Lessons Learned Evaluating and Deploying Electronic Laboratory Notebooks at NIST

Joshua A. Taillon

PlantMicrobe Seminar

Tuesday, Sept 12, 2023 - 13:00 CEST

NIST MATERIAL MEASUREMENT LABORATORY

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.

Outline

- Introductions
- What is an ELN and why would I want one?
- Two case-studies:
  - Early forays into ELNs at NIST
  - Implementing a centralized ELN server system

About NIST

- NIST is a U.S. federal research lab part of the U.S. Department of Commerce
  - The United States' National Metrology Institute (NMI)
  - Founded in 1901 as the National Bureau of Standards
  - Budget of around $1.25B
- Primary mission is to advance innovation and industrial competitiveness
- We improve the nation's measurement system and develop standards
- NIST scientists and engineers are often the world's experts in specific measuring fields

NIST at a Glance

"Industry's National Laboratory"

3,400+
FEDERAL EMPLOYEES

5
NOBEL PRIZES

2 CAMPUSES
GAITHERSBURG, MD [HQ]
BOULDER, CO

3,500+
ASSOCIATES

10
COLLABORATIVE INSTITUTES

400+
BUSINESSES USING NIST FACILITIES

ManufacturingUSA

16
NATL OFFICE FOR MANUFACTURING INSTITUTES

51
MANUFACTURING EXTENSION PARTNERSHIP CENTERS

U.S. BALDRIGE PERFORMANCE EXCELLENCE PROGRAM

Calibrations, Reference, and Measurement Service

Million-Pound Deadweight Machine

Credit: NIST

1,100 Standard Reference Material (SRM) products

100+ Standard Reference Data (SRD) products

550 measurement services

Every year:

32,000 SRM units sold

13,000 calibrations and tests

650 accreditations of testing and calibrations of laboratories

20,000 SRD products downloaded or purchased

Strategic Priorities and National Impacts

9/11/2001

WTC Towers Collapse

5/22/2011

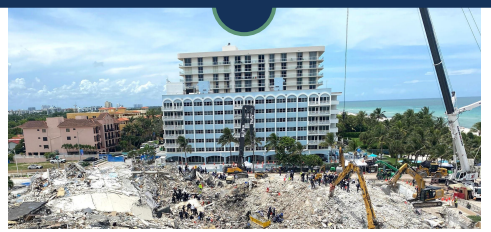
Joplin, MO Tornado

6/23/2012

Waldo Canyon Wildfire

Underway:

Technical investigation of effects of Hurricane Maria on Puerto Rico



Underway:

Technical investigation of cause of Champlain Towers collapse

Infrastructure Resilience

40+ NIST-led investigations of disaster and failure events since 1969

Resulting in **>40** significant changes to (inter)national building codes and design guidelines

About NIST

Seven locations around the United States:

Boulder, CO

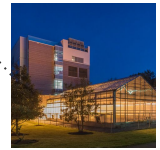
Center for Hierarchical Materials Design

Gaithersburg, MD

Brookhaven National Lab

Charleston, SC

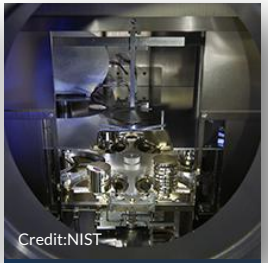
Center for Marine Debris Research



Institute for Bioscience and Biotechnology Research

NIST Laboratory Programs

Material Measurement Laboratory



Physical Measurement Laboratory

Engineering Laboratory

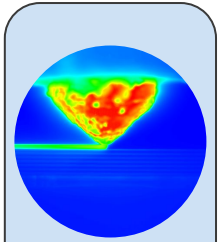
Information Technology Laboratory

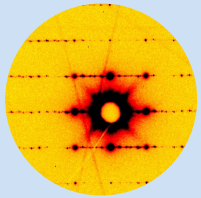
Communication Technology Laboratory

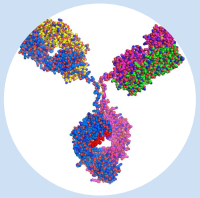
NIST Center for Neutron Research

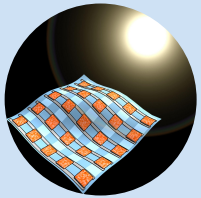
About MML (Material Measurement Laboratory)

- MML is one of the six laboratory organizations within NIST
- MML performs applied measurement science and provides measurement services across many program areas:


Advanced Manufacturing


Advanced Materials


Biomedical and health care


Energy


Environment and climate

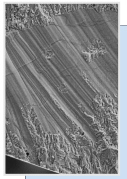
Food safety and nutrition


Physical infrastructure

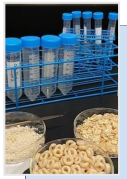

Safety, security, and forensics

Data and Informatics | Artificial Intelligence and Machine Learning

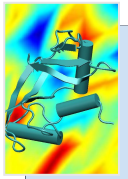
About MML (Material Measurement Laboratory)



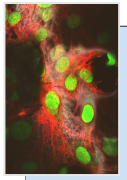
Applied Chemicals and Materials



Chemical Sciences

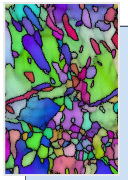


Biomolecular Measurement

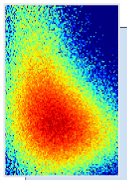


Biosystems and Biomaterials

Materials Measurement Science



Materials Science and Engineering



Office of Data and Informatics

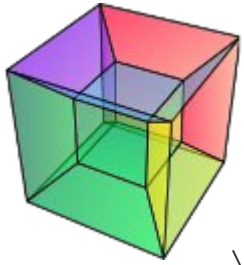
Office of Reference Materials

Office of Data and Informatics (ODI)

- Provides leadership and expertise to meet the data challenges of modern research
- ODI works with the research divisions to provide solutions and recommendations for FAIR data
- Staffed by data scientists that are also subject matter experts within the program areas of MML

About Me!

Academic background in Materials Science & Characterization



Hyperspy

python™

General scientific programming

SQLite

Databases (SQL)

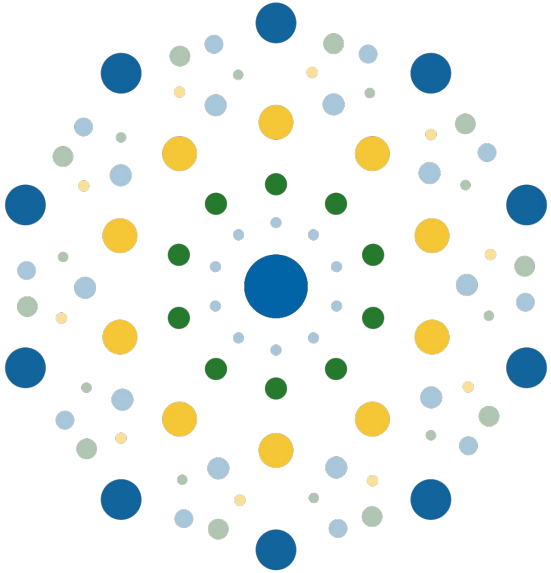
plotly | Dash

Dashboarding/app design

Findable Accessible Interoperable Reusable

An example of what I do at NIST

- Starting in 2019, we've worked with a multi-user electron microscopy facility at NIST to address data needs
- Built a microscopy LIMS mostly from scratch
  - Open-sourced at https://github.com/usnistgov/NexusLIMS
  - DOI: 10.18434/mds2-2355
  - Described in detail in *Microscopy and Microanalysis*, 27 (3), 2021. pp. 511 - 527. 10.1017/S1431927621000222



What is an ELN and why would I want one?

First, a paper recommendation

A great review of ELN history, considerations, and implementation suggestions (from Imperial College, London):

> Stuart Higgins, Akemi Nogiwa-Valdez, & Molly Stevens.
> "Considerations for implementing electronic laboratory notebooks in an academic research environment."
> *Nature Protocols* (Vol. 17, Issue 2, pp. 179–189), 2022.
> https://doi.org/10.1038/s41596-021-00645-8

Many of my points today are informed by this groups' wonderful paper and their data published on Zenodo: https://doi.org/10.5281/ZENODO.5012729

What is an ELN?

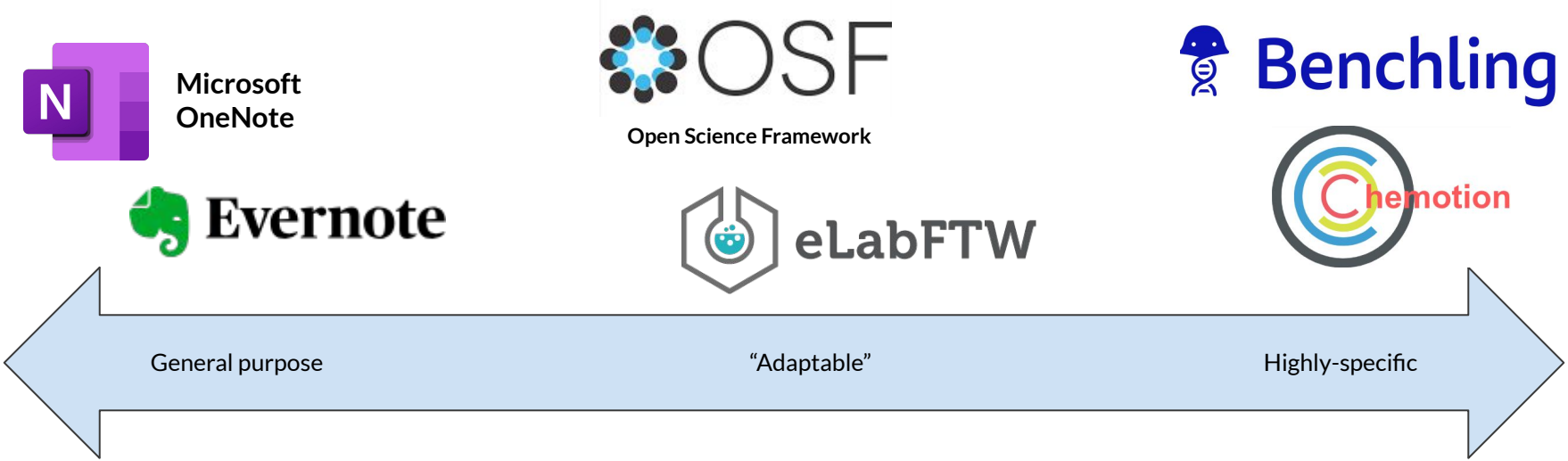
- Fundamentally, a digital version of the lab notes we are all used to taking
- Contains technical and general (meta)data about experiments
- Digitization at the time of creation, rather than *afterwards*

Fig. 1. Examples of early lab notebooks: Leonardo da Vinci's Vitruvian Man illustration, Thomas Jefferson's portable/re-writable ivory leaf note pad, Darwin's first rough sketch of the evolutionary tree concept, and his illustrations of Galapagos Finches.

Richard Gates, Mark McLean, & William Osborn, *J Res. NIST*, *120*, 293 (2015). https://doi.org/10.6028/jres.120.018

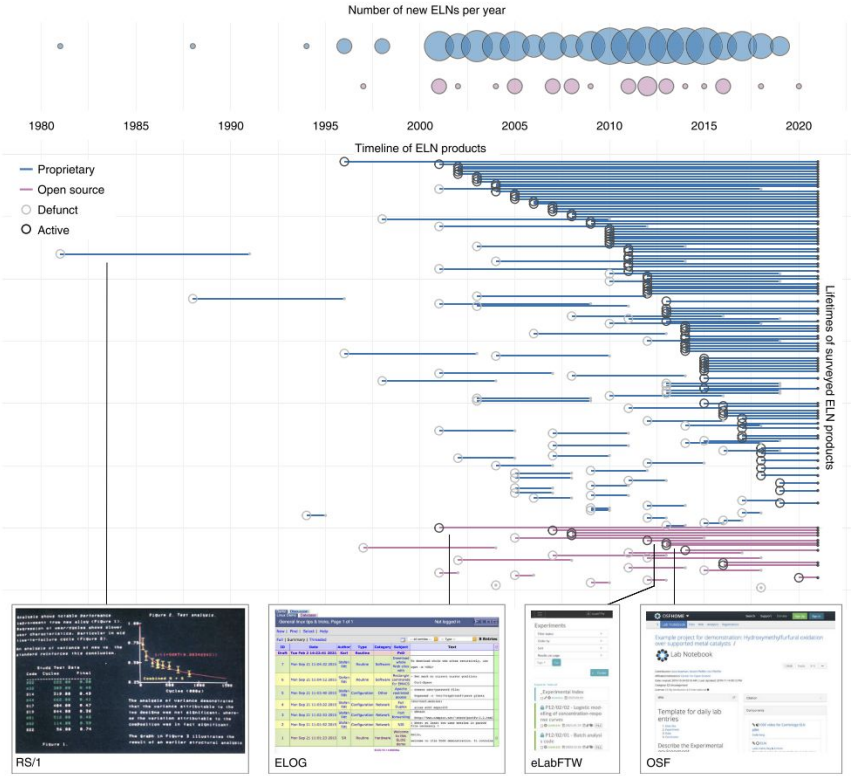
What sorts of ELNs are available?

Intended uses range from general-purpose note-taking to highly-specific tools for specialized applications



What sorts of ELNs are available?

- Higgins *et al.* (2022) identified 96 active and 76 defunct ELN software packages
- Open-source and commercial systems are both available
- Average lifetime of 7 ± 4.4 years
  - Open-source: 6 ± 4.4 years (*n* = 25)
  - Commercial: 7 ± 4.4 years (*n* = 147)



Benefits of ELNs vs. physical notebooks (1)

Electronic Notebooks

- Information can be found and shared easily
- Entries can be accessible across project teams
- References to other entries can be dynamically linked
- Typically provide advances querying tools
- Easy “templating” of entries to control what information is recorded – reproducibility!

Physical Notebooks

- Information is only accessible in one place (the actual notebook)
- Project teams have to share one single notebook in the lab
- References are one-time static entries typically by book and page numbers
- Cannot search other than manual browsing by chronology
- Requires custom printing or some other labor-intensive way to “template”

Benefits of ELNs vs. physical notebooks (2)

Electronic Notebooks

- Easy to add non-text or computer-generated content to an entry
- Legibility does not depend on handwriting
- Legal record (for patents, etc.) on-demand via digital signatures/timestamping
- Backing up is a simple database backup
- If desired, open science/data is possible with many systems
- Easy compliance with open data mandates

Physical Notebooks

- Requires printing and gluing into a physical notebook to add photos, tables, etc.
- Researchers may not have ideal penmanship
- Legal record requires multiple sign-offs and secure physical storage of notebooks
- Takes a long time to scan whole notebooks
- This would be impossible with physical notebooks

Drawbacks of ELNs vs. physical notebooks

Electronic Notebooks

- Adding entries requires opening computer, logging in, etc. – PPE can make this harder
- Inputting drawing, sketches, or equations can be difficult
- Potential distraction from requiring a computing device
- Potential (more so with proprietary applications) for lock-in or inaccessible data

Physical Notebooks

- Writing in a paper notebook is trivial
- Drawing is as easy as writing
- Single-use device good for focusing on the tasks at hand
- A paper notebook has no vendor lock-in

Early pilots of ELNs at NIST

Initial effort for ELNs within MML

- Started as a “wish list” in 2013 by a microfabrication research group
- Formalized into a list of desired capabilities
- Evaluated both various hardware and software platforms:
  - Tablet computers (Microsoft Surface)
  - Ultrabook laptops
  - Hybrid detachable computers
  - Microsoft OneNote
  - Evernote
- Documented experience in a 2015 manuscript:

> Richard Gates, Mark McLean, & William Osborn, J Res. NIST, 120, 293 (2015). https://doi.org/10.6028/jres.120.018

Guiding concepts identified for ELN at NIST

Guiding Concept	Capability	Details
Portable	Take with you anywhere	• Lightweight hardware • Long battery life
Intuitive	Touchscreen hardware. Familiar OS and application interface	• Graphical interface OS • Notebooks/Folders/Pages format
Clean	Cleanroom compatible	• Easily wiped down for cleanroom entry
Creative	Flexible and easy to use	• Multiple formats: • Text, images, audio, video, handwriting, etc.
Collaborative	Resource sharing	• Variable control for “circle of trust”
Live	Information available anywhere at anytime	• Wireless and Cloud
Smart	Features not normally associated with a lab notebook	• Camera, microphone, communications, sensors, web access, tagging, and searchable indexing
Secure	NIST IT security constraints	• Password protection, disk encryption, firewall, domain authentication, antivirus

Adapted from: Richard Gates, Mark McLean, & William Osborn, *J Res. NIST*, *120*, 293 (2015).

The Pilot Solution

Microsoft OneNote on Microsoft Surface Pro computer:

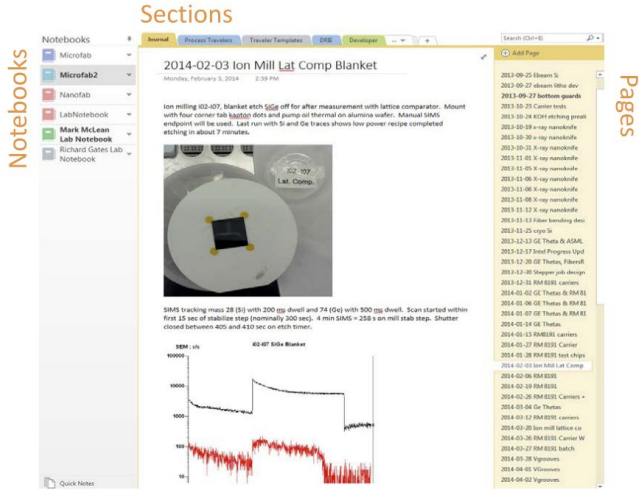


Fig. 3. OneNote organizational layout of the nested Notebook/Section/Page hierarchy with the contents of the selected page displayed in the center.

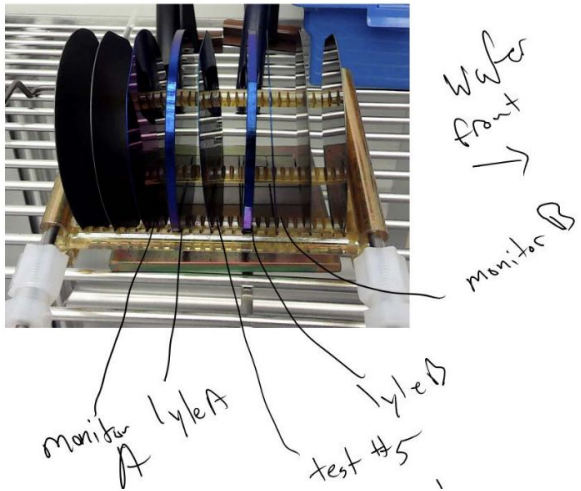


Fig. 4. Typical experiment documentation capability of the SELN combining the rear facing camera and hand annotation.

State of ELNs at NIST in 2022

- Rolled out this framework to ~100 researchers with demo notebooks and training
- No absolute numbers to date, but uptake estimated in the 5 - 10 % range
- OneNote solution works very well for those that use it, but has some key limitations….
    - Free-form notes do not allow “advanced” structuring of data, or querying that data
    - Notes are effectively “locked in” to OneNote format
    - Difficult to built “add-ons” for automated data workflows
    - Desktop application is not cross-platform (web-based version available, but limited)

Implementing a centralized ELN system

Revisiting Modern ELNs for NIST Researchers

Prior to July 2022

Initial Discussions

Limitations of OneNote identified. Researchers want an easy-to-use "database" without a formal DB

Sept. 2022

Presentation about eLabFTW to LIMS COI

Clemens Mangler (University of Vienna) presents to the MML LIMS COI about his laboratory's use of eLabFTW

Oct. 2022

Initial Use Cases Identified

Two research groups in MML's Biosystems and Biomaterials Division agree to be "guinea pigs"

2023

Pilot Implementation Underway

eLabFTW deployed on local infrastructure, iteration on templates and database types, etc.

Updated Requirements

- The OneNote solution was too general, but domain-specific tools do not allow interconnection and are a heavy administrative burden
- The “ideal” ELN thus sits in the center of the “generic – specific” spectrum
- In addition to everything from the earlier pilot, an ideal ELN must *also provide*:
  - Representation of arbitrary data types (samples, instruments, etc.) and a straightforward way for researchers (not data scientists) to create these
  - Linking of note entries to build a “knowledge graph”
  - Support for both “free text” notes, diagrams, etc., as well as structured metadata
  - Application Programming Interface (API) for advanced use cases

Trying out ELabFTW at NIST

- After discussions with existing users, a pilot was launched to evaluate ELabFTW
- What is ELabFTW?
  - A web-based, self-hosted electronic laboratory notebook platform
  - At its core, tracks experiments and manages lab information for many teams with a flexible database model

Reminder 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.

https://elabftw.net

ELabFTW Capabilities

- Secure, cross-platform, responsive design from any device
- Can handle independent "teams" (more on this later)
- Flexible and linkable research notes (more on this later)
- Simple to maintain database of "items" or “resources”
- Completely API-accessible for advanced use cases
- Enterprise-integrated login (SAML)

How does it work?

- Users are separated into one or more “teams”
- Users can create “experiments” based off templates
  - “Experiments” correlate to pages in a notebook, but everything in the app is linkable
  - Much like OneNote: add arbitrary text, attach files, drag-and-drop images, scribble drawings, etc.
  - Templates can be individual to a user, or shared among a team

How does it work? (continued)

- The "database" or "inventory" provides a flexible but powerful way to represent your particular data model
  - Represent samples, people, instruments, procedures, etc.
  - There are no limitations on types, so any model is possible
- Types are unique to each individual team
  - This means different research groups can have different models

Example use case 1: Microbial Strain Measurements

- The NIST Microbial Team characterizes many different microbes with multiple methods
- They need to:
  - Document characterizations
  - Synthesize results for decision making
  - Provide characterization results to external stakeholders via another API-enabled web tool
  - Have many staff using multiple shared protocols for experiments

Example use case 1: Microbial Strain Measurements

Experimental workflows

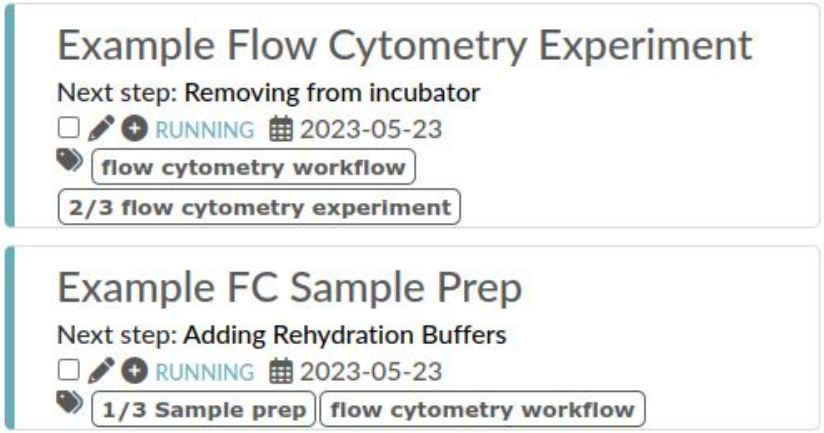
- Flow cytometry (multistage):
  - 1/3 - Sample prep
  - 2/3 - Cytometry
  - 3/3 - Analysis
- Sample handling
- CFU (colony forming unit) measurement

Database Items:

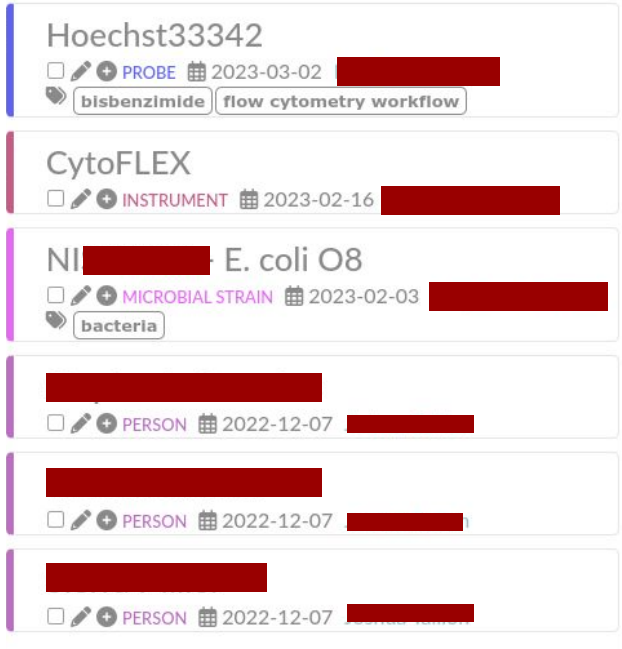
- Instruments
- Staff
- Microbial Strains
- Manufactured Materials
- Probes
- "Study ID"
- Manufacturers
- and more...

Example use case 1: Microbial Strain Measurements

Experimental workflows

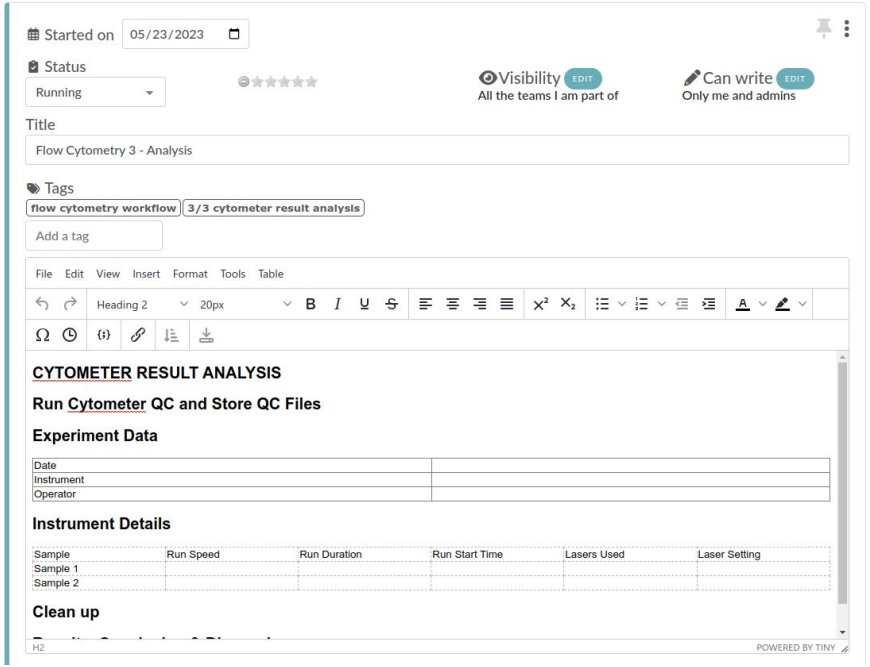


Database Items:

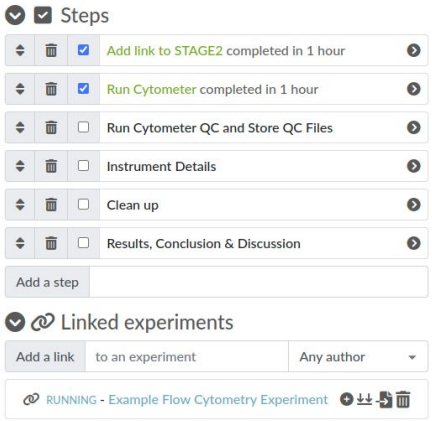


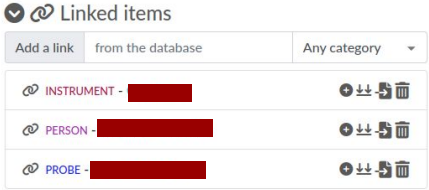
Example use case 1: Microbial Strain Measurements

Creating a new experiment from template



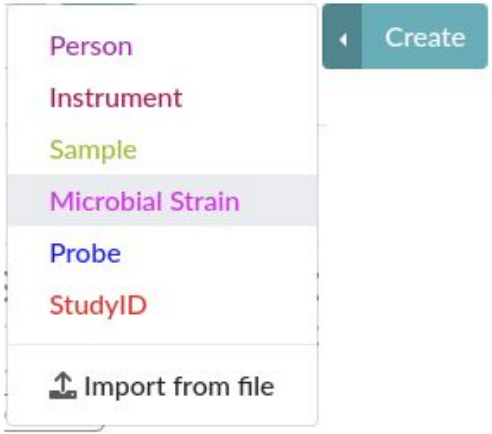
Workflow tracking and linking





Example use case 1: Microbial Strain Measurements

Creating new items with structured metadata:



Genus

Strain

species

Gram Strain

Positive Negative

Growth Media

The medium this strain was grown in

THA
THB
LB
BHI

Strain Number

Biosafety Level

1 2

Alternative Name

Growth Atmosphere

Aerobic
Anaerobic
Microaerophilic/Capnophilic (5% CO2)

Additional Information

Original Isolation Source

Example use case 1: Microbial Strain Measurements

Automated ingest of existing spreadsheets via API:

	A	B	C	D	E	F	G
1	Strain number	Genus	species	Strain	Alternative_name	Original isolation source	Growth Media
2	[redacted]	*Escherichia*	*coli*	O8	7.1994	[redacted]	THA/THB
3		*Escherichia*	*coli*	EHEC O11:H8	DEC8B		LB/BHI/TSA
4		*Escherichia*	*coli*	STEC x03:NM	90-1787		LB/BHI/TSA
5		*Escherichia*	*coli*	UPEC O6:H31	536		LB/BHI/TSA
6		*Escherichia*	*coli*	ETEC			LB/BHI/TSA
7		*Escherichia*	*coli*	ETEC	CG146		LB/BHI/TSA
8		*Escherichia*	*coli*	EHEC O157:H7	33		LB/BHI/TSA
9		*Escherichia*	*coli*	EPEC 0127:H6	E2348/69		LB/BHI/TSA
10		*Escherichia*	*coli*	EHEC O103:H2	SJ10		LB/BHI/TSA
11		*Escherichia*	*coli*	EHEC O121:H19	SJ18		LB/BHI/TSA
12		*Brenneria*	*nigrifluens*		SAFE118/ 1391		TSA
13		*Delftia*	*acidovorans*		B-769		TSA
14		*Dienococcus*	*radiodurans*	R1			TGY/R2A/TSA
15		*Lactobacillus*	*rhamnosus*				MRS/Blood
16		*Bacteroides*	*fragilis*	B-23622			Blood
17							

```python
def create_item_from_row(row, category_id):
    """
    Using the ELabFTW API, create an item in the database with a given set of metadata
    values specified by a row of the provided CSV file. This will require three steps:
    (1) a POST request to create the new item and get it's ID number, (2) a GET request
    to obtain the default metadata structure (as specified in the item type definition),
    and (3) a PATCH request to update the title and metadata of the newly created item
    according to what is provided in the metadata ``row`` value.

    This method does not need to return anything, since it only exists to create new
    items.

    Parameters
    ----------
    row : dict
        A dictionary representing a row of the CSV file (read by ``csvreader``) that
        contains the metadata to be inserted in this item
    category_id : int
        The category (type) of item to create, as an integer. e.g. item type
        "Microbial Strain" would be ``32``
    """
    # post request to create a new item and get its ID number
    res = requests.post(
        f"{API_HOST_URL}/items",
        json={"category_id": category_id},
        headers=headers,
        verify=False
    )
    item_id= int(res.headers["Location"].split("/")[-1])
```

Example use case 2: Cell Provenance

- NIST Cell Systems Science Group
- Cells are self-replicating and present unique measurement challenges
- They need to:
  - Demonstrate quality in cell expansion process
  - Log of process control measurements and specifications
  - Provide results to external stakeholders via another API-enabled web tool

Example use case 2: Cell Provenance

Experimental workflows

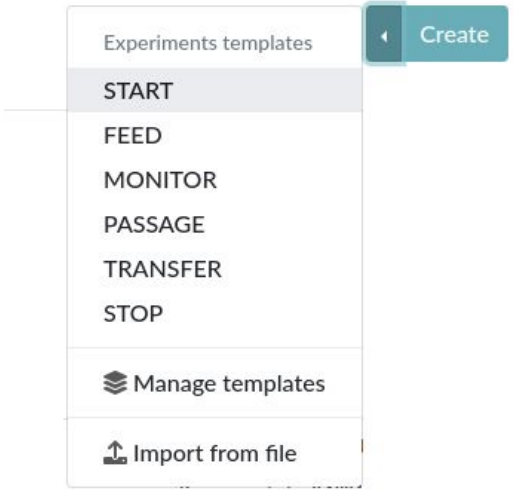
- Expansion protocol:
  - Start
  - Feed
  - Monitor
  - Passage
  - Transfer
  - Stop
- Each protocol needs its own interface (template) so consistent information is collected at every stage

Database Items:

- Instrument
- Operator
- Mammalian Cell
- Container
- Medium
- Serum
- Reagent
- and more...

Example use case 2: Cell Provenance

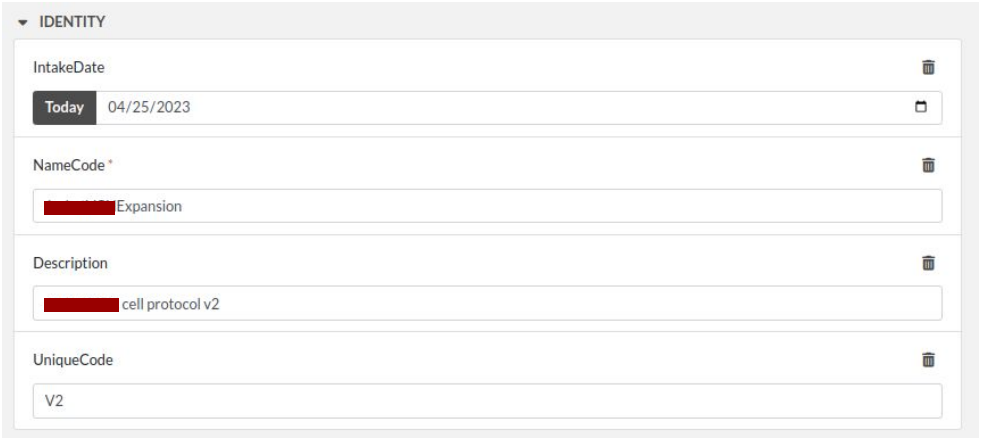
Experimental workflows

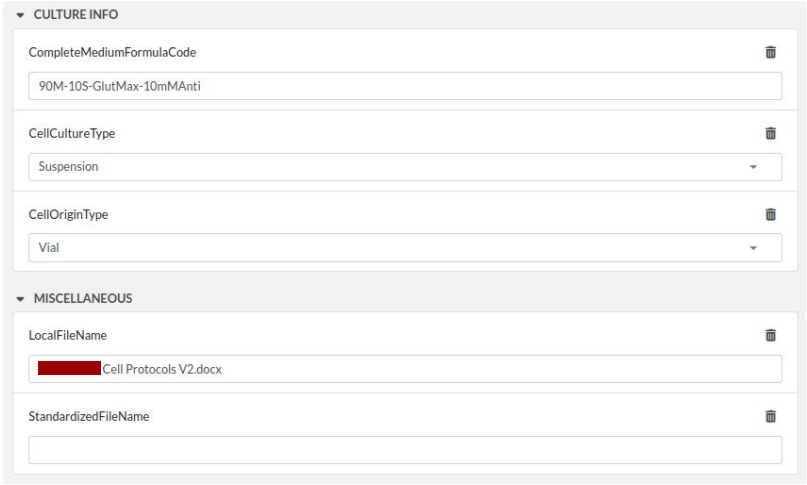


Database Items:

Example use case 2: Cell Provenance

Heavy use of structured metadata to represent their data model:





Pilot management model

- ODI administers single instance for MML
- A group, division, or project can register a "team"
- ODI trains a one or more representative(s) on team administration and basic data modeling/workflows
- Division/project level admins administer their own users

Conclusions

Where are we now?

- "Pilot" use cases are just starting real data collection after initial model development during the summer
- Received IT security clearance to run on NIST's internal network
- Currently working to train users and receive feedback on initial operations

Lessons learned

- There is no “one-size-fits” all solution in multi-disciplinary research environments
- The ease with which data models can be customized to individual or shared workflows is very important
- Specific is better than generic, but not too specific!
- Successful implementation of an ELN requires both technical and research champions

Thank you for your attention!
Questions?

joshua.taillon@nist.gov
https://orcid.org/0000-0002-5185-4503

Use case: NexusLIMS (multi-division)

- NexusLIMS is fairly mature, but has no editing capabilities and zero ability to capture information from a user during an experiment
- User workflows are extremely varied and tool/methodology specific
- Investigating eLab as more general note-taking tool and for linkages to auto-generated research records

Use case: NexusLIMS (multi-division)

More use of free-text notes:

Owned by [redacted]

Experiment was timestamped by [redacted] on 2023-04-26 at 11:43:10

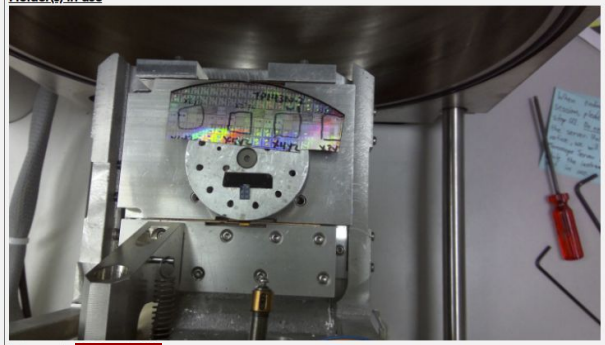
Success ★★★★★
Started on 2023-04-24

Visibility: All the teams I am part of
Can write: Only me and admins

NexusLIMS-WIPP project

Sample(s)
Old MRAM sample from [redacted]
Old nanodot samples on Hall crosses
Sample - XYZ

Holder(s) in use



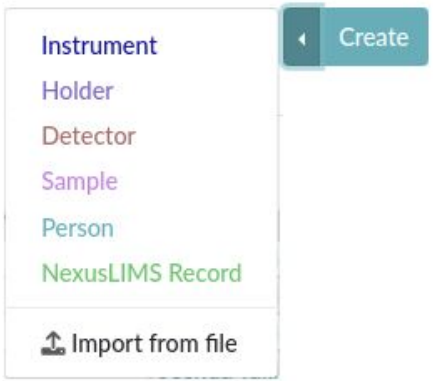
Instrument - [redacted]

Experiment Plan
Take high, medium low mag images in SE mode for these two samples. The images will re recorded by NexusLIMS, ITL team will try to do image registration.

Notable microscope/peripheral settings:

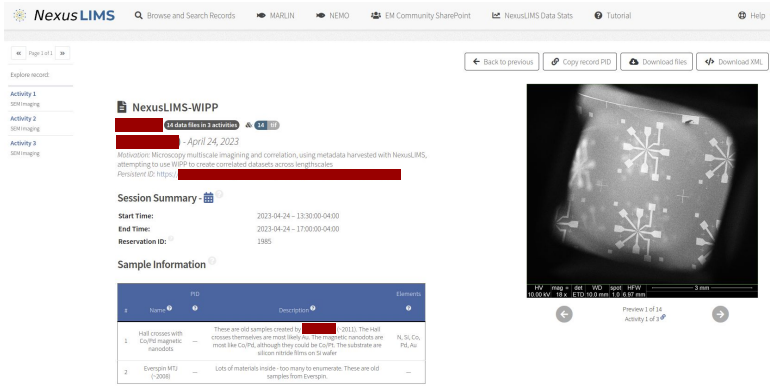
Use case: NexusLIMS (multi-division)

Database items of Instruments, Holders, Detectors, etc.
(eventually with PIDs)



Use case: NexusLIMS (multi-division)

Representation of auto-generated NexusLIMS Records
Uploaded via API at same time records are created in CDCS



Interactive notes can be linked to eLab record, which links to NexusLIMS CDCS for full context