LAPS – A Digital Data Repository Workflow in Experimental Rock Deformation

Ulrich Mok[1] and Matej Pec[1]

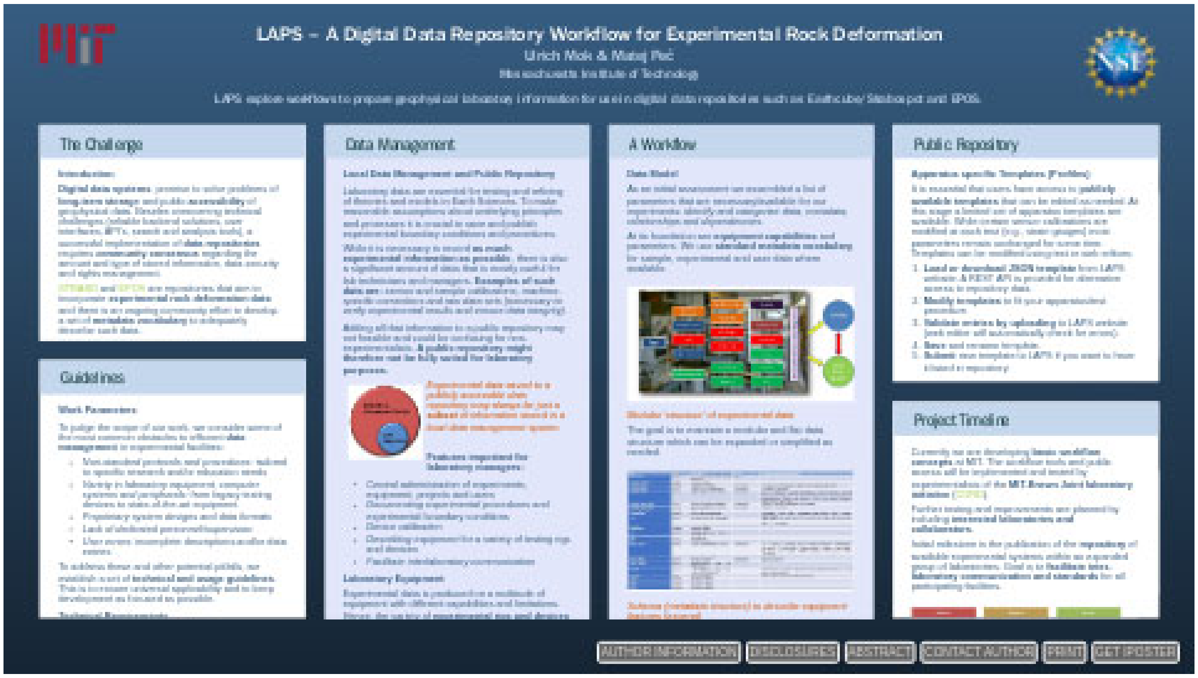
[1]Massachusetts Institute of Technology

November 21, 2022

Abstract

Laboratory data are essential for testing and refining of theories and models in Earth sciences. Recent developments in data mining techniques and machine learning have made it feasible to utilize and digest large amounts of information; yet such data must be initially prepared and structured in a meaningful way. Recognizing the potential and challenges of data access, multiple efforts are underway in the development of digital data repositories (Strabospot, Epos). Currently most information in experimental Geophysics is not accessible in digital, searchable form. Such information may include: equipment capabilities and configurations, original and edited experimental data, laboratory calibrations as well as information regarding testing protocols and procedures. The LAPS project aims to design and develop resources to facilitate data workflow and access. It specifically focuses on the needs of laboratory researchers, students and managers to prepare data for use in digital data repositories. As laboratories use a wide variety of hardware and software solutions to acquire and process data, we focus our efforts on the development of web based tools that do not require specific local infrastructure and software. One of the main objectives of LAPS is to establish a coherent and effective way to describe equipment and experiments across a variety of testing rigs and devices. To simplify the workflow we are proposing a combination of selecting pre-configured equipment and experimental profiles and manual data entry via web form. A completed test protocol containing all experimental metadata may then be saved locally (e.g., as JSON file) or (optionally) to a database. Such test protocols can be re-loaded and modified as needed. To complete the workflow of a successful test, a link to original or processed data files may be inserted. The data file layout can be defined in the experimental profile. Upload to a digital data repository is optional but by adhering to the proposed Strabo/Epos data model it will not require additional input. While the web based workflow will be accessible to all users, we also recognize the need to integrate the system into specific work flow solutions in rock deformation laboratories. We therefore provide the framework that simplifies local data management systems and analytical applications.

LAPS – A Digital Data Repository Workflow for Experimental Rock Deformation

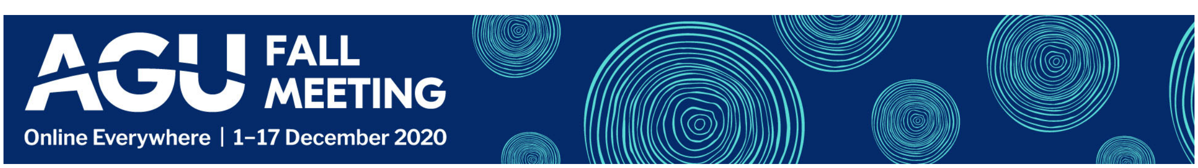


Ulrich Mok & Matej Peč

Massachusetts Institute of Technology

LAPS explores workflows to prepare geophysical laboratory information for use in digital data repositories such as Earthcube/Strabospot and EPOS.

PRESENTED AT:

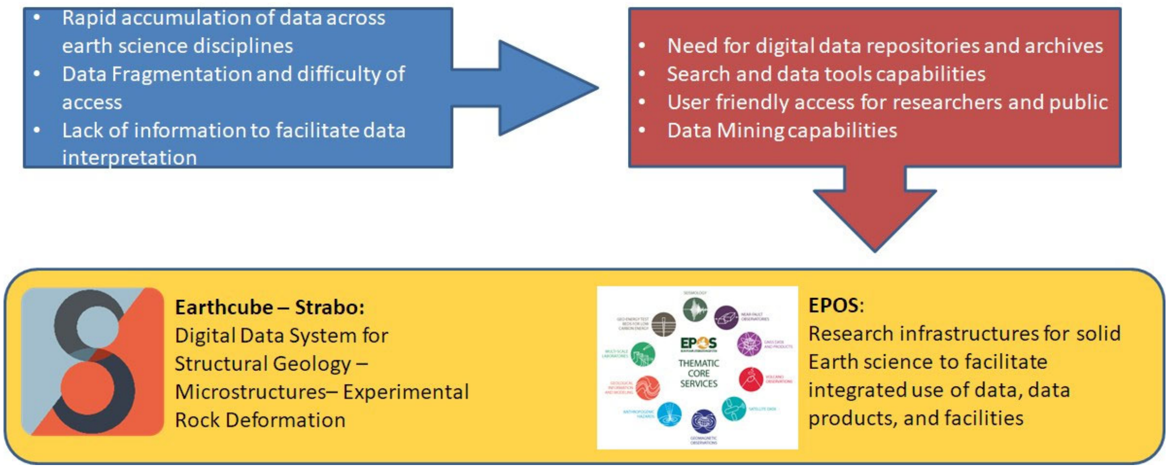


THE CHALLENGE

Introduction

Digital data systems promise to solve problems of **long-term storag**e and public **accessibility** of geophysical data. Besides overcoming technical challenges (reliable backend solutions, user interfaces, API's, search and analysis tools), a successful implementation of **data repositories** requires **community consensus** regarding the amount and type of stored information, data security and rights management.

STRABO (http://www.strabospot.org/)and EPOS (http://www.epos-ip.org/)are repositories that aim to incorporate **experimental rock deformation data** and there is an ongoing community effort to develop a set of **metadata vocabulary** to adequately describe such data.



*Community-driven Development of Digital Data Systems for Experimental Data (**Earthcube-Strabo/EPOS**)*

While most work is focused on publicly accessible data systems, standard vocabulary and user interfaces, little effort has been made so far to link local **laboratory workflows** to repositories.

In order to bridge the gap between data gathering and publication, we propose to examine existing workflows in rock deformation laboratories. Using **STRABO/EPOS vocabulary** we plan to develop tools and procedures to best describe and prepare experimental data for storage and help facilities with **local data management**.

"LAPS (http://verve.mit.edu/test/index.html)" (Laboratory Acquisition Protocols and Standards)

GUIDELINES

Work Parameters

To judge the scope of our work, we consider some of the most common obstacles to efficient **data management** in experimental facilities:

- *Non-standard protocols and procedures:* tailored to specific research and/or education needs
- *Variety in laboratory equipment, computer systems and peripherals:* from legacy testing devices to state-of-the-art equipment
- *Proprietary system designs and data formats*
- *Lack of dedicated personnel/supervision*
- *User errors:* incomplete descriptions and/or data entries

To address these and other potential pitfalls, we establish a set of **technical and usage guidelines**. This is to ensure universal applicability and to keep development as focused as possible.

Technical Requirements

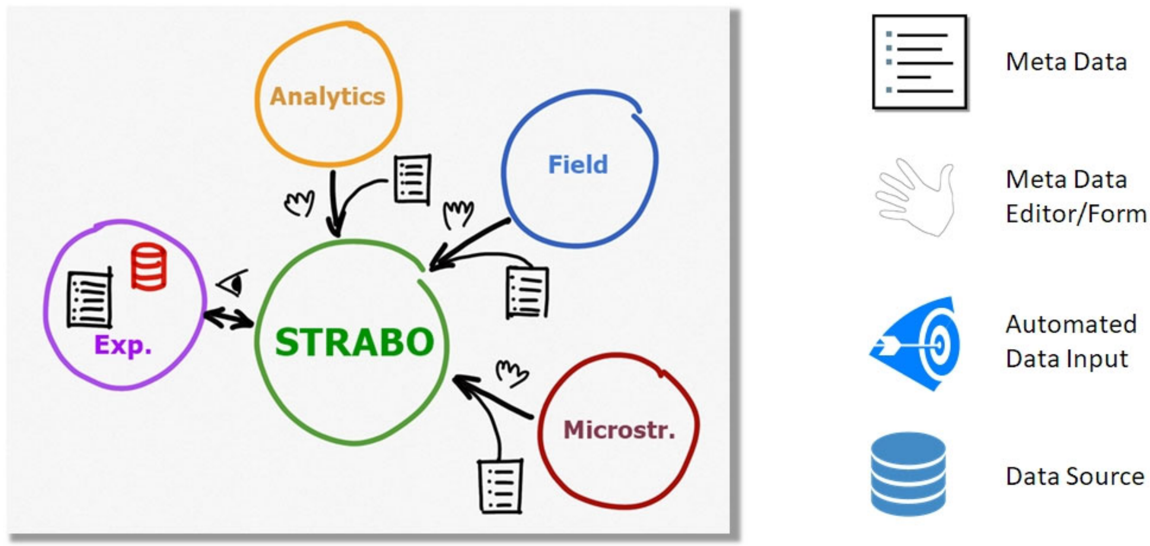
- Open-Source
- Multi-Platform (OS Agnostic)
- Modular components
- Metadata validation

Project Targets

- **Flexible** – simple or comprehensive if needed
- **Ease of Use** - intuitive interface
- To be used **locally and/or remotely**
- **Compliance** with existing protocols and funding agencies requirements
- Use **standard vocabulary**: agreed upon nomenclature (**STRABO/EPOS**)
- **Improvement** to existing insular solutions
- **Integration** into existing laboratory workflows (automated or otherwise)

Compatibility to Strabospot

Strabospot (https://www.strabospot.org/)(Strabo Data System) proposes to extend its capability from tectonic data to allow inclusion of experimental data.



*Data Input for **Strabo** digital data system.*

Metadata

- **Experiment:** Experimental results with comprehensive test information and boundary conditions
- **Field:** Observational data/field tests
- **Analytics:** Analytical data pertaining to field and/or laboratory tests (e.g., Chemistry, MicroCT)
- **Microstructures**

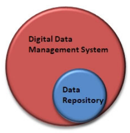
DATA MANAGEMENT

Local Data Management and Public Repository

Laboratory data are essential for testing and refining of theories and models in Earth Sciences. To make reasonable assumptions about underlying principles and processes it is crucial to save and publish experimental boundary conditions and procedures.

While it is necessary to record **as much experimental information as possible**, there is also a significant amount of data that is mostly useful for lab technicians and managers. **Examples of such data are:** *sensor and sample calibrations, machine-specific corrections and raw data sets (necessary to verify experimental results and ensure data integrity).*

Adding all that information to a public repository may not feasible and could be confusing for non-experimentalists. **A public repository might therefore not be fully suited for laboratory purposes.**



*Experimental data saved to a publicly accessible data repository may always be just a **subset** of information stored in a local data management system.*

Features important for laboratory managers:

- *Central administration* of experiments, equipment, projects and users
- *Documenting experimental procedures* and experimental boundary conditions
- *Device calibration*
- *Describing equipment* for a variety of testing rigs and devices
- *Facilitate interlaboratory communication*

Laboratory Equipment

Experimental data is produced on a multitude of equipment with different capabilities and limitations. Hence, the variety of **experimental rigs and devices** needs to be considered when examining workflow solutions.

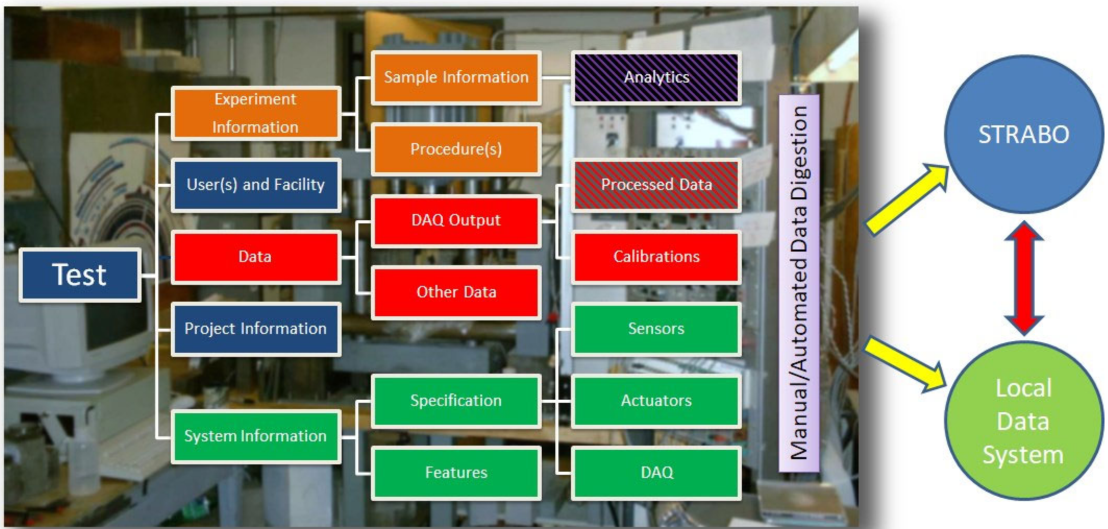
Experimental equipment at the MIT Rock Deformation Laboratory

A WORKFLOW

Data Model

As an initial assessment we assembled a list of parameters that are necessary/available for our experiments: *identify and categorize data, metadata, relationships and dependencies*.

At its foundation are **equipment capabilities** and parameters. We use **standard metadata vocabulary** for sample, experimental and user data where available.



Modular 'structure' of experimental data

The goal is to maintain a modular and flat data structure which can be expanded or simplified as needed.

name	type	description	Items	enumerations/restrictions/default
data_channel	object	List of DAQ and Data Channel Attributes	object	
_channel_number	integer	Channel Number		
_channel_type	string	Type of Channel Data	enumeration	"Analog Input","Analog Output","Digital Input","Digital Output","Data
_channel_format	string	Measured or Processed Data?	enumeration	"Processed","Measured"
_channel_name	string	Channel Description -> STRABO	enumeration	"Time", "Temperature", "Sample Temperature", "Fluid Temperature", "Confining Pressure", "Confining Pressure Intensifier-Displacement", "Confining Pressure Intensifier-Stroke", "Pore Fluid Pressure", "Pore Fluid Pressure Intensifier-Displacement", ...continued...
_channel_name_add1	string	Additional Channel Information	enumeration	"Internal", "External", "Upstream", "Downstream"
_channel_name_add2	string	Additional Channel Information	enumeration	"Low Gain", "High Gain"
_channel_name_note	string	Channel Notes		
_configuration	string	Input/Output Configuration		"Differential", "Single Ended", "Referenced Single Ended", "Serial", "Parallel", "Single", "Line"
_unit	string	Physical Units or Raw Data (Volt, A)		"degC", "degK", "sec", "min", "hour", "volt", "mV", "mA", "Ohm", "Pa", "MPa", "bar", "kbar", "N", ..continued...
_range_low	number	Low Input Range		
_range_high	number	High Input Range		
_rate	string	Acquisition/Sampling Rate in Hz/kHz/MHz		
_gain	string	Amplifier Gain		
_filter	string	HW Filter Settings		
_sensor_name	string	Sensor/Controller Name -> STRABO	enumeration	"LVDT", "Load Cell", "Capacitive Load Cell", "Pressure Transducer", "Thermocouple", "Hall Sensor", "P-Wave Sensor", "S-Wave Sensor", "Encoder", "Strain Gauge", "Thermistor", "Force Gauge", "DCDT", "pH-Meter", "Flow Sensor", "None", ...continued...
_sensor_type	string	Sensor or Controller?	enumeration	"Sensor", "Controller"
_sensor_detail	string	Passive or Active?	enumeration	"Passive", "Active"
_sensor_model	string	Sensor/Controller Model and Manufacturer		
_cal_info	string	Linear Calibration; u[unit], x[volt]: e.g., u=(x*a0+a1)*a2+a3		"if the formula is entered in the proper syntax (LabView Mathscript) it can be programmatically used"
_cal_date	date	Date of Calibration		default: 2019-07-26 (2019-07-26T18:25:43.511Z)
_cal_a0	number	Calibration value a0		
_cal_a1	number	Calibration value a1		
_cal_a2	number	Calibration value a2		
_cal_a3	number	Calibration value a3		

Schema (metadata structure) to describe equipment features (excerpt)

Possible Solution

Entering large amounts of data during a test is time consuming, prone to user errors, disruptive, and leads to acceptance issues. Any workflow that increases user workload might result in ***missing information and ultimately fewer datasets in a repository!***

Fortunately, much information is already defined by **equipment choice** and **experimental setup**. This potentially allows **semi-automated data input/access**. Related information is grouped into templates that can be **adapted and re-used**.

After analyzing the data model, we decided on three distinct **user templates**:

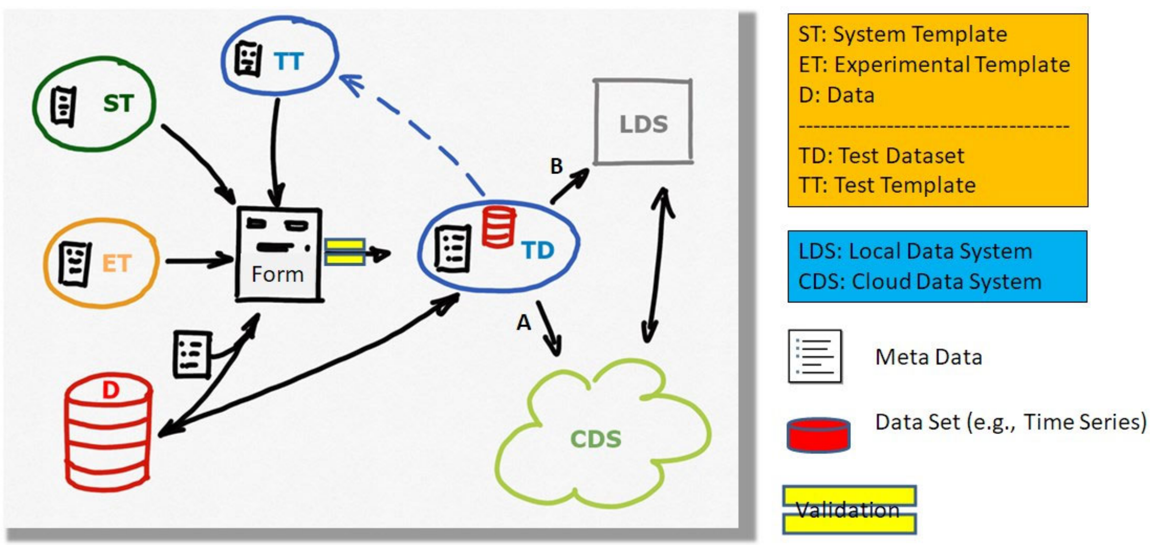
- *Apparatus Specific System Templates*
- *Lab Specific User Templates*
- *Experimental Templates* (Sample and Procedures)
- *Test Template* (a combination of the above)

These templates are **pre-populated** with as much information as possible to reduce test specific data entries. **Schema definitions** for JSON allow the flexibility in reducing or expanding datasets as needed without violating the underlying data structure. **Model templates** can be downloaded from a **public repository** and adapted.

Workflow (Web Form)

In one scenario, template information is accessible to the user via a local form (JavaScript) in a browser window. To reduce the complexity of the web form, the schema includes **flags** for editable or hidden data. Templates and resulting data sets are stored as readable JSON on the local system or network.

After adding information via editor and templates, the results are saved as text file (test instance) or as a new **test template** for future use. Besides being a **data collection tool**, the web form also provides **schema validation**.



Web based form central to lab workflow.

Steps

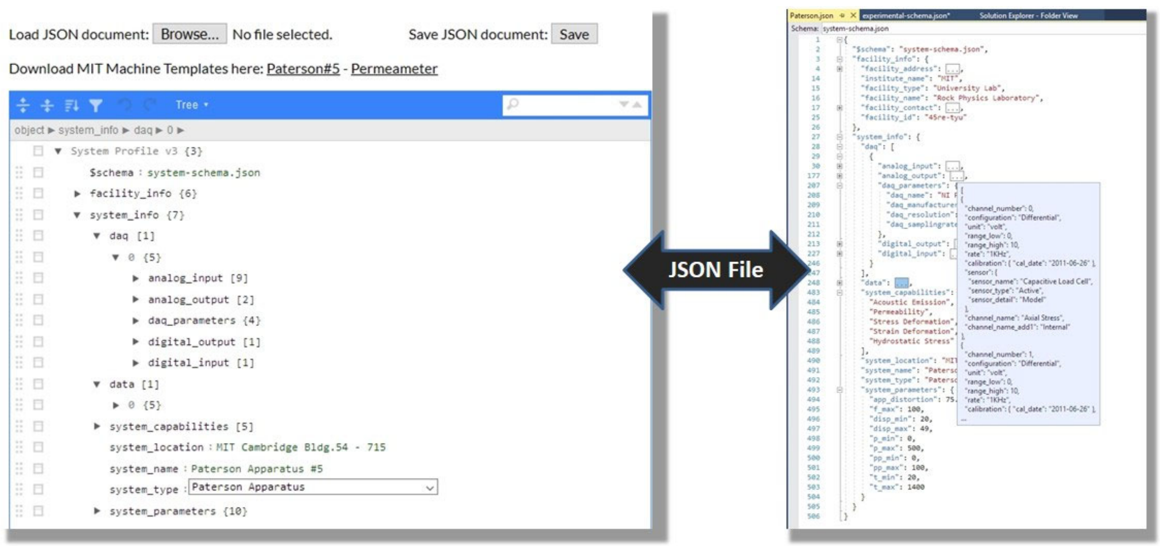
- **Select templates:** *Systems – Experiment – Facilities* from local storage or public repository
- **Enter and validate test information** - the appearance can be tailored to specific experiments via entry mask
- **Save dataset** to local and/or remote system (or reuse as template for other tests)
- Additional information (data, measurements) may be added to the dataset when available

PUBLIC REPOSITORY

Apparatus specific Templates (Profiles)

It is essential that users have access to **publicly available templates** that can be edited as needed. At this stage a limited set of apparatus templates are available. While certain sensor calibrations are modified at each test (e.g., strain gauges) most parameters remain unchanged for some time. Templates can be modified using text or web editors:

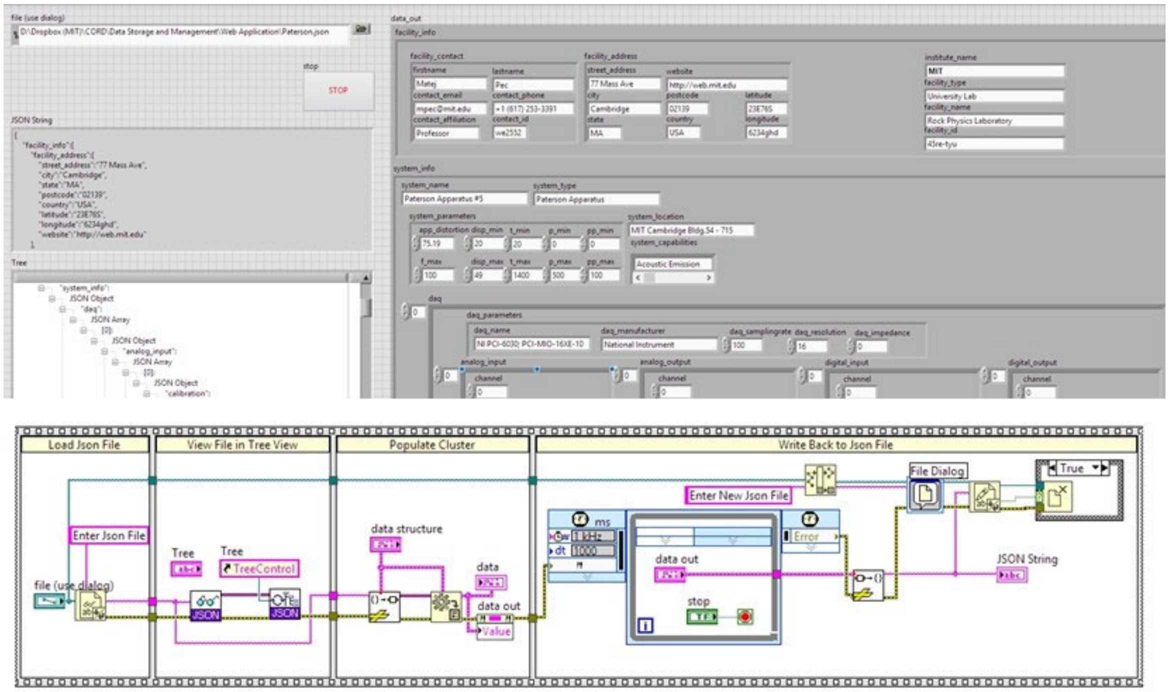
1. **Load or download JSON template** from LAPS website. A REST API is provided for alternative access to repository data.
2. **Modify templates** to fit your apparatus/test procedure.
3. **Validate entries by uploading** to LAPS website (web editor will automatically check for errors).
4. **Save** and rename template.
5. **Submit** new template to LAPS if you want to have it listed in repository.



Editing apparatus specific templates

Additional Tools

Emphasis is on **open source solutions** with web tools developed in **JavaScript**. For **parsing** existing datasets, MATLAB and Python scripts as well as LabVIEW routines are being provided.



LabVIEW program for data import/export from JSON

Furthermore:

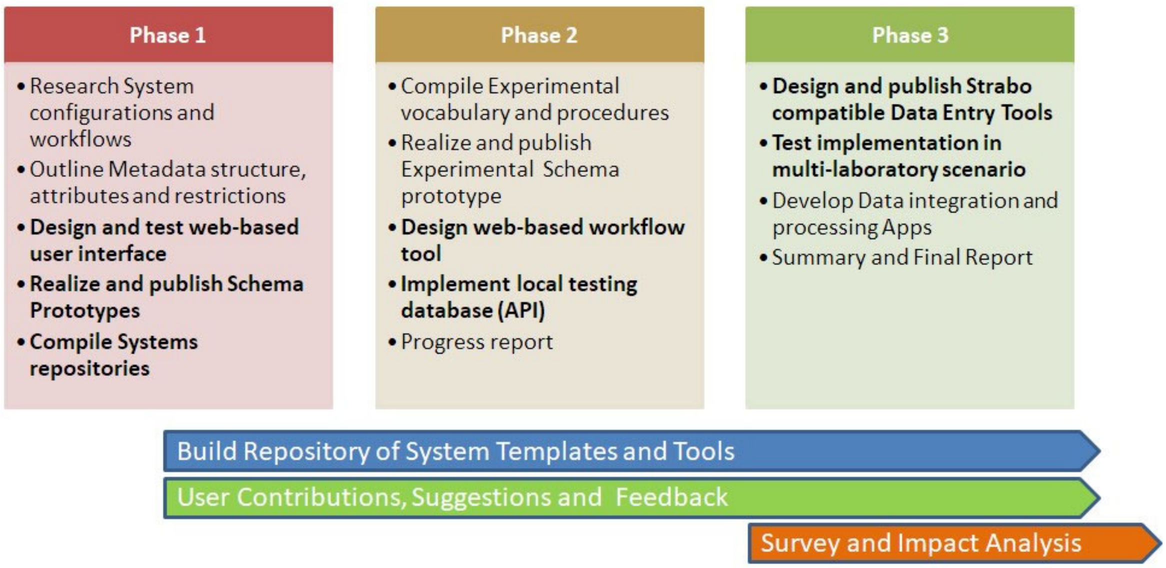
- **Integration** into existing user interfaces/hardware/data workflow
- **Data collection tools** for previously stored data files.
- **Database** option (for metadata only)

PROJECT TIMELINE

Currently we are developing **basic workflow concepts** at MIT. The workflow tools and public access will be implemented and tested by experimentalists of the **MIT-Brown Joint laboratory initiative** (CORD (http://cord.mit.edu)).

Further testing and improvements are planned by including **interested laboratories and collaborators**.

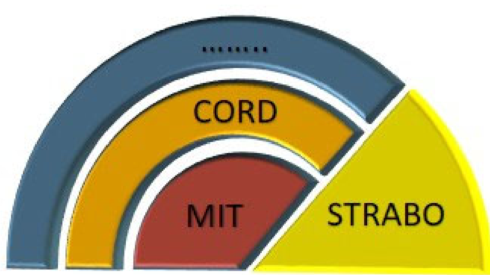
Initial milestone is the publication of the **repository** of available experimental systems within an expanded group of laboratories. Goal is to **facilitate inter-laboratory communication and standards** for all participating facilities.



Project Timeline (2020-2023)

Priorities

1. Collect equipment information and expand repository.
2. Add and test web-based workflow tool
3. Invite public comments
4. Strabo Interface (API)



Associated Project Partners

Final Takeaway

Improving local workflows can help with data transitioning until suitable public depositories are readily available.

DISCLOSURES

This Project is funded by the National Science Foundation (NSF) - Reward #1948453

AUTHOR INFORMATION

Ulrich Mok (u_mok@mit.edu)

Matej Pec (mpec@mit.edu)

ABSTRACT

Laboratory data are essential for testing and refining of theories and models in Earth sciences. Recent developments in data mining techniques and machine learning have made it feasible to utilize and digest large amounts of information; yet such data must be initially prepared and structured in a meaningful way. Recognizing the potential and challenges of data access, multiple efforts are underway in the development of digital data repositories (STRABOSPOT, EPOS).

Currently most information in experimental Geophysics is not accessible in digital, searchable form. Such information may include: equipment capabilities and configurations, original and edited experimental data, laboratory calibrations as well as information regarding testing protocols and procedures. The LAPS project aims to design and develop resources to facilitate data workflow and access. It specifically focuses on the needs of laboratory researchers, students and managers to prepare data for use in digital data repositories.

As laboratories use a wide variety of hardware and software solutions to acquire and process data, we focus our efforts on the development of web based tools that do not require specific local infrastructure and software. One of the main objectives of LAPS is to establish a coherent and effective way to describe equipment and experiments across a variety of testing rigs and devices. To simplify the workflow we are proposing a combination of selecting pre-configured equipment and experimental profiles and manual data entry via web form. A completed test protocol containing all experimental metadata may then be saved locally (e.g., as JSON file) or (optionally) to a database. Such test protocols can be re-loaded and modified as needed.

To complete the workflow of a successful test, a link to original or processed data files may be inserted. The data file layout can be defined in the experimental profile. Upload to a digital data repository is optional but by adhering to the proposed STRABO/EPOS data model it will not require additional input. While the web based workflow will be accessible to all users, we also recognize the need to integrate the system into specific work flow solutions in rock deformation laboratories. We therefore provide the framework that simplifies local data management systems and analytical applications.