A Schema for Storing Synthesis Recipes

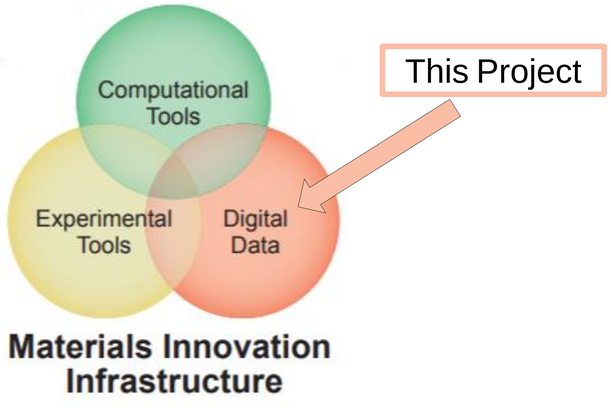
Ruby Aidun[1], Yevgeny Rakita[1], and Simon J. L. Billinge [1,2]

[1] Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027

[2] Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973

Background: Materials Genome Initiative

It currently takes 10-20 years to bring a new material from initial research to market. Incorporating advanced computing and data sharing into material research has the prospect of accelerating novel material discovery and deployment. This is the goal put forth by the Materials Genome Initiative (MGI).



National Science and Technology Council. (2011). *Material Genome Initiative for Global Competitiveness* [White Paper]. https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf

Problem Statement

Can predict a new material from:
Databases of Material Properties
&
Computational Prediction of Novel Materials

MATERIALS DATA FACILITY

The Materials Project
materialsproject.org

AFLOW
Automatic - FLOW for Materials Discovery

How can this novel material be synthesized?

Can we teach scientific intuition to a computer so it can suggest a recipe?

Need:

A large dataset of synthesis protocols for Machine Learning

Therefore Need:

A well structured database to store these protocols

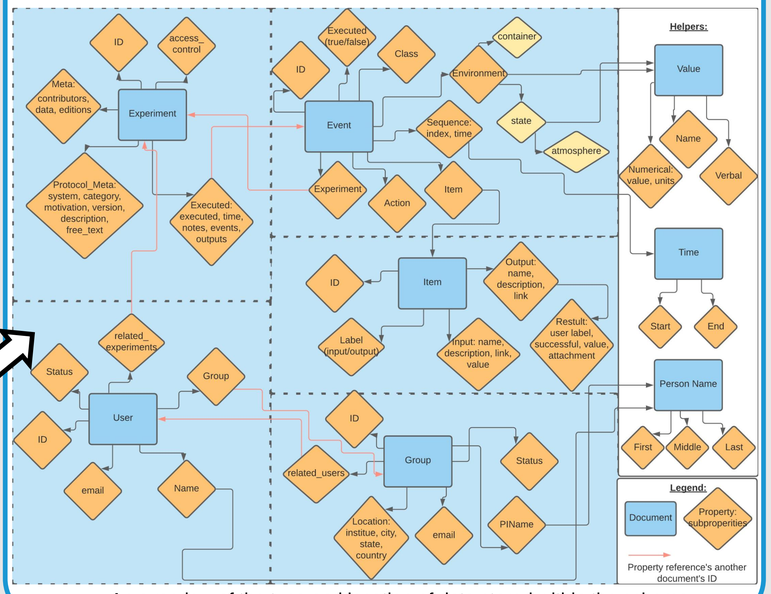
Blaiszik, Ben, Logan Ward, Marcus Schwarting, Jonathon Gaff, Ryan Chard, Daniel Pike, Kyle Chard, and Ian Foster. "A Data Ecosystem to Support Machine Learning in Materials Science." MRS Communications (October 10, 2019): 1–9. doi:10.1557/mrc.2019.118.

Jain*, S.P. Ong*, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson *The Materials Project: A materials genome approach to accelerating materials innovation* APL Materials, 2013, 1(1), 011002. doi:10.1063/1.4812323

AFLOW: http://www.aflowlib.org/

Schema Development

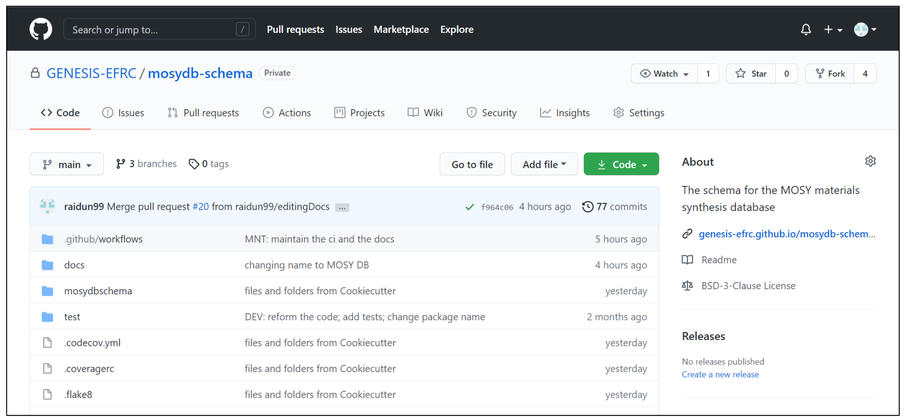
The idea began as a set of spread sheets, a relational database. However, solid state syntheses are diverse, requiring a variety of data to be stored within the database. This requires a non-relational (NoSQL) database, so we developed a JSON based schema from use-cases in the relational database.



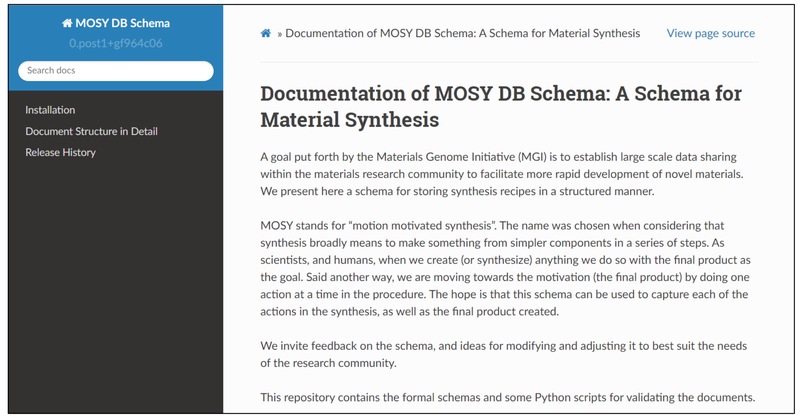
An overview of the type and location of data stored within the schema

Publishing and Maintaining the Schema

The development of a database and schema for storing synthesis recipes will, ultimately, be completed by the materials community collectively. What is published is only a work in progress. Over time, with feedback from researchers and ideas from the community, the schema will be modified and updated with new releases.



Repository containing the schema and schema validation scripts: https://github.com/GENESIS-EFRC/mosydb-schema



The documentation of the schema and how to install it: https://genesis-efrc.github.io/mosydb-schema/

Acknowledgements

The Billinge Group
https://billingegroup.github.io/

Top: Connor Bracy, Simon Billinge, Yevgeny Rakita, Daniela Yano
2nd row: Zachary Thatcher, Lauren Kranis, Sandra Skjaervoe, Berrak Ozer
3rd Row: Jaylyn Umana, Ruby Aidun, Ahmed Shaaban, Songsheng Tao
Bottom: Eric Shen
Not Pictured: Ling Lan, Priyanka Nehra, Martin Karlsen, Louis Cheng

This work in the Billinge group was supported by the U.S. National Science Foundation through grant DMREF-1922234

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science