

IN PARTNERSHIP WITH UNCOUNTABLE,
C&EN PRESENTS

c&en | WHITE PAPERS

Creating The Proper Data Infrastructure for AI-Driven Formulation & Measurement in R&D

BROUGHT TO YOU BY





Creating The Proper Data Infrastructure for AI-Driven Formulation & Measurement in R&D

UNCOUNTABLE WHITE PAPER | 2023

TABLE OF CONTENTS

<u>3 - 4</u>	I. INTRODUCTION	
	Structured Data vs. Unstructured Data	4
<u>5 - 6</u>	II. IMPORTANCE OF STRUCTURING LAB DATA FOR AI	
<u>7</u>	III. EXAMPLES OF LAB DATA: UNSTRUCTURED VS. STRUCTURED	
	Brookfield Viscosity: The Standard Method of Data Capture	7
	Brookfield Viscosity: The Best Practice for Structuring Lab D Metfor AI.....	7
<u>8 - 10</u>	IV. R&D DATA MANAGEMENT SYSTEMS	
	Laboratory Information Management Systems LIMS.....	8-9
	Electronic Laboratory Notebooks (ELNs).....	9
	Other Data Systems / Sources in The R&D Ecosystem.....	10
<u>11 - 12</u>	V. DEPLOYING AI PREMATURELY	
	Erroneous Insights.....	11
	Diminished Trust in AI Capabilities.....	11
	How to Avoid Early Deployment.....	11-12
	Ensure Robust Data Infrastructure.....	12
<u>13 - 14</u>	VI. PREPARING FOR AI: CONSIDERATIONS FOR R&D ORGANIZATIONS	
	What to Evaluate in Your Current Data Infrastructure.....	13
	How to Define Objectives & Expected Outcomes.....	14
<u>15 - 17</u>	VII. CREATING A ROADMAP FOR AI-DRIVEN DEVELOPMENT & FORMULATION	
	Preparatory Phase.....	15
	Deployment Phase.....	16
	Maintenance Phase.....	16-17
<u>17 - 22</u>	VIII. EVALUATING DIFFERENT PROVIDERS	
	Establish Your Criteria for Evaluation.....	18
	Ask The Right Questions.....	19-20
	Review Customer Case Studies & Testimonials.....	20
	Get Personalized Demos & Pilot Tests.....	20-22
<u>22 - 24</u>	IX. UNCOUNTABLE: THE ALL-IN-ONE PLATFORM	
	Uncountable's AI Solutions.....	22-24
<u>25</u>	X. CONCLUSION	

ABSTRACT

In the evolving landscape of innovation, utilizing Artificial Intelligence (AI) and Machine Learning (ML) in Research and Development (R&D) has transitioned from being a novelty to a necessity. Yet, the vital ingredient that fuels this transition rests on the bedrock of structured data. This white paper underscores the significance of structured lab data in fostering AI-driven innovations and maps out a strategic roadmap for organizations to fully capitalize on AI's prowess in optimizing R&D outcomes.

I. INTRODUCTION

Today, more than ever, the emphasis on Artificial Intelligence (AI) and Machine Learning (ML) is ever-growing, reshaping how we analyze, interpret, and leverage large volumes of data. A crucial precursor to harnessing the transformative power of AI in research and development is the meticulous structuring of laboratory data.

As industries strive to maintain a competitive edge in a rapidly evolving marketplace, leveraging AI in R&D activities has become a powerful catalyst to spur innovation and efficiency. However, integrating AI and ML within the R&D sector necessitates a foundational shift towards structured data management.

Before delving deeper into the subject, it is imperative to understand the fundamental distinction between structured and unstructured data.

- **Structured data** refers to information that adheres to a specific format or model, thus enabling easier analysis and processing. This type of data, often found in databases, follows a clear pattern, facilitating streamlined storage and retrieval.
- **Unstructured data** is more free-form, lacking a definite structure, which makes its analysis more complex and time-consuming. It encompasses a variety of formats, such as text documents, videos, and others.

The graphic below further breaks down the differences between structured and unstructured data, outlining both examples, advantages, and disadvantages.

Unstructured	Structured
Examples <ul style="list-style-type: none">• Spreadsheets• Word Documents• PDFs• Lab Journals / ELNs• SharePoint / Shared Drives Advantages <ul style="list-style-type: none">• Free• Unrestricted Entry of Information• Known / Second Nature "Habitual" Disadvantages <ul style="list-style-type: none">• Limited Scope & Scalability For Application Of Info• Ctrl+F Keyword Searching• Limited Collaboration• Inability to Innovate Efficiently & at Market-Pace	Examples <ul style="list-style-type: none">• Databases• LIMS• Inventory Systems• Uncountable Advantages <ul style="list-style-type: none">• Instant Access To Specific Information / Data• Shareable & Scalable Information• Intelligent Insights & Reporting• Preservation of Institutional Knowledge Disadvantages <ul style="list-style-type: none">• Requires Intentional / Deliberate Entry Of Information• Change Management• Migration of Historical Data Into New System• Disciplined Use

As we transition into a more data-driven era, the delineation between these two forms of data becomes critically significant, establishing the groundwork for integrating AI and ML effectively within the R&D sector.

This paper delves into the multi-faceted implications of structured lab data, popular data systems used in today's lab environment, how to determine and prepare your organization for compatible and successful AI-driven development, how to prepare and find the right AI solution for your organization, and best practices for R&D organizations aiming to pave the way for AI-driven formulation and measurement in R&D.

II. IMPORTANCE OF STRUCTURING LAB DATA FOR AI

To appreciate the magnitude of impact structured data can have, it is vital to dissect several aspects that underscore its critical role in AI implementation.

- **Standardization & Consistency:** Implementing a uniform format across all data mitigates errors and facilitates smoother data processing. AI models often depend on recognizing patterns and trends; thus, having consistent data is paramount to developing a robust AI model that can accurately interpret data and provide reliable insights. Structured data promotes uniformity and consistency, facilitating the development of predictive models that can assimilate various data forms seamlessly, thereby fostering enhanced decision-making processes.
- **Reliability:** Reliability surfaces as a core facet of structured data as structured data is more reliable as it undergoes a series of validations, significantly reducing the chances of inconsistencies and errors and fostering trust in the data being used. Structured data, characterized by meticulous organization and management, is a reliable source for insights, bolstering accurate analyses and developing precise predictive models.
- **Feature Engineering:** Structured data simplifies the feature engineering process enabling robust AI models to effectively parse through intricate data patterns where significant data characteristics are identified and utilized to create more complex, predictive models.
- **Scalability:** One cannot overlook the role of scalability in this context. As organizations grow, so does the volume of data. Structured data can be scaled more readily, allowing AI models to adapt and grow without losing effectiveness or accuracy. This scalability ensures that organizations can maintain growth without being bogged down by data management issues. A structured data framework allows for scalable solutions, permitting comprehensive analyses as the data volume increases, thereby fostering progressive improvements in R&D processes.

- **Data Integrity:** Data integrity ensures that data remains accurate and unaltered over its lifecycle. Maintaining data integrity is central to developing AI models that can be relied upon to provide accurate insights and facilitate informed decision-making processes. Structured data mitigates errors and inconsistencies that can derail the AI model's performance.
 - **Interoperability:** Interoperability comes into play, permitting different systems and organizations to use the data effectively and fostering collaboration and information sharing. Structured data engenders interoperability, enabling a synergistic interaction between various data systems and tools and promoting a collaborative and cohesive work environment.
 - **Reduction in Noise:** Structured data inherently reduces noise – a common culprit in misleading results and inefficient model performance, thus paving the way for more accurate and reliable outcomes.
 - **Easier Analysis & Debugging:** Structured data allows for easier analysis and debugging, helping teams identify and rectify issues swiftly. Structured data fosters a model of data governance that ensures high data quality and management – critical aspects to safeguarding data and ensuring compliance with regulatory requirements.
- Data Governance:** Implementing comprehensive data governance strategies becomes feasible with structured data, ensuring compliance with regulations and fostering
- efficient data management policies.

Illustratively, consider lab data that notes viscosity measurements. In an unstructured environment, this data might be scattered across different mediums, recorded in inconsistent units, and lack contextual information. When structured, this data would be housed centrally, with uniform units, accompanied by relevant contextual details, thus making the task of analyzing this data through AI tools streamlined and more insightful.

III. EXAMPLES: UNSTRUCTURED VS. STRUCTURED DATA

To illustrate the significance of structured data, let's consider the example of Brookfield Viscosity, a common parameter measured in labs.

- **UNSTRUCTURED DATA**

- *The Standard Method Data Gets Recorded in Spreadsheets & Notebooks:*

- The viscosity data might initially exist in disparate systems, with varying units and inconsistent recording methods. This might lead to a chaotic data environment where drawing meaningful insights becomes arduous.

For example:

- Viscosity, 7D = 3000
 - Brookfield Visc. Sp #4 = 5500
 - BV, ON = 1800

- **STRUCTURED DATA**

- *The Best Practice for Structuring Lab Data for AI with a Structured Data System:*

- When structured, the viscosity data would be centralized, with standardized units and consistent data recording protocols. This structured framework enables easier analysis, fostering quicker and more accurate insights, which is critical for informed decision-making in R&D processes.

For Example:

- Brookfield Viscosity = 5000
 - Liquid Aging Time + Temperature: 7D at 23°C
 - Spindle #4
 - RPM: 150
 - Test Temperature: 23°
 - Exact temperature and time
 - Machine SN, Operator

IV. R&D DATA MANAGEMENT SYSTEMS

Laboratory Information Management Systems (LIMS) and Electronic Lab Notebooks (ELNs) are quintessential platforms that can foster structured data management, acting as a reservoir of data that can be optimized through AI integration. These platforms centralize data, making it easier to effectively implement AI tools for enhanced insights and outcomes.

In the rapidly evolving landscape of research and development, the role of sophisticated data management systems cannot be overstated. Laboratories and research facilities globally recognize the need to centralize and structure their burgeoning data repositories. LIMS and ELNs have emerged as frontrunners in this revolution. These platforms transform data management practices and pave the way for AI integration in the R&D sector. Let us delve deeper into the capabilities and benefits these platforms bring to the table:

- **LABORATORY INFORMATION MANAGEMENT SYSTEMS (LIMS)**

- LIMS platforms are the backbone for structured data management within modern laboratories. These systems excel in managing complex processes, tracking a plethora of data, and ensuring compliance with various regulatory requirements. Here, we explore the multi-faceted benefits of LIMS:
 - **Centralization & Standardization:** LIMS enables data centralization, providing a unified platform where data from various sources can be aggregated, standardized, and accessed seamlessly. This facilitates a consistent approach to data handling, which is vital for AI-driven analyses.
 - **Automation & Efficiency:** With capabilities to automate routine processes, LIMS significantly reduces manual errors and boosts operational efficiency. This automation is a precursor to AI integration, setting the stage for sophisticated data analyses and predictions.
 - **Compliance & Security:** LIMS are designed to adhere to stringent regulatory compliances, ensuring data security and integrity. These platforms offer robust features safeguarding sensitive data – critical to fostering scientists' trust in AI-driven insights.

- **Integration & AI Tools:** LIMS can be seamlessly integrated with AI tools, creating a powerful synergy to drive innovation and progress in R&D projects. LIMS can offer predictive analytics through AI, enabling data-driven decision-making and fostering innovation.

[DOWNLOAD OUR LIMS GUIDE](#)

- **ELECTRONIC LABORATORY NOTEBOOKS (ELNs):**

- Parallel to LIMS, ELNs have emerged as powerful platforms transforming how data is recorded and managed in laboratories. Let's explore the diverse capabilities of ELNs:
 - **Digitalization & Collaboration:** ELNs facilitate the transition from paper-based records to digital formats, fostering collaboration and data sharing across teams and organizations. This digital transition is vital in preparing data for AI analyses, offering structured data that can be mined for insights.
 - **Data Retrieval & Analysis:** ELNs offer advanced data retrieval and analysis features, providing tools to sift through large datasets to extract relevant information quickly. This capability is particularly beneficial in AI integrations, where quick data retrieval can facilitate real-time analyses and insights.
 - **Interoperability & Flexibility:** ELNs are known for their interoperability, offering flexibility in integrating with other platforms and systems. This facilitates a cohesive data management strategy, creating a structured data repository ready for AI-driven analyses and predictions.
 - **Facilitating AI-Driven Innovations:** ELNs can foster AI-driven innovations in laboratories through seamless integration with AI tools. AI can leverage the structured data housed within ELNs to generate predictive models, optimize processes, and guide R&D projects to success.

[DOWNLOAD OUR ELN GUIDE](#)

While LIMS and ELNs have become today's most recognized digital lab solutions, it's important to recognize the various, siloed, and disparate structured and unstructured data management systems that still exist and, in most cases, are still being utilized in the R&D landscape today. More often than not, as organizations introduce additional and separate data systems, utilizing them simultaneously throughout R&D workflows, data becomes increasingly fragmented – increasing the margin for error and stifling the ability to apply AI effectively.

Other Data Systems / Sources in The R&D Ecosystem

Spreadsheets



ELNs / Lab Journals



LIMS



Inventory Management



Visualizations / Analysis



Statistical Tools



Predictive Tools



ERPs



Today, LIMS and ELNs stand out as two of the most common systems that have transformed the data landscape within laboratories, which if implemented properly, can foster structured data management conducive to AI integration. By adopting these platforms, organizations can pave the way for AI-driven advancements, unlocking unprecedented avenues for innovation and growth in R&D projects. Through centralized, structured, and secure data repositories, LIMS and ELNs set the stage for a future where AI and data become the cornerstone of progress and success in the R&D sector.

V. DEPLOYING AI PREMATURELY

Implementing AI prematurely, while the data is not adequately structured, can lead to many pitfalls, including erroneous insights and diminished trust in AI capabilities. As such, it's imperative that R&D organizations understand and further explore strategies to avoid early deployment and ensure a robust data infrastructure first.

While the advent of Artificial Intelligence in the R&D sector is a remarkable milestone with a promising future filled with innovation and efficiency, the success of AI integration hinges critically on the readiness of the existing data infrastructure. Premature deployment of AI technologies, especially before establishing a robust and structured data infrastructure, can backfire, leading to inaccurate predictions, wasted resources, and even a loss of trust in the capabilities of AI. This section delineates the adverse consequences of early AI deployment and outlines strategies to prevent these pitfalls and foster a conducive environment for successful AI integration.

1. ERRONEOUS INSIGHTS: Deploying AI without well-structured data can often result in misleading insights. AI models are only as good as the data fed into them; poor data quality invariably translates to unreliable outcomes. Moreover, unstructured data can create noise, confusing AI algorithms and impairing model performance. It is pivotal to ensure data accuracy, consistency, and reliability before letting AI algorithms interpret and learn from the data.

2. DIMINISHED TRUST IN AI CAPABILITIES: Premature deployment affects the immediate outcomes and can have long-lasting impacts on the perception of AI. When AI fails to deliver on its promises due to inadequate data infrastructure, it can foster skepticism and resistance among team members. Once lost, trust is difficult to regain, which can hinder future efforts to integrate AI into R&D workflows. Therefore, fostering trust through a methodical and phased approach to AI deployment is critical.

3. HOW TO AVOID EARLY DEPLOYMENT:

- **Data Audit:** Before considering AI integration, organizations should conduct a comprehensive data audit to assess the current state of their data infrastructure. This involves evaluating existing data's quality, consistency, and format to identify areas that require improvement or restructuring.

- **Data Governance & Quality Assurance:** Establishing a data governance framework can be instrumental in ensuring data quality and integrity. Implementing guidelines and standards for data collection, storage, and management can prevent inconsistencies and errors, paving the way for a smooth AI integration.
- **Pilot Testing:** Before a full-scale deployment, organizations should consider pilot-testing AI tools on a smaller scale. This allows teams to identify potential issues and fine-tune the system before broader implementation.
- **Setting Realistic Objectives & Expectations:** Setting realistic and achievable goals when deploying an AI tool is essential. This requires critically evaluating existing data and clearly understanding what the AI tool can do. Having a clear goal in mind and how you plan to measure success is helpful. Keep in mind the potential biases and limitations of the AI tool as well, and consider how to deal with them.
- **Training & Skill Development Skill Development:** Training the team to handle AI tools effectively is crucial. Equipping them with the necessary skills and knowledge can prevent mishaps and ensure that the AI integration proceeds smoothly.

4. ENSURE ROBUST DATA INFRASTRUCTURE: Before embracing AI, investing in building a robust data infrastructure that can support AI's complex requirements is essential. This includes creating centralized data repositories, fostering interoperability, and ensuring data security and privacy. Developing protocols for data labeling and model validation processes can further enhance the reliability of AI analyses. Moreover, collaborating with experienced providers who understand the nuances of AI integration can be a wise strategy.

While the urge to quickly adopt AI is understandable, given its potential benefits, a cautious and well-planned approach can prevent pitfalls and ensure a successful integration. Organizations should focus on laying a solid data foundation first, thereby facilitating a smoother, more effective transition into the AI-driven future of R&D.

VI. PREPARING FOR AI: CONSIDERATIONS FOR R&D ORGANIZATIONS

Before embarking on the AI integration journey, organizations should carefully evaluate their existing data infrastructure and readiness levels and clearly define their objectives and expected outcomes from AI deployment.

Embarking on a journey to integrate Artificial Intelligence within the research and development landscape necessitates a well-thought-out strategy grounded in a thorough understanding of the organization's current standing and the ultimate objectives of AI deployment. Implementing AI is not just about adopting cutting-edge technology; it's about aligning technology with the specific needs and goals of the organization. Here, we elaborate on critical considerations that R&D organizations should ponder upon before setting out on the AI integration journey:

• WHAT TO EVALUATE IN YOUR CURRENT DATA INFRASTRUCTURE

- **Infrastructure Robustness:** Assess the robustness of current data infrastructure to ensure it can handle the large volumes of data that AI algorithms require for effective learning and predictions. This includes evaluating storage capacities, data retrieval speeds, and data security protocols.
- **Data Quality & Consistency:** The reliability of AI outcomes is intrinsically linked to the quality and consistency of the data. Organizations must evaluate the current state of data quality, focusing on completeness, accuracy, and timeliness. Take necessary steps to improve data quality where needed.
- **Data Interoperability:** Assess the level of interoperability within existing systems. A seamless data flow across various platforms and systems is vital to creating a cohesive data ecosystem conducive to AI integration.
- **Skills & Expertise:** It is also crucial to evaluate the organization's readiness regarding skills and expertise. This involves assessing whether the current team possesses the necessary skills to handle and manage AI tools effectively or if there's a need for training and development programs.

• HOW TO DEFINE OBJECTIVES & EXPECTED OUTCOMES

- **Set Clear Objectives:** Organizations must have well-defined objectives before deploying AI. This could involve optimizing existing processes, enhancing product quality, fostering innovation, or driving cost efficiencies. The goals should be Specific, Measurable, Achievable, Relevant, and Time-bound (SMART).
- **Use Outcome Measurement Metrics:** Develop metrics to measure the outcomes of AI integration. This helps track progress and ensures the deployment is aligned with the set objectives.
- **Risk Assessment & Mitigation:** A comprehensive risk assessment should be carried out to identify potential challenges and develop strategies to mitigate these risks. This includes evaluating the potential impact on existing workflows and developing contingency plans to address unforeseen challenges.
- **Get Stakeholder Engagement:** Engage with various stakeholders, including scientists, researchers, and technicians, to gather insights and inputs. Their engagement can provide a deeper understanding of the present conditions and the potential areas where AI can bring significant improvements.
- **Evaluate Vendors Thoroughly:** A detailed evaluation process should be undertaken to choose the right partners if the plan involves collaborating with external vendors. This includes assessing the vendors' expertise, experience, and the solutions they offer to ensure alignment with the organization's objectives.
- **Outline Budget & Resource Allocation:** Clearly outline the AI project's budgetary requirements and resource allocations. A well-planned budget can prevent cost overruns and ensure the smooth progression of the initiative.

By meticulously evaluating the existing data infrastructure and clearly defining the objectives and expected outcomes, R&D organizations can create a well-charted pathway to AI integration. The process, although complex, can lead to transformative results, ushering organizations into a new era of innovation and efficiency. Thus, a thoughtful preparation phase forms the cornerstone of a successful AI integration journey, paving the way for a future where data-driven insights fuel groundbreaking developments in the R&D sector.

VII. CREATING A ROADMAP FOR AI-DRIVEN DEVELOPMENT & FORMULATION

As organizations navigate the complex journey of incorporating AI into their R&D processes, having a structured and well-articulated roadmap can guide them to successful integration. A strategic roadmap helps identify critical milestones and outline actionable steps to be undertaken at various phases of the integration process.

While planning and following a strategic roadmap is pivotal to integrating AI in R&D successfully, it's important to ensure that the type of data and how it's collected, captured, and structured is regularly and precisely considered at each stage in the roadmap. This section outlines the crucial phases of implementing AI within your R&D efforts – the preparatory, deployment, and maintenance phases – and offers insights into the best practices at each stage to foster smooth and effective AI implementation.

1. PREPARATORY PHASE

- Identification of Goals & Objectives
 - **Gap Analysis:** Perform a gap analysis to understand the current limitations and identify areas where AI can add substantial value.
 - **Stakeholder Engagement:** Engage with all stakeholders to gather diverse perspectives and identify overarching goals and objectives.
- Data Analysis & Structuring
 - **Data Audit:** Conduct comprehensive data audits to ascertain the quality and structure of existing data.
 - **Data Cleaning and Structuring:** Invest time and resources in cleaning and structuring data, focusing on data integrity and interoperability to create a robust base for AI algorithms.
- Resource Allocation
 - **Budgeting:** Define budget allocations clearly, keeping in mind both technological and human resources needed for the project.
 - **Skills Development:** Develop training programs to upskill the existing workforce in line with the anticipated demands of AI integration.

2. DEPLOYMENT PHASE

◦ Selection of Tools & Technologies

- **Vendor Evaluation:** Conduct meticulous vendor evaluations to select the right tools and technologies that align with organizational objectives.
- **Pilot Testing:** Before full-fledged deployment, undertake pilot testing to identify potential bottlenecks and make necessary adjustments.

◦ Implementation Strategy

- **Change Management:** Develop a change management strategy to facilitate a smooth transition and minimize disruptions.
- **Collaborative Approach:** Foster a collaborative approach, encouraging cross-functional teams to work together to implement AI effectively.

◦ Monitoring & Adjustments

- **Real-time Monitoring:** Implement real-time monitoring mechanisms to track the progress of the deployment phase.
- **Feedback & Adjustments:** Establish feedback loops to make necessary adjustments based on real-time insights and observations.

3. MAINTENANCE PHASE

◦ Continuous Learning & Improvement

- **Feedback Analysis:** Analyze feedback continuously to make iterative improvements and optimize performance.
- **Knowledge Sharing:** Foster a culture of knowledge sharing where teams can learn from each other's experiences and insights.

◦ Scalability & Evolution

- **Scalability Planning:** Plan for scalability, ensuring your data systems can evolve with the organization's growing demands and changing dynamics.
- **Evolutionary Development:** Encourage an evolutionary development approach, where AI systems are continuously updated and improved based on changing needs and technological advancements.

- **Compliance & Governance**

- **Regulatory Compliance:** Ensure adherence to regulatory compliance and industry standards to avoid legal complications.
- **Data Governance:** Establish robust data governance mechanisms to safeguard data integrity and privacy.

Creating a detailed roadmap with clear milestones and actionable steps at each phase can guide R&D organizations to successful AI integration.

VIII. EVALUATING DIFFERENT SOLUTION PROVIDERS

Choosing the right provider is a linchpin in successfully implementing AI within an R&D organization. Today's market is flooded with many providers offering varying expertise and capabilities. It becomes crucial for organizations to devise a strategy that helps in evaluating and selecting the most appropriate provider that aligns with their objectives and requirements.

Successfully implementing a robust solution that will allow you to leverage the benefits of AI within an R&D organization hinges on looking for and selecting a provider that understands the critical importance of structured data groundwork. The marketplace is teeming with providers touting "AI-first" solutions. However, AI is only as potent as the structured data it operates on. As such, organizations should recognize the AI's dependency on having a foundation for a structured data system. As such, when evaluating different providers to leverage the power of AI, it's important to prioritize vendors who prioritize data capture, management, and structure as the foundation, with AI capabilities as the icing on the cake.

This section breaks down the integral elements and questions organizations should consider during their evaluation. This guidance ensures the solution integrates seamlessly with existing systems and amplifies the organization's objectives.

1. ESTABLISH YOUR CRITERIA FOR EVALUATION

◦ Data Management & Structuring

- **Structured Data Groundwork:** Ascertain the provider's capability in creating, managing, and maintaining structured data systems that form the bedrock of any AI application.
- **Data Integration:** Understand the provider's expertise in merging data from various sources and systems to create a cohesive data pool.
- **Data Analytics Capabilities:** Gauge how the solution can assist in deriving insights before applying AI.
- **Organizational Data Capture:** Ask the provider how they ensure experimental data is being captured and standardized across individuals and organizationally in a way that provides AI/ML models with accurate training data.

◦ Technical Expertise

- **Track Record:** Investigate the provider's history in implementing AI tools, particularly in settings where structured data management is paramount.
- **Technological Innovations:** Determine the provider's proficiency in staying ahead in AI and integrating technological advancements.

◦ Solution Compatibility

- **Integration Capabilities:** Assess the provider's ability to integrate their solutions seamlessly within your infrastructure.
- **Customization Options:** Evaluate how much the provider offers customization to meet specific organizational needs and requirements.

◦ Security & Compliance

- **Data Security Protocols:** Review the provider's data security protocols to ensure the safety and integrity of your data.
- **Regulatory Compliance:** To mitigate legal and compliance risks, check the provider's adherence to industry regulations and standards.

2. ASK THE RIGHT QUESTIONS

◦ Solution Specific Questions

- What sets your solution apart, especially concerning data management and structuring?
- How do you compare to other “AI-first” solution providers?
- Can the solution scale as the organization grows, and to what extent?
- What kind of data management and analytics capabilities does the solution offer?
- Can the solution be customized to fit business and industry requirements?
- How do you handle data privacy and security concerns?

◦ Implementation & Support Questions

- What timelines are we looking at for the solution's implementation?
- What support and training avenues are available during and after?
- How does the provider handle data migration and integration with historical data and existing systems?
- What is your approach to updates and continuous maintenance?

◦ Financial & Cost Consideration Questions

- Can you break down the solution's pricing structure for us?
- What tools are included in the platform's total cost?
- What Return-on-Investment "ROI" can we realistically anticipate?
- What are the terms and conditions of the contract, including cancellation policies and service level agreements?
- Are there any initial setup costs or fees?
- Is the cost of implementation and customer support included?

◦ Future Readiness & Planning Questions

- How does the provider plan to keep the solution updated with the latest technological advancements?
- Could you share a roadmap of your planned developments?
- How do you ensure the solution remains competitive as market dynamics evolve?

Remember, the value AI promises is intrinsically linked to the data it's fed. So, in your quest for an AI solution, ensure that your chosen provider comprehends and emphasizes the symbiotic relationship between data structure and AI efficacy.

3. REVIEW CUSTOMER CASE STUDIES & TESTIMONIALS

Encourage providers to share case studies and testimonials from clients in similar industries, offering insights into the real-world impact of their solutions.

Request references to engage in discussions with existing clients to understand their experiences and satisfaction with the provider's solution and support.

Organizations can make informed decisions aligning with their strategic objectives by systematically evaluating providers based on well-defined criteria and asking pertinent questions. It ensures the successful integration of the AI tool and fosters a fruitful and long-lasting partnership with the provider, facilitating continual growth and innovation in the R&D domain.

4. GET PERSONALIZED DEMOS & PILOT TESTS

The decision to integrate a new solution within an organization is significant and multifaceted. While vendor presentations, case studies, and testimonials can provide valuable insights, there's no substitute for first-hand experience. Embarking on personalized platform demos and pilot tests with potential solution providers is crucial during the evaluation process. Here's why:

- **Understanding Real-World Application:** Personalized demos and pilot tests allow organizations to assess how the solution would function in their unique environment. This first-hand experience can clarify whether the solution truly meets specific needs or if it's more hype than substance.
- **Mitigating Risk:** Pilots and trials are a window into possible issues during full-scale integration. Identifying and addressing these during evaluation can save time, resources, and potential setbacks.
- **Gauging Ease of Use & Integration Capabilities:** No matter how sophisticated, a solution must be user-friendly and compatible with existing systems. Personalized demos can help stakeholders gauge the ease of integration, intuitiveness, and user experience, ensuring smoother transitions and better adoption rates.

- **Assessing Vendor Support:** Pilots often give a glimpse into the quality of customer support a vendor provides. How responsive are they during the trial phase? How effectively do they handle queries or challenges? This experience can indicate the long-term relationship and support you can expect.
- **Quantifying Value:** Beyond a solution's theoretical benefits, a pilot allows for tangible evaluation. You can measure specific outcomes, efficiencies, or benefits, translating promises into measurable results.
- **Gathering Stakeholder Feedback:** Various stakeholder involvement in demos and pilots ensures diverse feedback. This collective input can be instrumental in making informed decisions that account for different departmental and functional needs.
- **Building Confidence in Investment:** Investing in a new platform is a substantial commitment. Pilot tests and demos can provide the assurance needed for full-scale implementation, confirming the solution's worthiness as a long-term partner.

4.1. Best Practices for Effective Platform Pilots & Demos:

When running a pilot test, trial, and/or demo with a vendor, it's important that you consider the key steps that will ensure you gain the most value and knowledge from your experience. Following these best practices will ensure that you're taking a comprehensive approach to evaluating the features and functionality of the platform and how well the platform fits your organization's short and long-term needs.

- **Define Clear Objectives:** Before initiating a demo or pilot, outline what you intend to achieve or evaluate. Having clear objectives ensures that your assessment is focused and comprehensive.
- **Engage Relevant Stakeholders:** Ensure that the right people, from end-users to decision-makers, are involved in the evaluation process. Their collective feedback will provide a holistic view.
- **Document Experiences & Outcomes:** Record all observations, issues, benefits, and feedback during the demo or pilot phase. This documentation will be invaluable when comparing vendors or seeking organizational buy-in.

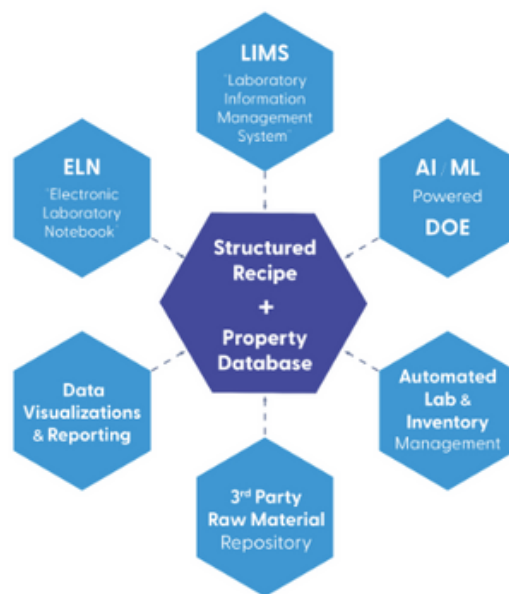
- **Seek Vendor Feedback:** Post-demo or pilot, ask for feedback sessions with the vendor. Their insights and willingness to adapt or improve can be telling.
- **Review & Reflect:** Gather all stakeholders for a debrief once the demo or pilot concludes. Discuss experiences, outcomes, and potential next steps. This collective reflection can guide the decision-making process.

While researching and conversations lay the essential groundwork, nothing can replicate the insights from hands-on experience. Prioritizing personalized demos and pilots in your evaluation strategy ensures that you're not just buying into a vision but investing in a solution that has proven its mettle in your real-world context.

IX. UNCOUNTABLE: THE ALL-IN-ONE STRUCTURED DATA PLATFORM

Uncountable offers a comprehensive data management system that connects inputs to outputs and structures and centralizes all experimental data.

With Uncountable's all-in-one structured data platform, scientists can seamlessly connect their formulation and measurement data in one centralized repository – gaining access to both a structured recipe and property database equipped to handle the complexities and requirements of AI.















Uncountable's AI Solutions


Uncountable's platform enables the utilization of AI tools for a range of functions, including output fits, linear coefficients, and suggesting experiments, paving the way for groundbreaking innovations in R&D. As shown in the examples below, Uncountable's platform offers advanced AI capabilities, such as:


1. OUTPUT FITS


Uncountable's platform can make predictions on data. Every output you're interested in will have a prediction and a confidence interval. Our easy-to-use dashboard uses real data allowing users to view accuracy of models created and the model's performance

Training Accuracy

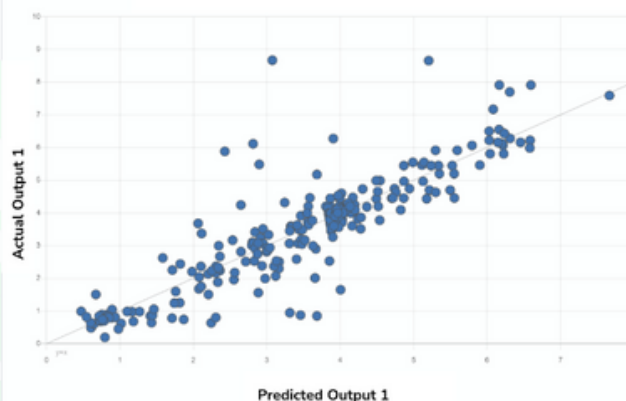
Numeric Outputs							
Plot	Name	Summary Statistics (Training Data)			Model Predictions		
		# Samples	Mean μ	Std. Dev. σ	RMSE	r ² Score	Explained Error % 
	Output 1	1360	172	140	54.4	0.86	62.6%
	Output 2	319	6.17	10.3	3.39	0.891	67.1%
	Output 3	254	16.4	7.36	5.03	0.533	31.8%
	Output 4	541	65.2	71.4	17.7	0.939	75.2%
	Output 5	359	8.02	5.4	2.96	0.696	45.1%
	Output 6	387	41.6	31.9	8.23	0.933	74.2%
	Output 7	1511	7520	12200	6070	0.752	59.2%
	Output 8	251	3.51	1.7	0.818	0.767	51.8%
	Output 9	895	7.25	4.03	2.12	0.725	47.6%
	Output 10	1475	6.26	6.71	2.9	0.813	56.8%
	Output 11	1492	0.00279	0.00112	0.000767	0.533	31.7%

 High Predictive Accuracy

 Low Predictive Accuracy

 Add Consistent Model

High Predictive Accuracy Low Predictive Accuracy Add Comparison Model



2. LINEAR COEFFICIENTS

The major benefit of traditional factorial Design-of-Experiments "DOE" models and linear regressions are the coefficients produced, providing explainability on how different inputs contribute to the output in question. Uncountable provides similar coefficients for its models, allowing users to learn which inputs are important for their final formulation.

Effect Sizes

Show linear coefficient approximations

Type	Name	TE (kcal)	Output 1	Output 2	Output 3	Output 4	Output 5	Output 6	Output 7	Output 8	Output 9	Output 10	Output 11	Output 12
Ingredient 1										-0.753		-1.51	0.000365	
Ingredient 2										0.231		-0.0274	0.000324	
Ingredient 3			10.8							0.0401		0.153	0.000214	
Ingredient 4			27.1		1.32	-34.7				-1.59	-0.86	-0.872	-0.000198	
Ingredient 5			-1.25		-0.0571		-0.223	0.233		0.229	-0.615	12.3	-0.781	-0.000187
Ingredient 6			4.96							3.21	-11.7	0.199	-0.000177	
Ingredient 7			3.48							0.5483		0.18	0.00017	
Ingredient 8					-0.187					-0.445	0.487	-3.6	0.213	0.000196
Ingredient 9					-1.13	4.96	15.5		-3.46	2.95	20.8	-5.88	-0.000196	
Ingredient 10												-1.35	-0.803	-0.000193
Ingredient 11			8.02		7.32	11.2			5.34	-1.12	5.63		-10.7	0.000675
Ingredient 12		-26.3	0.0219								-0.762	0.0724	0.000671	
Ingredient 13			5.72							-0.0291	-5.8	0.206	-0.000477	
Ingredient 14			3.5							-0.0917	-4.96	0.387	0.000437	
Ingredient 15			1.38		0.1706	5.793	-0.437	3.422	-0.0411	-0.893	4.04	-0.209	-0.00042	
Ingredient 16			1.92						0.0139	0.192	-7.79	0.0884	0.000417	
Ingredient 17		2.2+2.4+*	-1.83			-0.054					0.0967	3.92	-0.158	0.000462
Ingredient 18										39.5		0.171	0.000366	
Ingredient 19		crion	-25.1		0.00226	-4.12	-1.96		0.0562	-1.25	-1.54	3.47	7.37	0.000344
Ingredient 20		-25.3	0.224							0.5918	-0.561	0.0222	0.000326	
Ingredient 21			2.98							-0.276	-4.03	-0.5774	-0.000317	
Ingredient 22										0.189	0.183	6.13	0.373	-0.000294
Ingredient 23			2.38								-2.82	0.335	0.000285	
Ingredient 24			0.148							-3.67	-0.5553	-0.000295		
Ingredient 25			1.73							-0.0291	-1.21	0.263	0.000266	
Ingredient 26			-0.915		0.0203	-0.41	-0.114		9.12	-0.9618	-0.987	-0.473	-0.36	-0.000236
Ingredient 27			1.49							0.488	-3.2	0.526	0.000232	

3. SUGGESTED EXPERIMENTS

When users have a specific target in mind, Uncountable can suggest different experiments to make progress toward the goal. Balancing the objective of exploring new spaces with exploiting spaces already performing well, Uncountable's platform uses cutting-edge Bayesian Optimization techniques to suggest experiments.

Suggested Formulations

	Recipe Name Import Recipe?	Recipe 1	Recipe 2	Recipe 3	Recipe 4	Recipe 5	Recipe 6
Ingredient 1	= 2.5	2.5	2.5	2.5	2.5	2.5	2.5
Ingredient 2	= 0.202	0.2021	0.2021	0.2021	0.2021	0.2021	0.2021
Ingredient 3	= 0.0121	0.01209	0.01209	0.01209	0.01209	0.01209	0.01209
Ingredient 4	[6.52, 10.2]		9.709	10.11	8.072	8.411	9.5
Ingredient 5	[10.3, 25]	22.71	22.71	17.27	18	18.62	18.69
Ingredient 6	[4.78, 10]	9.658	8.146		6.247	6.307	
Ingredient 7	[5.84, 10.1]	6.138	6.347	6.328	6.235	5.906	5.877
Ingredient 8	[5.2, 50]	27.86	39.03	32.46	30.69	46.62	49.81
Ingredient 9	[5.1, 14.3]	10.26	11.35	9.108	10.64	11.42	13.41
Ingredient 10	[14.8, 32]	20.67		22	17.4		
Ingredient 11		100	100	100	100	100	100
Calculation 1	[2, 4]	2.25	3.83	2.09	2.39	2.98	2.78
Calculation 2	[0.8, 1.2]	0.804	1.11	1.02	1.08	1.13	1.08
Calculation 3		3.16	3.87	2.89	3.21	3.01	2.81
Calculation 4		1.91	9.82	0.595	3.04	6.36	5.41
Goal							
Predicted Output 1	≥ 300	221 ± 111	236 ± 111	224 ± 109	216 ± 107	210 ± 108	215 ± 109
Predicted Output 2	≥ 5	7.5 ± 2.75	5.82 ± 3.11	7.71 ± 2.88	7.39 ± 2.84	5.71 ± 3.08	5.94 ± 3.11
Predicted Output 3	≤ 4000	2100 ± 1390	1860 ± 1630	2140 ± 1580	2050 ± 1460	1810 ± 1760	1900 ± 1840
Predicted Output 4	≥ 2	2.5 ± 1.07	2.68 ± 1.15	1.94 ± 1.12	2.56 ± 1.07	2.46 ± 1.1	3.85 ± 1.18

X. CONCLUSION

In conclusion, this white paper highlights the critical role of structured data in fostering AI-driven innovations in R&D. By understanding and establishing the proper foundation of data upon which AI / ML solutions are built, R&D organizations will be well equipped and confident that the data they collect today, continues to add value to the data and experiments they run in the future.

By implementing a structured data system, organizations can leverage AI to unlock unprecedented avenues for growth and success in their R&D endeavors. Structured data is the key to ensuring better decision-making and a competitive advantage today and in the long term.

WANT TO LEARN MORE?

Contact Us Today!

General Inquiries: info@uncountable.com

Sales: sales@uncountable.com

www.uncountable.com

GET A DEMO OF UNCOUNTABLE