NATIONAL HIGH MAGNETIC FIELD LABORATORY

A user-centric approach to building new CI to support AI-readiness at the MagLab

David S. Butcher

January 12th, 2026

NSF

FSU FLORIDA STATE UNIVERSITY

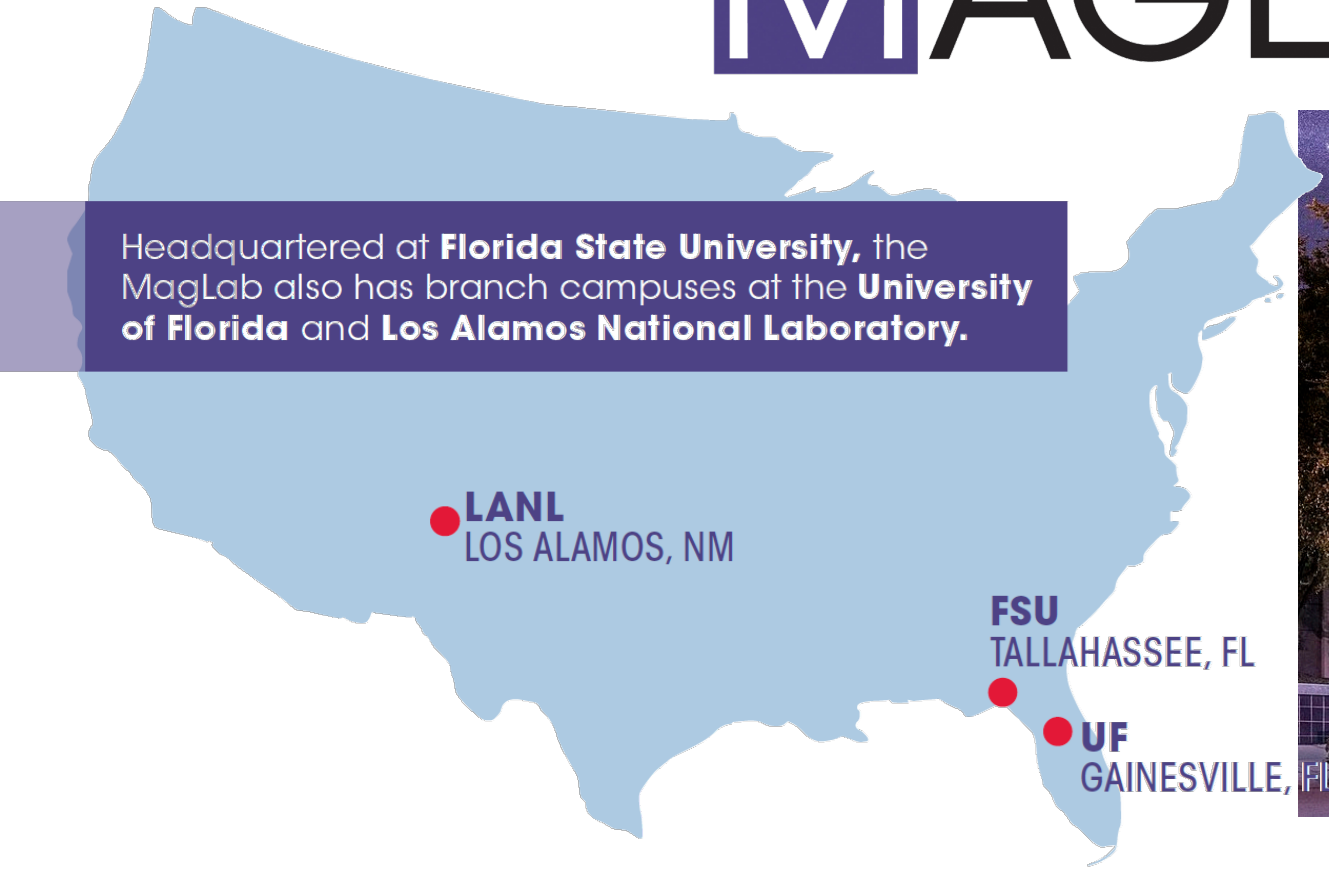
UF UNIVERSITY of FLORIDA

Los Alamos NATIONAL LABORATORY

MagLab Overview

NATIONAL MAGLAB

Headquartered at **Florida State University,** the MagLab also has branch campuses at the **University of Florida** and **Los Alamos National Laboratory.**



MagLab Overview

1 Lab. 3 Sites.
7 User Facilities.

- **DC Field User Facility (at Florida State University):** *continuous fields up to 45T.*
- **Electron Magnetic Resonance User Facility (at Florida State University):** *continuous and pulsed magnetic resonance techniques associated with unpaired electron spins.*
- **Nuclear Magnetic Resonance User Facility (at Florida State University):** *solution and solid state NMR and animal MRI.*
- **Ion Cyclotron Resonance User Facility (at Florida State University):** *ultra-high resolution and high mass accuracy Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry.*
- **Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) (at the University of Florida):** *high-resolution solution and solid state NMR,* in vivo *imaging and spectroscopy, MRI microscopy, and diffusion.*
- **High B/T (at the University of Florida):** *magnetic fields up to 15T combined with ultra-cold electron temperatures to 0.4mK.*
- **Pulsed Field User Facility (at Los Alamos National Laboratory):** *short, ultra-powerful magnetic fields up to 100T.*

FSU | FLORIDA STATE UNIVERSITY UF | UNIVERSITY of FLORIDA Los Alamos NATIONAL LABORATORY

Making MagLab data AI-ready

Center for FAIR & Open Science

"The MagLab's Center for FAIR and Open Science aims to ensure that all products derived from research done at the MagLab are shared according to the principles of FAIR data and open science."

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier
F2. Data are described with rich metadata (defined by R1 below)
F3. Metadata clearly and explicitly include the identifier of the data they describe
F4. (Meta)data are registered or indexed in a searchable resource

Accessible

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
A1.1 The protocol is open, free, and universally implementable
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
A2. Metadata are accessible, even when the data are no longer available

Interoperable

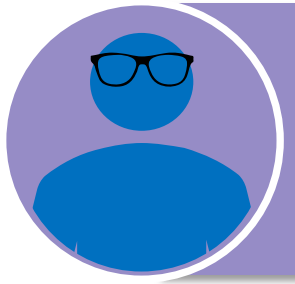
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (Meta)data use vocabularies that follow FAIR principles
I3. (Meta)data include qualified references to other (meta)data

Reusable

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (Meta)data are released with a clear and accessible data usage license
R1.2. (Meta)data are associated with detailed provenance
R1.3. (Meta)data meet domain-relevant community standards

Obstacles to AI-readiness

What are the primary obstacles to capturing metadata for AI/ML applications at the MagLab?



PI-centric approach that grants users substantial leeway in acquiring and transferring data

Lack of critical cyberinfrastructure to enable AI-readiness

Practical limitations to the types and comprehensiveness of metadata that can be collected

CI Solutions for AI-readiness

To address the need for FAIR/AI-ready data, the MagLab needs a new cyberinfrastructure solution that:

- ✅ Generates rich, structured metadata for diverse datasets and constituent files of many types
- ✅ Allows for additional metadata to be entered by users
- ✅ Is platform agnostic – can be run on data acquisition systems of various kinds
- ✅ Can work on data in real-time (as it is generated) or after dataset acquisition is complete

FOXDEN

CHESS
CORNELL HIGH ENERGY SYNCHROTRON SOURCE

FOXDEN[1]
FAIR Open-Science eXtensible Data Exchange Network

FOXDEN provides a variety of services for creation and management of data, metadata, and provenance records.

The MagLab is working to develop custom applications and considering integration of FOXDEN services into its existing data acquisition software.


Valentin Kuznetsov


Werner Sun

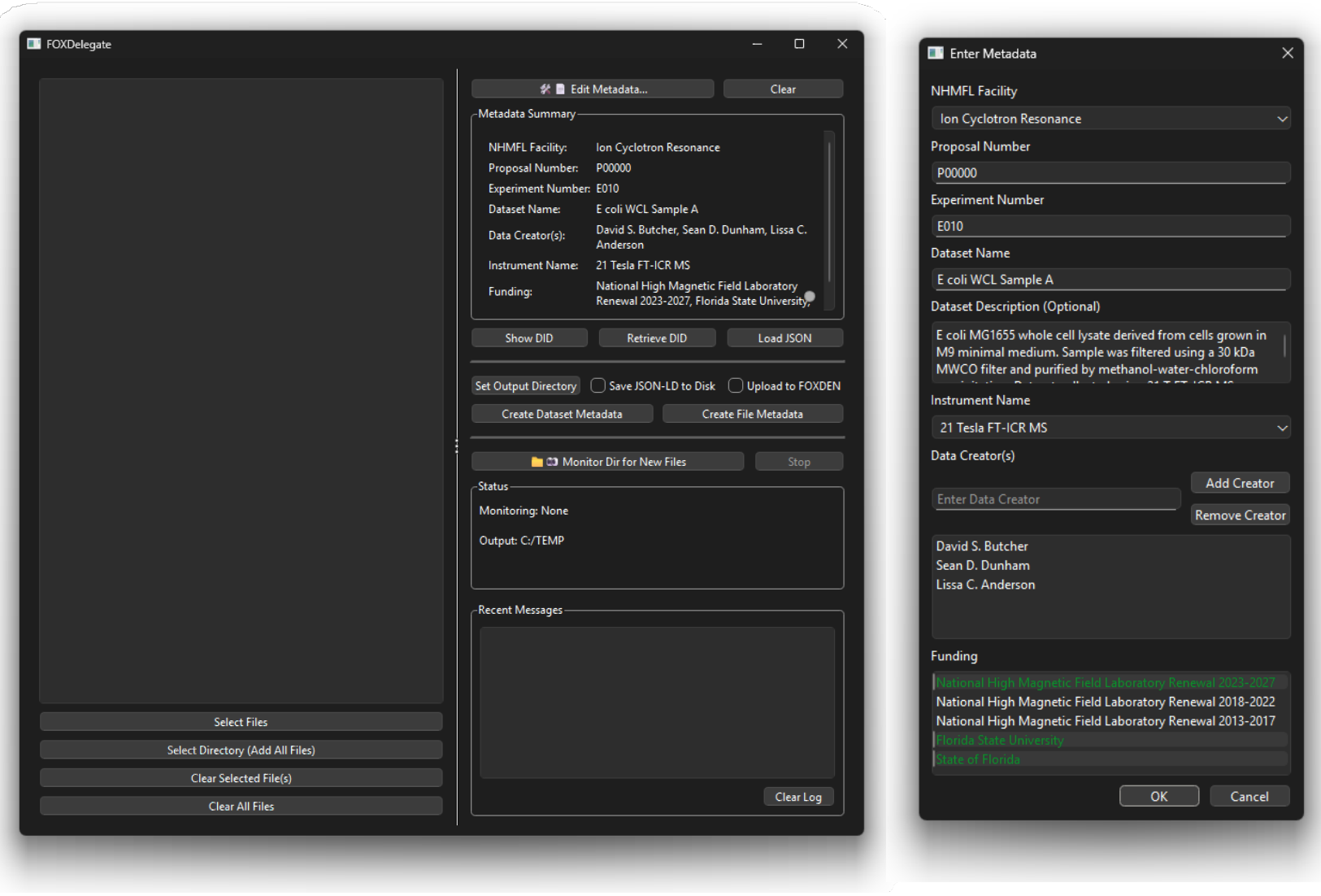

Keara Soloway


Rolf Verberg

1. https://github.com/CHESSComputing/FOXDEN

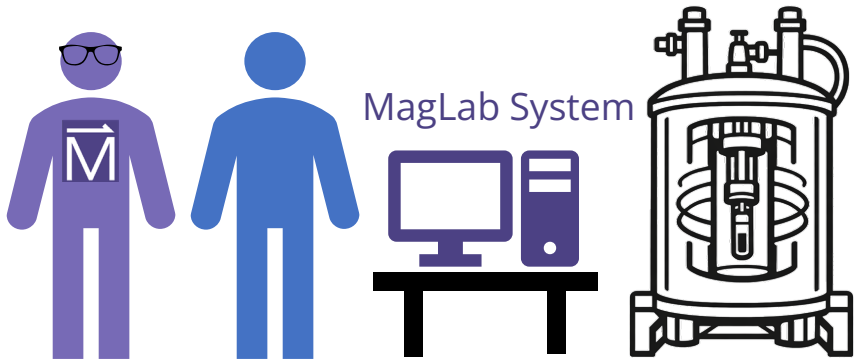
Custom Application

FOXDelegate is a custom PyQt6[1,2] application built around FOXDEN that allows MagLab researchers to automatically generate rich structured JSON-LD metadata for datasets and files and uploads it to a local database.



1. https://www.riverbankcomputing.com/software/pyqt/
2. https://www.qt.io/

AI-ready data/metadata collection



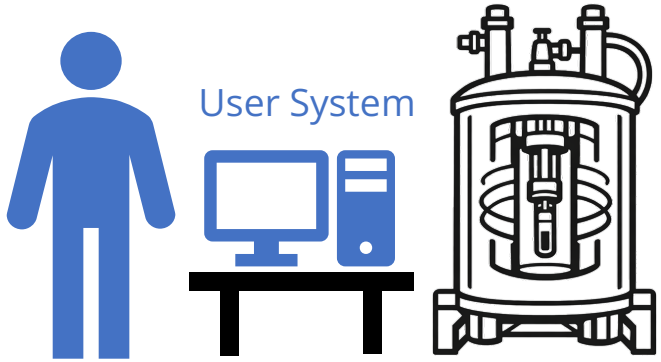
A. User works with MagLab systems with staff support to collect data and generate FAIR metadata

FOXDelegate

Custom Data Acq. Software

FOXDEN

MagLab User Portal



B. User provides data to MagLab staff who generates FAIR metadata.

Metadata example

Thermo .Raw file

RawFileReader Metadata Extraction by FOXDelegate[1]

- **Standards and semantic context:** JSON-LD context definitions; use of Schema.org/Science-on-Schema.org, PROV-O, MODSCI, and controlled vocabulary terms; ontology-based identifiers for techniques and properties
- **Data identification and integrity:** file name and original file path; file format and encoding (Thermo Scientific RAW); file creation and modification timestamps; cryptographic checksum (SHA-256) and associated checksum algorithm
- **Scientific dataset description:** dataset name and textual description; optional sample information; experimental context; relationship between the file and the parent dataset
- **Provenance and attribution:** data creators and contributors; explicit provenance relationships (what generated the data and what it is part of); associated facilities and organizations
- **Instrument and facility metadata:** mass spectrometer type, model, and magnetic field strength; instrument serial number; hosting facility; institutional affiliations; persistent identifiers for instruments and organizations (RRID, ROR)
- **Acquisition and method parameters:** mass spectrometer parameters that enable full experimental reproducibility
- **Software and configuration:** acquisition and control software name and version; instrument method and configuration file paths; firmware or method-specific settings
- **Funding and institutional support:** funding sources and grant identifiers; links between the dataset and supporting agencies or institutions

1. https://github.com/thermofisherlsms/RawFileReader

Fun with LLMs

Task	Grade
Generate a 100-word summary of the experiment that produced the file described by the metadata.	B
Describe the instrument method based on the metadata provided.	B+
Convert the metadata into a manuscript-ready methods paragraph.	C+
Compare it against standard top-down proteomics best practices.	C

Copilot

OVERALL GRADE:

C+

STUDY HARDER!

Through implementing policy changes and building new CI, the MagLab can support AI-readiness for users and facilitate future AI-driven operations.

NATIONAL HIGH MAGNETIC FIELD LABORATORY

David S. Butcher
(dbutcher@magnet.fsu.edu)