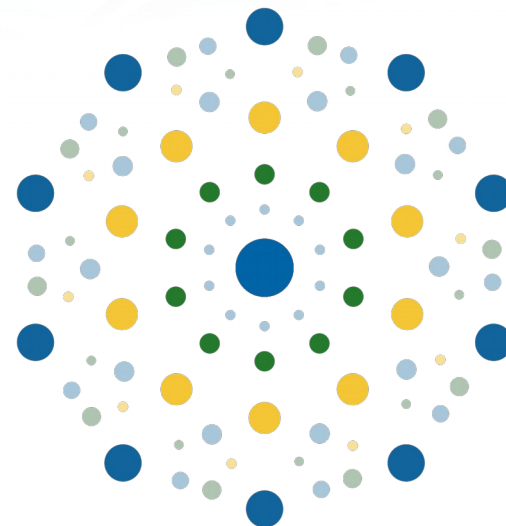


# Harvesting Microscopy Experimental Context with a Configurable Laboratory Information Management System

Joshua A. Taillon, *Thomas Bina*<sup>1</sup>, *Rachel F. Devers*<sup>2</sup>,  
Raymond L. Plante, Marcus W. Newrock, June W. Lau,  
and Gretchen Greene



*Nexus***LIMS**

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

(1)



**PennState**  
Materials Science  
and Engineering

(2)



Electrical and Computer  
Engineering

**MATERIAL  
MEASUREMENT  
LABORATORY**

# Summer Undergraduate Research Fellowship

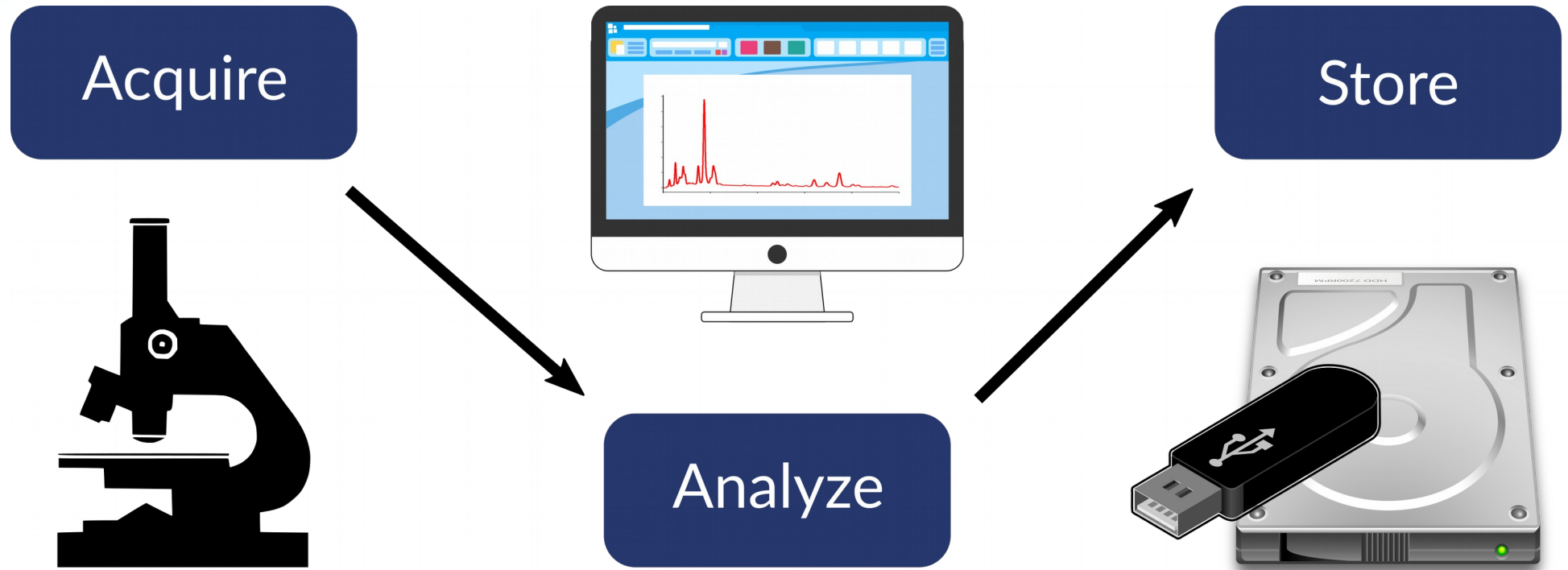


- **When?**
  - Applications due in February of each year
  - Program runs from late May to early August
- **Where?**
  - Gaithersburg, MD or Boulder, CO – *US Citizens only*
- **How much?**
  - ~\$500 per week and allowance for lodging and travel
- **How do I apply/get more info?**
  - <https://www.nist.gov/summer-undergraduate-research-fellowship-surf/>

# 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.*

# An example EM data lifecycle for an individual user



# An example EM data lifecycle for an individual user

Read, 2010

Idell, 2016



Morrow,  
2009

McMorran,  
2011

# Leads to UnFAIR data

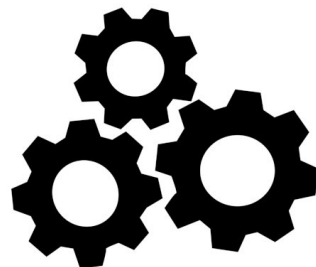
F<sub>indable</sub>



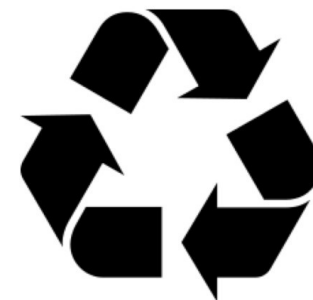
A<sub>ccessible</sub>



I<sub>nteroperable</sub>



R<sub>eusable</sub>



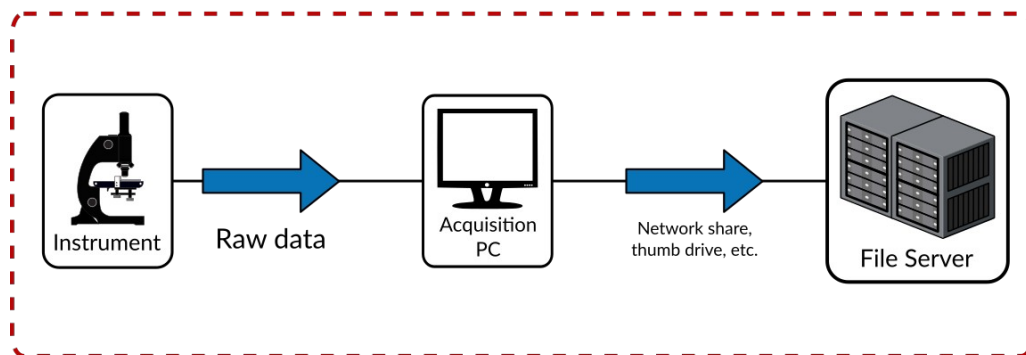
Wilkinson et al., Scientific Data, 3, 160018, 2016 ([link](#))

Image: Sangya Pundir - [CC-BY-SA 4.0](#)

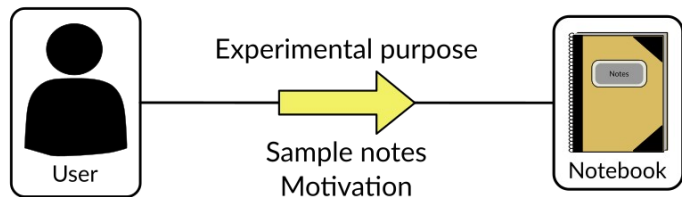


# The *status quo* of facility-level EM Data Management (at NIST)

- Data capture/transfer
- Experimental context

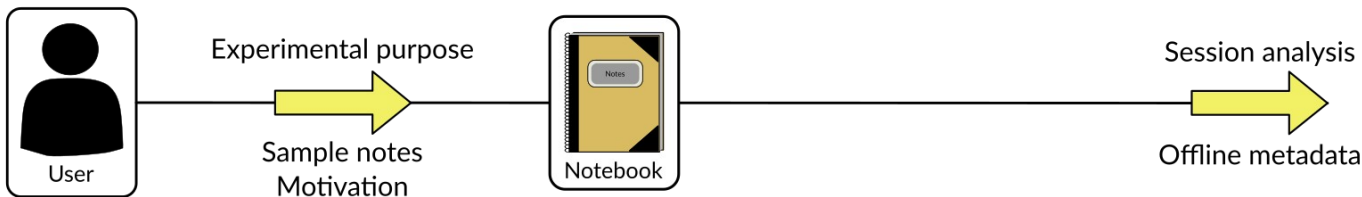
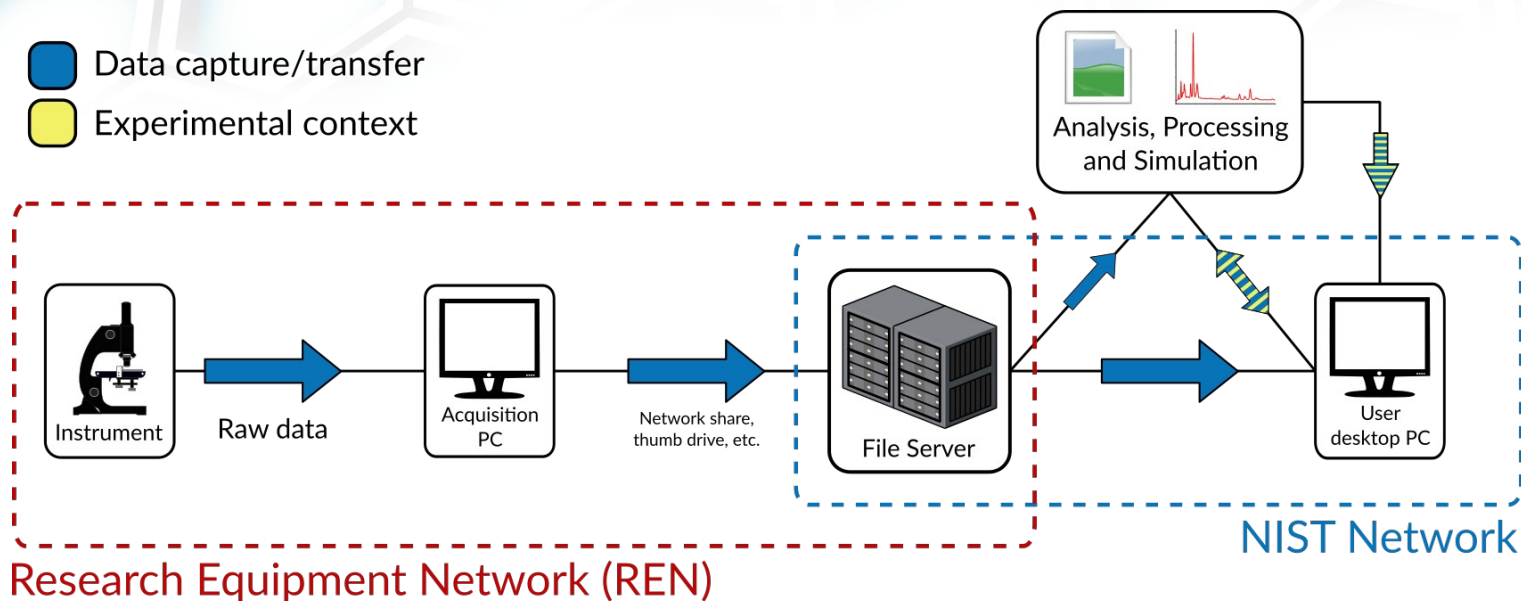


Research Equipment Network (REN)



# The *status quo* of facility-level EM Data Management (at NIST)

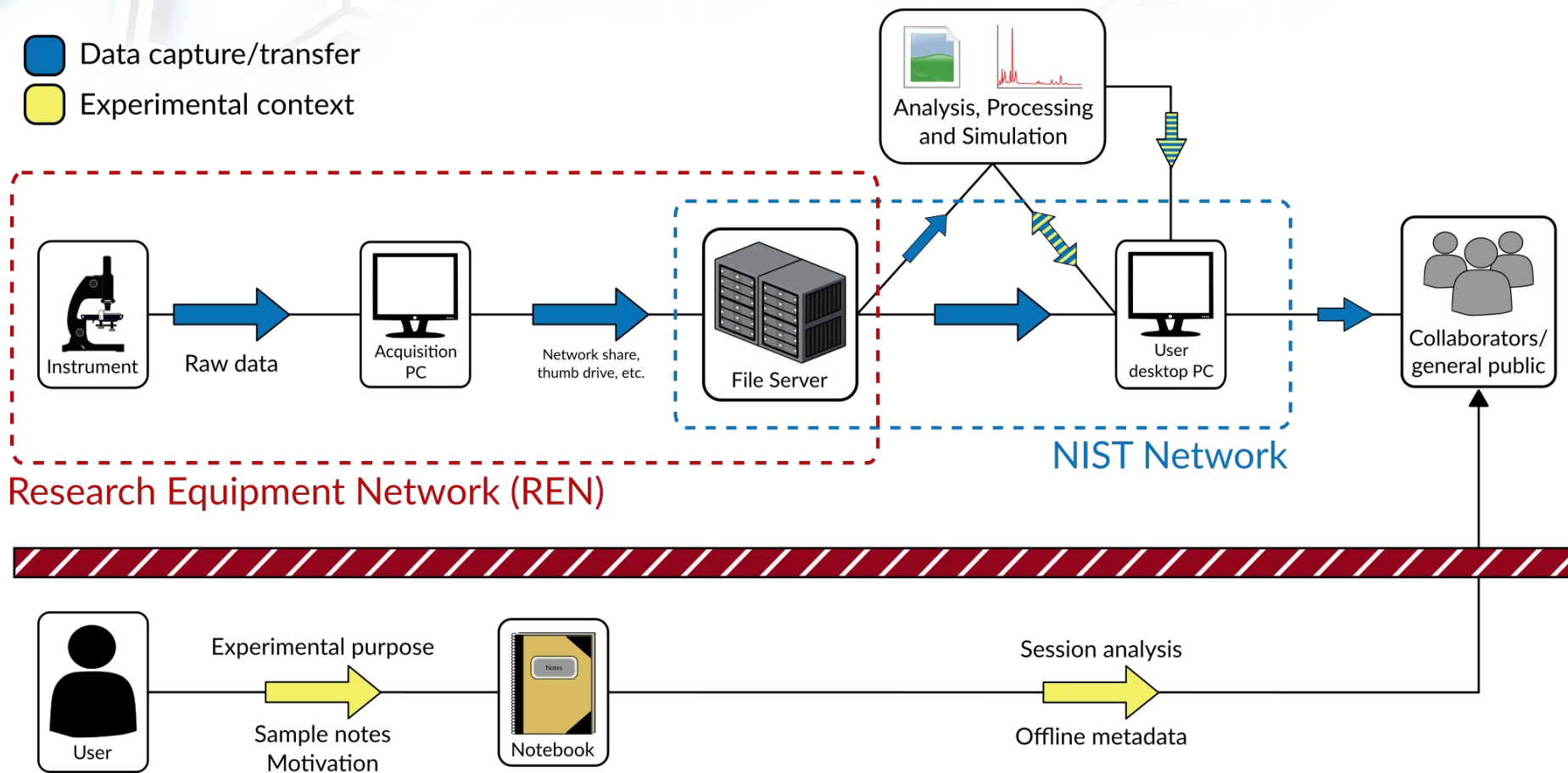
- Data capture/transfer
- Experimental context





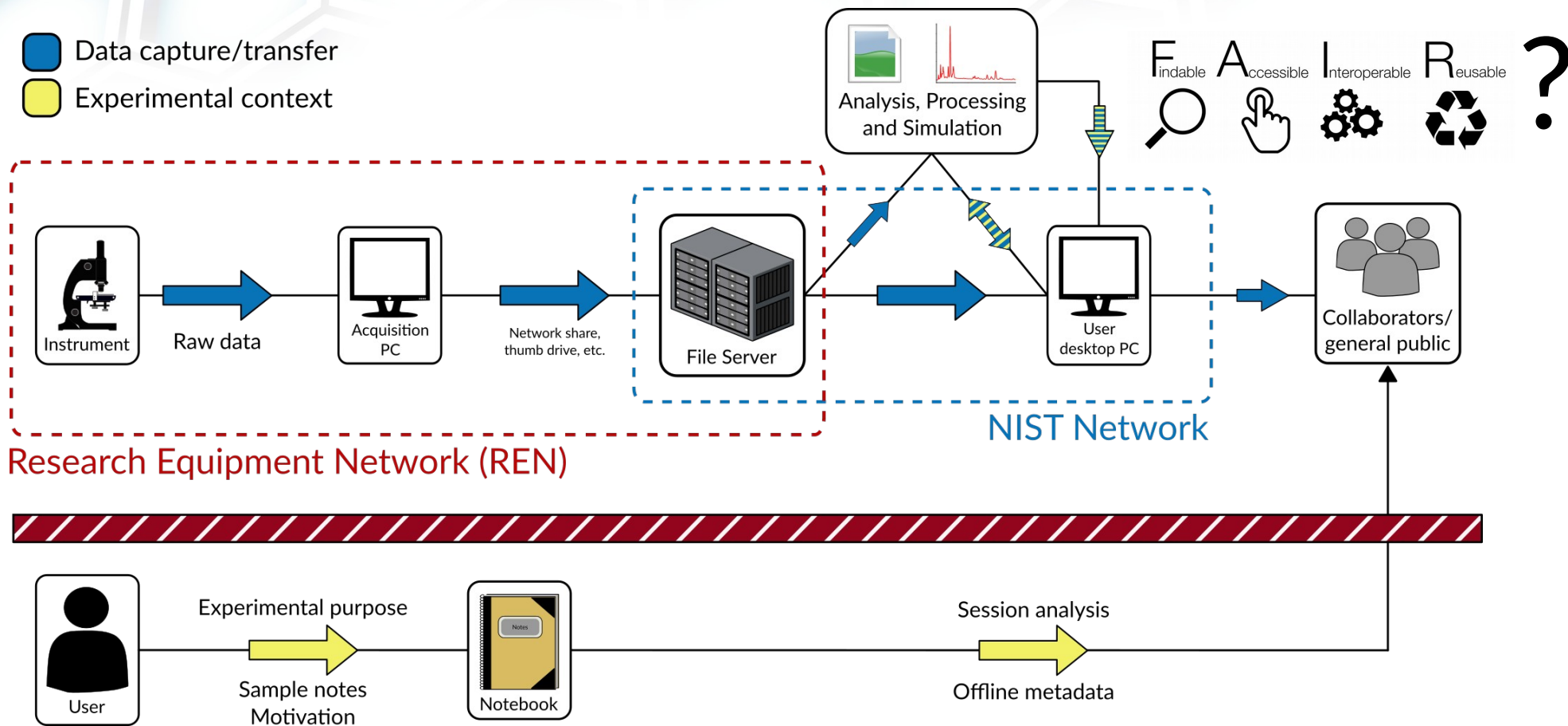
# The *status quo* of facility-level EM Data Management (at NIST)

- Data capture/transfer
- Experimental context



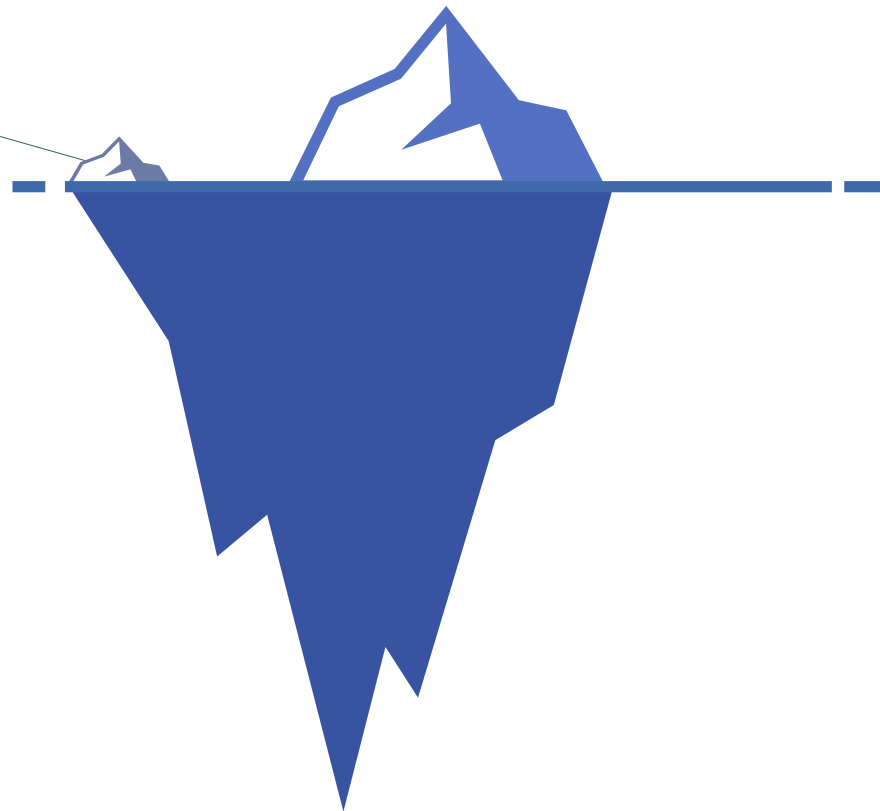
# The *status quo* of facility-level EM Data Management (at NIST)

- Data capture/transfer
- Experimental context



# The challenge of contextual data

**“Easy”  
Metadata**  
*(who, what, when, where)*

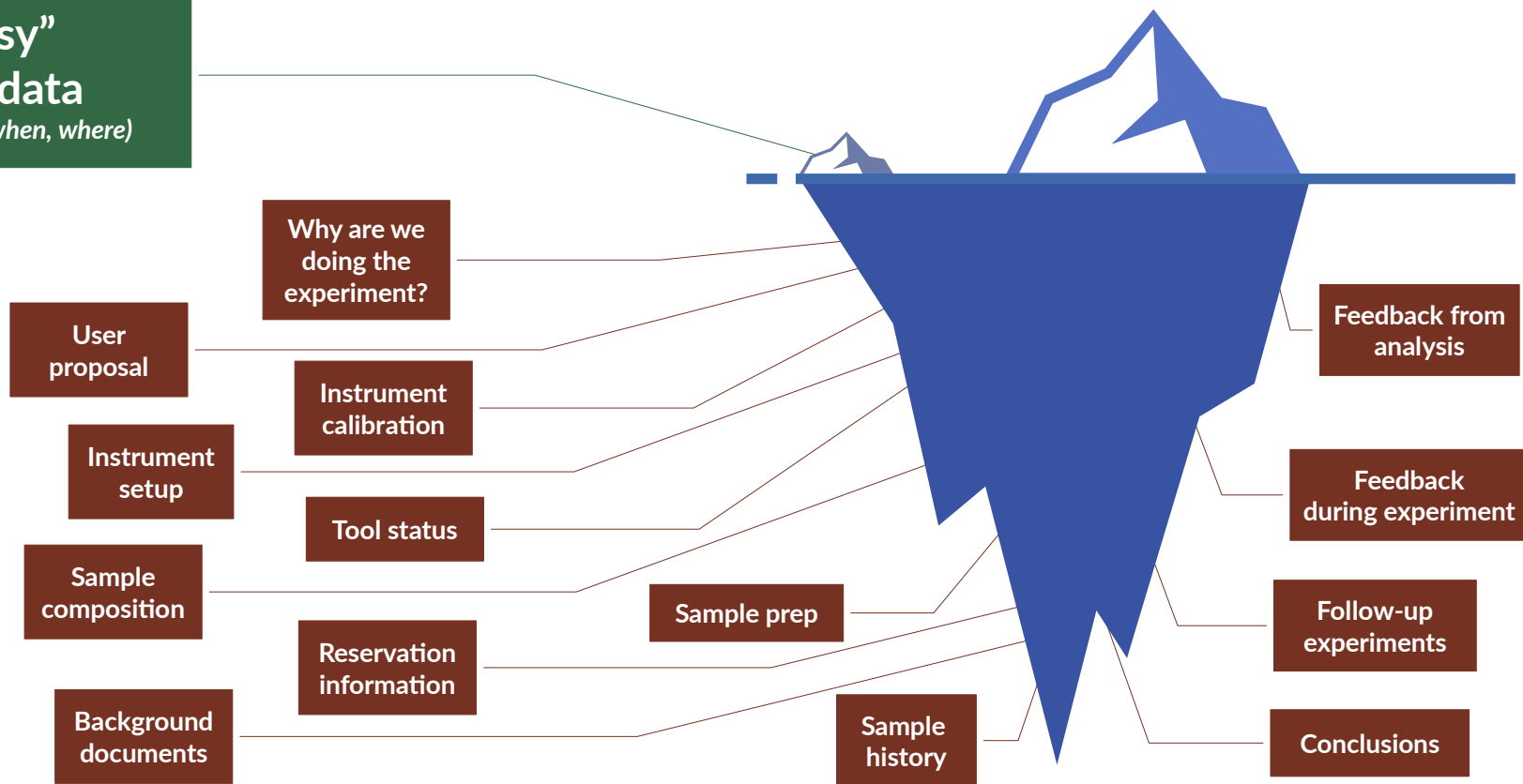
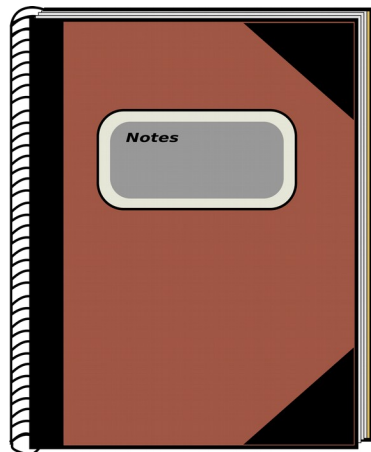


# The challenge of contextual data

## “Easy” Metadata

(who, what, when, where)

## The “Why”



# Our Goal – A LIMS for materials microscopy



**L**aboratory  
**I**nformation  
**M**anagement  
**S**ystem

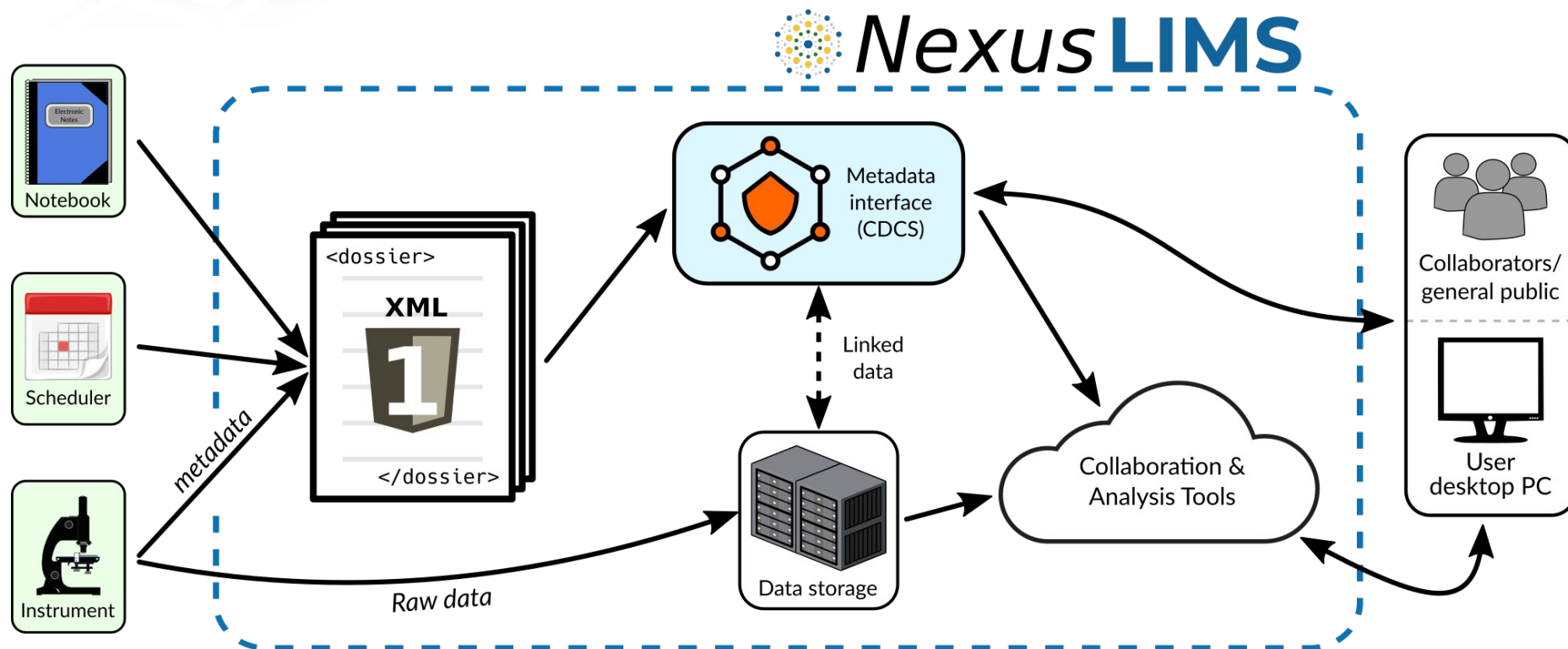
# **Our Goal – A LIMS for materials microscopy**

**Making it simple to capture the  
“why” of an experiment**

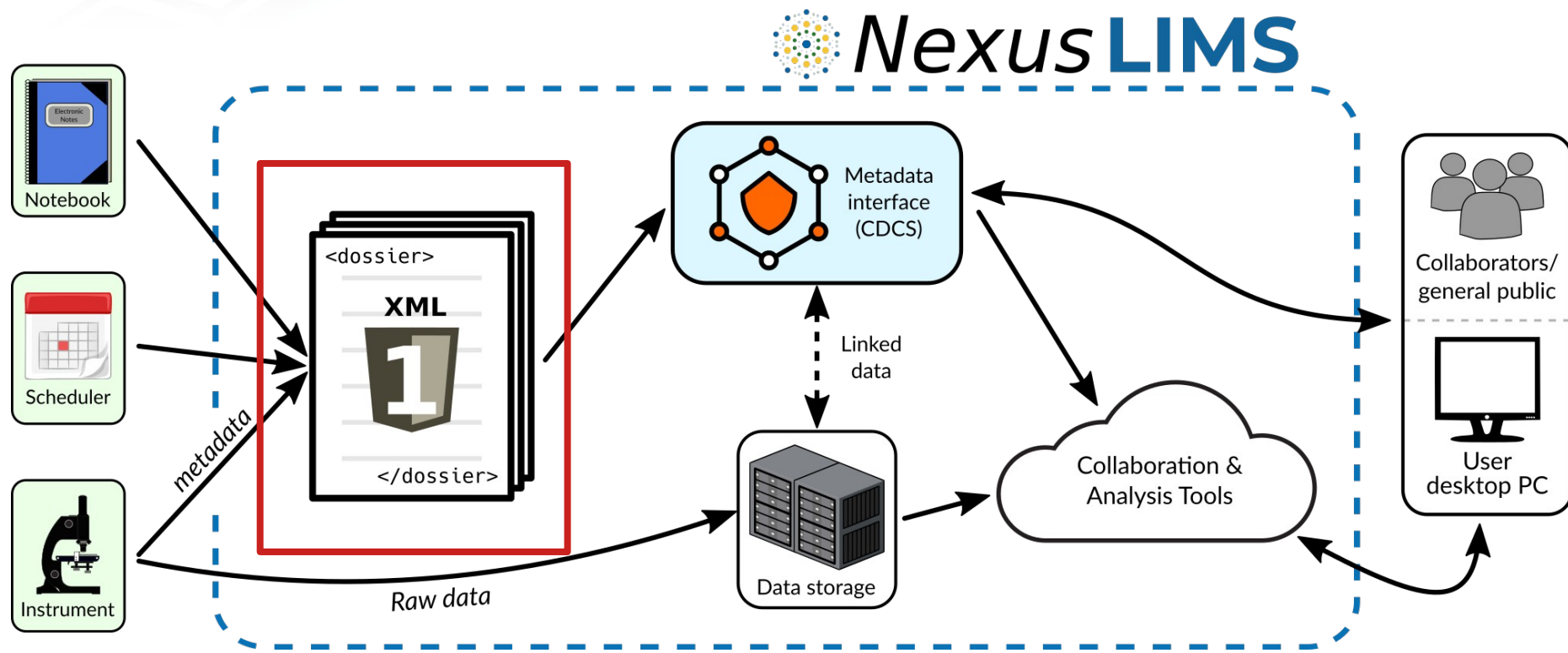
**Facilitating materials research with  
minimal input from the scientist**



# How do we accomplish this?



# How do we accomplish this?



# 1. A schema for materials microscopy – *conceptual*

- Data is most useful when intelligently structured
  - Allows browsing, querying, transforming, validating, etc.

# 1. A schema for materials microscopy – *conceptual*

- 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?

# 1. A schema for materials microscopy – *conceptual*

- 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?
- Schema instance represents a session on the microscope

Experiment

# 1. A schema for materials microscopy – conceptual

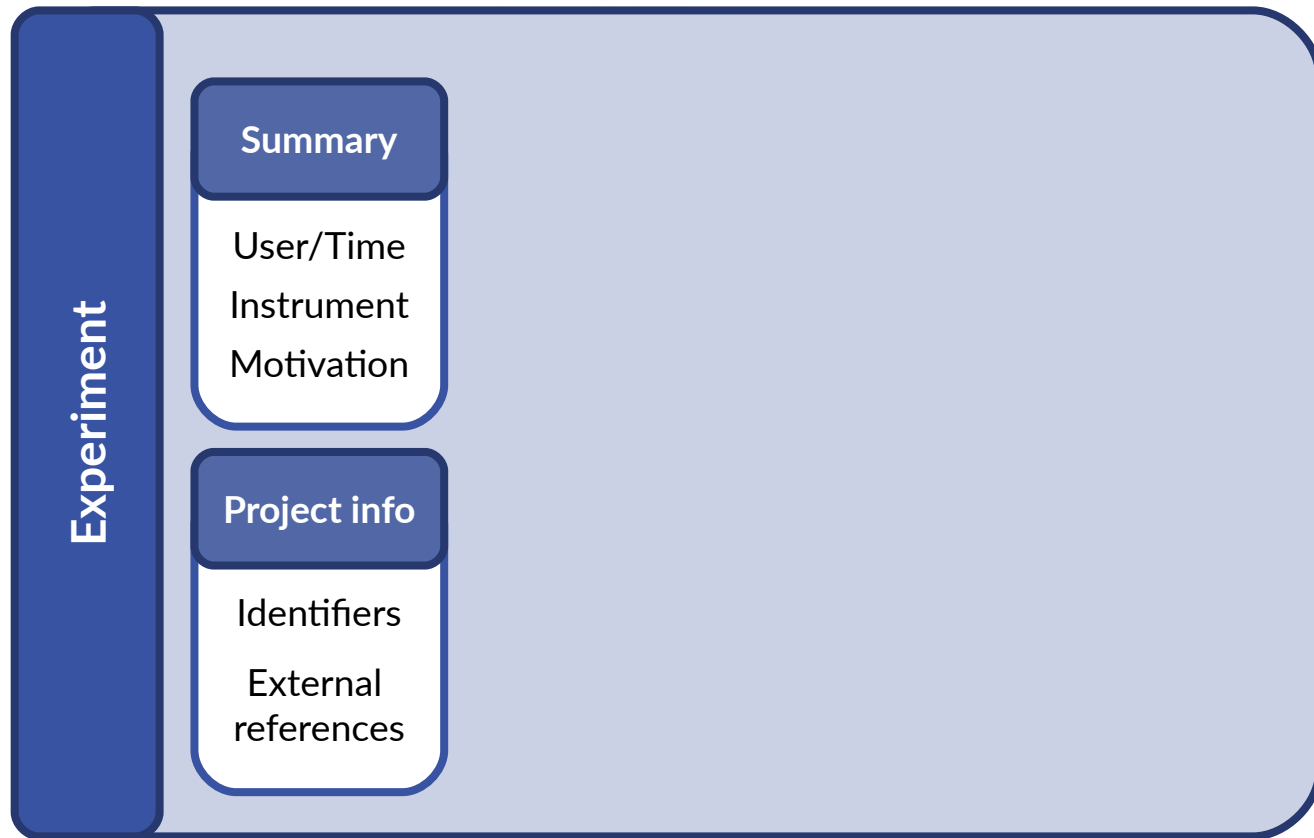
- Data is most useful when intelligently structured
  - Allows browsing, querying, transforming, validating, etc.
- Structure should be tailored to context
  - What information could a researcher/manager/auditor want to see?
- Schema instance represents a session on the microscope





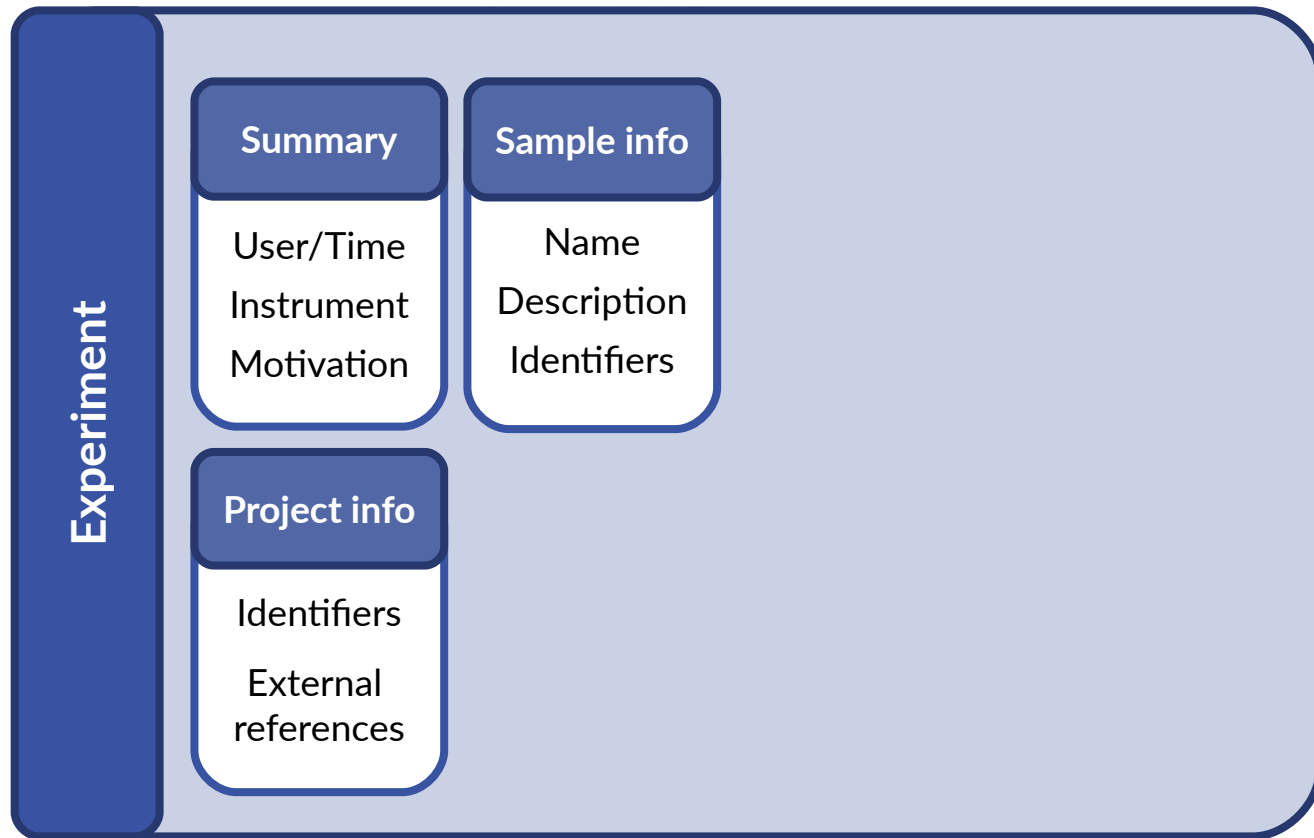
# 1. A schema for materials microscopy – conceptual

- 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?
- Schema instance represents a session on the microscope



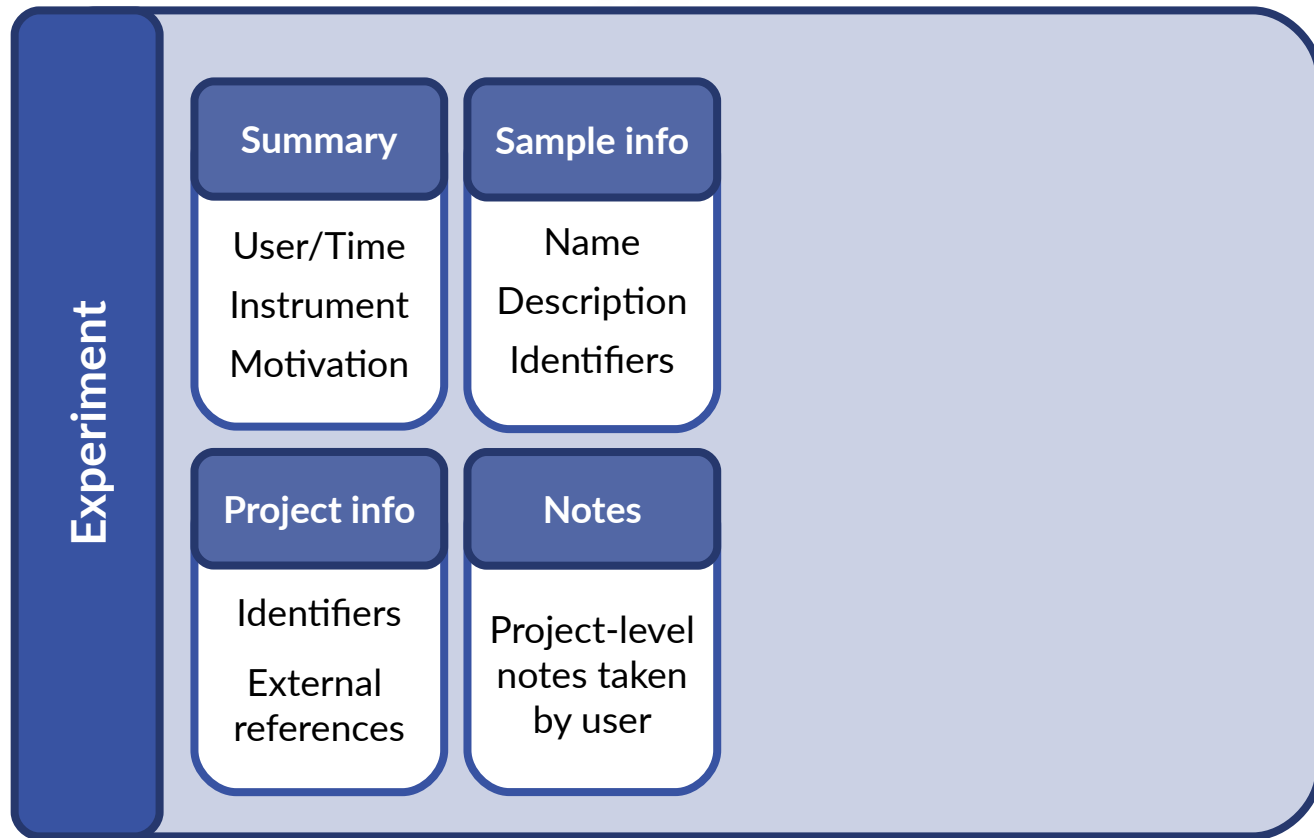
# 1. A schema for materials microscopy – *conceptual*

- 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?
- Schema instance represents a session on the microscope



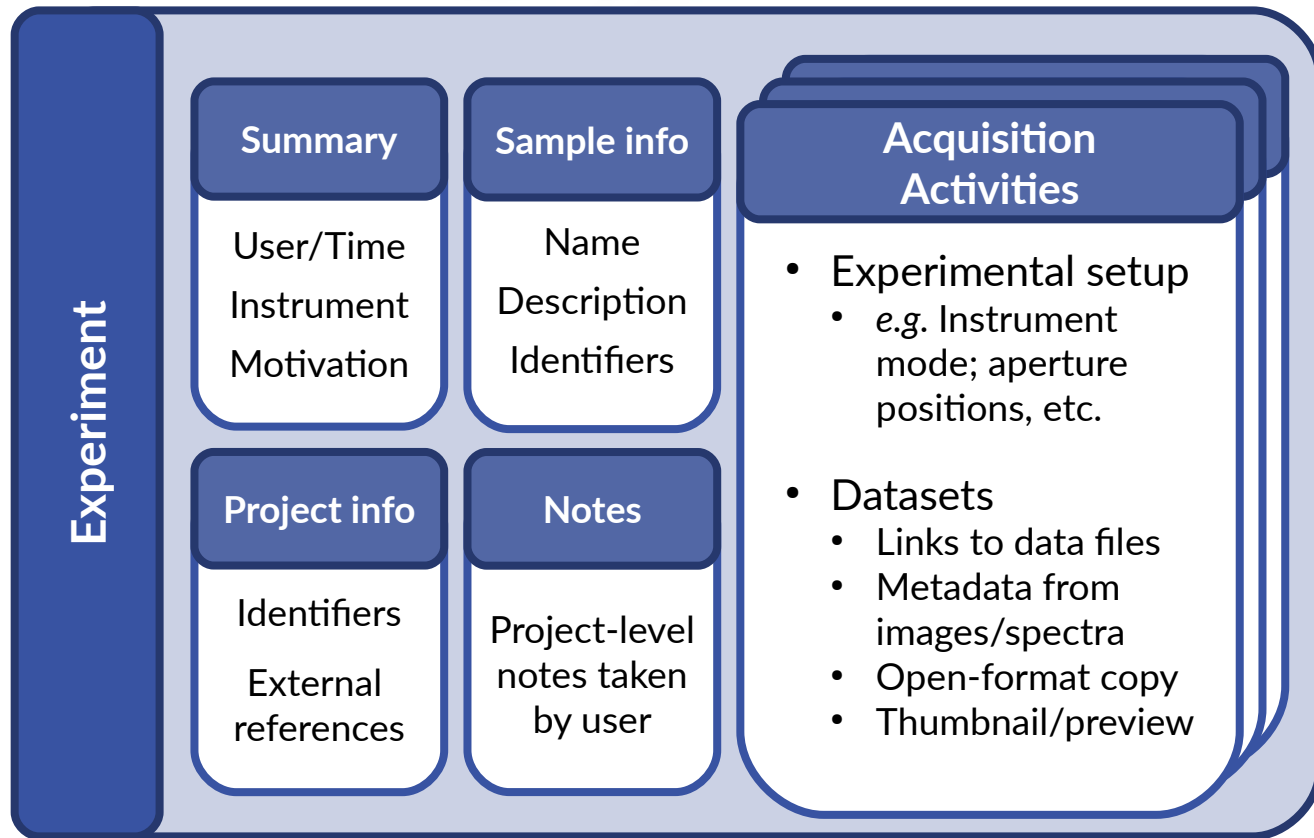
# 1. A schema for materials microscopy – conceptual

- 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?
- Schema instance represents a session on the microscope



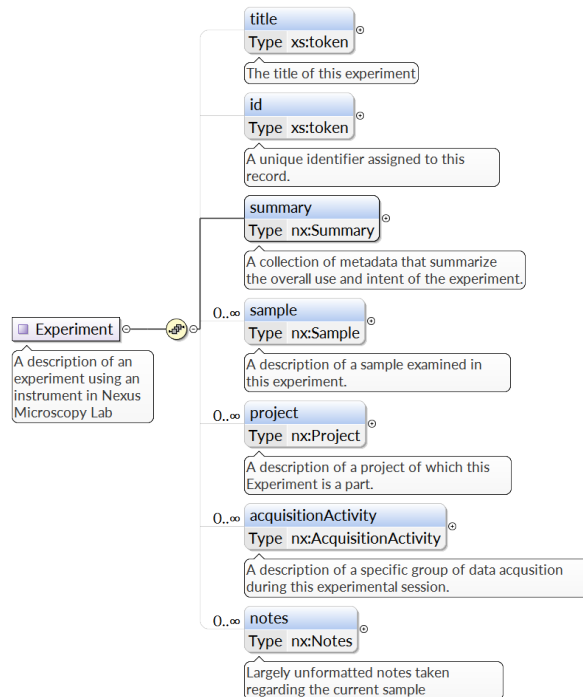
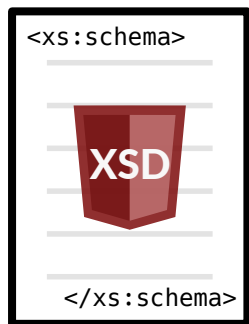
# 1. A schema for materials microscopy – conceptual

- 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?
- Schema instance represents a session on the microscope



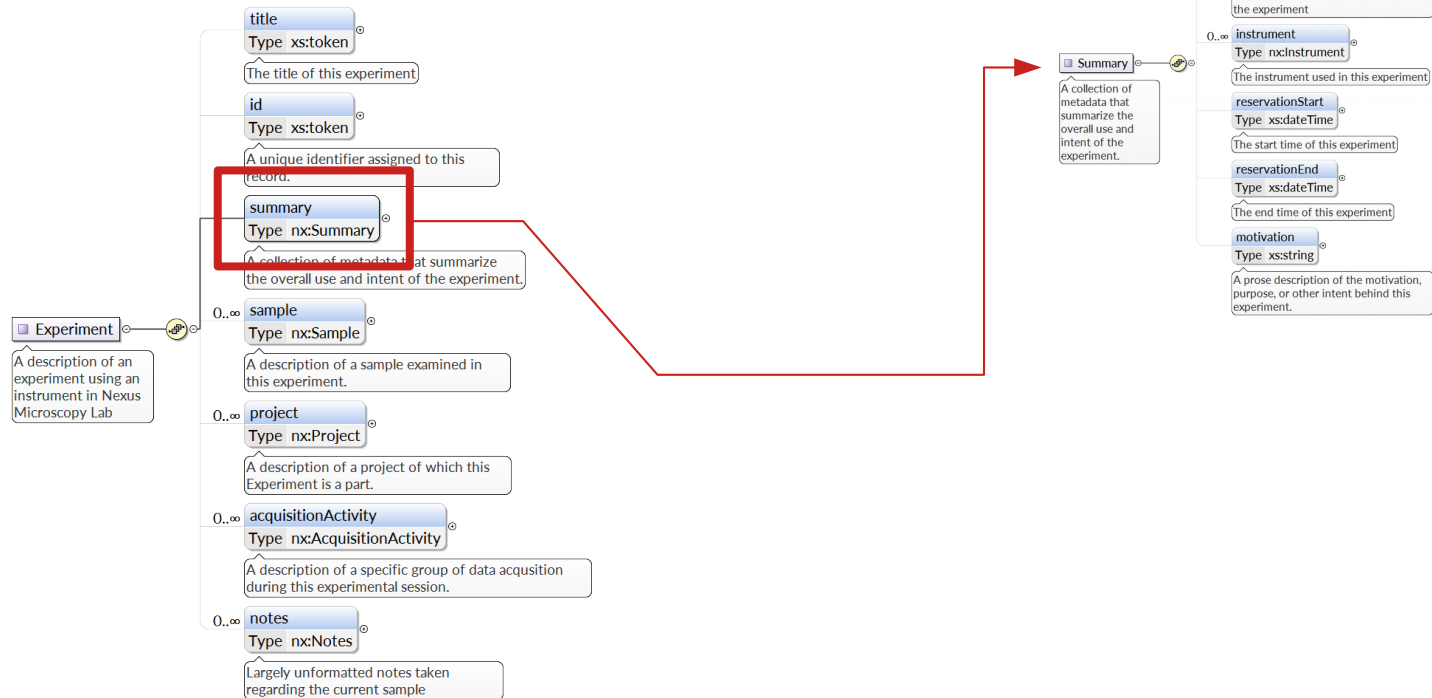
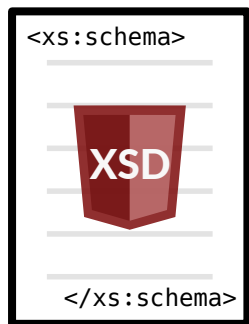
# 1. A schema for materials microscopy – *technical*

- XML schema:
  - Structured
  - Machine readable
  - Supports data validation (data integrity)



# 1. A schema for materials microscopy – technical

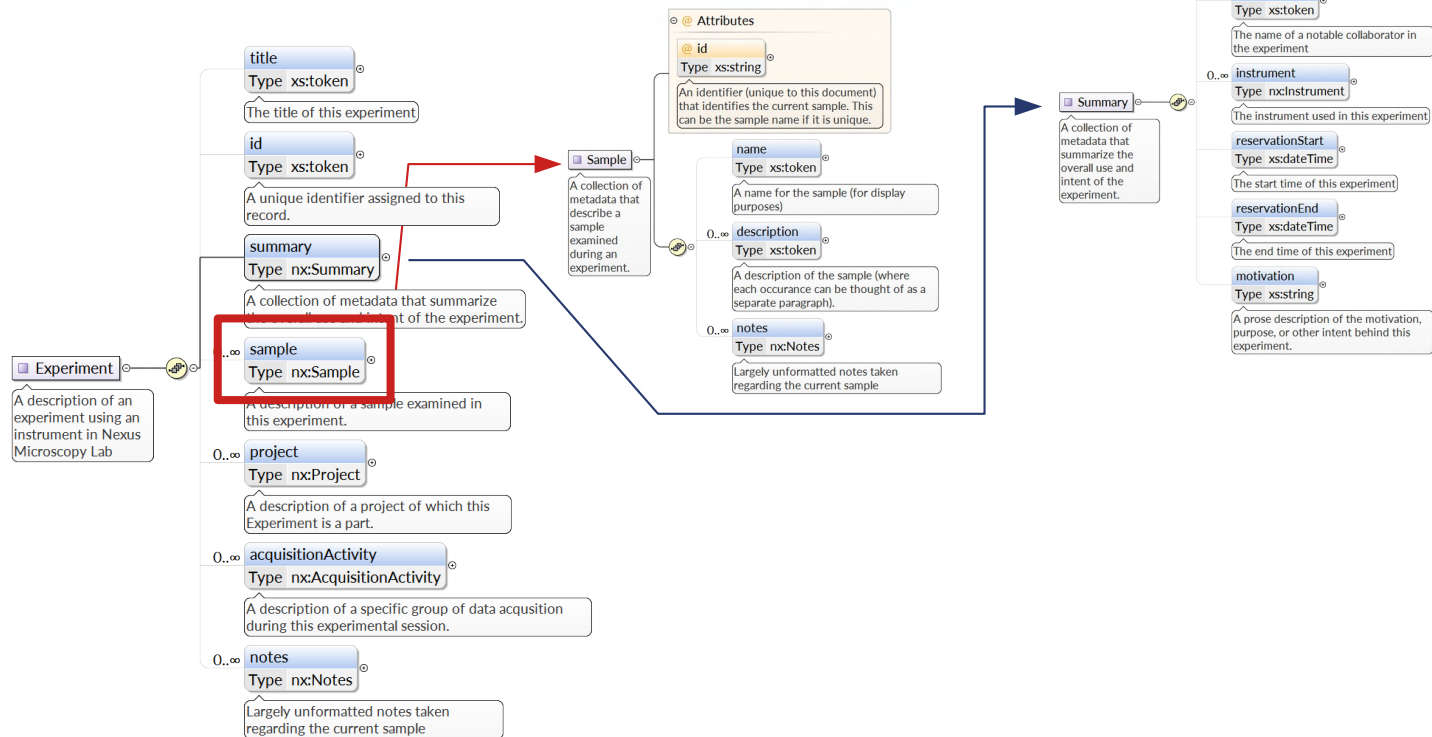
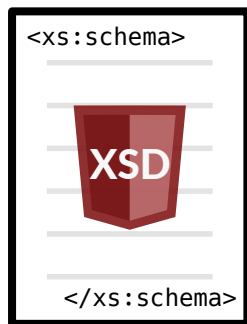
- XML schema:
  - Structured
  - Machine readable
  - Supports data validation (data integrity)





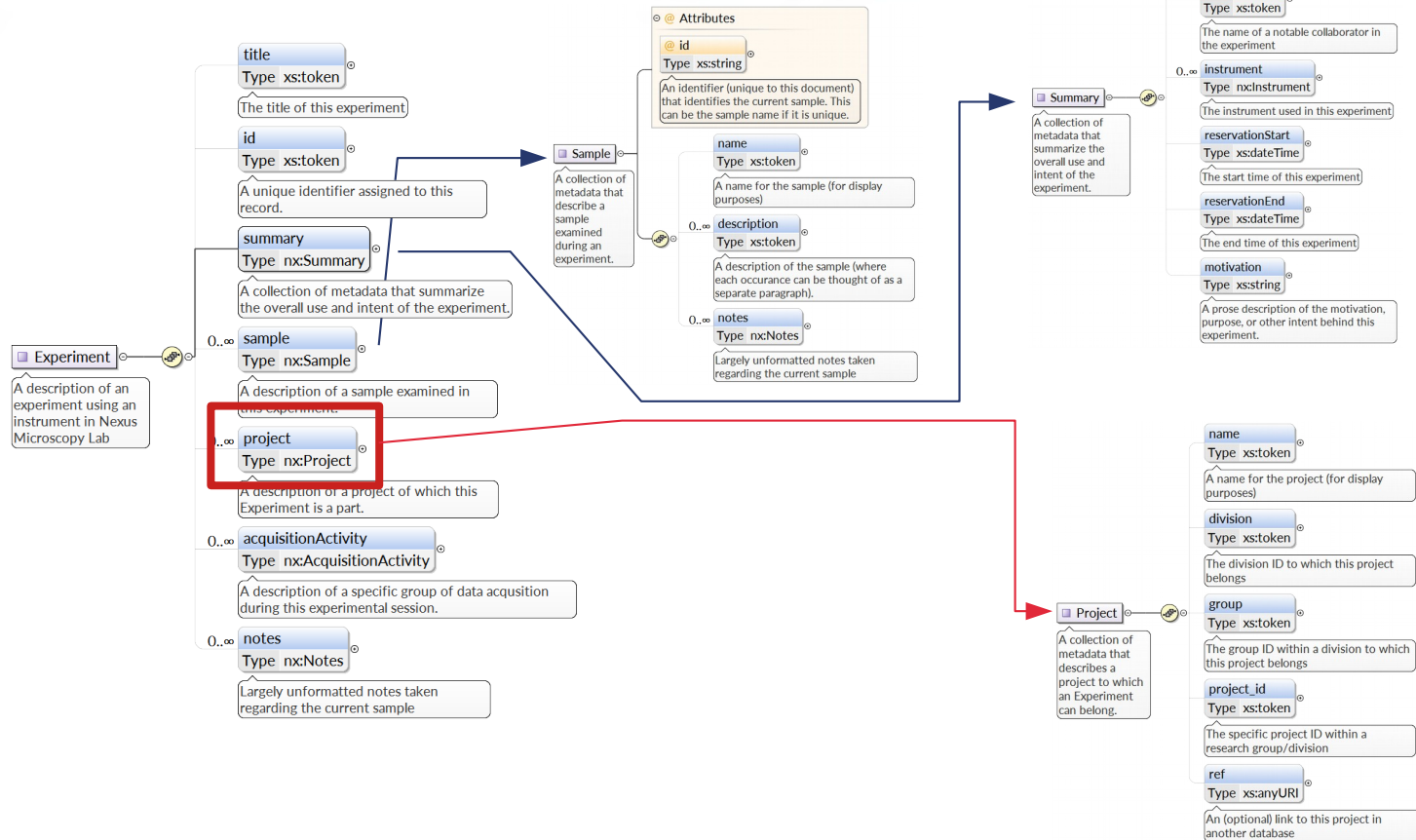
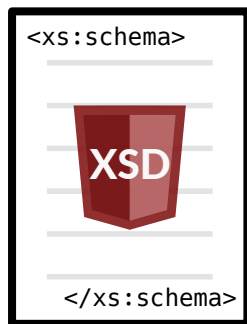
# 1. A schema for materials microscopy – technical

- XML schema:
  - Structured
  - Machine readable
  - Supports data validation (data integrity)



# 1. A schema for materials microscopy – technical

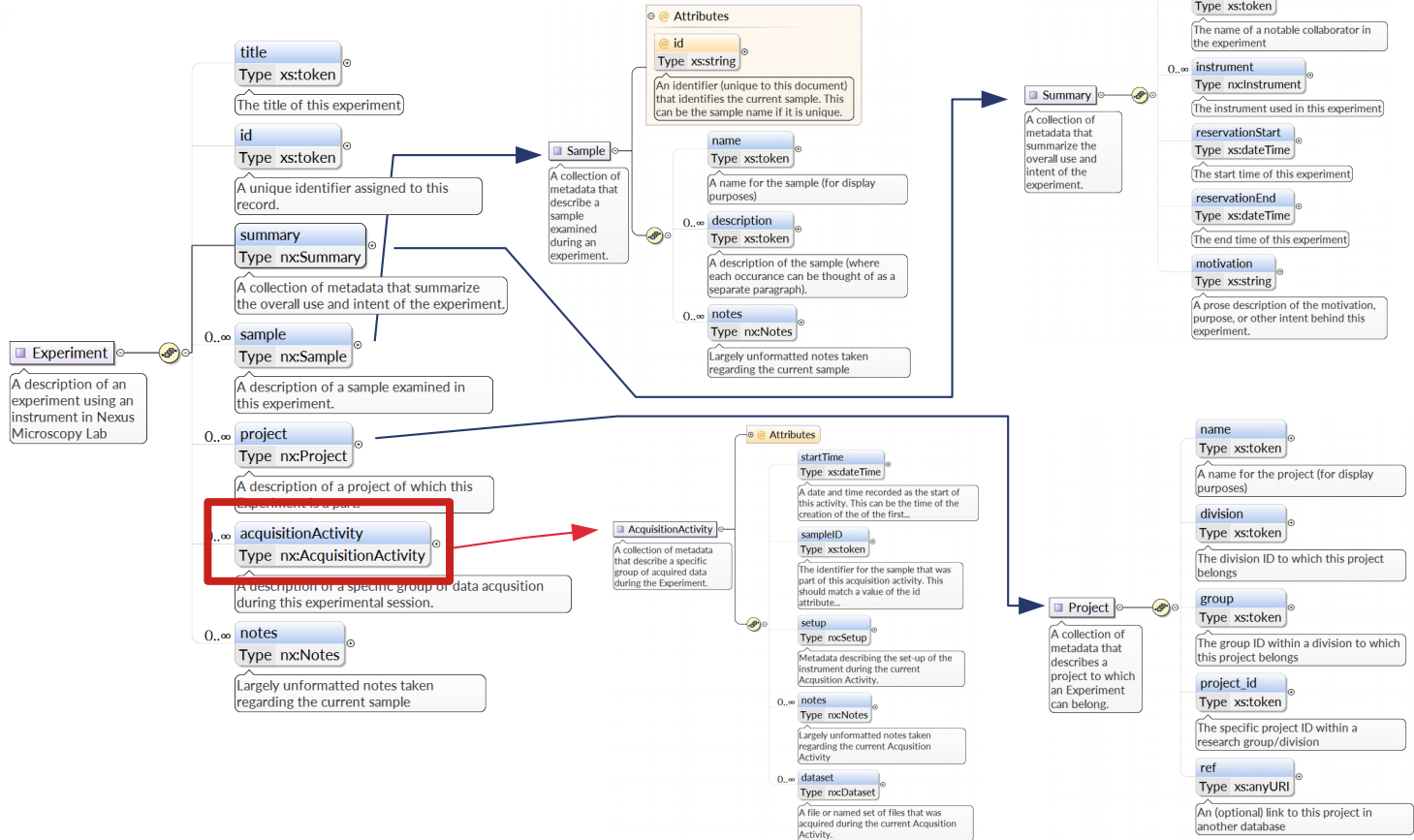
- XML schema:
  - Structured
  - Machine readable
  - Supports data validation (data integrity)



# 1. A schema for materials microscopy – *technical*

- **XML schema:**

- Structured
- Machine readable
- Supports data validation (data integrity)

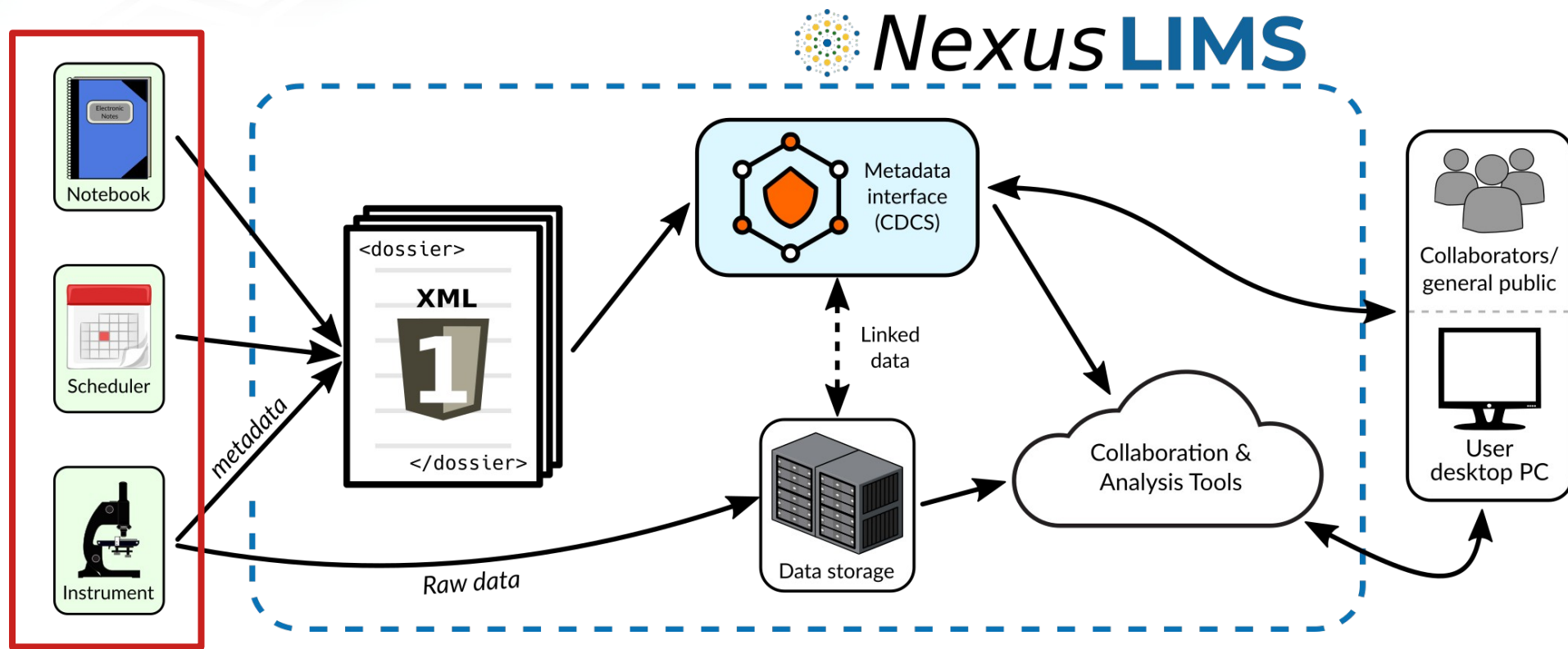


- **XML schema:**

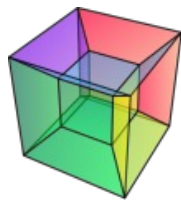
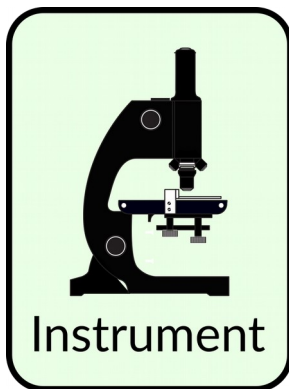
- 



# How do we accomplish this?



## 2. Harvesting and extracting metadata

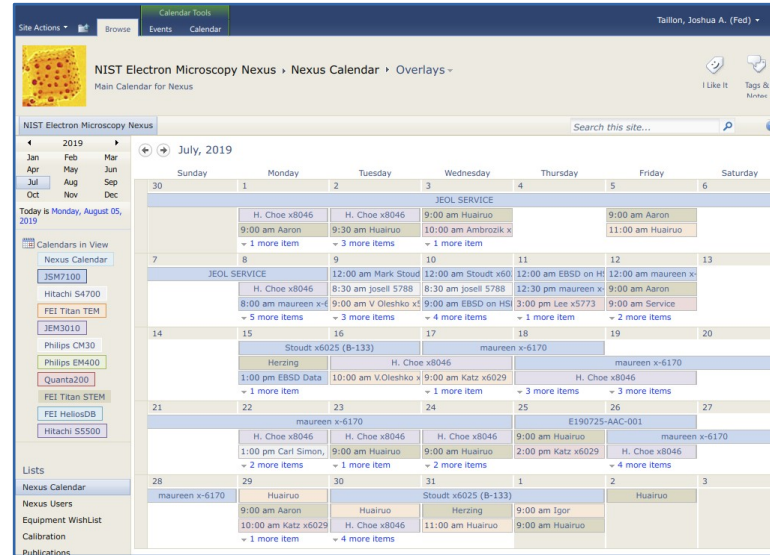
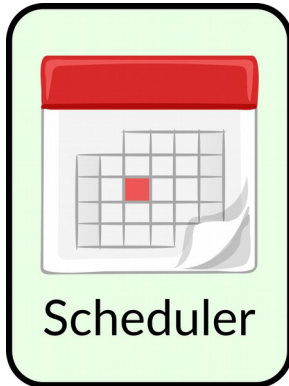


**HyperSpy**  
multi-dimensional data analysis

Instrument and image/spectral metadata  
from central file storage

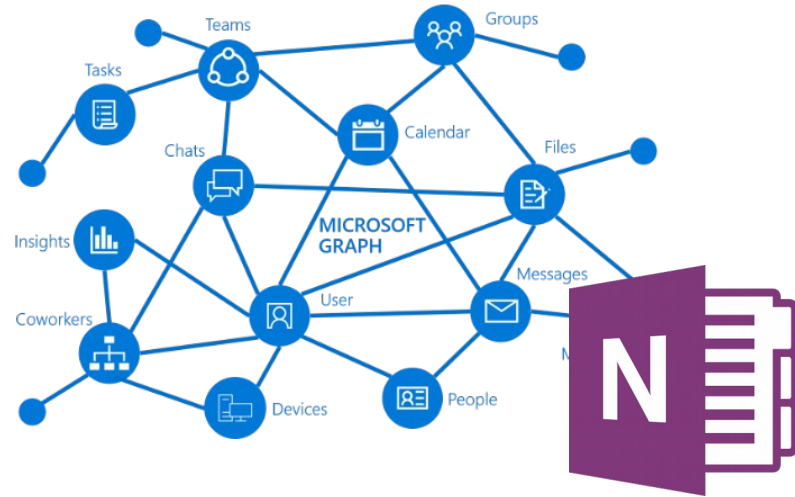
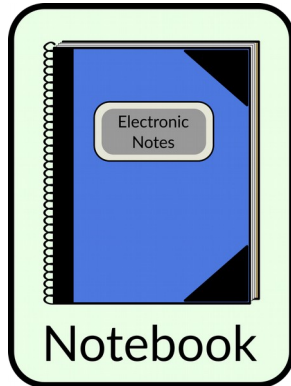


## 2. Harvesting and extracting metadata



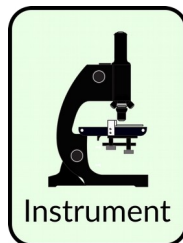
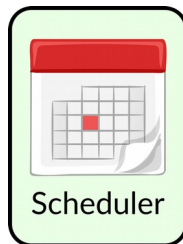
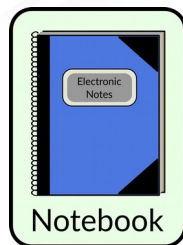
User and session information from SharePoint calendar

## 2. Harvesting and extracting metadata



Users' freeform and/or structured notes

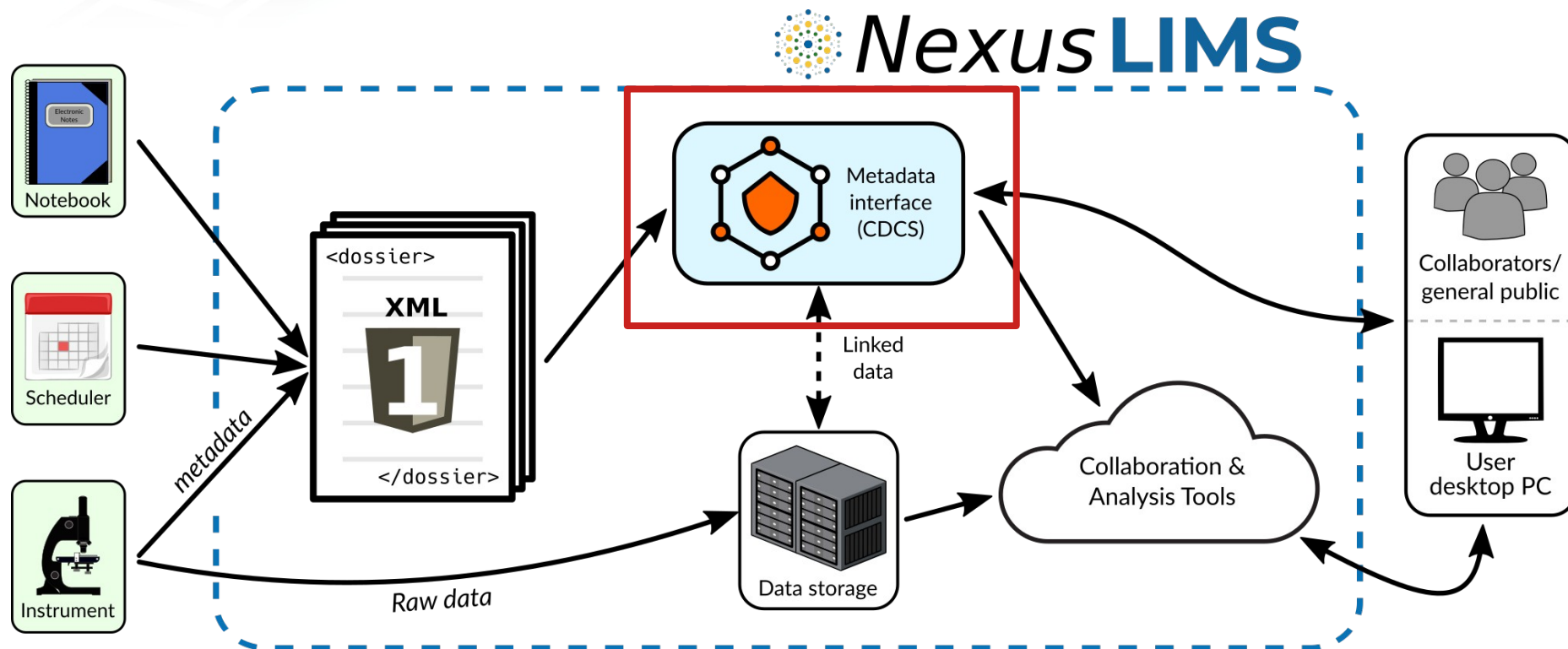
## 2. Harvesting and extracting metadata



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="../xslStylesheet.xsl"?>
3 <nx:Experiment xmlns=""
4 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5 xmlns:nx="https://data.nist.gov/od/dm/nexus/experiment/v1.0">
6   <event>
7     <dateSearched>2018-11-13</dateSearched>
8     <userSearched>mbk1</userSearched>
9     <title>Katz x6029</title>
10    <instrument>FEITitanTEMEEvents</instrument>
11    <user>
12      <userName>mbk1</userName>
13      <name>Katz, Michael B. (Fed)</name>
14      <email>michael.katz@nist.gov</email>
15      <phone>3019756029</phone>
16      <office>223/A163</office>
17      <link>https://share.nist.gov/sites/MML/Div/msed/MSED-MMF/_vti_bin/Li
18      <userId>2815</userId>
19    </user>
20    <purpose>Trying to find us some martensite!</purpose>
21    <sampleDetails>AM 17-4</sampleDetails>
22    <description/>
23    <startTime>2018-11-13T09:00:00</startTime>
24    <endTime>2018-11-13T16:00:00</endTime>
25    <link>https://share.nist.gov/sites/MML/Div/msed/MSED-MMF/_vti_bin/ListDa
26    <eventId>470</eventId>
27  </event>
28  <acquisitionActivity seqno="0">
29    <startTime>2018-11-13T13:01:28.179682</startTime>
30    <sampleID>f81d3518-10af-4fab-9bd3-cfa2b0aea807</sampleID>
31    <setup>
32      <param name="Actual_Magnification">17677.0</param>
33      <param name="C1_Aperture">2000</param>
34      <param name="C2_Aperture">150</param>
35      <param name="Csmm">1.2</param>
36      <param name="Dif_Strength">44.625</param>
37      <param name="Extractor_Voltage">4400</param>
```

XML Record/"Dossier"

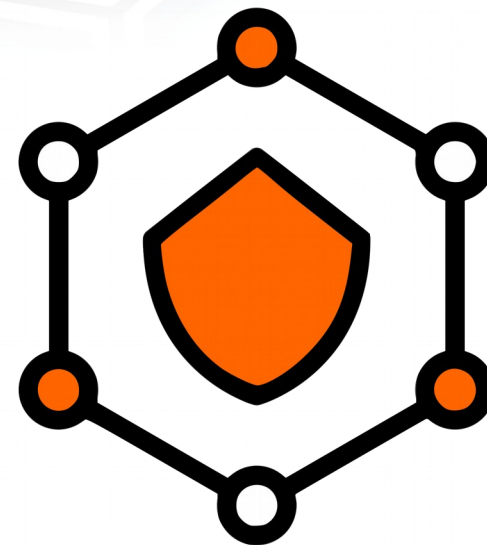
# How do we accomplish this?



### 3. Providing an interface for users

Once the records are built:

- Organize, curate, and display using CDCS – the Configurable Data Curation System



**CDCS**

Dima, A., Bhaskarla, S., Becker, C. *et al.* JOM (2016) 68: 2053.  
<https://doi.org/10.1007/s11837-016-2000-4>

<https://www.nist.gov/itl/ssd/information-systems-group/configurable-data-curation-system-cdcs/about-cdcs>

# 3. Providing an interface for users

- CDCS used as the LIMS “backbone”

```

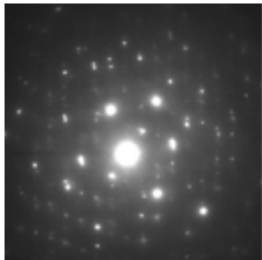
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="../../xsl/stylesheet.xsl"?>
3 <mx:Experiment xmlns="">
4   xmlns:xsl="http://www.w3.org/2001/XMLSchema-instance"
5   xmlns:mx="https://data.nist.gov/od/ds/nexus/experiment/v1.0"
6   <event>
7     <dateSearched>2018-11-13</dateSearched>
8     <userSearched>mbk1</userSearched>
9     <title>Katz x6029</title>
10    <instrument>FEITitanTEMEvents</instrument>
11    <user>
12      <userName>mbk1</userName>
13      <name>Katz, Michael B. (Fed)</name>
14      <email>michael.katz@nist.gov</email>
15      <phone>3019756029</phone>
16      <office>223/A103</office>
17      <link>https://share.nist.gov/sites/WWW/Div/msed/WSED-MMF_vti_bin/Listba
18      <userId>2815</userId>
19    </user>
20    <purpose>Trying to find us some martensite</purpose>
21    <sampleDetails>AM 17-4</sampleDetails>
22    <description>
23      <startTime>2018-11-13T09:00:00</startTime>
24      <endTime>2018-11-13T16:00:00</endTime>
25      <link>https://share.nist.gov/sites/WWW/Div/msed/WSED-MMF_vti_bin/Listba
26      <eventId>470</eventId>
27    </event>
28    <acquisitionActivity seqno="0">
29      <startTime>2018-11-13T13:01:28.179682</startTime>
30      <sampleID>f81d3518-10af-4fab-9bd3-cfa2b0aea807</sampleID>
31      <setup>
32        <param name="Actual_Magnification">17677.0</param>
33        <param name="C1_Aperture">2000</param>
34        <param name="C2_Aperture">150</param>
35        <param name="Cmm">1.2</param>
36        <param name="Drift_Strength">44.625</param>
37        <param name="Extractor_Voltage">4400</param>
  
```



```

xsl:stylesheet xsl:template html body div
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
3   xmlns:xs="http://www.w3.org/2001/XMLSchema"
4   xmlns:mx="https://data.nist.gov/od/ds/nexus/experiment/v1.0"
5   exclude-result-prefixes="xs"
6   version="2.0">
7
8 <xsl:template match="/mx:Experiment">
9
10  <!-- CSS Styling -->
11  <!-- Main Generation of the Page -->
12  <!-- HTML Wrapper -->
13  <!-- Execute showButtonOnScroll() whenever page is scrolled to have the button which jumps back to the
14  <!-- Add sidebar to the page -->
15  <!-- Include sidebar heading -->
16  <!-- Navigation -->
17  <!-- Add a horizontal line to separate sections in the sidebar -->
18  <!-- Define site title for the page -->
19  <!-- Create floating button in bottom right which jumps to the top of the page when clicked -->
20  </div>
  
```



<b>Navigation</b> Summary	<b>Katz x6029</b> Katz, Michael B. (Fed) michael.katz@nist.gov																																																		
<b>Activity 1</b> Mode: TEM UP SA Zoom Image	<b>Motivation</b> Trying to find us some martensite!																																																		
<b>Activity 2</b> Mode: TEM UP SA Zoom Diffraction	Instrument: FEITitanTEMEvents Date: 2018-11-13 Start Time: 09:00:00 End Time: 16:00:00 Session ID: 470																																																		
<b>Activity 3</b> Mode: TEM UP SA Zoom Image	<b>Sample Information</b> Sample Name: AM 17-4 Sample ID: f81d3518-10af-4fab-9bd3-cfa2b0aea807 Description:																																																		
<b>Activity 4</b> Mode: TEM UP SA Zoom Diffraction	<b>Acquisition Activity 1</b> TEM UP SA Zoom Image (Original Data - placeholder)																																																		
<b>Activity 5</b> Mode: TEM UP SA Zoom Image	<b>Activity Parameters</b> Start time: 13:01:28.179682																																																		
<b>Activity 6</b> Mode: TEM UP SA Zoom Diffraction	<table border="1"> <thead> <tr> <th>Setup</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>Actual_Magnification</td><td>17677.0</td></tr> <tr><td>C1_Aperture</td><td>2000</td></tr> <tr><td>C2_Aperture</td><td>150</td></tr> <tr><td>Cmm</td><td>1.2</td></tr> <tr><td>Drift_Strength</td><td>44.625</td></tr> <tr><td>Extractor_Voltage</td><td>4400</td></tr> <tr><td>Filter_Aperture</td><td>3.0</td></tr> <tr><td>Filter_Dispersion</td><td>0.1</td></tr> <tr><td>Filter_Drift_Tube</td><td>0.0</td></tr> <tr><td>Filter_Mode</td><td>Spectroscopy</td></tr> <tr><td>Filter_Prism_Shift</td><td>-0.0</td></tr> <tr><td>Filter_Total_Energy_Loss</td><td>0.0</td></tr> <tr><td>Gun_Lens_Ma</td><td>3</td></tr> <tr><td>Gun_Name</td><td>FEG HT 300</td></tr> <tr><td>Image_Shift_x</td><td>0.0</td></tr> <tr><td>Image_Shift_y</td><td>0.0</td></tr> <tr><td>Indicated_Magnification</td><td>13000.0</td></tr> <tr><td>Mode</td><td>TEM UP SA Zoom Image</td></tr> <tr><td>Obj_Strength</td><td>87.685</td></tr> <tr><td>STEM_Camera_Length</td><td>0.0</td></tr> <tr><td>Spot</td><td>2</td></tr> <tr><td>Stage_Position_phi</td><td>0.0</td></tr> <tr><td>Stage_Position_theta</td><td>1.85</td></tr> <tr><td>Voltage</td><td>300000.0</td></tr> </tbody> </table>	Setup	Value	Actual_Magnification	17677.0	C1_Aperture	2000	C2_Aperture	150	Cmm	1.2	Drift_Strength	44.625	Extractor_Voltage	4400	Filter_Aperture	3.0	Filter_Dispersion	0.1	Filter_Drift_Tube	0.0	Filter_Mode	Spectroscopy	Filter_Prism_Shift	-0.0	Filter_Total_Energy_Loss	0.0	Gun_Lens_Ma	3	Gun_Name	FEG HT 300	Image_Shift_x	0.0	Image_Shift_y	0.0	Indicated_Magnification	13000.0	Mode	TEM UP SA Zoom Image	Obj_Strength	87.685	STEM_Camera_Length	0.0	Spot	2	Stage_Position_phi	0.0	Stage_Position_theta	1.85	Voltage	300000.0
Setup	Value																																																		
Actual_Magnification	17677.0																																																		
C1_Aperture	2000																																																		
C2_Aperture	150																																																		
Cmm	1.2																																																		
Drift_Strength	44.625																																																		
Extractor_Voltage	4400																																																		
Filter_Aperture	3.0																																																		
Filter_Dispersion	0.1																																																		
Filter_Drift_Tube	0.0																																																		
Filter_Mode	Spectroscopy																																																		
Filter_Prism_Shift	-0.0																																																		
Filter_Total_Energy_Loss	0.0																																																		
Gun_Lens_Ma	3																																																		
Gun_Name	FEG HT 300																																																		
Image_Shift_x	0.0																																																		
Image_Shift_y	0.0																																																		
Indicated_Magnification	13000.0																																																		
Mode	TEM UP SA Zoom Image																																																		
Obj_Strength	87.685																																																		
STEM_Camera_Length	0.0																																																		
Spot	2																																																		
Stage_Position_phi	0.0																																																		
Stage_Position_theta	1.85																																																		
Voltage	300000.0																																																		
<b>Activity 7</b> Mode: TEM UP SA Zoom Image																																																			
<b>Activity 8</b> Mode: TEM UP SA Zoom Diffraction																																																			
<b>Activity 9</b> Mode: TEM UP SA Zoom Image																																																			

XML

XSL Transformation

HTML Output



# Summary

- We have developed *NexusLIMS*
  - A system to automatically harvest experimental metadata and context to create records of individual microscopy sessions
- Based on structured data and XML transformations
  - Allows us to valid data structure using a schema
  - Transform from raw data to useful display
- Using NIST's CDCS as key infrastructure
  - Leveraging existing systems to accelerate materials knowledge

# Where are we headed?

- Initial deployment of *NexusLIMS*
  - Gathering feedback from users and tailoring to needs
  - Tweaking of schema and display; expanding to more tools



# Where are we headed?

- Initial deployment of *NexusLIMS*
  - Gathering feedback from users and tailoring to needs
  - Tweaking of schema and display; expanding to more tools
- Using *NexusLIMS* to enable research
  - Cloud-based analysis tools – close to the data
  - e.g. Machine learning and predictive analysis using large volumes of structured microscopy data

# Acknowledgements

## Guinea pigs (for data):

- Mike Katz
- Vladimir Oleshko

## HyperSpy and the scientific Python ecosystem:

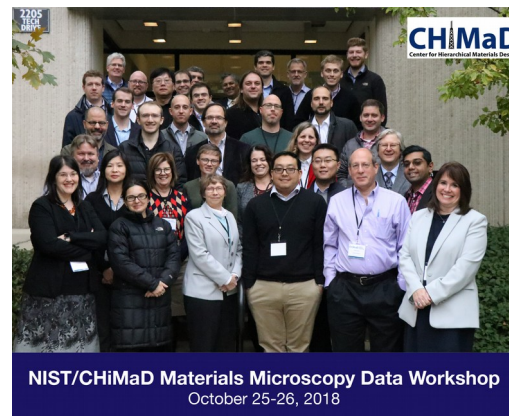
- Wonderful resources for projects such as this

## CDCS:

- CDCS is developed by NIST's Information Systems Group within the Information Technology Lab

## CHiMAD:

- Materials Microscopy Data conference/working groups:



A decorative pattern of white hexagons with blue outlines, arranged in a honeycomb-like structure, occupies the top portion of the slide.

# Thank you!

[joshua.taillon@nist.gov](mailto:joshua.taillon@nist.gov)



# CDCS as the LIMS backbone

## Materials Data Curation System

Welcome, jlau. Thanks for logging in.

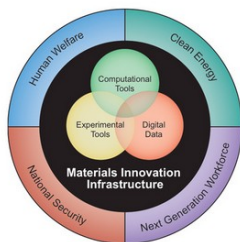
Logout | Dashboard | Help

Home | Data Curation | Data Exploration | Composer

## Materials Data Curator

This system allows for the curation of Material Data in a repository using predefined templates.

This is being developed at the National Institute of Standards and Technology and is made available to solicit comments from the Material Science community. Please do not enter any proprietary data into this system.



## Available Options

[All Options »](#)

### Curate your Materials Data

Click here to select a form template and then fill out the corresponding form.



### Explore the repository

Click here to search for Materials Data in the repository using flexible queries.



### Compose a template

Click here to compose your own template.

## Most Recent Templates

[Browse All »](#)

WebFF-WaterModel-Rigid-25SEPT18 | WebFF-FF-WaterModel-Flattened.xsd

Test-Strength | Strengthmodel\_3.xsd

EOS-Data | EOS\_Data.xsd

example | mod.demo.diffusion.xsd

Polymer-AM | Polymer-AM3.xsd

Test-Strength | strength model2.xsd

Polymer-AM | Polymer-AM2.xsd

## Data Entry

Here you can fill in the Materials Data form. Once it is completed, you can go to 'View Data' to review what you have entered. You won't be able to reach the review page before the document is valid according to the selected template. From the review page, you will be able to curate the data. The 'Save Form' button allows you to save partial data that you may want to edit later. This will only save a temporary document and won't actually curate data. All grayed elements are optional. Thus, all elements written in black are required. The document may still be valid with empty elements. There are no validation on empty fields if no such constraint is defined in the template. Thus, an empty string of characters may not raise a validation error, but an empty number will.

[Clear Fields](#) [Save Form](#) [Download](#)

### • collection

• Name

• Alternate Name (e.g. acronym, nickname)

• Identifier

• Keyword

• Description

• Date

• Access

• Associated Persons, Organizations, and Resources

• Associated Organization

• Associated Person

• Associated Resource

• Attribute

• Comment

• Digital Artifact (e.g. dataset, image)

• Subcollection

# The “Dossier”

**What we propose:** creating an electron microscopy session dossier

**Dossier-** “a collection or file of documents on the same subject, especially a complete file containing detailed information about a person or topic”

## **The Electron Microscopy Dossier:**

- A high-level summary view of the microscopy session
- Reunites an electron microscopy sessions’ data and context within a structured data model
- Minimal researcher intervention
  - Records are automatically generated using Python scripts
  - Yet, researcher has the freedom to correct records if they are wrong.
- Each session can be represented by a single XML record

## **We are using XML because:**

- Structured
- Machine readable
- Supports data validation (data integrity)
- XSL transformations enable creation of human readable records

