HydroGEN: a CKAN-based Data Hub for Advanced Water Splitting Devices

Violet Yao[1], Nemanja Danilovic[2], Dan Gunter[2]

1. University of California, Berkeley 2. Lawrence Berkeley National Laboratory

Problem Statement

- The development of efficient water-splitting devices is valuable because hydrogen is a versatile energy carrier that enables domestic energy and environmental security with large-scale market potential, job creation, and economic growth opportunities.
- **It generally takes fifteen to twenty years for a new energy material to go from discovery to production**.

HydroGEN
Advanced Water Splitting Materials

Background

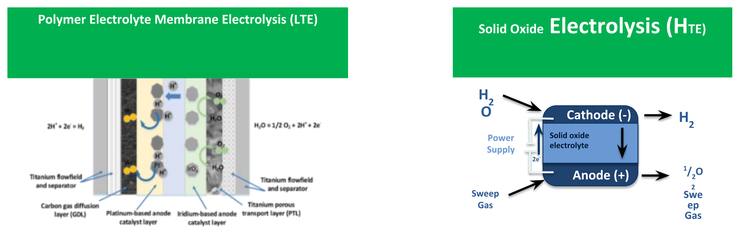
The HydroGEN Energy Materials Network (EMN) is a Department of Energy Consortium created to discover new materials for water-splitting devices. 19 external projects are funded to collaborate with 91 National Lab capabilities on projects using experimental, simulation synthesis and design tasks. A Data Hub was created with the intention of gathering and sharing important experimental data in order to accelerate innovation.

Project Goals

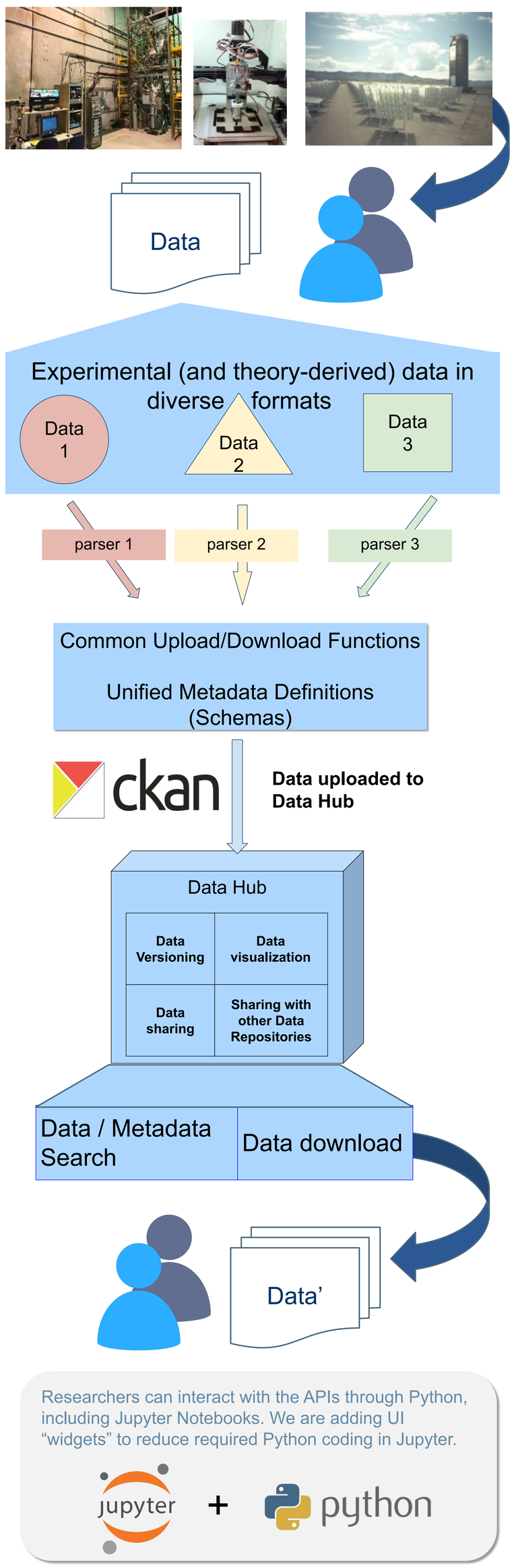
The HydroGEN Data Hub supports collaborative science through the establishment of an accessible, searchable data resource.

- Secure sharing of data among project team members.
- Advanced search across all data using common, defined metadata.
- Facilitate access to advanced data tools for analysis.
- Make selected datasets publically available.

This enables cross-cutting analysis of experimental and simulated data from multiple Advanced Water Splitting Material (AWSM) technologies



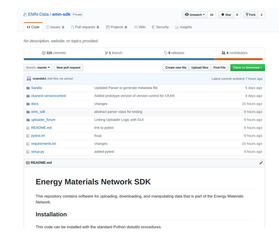
System Design



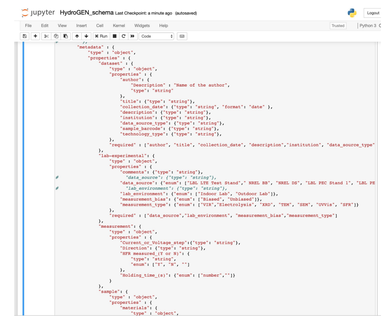
Highlights

Progress has been made in three essential areas:

1. Collaborative development with GitHub
   - Single repository across Labs
   - Code packaged with distutils
   - Using Github Issues and Pull Requests



2. Unified data packaging and description
   - Codified in JSON-Schema
   - Validate Metadata formats with an API in Jupyter Notebooks or Python Scripts



3. Extensible parsing framework
   - Implemented parsers for Excel and CSV
   - Excel templates for easy data entry from laboratory settings

Put screenshot of Nem's Excel Spreadsheet here

parse the spreadsheet

can either upload an xlsx or a csv

```
[3]: # parse excel
     input_file = 'LTE_data_metadata_example_11132018.xlsx'
     sheet = parser.Spreadsheet(input_file)

     input_format is :  xlsx
     Name of Spreadsheet is : LTE_data_metadata_example_11132018

[4]: # parse csv
     input_file = 'LTE_data_metadata_example_11132018.csv'
     sheet = parser.Spreadsheet(input_file)

     input_format is :  csv
```

Future Work

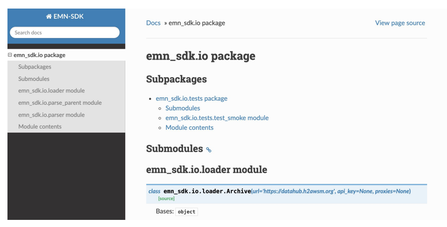
- Add more data-specific parsers (PDF, Xml, TEXT...)
- Enhance the searchability of Metadata key fields and datasets through solr, CKAN's search platform

Solr

lab-experimental-metadata		measurement-metadata		sample-metadata-material	
Key	Value	Key	Value	Key	Value
comments		direction		anode_catalyst_material	Pt/C
data_source	LBL LTE Test Stand	current_or_voltage_step		anode_catalyst_particle_size	5 nm
lab_environment	Indoor Lab	holding_time_(s)		anode_catalyst_support	Carbon
measurement_bias	Biased	hfr measured_(y or n)		anode_catalyst_synthesis	n/a
measurement_type	VIR			cathode_catalyst_material	Pt/C
				cathode_catalyst_particle_size	5 nm
				cathode_catalyst_support	Carbon
				cathode_catalyst_synthesis	n/a
				membrane_ionomer	Nafion
				membrane_thickness	2 mil
				membrane_iec	1100 EW
				membrane_preparation_conditions	as received

- Develop straightforward document based on Sphinx

SPHINX
PYTHON DOCUMENTATION GENERATOR



Acknowledgement

Special thanks to my mentors Dan Gunter and Nemanja Danilovic, my collaborators Keith Kozlosky and Srinath Nandakumar from Sandia National Laboratories.

The author gratefully acknowledges research support from the HydroGEN Advanced Water Splitting Materials Consortium, established as part of the Energy Materials Network under the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Fuel Cell Technologies Office, under Contract Number DE-AC02-05CH11231.

Energy Materials Network
U.S. Department of Energy

Sandia National Laboratories

NREL
National Renewable Energy Laboratory
Innovation for Our Energy Future