A Roadmap for LIMS at NIST Material Measurement Laboratory

Gretchen Greene
Jared Ragland
Zachary Trautt
June Lau
Raymond Plante
Joshua Taillon
Adam Creuziger
Chandler Becker
Joseph Bennett
Niksa Blonder
Lisa Borsuk
Carelyn Campbell
Adam Friss
Lucas Hale
Michael Halter

Robert Hanisch
Gary Hardin
Lyle Levine
Samantha Maragh
Sierra Miller
Christopher Muzny
Marcus Newrock
John Perkins
Anne Plant
Bruce Ravel
David Ross
John Henry Scott
Chris Szakal
Alessandro Tona
Peter Vallone

NIST

National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Technical Note 2216

A Roadmap for LIMS at NIST Material Measurement Laboratory

Gretchen Greene
Jared Ragland
Zachary Trautt
June Lau
Raymond Plante
Joshua Taillon
Adam Creuziger
Chandler Becker
Joseph Bennett
Niksa Blonder
Lisa Borsuk
Carelyn Campbell
Adam Friss
Lucas Hale
Michael Halter

Robert Hanisch
Gary Hardin
Lyle Levine
Samantha Maragh
Sierra Miller
Christopher Muzny
Marcus Newrock
John Perkins
Anne Plant
Bruce Ravel
David Ross
John Henry Scott
Chris Szakal
Alessandro Tona
Peter Vallone

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

Foreword

Over the past decade, emerging technology in laboratory and computational science has changed the landscape for research by accelerating the production, processing, and exchange of data. The NIST Material Measurement Laboratory community recognizes that to keep pace with the transformation of measurement science to a digital paradigm, it is essential to implement Laboratory Information Management Systems (LIMS). Effective introduction of LIMS early in the research life cycle provides direct support for planning and execution of experiments and accelerating research productivity. From this perspective, LIMS are not passive entities with isolated interaction, but rather key resources supporting collaboration, scientific integrity, and transfer of knowledge over time. They serve as a delivery system for organizational contributions to the broader federated data community, supporting both controlled and open access determined by the sensitivity of the research.

The overall goal of a successful LIMS is to empower a research community by establishing common tools providing access to laboratory data resources. Modern LIMS should therefore provide several core functions.

Core Functions and Touchpoints

- *Workflow management* – A research workflow describes steps to be performed to derive results. These patterns serve as a prescription for LIMS to control the progression of data and associated services or tools. Automation of a workflow simplifies the transfer of information through defined interfaces from a network of systems.

- *Repository of data* – Effective storage and retrieval of data (raw and derived) – including associated metadata, data products, calibration, software, logs, etc. – facilitates data discovery, processing, collaboration, and dissemination.

- *Creation of data products and tools* – A LIMS should support storage and processing of raw data leading to products which can be shared and consumed. Examples would include sample data, instrument generated data, and algorithms generating defined outputs. Inclusion of data models provides context and structure, and Machine Learning integration may generate related data which could be combined into more comprehensive data models. Tools may include visualization, evaluation, and analysis packages offering users advanced capabilities for their research projects.

- *Organization of data for search and retrieval* – Tools and interfaces give users access to sophisticated searches of data holdings and efficient mechanisms for data transfer in standardized formats. Searching should extend to domain or project specific semantics, be coupled closely with related data, and go beyond individual research projects to include super-searches (e.g., use case driven interoperability between LIMS).

- *Long-lived, stable, and agile structures* – LIMS require institutional and architectural sustainability for long baseline research and curatorship. Technology tends to change faster than the practical lifetime of research programs, so paths must exist for maintaining IT infrastructure and introducing faster and more complex processes.

- *Standards and best practices* – LIMS benefit from standardization to support collaborations among research communities and make data workflows efficient and affordable. Community buy-in for standards and best practices is an essential part of LIMS and organizational shared expertise naturally serves as a means for coordination and adaptation of standards.

- *User involvement* – In all the core functions listed above, it is critical to involve the subject matter experts from the beginning. LIMS should establish a working team that explicitly includes representatives from the end user community.

Abstract

Instrumentation generates data faster and in greater quantity than ever before, and interlaboratory research is in historic demand domestically and internationally to stimulate economic innovation. Strategic mission needs of the NIST Material Measurement Laboratory (MML) to support a wide array of research disciplines therefore compel our organization to adopt advanced strategies for research data management. Laboratory Information Management Systems (LIMS) provide a framework for managing data from the outset of the research life cycle delivering new capabilities for machine learning (ML), data analysis, collaboration, and dissemination. This roadmap describes our current understanding and strategy for adapting our research workflows for LIMS throughout MML by embracing the use of standards and best practices from data science communities. The NIST research data cyber infrastructure complements these goals for MML by providing a secure environment to host LIMS solutions. Additionally, integration of scientific workflows requires ongoing collaboration to bridge organizational LIMS with external scientific communities. Thus, MML LIMS will evolve over time in synergy with the technology and experimental environments delivering new science. LIMS will broaden our mission impact through adoption of FAIR data principles.

Key words

Table of Contents

List of Tables

List of Figures

1. Introduction

Beginning late 2019, MML initiated as part of its strategic plan the development of "*next generation"* data and informatics with a focus on LIMS as a key resource to support research data and science. This effort was complemented by initiatives for enhancing data management planning and data systems infrastructure. The first year's groundwork established common needs for both individual researchers and teams to engage more readily with LIMS with a goal of building greater capacity for interaction and use of data. A vision for LIMS was written to convey the purpose of these collective efforts:

"*Laboratory Information Management Systems will provide MML scientists a practical means for repeatability, traceability, reproducibility, efficiency, and compliance of research, serving as a beacon to both intramural and extramural community stakeholders.*"

The MML approach to implementing LIMS started with defining goals for specific division research laboratory projects and established a cross-divisional Community of Interest (COI) group for sharing solutions, services, practices, and challenges. Comprehensive LIMS solutions have been successfully implemented in several NIST laboratories. Several shared resources have successfully demonstrated use of LIMS components including repository platforms, a standard persistent identifier service, a centralized research data storage with networked data transfer nodes, and data transfer services, in addition to expertise in data modeling and semantics. These resources along with community best practices contribute to a basic LIMS architecture for research.

A system view and architecture model provide the foundation for planned future outcomes. In this Roadmap, we define a set of research-oriented LIMS capabilities which serve to guide implementation along with components to deliver these capabilities. More detailed guidance for use of specific LIMS resources is available internally to NIST and where possible shared on external repository websites. It is also anticipated that LIMS implementation will provide an important contribution to the goals of NIST program areas such as Artificial Intelligence (AI), biosystems, chemical informatics, additive manufacturing, and the materials science areas which spearheaded early innovation in data systems through the Materials Genome Initiative.

2. Roadmap Objectives

This roadmap provides a framework for manifesting the MML LIMS vision and outlines the key objectives highlighted by the MML LIMS COI project goals. These are grouped into near- and long-term objectives for MML LIMS prioritization. These objectives along with broader community efforts will strengthen the NIST data-as-an-asset[1] strategic approach to research. More comprehensive goals such as the development of a *Digital Twin[2]* will enable models to probe the measurement science space to further analyze the physics and gain understanding leading to new science.

1

Near Term Objectives:
- Establish a LIMS COI for MML[#]
- Develop pilot LIMS solutions for targeted research workflows [#]
- Design tiered LIMS architectures to support a range of research workflow implementations
- Establish core infrastructure services to support LIMS
- Develop data acquisition and experimental activity capture solutions
- Deploy key functional support services such as Handle Service (supports Persistent Identifiers) and Data Transfer Service (see Section 4.3)
- Prototype and exchange LIMS components as a basis for shared resources (e.g., repository platforms for experimental activity, instrumentation, samples, Extract Transform Load (ETL))
- Establish best practices in digital object management to support standards and community practice

[#] As of the time of this writing, a LIMS COI is in place and several solutions have been piloted and deployed for laboratory operations.

Long Term Objectives
- Establish best practices for development and implementation of data model and semantics
- Deploy prioritized LIMS end-to-end solutions (achieving multi-component level functionality)
- Develop use case driven solutions for cross-laboratory LIMS interconnectivity
- Provide on-demand system level LIMS resources for research teams at NIST for rapid engagement
- Develop automated workflow integration between LIMS and computational platforms (High Performance Computing (HPC), ML, analysis applications, e.g., SciServer)
- Develop integration between LIMS and public data access systems
- Establish methodology for building and applying *Digital Twin models*
- Provide NIST leadership a strategy for adoption and implementation of LIMS to promote innovative data-driven science

2.1.   Challenges

Within the MML LIMS COI exchange forum, lessons learned and shared by early adopters of LIMS highlighted several key challenges. These will be factored into data management plans for implementation and extended to the broader MML community which relies on either operation of LIMS or end-usage LIMS outputs.

One common challenge with instrumentation data is generation of vendor-proprietary output formats. A repository for sharing data format exchange software tools is a good example of a solution benefiting LIMS by supporting the need to transform vendor data into more

2

consumable open data formats for downstream analysis and computation. A prototype repository was created by the Office of Data and Informatics (ODI) with a few extractor tools and efforts are underway to explore how this may achieve wider utility. Community repositories such as Bio-Formats and Materials.IO, are examples of resources which support tools for conversions of third-party data into open data models. These community-oriented solutions successfully demonstrate methods to lower the barrier for LIMS through shared software.

Another challenge is finding appropriately skilled labor resources required for engineering domain-specific LIMS workflows. This is a common barrier to LIMS prioritization for research organizations. Integrating data structures requires close collaboration between domain and data science subject matter experts (SMEs) for modeling and mapping of multiple source data to repository storage.

Integration or migration of legacy systems and bespoke tools with next-generation LIMS architecture presents another challenge, especially for those with limited resources supporting maintenance. Legacy systems commonly lack sustainability due to factors such as end of funding support or unavailable expertise.

Data provenance is commonly required for sample tracking and traceability across laboratory processes (e.g., sample transformations, generation of parts, or interlaboratory sample exchange). The latter is a challenge for interlaboratory study because data management systems (including LIMS) are most commonly not standard or normalized. Supporting common data exchange protocols and chain of custody workflows will be an ongoing design consideration for interoperability, including concepts such as data trust and integrity.

Another common challenge is operational security compliance for IT infrastructure. NIST adheres to CIS Controls (Critical Security Controls) and as LIMS architectures rely on networked systems, this translates to requirements for vigilant monitoring of service and platform deployments to ensure organizational security.

3. MML LIMS Stakeholders

MML LIMS has a number of stakeholders with differing needs and priorities. ODI and research project leads work with all stakeholder categories, balancing goals to develop options that provide the best overall benefit.

Research Community
Each research discipline area involves a community benefiting from the generation and exchange of data outcomes.

MML Laboratory Office
The MML Lab Office ensures support of the NIST organizational mission through the productivity achieved from the use of LIMS.

3

Program funding sources
Funding for strategic initiatives and research priorities for NIST and MML address both internal and external stakeholder needs and account for LIMS resources.

Collaborative Partners
Critical stakeholders help foster research innovation and science through exchange of LIMS data and software products. A key benefit to LIMS is the ability to support collaboration through exchange and access to data, data products, software, and related resources.

3.1. External Solutions and Partnering

Working with the external community has resulted in a more robust MML LIMS knowledge base benefiting from partnerships with external researchers during LIMS development and implementation, commercial procurement, and collaboration. Several organizational partners have contributed to MML efforts including National Research Energy Laboratory[3], Oak Ridge National Laboratory (ORNL), Brookhaven National Laboratory (BNL), Air Force Research Laboratory, 3M Corporation, NASA, University of Illinois, Champaign and national forensic crime labs in collaboration with the National Institute of Justice[4]. Many NIST strategic research areas rely on collaborative data engagement, thus having a LIMS with capabilities for supporting external access is important in many cases.

Collaboration with external partners may involve co-development of a LIMS platform system (or component), adoption of a community LIMS, shared access to data/code, or support for commercial vendor solution customization. Such collaborations often include harmonizing requirements and goals to inform the design, architecture, and engineering of a solution. Open-source solutions often provide more mechanisms supporting the complexity of many research workflows; examples include flexible configuration, customization, and independent software enhancements. A few examples include ORNL DataFed[5], SynBioHub, BNL's BlueSky[6] platform, and SciServer. While these may require engineering expertise to fit within the NIST infrastructure, each adds meaningful user capabilities.

As one example, NIST synchrotron beamline stations at the BNL National Synchrotron Light Source II (NSLS-II) Facility have implemented BlueSky in partnership with BNL's Data Science and Systems Integration group.

Solutions have also been developed by partnering within NIST such as (1) the NEXUS Electron Microscopy LIMS (NEXUS-LIMS[7]) workflow based on the Configurable Data Repository System (CDCS), a NIST-developed open platform, and (2) a LIMS supporting real-time biosystem cell line sample tracking LIMS with a custom Excel application used for experimental activity capture. Both are highly customized to the research requirements of their respective user groups.

A few commercial solutions[1] were adopted in part or in full. In one successful example, the NIST Center for Automotive Lightweighting (NCAL) successfully implemented the commercial platform Ansys Granta. Other community and vendor solutions have been and

---

[1] Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

continue to be evaluated for pilot use such as 4CeeD[8], Tadabase no code solution, Benchling cloud platform, Microsoft platforms and others. Due to the challenging nature of customized research workflows and complexity in the secure integration with government networked infrastructure, commercial solutions may pose additional cost and skill set requirements for successful adaptation into the laboratory working environment. In most cases at NIST, use case development leads to the adoption of hybrid solutions. Closed-source solutions widely adopted in the scientific community, e.g., Globus, provide unique and robust capabilities that are difficult to recreate. In the instance of Globus, the linkage to several research community services and best practices such as GridFTP, linked identity management with institutions authentication including InCommon, a python software development kit, and multi-platform storage connections provide high value and ease of adoption.

4. Implementation of the Roadmap

4.1. System Level Solutions

LIMS, as we define it, is a *system* of components which delivers the capabilities for the early stages of a research life cycle[9]. It is widely recognized there is no singular solution for LIMS across all disciplines of research, yet shared components (and data assets) provide greater economy of scale and consistent usage across the organization and beyond. A LIMS design implements research workflow requirements and provides context for assembling an architecture supporting component integration to produce desired outcomes. Off-the-Shelf (OTS) LIMS solutions are often challenging to adopt, primarily due to the limitations in both configuration and customization of components to match workflows. Monolithic solutions have demonstrated challenges whether they are homegrown or OTS by constraining interfaces between components. With service-oriented solutions, workflow flexibility, orchestration, and evolution can be managed with lower risk[10] to overall performance. Therefore, implementations may vary as to which components are used and in which sequence to support the required research workflow. A tiered model ranging from basic components and plug-in architecture to instantiation of more complex data models with computational support can lower the barrier for entry. Given this context and the goal of flexibility, a few commonly used LIMS workflow components provide critical functionality.

4.2. LIMS Tiers

Requirements for building or deploying LIMS are dependent upon the complexity of the research workflow and project needs. In the consideration of the system, a basic tiered model serves as a general guide for capabilities mapping and implementation, to level of effort (Fig. 1). As with many aspects of LIMS, this model may have variations in the strata depending on the optimal architecture for achieving desired outcomes.

Tier 1 is primarily reliant on infrastructure and support services and can be readily adopted with available "on-demand" resources, meaning these exist and can be provisioned for use upon request. The main functions for Tier 1 LIMS include data acquisition, near-line instrument data collection, sample tracking, and data movement, possibly automated, to a

storage location, ideally with access for processing and analysis. Tier 1 LIMS support unstructured data and may provide more flexibility for researchers to experiment with data structures, formats, and tuning for workflow.

Tier 2 integrates research workflow design including data structuring, formatting, metadata, and possible integration to a data repository solution, and requires a greater level of effort and engagement between a research SME and data science engineers for design, installation, and operational configuration.

Tier 3 LIMS require the highest level of commitment and collaboration between the research SME and a data science expert to factor in all the functionality for more complex workflows. This might involve functionality such as computational system integration, support for external client tools through an application programming interface (API), and semantic modeling to factor in community adoption for data consumption. An example of a Tier 3 LIMS would include data generated from multiple instruments, each with variable processes, interconnected to computational tools and applications to produce data products, with dependencies on analytical requirements for external interoperability. Both the MML NEXUS LIMS and Granta NCAL systems are examples of Tier 3.
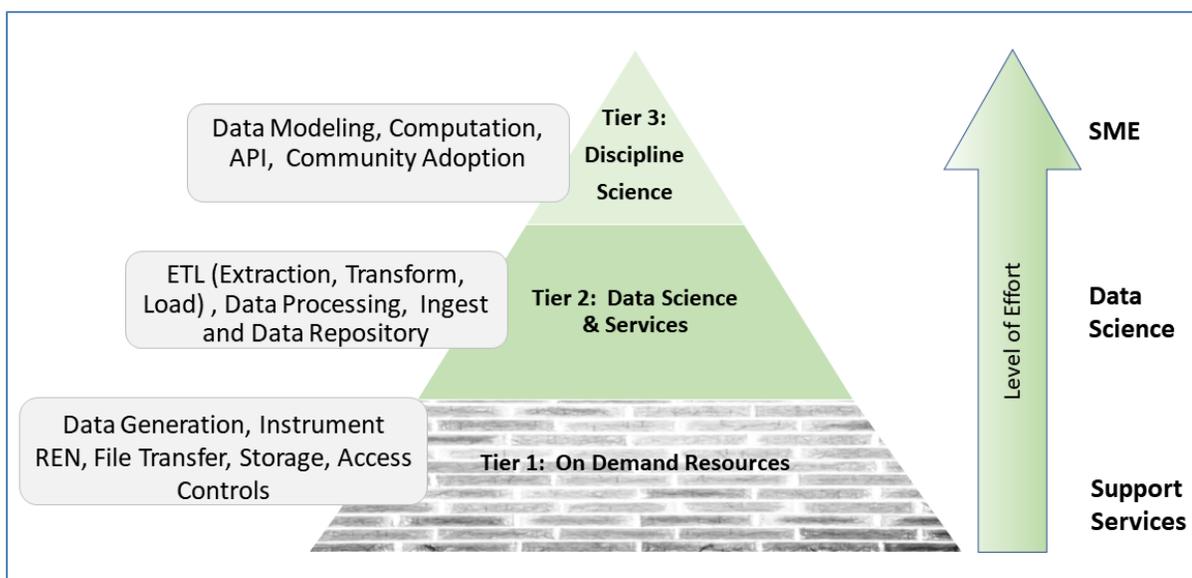
Fig. 1. LIMS three tiered model for implementation

4.3. Supporting Services

Common services support LIMS at various touch points throughout the research data workflow. These may be considered part of the infrastructure, i.e., they are underlying services shared and configured to support more than one stage of a LIMS workflow, yet by nature they often require a contextual prescription for how they interface with a LIMS workflow of component. Examples of these key support services include:

- Data transfer services: manual, automation, and tools for file movement between storage locations
- Handle service: generates and resolves a standard Persistent Identifier (PID) known as a Handle
- Repository Systems: OTS solutions or well-established and supported repository platforms (e.g., CDCS, Cordra, GitLab, GitHub) and open-source solutions supporting customization for (meta)data
- Containerized Deployments: Docker, Kubernetes, virtual machine/Cloud services supporting deployments of LIMS applications and tools
- Solutions Brokering: Documentation and communications including instructional guides ("playbooks") and consultation for design and use of LIMS systems, services, and components

4.4. Infrastructure Resources for LIMS

Organizationally MML and NIST both provide infrastructure-as-a-service for networking, storage, and compute resources. These systems may be requested through internal NIST IT services and are readily available.

As part of the MML Data and Informatics strategic plan, both network and storage have been significantly expanded to support higher bandwidth for data transfer between laboratory instruments and storage (see Fig. 2). Data space allocations are designed to be flexible and can be established on request for cross-organizational projects, research teams, and instrument dedicated endpoints. Several network attached storage (NAS) solutions have been implemented to support both localized (Data Collection Nodes) and a Central Data Store (a dedicated MML Research Data storage array). Additionally, the NIST Amazon Web Service (AWS) cloud environment is available and integrated with the NIST VPN (virtual private network) providing both storage and compute on demand.

A Research Equipment Network (REN) is used to manage pass through and routing between instrument laboratory equipment and data storage. Network configuration via the REN ensures that instrument-control computers are protected and isolated as needed for secure operations. Higher speed network backbones and a Science DMZ[11] is currently in a planning phase to provide higher throughputs and secured zones for operating with LIMS architectures.
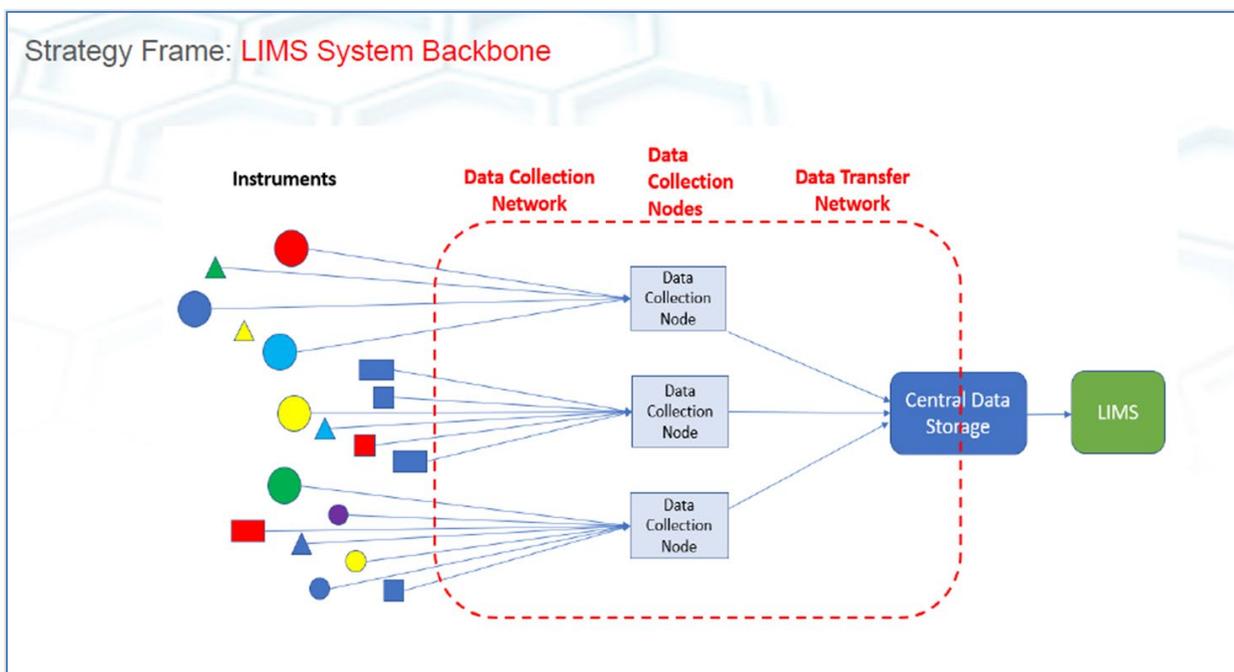
Fig. 1. MML LIMS data plumbing model for harvesting instrumental raw data

Current infrastructure resources are highlighted in bold, while ongoing efforts are in progress to expand capabilities in support of LIMS.

- Research Data Storage (**DCN, NAS, CDS, AWS, Local**)
- Computational Resources (**ENKI**, **AWS**, **SciServer**), **HPC**)
- Security Controls (CIS)

Security for government research networks is a top priority for NIST, and all LIMS deployments must adhere to cyber security controls and processes for monitoring and managing access. This critical infrastructure element requires coordination between LIMS developers and IT security offices to ensure systems are protected and compliant with CIS including monitoring, security patching, and notification of problems for mitigation. In developing a plan to implement LIMS, resource needs to complete security assessment and authorization should be taken into consideration.

4.5. Community Standards and Best Practices

Several standards and best practices in the research data community are key to building an effective LIMS workspace. While we list a few concepts core to LIMS and examples of use, these are only a small subset in the growing field of Data Science. Standards and best practices will continue to evolve, requiring resources for maintenance, expansion, community engagement, and user training, though one goal of a successful LIMS implementation is to minimize these burdens for end users.

Table 1. Data Standards and Practices used with LIMS

Community Standard/Practice	Examples of Use
Persistent Identifiers	DOIs, ORCIDs, ARKid; Handles
Open-source code	USNISTGOV GitHub (ETL pipelines, libraries, tools)
FAIR[12] (Findable – Accessible – Interoperable – Reusable) Data Principles	Go-Fair.org, FAIR Maturity Model[13]
Semantic Standards	DCAT, Schema.org, DataCite, Linked-Data, Discipline specific taxonomy, ontological
Standardized communications network protocols	RESTful API–Open API, e.g., Swagger Docs

5. LIMS Architecture

5.1. Capabilities of LIMS

LIMS helps a researcher by ***providing capabilities*** in an integrated way. These capabilities are delivered by one or more of the system components and serve as drivers for LIMS requirements. Common data-oriented functions, such as "sample-tracking", from the user perspective might also be more readily identified in terms of the capabilities. These may take on slightly different naming convention and forms across implementations, yet in general they can be described in the context of the function they support within research environments. Capabilities may be mapped into the design architecture through implementation of modular components as shown in Fig. 3. For external vendor and community open LIMS platforms or sub-components, understanding how and where in the system these capabilities are supported is also key to evaluating their compatibility with research workflow.

Core capabilities generally fall into one of several categories. A few example aspects supported by each category are provided.

> **Data Generation** – provenance of data, software descriptions and custom code, instrument configuration, data acquisition descriptions

> **Digital Information Management –** data models, data formats and reformatting, file and data set access, persistent identifier resolution, curation, sample tracking

> **Data Ingestion –** absorption of data, attachment of metadata, digital transformations, support for combinations of structured and non-structured data sources

> **Data Archiving –** long term preservation, automated backup, user permission access controls, life cycle maintenance

Processing and Analysis – instrument and model calibration records, data pipelines for computation of derived or final results, visualization and reporting tools

Data Publishing and Sharing – table or figure preparation, organization, cross-holdings search, data as a service

Digital Asset Movement – internal and external sharing, migration between platforms, exports to other platforms, life cycle management, and others that support:

      … capture and organization of data from equipment;
      … preservation and recovery data;
      … tracking of samples;
      … automated conversion of data to open formats;
      … linking information between systems;
      … data querying for download; and
      … countless others as needs evolve.

5.2.   LIMS Components

The architecture of LIMS is built with functional components illustrated in Fig. 3, which may be implemented using different platform services, tools, and interfaces. They are shown here grouped into a set of four functional categories (generation, persistence, communication, consumption, and distribution) which are also illustrated in the LIMS networked architecture (see Fig. 6). Systems engineering factors into capabilities, components, and research requirements to build out an integrated LIMS such as the sample model illustrated in Fig. 4. The model shows examples of data generation: instruments, Electronic Laboratory Notebook (ELN)[14], or Simulation which are managed by intake to repository and file storage to support processing, analysis and user search and access.
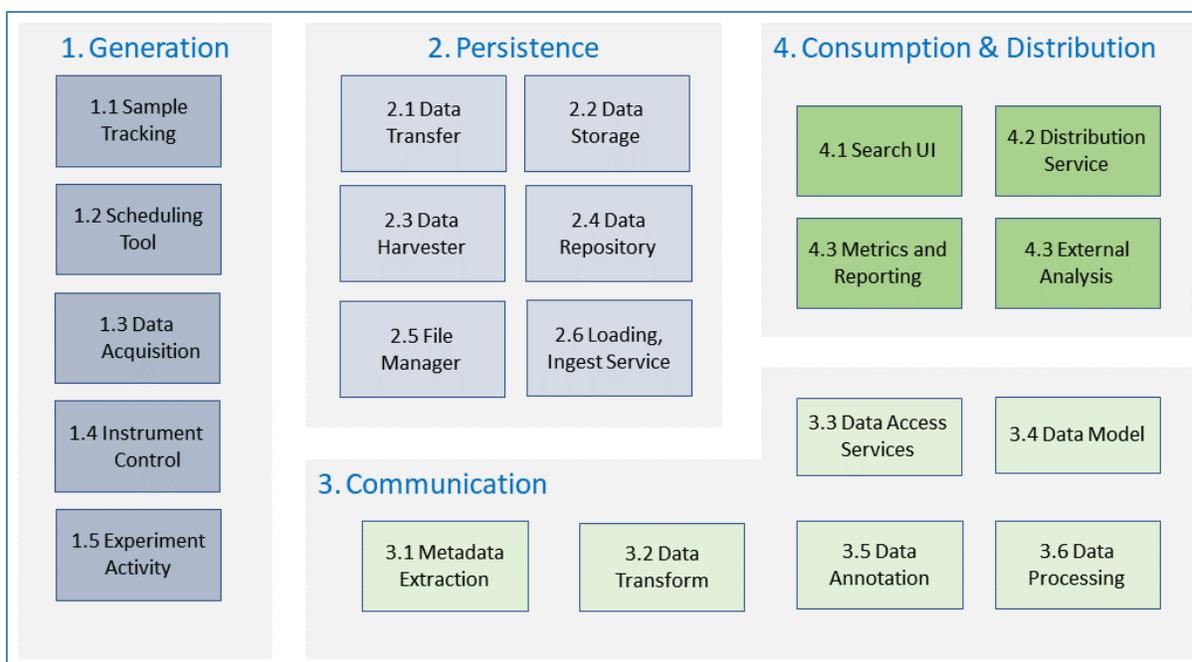
**Fig. 2.** LIMS components grouped by major functional category
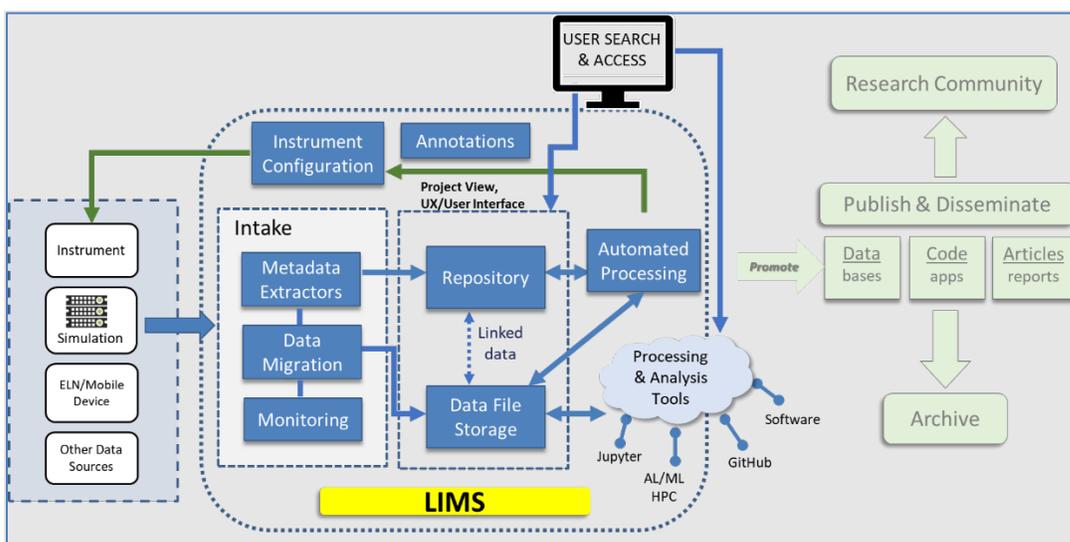
Fig. 3. A basic LIMS functional workflow model

5.3. Research Workflows

Requirements for each research discipline may be unique and require customized workflows. There are many common patterns, though each instrument and component implementation may have a unique interface for the input and output. It is critical when defining workflows to incorporate the perspective of experimentalists (or theoreticians and computational scientists) such that key inputs and outputs are captured in the correct sequence and factor into the human interaction dependencies and touchpoints. Several common themes include

activity capture, management and recall of varied data, availability of infrastructure and resource skills, and iterative data planning to improve upon practice. One approach is to define the native workflow (perhaps through a white-boarding exercise) to identify the principle experimental attributes, e.g., instrument, calendar (time), sample preparation and characterization, laboratory notes and observations, and processes. Another exercise complementary to the research workflow is the definition of downstream data queries which SMEs would routinely use. The emerging theme of "decision-science" has been introduced as the concept which helps use-case driven design to factor in what questions (query) will the LIMS support through access to holdings/services/tools. These queries motivate the process of analyzing, describing, and defining workflows. Furthermore, they provide scientifically motivated context for LIMS design such that the granularity of structured data and operational utility of the system will provide appropriate functionality for analytical tools.

Examples of research workflow are illustrated in Fig. 5, showing how native workflows serve as use cases to identify which LIMS components are required to support specific capabilities.
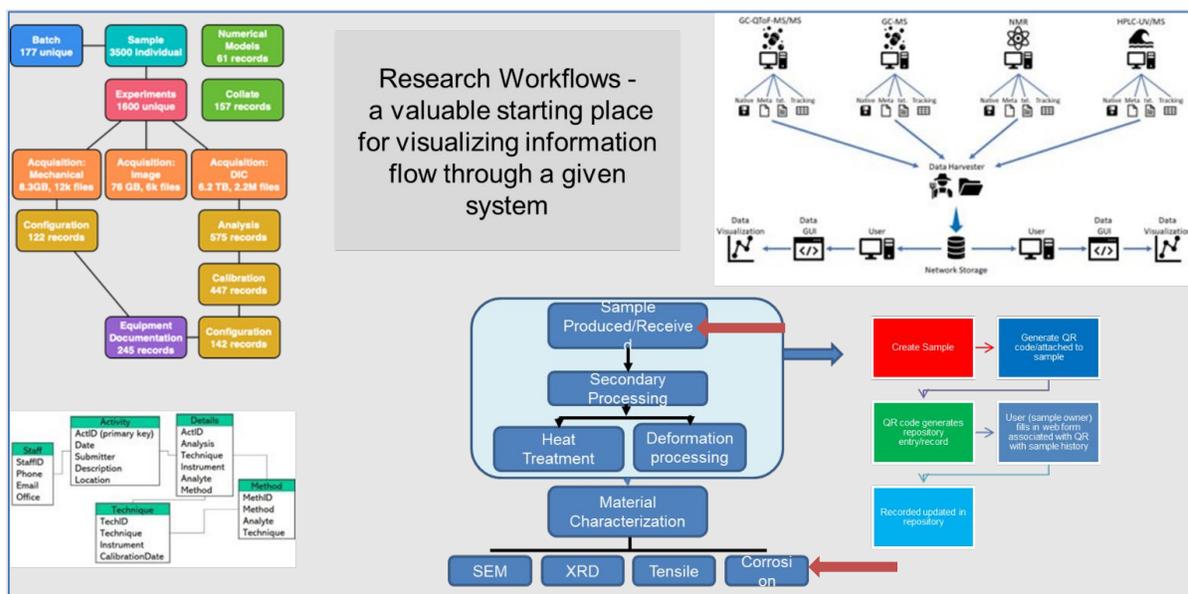
Fig. 4. Assorted research workflows in native model representations

5.4. LIMS Networked Architecture

An end-to-end architecture workflow view is shown in Fig. 6 with networked functionality. The data progresses from generation through persistence via communication and leads to resulting consumption and distribution. This illustrates at a conceptual level a generalized workflow with networked systems. The operational requirements for an end-to-end architecture to support scientific application requires infrastructure reliability, performance, reference "playbooks" for developers, and consistent availability of microservices which provide the flexibility for integration and functionality.
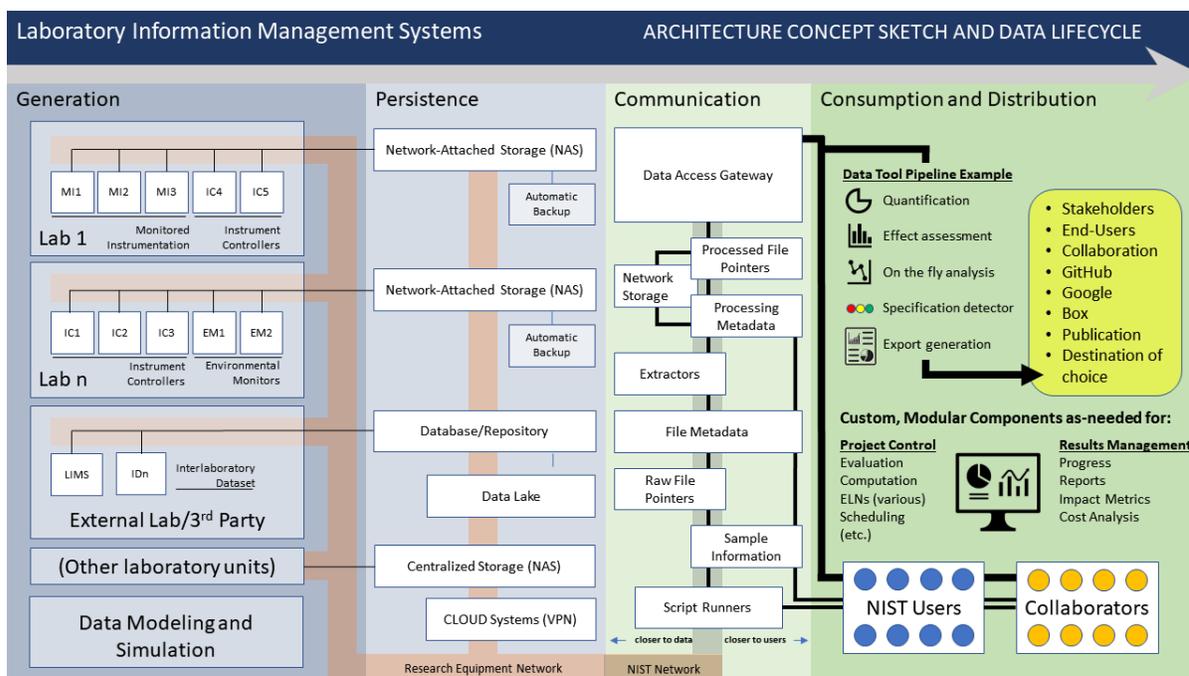
Fig. 5. LIMS networked architecture concept with four stages of operation

The supporting LIMS infrastructure takes into consideration the supporting services and sources of information and is critical in providing the foundation for applications. Sub-LIMS - i.e., specialized systems such as freezer systems, inventory and tracking systems - also need to be accounted for, in particular those which will continue to involve manual inputs and operations. The "people" resources for LIMS, shared infrastructure, and specialized systems are critical resources for building and maintaining an environment for successful operation.

6. Metrics and Qualitative Measures

As important as the establishing and vetting of the vision, strategic goals and operating principles outlined here is the establishment of a set of metrics that can be used internally and externally to quantify how well NIST MML LIMS efforts are progressing.

- *LIMS Usage Statistics* – MML and NIST are beginning the process of collecting statistics on usage and impact of data and software systems. These are expected to be useful for project implementation as well as for all activities in which MML promotes the benefits of data-driven research. To achieve this, LIMS efforts will require supporting systems for sharing usage of instrumentation and technology in MML.

- *High Science Impact Capabilities* – While we cannot guarantee publication or adoption in high impact science areas, we should be able to demonstrate the potential for such results to be produced. Metrics should focus on data products, features, and processes that

13

increase the potential for High Impact Science. Peer reviewed papers, data publications, and industry impact measurements will be used to demonstrate this measure.

- *Collaborative Activities and Capabilities* – As a major focus for our future, tracking the number and extent of collaborative activities will be important. Tracking should include quality and cost effectiveness assessments.

- *Cost, Schedule, and Scope Targets for Milestones* – These metrics will be used for demonstrating value as well as for highlighting potential efficiency areas.

7. Conclusion

In the MML research environment, a wide array of applications and digital tools are adopted to conduct both experimental and theoretical measurement science. We have embarked upon a new mission space for data management and sharing which we believe will significantly enhance the scientific return on investment in our laboratory. Adoption of LIMS provides a modern approach toward this goal by embedding data and software systems supporting research workflows and capturing information in a FAIR manner. In this Roadmap we introduce important concepts for the design and implementation of LIMS. Benefits of LIMS are already beginning to demonstrate their value to science through more efficient access to data for analysis over time. Our next-generation LIMS will provide standards for interoperability and collaboration, further enabling scientific investigation spanning across experimental groups. With proper LIMS design in place, solutions will evolve in tandem with research domain knowledge, providing an essential resource to stimulate scientific innovation.

References

[1] U.S. General Services Administration, Technology Transformation Services, *Federal Data Strategy: Leveraging Data as a Strategic Asset*, Available at https://strategy.data.gov/resources/
[2] Knapp GL (2017) Building blocks for a digital twin of additive manufacturing, *Acta Materialia*,Volume 135, ,pp 390-399, https://doi.org/10.1016/j.actamat.2017.06.039

[3] Talley K (2021) Research data infrastructure for high-throughput experimental materials science, *Patterns*, Volume 2, Issue 12, 100373, ISSN 2666-3899, https://doi.org/10.1016/j.patter.2021.100373

[4] National Institute of Justice Forensic Technology Center of Excellence, *A Landscape Study of Laboratory Information Management Systems (LIMS) for Forensic Crime Labs*, Available at https://nij.ojp.gov/library/publications/landscape-study-laboratory-information-management-systems-forensic-crime

[5] Oak Ridge National Laboratory Leadership Computing Facility. *Datafed – A Federated Scientific Data Management System*. Available at https://www.olcf.ornl.gov/olcf-resources/rd-project/datafed-a-federated-scientific-data-management-system/

[6] Arkilic D (2017) Towards Integrated Facility-Wide Data Acquisition and Analysis at NSLS-II, *Synchrotron Radiation News*, 30:2, pp 44-45, https://doi.org/10.1080/08940886.2017.1289810

[7] Taillon J (2021) NexusLIMS: A Laboratory Information Management System for Shared-Use Electron Microscopy Facilities. *Microscopy and Microanalysis*, 27(3), 511-527. https://doi.org/10.1017/S1431927621000222

[8] Lau J (2019), Laboratory Information Management Systems for Electron Microscopy: Evaluation of the 4CeeD Data Curation Platform, *Journal of Research (NIST JRES)*, National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/jres.124.034

[9] University Santa Cruz University Library, *Research Data Management Lifecycle*. Available at https://guides.library.ucsc.edu/datamanagement/

[10] U.S. General Services Administration, *De-Risking Government Technology: A Federal Field Guide*, Available at https://derisking-guide.18f.gov/federal-field-guide/

[11] Dart E (2014). The Science DMZ: A Network Design Pattern for Data-Intensive Science. *Scientific Programming*. 22. https://doi.org/10.3233/SPR-140382

[12] Wilkinson M (2019) Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data* 6, 174. https://doi.org/10.1038/s41597-019-0184-5

[13] Wilkinson M (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

[14] Gates R (2015), Smart Electronic Laboratory Notebooks for the NIST Research Environment, *Journal of Research (NIST JRES)*, National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/jres.120.018

Appendix A: Acronyms and Initialisms

AFRL	Air Force Research Laboratory
AI	Artificial Intelligence
ARK	Archival Resource Key
API	Application Programming Interface
APPS	Applications
AWS	Amazon Web Services
BNL	Brookhaven National Laboratory
CDCS	Configurable Data Curation System
CDS	Central Data Storage
CIS	Critical Security
COI	Community of Interest
DCAT	Data Catalog Vocabulary
DCN	Data Compute Node
DOI	Digital Object Identifier
DSSI	Data Science and Systems Integration
ELN	Electronic Laboratory Notebook
EM	Environmental Monitor
ESNET	Energy Sciences Network
ETL	Extract Transform Load
FAIR	Findable, Accessible, Interoperable, Reusable
FTP	File Transfer Protocol
HPC	High Performance Computing
I2	Internet 2
IaaS	Infrastructure-as-a-Service
IC	Instrument Controller
IDn	Interlaboratory Dataset
IT	Information Technology
LIMS	Laboratory Information Management System
MI	Monitored Instrumentation
ML	Machine Learning
MML	Material Measurement Laboratory
NAS	Network Attached Storage
NASA	National Aeronautics and Space Administration
NCAL	NIST Center for Automotive Lightweighting
NIST	National Institute of Standards and Technology
NREL	National Renewable Energy Laboratory
NSLS-II	National Synchrotron Light Source-II
ODI	Office of Data and Informatics (in MML, NIST)
ORCID	Open Researcher and Contributor IDentifier
ORNL	Oak Ridge National Laboratory
OTS	Off-the-Shelf
PID	Persistent IDentifier
REN	Research Equipment Network
REST	REpresentational State Transfer
SDK	Software Development Kit

SEM	Scanning Electron Microscope
SME	Subject Matter Expert
UIUC	University of Illinois, Champaign
UX	User Experience
VM	Virtual Machine
VPN	Virtual Private Network
XRD	X-Ray Diffraction