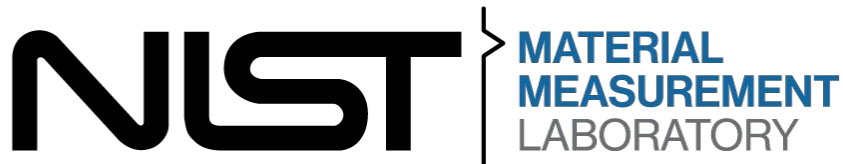


Overview of NexusLIMS Usage and Development

Joshua A. Taillon

Northwestern University Meeting

Monday, May 1, 2023



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.

What do we mean by LIMS?

LIMS:

Laboratory Information Management System

Ideally start at the bottom of the pyramid, but scientific value comes at the top

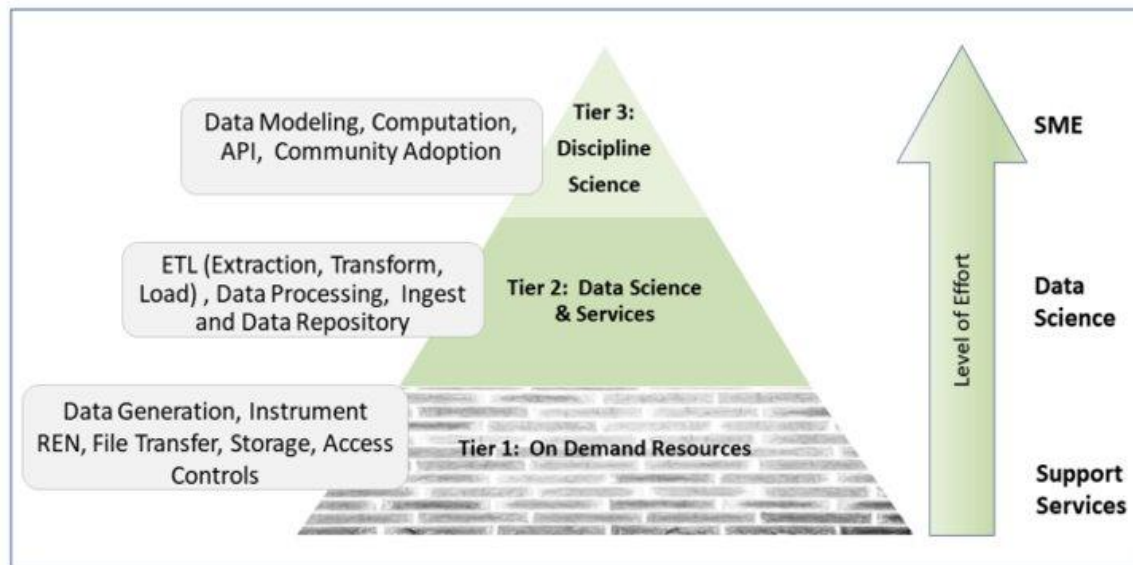


Fig. 1. LIMS three tiered model for implementation

NIST Technical Note 2216 - <https://doi.org/10.6028/NIST.TN.2216>

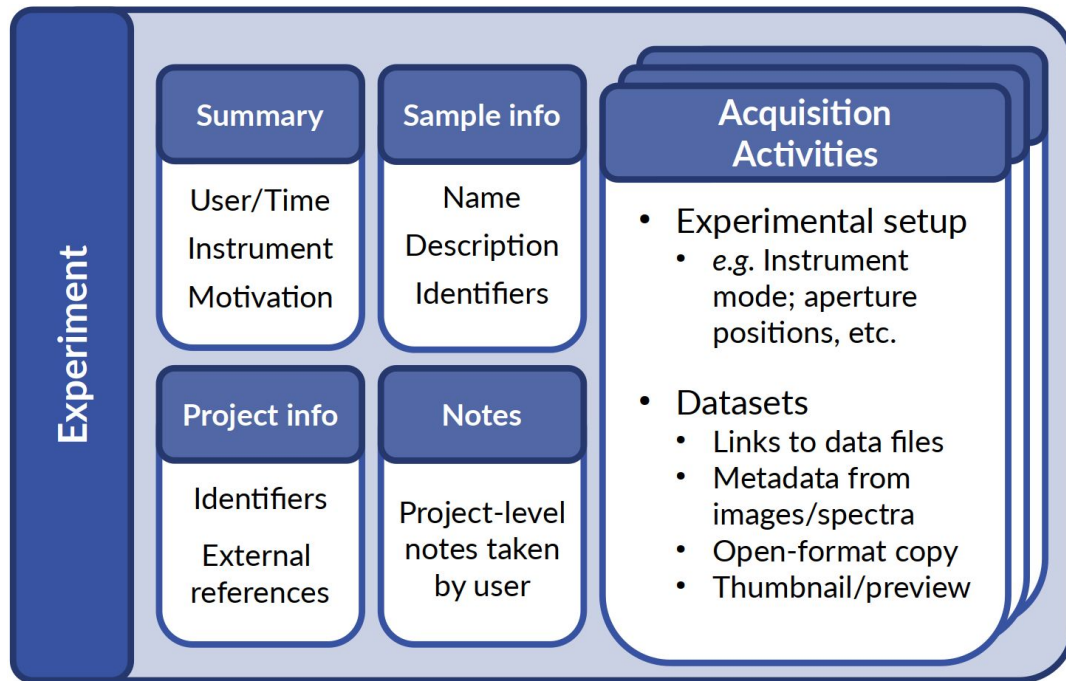
NexusLIMS attempted to build from scratch

- Prior to community efforts (ca. 2018), we wanted to solve these issues for our shared microscopy facility
- Built NexusLIMS, a microscopy LIMS mostly from scratch
 - Open-sourced at <https://github.com/usnistgov/NexusLIMS>
 - DOI: [10.18434/mds2-2355](https://doi.org/10.18434/mds2-2355)
 - Described in detail in *Microscopy and Microanalysis*, 27 (3), 2021. pp. 511 - 527. [10.1017/S1431927621000222](https://doi.org/10.1017/S1431927621000222)



Mapping EM workflows into a data model

- Data is most useful when intelligently structured
 - Allows browsing, querying, transforming, validating, etc.
- Structure should be tailored to context
 - What information could a researcher/manager/auditor want to see?
- A “record” represents an individual experimental session on microscope
- Schema published at <https://doi.org/10.18434/M32245>



J.. Taillon, et al., *Microscopy and Microanalysis*, vol. 25, no. S2, pp. 140–141, 2019.

What does it look like for users?

These slides (6-10) show features of NEMO/MARLIN, our facility management system

When a user wants to use a tool, they log in to “MARLIN” and make a reservation:

The screenshot displays the MARLIN interface. At the top, there are navigation tabs: MARLIN, Calendar, Tool control, Status dashboard, and Administration. Below these are buttons for 'Reservations', 'Today', 'Reserve for someone else', and 'Schedule an outage'. A search bar is labeled 'Search for a tool or an area'. On the left, there are sections for 'Personal schedule', 'Areas', and 'Tools'. Under 'Tools', a list includes Boulder, Gaithersburg, Bio, Chemistry, Demonstration, Demonstration Tool (highlighted), and Materials Science. The main area is a calendar grid with columns for Sun 4/30, Mon 5/01, and Tue 5/02, and rows for 7am, 8am, 9am, 10am, 11am, and 12pm. A blue reservation block is visible on Mon 5/01 from 10:45 to 11:45, labeled 'Joshua Tallon (jat)'. A yellow shaded area covers the entire Mon 5/01 column.

See <https://github.com/usnistgov/NEMO>

Select a tool, then click and drag to reserve a time

What does it look like for users?

These slides (6-10) show features of NEMO/MARLIN, our facility management system

When a user wants to use a tool, they log in to “MARLIN” and make a reservation:

The screenshot shows a web browser window with the title 'Reservation questions'. The form contains the following fields and options:

- Project ID:** A text input field with a red border. Below it, examples are listed: 'Additive Manufacturing, 642.01.03'.
- Title of Experiment: *** A text input field with a red border.
- Experiment Purpose: *** A larger text input field with a red border.
- Curate this session's data using NexusLIMS:** Two radio buttons: 'Agree' (selected) and 'Disagree'.
- Sample information:** A link to 'Explore other LIMS / Specimen Repositories:'.
- Sample:** A dropdown menu showing 'NCAL: Mi' with a magnifying glass icon.
- Below the dropdown, a note says: 'Enter information for at least 1 sample below, and click the "Add" button to add up to 5 samples.'

The screenshot shows a web browser window with the title 'Sample Name / PID: *'. The form contains the following fields and options:

- Sample Name / PID: *** A text input field with a red border.
- Below the field, a note says: 'Write in a sample name, or paste a persistent identifier (PID) for your specimen from one of the LIMS/repositories above.'
- Is this a sample name or a PID? *** Two radio buttons: 'Sample Name' and 'PID'.
- Sample Details:** A text input field with a red border.
- What elements do you expect to be in your specimen? *** A dropdown menu that is expanded to show a periodic table of elements. The table is color-coded by groups.
- Below the table, there is an 'Add' button.
- At the bottom of the form, a note says: 'Please answer the required questions (above) to proceed'.

User fills out basic metadata about their experiment at reservation time

What does it look like for users?

These slides (6-10) show features of NEMO/MARLIN, our facility management system

When it comes time to use a tool, they log in to “MARLIN” and enable it:

The screenshot displays the MARLIN web interface. At the top, there is a navigation bar with links for 'MARLIN', 'Calendar', 'Tool control', 'Status dashboard', and 'Administration'. Below this is a search bar labeled 'Search for a tool' and a list of categories: Boulder, Gaithersburg, Bio, Chemistry, Demonstration, **Demonstration Tool** (highlighted), and Materials Science. The main content area is titled 'Demonstration Tool' and includes buttons for 'Summary', 'Details', 'Report a problem', and 'Post a comment'. A large green checkmark and text state: 'This tool is operational and idle.' Below this, a section titled 'What would you like to do?' contains three radio button options: 'Use this tool for my own project' (selected), 'Use this tool on behalf of another user', and 'Use this tool for a remote project'. A note indicates 'Tool usage will be billed to the project named "641"'. At the bottom, a green button labeled 'Start using the Demonstration Tool' is highlighted with a red arrow.

What does it look like for users?

These slides (6-10) show features of NEMO/MARLIN, our facility management system

When they are done, they log in to “MARLIN” and disable it:

MARLIN Calendar Tool control Status dashboard Administration ▾ Welcome, Joshua

Search for a tool

Boulder
Gaithersburg
Bio
Chemistry
Demonstration
Demonstration Tool
Materials Science

Demonstration Tool Summary Details Report a problem Post a comment

You are using this tool for the project named 641 since Monday @ 9:37 AM.

■ Stop using the Demonstration Tool

Access control (if you want)

These slides (6-10) show features of NEMO/MARLIN, our facility management system

- Depending on desired level of control, NEMO/MARLIN can physically lock-out tools that are not enabled
- This can be done for billing
 - Could be used to ensure metadata entries are collected prior to tool use
 - May be otherwise useful for group/division management



Building of research records

More details in later slides ([slide 21](#)); but once a user finishes their session, data is copied automatically to centralized file storage and a “record” of that session (matching the schema from [slide 5](#)) is built automatically and loaded into the NexusLIMS web interface (next slide)

Web-based user interface is based off open-source [CDCS platform](#)

Querying the database - Searching for records

Free-text search for metadata terms

The screenshot shows the NexusLIMS search interface. At the top, there is a navigation bar with the NexusLIMS logo and several menu items: 'Browse and Search Records', 'MARLIN', 'NEMO', 'EM Community SharePoint', 'Tutorial', 'Help', and a user profile 'jat'. Below the navigation bar is a search bar containing the text 'EDS' and a 'Search' button. The search results are displayed below the search bar, showing 59 results. The results are listed in a table-like format with columns for checkboxes, record titles, dates, and actions. The records are:

- Tom's Wire #1 EDS** (JEOL JSM7100) 4 data files in 1 activity & 4 tif - July 19, 2022
William [redacted] - July 19, 2022
Motivation: EDS mapping of Cu wire cross-section
- STEM** (FEI Titan STEM) 149 data files in 6 activities & 149 dm3 - July 08, 2022
Huairuo [redacted] - July 08, 2022
Motivation: HAADF
- STEM** (FEI Titan STEM) 224 data files in 11 activities & 224 dm3 - June 28, 2022
Huairuo [redacted] - June 28, 2022
Motivation: HAADF
- slow cool** (JEOL JSM7100) 2 data files in 1 activity & 2 tif - June 21, 2022
Maureen [redacted] - June 21, 2022
Motivation: EDS
- EDS of Sn on Graphene** (FEI Quanta200) 4 data files in 1 activity & 4 tif - June 17, 2022
David [redacted] - June 17, 2022
Motivation: How much Sn is on Graphene!!!

The last record, 'EDS of Sn on Graphene', is highlighted with a blue border. At the bottom of the search results, there are several utility buttons: 'Sort', 'Share Query', 'Share PIDs', 'Download', and 'Date'.

Summary display of record contents

Querying the database - Searching for records

Further refinement allows for quickly finding record of interest

The screenshot displays the NexusLIMS search interface. At the top, the navigation bar includes the NexusLIMS logo, a search bar with the text "Browse and Search Records", and several utility icons: MARLIN, NEMO, Tutorial, Help, and a user profile icon labeled "jat". Below the navigation bar is a search input field containing the terms "EDS" and "david", with a blue "Search" button to the right. Below the search bar, it indicates "Found 4 Results:" and provides options for "Sort", "Share Query", "Share PIDs", "Download", and "Date".

The search results are listed as follows:

- EDS of Sn on Graphene** (FEI Quanta200) 4 data files in 1 activity & 4 tif
David [redacted] - June 17, 2022
Motivation: How much Sn is on Graphene!!!
June 17 2022 1:00PM
- EDS W, Ag post-echem** (FEI Quanta200) 25 data files in 2 activities & 25 tif
David [redacted] - May 07, 2021
Motivation: Morphology and species identification
May 07 2021 1:41PM
- EDS W, Ag post-echem** (FEI Quanta200) 3 data files in 1 activity & 3 tif
David [redacted] - May 07, 2021
Motivation: Morphology and species identification
May 07 2021 9:20AM
- W Ref** (FEI Quanta200) 6 data files in 2 activities & 6 tif
David [redacted] - December 10, 2020
Motivation: EDS
Dec. 10 2020 10:47AM

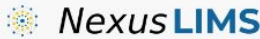





Viewing a record

Dataset
preview
gallery





Summary
info

Session/
reservation
info

Sample
and
project



 [Browse and Search Records](#)  MARLIN  NEMO  Tutorial  Help  jat


« Page 1 of 1 »


← Back to previous  Copy record PID  Download files  Download XML  Edit this record

Explore record:

Activity 1
SEM Imaging

 **EDS of Sn on Graphene**
FEI Quanta200 4 data files in 1 activity  4 **tif**

David  - June 17, 2022
Motivation: How much Sn is on Graphene!!!
Persistent ID: <https://hdl-i.nist.gov/8000/20.500.13060/b1df6f99-8350-40ae-a55d-e3d8d339830c>

Session Summary - 

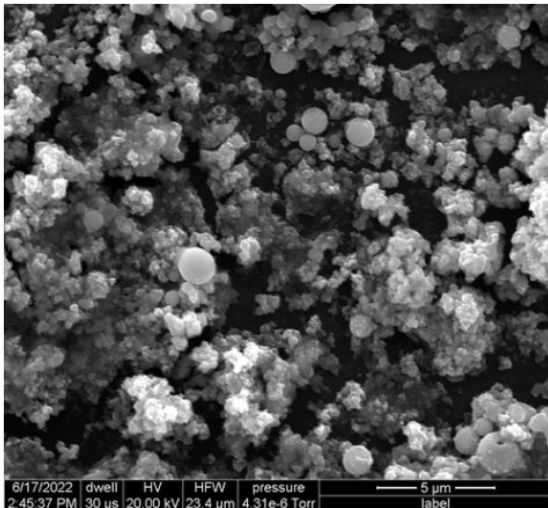
Date: 2022-06-17
Start Time: 14:00:00-04:00
End Time: 16:00:00-04:00
Reservation ID: 724

Sample Information

Name: Sn on Graphene

Project Information

ID: 642.03



6/17/2022 dwell HV HFW pressure
2:46:37 PM 30 µs 20.00 kV 23.4 µm 4.31e-6 Torr

5 µm
label

Dataset 4 of 4
Activity 1 of 1

File downloading via Web UI

Explore record:

Activity 1

SEM Imaging

EDS of Sn on Graphene

4 tif

Complete filelisting for:

EDS of Sn on Graphene - June 17, 2022

Root path: /Quanta/SA20220617_Au patterned Graphene Sn02/

Search:

1

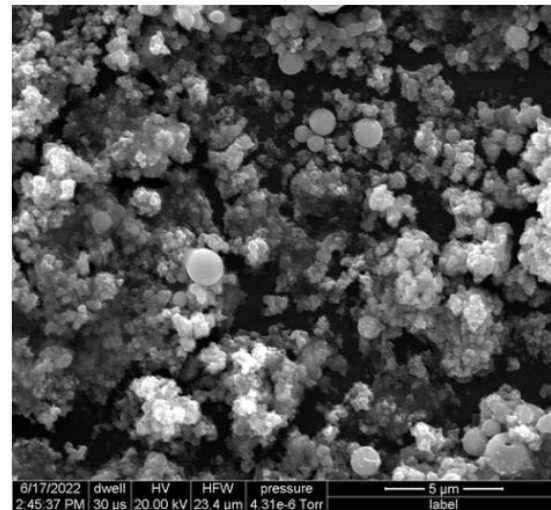
Select all Select none Download all as .zip Download selected as .zip

Copy CSV Excel Print

Total size of all datasets: 3.7 MiB.

Dataset Name	Path	Size	Type	Meta	D/L
<input type="checkbox"/> Region 1.tif	/	953.4 KiB	Image		
<input type="checkbox"/> Region 2.tif	/	953.4 KiB	Image		
<input type="checkbox"/> Region 3.tif	/	953.4 KiB	Image		
<input type="checkbox"/> Other chip region 1.tif	/	953.4 KiB	Image		

Showing 1 to 4 of 4 datasets



Dataset 4 of 4
Activity 1 of 1



Viewing individual file extracted metadata

Explore record:

Activity 1

SEM Imaging

EDS of Sn on Graphene

Complete filelisting for:

EDS of Sn on Graphene - June 17, 2022

Root path: /Quanta/SA20220617_Au patterned Graphene Sn02/

Search:

Select all Select none Download all as .zip Download selected as .zip

Copy CSV Excel Print

Total size of all datasets: 3.7 MiB.

Dataset Name	Path	Size	Type	Meta	D/L
<input type="checkbox"/> Region 1.tif	/	953.4 KIB	Image		
<input type="checkbox"/> Region 2.tif	/	953.4 KIB	Image		
<input type="checkbox"/> Region 3.tif	/	953.4 KIB	Image		
<input type="checkbox"/> Other chip region 1.tif	/	953.4 KIB	Image		

Showing 1 to 4 of 4 datasets



```
{
  "nx_meta": {
    "Acquisition Date": "06/17/2022",
    "Acquisition Time": "02:45:37 PM",
    "Beam Name": "EBeam",
    "Beam Tilt X": 0.0704791,
    "Beam Tilt Y": -0.0416927,
    "Chamber ID": "XL30SB",
    "Chamber Pressure (mPa)": 0.574172,
    "Column Type": "FEG SEM",
    "Creation Time": "2022-06-17T14:46:00.247976",
    "Data Dimensions": "(1024, 884)",
    "Data Type": "SEM_Imaging",
    "DatasetType": "Image",
    "Detector Brightness Setting": 40.3927,
    "Detector Contrast Setting": 38.8255,
    "Detector Grid Voltage (V)": 250,
    "Detector Name": "ETD",
    "Detector Signal": "SE",
    "Drift Correction Applied": true,
    "Emission Current (uA)": 130,
    "Horizontal Field Width (um)": 23.4009,
    "Instrument ID": "FEI-Quanta200-ESEM-633137",
    "Magnification Mode": 3,
    "Operator": "draciti",
    "Pixel Dwell Time (us)": 30,
    "Pixel Height (nm)": 22.8525,
    "Pixel Width (nm)": 22.8525,
    "Software Version": "4.1.15.2218 (build 2218)",
    "Spot Size": 4.5,
    "Stage Description": "50 x 50 manual tilt",
    "Stage Position": {
      "R": -0.0429575,
      "X": -0.000458187,
      "Y": -0.0128239,
      "Z": 0.0100184,
      "a": 0.000391165
    },
    "Stigmator X Value": 0.00128282,
    "Stigmator Y Value": 0.00154145,
    "System Type": "Quanta FEG",
    "Total Frame Time (s)": 29.5667,
    "Vacuum Mode": "High vacuum",
    "Vacuum Pump": "TMP",
    "Vertical Field Width (um)": 20.2016,
    "Voltage (kV)": 20,
    "warnings": [

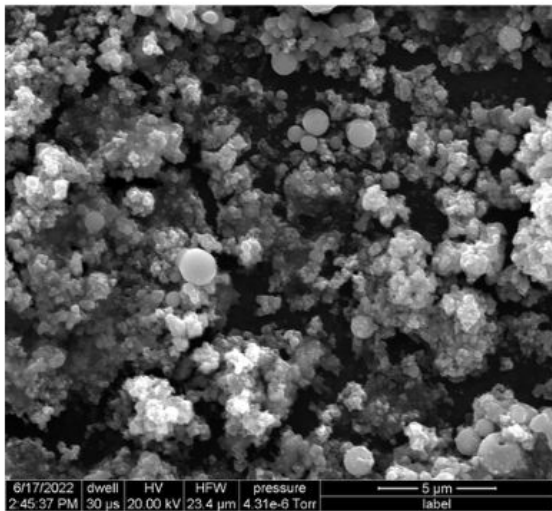
```


Viewing metadata specific to individual “activities”

Experiment activity 1

Activity contents: *SEM Imaging*

4 data files  4 tif



Dataset Name [?]	Creation Time	Type [?]	Role [?]	Meta	D/L
Region 1.tif	2022-06-17 14:21	Image	Experimental	 	
Region 2.tif	2022-06-17 14:32	Image	Experimental	 	
Region 3.tif	2022-06-17 14:39	Image	Experimental	 	
Other chip region 1.tif	2022-06-17 14:46	Image	Experimental	 	

Viewing metadata specific to individual “activities”

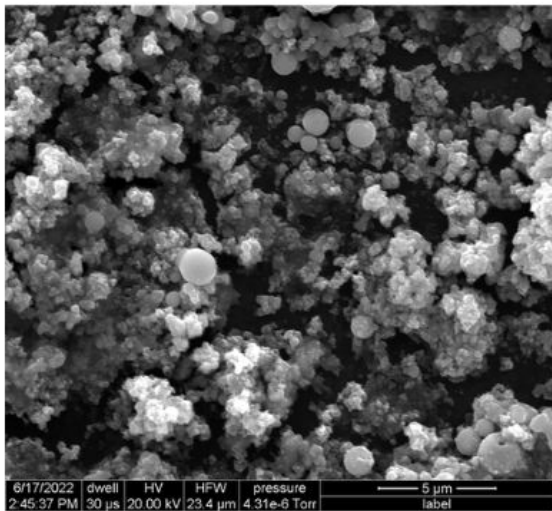
Experiment activity 1



Activity contents: SEM Imaging

4 data files

4 tif



Experiment activity 1

Activity contents: SEM Imaging

Search:

« 1 2 3 »

Setup Parameter	Value
Start time	14:21:06
Acquisition Date	06/17/2022
Beam Name	EBeam
Beam Tilt X	0.0704791
Beam Tilt Y	-0.0416927
Chamber ID	XL30SB
Column Type	FEG SEM
Data Dimensions	(1024, 884)
Data Type	SEM Imaging
Detector Grid Voltage (V)	250.0

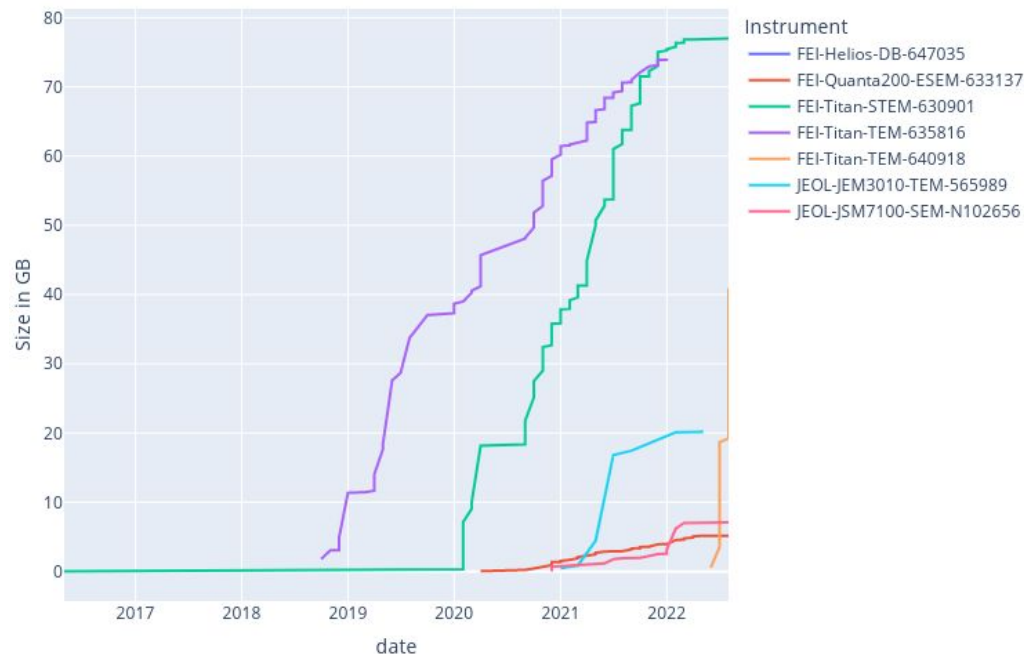
[Missing metadata?](#)

Time	Type	Role	Meta	D/L
7 14:21	Image	Experimental		
7 14:32	Image	Experimental		
7 14:39	Image	Experimental		
7 14:46	Image	Experimental		

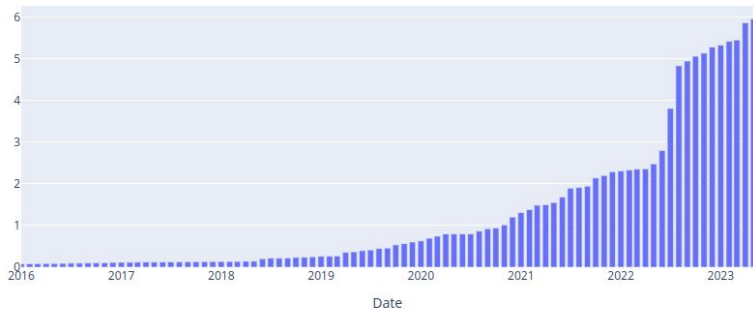
How's it going?

As of May 2023:

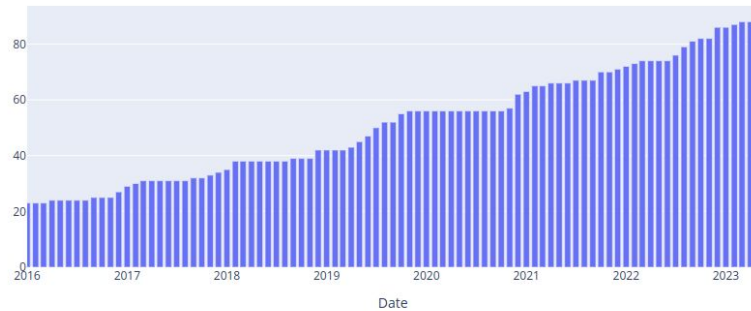
- 16 instruments “under management”
- ~ 800 individual “records” from ~ 90 users
- ~ 500 GB of files processed (mostly .dm3/4 and .tif)
- Lots of types of files (EBSD, 4D-STEM, etc.) we're still not “capturing”



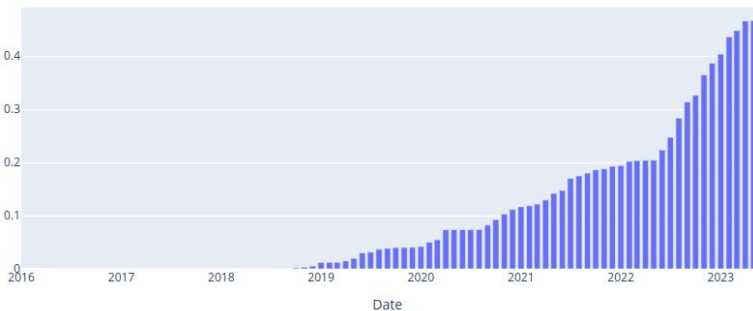
Total CFS data usage (TB)



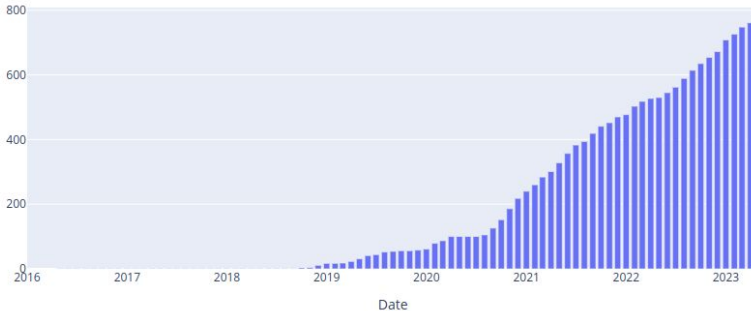
Users with files



Total Nexus data usage (TB)



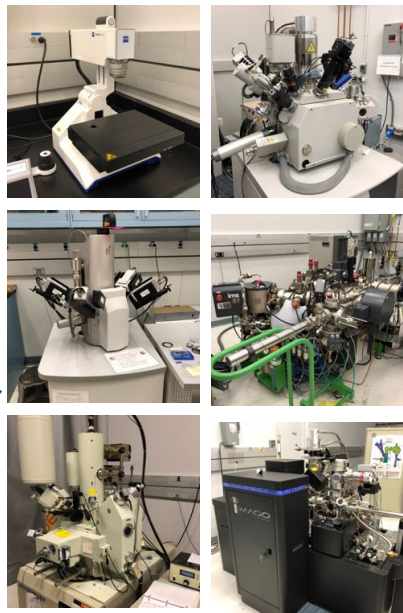
Number of Nexus records



How the sausage gets made...

- Make sure to check out documentation: <https://pages.nist.gov/NexusLIMS>
 - This should stay up-to-date
- In particular:
 - The record building docs: https://pages.nist.gov/NexusLIMS/record_building.html
 - The development docs: <https://pages.nist.gov/NexusLIMS/development.html>
- Assumption: instrument data somehow ends up in one central place
 - Diversion on next two slides describing what this looks like at NIST

Data “Plumbing”



Data Flow Server



Centralized storage; one folder per instrument PC with persistent names

Name	Size	Modified
ABSciex-QTrap_MS-G000019	8 items	3/8/22 10:12 AM
Dell-servo-hydraulic_imaging_computer-G000003	4 items	1/4/22 10:46 AM
EDAX-Gemini_300_EBS-000025	1 item	4/11/22 4:40 PM
EDAX-LEO_1525_EDAX-000022	1 item	4/11/22 3:53 PM
FEI-Helios_FIB_SEM-G000025	63 items	7/28/22 2:57 PM
FEI-Quanta_200F_SEM-G000007	57 items	7/15/22 12:17 PM
FEI-Quanta_400_SEM-000023	1 item	4/7/22 3:29 PM
FEI-Quanta_Bruker-G000008	70 items	5/19/22 9:03 PM
FEI-Titan_80_300_STEM-G000020	18 items	7/15/22 4:42 PM
FEI-Titan_TEM-G000021	26 items	4/15/22 6:05 PM
Gatan-K2_IS-G000022	5 items	7/7/22 8:12 AM
Hitachi-S4700-SEM-606559	2 items	3/5/21 9:35 AM
Illumina-MiSeq_FGx_DNA_Sequencer_Server-G000023	2 items	7/27/22 4:40 PM
Illumina-MiSeq_FGx_DNA_Sequencer-G000023	8 items	7/5/22 10:39 PM
JAWoollam-A330_glove_box_ellipsometer-G000001	81 items	6/21/22 12:07 PM
JAWoollam-A330_insitu_ellipsometer-G000002	10 items	3/3/22 11:00 AM
JEOL-3010_Gatan_S_TEM-G000012	4 items	3/30/22 4:37 PM
JEOL-3010_Strobo_S_TEM-G000013	7 items	3/30/22 5:08 PM

As of July 2022:

- 36.7 TB of data harvested from 66 instruments on 2 campuses

Data “Plumbing”

- Automates data flows from instruments across MML’s scientific laboratories into one or more centralized location(s)
- Each PC shares a read-only folder
 - This folder becomes the new “data” folder for users on the instrument
 - Users can use any folder hierarchy they wish - helpful to use usernames
- Networked server periodically copies all data (rsync) to centralized storage
- Instruments are added via user-submitted form and automated script

The core pieces:

Backend Database

Record building server

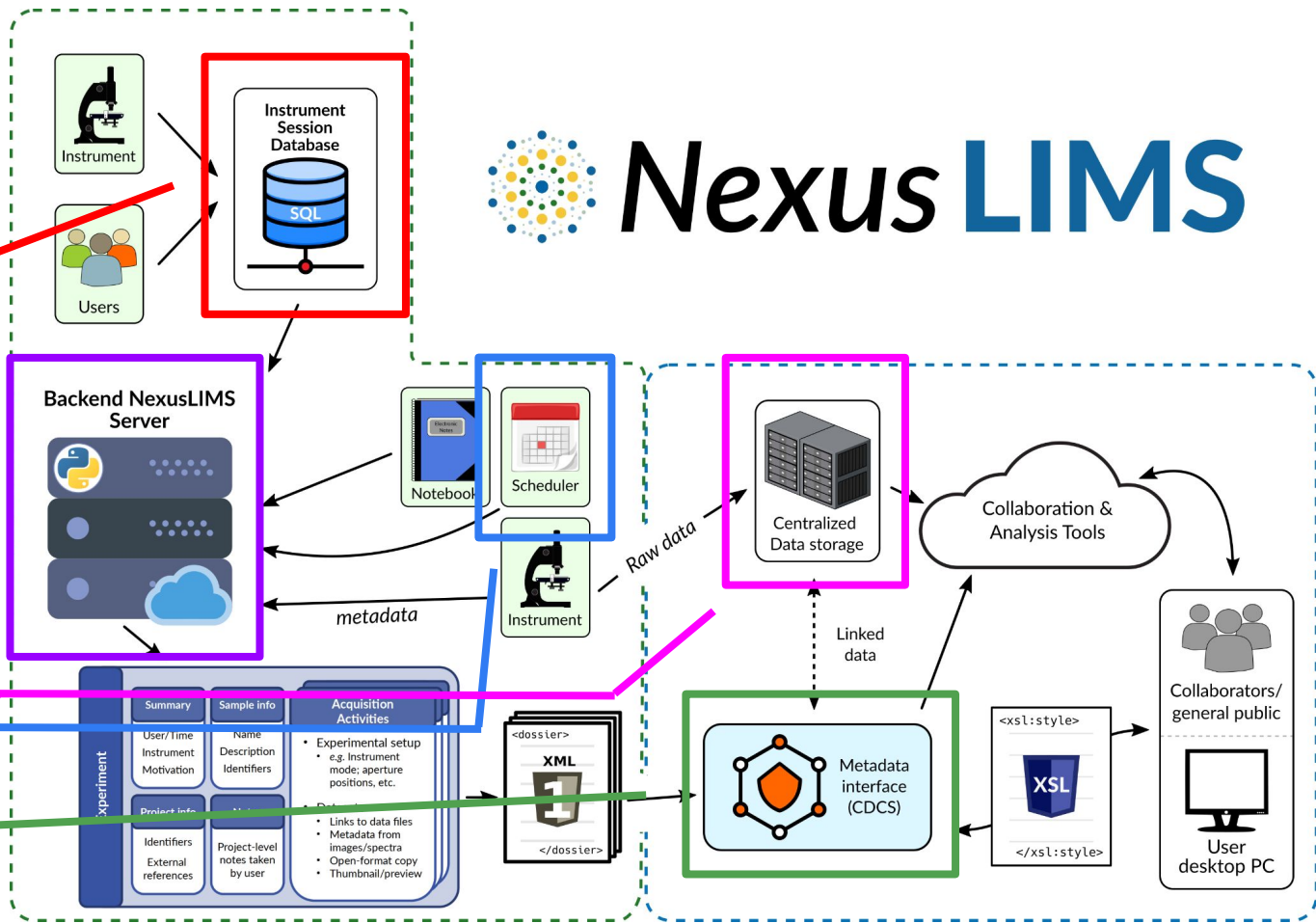
Data storage

Equipment scheduler

Front-end Web UI



Nexus LIMS



NexusLIMS Backend Database

- Very simple model implemented
In SQLite:



session_log	
id_session_log	INTEGER
session_identifier	VARCHAR
instrument	VARCHAR
timestamp	DATETIME
event_type	TEXT
record_status	TEXT
user	VARCHAR

An arrow points from the 'instrument' field in the session_log table to the 'instrument_pid' field in the instruments table.

instruments	
instrument_pid	VARCHAR
api_url	TEXT
calendar_name	TEXT
calendar_url	TEXT
location	VARCHAR
schema_name	TEXT
property_tag	VARCHAR
filestore_path	TEXT
computer_name	TEXT
computer_ip	VARCHAR
computer_mount	TEXT
harvester	TEXT
timezone	TEXT

Scheduler Data Access



Tools

API access is critical for non-interactive access (NEMO has this baked in)

Reservations

```
{
  "id": 246,
  "question_data": {
    "project_id": "Hydrogen",
    "experiment_title": "Deformation evolution",
    "experiment_purpose": "Compare
microstructures
after various ...",
    "data_consent": "Agree",
    "sample_group": {
      // could have additional samples defined
      "0": {
        "sample_name": "4130-no strain",
        "sample_or_pid": "Sample Name",
        "sample_details": ""
      }
    }
  },
  "creation_time":
"2022-01-18T15:48:10.987314-07:00",
  "start": "2022-02-02T08:00:00-07:00",
  "end": "2022-02-03T16:00:00-07:00",
  "user": 2,
  "tool": 15,
  "project": 14
}
```

Usage Events

```
{
  "id": 51,
  "start":
"2022-01-21T08:20:53.879161-07:00",
  "end":
"2022-01-24T06:45:55.363185-07:00",
  "run_data": "",
  "user": 2,
  "operator": 2,
  "project": 13,
  "tool": 15
}
```



```
{
  "id": 15,
  "timezone": "America/New_York",
  "name": "642 JEOL 3010",
  "_description": "Stroboscopic TEM,
Thermionic
LaB6 emitter, 300 keV",
  "_image": "http://*****.nist.gov/media/
tool_images/642-jeol-3010.png",
  "_tool_calendar_color": "#33ad33",
  "_category": "Gaithersburg/(S)TEM",
  "_location": "223 A132",
  "_phone_number": "301-975-2000, x12345",
  "_notification_email_address":
"xyz.abc@nist.gov",
  "_superusers": [ 2 ]
}
```

Record building process

```
python -m nexusLIMS.record_builder
```

Find new sessions

Check for new `usage_events` for our instruments using NEMO API

Add each new usage event to NexusLIMS `session_log` table

For each, return `session_handler.Session` object with instrument, timestamps, and user info

For each Session...

Find matching reservation for this instrument and timespan (with reservation question answers)

Find files created on by this instrument during the usage event timespan

Cluster files by creation time, and extract metadata/create preview for each

Generate XML record from calendar information and files

Upload to front-end

Using web interface API, upload XML directly to CDCS

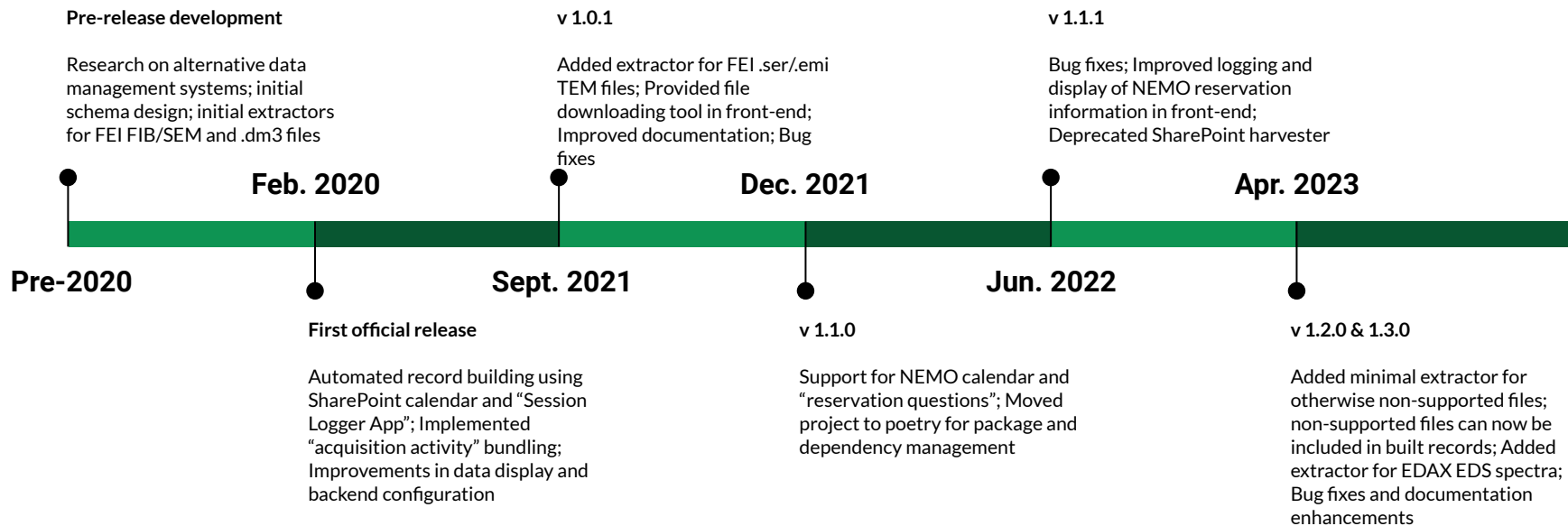
CDCS assigns a persistent identifier (PID) to each record via a local [handle server](#) deployment

Entire process runs without user interaction on a configurable 15 minute interval

Development Process

- Development takes place on using internal Gitlab project with issues, merge requests, CI/CD pipelines, etc.
- Releases periodically pushed to public repo at <https://github.com/usnistgov/NexusLIMS>
- Backend is a Python 3.9+ library with 100% test coverage (*caveat*) and auto-built documentation
- Most development has been done by me, with a few student contributions
- Recently added another part-time resource to work on new features (mostly extractors, at this point)

Development timeline



What else can/should we do?

- Automated metadata extraction from *all* research files, not just those managed by NexusLIMS
- Tools to query and find data by user, instrument, or any other arbitrary metadata
- Additional institutional data sources:
 - Organization-wide instrument database with persistent identifiers
 - Project database; Sample database
- Generalizing capabilities across MML and lowering barrier to entry

What have we learned from NexusLIMS?

- It's extremely hard to do everything yourself!
- If you want to use it, data must be centralized and accessible
- Our problems (mostly) are not particularly unique to microscopy
- As an organization, we need to invest in data-first infrastructure
 - Infeasible to repeat NexusLIMS process for every project, group, etc.

Thank you for your attention!
Questions?

joshua.taillon@nist.gov

<https://orcid.org/0000-0002-5185-4503>