Harvesting Microscopy Experimental Context with a Configurable Laboratory Information Management System

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

NexusLIMS

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

(1) PennState Materials Science and Engineering

(2) University of Maryland 18 56 Electrical and Computer Engineering

MATERIAL MEASUREMENT LABORATORY

Summer Undergraduate Research Fellowship

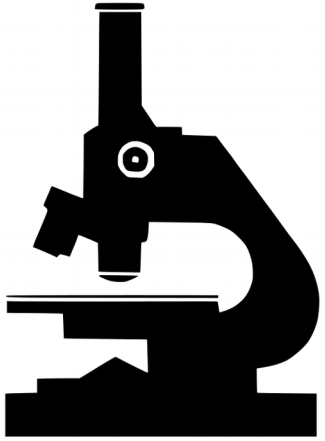
- **When?**
  - Applications due in February of each year
  - Program runs from late May to early August
- **Where?**
  - Gaithersburg, MD or Boulder, CO – *US Citizens only*
- **How much?**
  - ~$500 per week and allowance for lodging and travel
- **How do I apply/get more info?**
  - https://www.nist.gov/summer-undergraduate-research-fellowship-surf/

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

Acquire



Analyze

Store

An example EM data lifecycle for an individual user

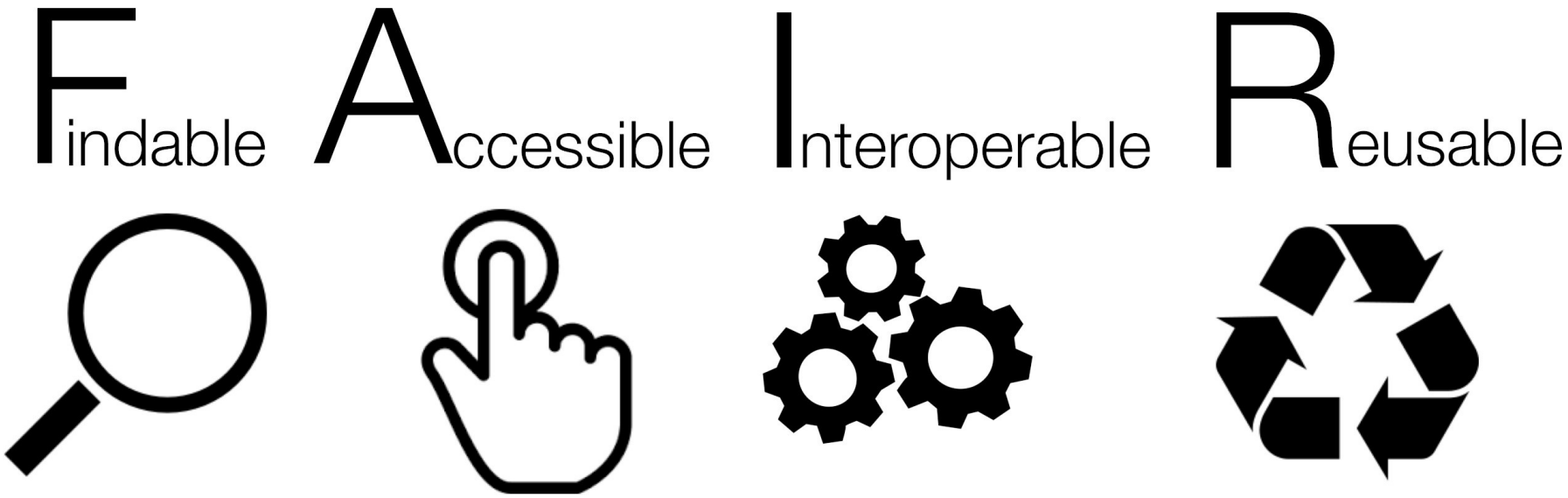
Morrow, 2009

Read, 2010

McMorran, 2011

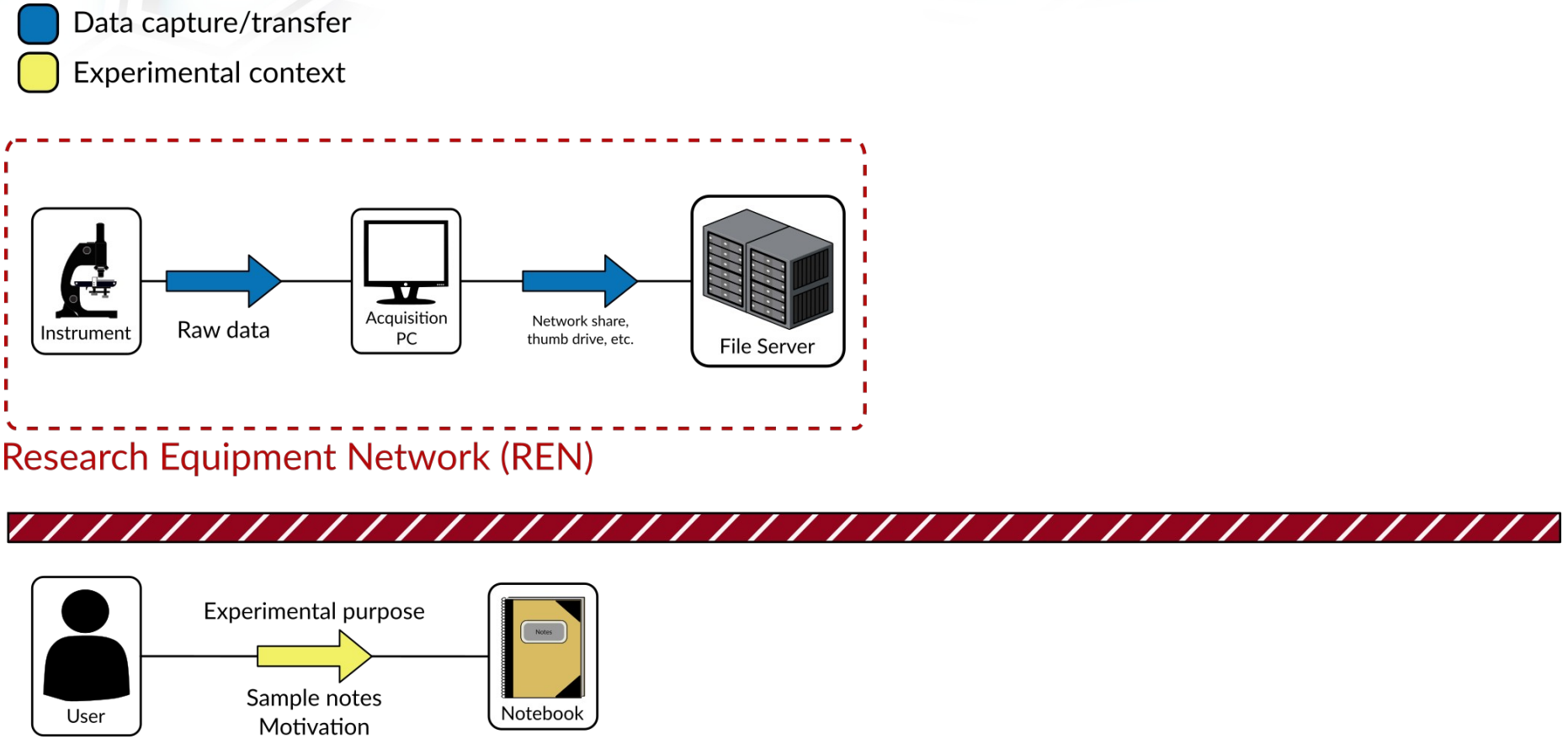
Idell, 2016

Leads to UnFAIR data

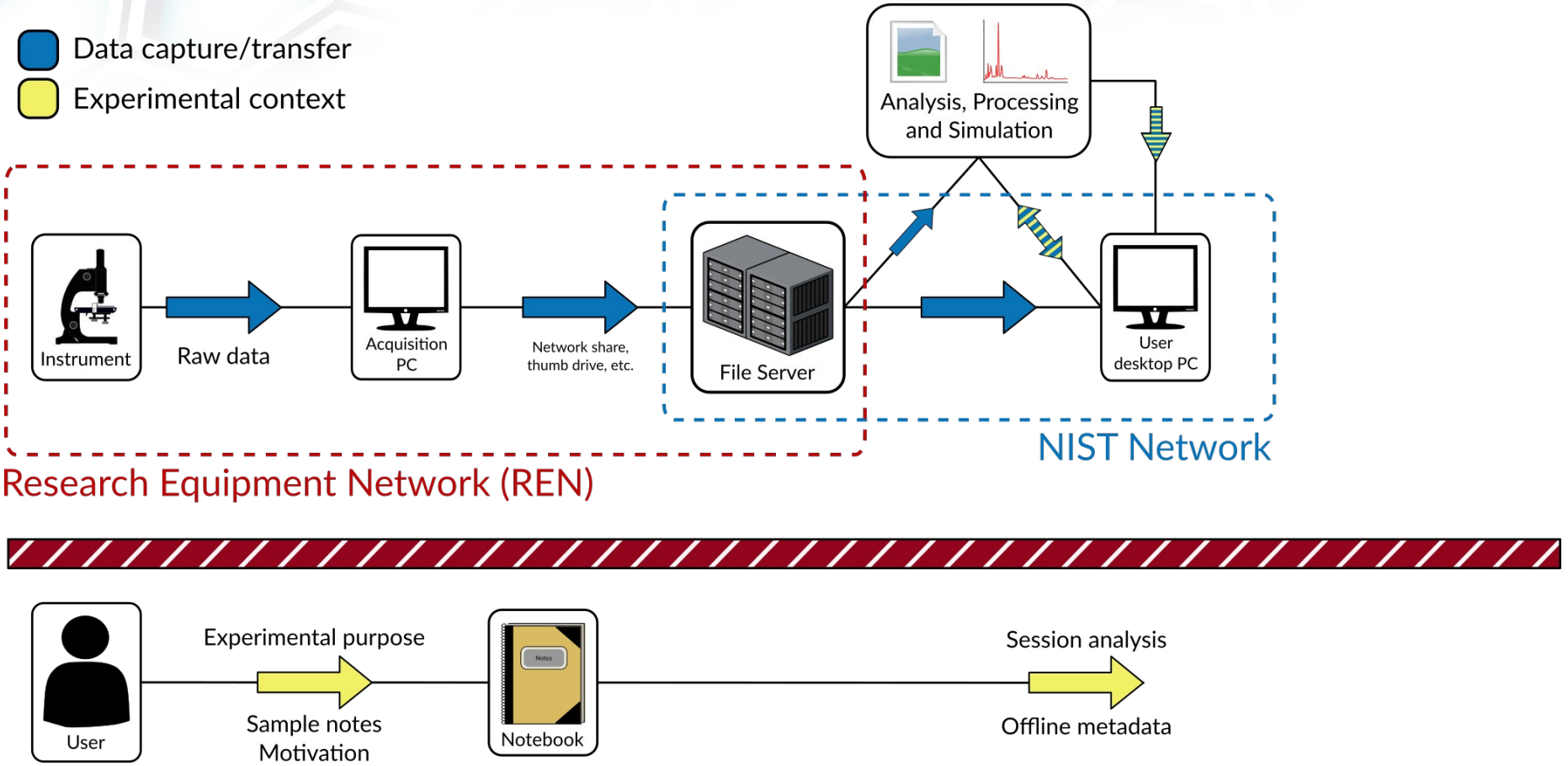


Wilkinson et al., Scientific Data, **3**, 160018, 2016 (link)
Image: Sangya Pundir - CC-BY-SA 4.0

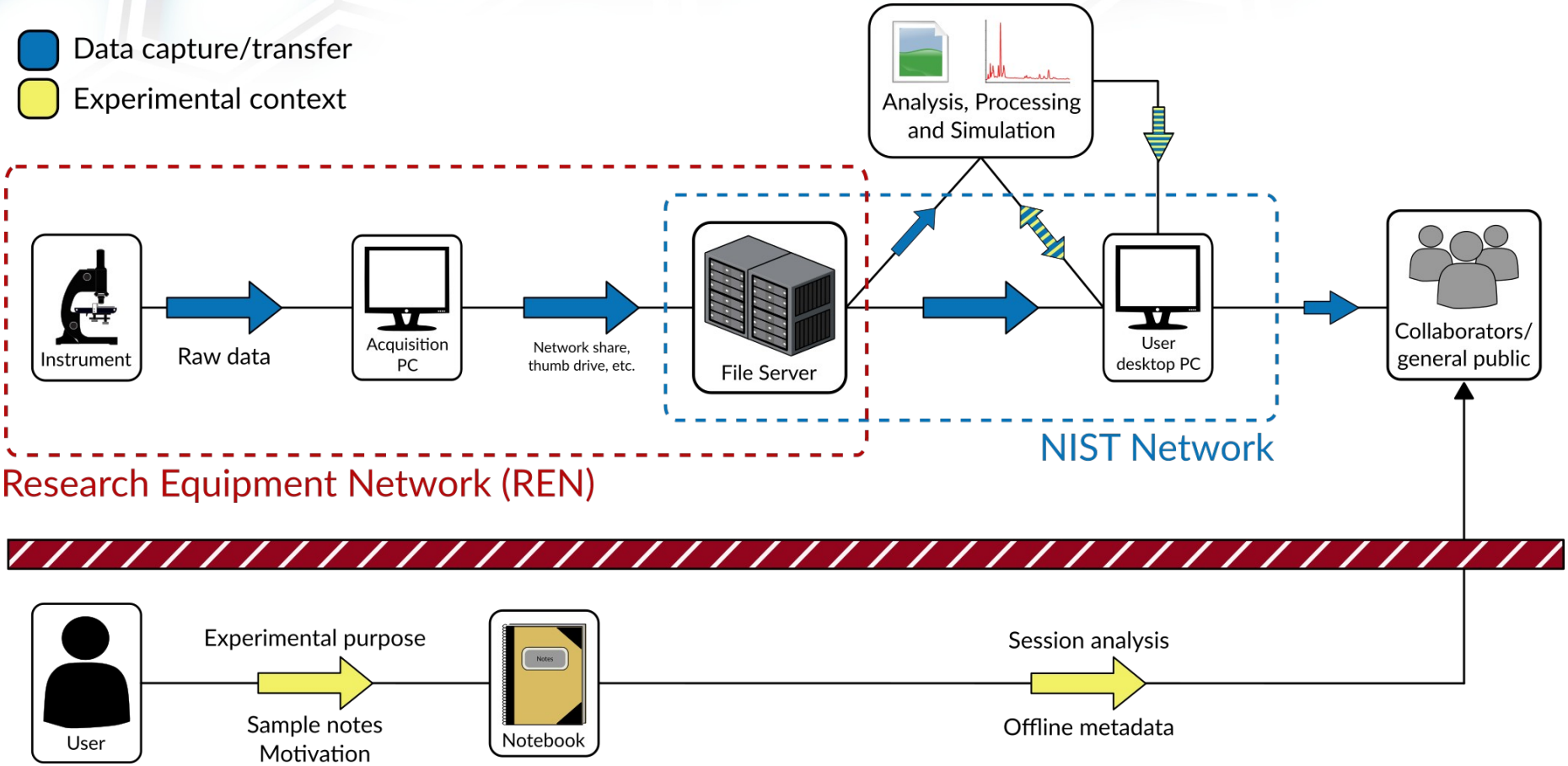
The *status quo* of facility-level EM Data Management (at NIST)



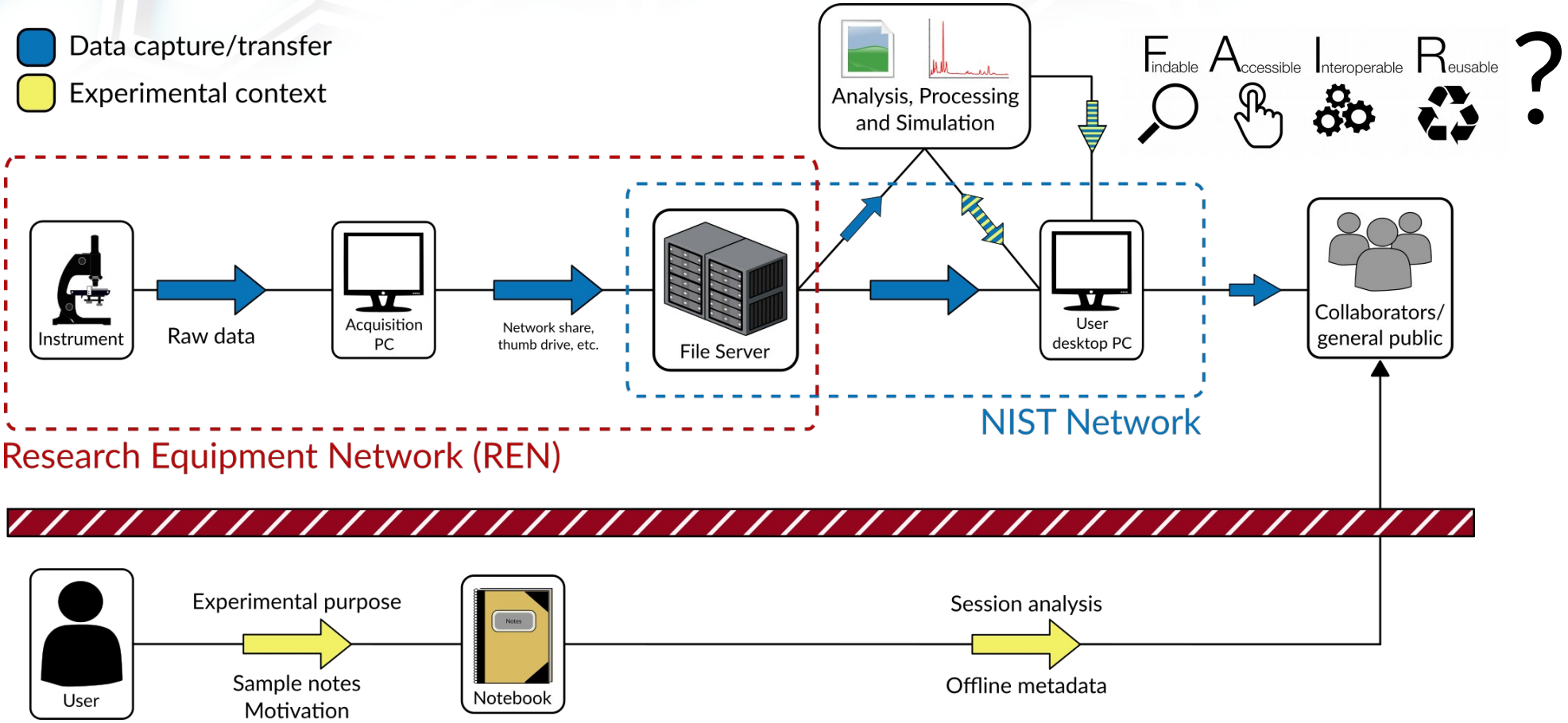
The *status quo* of facility-level EM Data Management (at NIST)



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

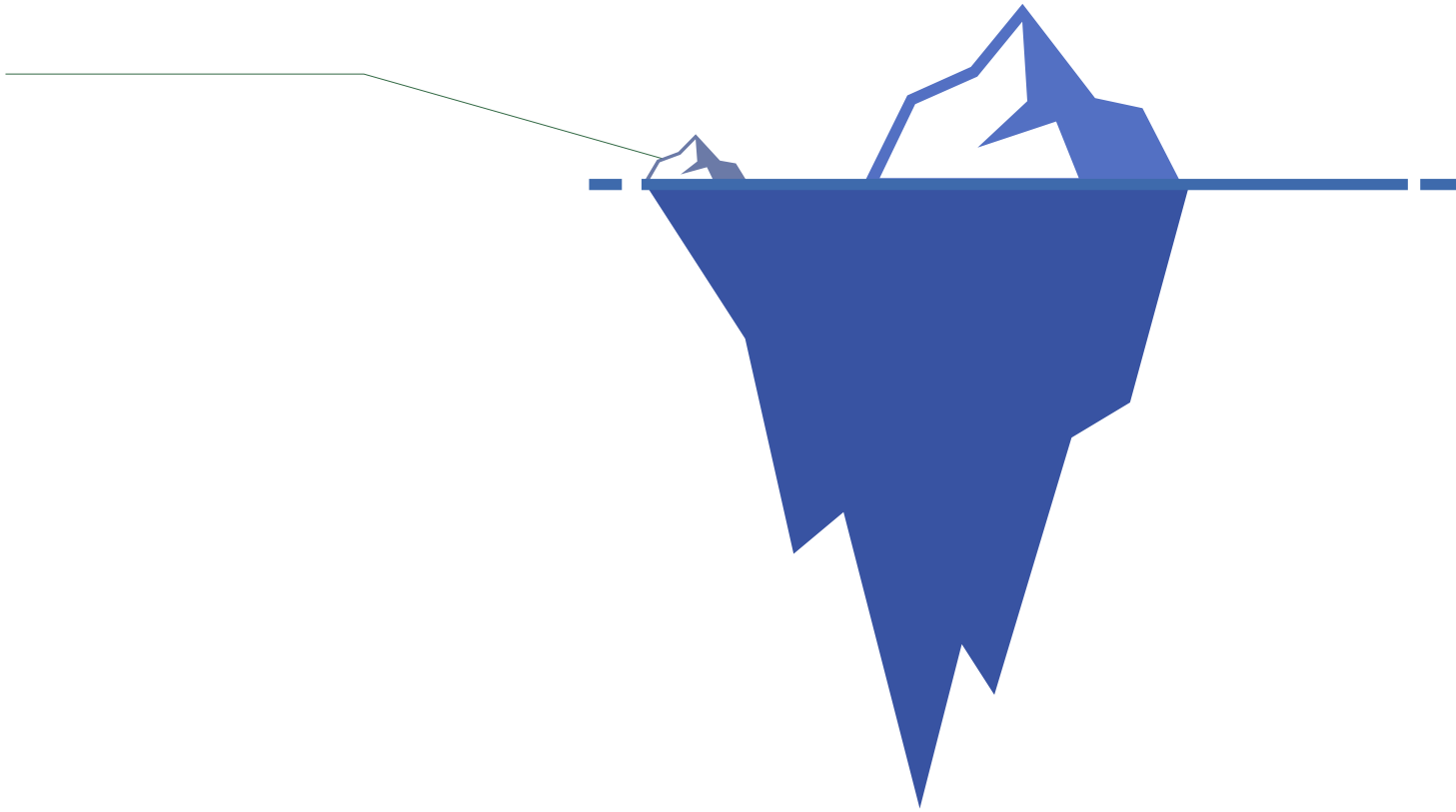


The *status quo* of facility-level EM Data Management (at NIST)



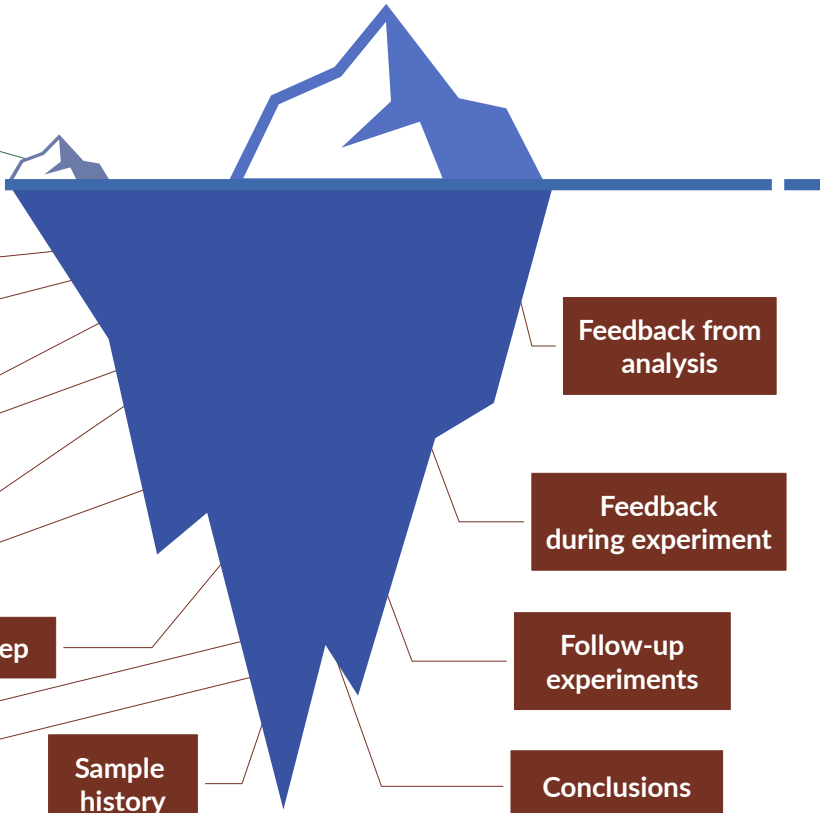
The challenge of contextual data

"Easy" Metadata
(who, what, when, where)

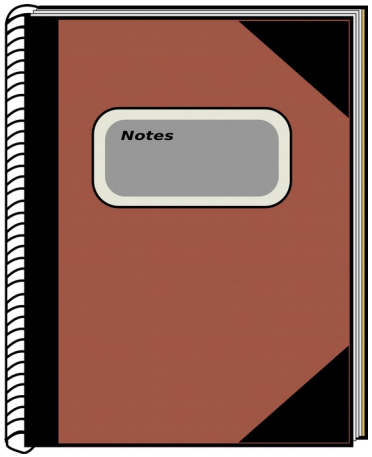


The challenge of contextual data

"Easy" Metadata
(who, what, when, where)

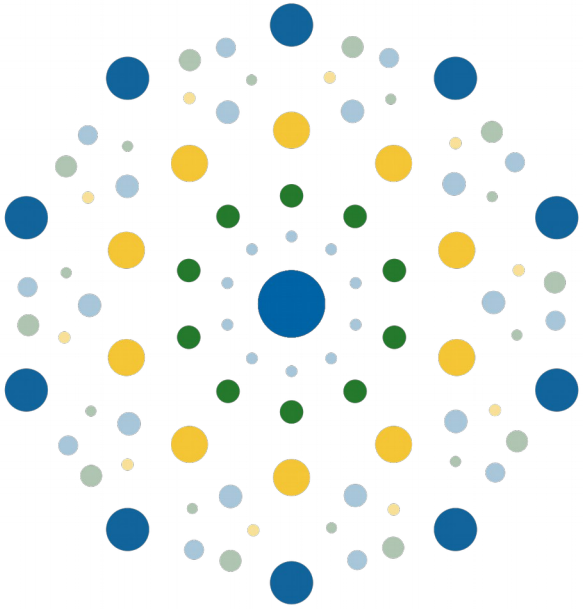


The "Why"



- Why are we doing the experiment?
- User proposal
- Instrument calibration
- Instrument setup
- Tool status
- Sample composition
- Sample prep
- Reservation information
- Background documents
- Sample history
- Feedback from analysis
- Feedback during experiment
- Follow-up experiments
- Conclusions

Our Goal – A LIMS for materials microscopy



Laboratory

Information

Management

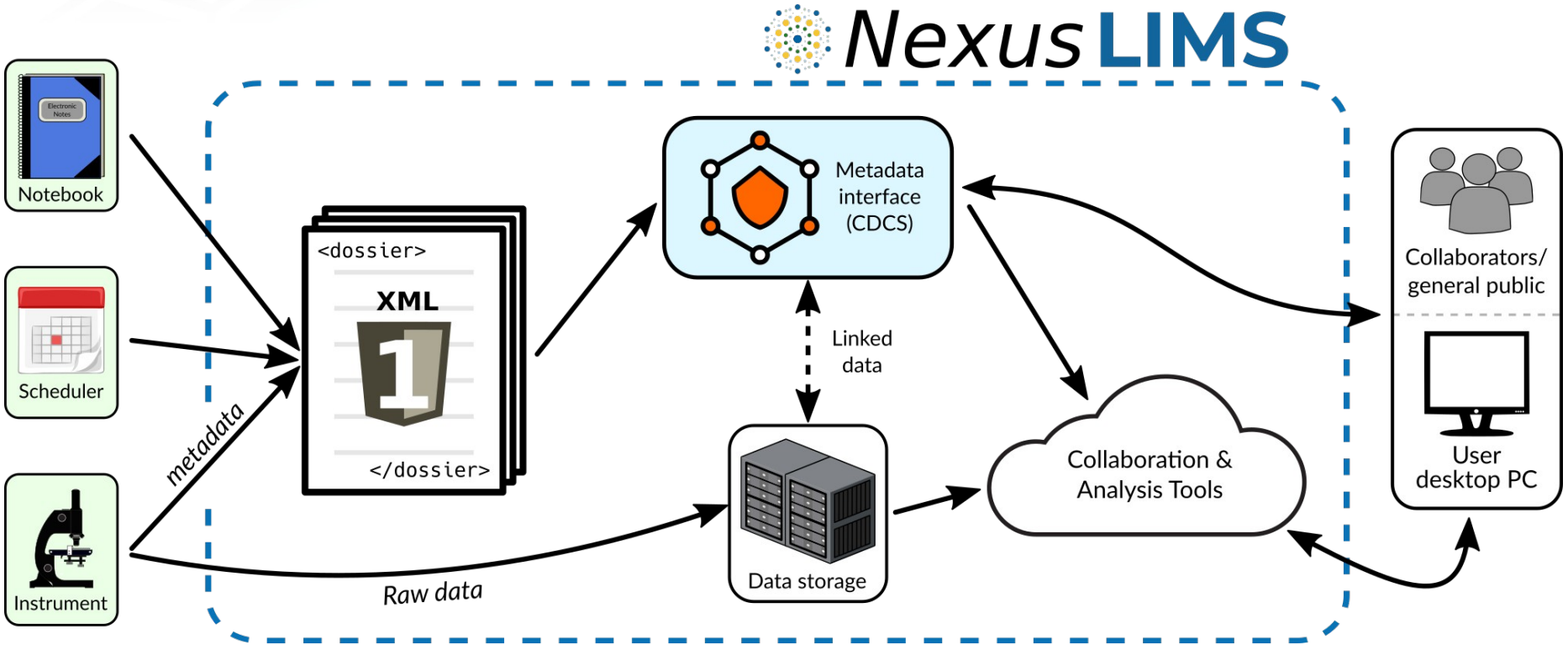
System

Our Goal – A LIMS for materials microscopy

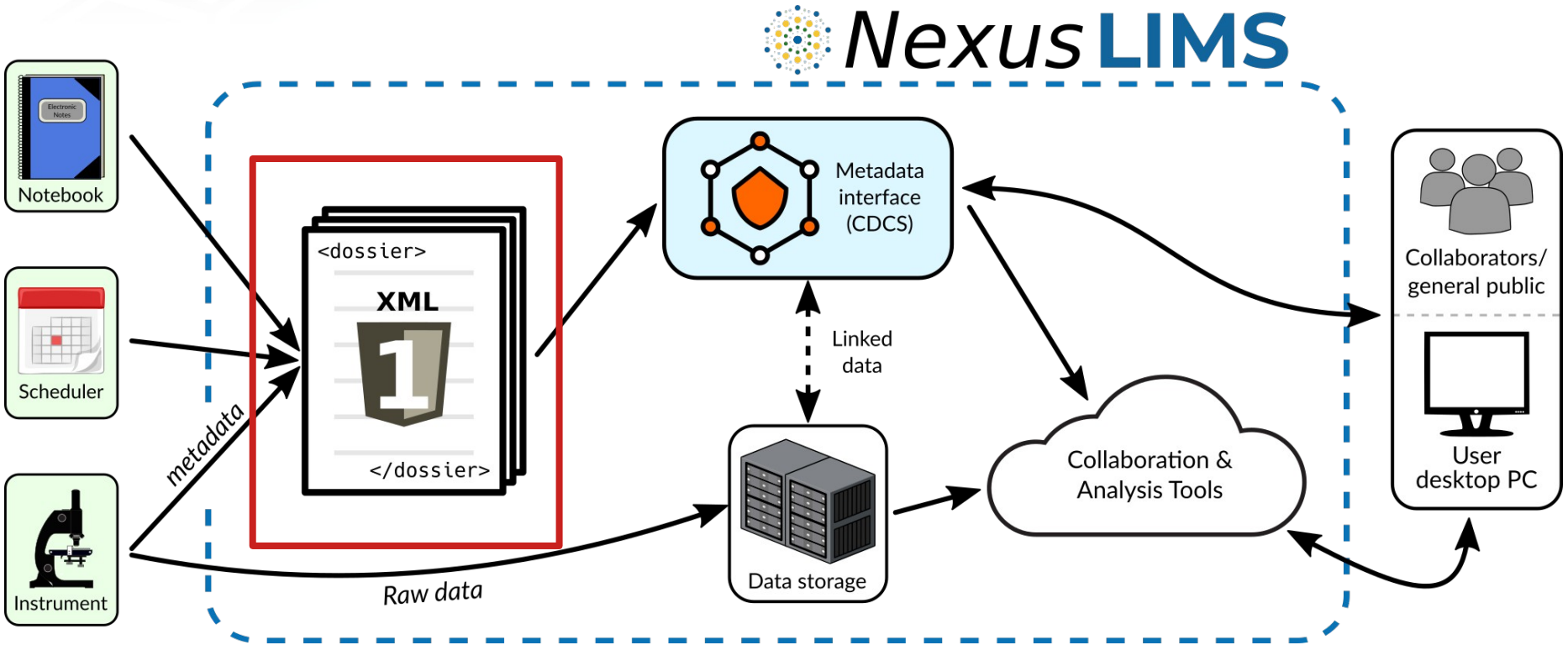
Making it simple to capture the "why" of an experiment

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*

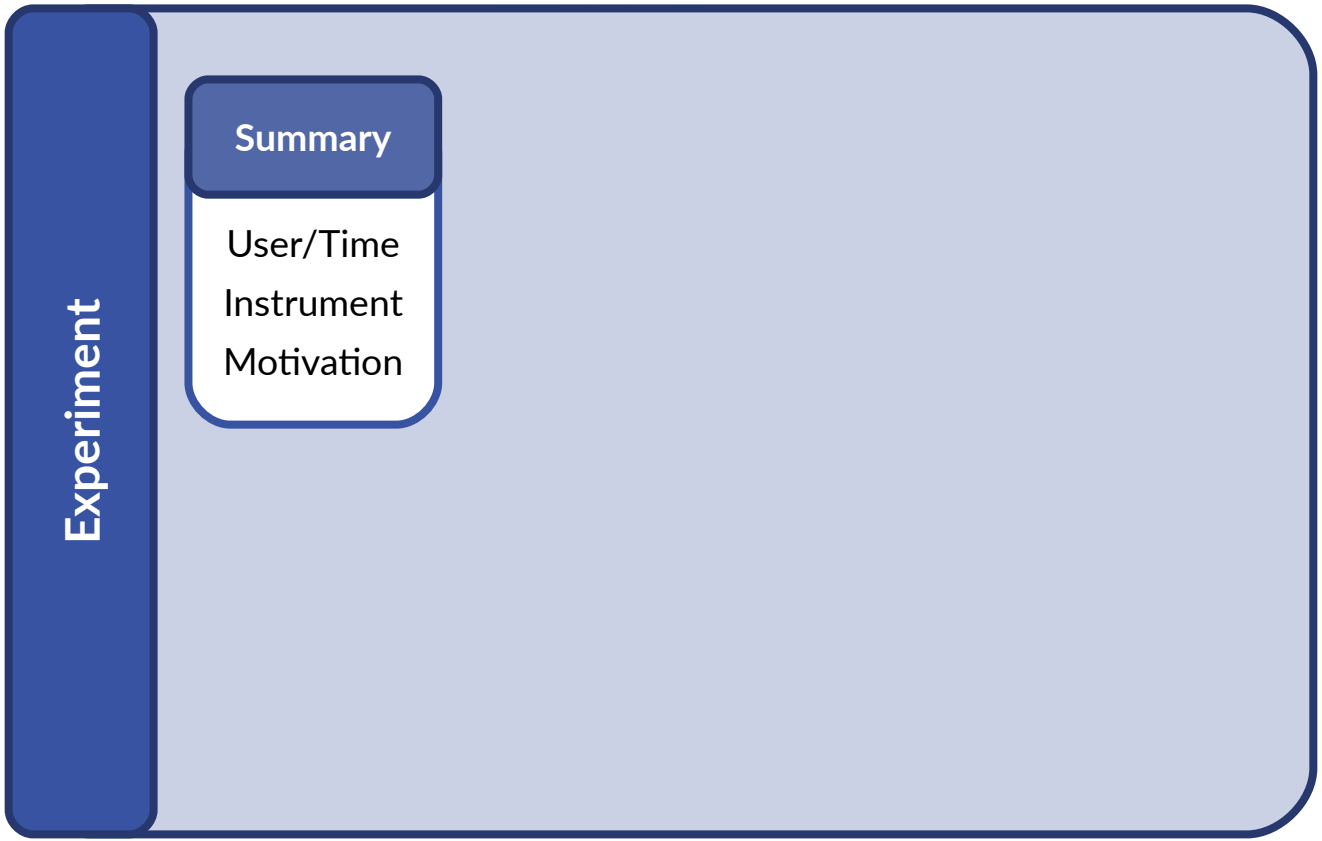
- Data is most useful when intelligently structured
  - Allows browsing, querying, transforming, validating, etc.
- Structure should be tailored to context
  - What information could a researcher/manager/auditor want to see?

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

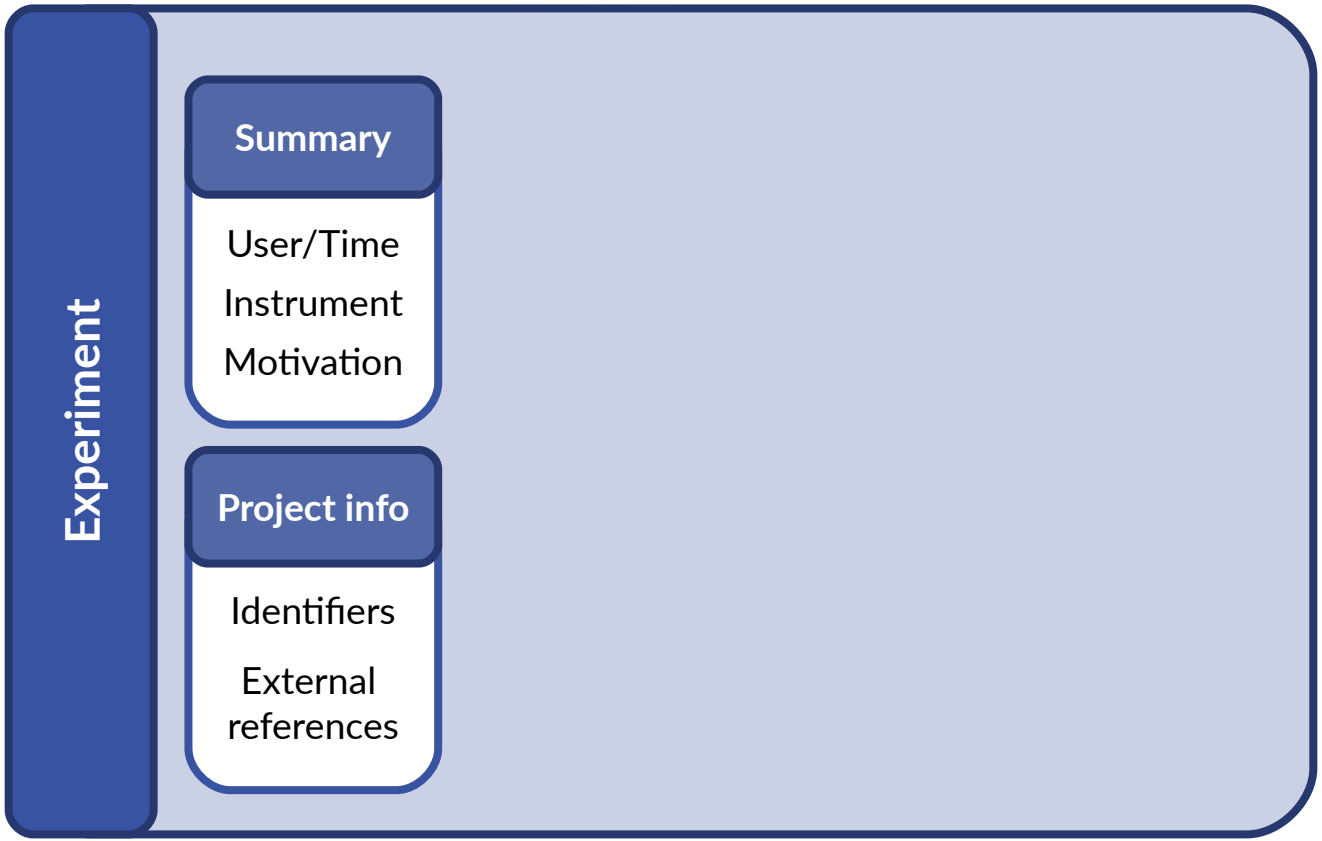
1. A schema for materials microscopy – *conceptual*

- 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



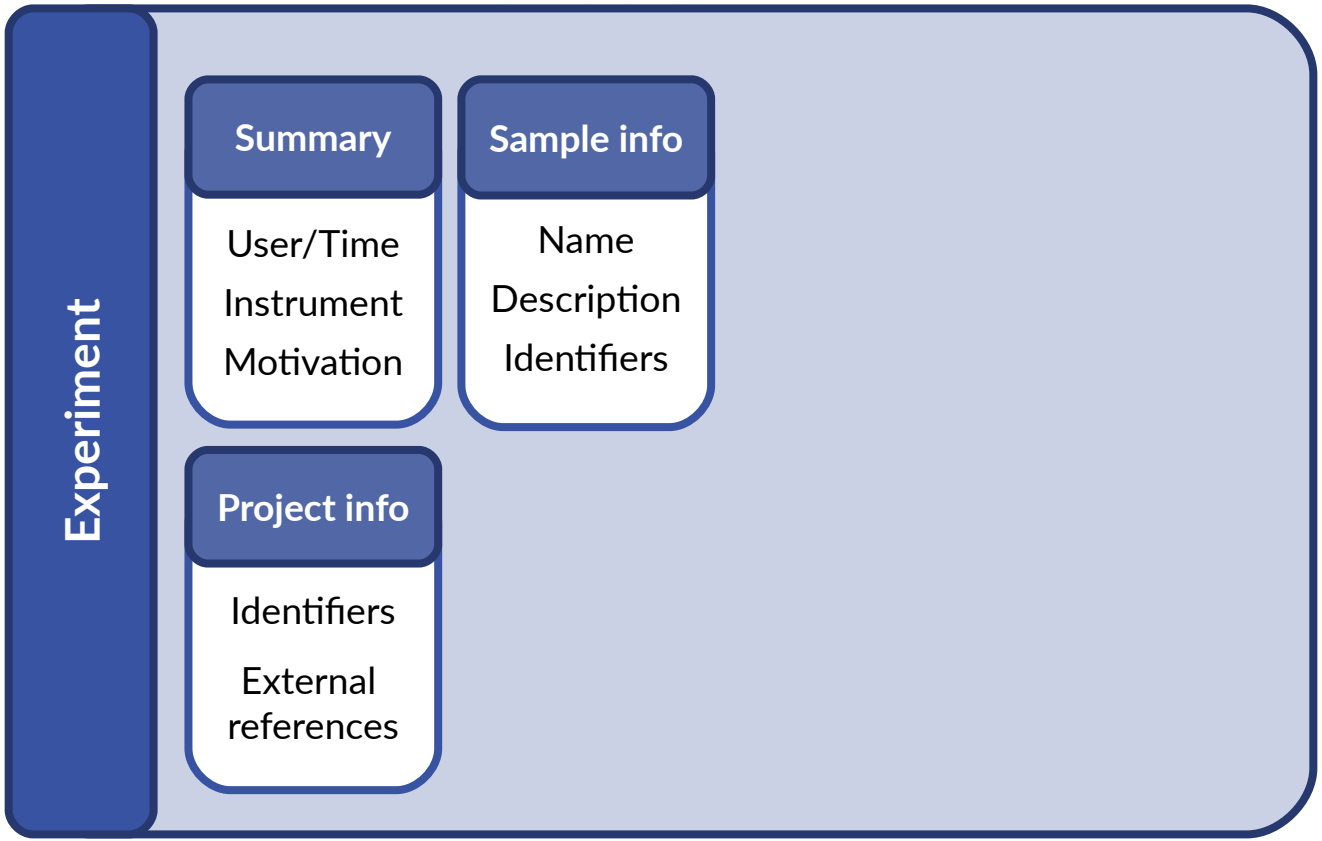
1. A schema for materials microscopy – *conceptual*

- 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



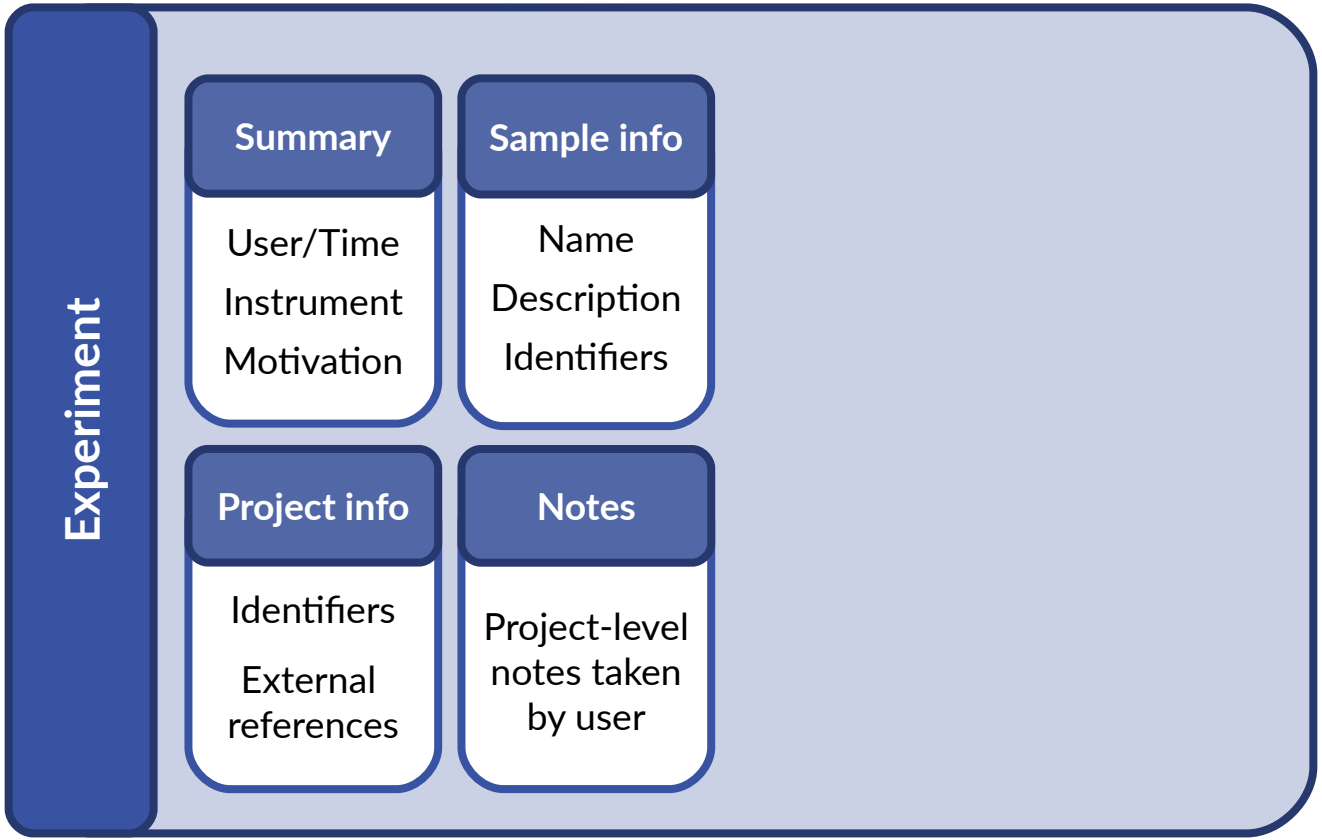
1. A schema for materials microscopy – *conceptual*

- 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



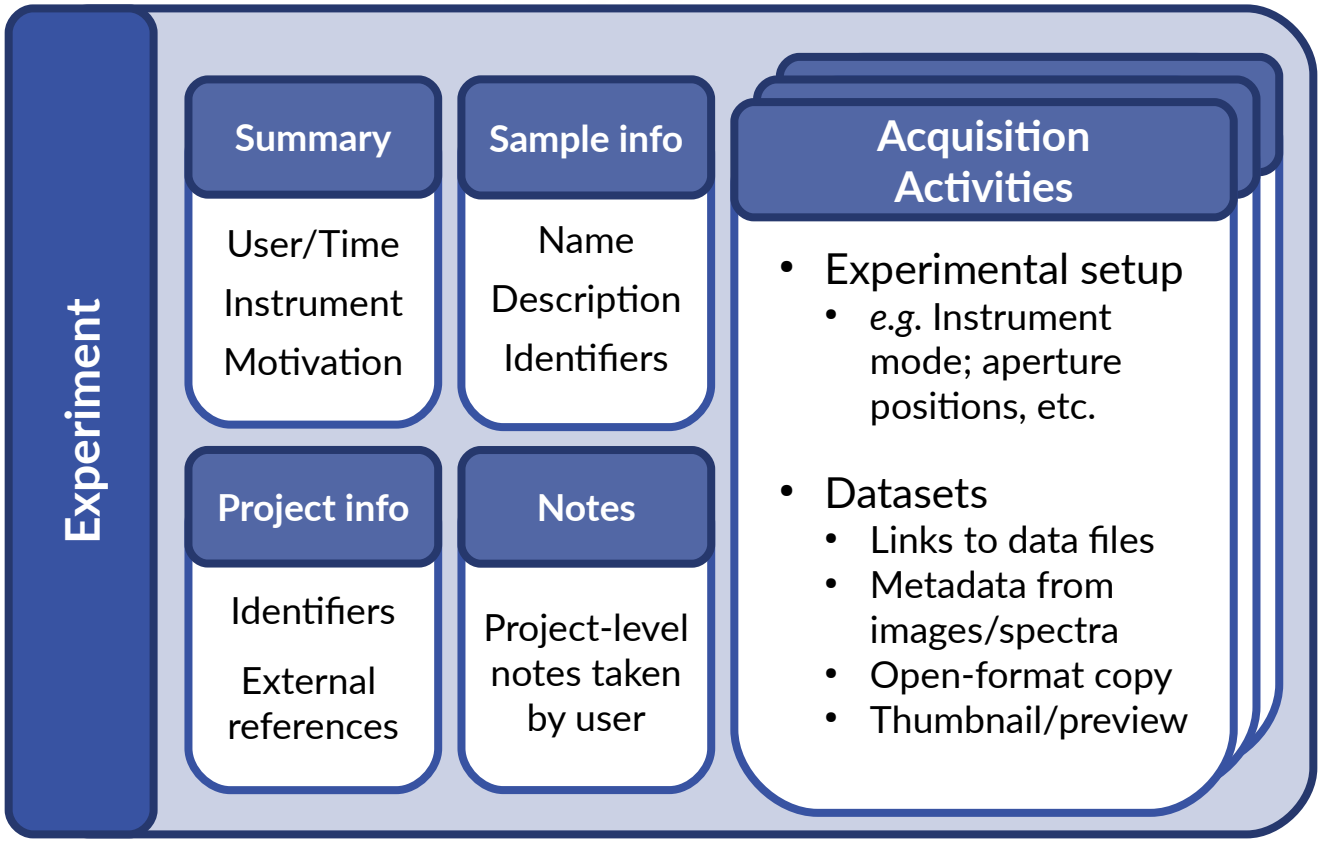
1. A schema for materials microscopy – *conceptual*

- 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



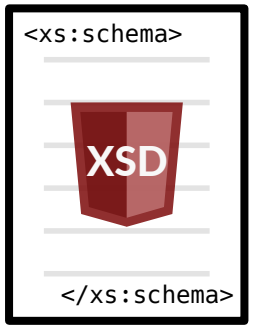
1. A schema for materials microscopy – *conceptual*

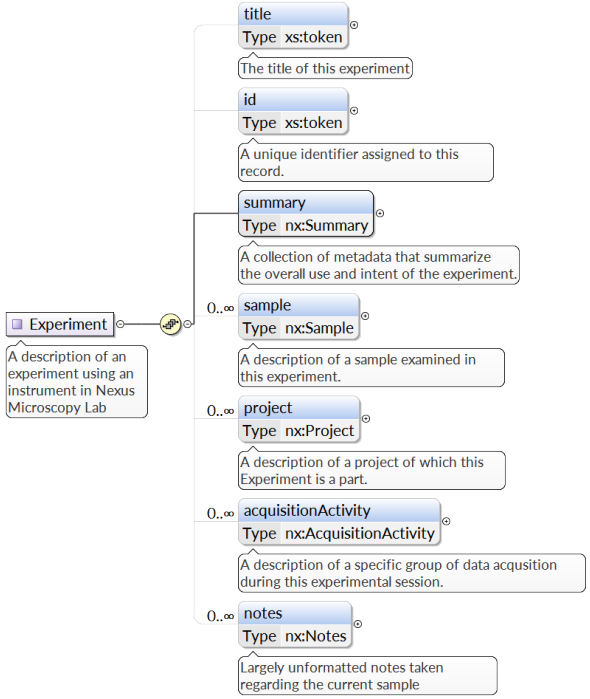
- Data is most useful when intelligently structured
  - Allows browsing, querying, transforming, validating, etc.
- Structure should be tailored to context
  - What information could a researcher/manager/auditor want to see?
- Schema instance represents a session on the microscope



1. A schema for materials microscopy – *technical*

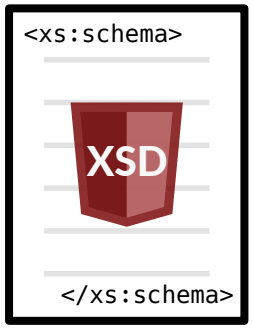
- **XML schema:**
  - Structured
  - Machine readable
  - Supports data validation (data integrity)

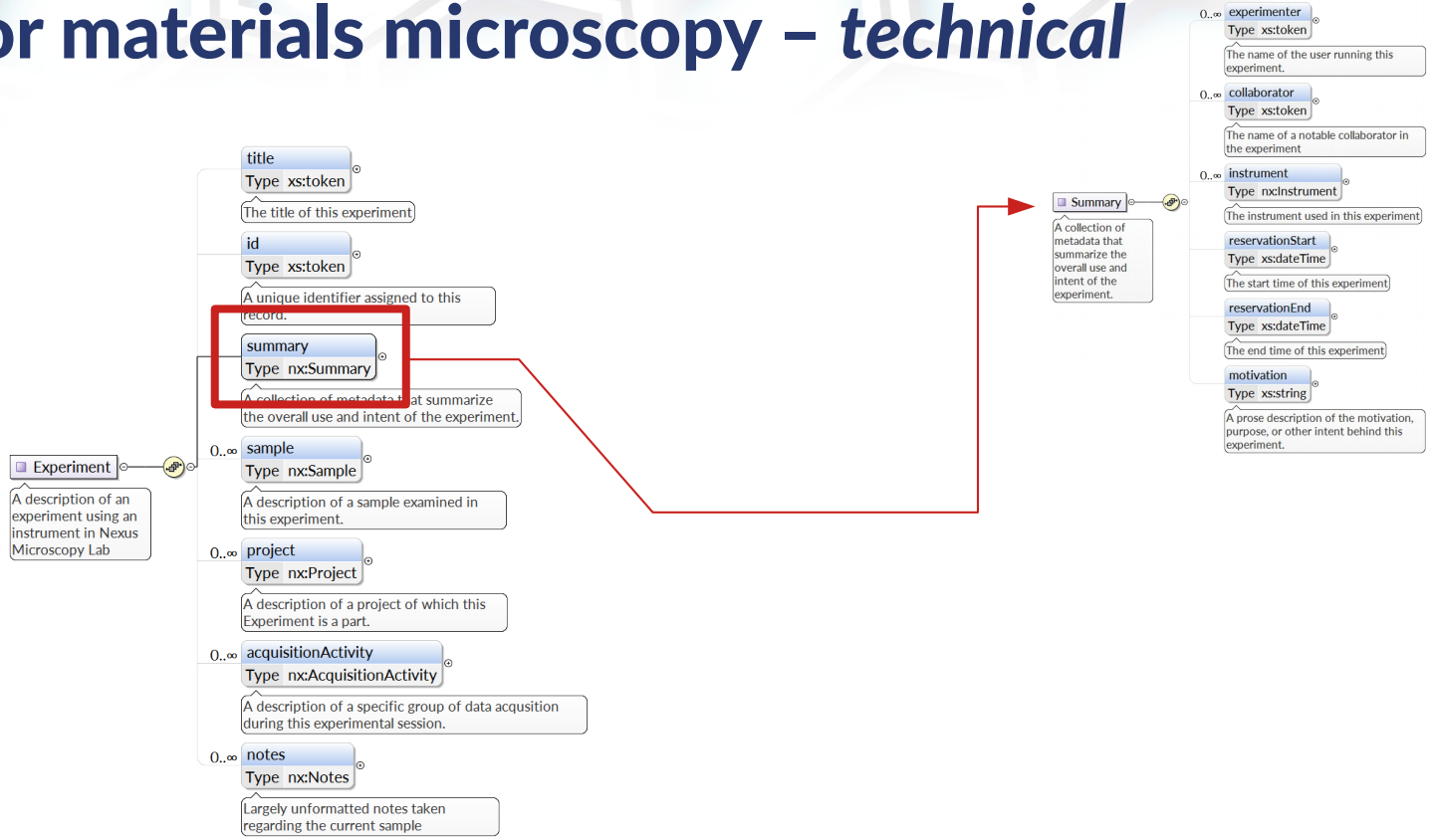




1. A schema for materials microscopy – *technical*

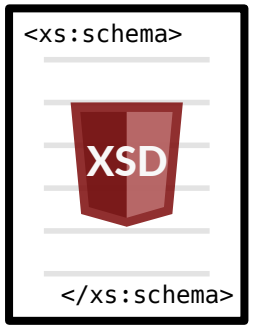
- **XML schema:**
  - Structured
  - Machine readable
  - Supports data validation (data integrity)

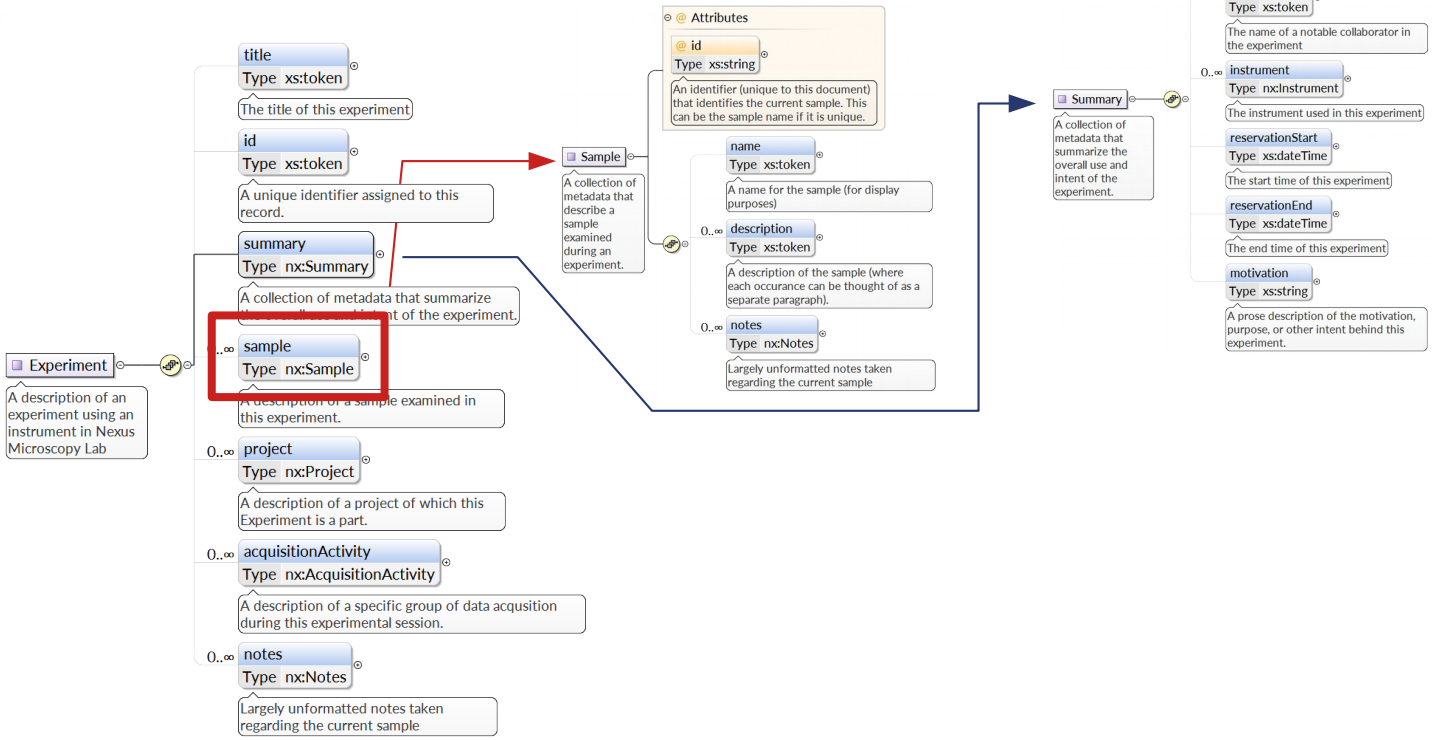




1. A schema for materials microscopy – *technical*

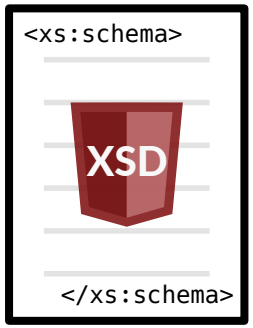
- **XML schema:**
  - Structured
  - Machine readable
  - Supports data validation (data integrity)

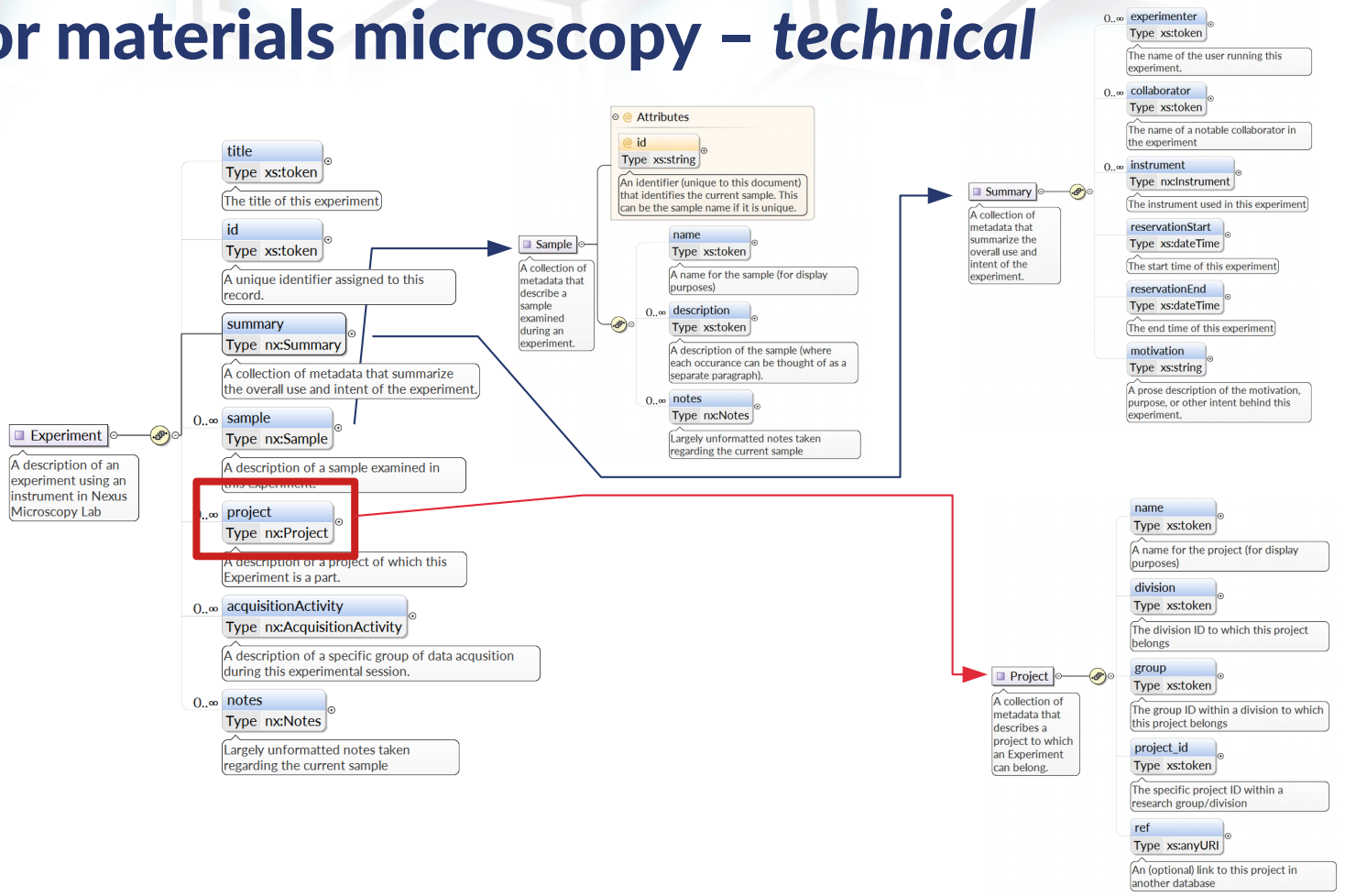




1. A schema for materials microscopy – *technical*

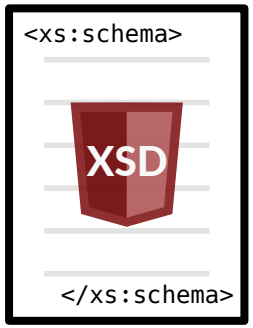
- **XML schema:**
  - Structured
  - Machine readable
  - Supports data validation (data integrity)

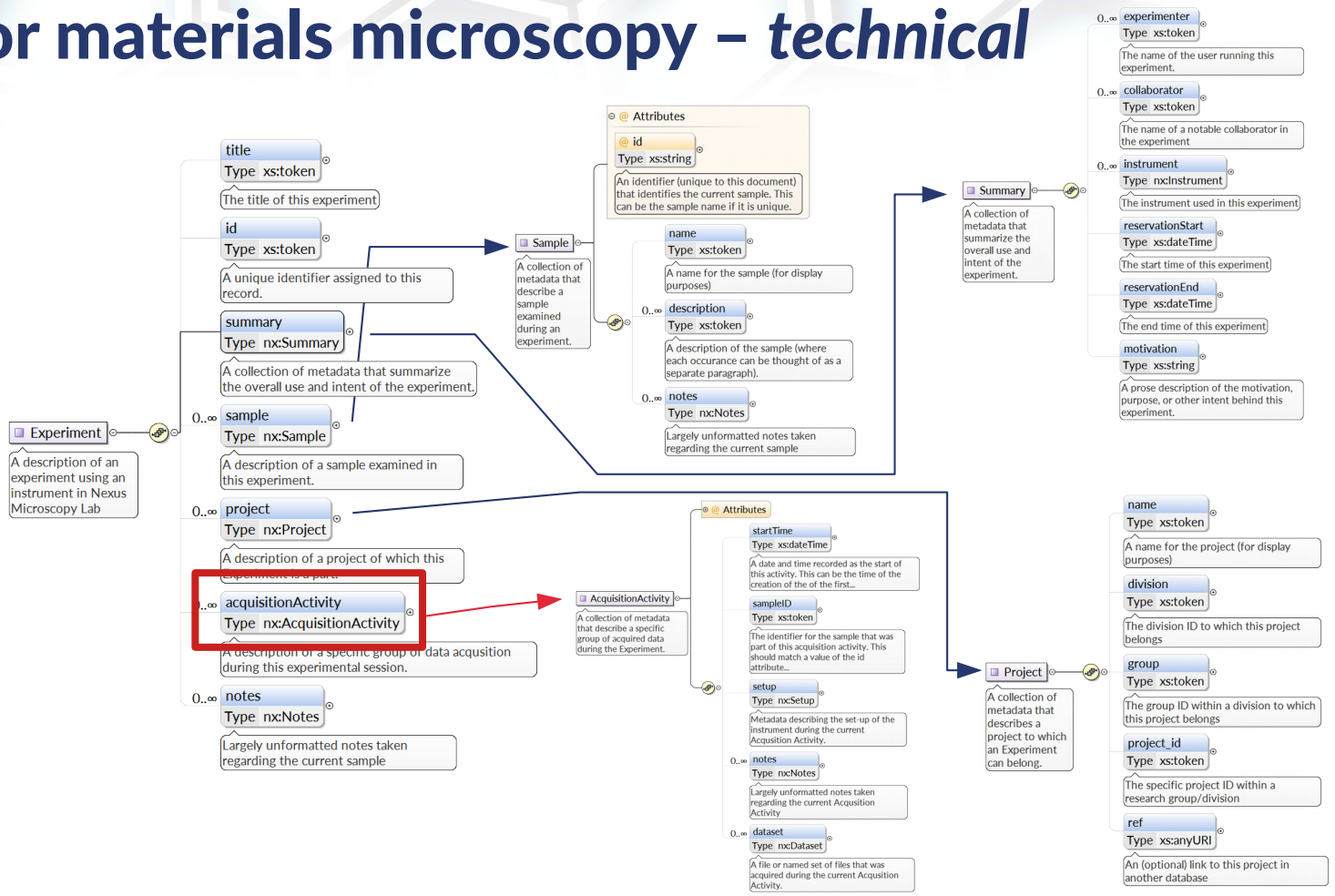




1. A schema for materials microscopy – *technical*

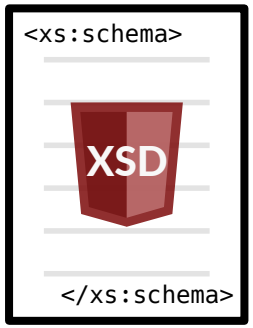
- **XML schema:**
  - Structured
  - Machine readable
  - Supports data validation (data integrity)

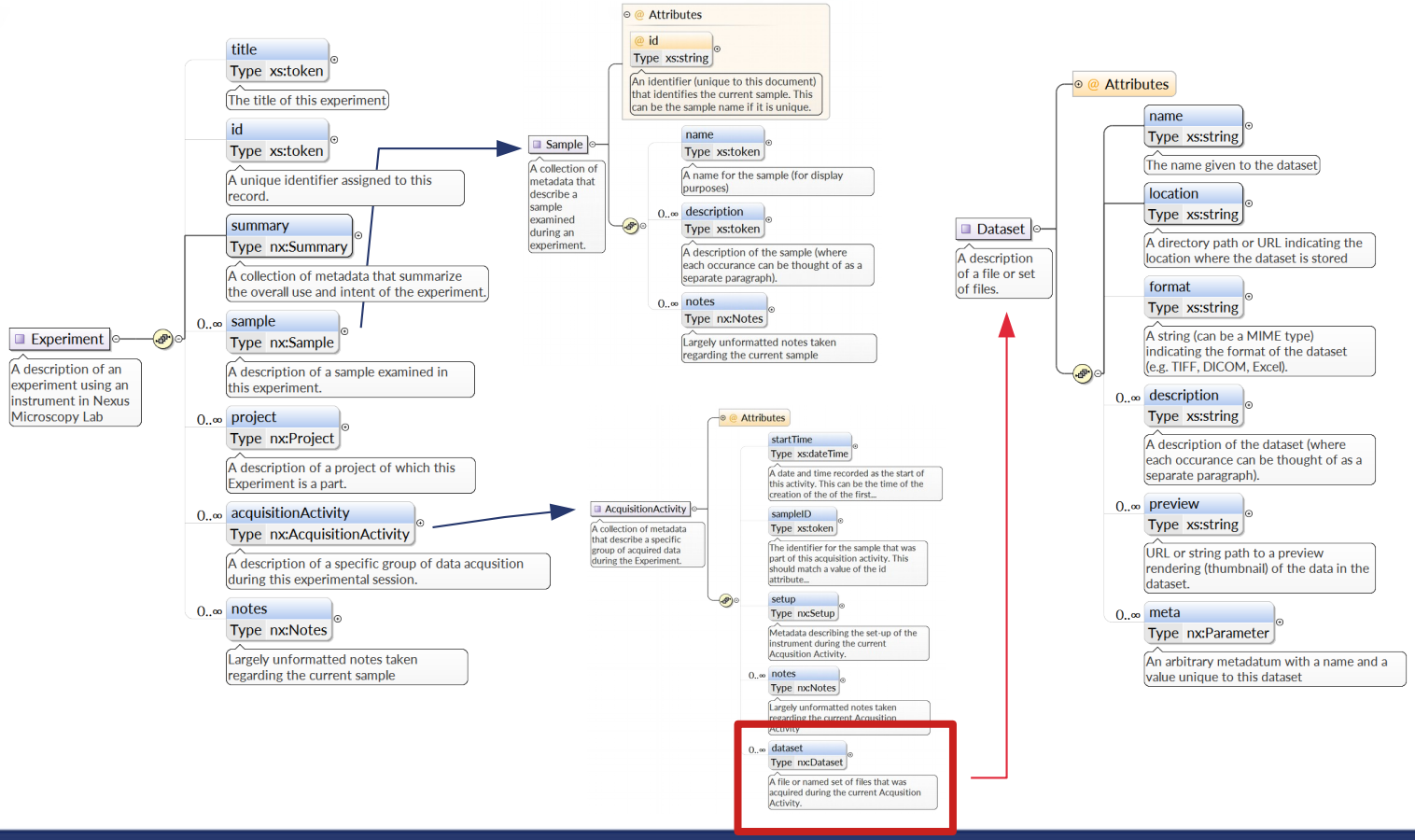




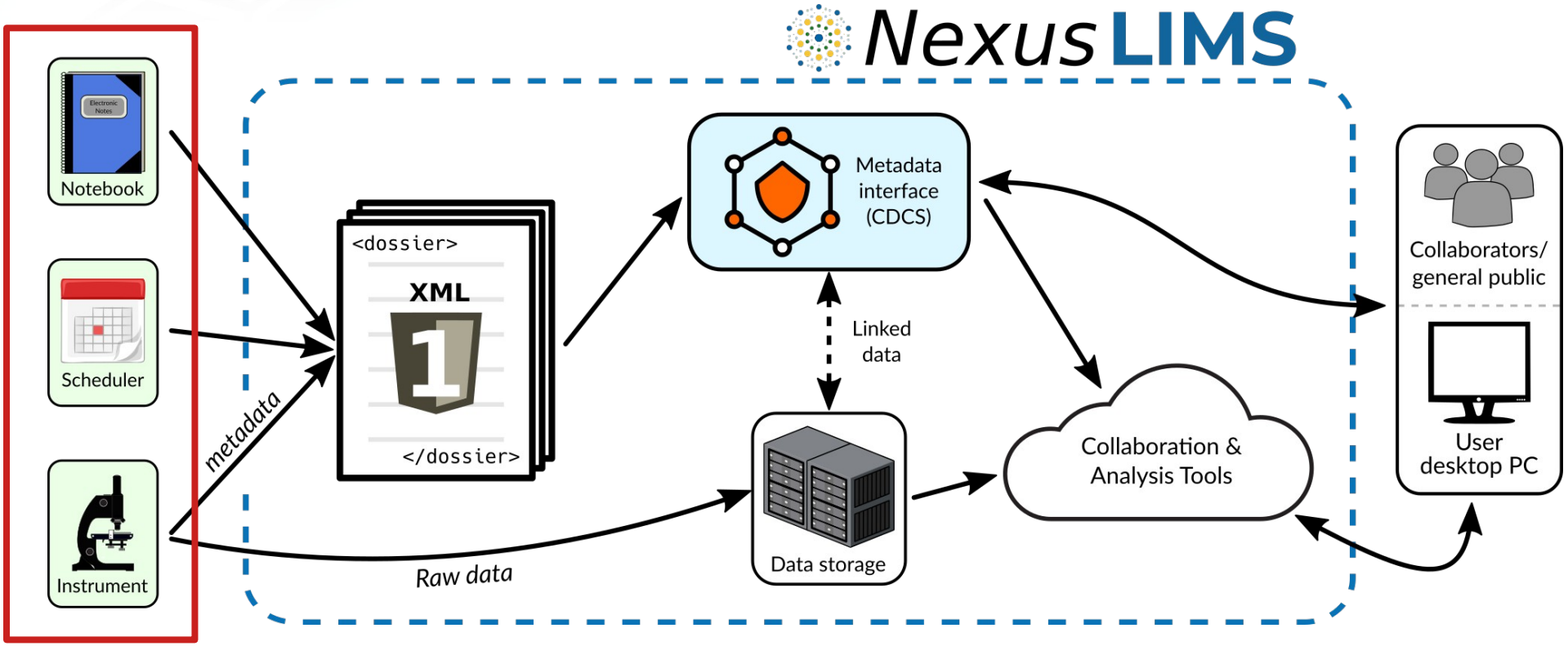
1. A schema for materials microscopy – *technical*

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

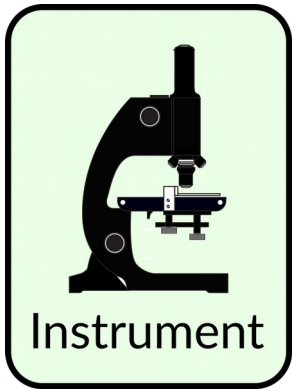




How do we accomplish this?



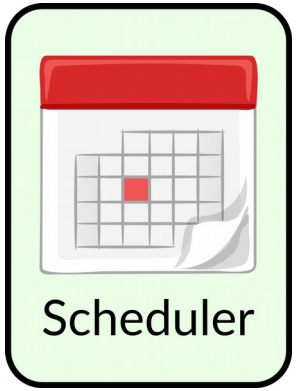
2. Harvesting and extracting metadata

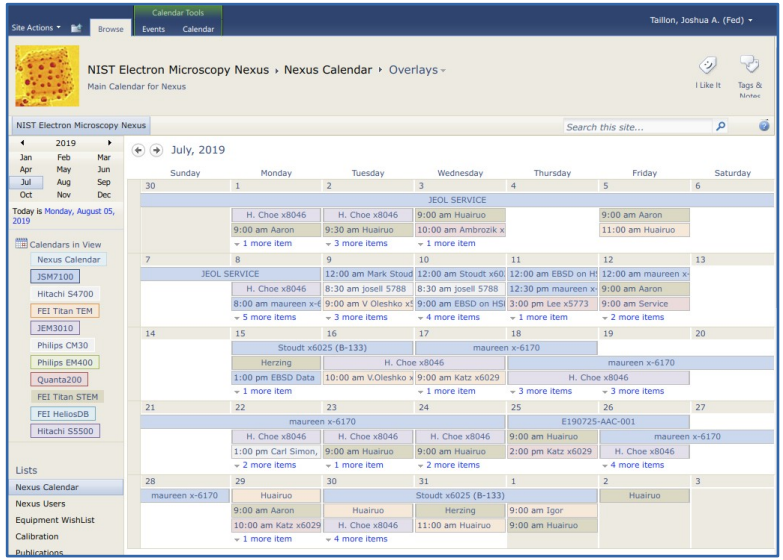


HyperSpy
multi-dimensional data analysis

Instrument and image/spectral metadata
from central file storage

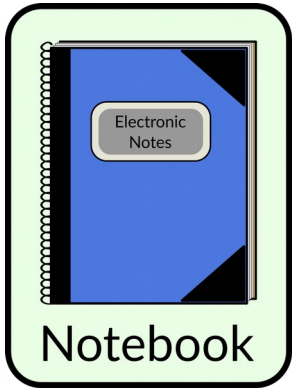
2. Harvesting and extracting metadata

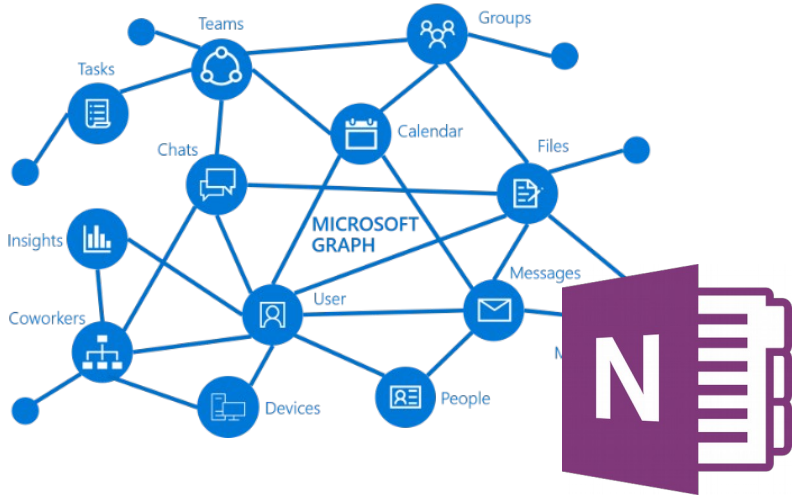




User and session information from SharePoint calendar

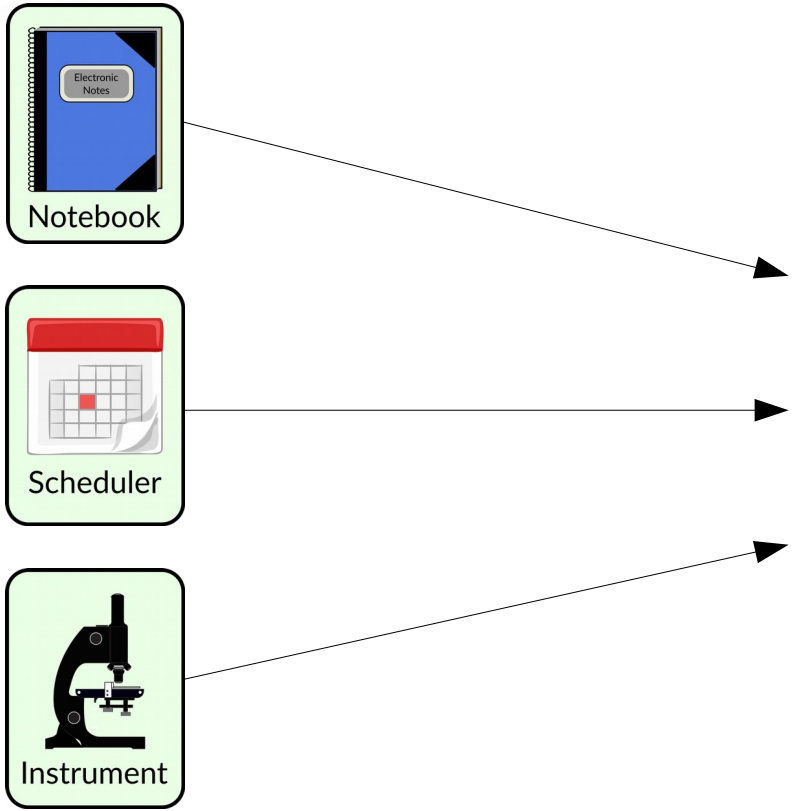
2. Harvesting and extracting metadata





Users' freeform and/or structured notes

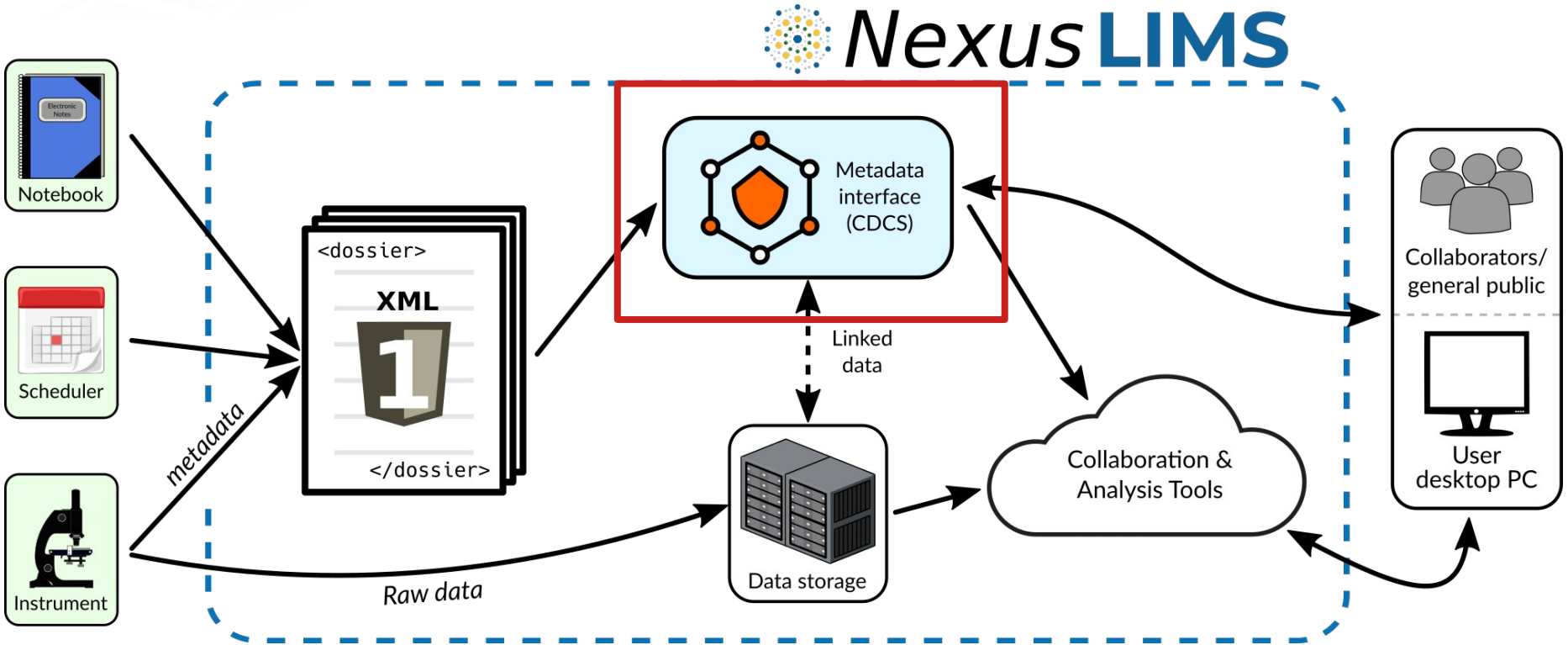
2. Harvesting and extracting metadata




```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="../xslStylesheet.xsl"?>
<nx:Experiment xmlns=""
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:nx="https://data.nist.gov/od/dm/nexus/experiment/v1.0">
    <event>
        <dateSearched>2018-11-13</dateSearched>
        <userSearched>mbk1</userSearched>
        <title>Katz x6029</title>
        <instrument>FEITitanTEMEvents</instrument>
        <user>
            <userName>mbk1</userName>
            <name>Katz, Michael B. (Fed)</name>
            <email>michael.katz@nist.gov</email>
            <phone>3019756029</phone>
            <office>223/A163</office>
            <link>https://share.nist.gov/sites/MML/Div/msed/MSED-MMF/_vti_bin/Li
            <userId>2815</userId>
        </user>
        <purpose>Trying to find us some martensite!</purpose>
        <sampleDetails>AM 17-4</sampleDetails>
        <description/>
        <startTime>2018-11-13T09:00:00</startTime>
        <endTime>2018-11-13T16:00:00</endTime>
        <link>https://share.nist.gov/sites/MML/Div/msed/MSED-MMF/_vti_bin/ListDa
        <eventId>470</eventId>
    </event>
    <acquisitionActivity seqno="0">
    <startTime>2018-11-13T13:01:28.179682</startTime>
    <sampleID>f81d3518-10af-4fab-9bd3-cfa2b0aea807</sampleID>
    <setup>
      <param name="Actual_Magnification">17677.0</param>
      <param name="C1_Aperture">2000</param>
      <param name="C2_Aperture">150</param>
      <param name="Csmm">1.2</param>
      <param name="Dif_Strength">44.625</param>
      <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

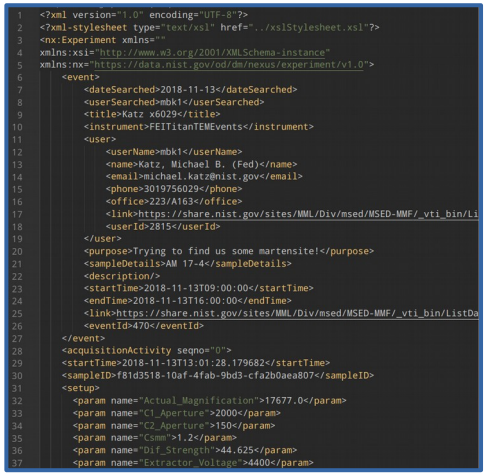
Dima, A., Bhaskarla, S., Becker, C. *et al*. JOM (2016) 68: 2053. https://doi.org/10.1007/s11837-016-2000-4

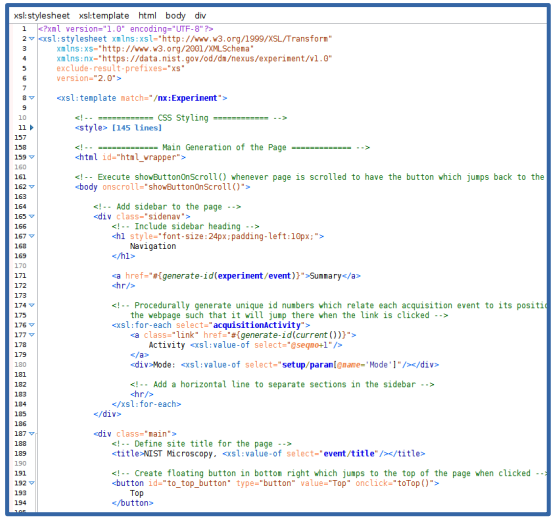
https://www.nist.gov/itl/ssd/information-systems-group/configurable-data-curation-system-cdcs/about-cdcs

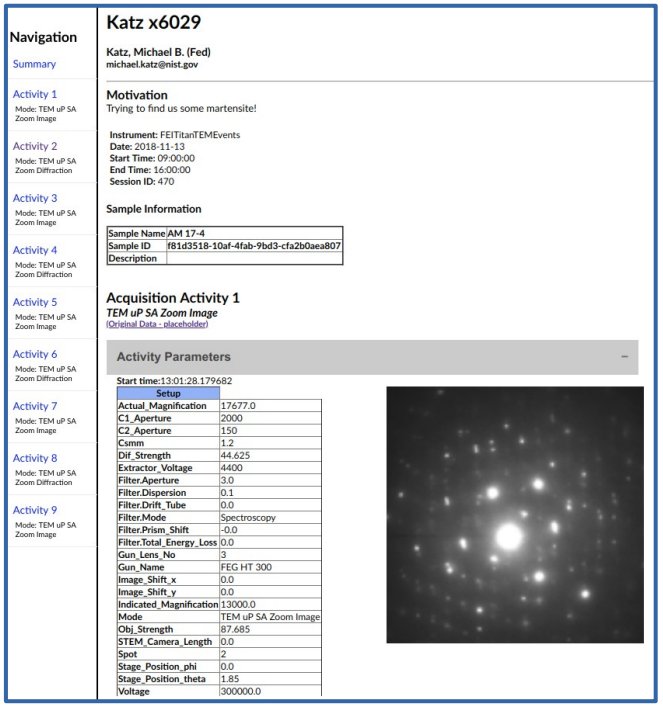
3. Providing an interface for users

CDCS

- CDCS used as the LIMS “backbone”







XML | XSL Transformation | HTML Output

Summary

- We have developed *Nexus***LIMS**
  - 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 *Nexus***LIMS**
  - Gathering feedback from users and tailoring to needs
  - Tweaking of schema and display; expanding to more tools

Where are we headed?

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

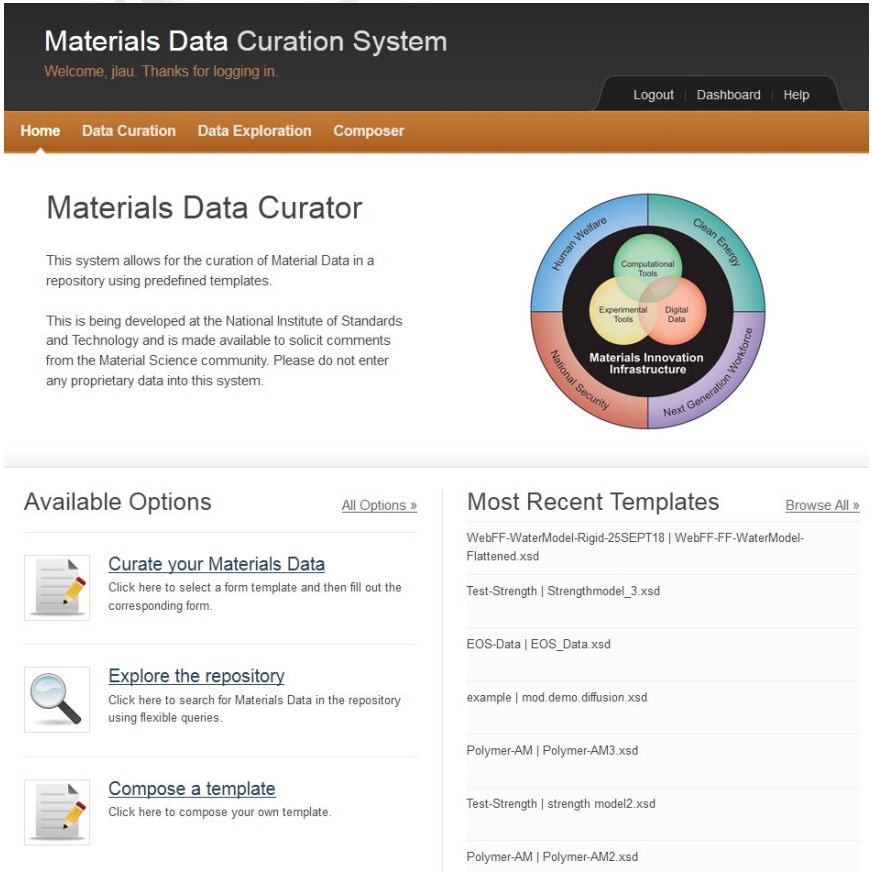
CHiMAD:

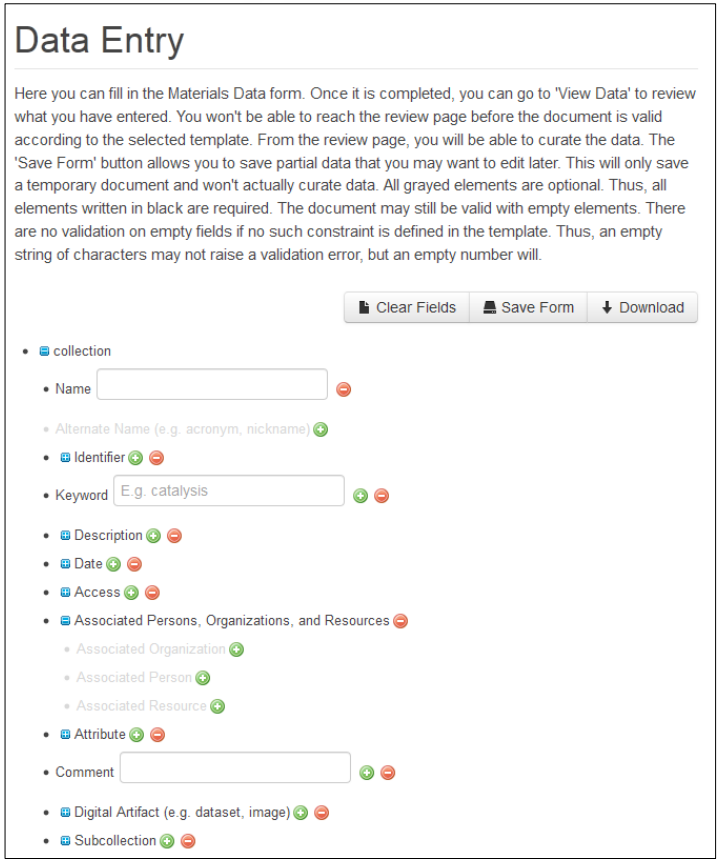
- Materials Microscopy Data conference/working groups:

Thank you!

joshua.taillon@nist.gov

CDCS as the LIMS backbone





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