

An update from the MaRDA LIMS Working Group

Eric Stach, Joshua Taillon, & the WG members

2024 Spring MRS

Monday, Apr 22, 2024 - 16:00



NIST Disclaimer

Certain commercial equipment, instruments, materials, vendors, and software are identified in this talk for example purposes and to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Any opinions expressed are my own, and not a statement on behalf of the U.S. Government.

Eric Stach Disclaimer

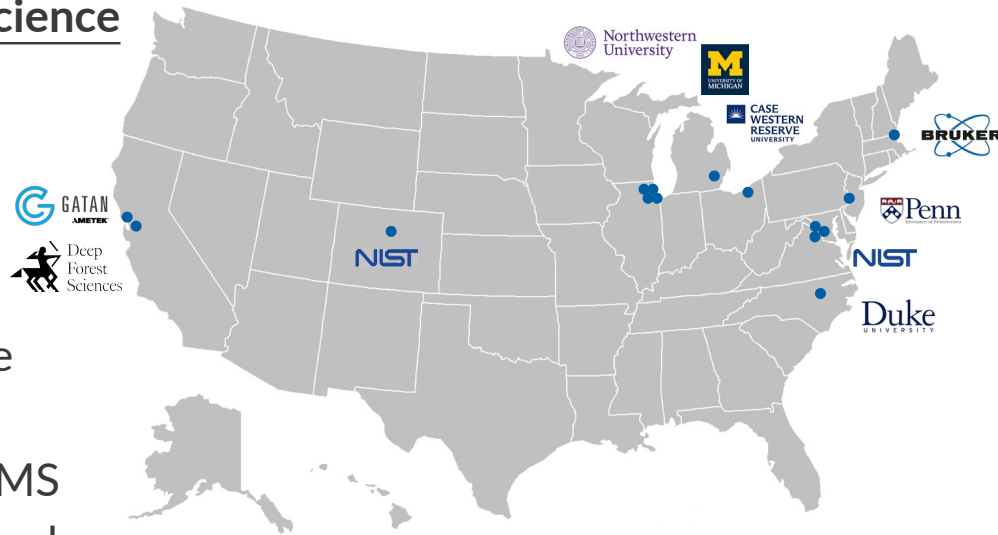
I became involved in this activity prior to becoming MRS Vice President.

This means that I am not representing the MRS or speaking on behalf of the MRS.

Introducing the FAIR LIMS Data Working Group

Brings together nationwide experts in materials science and engineering from across academia, government, and industry

Primary aim is to make recommendations to facilitate the use of LIMS in materials research and engineering



Co-Chairs

Eric Stach *University of Pennsylvania*

Josh Taillon *NIST*

Ex Officio

June Lau, *NIST*

Laura Bartolo, *Northwestern*

Members

John Allison *University of Michigan*

Jennifer Carter

Case Western Reserve University

Carelyn Campbell *NIST*

Kamal Choudhary *NIST*

Cory Czarnik *Gatan, Inc.*

Dieter Isheim *Northwestern University*

Derk Joester *Northwestern University*

Bharath Ramsundar *Deep Forest Sciences*

Roberto dos Reis *Northwestern University*

Richard Sheridan *Duke University*

Doug Stauffer *Bruker Corporation*

What do we mean by LIMS?

Laboratory Information Management Systems

“ LIMS are key resources supporting collaboration, scientific integrity, and transfer of knowledge over time [and can] empower a research community by establishing common tools providing access to laboratory data resources [1] ”

Some example functions of LIMS (non-exhaustive):



[1] “A Roadmap for LIMS at NIST Material Measurement Laboratory” NIST Technical Note 2216 (2022), National Institute of Standards and Technology, Gaithersburg, MD, [online], <https://doi.org/10.6028/NIST.TN.2216>

Working group goals

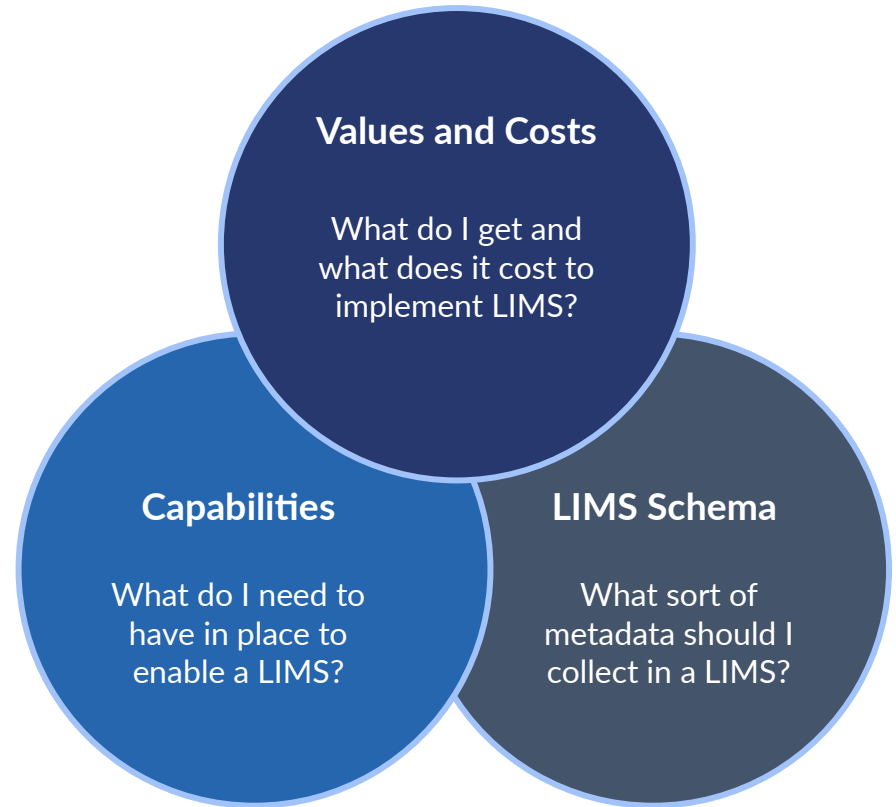
- Document best practices for recording experimental metadata in materials analysis (both research and engineering)
- Foster awareness of existing recognized schemas used in LIMS
- Identify (any) key gaps in existing schemas and metadata recording methods in current materials research practices
- Document how LIMS can benefit individuals, institutions, and the community as a whole through better practices in recording and sharing of data

Deliverables

- Documentation of various capabilities required for implementation of a LIMS
- Description of value added by LIMS, together with associated costs for various stakeholders (researchers, project leaders, facility managers, vendors, etc.)
- Review of existing recognized metadata practices in materials science and cognate disciplines (with references and links)
- Development and publication of an actionable set of recommendations for implementing LIMS in support of materials research and engineering

Working group focus areas

- Working group divided into three focus areas depending on member interests
- Sub-groups work independently and periodically report back to the larger group
- Outputs of each group will be collected into a set of recommendations for final publication to the community



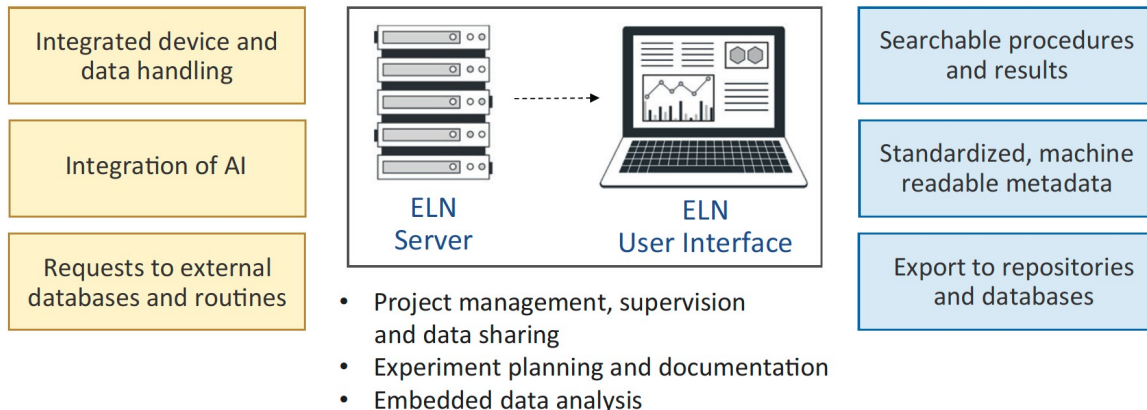
Progress to date and expected timeline

| Early 2023 | May 2023 | Oct 2023 | Early 2024 | Spring 2024 | Fall 2024 |
|--|---|---|--|---|---|
| Working Groups Founded Members and chairs identified, initial in-person meeting scheduled, virtual meetings, WG proposal drafted | In-person meetings to finalize proposal Working groups met at Northwestern University to hone working group proposals and expected outcomes | Instrument vendor WG meeting Working groups meet to gather feedback from instrument vendors Vendor representatives added to working groups | Preparation of draft recommendations Working groups will prepare drafts of their recommendations and request feedback from MaRDA membership and interested parties | Gather feedback Drafts will be presented at MRS Spring Meeting and other venues Feedback will be gathered and incorporated to generate final recommendations | Publish final recommendations Final recommendations will be published online and in a suitable journal by August 2024 |

Costs & Benefits

- Costs: Software implementation & storage costs
 - Can range from as low as \$30k to \$150k depending on storage size, onsite versus cloud
 - One-time vs. ongoing costs for maintenance
- Benefits:
 - Increase reliability and reproducibility
 - Makes compliance with FAIR data principles easier
 - Allows data to be better utilized in automated data analysis and eventually autonomous experimentation

Benefits and characteristics of ELNs



The benefits of using an ELN for the management, documentation, and analysis of scientific data. Left in yellow: benefits related to the support of the scientists' work processes; Right in blue: benefits related to a later re-use of the data and metadata. Images used for the figure were generated by C. Henken, Karlsruhe Institute of Technology (KIT), ZML - Center for Technology-Enhanced Learning, License: CC-BY.

[2] The Impact of Digitalized Data Management on Materials Systems Workflows
Frank Tristram, et al., Adv. Func. Mater., 2303616, 2023
<https://doi.org/10.1002/adfm.202303615>

Prerequisites, Capabilities, and Roles

- Prior to LIMS implementation, an organization needs certain pieces in place
- Planning stages:
 - Clearly defined goals and objectives
 - A project team representing all relevant departments and stakeholders
 - Understanding of the laboratory's workflows and processes
 - Comprehensive inventory of all laboratory equipment, instruments, and software
 - Data migration and backup plan from existing systems to the new LIMS
 - Plans for user training and support
 - Plan for ongoing maintenance and system updates

Prerequisites, Capabilities, and Roles

- A LIMS implementation should provide as many of the following technical capabilities as possible (but start even if you don't have them all!):
 - Centralized automated collection and storage of research data and metadata
 - Data and metadata collection integrate with research workflows and project management (may include ELN for metadata collection)
 - Interfaces with laboratory scheduler and laboratory management software
 - Makes use of a handle server (or alternative) for assigning persistent identifiers
 - Supports creation of derivative data, either directly or by integration with visualization, analysis, and data evaluation tools
 - Data and metadata searchable for retrieval, analysis and metrics
 - Interoperability with other systems on data and metadata levels to support inter-laboratory networked LIMS systems and data exchange

Prerequisites, Capabilities, and Roles

- Referencing the NIST Research Data Framework (RDaF) [3], identify relevant stakeholders and their roles pertaining to the implementation and use of LIMS:

| Stakeholders |
|-------------------|
| Researcher |
| Data Manager |
| Facility Manager |
| IT Manager |
| Instrument Vendor |

| Topics | Subtopics |
|-----------------------|------------------------------------|
| Planning | Data management expertise/planning |
| | Data/Metadata considerations |
| Generation | Generated Experimental Data |
| | Community-based Standards |
| | Acquisition Software |
| Processing / Analysis | Types of processed data |
| | Instrument outputs |



[3] NIST Research Data Framework (RDaF): Version 1.5. NIST SP 1500-18r1. Robert Hanisch, *et al.* (2023). <https://doi.org/10.6028/NIST.SP.1500-18r1>

Prerequisites, Capabilities, and Roles

- These recommendations will be presented as a series of “checklists” or “blueprints” for an group, department, or organization
- By providing a concise and actionable list of “things to do/have”, we hope to make LIMS more accessible throughout the materials research community

Schema / Metadata

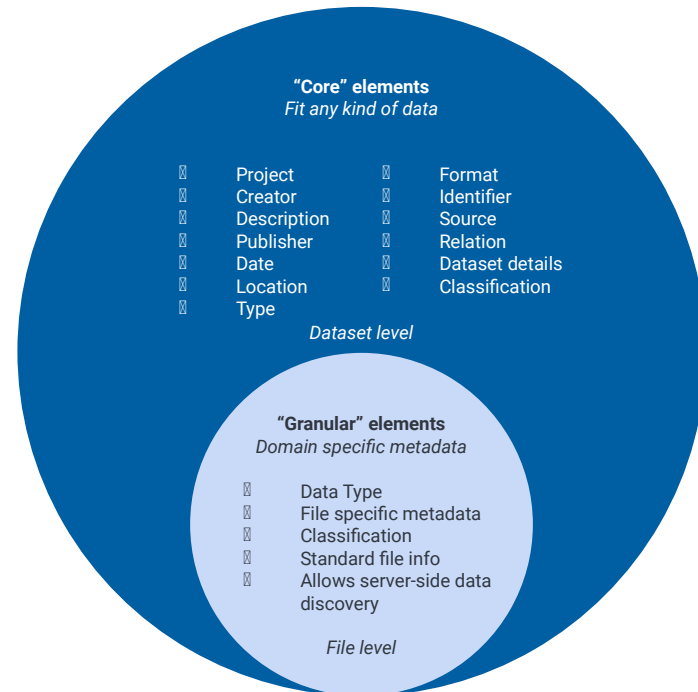
- Goal of this group is to identify, analyze, and recommend (if possible) relevant schemas describing *what* metadata should be collected and stored in a LIMS
- Schemas examined:

| | | |
|---|--|---|
| NIST Nexus Experiment (10.18434/M32245) | Materials Data Facility  (materials-data-facility/data-schemas) | Dublin Core Metadata Initiative (dublincore.org) |
| Foundry-ML (10.5281/zenodo.10884279) | Figshare (figshare.com) | Schema.org (schema.org) |
| Open Science Framework (10.17605/OSF.IO/8YCZR) | Sandia National Laboratories Ecosystem for Open Science (DCAT-eOS-AP) (10.2172/1777073) | Data Catalog Vocabulary (DCAT) (DOI-DO/dcat-us)  |
| PIDInst (instrumentation) (https://www.pidinst.org) | IGSN (sample identification) (https://igsn.org) | ORCID (people identification) (https://orcid.org) |

Persistent identifiers

Schema / Metadata

- Initial recommendations:
 - Need to capture the who, what, when, where (and ideally why)
 - Enough distinct metadata schemas exist to generally describe experimental datasets
 - The “core organizing unit” should be a Dataset, which consists of one or more files
 - Metadata can be defined at the dataset and file levels, where allowable metadata can change depending on file
 - An “extensible” model with minimal *required* metadata provides most utility to widest group
 - Use existing community standards where possible for things such as *instruments, samples, people/organizations*



[4] Example of extensible metadata structure from DCAT-eOS-AP.
<https://doi.org/10.2172/1777073>

How to learn more

MaRDA Working
Groups Website



<https://www.marda-alliance.org/working-groups/>

Draft Publication of
Recommendations Online



June 2024

Published Article
describing WG Results



Late 2024