal Counsel • Archival and Records Staff • Analytics and AI Engineers
■ Liability Track Refer to section 3.4 for details.	Depends on the organization	• Legal Counsel • Archival and Records Staff

Table 3-1: Description of lifecycle tracks of enterprise data assets.

The principles described in this table have been adopted by global business enterprises for over a decade, primarily applied to business data within IT systems. There is no question that these principles should be applied to engineering and operational data. In fact the case for doing so is more pressing and relevant. As enterprise systems incorporate more digital twins, IoT, AI, and XR sub-systems, the data volumes produced and consumed by these systems are beginning to dwarf those generated by traditional IT-based business systems. Therefore, the lifecycle methodology described earlier must be refocused on managing valuable enterprise data assets in industry-ready knowledge repositories that can scale and adapt to new data sources.

3.3 Data Lifecycle Policies

Organizations must develop lifecycle management policies that comply with diverse regulations and laws across different jurisdictions. These policies should also address the needs of different stakeholders across the lifecycle tracks. Typically, the policies define minimum retention periods, after which disposition may be optional. However, in cases involving privacy and confidentiality requirements, disposition at the end of the retention period becomes mandatory.

[20] US Federal Rules of Civil Procedures *https://www.law.cornell.edu/rules/frcp*

3.4 Data Liability

Electronic discovery or eDiscovery[21] refers to the process of identifying, collecting, and producing *Electronically Stored Information*[22] (ESI) for use in legal proceedings. This can include emails, documents, databases, and other forms of digital data (IoT data for example).

Enterprise Data assets can hold significant evidentiary value, supporting[23] an organization's legal, regulatory compliance, or insurance claims. Conversely, these assets may also bolster the adversary's position[24] in such eDiscovery scenarios, rendering them liabilities.

This raises a critical question: once data has met its minimum retention requirements as stipulated by laws and regulations, can it be deleted despite its liability status? This dilemma involves major ethical, moral, and legal implications, particularly in industries where safety considerations[25] are paramount. Mishandling this issue can lead to severe, sometimes disastrous, repercussions[26] for enterprises.

The *US Federal Rules of Civil Procedure*[27] (FRCP) 2023 mandates strict guidelines regarding the scope and timeframes of eDiscovery processes. They also mandate the suspension of deletion of *responsive data*[28] (ESI) upon the anticipation of litigation. Failure to comply with this requirement constitutes *spoliation*[29], or the destruction of evidence.

Very Important: This article and all its contents do not constitute legal advice regarding data retention policies and enforcement. The organization's Legal Counsel is the ultimate authority on these matters, and IT and OT teams must adhere to their guidelines and instructions.

Organizations should begin with the development and publishing of well-researched and legally defensible data retention policy for all data assets, including OT data. They should then define their disposition enforcement activities based on the following considerations:

- Identify data assets subject to litigation, regulatory, and insurance investigations, and place them "on hold" (suspend disposition activities on them).

---

21 *https://www.proofpoint.com/us/threat-reference/e-discovery*
22 *https://powerhouseforensics.com/ediscovery/types-of-esi/*
23 Innocence
24 Guilt
25 *https://alexanderlaw.com/blog/gas-tank-fires-the-4-9-billion-verdict-against-general-motors-for-the-explosion-of-a-1979-chevrolet-malibu/*
26 What happened to Enron? *https://www.investopedia.com/terms/e/enron.asp*
27 *https://www.uscourts.gov/rules-policies/current-rules-practice-procedure/federal-rules-civil-procedure*
28 Data deemed to be relevant to a case
29 Spoliation is a criminal offense

- Enforce disposition actions on data assets whose lifecycle has expired (and not on hold), ensuring this activity aligns with the published policy and is consistently applied over time.
- Group disposition activities by date range (e.g., monthly) to significantly reduce IT burdens.
- Considering the rapid advances in AI technology and the potential value (tangible and intangible) that may be extracted from data in the future, enterprises should exercise restraint in their disposition activities. Refer to section 2.1.

3.5 Expectations of Legal and Regulatory Authorities

Legal and regulatory authorities expect organizations to *make reasonable efforts* to comply with data lifecycle rules (primarily business data). This term is technology-agnostic and industry-specific. Expectations have evolved over the years to go from merely having up-to-date and published data lifecycle policies and programs all the way to having a legally defensible track record of consistent enforcement. These rising expectations have been driven by:

- High profile compliance events and litigations[30] [31] [32]
- Increasing maturity of data lifecycle management practices in large enterprises
- Emergence of practical, cost-effective solutions that facilitate compliance and administer it

A similar and rapid transition is expected for OT data due to several factors:

- Laws, regulations and best practices for the above have already been developed
- Industry's rich history and culture in compliance with safety and environmental regulations
- Severe implications of non-compliance in industry (safety and environmental impact)

3.6 Migration Across Systems

As mentioned earlier, it is not uncommon in industry for the lifecycle of enterprise data assets to be longer than the lifecycle of the systems that create them. This raises the question about what to do with the data assets when the system reaches the end of its lifecycle. Enterprises have long faced this dilemma. The options are as follows.

Choose the lowest common denominator, which is to "do nothing." By default, this means extending the lifecycle of obsolete systems:

- This keeps data assets "scattered" and "captive" in obsolete applications and systems,
- The value of these data assets declines over time,
- Operational systems may lack the ability to enforce lifecycle management policies,
- The cost of operating and maintaining old systems increases over time,
- This approach becomes a barrier to the evolution of the infrastructure,

---

30 Armstrong vs Executive Office (Iran-Contra) *https://dictionary.archivists.org/entry/profs-case.html* ,

31 Zubulake vs UBS Warburg *https://en.wikipedia.org/wiki/Zubulake_v._UBS_Warburg*

32 Toyota Unintended Acceleration Liability Litigation. *https://www.hbsslaw.com/cases/toyota-sudden-unintended-acceleration*

- This approach can introduce major security risks as new systems tend to be more secure.

A better strategy is to gradually migrate eligible data assets to other systems after their transactional lifecycle track has ended. However, this raises several important observations and questions:

- Can lifecycle policy enforcement processes transcend individual systems and be applied consistently across systems?
- The majority of transactional and operational systems cannot act as Systems of Records[33], therefore the migration of valuable data may be a necessity.
- To ensure the integrity of the lifecycle management process, it is essential to carefully consider which contextual metadata should be migrated with the data asset.

3.7 COST JUSTIFICATION

Managing the lifecycle of enterprise data presents significant cost-justification challenges, particularly if it is done at a granular level. Recognizing this challenge some two decades ago, US NARA recommended a "big bucket" approach that broadened the scope of data and record types and significantly reduced their numbers. This method has since been adopted by many global corporations.

Despite this, the plummeting costs of on-prem and cloud storage have made it increasingly difficult to justify data lifecycle management programs based on complex operational, business, legal and compliance requirements. So organizations are tempted to base their enterprise data lifecycle management programs on pragmatic IT cost considerations alone.

The reality is that the loaded costs of owning and managing valuable enterprise data far exceeds direct storage costs. This includes the cost of disaster recovery systems, IT staff, IT physical infrastructure, energy, security, etc., all of which are rising. Additionally, keeping valuable data in “hot” silos beyond its operational lifecycle burdens IT and OT infrastructure (edge, digital twins, IoT endpoints, etc.), complicates data retrieval and application decommissioning, and challenges disaster recovery. It also hampers compliance and litigation responses. Effective data lifecycle management programs grounded in operational, business, legal and compliance defensibility considerations function as macro filters that enhance the tangible and intangible value of data. In contrast, moving unfiltered data to the cloud results in a growing mess.

Assigning precise cost figures to the above depends on organizational specifics, including its data growth volumetrics, its IT and OT infrastructure, business domains and jurisdictions regulations. The annual cost of owning and managing valuable data can be anywhere from a few thousand

---

[33] *https://en.wikipedia.org/wiki/System_of_record*

US$ per TB to several times that amount[34] [35] [36]. These figures exclude intangible costs such as loss of data value and missed opportunities to leverage data when left scattered.

Organizations should also avoid viewing data and information lifecycle management programs as *solutions in search of problems*. The real cost-justification for these programs lies in the critical role they play in the management of valuable enterprise data assets. Digital transformation initiatives have generated massive amounts of new data, scattered across highly distributed IT and OT infrastructures, including edge, cloud, digital twins, and AI models. By adopting data-centric and value-centric perspectives, enterprises can improve their control over these assets. In this context, lifecycle management is justifiable within a data-centric framework, provided it is implemented pragmatically, effectively and consistently.

3.8 Security Justification

Reducing security risks is one of the key drivers for implementing data lifecycle management controls. With advancements in digital transformation, data should not remain dispersed across IT and OT infrastructures without proper lifecycle management controls. Failure to do so can lead to significant security risks impacting privacy, confidentiality, safety, reliability, and resilience. The subject of IT and OT data security is extensively covered in the market[37].

4 OAIS: Archival Framework (ISO 14721)

The Reference Model for an Open Archival Information System[38] (OAIS) provides hints about effective ways to deal with the question of managing the lifecycle of enterprise data assets. Developed over a decade ago by the Consultative Committee for Space Data Systems (CCSDS)[39], this framework outlines responsibilities and functions for managing and preserving digital information over the long term in a manner that is independent of the applications that created the data and applications used to consume the data.

The OAIS framework recognizes that information can outlive the applications that create them. Therefore, in order for this information to be retrievable in the future, it must be archived outside the applications in an application-agnostic and self-describing manner.

Refer to Figure 4-1 for a high-level functional model diagram:

---

34 *https://www.forbes.com/councils/forbestechcouncil/2021/11/03/uncovering-the-true-costs-of-it-infrastructure/?form=MG0AV3*

35 *https://www.intelligentcio.com/eu/2022/08/16/three-steps-organisations-can-take-to-store-and-manage-data-more-cost-effectively/?form=MG0AV3*

36 *https://www.cloudzero.com/blog/cloud-storage-pricing/?form=MG0AV3*

37 IIC Industry Internet Security Framework 2.0 *https://www.iiconsortium.org/iisf/*

38 *https://en.wikipedia.org/wiki/Open_Archival_Information_System*

39 Includes US NASA, European Space Agency (ESA) and many other global space agencies.

- SIP: Submission Information Package
- AIP: Archival Information Package
- DIP: Dissemination Information Package

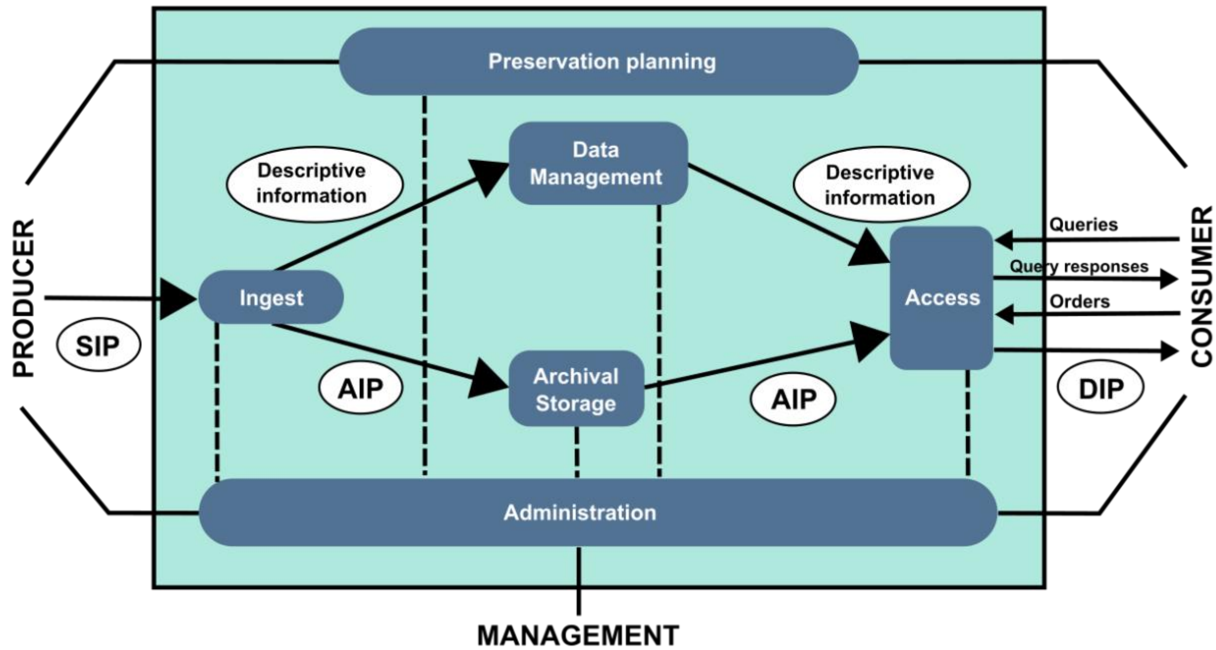

Figure 4-1: OAIS functional model diagram. (Source: CCSD - OAIS)

A detailed description of this framework is beyond the scope of this article. However, it is worth noting that its principles have been adopted in several major information archival projects, some publicly known and some private:

- US National Archives Electronic Records Archive (see below)
- PSA Peugeot Citroën Quality System (see below)
- CEDARS UK
- SIPAD French Space Agency
- NSSDC National Space Science Data Center

Most of these projects predate (at least in their first iterations) some of the modern technologies and architecture that are common in today's IT and Internet-enabled OT environments. However, the principles these projects adopted remain valid and highly relevant to the development of modern lifecycle management solutions for Enterprise Data Assets.

4.1 US NARA ERA

The Electronic Records Archives (ERA) is a program by the National Archives and Records Administration (NARA) designed to preserve and provide access to U.S. Federal Government electronic records. It ensures that these records, which are "born-digital" (created in digital format), are safeguarded and accessible for future generations[40].

ERA was implemented to manage the increasing volume of electronic records and to provide a systematic way to manage, preserve, and access these records long after the original applications are decommissioned. This ensures that vital government records are preserved and accessible, aligning with NARA's mission in the digital age.

NARA adopted the OAIS Archival Framework when developing the ERA:

- Ensure that it can effectively ingest, store, manage, and provide access to electronic records
- Adhere to the principles and processes outlined in the OAIS model to ensure that digital records are preserved authentically and remain accessible over time
- Guide its approach for digital preservation to ensure that it meets future needs

In essence, the OAIS Archival Framework serves as the foundation for NARA's ERA.

The future of the ERA involves continuous improvements and updates to keep up with technological advancements and the ever-increasing volume of electronic records. The system is evolving to better support federal agencies in managing their records and to enhance public access to these records. The recent launch of ERA 2.0 in 2023 marked a step towards more efficient records scheduling and transfer processes.

4.2 PSA Peugeot Citroën Quality System

In the early 2010s, the automotive manufacturer PSA Peugeot Citroën France was faced with an exponential growth of its physical and electronic archives coupled with the necessity to migrate engineering and safety records—about components such as throttle control systems and brakes—across multiple systems in different jurisdictions and throughout their lifecycle. A critical factor in this project at the time was the sudden acceleration safety issue plaguing a major competitor[41] and the fact that in the automotive industry, it is common for component suppliers to provide the same components to multiple manufacturers.

Recognizing this challenge, PSA’s Quality Assurance department initiated a large-scale project to integrate multiple design, manufacturing and administration systems (including SAP, content management, RSD Folders Archive, production line systems, messaging systems, etc.) with a federated information governance platform named RSD GLASS, culminating in their Système d’Archivage d’Entreprise (SAE)[42].

[40] *https://www.archives.gov/research/electronic-records*

[41] Toyota’s Sudden Acceleration Problems *https://spectrum.ieee.org/toyota-sudden-acceleration-problems-what-did-it-know-and-when-did-it-know-it-*

[42] *https://www.alliancy.fr/psa-embraye-sur-le-records-management*

The overall concept was inspired by the OAIS Archival Framework described earlier. The vision was to implement the SAE solution globally and configure the RSD GLASS platform with governance policies capable of accommodating varying legal and regulatory requirements across different jurisdictions.

The initial implementation prioritized safety records for new vehicles, beginning with brake records, encompassing the supply chain, design, inventory, and more.

The SAE project did not incorporate technologies unavailable at the time nor did it focus on managing the value of enterprise data assets. However, many principles and concepts discussed in this article were integral to the solution's vision and implementation.

5 PUTTING IT TOGETHER

This article discussed two interrelated topics: the value of Enterprise Data Assets and the need to manage their lifecycle. Keeping data assets in silos and obsolete systems that are scattered across the IT and OT infrastructures presents significant technical, organizational, and logistical challenges, which are exacerbated by the "illusion" that storage is cheap.

Enterprise data (IT and OT) is a critical resource in digital enterprises. Hence, the question of how to leverage the value of this data and manage its lifecycle as an asset has executive level visibility. CIOs and CISOs should collaborate with their business, operations, legal and compliance counterparts to achieve this.

On the other hand, the expectations of the legal and regulatory authorities regarding compliance are growing and they expect organizations to make "reasonable efforts" in this area. The term "reasonable efforts" is a technology agnostic term that is a moving target evolving with the growing maturity of data lifecycle management practices in organizations.

Enterprises will struggle to harness their data assets' value strategically if they do not incorporate the concepts of data value and uniform data lifecycle management into their enterprise data management strategies. In addition, enterprises need to regularly re-evaluate the value of their data to understand which data need to be refreshed and upkeep – just like any valuable assets which require regular maintenance and upkeep. The exponential growth of data volumes in digital enterprises makes this issue urgent and time critical.

The article proposes that enterprises in industry take the following steps[43]:

- Assess the current state of enterprise data assets, including their data categories, and whether they have value and a lifecycle beyond the transactional phase (refer to section 3.2), and much more.

[43] The author intends to publish additional articles exploring practical methods and systems to address the issues mentioned above.

- Develop a data strategy (or update the current one) to cater for the notion of value of enterprise data assets and their lifecycle as described in Chapters 2 and 3.
- Develop lifecycle management policies that cater for the legal/regulatory (across jurisdictions) and data value considerations and the different lifecycle tracks.
- Develop and implement methods and systems that can enforce these lifecycle policies on enterprise data assets in a legally defensible manner.
- Align the defined systems to enterprise application life cycle management to ensure new deployments integrate by default to these processes.
- Apply these methods in a systematic way with an audit trail.

What is new here is that these steps must be applied to OT data and not only IT data. Unfortunately, most organizations in industry are significantly behind their business counterparts in this regard. In all fairness to them, over the past decade, these organizations have been digitally transforming their OT infrastructure and implementing a completely new generation of solutions based on IoT, digital twins, and AI. Going forward these organizations must deal with their data asset considerations. It is highly unlikely that the legal, compliance and insurance communities will exempt enterprises in industry from their responsibilities towards their industry data, especially when one considers the safety risks.

The time to act is now.

ACKNOWLEDGEMENTS

The views expressed in the *OMG Journal of Innovation* are the author's views and do not necessarily represent the views of their respective employers nor those of the Object Management Group® (OMG®).

➢ Return to *OMG Journal of Innovation landing page* for more articles and past editions.