



# Foundational Scientific Data: Enabling the Modern Scientific Data Stack

Foundate Systems

<https://foundatesystems.com/>

Benjamin Flores

11-May-2026

## 1. Executive Summary

Laboratories are investing heavily in AI, analytics, visualization platforms, and digital transformation initiatives. Yet many organizations still rely on manual transcription, disconnected instruments, spreadsheets, and fragmented workflows to move foundational scientific data into data stores. This reliance on manual and fragmented workflows results in increased manual transcription errors, delayed data availability, reduced trust in the data, limited interoperability, and difficulty scaling for AI.

This challenge is pervasive across modern scientific organizations. From hands-on experience in the laboratory, Foundate Systems understands the real pains that organizations feel. Through a Scientific Data Capture Layer, organizations can close this gap and enable organizations to establish solid foundational data.

## 2. The Hidden Problem in Modern Laboratories

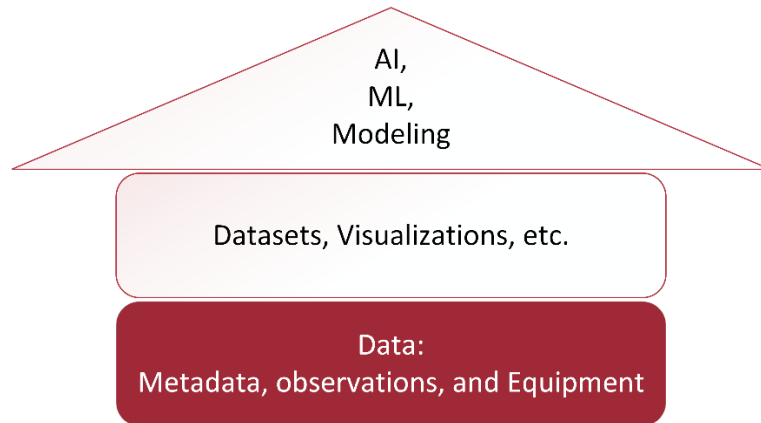
Modern laboratories often possess advanced analytical equipment; however, fragile and manual data capture processes still remain. Workflows may require scientists to manually transcribe data from a printout, copy and paste data between files and LIMS, or acquire data from a piece of an equipment through a USB drive. Some equipment vendors offer proprietary data stores, but these may often leave the data siloed. Some LIMS vendors offer connectors to retrieve data from specific pieces of equipment. However, not all of your equipment may have a connector, and what happens when the equipment firmware is updated? How quickly will a new connector be available? The hidden challenge within modern laboratory data capture is the deceptively complex variety of transmission mediums, communication protocols, and data formats.

## 3. The Cost of Manual Transcription

No matter how the data from a study is moved into a LIMS or an ELN, it must ultimately rest there. Unfortunately, the norm is that highly trained scientists spend non-value-added time manually entering data. This manual entry doesn't just require additional time, it also creates opportunities for transcription errors, delayed reporting, and propagation of bad data. Studies from PubMed demonstrate that manually transcribed data can create errors in the data at a rate of ~3% (Mays & Mathias, 2019) - ~9% (Feng et al., 2020) in scientific and medical data.

## 4. Data as a House

A data system can initially be thought of as a house. Organizations usually want to focus on the “roof” where there is AI, machine learning, predictive modeling, etc.



*Figure 1 - Data as a House*

However, like a house, the structure of the house is dependent on the foundation that it sits upon. With process and analytical data, metadata, observations, and contextual scientific information, a foundation is created that higher technologies sit upon. From the foundational data, organizations can begin to build datasets, visualizations, dashboards, etc. that create a digital twin of the scientific process running in the laboratory or manufacturing floor. Finally, the “roof”, AI and predictive models can ingest the datasets, data models, and other data that supports the application of these higher technologies for advanced insights.

## 5. Data as a Shared Resource

In some laboratories today, data is still treated as application-owned, or even scientist-owned, rather than organization-owned. As a result, scientific observations become trapped within individual instruments, spreadsheets, or departmental systems. However, mature scientific organizations evolve beyond this model by treating data as reusable infrastructure.

As the data system and laboratory mature, the house develops into something that more resembles a metaphoric office building. Data becomes a shared resource, like the electricity in the office building. The data flows automatically and users perform their functions without the pains of entering the data into the system.

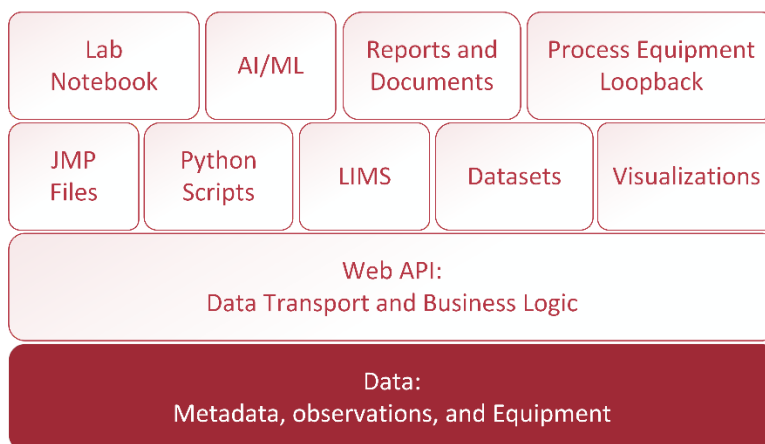


Figure 2 - Data as a Shared Resource

The data still serves as the foundation of the structure, the layers upon the data are able to evolve into more robust capabilities. Transport layers and business logic layers can support external and more robust use cases, such as data ingestion into LIMS, user developed and shared Python scripts, or protein concentration from an HPLC being fed into a bioreactor system. Data in this matured system can become FAIR (Findable, Accessible, Interoperable, and Reusable) and governed to ensure that those individuals that need access have access to the data, such as bioinformaticians, project managers, or even other collaborating scientists at other sites or running at different scales.

## 6. The Modern Scientific Data Stack

Current understandings of data acquisition are either vague or don't support a complete set of analytical laboratory equipment. With hands-on experience, one will find that a new data stack and understanding is required. For this reason, Foundate Systems has defined the Modern Scientific Data Stack.

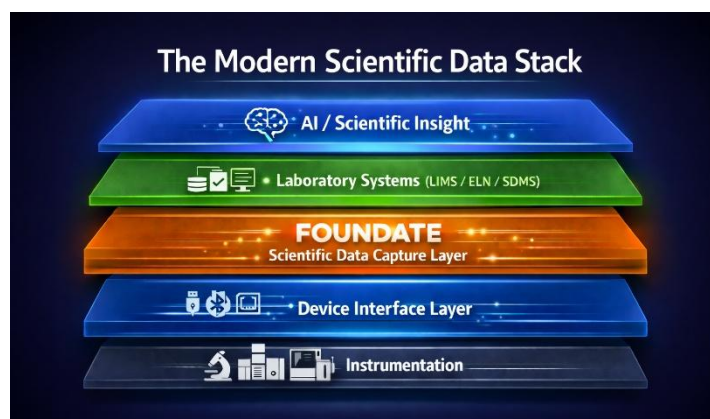


Figure 3 - Modern Scientific Data Stack

With the Modern Scientific Data Stack, the analytical instruments sit at the bottom of the stack, followed by the Device Interface Layer. By acknowledging the device interface layer, the medium in which data is transferred is defined and is standardized. With the standardization of incoming data, in the Scientific Data Capture Layer, the data can be normalized and easily stored in the Laboratory Systems layer. Then the final layer, the AI / Scientific Insight layer can enable businesses and scientists to leverage those desired, higher technologies.

## 7. The Scientific Data Capture Layer

The Scientific Data Capture Layer serves as the connective tissue between laboratory instrumentation and modern digital ecosystems. In this layer, the data is automatically captured at the point of generation. Because the data is captured at the point of generation, opportunities for manual manipulation or alteration are significantly reduced. The data moves into the scientific data capture layer from the device interface layer and an audit trail is immediately created, generating metadata around the equipment that generated the data. The data is normalized and parsing maps are applied, captured in the audit file, allowing the data to easily be sent to its desired data store. By ingesting and normalizing the data, the system is able to remain agnostic to both inputs and outputs.

For example, a scientist may generate protein concentration data from an HPLC. In many laboratory environments, the result may be exported to a CSV or Excel file, copied into another spreadsheet, and later entered into a LIMS. Within the Scientific Data Capture Layer architecture, the data is instead captured automatically at the point of generation, associated with equipment metadata and audit information, normalized into a standardized format, and securely made available to downstream systems, dashboards, reporting tools, or process equipment. This contextual metadata may include instrument identity, timestamps, operator information, method versions, batch identifiers, or network traceability information. The scientist is no longer responsible for transcribing data between disconnected systems.

Foundate Systems not only understands the equipment generating the data, but also the Quality and Security concerns. The audit file that is generated for each sample tracks information as the data moves through the system and is encrypted to safeguard against tampering. The system has components for segmented networks with a DMZ network between the Process Control Network and the Corporate Network. Transmitted through secure channels, the data comes to rest and is virus scanned before moving through the network layers, to help align with IT Security policies. The Scientific Data Capture Layer

moves data from equipment to the data store, enabling scientific insight, while keeping traceability and security.

## 8. Enabling the Next Generation of Scientific Computing

With the application of the Scientific Data Capture Layer, data is able to move from analytical equipment to the data store, and possibly back into process equipment, with confidence. As scientific data becomes centralized, structured, and reusable, it begins to create “data gravity,” enabling increasingly advanced applications and collaboration around a common trusted foundation. Organizations can then begin to conduct cross site/scale analytics, AI/ML implementations, process optimization, automated reporting, and many other tasks. The Scientific Data Capture Layer enables the next generation of scientific computing to be a possibility.

## 9. Conclusion

As organizations accelerate AI initiatives and digital transformation programs, the reliability and accessibility of foundational scientific data has become increasingly strategic. The challenge is no longer simply collecting data, but ensuring that it is trustworthy, traceable, interoperable, and available for downstream scientific computing. The future of scientific computing will not be defined solely by advances in artificial intelligence. It will be defined by the quality, availability, and trustworthiness of the foundational scientific data that powers those systems. But before organizations can fully realize AI-driven laboratories, they must first solve the foundational challenge of scientific analytical data capture.

## References

Feng, J. E., Anoushiravani, A. A., Tesoriero, P. J., Ani, L., Meftah, M., Schwarzkopf, R., & Leucht, P. (2020). *Transcription error rates in retrospective chart reviews*. *Orthopedics*, 43(5), e404–e408. <https://doi.org/10.3928/01477447-20200619-10>

Mays, J. A., & Mathias, P. C. (2019). Measuring the rate of manual transcription error in outpatient point-of-care testing. *Journal of the American Medical Informatics Association*, 26(3), 269–272. <https://doi.org/10.1093/jamia/ocy170>

Wilkinson, M. D., et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>