OvationData

# From forgotten archives to business value –

## the case for clean, structured data in a world of digital clutter

ovationdata.com

# Contents


| | |
|---|---|
| Foreword | 3 |
| Executive summary | 4 |
| Introduction - why clean, structured data matters | 5 |
| The data readiness gap | 7 |
| The business case for structured, clean data | 9 |
| Energy and geoscience - a high-stakes case for data readiness | 12 |
| Heritage - safeguarding memory through structured data | 15 |
| Life sciences and biotech - accelerating discovery through structured knowledge | 17 |
| Building clean, structured data foundations – what it takes | 19 |
| Case study from the market - the value of clean, structured data to Moderna | 21 |
| Ovation's approach - more than just transcription | 22 |
| Conclusion | 23 |

# Foreword

For almost 40 years with Ovation Data Services, I've seen firsthand the impact of clean, structured data in the organisations we work with. The transformation isn't just about files and formats; it's about what data can do, and the business benefits it delivers. In many cases, it means people can make faster, more informed decisions and trust what they see. It's about being prepared when others aren't.

The organisations that will thrive in the next decade are now investing in their information infrastructure. That doesn't mean chasing every new AI tool or digital trend, but starting with the basics. Understanding your data, where it is, and how to make it usable.

In this paper, for the first time, we capture decades of data expertise and reflect on the challenges and opportunities in some of the sectors we serve. We've attempted to speak plainly about the realities many organisations face in managing their data: legacy systems, inaccessible formats, regulatory pressure, and wasted effort. This paper also offers a way forward - practical, scalable, and grounded in real-world experience. At Ovation, we don't approach this work as a vendor ticking boxes but as partners who are curious, methodical, and deeply committed to quality. We know that organisations' data is irreplaceable and treat it with utmost care. We hope this paper will be essential reading for all those thinking about how to get their data into a more useful (and potentially valuable) state, including governments and private enterprises.

If your organisation is serious about getting ahead, this is the right place to start. Read on, and then take action.

**Gregory Servos, Executive Chairman,**
**Ovation Data Services Inc.**
June 2025

# Executive summary

In a world increasingly reliant on digital intelligence, the value of clean, structured data has never been higher. The consequences of neglect can be enormous. Data can be lost, compliance requirements missed, and the potential of AI to transform business operations can go unrealised, leaving companies struggling to compete. Yet many organisations remain unprepared across sectors as diverse as energy, heritage, and biotech. Their data is fragmented, poorly documented, or locked in obsolete formats, undermining strategic goals and limiting their innovation capacity.

This white paper sets out the case for data readiness,  turning dormant, damaged or disordered information into an asset that supports operational efficiency, compliance, and AI-powered insight. It shows how organisations can unlock the value of their data through disciplined structuring and how failing to do so results in lost opportunities, duplication of effort, and reputational or regulatory risk.

Case studies from the energy industry, cultural institutions, and biomedical research highlight the practical implications of clean, structured data. In each case, the transformation was not simply technical; it was strategic. The results include faster decision-making, reduced costs, improved public and stakeholder trust, and stronger resilience.

The paper also outlines a pragmatic approach to data readiness. From audit and recovery to metadata tagging and secure hosting, it sets out the steps required to build a sustainable, accessible data foundation. It shows how Ovation Data, with its cross-sector experience, is helping organisations of all types to move from data chaos to clarity.

For any organisation looking to improve compliance, enable trustworthy AI, or use its intellectual and operational memory better, the message is clear: clean, structured data is not optional, it’s fundamental.

# Introduction – why clean, structured data matters

We are living through the most enormous explosion of data in human history. By 2026, the world is expected to generate more than 221 zettabytes of data annually.[1] To put that into perspective, 221 zettabytes is more data than humanity created in the first 5,000 years of recorded history, being generated every twelve months. But volume is not the same as value. Without structure, much of it is noise. Only a fraction of this information is fit for analysis, insight, or decision-making. Organisations are investing heavily in AI and digital transformation, but most still lack the foundational asset that makes these efforts work: clean, structured, interoperable data.

## Business risks: The cost of data neglect

### Energy and geoscience

**Regulatory non-compliance:** Inaccurate emissions or reporting data can lead to fines, halted projects, or reputational damage.

**Inefficient decision-making:** Disparate and inconsistent data from exploration, drilling, and production reduce forecasting accuracy and increase financial risk.

**Operational downtime:** Inability to track equipment performance and maintenance data consistently increases downtime and safety risks.

**Lost investment opportunities:** Inconsistent data prevents accurate ROI modelling for new wells, fields, or infrastructure.

**Cybersecurity vulnerabilities:** Disorganised systems are harder to secure, increasing the risk of ransomware or data breaches.

### Heritage

**Loss of heritage assets:** Incomplete or inconsistent records of collections increase the risk of physical or digital asset loss.

**Reduced public engagement:** The lack of searchable digital records limits virtual tours, exhibitions, and educational outreach.

**Funding challenges:** Grant applications and donor trust require accurate, well-managed data on collections and visitor impact.

**Inefficient staff workflows:** Staff waste time searching across fragmented databases, slowing curation, research, and programming.

**Data degradation:** Without standardised formats and storage, valuable digital assets can become obsolete or unreadable.

### Lifesciences and biotech

**Regulatory delays or failures:** Dirty or inconsistent clinical trial and compliance data can result in rejected drug approvals or delayed market entry.

**Slower R&D cycles:** Disconnected lab, clinical, and commercial data systems impede insight generation and innovation.

**Increased cost and waste:** Without interoperable systems, redundant tests, manual reconciliation, and inefficiencies increase operational costs.

**Data integrity concerns:** Unclean data jeopardises scientific validity, damaging trust with regulators, partners, and investors.

**Missed AI opportunities:** AI models require high-quality, structured data. Bad data blocks predictive modelling and precision medicine advancements.

1 LTO Program. "Help Me Ronna: AI is Only as Good as the Data It's Given." *LTO.org*, 2 March 2023. https://www.lto.org/2023/03/help-me-ronna

The urgency around AI readiness has brought this problem into sharp relief. Models trained on limited, poor-quality or unstructured data, not only simply underperform; they generate unreliable outputs that can misinform critical decisions. The consequences range from minor inefficiencies to significant financial loss, reputational damage, or regulatory breaches. This risk is amplified in sectors like energy, biotech, and government, where the stakes are high and the datasets are vast. The assumption that digital equals useful is misleading. Without structure, digitised data remains inert. Without context, it becomes a liability.

This paper is not about digitisation in the abstract. It is about something far more specific and consequential - data readiness.

**Clean, structured, interoperable data can be trusted, used, and shared. It is machine-readable and human-understandable. It includes metadata, is standardised against accepted schemas, and exists in formats allowing platform integration. This is what AI needs to function correctly. It is also what auditors, regulators, researchers, and analysts require to do their jobs effectively.**

The distinction between data as an asset and a liability is strategic. When properly structured, data becomes a source of value. It enables better decision-making, accelerates discovery, and supports regulatory compliance. It becomes a burden when left in legacy formats, unstructured folders, or undocumented archives. It incurs hidden costs through duplication of effort, missed opportunities, and loss of institutional knowledge.

Clean, structured data is not just a prerequisite for AI, it is a foundation for long-term resilience, institutional memory, and competitive advantage. This paper explores what it means to achieve that state, why so few organisations have done so, and how to begin the journey from forgotten archives to operational value.

# The data readiness gap

Across industries, the promise of AI and advanced analytics is outpacing the underlying preparedness of data infrastructure. Many organisations claim to be on a digital transformation journey, but their data tells a different story, one of fragmentation, inaccessibility, and disrepair. This disconnect is both a technical inconvenience and a structural obstacle to progress.

In 2023, McKinsey reported that only 15% of organisations fully realise their AI objectives, with the leading cause of failure being poor quality data and siloed systems.[2] The 2025 edition of their report emphasises that most respondents to their survey have yet to see organisation-wide, bottom-line impact from generative AI use. Many organisations are still in the early stages of implementing the adoption and scaling practices necessary to capture value from AI.[3] So while AI adoption is increasing, significant challenges remain in achieving full value realisation, frequently due to persistent data quality and integration matters.

While headlines often highlight breakthroughs in machine learning and predictive modelling, the day-to-day reality inside most enterprises is quite different. Much of their legacy data, whether in oil exploration logs, pathology reports, archival records, or customer files, remains locked in proprietary formats, duplicated across departments, or is missing crucial metadata.

This issue of fragmentation is especially acute in sectors with long operational histories. Energy firms, for example, may hold decades' worth of seismic surveys, production records, and licensing documents. These are regularly found on outdated media such as magnetic tapes or early-generation digital storage. Without proper cataloguing and formatting, these archives are unreadable by modern systems. The financial cost of this gap can be vast: recreating seismic surveys that have been lost or degraded may cost millions, delay exploration, and duplicate effort. In one case shared with Ovation Data, a major offshore operator conducted a full re-acquisition of the survey after discovering that the original data was unavailable in a usable form, setting back the project by over a year.

The heritage sector faces a different version of the same problem. Institutions often possess rich troves of analogue materials - photographs, manuscripts, audio reels - yet frequently lack the tools to convert them into searchable, shareable formats. Without structured digitisation, these materials remain invisible to modern audiences and algorithms alike. The absence of metadata - for example, who created a document, when, in what context - renders even digitised assets opaque. The potential for public engagement, academic research, and educational use is lost simply because the underlying data is not ready.

Biotech firms and research networks confront the challenge from yet another angle. The issue is not analogue-to-digital conversion, but 'digital sprawl'. Laboratories generate vast datasets from clinical trials, field studies, and genomics platforms, which are often stored in inconsistent formats across incompatible databases. Public health data is another resource with significant potential for discovery, including a better understanding of diseases. It too frequently sits in legacy formats, across siloed systems. Even routine analysis becomes time-consuming without a shared schema or data governance framework. According to one article in the Computational and Structural Biotechnology Journal, "data scientists spend a great amount of time and effort on cleaning and organising data, taking up 50–80% of

2 McKinsey & Company. *The State of AI in 2023: Generative AI's Breakout Year*. August 2023. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

3 McKinsey & Company. *The State of AI: How organisations Are Rewiring to Capture Value*. March 12, 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

their work time."[4]

The readiness gap is not only technical but cultural. In many organisations, data responsibility and ownership is unclear. No single person or team owns the lifecycle of legacy information. IT departments may handle storage, but not structure. Business units may know the content, but not the systems. As a result, accountability is diffused, and so too is urgency. Investment in data preparation is often deferred in favour of more visible front-end applications or AI prototypes, even though these depend entirely on what lies beneath.

There is also a psychological cost to data neglect. When teams do not trust their data, they stop using it. This leads to an over reliance on anecdotes, gut instincts, or redundant workarounds. In the worst cases, it breeds a culture of avoidance. The longer this persists, the harder it becomes to restore confidence.

Bridging the readiness gap begins with recognising that legacy data is not historical clutter. It is potential, untapped, dormant, and often irreplaceable. But realising that potential requires structure, not just storage, but accessibility; not just format conversion, but intelligent metadata; not just digitisation, but interoperability.

The following sections of this paper explore what that looks like in practice and what it takes to get there.

4 Caliskan.A., Dangwal.S., Dandekar.T (2023) *Metadata integrity in bioinformatics: Bridging the gap between data and knowledge,* https://www.sciencedirect.com/science/article/pii/S2001037023003616

# The business case for structured, clean data

**"Structured data transforms research outcomes across biotech, medicine, and water treatment. It cuts costs by removing the need to repeat earlier experiments and accelerates discovery by enabling direct access to prior findings. Researchers can identify therapeutic side effects with potential new applications or build on clean datasets to fast-track public health solutions, without re-establishing baselines. In a sector where processing infrastructure is costly, structured data enables targeted analysis, efficient resource use, and evidence-based decision-making. Without it, insights are buried, timelines are extended, and opportunities are lost."**

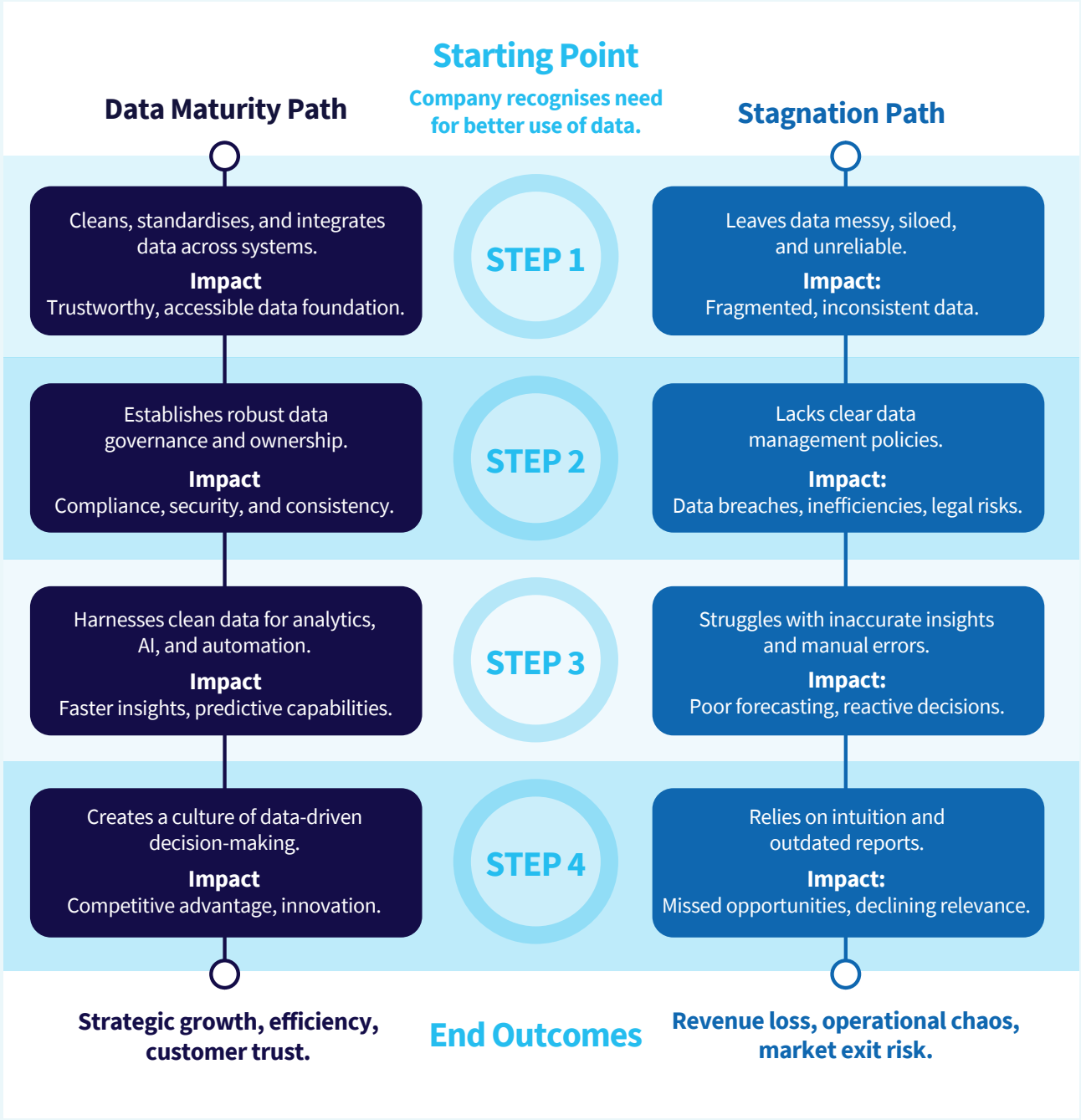
Dr. Anindito Sen, Ph.D., Research Scientist,
Microscopy and Imaging Center, Texas A&M University

Data is more than a digital asset; it is the operational backbone of modern business. For decision-makers in both public and private sectors, clean, structured data translates directly into improved financial performance, operational efficiency, and resilience. The return on investment is measurable and compelling. Moreover, efforts are underway, including by global accountancy bodies, to try and place a value on data (as an intangible asset) so it can be represented on the balance sheet. Efforts in the UK public sector include an analysis by National Highways, which has calculated the value of its data assets at around £60 billion. This is more than half the value of their physical road infrastructure.[5]

Structured, clean, interoperable data also builds efficiency as organisations avoid costly duplication of effort. Industry discussions suggest that poor data practices may cost upstream oil and gas companies up to 10% of their annual capital budgets.[6] These losses are driven not only by repeated surveys and data held in legacy formats but by inefficiencies in workflows that rely on incomplete or inconsistent data inputs. In contrast, Ovation Data's work with clients in the Gulf and North Sea regions has demonstrated that once properly recovered and structured, legacy data can be reused across multiple asset evaluations, reducing expenditure on new data acquisition and accelerating project timelines.

5 Anmut, *Why You Should be Treating Your Data as an Asset,* https://www.anmut.co.uk/why-you-should-be-treating-your-data-as-an-asset/#:~:text=In%20National%20Highways'%20case%20we,National%20Highways'%20physical%20asset%20value.

6 Kinsmen Group, *If data is the New Gold, Where do we Dig?*(2022) https://kinsmengroup.com/wp-content/uploads/2022/04/Data-Gold-Whitepaper.pdf#:~:text=manufacturing%20companies%20,start%20for%20a%20business%20case


Starting Point
Company recognises need for better use of data.
Data Maturity Path
Stagnation Path
Cleans, standardises, and integrates data across systems.
Impact
Trustworthy, accessible data foundation.
STEP 1
Leaves data messy, siloed, and unreliable.
Impact:
Fragmented, inconsistent data.
Establishes robust data governance and ownership.
Impact
Compliance, security, and consistency.
STEP 2
Lacks clear data management policies.
Impact:
Data breaches, inefficiencies, legal risks.
Harnesses clean data for analytics, AI, and automation.
Impact
Faster insights, predictive capabilities.
STEP 3
Struggles with inaccurate insights and manual errors.
Impact:
Poor forecasting, reactive decisions.
Creates a culture of data-driven decision-making.
Impact
Competitive advantage, innovation.
STEP 4
Relies on intuition and outdated reports.
Impact:
Missed opportunities, declining relevance.
Strategic growth, efficiency, customer trust.
End Outcomes
Revenue loss, operational chaos, market exit risk.

Productivity also improves significantly. A Forrester Consulting survey of large organisations found that knowledge workers spend nearly 29% of their workweek (≈11.6 hours out of 40) just searching for information needed to do their jobs.[7] This figure can be reduced significantly when data is structured and tagged. The ability to retrieve relevant information quickly allows teams to focus on high-value analysis and decision-making rather than data wrangling. In biotech, where researchers operate under tight funding and regulatory constraints, this productivity gain can directly affect the speed at which discoveries are made and products reach the market.

Decision quality improves as well. Structured data enables decision-makers to view complete, consistent, and contextualised information, reducing the risk of errors based on partial or outdated inputs. This can mean the difference between proactive intervention and regulatory failure in sectors such as finance or healthcare.

Risk is another primary consideration. Structured data enhances traceability and auditability, making it easier to comply with industry regulations and internal governance standards. The Financial Conduct Authority in the UK and other bodies have increasingly signalled that data integrity will become central to compliance enforcement, particularly in high-risk sectors such as finance, energy, and pharmaceuticals. Ovation's clients have found that preparing legacy data for inspection, rather than simply storing it, is often the difference between a smooth audit and significant penalties.

Finally, structured data underpins the effectiveness of AI and machine learning systems. These technologies rely on clear, well-labelled inputs to function reliably. When data is unstructured, AI outputs are less accurate and harder to interpret. As the Open Data Institute (ODI) has noted, data-centric approaches consistently outperform model-centric ones.[8] Even the most advanced algorithms are rendered ineffective or misleading without clean data.

In the next section, we examine a real-world example of these principles in action, focusing on the energy sector, where the stakes of poor data quality are exceptionally high.

7 VentureBeat, *Report: Data Silos Cause Employees to Lose 12 Hours a Week Chasing Data*, https://venturebeat.com/data-infrastructure/report-data-silos-cause-employees-to-lose-12-hours-a-week-chasing-data/#:~:text=The%20survey%20uncovered%20that%20the,employees%20say%20they%E2%80%99re%20feeling%20disengaged

8 The Open Data Institute, Data-centric AI (2023) https://theodi.org/insights/projects/data-centric-ai/

# Energy and geoscience – a high-stakes case for data readiness

Few industries are as data-intensive, capital-heavy, and risk-sensitive as energy. The sector generates vast quantities of data over the exploration, production, and asset management lifecycle, from seismic surveys and well logs to environmental reports and licensing documents. Yet, much of this data remains trapped in inaccessible formats or scattered across unconnected repositories. This lack of structure doesn't just limit efficiency; it compromises commercial decision-making and compliance readiness at scale.

Take, for example, a recent international example where Ovation Data was engaged with a multi-year geo-spatial data management project. The organisation held extensive subsurface and seismic data archives for years, spanning decades of exploration activity. However, large portions of this data were stored on magnetic tapes and early digital media, with little accompanying metadata. Without intervention, these records were on the verge of becoming unusable.

Working with Ovation Data, the organisation undertook a comprehensive recovery and structuring programme. Data was retrieved from legacy formats, verified, standardised, and catalogued using internationally recognised schemas. The resulting dataset not only supported internal planning and resource management but was also used to underpin a new data business model, increasing transparency and attracting international investment. What had once been an operational liability became a strategic asset.

Conversely, failure to act - perhaps due to short-term cost aversion - can mask deeper strategic vulnerabilities and result in reputational damage, financial penalties and lost competitive advantage. An example was shared by a North Sea operator that faced a costly dilemma. With a significant inspection deadline looming, the organisation needed to retrieve historical geophysical data to demonstrate compliance with environmental and licensing obligations. However, the relevant information had been archived across multiple locations in incompatible formats. This gave them the choice of whether to digitise the records or pay the penalty, weighing up the costs associated with organising data into the required formats versus the likely fine for non-compliance. As Ovation Data's Vice President of Global Operations, James Laming says:

"Regulation alone has not helped to achieve the strategic ambitions of the National Data Repository (NDR). There are numerous reasons for this. The cost of re-mastering legacy data into a digital format and adhering to the standards required, the effort and resources needed to achieve this by the operators, and the notion that data that has required significant investment to acquire in the first place will be (by inclusion in the NDR) made freely available, have disincentivised compliance. In our experience, the sanctions for non-compliance have been ineffective because they represent a fraction of the costs of meeting the requirements."

## Data as part of the energy value chain

These examples highlight a recurring issue in energy: the inability to treat legacy data as part of the value chain. Based on conversations with UK operators, seismic surveys, for example, can cost up to $1 million per day offshore, making their loss or duplication financially severe. Yet, over more than 40 years, Ovation has witnessed organisations routinely repeating surveys simply because the originals cannot be found, accessed, or interpreted in time. According to the International Association of Oil & Gas Producers (IOGP), poor data management practices cause project overruns, increased safety risk, and reduced asset

performance. Strengthening data management, through complete and quality information, and industry-standard data processes, is therefore seen as essential to controlling project costs, ensuring safety, and maximising asset value.[9]

Structured data is not only about efficient retrieval. It also plays a critical role in mergers and acquisitions, joint ventures, and asset divestment. Buyers and regulators demand clear, auditable records to assess risk and validate operational claims. Firms that cannot deliver structured datasets face higher legal and due diligence costs, and often command lower valuations. For example, during a recent M&A event where a large organisation was acquiring the assets of another large organisation, it was identified during due diligence that one significantly important piece of data could not be adequately attributed in ownership. The result was a $50m fine.

Data readiness quickly becomes a competitive differentiator as the sector confronts the twin pressures of decarbonisation and digitalisation. Structured information allows for better modelling of environmental impacts, smoother integration of renewable assets, and more efficient planning for decommissioning and repurposing infrastructure.

> **"Oil and gas data doesn't arrive in neat spreadsheets, it comes from seismic sensors, aerial surveys, obsolete tape reels and decades-old storage formats. Recovery isn't just copying files; it involves assessing whether data is usable, correcting for loss or corruption, and converting it into structured, searchable formats. That means not only making sense of petabytes of raw inputs but also transforming them into actionable insights that can drive exploration, energy transition planning, and regulatory compliance. Without this step, vast archives remain inaccessible - valuable but effectively invisible."**
>
> Vijay Kumar, Project Manager QC,
> Seismic Data, Ovation Data

9 IOGP, Geomatics, https://www.iogp.org/workstreams/engineering/geomatics/ and IOGP, Data Standards Opportunity Survey, https://www.iogp.org/wp-content/uploads/2019/12/Data-Standards-Opportunity-Survey.pdf#:~:text=Analysis%20of%20all%20the%20information,positive%20impact%20on%20asset%20availability

## Energy and geoscience – turning legacy data into a strategic asset

**Too valuable to lose, too costly to ignore.** In upstream energy, poor data practices lead to duplicated surveys, missed deadlines, and reputational risk. But with the right approach, legacy seismic and geophysical data can become an engine for insight and investment.

### Your data maturity strategy

1. **Audit legacy data** - identify what exists, where it lives, and what formats it's in.
2. **Don't assume that digital = clean and structured** - legacy media that has been re-formatted is machine-readable. When much of this data was re-formatted, AI and Cloud were unlikely to have been a consideration.
3. **Prioritise value** - focus on high-cost, high-impact datasets (e.g. seismic surveys, compliance-critical files).
4. **Structure for reuse** - apply metadata standards and cataloguing to support licensing, asset evaluation, and M&A.
5. **Deliver transparency** - use structured archives to attract partners, support ESG goals, and comply with regulations.
6. **Reduce future risk** - avoid project delays, regulator penalties, and unnecessary rework through ongoing data governance.

Clients tell us that seismic data can cost up to $1 million daily to gather. It can quickly become unusable if not maintained in a structured format.

# Heritage – safeguarding memory through structured data

Where the energy sector deals in resource value and compliance risk, the heritage sector works with something irreplaceable - memory, truth, and public trust. Museums, archives, libraries, and cultural institutions hold immense volumes of material, including photographs, manuscripts, audio recordings, artefact catalogues, and architectural plans, that collectively document the human experience. Yet, much of this knowledge remains locked away without structured digital practices.

Many institutions have undertaken digitisation efforts, often spurred by preservation needs or public access goals. But digitisation without structure merely moves the problem from a physical shelf to a digital drawer! Scans of manuscripts, for example, remain hard to use unless they are described with meaningful metadata and indexed. Digital assets are hard to search and contextualise without structured formats, and nearly impossible to integrate with other collections.

The risks are not abstract. As the Organisation for Economic Co-operation and Development (OECD)[10] and United Nations Educational, Scientific and Cultural Organisation (UNESCO)[11] have noted, the erosion of reliable historical sources in a digital era contributes directly to the rise of disinformation and historical distortion. Structured data allows archives to act as public guardians of truth. Without it, cultural institutions lose relevance in a world increasingly shaped by algorithms and online narratives.

A case in point is the Camden Place[12] project. Ovation Data supported the recovery and digitisation of a treasure trove of historic material tied to French Emperor Napoleon III. Among the artefacts was a deteriorating copy of the lease signed by the Empress Eugenie in 1873 when she assumed rental of Camden Place after Napoleon's death. Now, in a structured, digital archive, this fragile document is accessible to a much wider audience without the risk of further damage to the original. Camden Place House Heritage Committee member, Angela Hatton, picks up the story: "Digitising our collection uncovered extraordinary documents we didn't know we had, like Napoleon III's funeral plans, filled with handwritten edits and forgotten details. However, the real value emerges when clean, structured data is shared beyond a single archive. Interoperable formats allow researchers to connect scattered fragments of information across collections, surfacing new insights that no one institution could find alone."

Structured cultural data also enables entirely new forms of engagement. Virtual exhibitions, AI-powered discovery tools, and immersive educational experiences all depend on underlying data that is clean, standardised, and interoperable. Without that, digital storytelling remains superficial. With it, institutions can surface forgotten histories, democratise access to heritage, and anchor public knowledge in verifiable sources. For smaller historic properties and institutions, catalogues can become accessible without needing to expand access to physical sites, and scholars worldwide can source data about collections they might otherwise struggle to analyse.

In an era when cultural narratives are being contested and rewritten in real time, structured data does more than preserve; it affirms the authority of evidence. It ensures that memory is not lost in the noise. It also gives institutions the tools to remain relevant in a digitally mediated world.

---

10 OECD, *The Power of Data and Culture on Citizen Participation* (2024), https://www.oecd.org/en/blogs/2024/09/the-power-of-data-and-culture-on-citizen-participation.html

11 UNESCO, *Memory of the World*, https://www.unesco.org/en/memory-world

12 https://www.historichouses.org/house/camden-place/visit/

**"Structured data helps data stewards work more efficiently and effectively. Beyond that, structured data in heritage is foundational to everything an organisation or their data users might want to do with the data; foundational to finding it, publishing it at scale, consuming it easily, sharing it, linking it with other data for improved context and discovery, building services on top of it, and preserving it. Data structure powers all of this."**

Allison Rivera, Director of Digital Heritage Services,
Ovation Data

## Heritage – structured data as a digital preservation strategy

It's not enough to digitise. Scanning archives and storing them isn't preservation, it's deferral. Cultural assets remain invisible and disconnected without structure, losing relevance with every passing year.

### Your data maturity strategy

1. **Digitise with purpose** - scan to help preserve, but structure to share, discover, and understand.
2. **Add context** - create metadata linking images, documents, people, and places.
3. **Standardise formats** - ensure interoperability across collections and platforms.
4. **Build public value** - enable new research, digital storytelling, and online access.
5. **Protect for the future** - preserve data in systems designed for long-term access and integrity.

Structured data is the difference between a scanned manuscript and a searchable, shareable piece of history.

# Life sciences and biotech – accelerating discovery through structured knowledge

Data is not a by-product of operations in life sciences; it is the raw material of progress. Research institutions, pharmaceutical companies, and biotech startups rely on vast and varied datasets to develop new therapies, track disease patterns, and identify promising molecular targets. Yet, fragmentation and inconsistency remain persistent obstacles to scientific discovery across this data-rich sector.

One of the clearest examples of this challenge comes from the National Institutes of Health (NIH) in the United States, which has long grappled with the problem of data silos. NIH-funded research often produces valuable outputs, including clinical trial results, genomic data, and longitudinal patient studies. Still, these are stored in unlinked repositories, formatted differently, and described using inconsistent terminology. The result is that vital data remains inaccessible to those who could build upon it. In response, NIH has launched several initiatives to improve interoperability, including adopting the FAIR principles (Findable, Accessible, Interoperable, Reusable) to make data more usable across teams, technologies, and time.[13]

At a more operational level, the benefits of structured data in biotech are tangible. Researchers can identify biological patterns and correlations more quickly when they work with harmonised datasets. This reduces time-to-discovery, shortens the path to publication, and accelerates the development of diagnostics and treatments. Structured data also improves collaboration between institutions. It allows different labs, often spread across countries and disciplines, to share findings, compare methodologies, and confidently build on each other's work. All too often, however, data is poorly handled, and that means opportunities for advancements in science and treatments are missed.

**"Coronavirus and its various strains were not unknown before the COVID-19 pandemic; coronaviruses were studied decades ago.[14] But the data wasn't stored or structured so that we could build on it; we couldn't analyse it - for example - to predict if and when a coronavirus might jump to humans. SARS-CoV-1 is a strain that caused the 2002–2004 SARS outbreak,[15] and is closely related to SARS-CoV-2 , yet the results from the last decades along with these were not organised enough to predict the potential disaster of a variant infecting humans that caused COVID-19. We see the same in cancer research; valuable findings are lost in paper archives, forcing scientists to repeat work instead of advancing it. Without structured data, time, money, and lives are wasted."**

Dr. Anindito Sen, Ph.D., Research Scientist,
Microscopy and Imaging Center, Texas A&M University

13 Kush.R.D., Warzel.D., Kush.M.A, et al, *FAIR data sharing: The roles of common data elements and harmonization* (2020) https://pubmed.ncbi.nlm.nih.gov/32407878/#:~:text=readable%20semantics,their%20efforts%20to%20identify%20the

14 Covid-19: First coronavirus was described in *The BMJ* in 1965, BMJ (2020) https://www.bmj.com/content/369/bmj.m1547#:~:text=One%20such%20sample%2C%20referred%20to,inoculation%20of%20the%20original%20specimen.%E2%80%9D

15 Wikipedia, https://en.wikipedia.org/wiki/SARS-CoV-2

Ovation Data works with clients in the research and healthcare space to bring order to this kind of complexity. In one case, a research network working on neurological conditions faced a backlog of disconnected studies, spreadsheet archives, and imaging files that could not be searched or linked. Ovation's team helped to recover and catalogue the materials, tag them with consistent metadata, and structure them into a platform that could support new research queries. Within months, the group reported a measurable drop in time spent preparing data for analysis and a marked improvement in interdepartmental collaboration.

Structured data in life sciences also supports public health objectives, enabling faster risk detection, more accurate modelling, and better decision-making by scientists and public health officials. As advanced technologies such as machine learning and natural language processing become more embedded in biomedical workflows, the importance of clean, standardised inputs will only grow.

## Life sciences and biotech – structuring data for faster breakthroughs

**In research, time is everything.** But scientists still spend up to 80% of their time cleaning and wrangling data. Structured datasets don't just save time - they accelerate discovery, collaboration, and public health impact.

### Your data maturity strategy

1. **Identify priority datasets** - clinical trials, lab outputs, genomics archives, etc.
2. **Apply FAIR principles** - make data Findable, Accessible, Interoperable, Reusable.
3. **Eliminate silos** - consolidate and harmonise across projects and departments.
4. **Enable collaboration** - support research networks with searchable, structured repositories.
5. **Prepare for AI** - ensure data is usable for pattern detection, diagnostics, and modelling.

Structured data helped Moderna scale up its vaccine R&D, and it can do the same for your research programme (see case study on page 21 for more insights about Moderna).

# Building clean, structured data foundations – what it takes

Data does not become clean and structured by accident. It requires deliberate action, expertise, and investment. Creating clean, structured data foundations can seem daunting for organisations with decades of accumulated records, scattered systems, or legacy formats. But the process is well-defined and increasingly essential.

The first step is an audit. Before anything can be structured, it must be discovered. This means identifying what data exists, in what formats, across which systems, and under whose ownership. In many cases, organisations are surprised by the volume and diversity of what they hold and how little is known beyond a handful of staff.

> **"In our experience, legacy systems and data silos make the task of standardising and integrating data difficult. While we see the problems and challenges being recognised by those using and working with data, there are, overwhelmingly, real issues grounded in a lack of clear strategy for data governance and preparation and a lack of awareness about the importance of clean data."**
>
> James Laming, Vice President of Global Operations,
> Ovation Data

The next stage is recovery. This may involve physical media, such as tapes, reels, paper files, or digital systems requiring extraction from outdated software environments. Here, the emphasis is on retrieval without loss or degradation. Ovation Data, for example, has worked on projects where data was successfully retrieved from formats no longer supported by modern hardware, using specialist equipment and technical procedures designed to preserve content and context.

Digitisation follows, but this is only the beginning. Without structure, digital files are inert. Proper data interoperability requires tagging, classification, and standardisation. Metadata is created or enhanced, file formats are converted where necessary, and naming conventions are unified. These steps are critical to ensuring that data is not only readable but also useful, both to humans and machines.

Once data is structured, it must be housed in systems allowing access, analysis, and security. These platforms must accommodate the dual need for protection and availability.

Equally important is the human component. Technology alone cannot guarantee success. Structured data projects require cross-functional collaboration between IT, subject-matter experts, records managers, and compliance leads. Governance frameworks must define roles, responsibilities, and quality control processes.

> **“Data issues are often a proxy for organisational issues. Solving them requires open, cross-disciplinary collaboration. When departments hold honest discussions about workflows and challenges, they improve both data quality and how they work together. That sort of trust is essential, and it often begins by acknowledging what’s missing, not just what’s broken.”**
>
> Jasmine Tran, Head of Strategic Partnerships and Global Key Accounts,
> Ovation Data

Perhaps most crucially, structuring data should not be a one-off effort. It is a continuous process. As new data is created or acquired, it must be integrated into existing frameworks. Policies must be maintained, systems monitored, and standards periodically reviewed to reflect evolving needs. The most effective organisations treat their data as a living asset - something to be curated, improved, and built upon.

A clean data foundation is not just a technical achievement. It is a shift in mindset. It signals a readiness not just for AI, but for resilience, transparency, and future growth.

# Case study from the market – the value of clean, structured data to Moderna

Moderna's rapid development of a COVID-19 vaccine demonstrated the utility of clean, structured data and a digital-first approach in biotechnology. Long before the pandemic, Moderna had invested in an integrated data infrastructure that connected its research, development, and manufacturing systems. This enabled real-time data access and decision-making across the organisation.

From its founding, Moderna aimed to build a platform company, not a single-drug firm. This meant developing a reusable data and automation backbone that could support multiple mRNA[16] programs. The company's operations ran on cloud-native infrastructure (primarily AWS), allowing seamless data sharing and the use of AI and machine learning tools across the value chain, from mRNA design to production scaling.[17]

Automation played a key role. Early digitisation of lab work allowed for robotic experimentation and standardised data capture, dramatically increasing throughput. Pre-pandemic, Moderna could produce around 30 mRNA constructs per month; with automation, that rose to 1,000.[18] This enabled scientists to test options in parallel, accelerating development timelines.

Moderna also applied AI to optimise mRNA sequence design and quality control. These models relied on structured, high-quality data and were integrated into user workflows. As Moderna's Chief Data and AI Officer put it, COVID-19 didn't require new algorithms it simply ran through their existing digital pipeline.[19]

When the SARS-CoV-2 genome was published in January 2020, Moderna initiated vaccine development within days. The company moved a candidate to clinical trials in just 63 days, enabled by its unified data environment and automation.[20]

16 Messenger ribonucleic acid (mRNA) is a single-stranded molecule of RNA that corresponds to the genetic sequence of a gene, and is read by a ribosome in the process of synthesizing a protein. Wikipedia, https://en.wikipedia.org/wiki/Messenger_RNA

17 AWS, 2021, https://aws.amazon.com/solutions/case-studies/innovators/moderna/

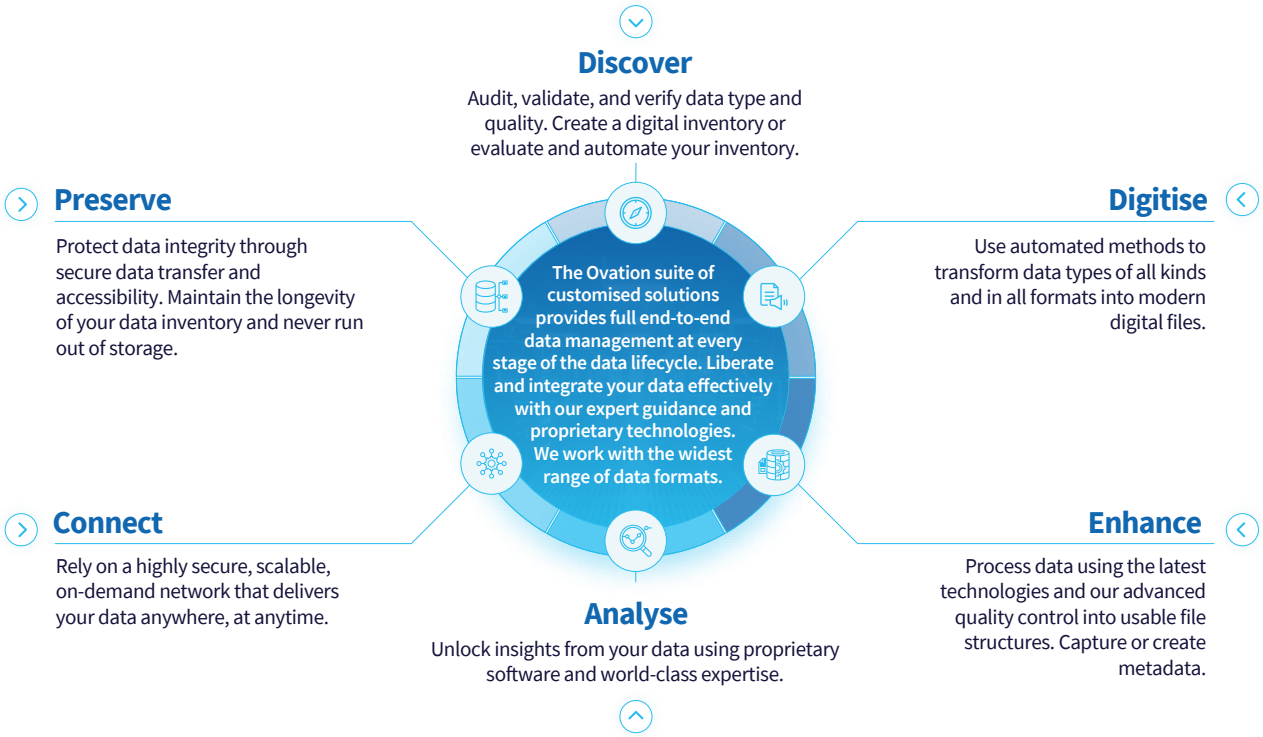
18 MIT Sloan, 2021, https://sloanreview.mit.edu/audio/ai-and-the-covid-19-vaccine-modernas-dave-johnson/

19 Ibid.

20 Iansiti, M. Et al., Moderna (A), HBS Case Collection, Harvard Business School (2020), https://www.hbs.edu/faculty/Pages/item.aspx?num=58942

# Ovation's approach – more than just transcription

Ovation Data distinguishes itself through deep sector understanding and disciplined methodology - The Ovation Method - in an industry often focused on tools and transactions. Ovation's work goes far beyond data digitisation. It encompasses the entire data discovery, retrieval, preparation, structuring, and integration lifecycle, bringing technical rigour and strategic foresight to each engagement.




What sets Ovation apart is its dual commitment to precision and purpose. The company does not treat legacy data as a bulk processing problem. Every dataset is approached as a potential source of insight. That means understanding the domain in which it originated, whether it's seismic data from a frontier basin, degraded film from a cultural archive, or lab results from a long-forgotten clinical study. Structuring begins with context.

Our approach and technology support high-value, historically significant, or regulation-sensitive data complexities. We enable metadata schemas, role-based access, and traceability features that align with compliance, operational, and research needs. But the tech is only as good as the expertise behind it. Ovation's teams combine data engineers, archivists, software developers, and sector specialists who understand both the value and vulnerability of the data they handle.

Clients frequently comment on Ovation's ability to bridge the gap between raw data and usable intelligence. This is not about mass conversion; it is about strategic enablement. By working with clients to understand the real use cases for their data, from licensing rounds to public exhibitions to AI model training, Ovation ensures that the final outputs are not just complete but usable.

What begins as a compliance project evolves into a wider capability improvement in many engagements. Once people see the utility of structured, accessible data, they begin to view information governance not as a constraint, but as a source of competitive and institutional strength.

# Conclusion

The evidence is clear. The message across every sector examined in this paper, energy, heritage, and life sciences, is consistent: clean, structured data is not a luxury but a necessity. The cost of inaction is rising, whether measured in financial cost, lost insights, reputational damage, or missed opportunities for innovation or, frankly, the total loss of data. In contrast, organisations that take data readiness seriously position themselves to lead.

Acting now is not just prudent, it is strategic. Structured data unlocks long-term operational efficiency, supports more reliable AI outcomes, enables smoother compliance, and ensures that knowledge survives technological change. The advantages are practical, measurable, and increasingly expected by partners, regulators, and investors alike.

Ovation Data is ready to support organisations at any stage of this journey. With a proven track record, specialist technical capability, and deep cross-sector understanding, Ovation delivers high-quality structured data and the systems, processes, and mindset shifts needed to sustain it. Whether preparing for an audit, considering a licensing round, embarking on AI deployment, or seeking to preserve cultural or scientific knowledge, Ovation can help ensure your data is ready to deliver.

**"As society increasingly relies on AI and machine learning to drive innovation, the importance of clean, structured, accessible data has never been greater. Yet, without prioritising data preservation, we risk a loss as profound as the destruction of the Library of Alexandria – valuable knowledge will be lost forever, setting back progress for generations. It's unacceptable that 80% of historical data still sits on shelves, unused and forgotten, while powerful technologies struggle to deliver their potential. We must act now to ensure society can fully benefit from this immense resource."**

Gregory Servos, Executive Chairman,
Ovation Data Services Inc.
June 2025


**To learn more or arrange a meeting with a member of the Ovation team,**
**visit: www.ovationdata.com/contact-us/**
**or solutions@ovationdata.com**