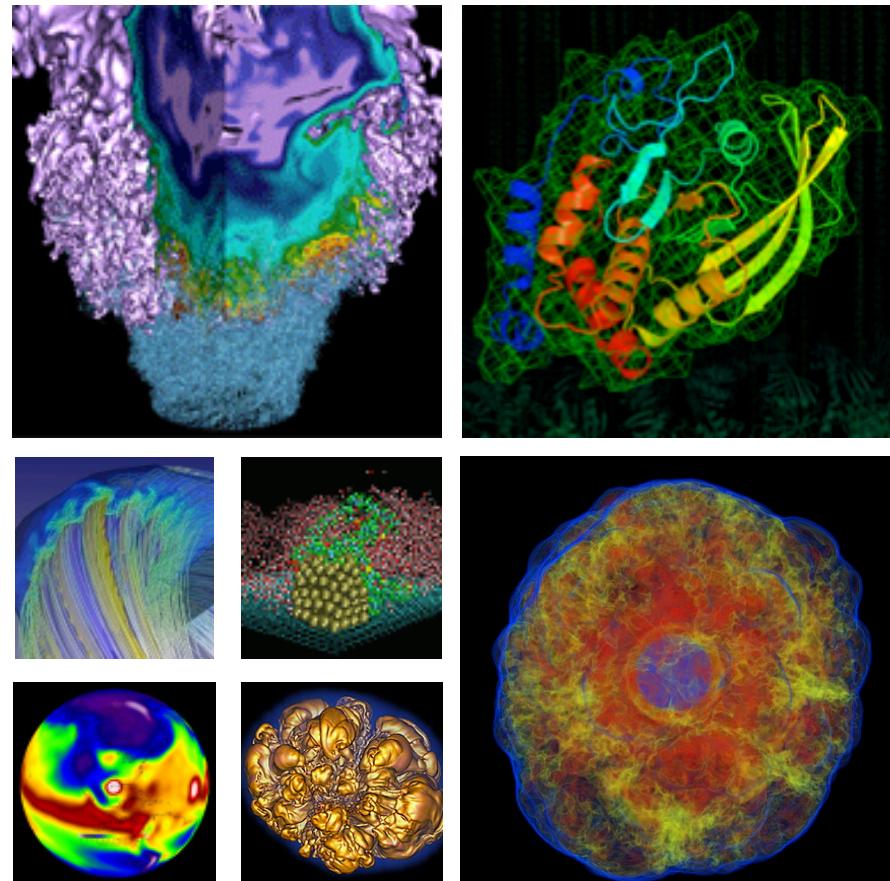


# The Materials Project



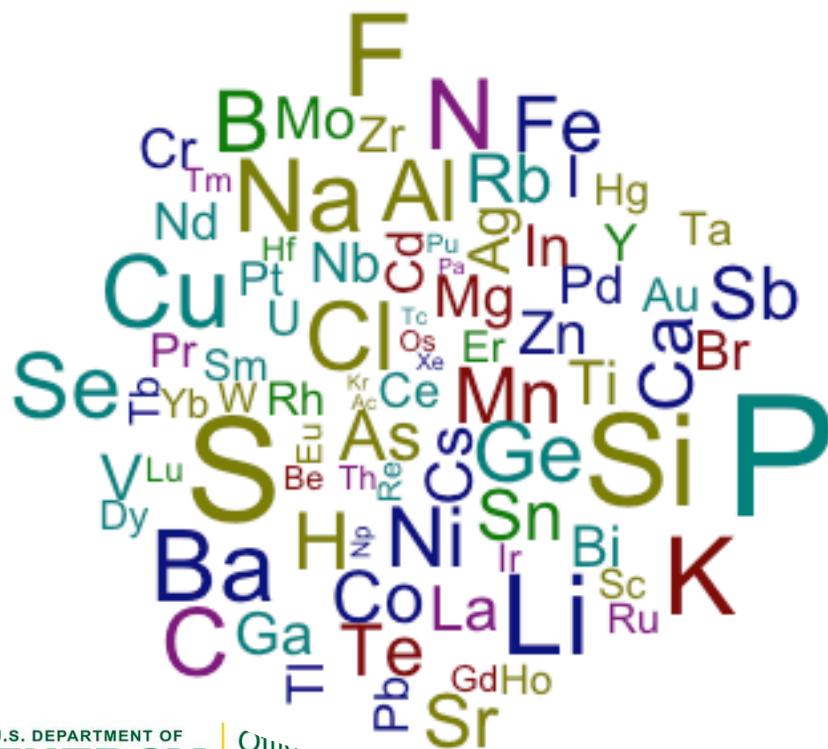
**Shreyas Cholia**  
Lawrence Berkeley National Laboratory

July 3, 2013



# The Materials Project

An open science initiative that makes available a huge database of computed materials properties for all materials researchers.



Wordcloud showing frequencies of elements in Materials Project's database

..except Oxygen, which appears 12,751 times (3.5x as much as the next most frequent, Phosphorus)



# Big Data

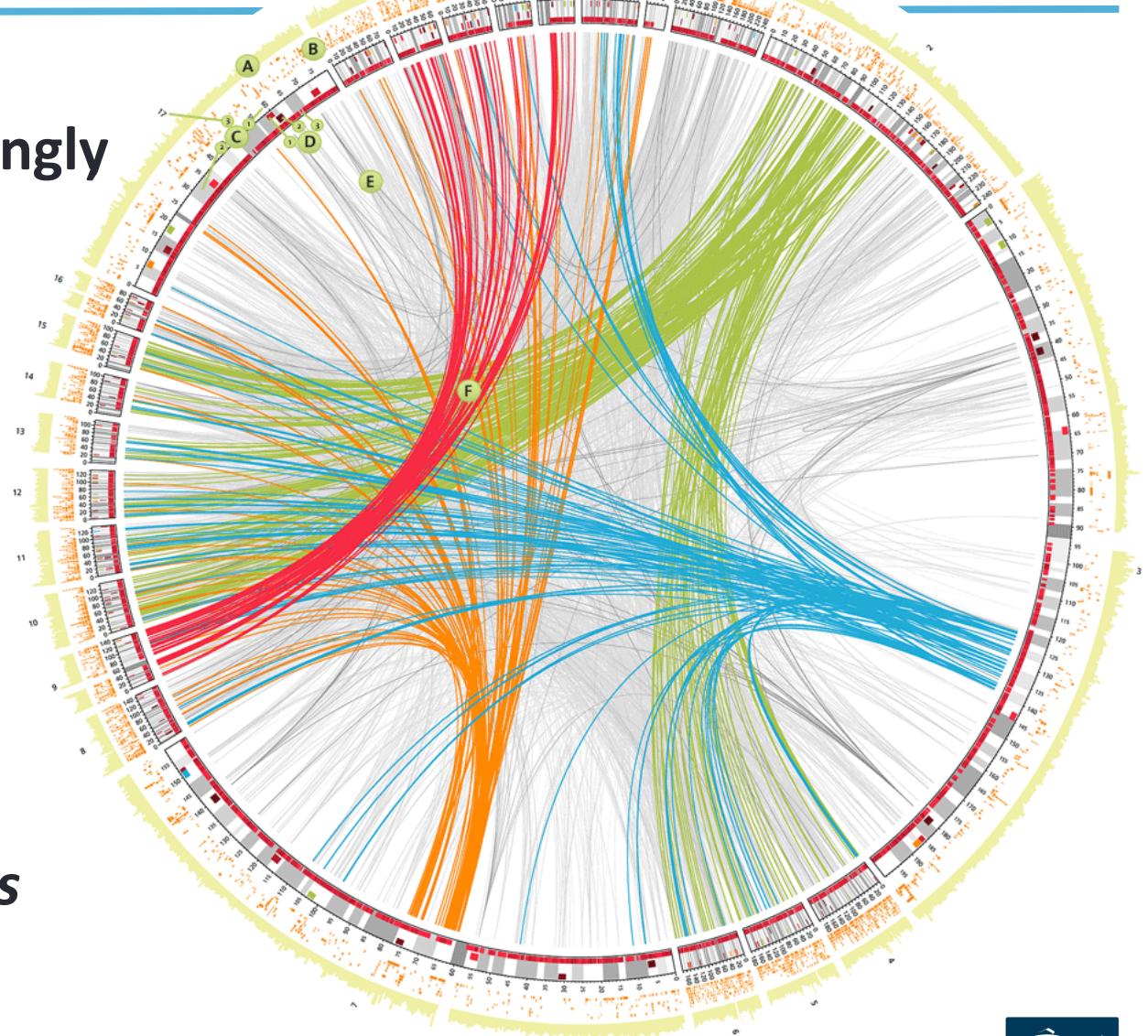


Science is increasingly  
data-driven

Computational  
cycles are cheap

Take an –omics  
approach to  
science

*Compute all  
interesting things  
first, ask questions  
later*

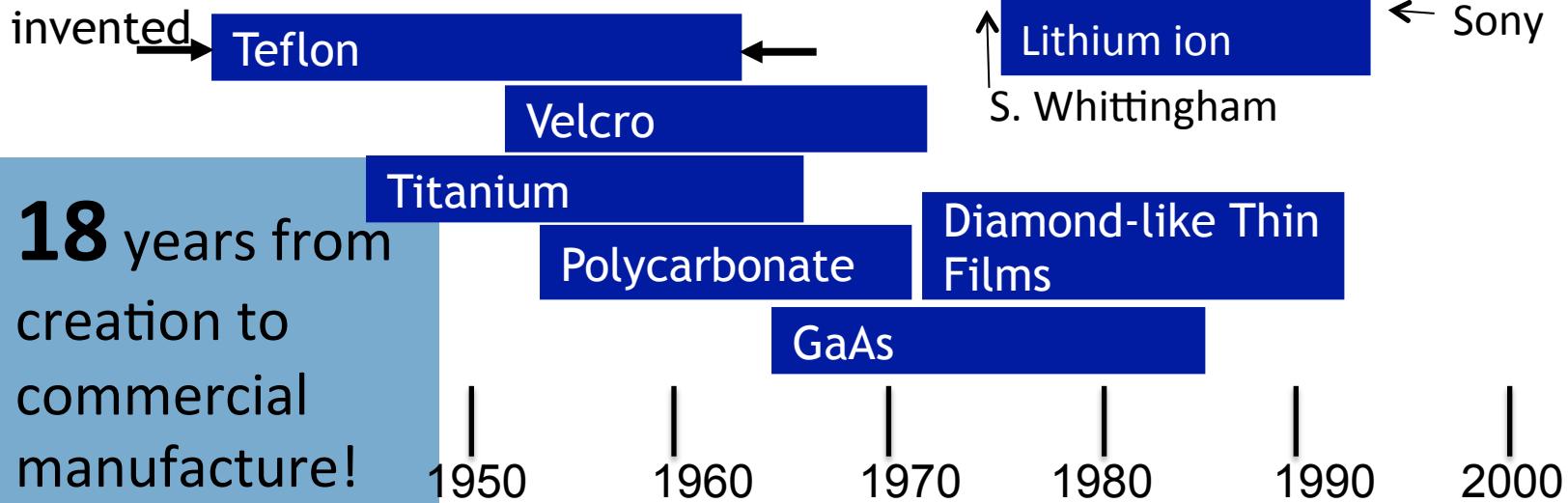
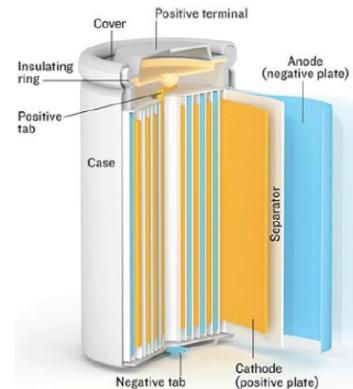


# The –omics approach

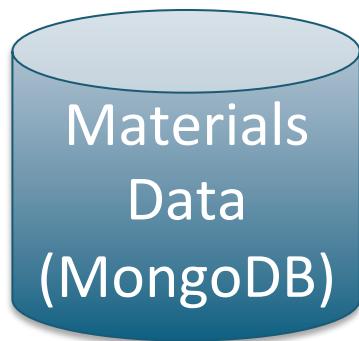
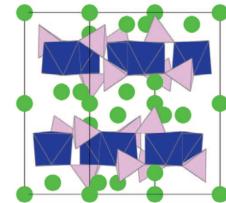
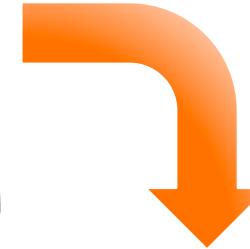
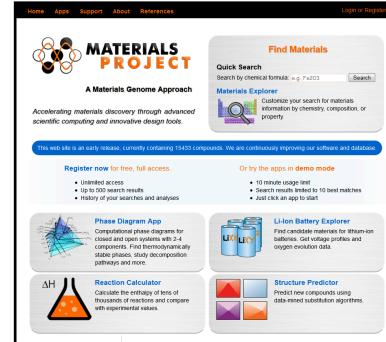
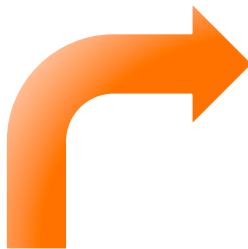
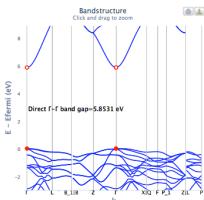


- Instead of trying to derive a solution and compute the results, just compute the space of all possibilities and look for the optimal result in there.

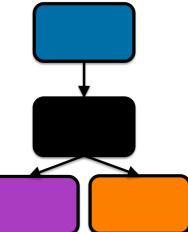
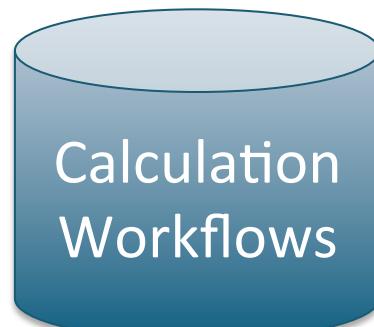
# "Need for speed" in new materials



# Infrastructure



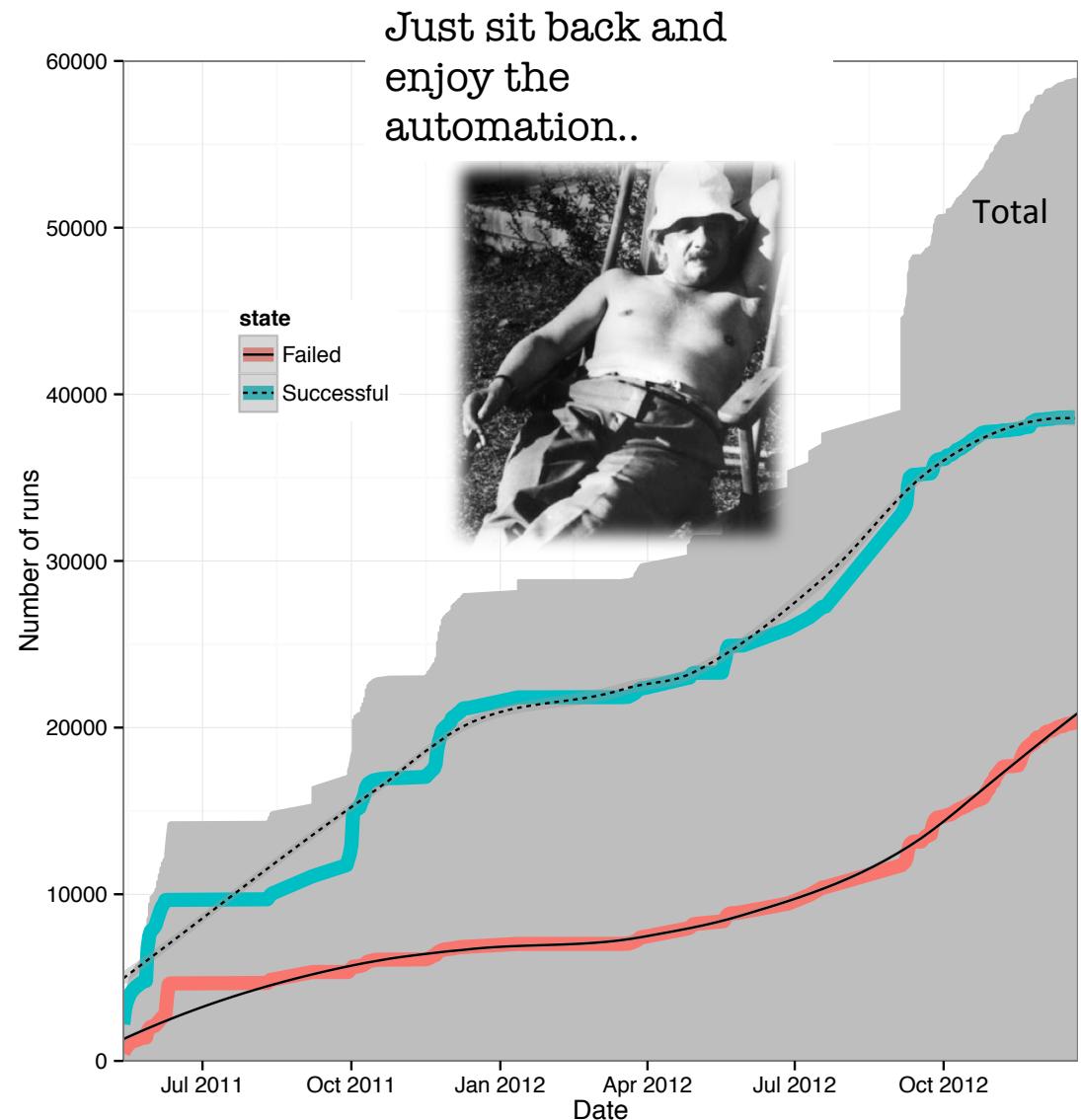
- Over 10 million CPU hours of calculations in < 6 months
- Over 40,000 successful VASP runs (30,000+ materials)
- Generalizable to other high-throughput codes/analyses



# Computation



- Run VASP on NERSC supercomputing resources
- Use Fireworks to manage large groups of runs
- Results in ... data for MP



# Application Stack



- Python Django Web Portal
- Pymatgen and other scientific Python libraries
- Fireworks + VASP



**MongoDB under all these layers**

- Currently at Mongo 2.2

# Why we like MongoDB



- **Flexible Schema**
- **Developer-friendliness**
- **JSON**
- **Great for read-heavy patterns**

# The Materials Project



<http://materialsproject.org/>

The screenshot shows the Materials Project homepage. At the top, there's a navigation bar with links for Home, Apps, Resources, About, and References, along with a Profile for scholia@lbl.gov and a Logout button. Below the navigation is the Materials Project logo and a tagline: "A Materials Genome Approach". A search bar is present with the placeholder "e.g. explore Fe2O3 or Li-Fe-O pd" and a "Search" button powered by MOOGLE. To the right, there are "Database Statistics": 30732 materials, 3044 bandstructures, 405 intercalation batteries, and 15175 conversion batteries. A section for "We're hiring!" offers postdoc and web developer opportunities. Below this are six app cards: Materials Explorer, Lithium Battery Explorer, Crystal Toolkit, Phase Diagram App, Reaction Calculator, and Structure Predictor. A note at the bottom of this section encourages users to learn more about the open Materials API and pymatgen library. On the left side, there are links for Tutorials, MATERIALS PROJECT Tube, and Press Highlights from The New York Times. On the right, there's a "Latest News" section with a pymatgen v2.5.2 update notice.

**MATERIALS PROJECT**  
A Materials Genome Approach

Accelerating materials discovery through advanced scientific computing and innovative design tools.

We're hiring! See our new opportunities for postdocs and web developers!

**Database Statistics**

30732 materials    3044 bandstructures  
405 intercalation batteries    15175 conversion batteries

**Materials Explorer**  
Search for materials information by chemistry, composition, or property.

**Lithium Battery Explorer**  
Find candidate materials for lithium batteries. Get voltage profiles and oxygen evolution data.

**Crystal Toolkit**  
Convert between CIF and VASP input files. Generate new crystals by substituting or removing species.

**Phase Diagram App**  
Computational phase diagrams for closed and open systems. Find stable phases and study reaction pathways.

**Reaction Calculator**  
Calculate the enthalpy of tens of thousands of reactions and compare with experimental values.

**Structure Predictor**  
Predict new compounds using data-mined substitution algorithms.

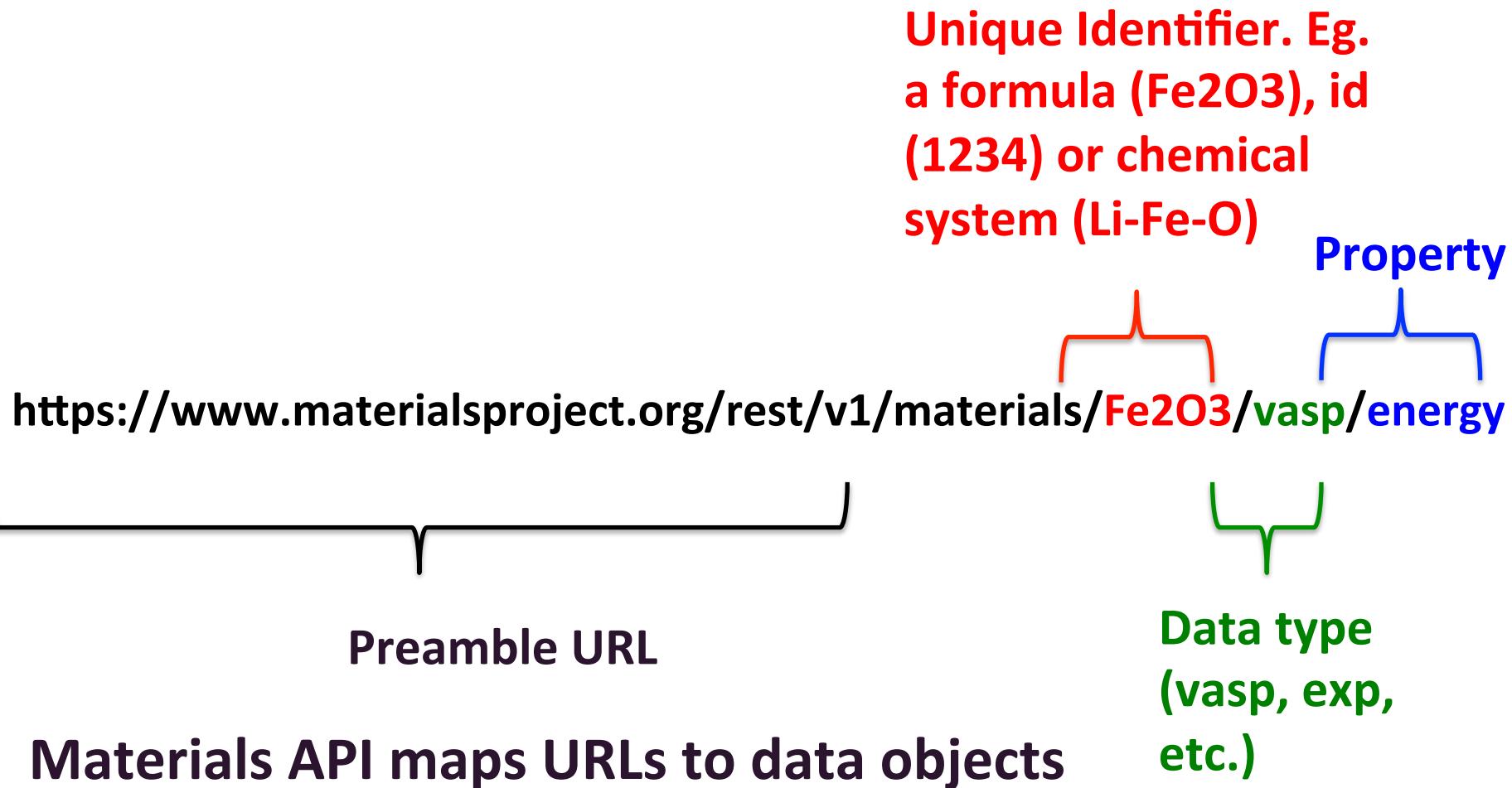
Find out more about our [open Materials API](#) and [pymatgen library](#) for querying large amounts of data.

**Tutorials**  
**MATERIALS PROJECT Tube**  
**Press Highlights**  
The New York Times

**Latest News**

pymatgen v2.5.2  
by Shyue Ping Ong - Feb 19, 2013  
Version 2.5.2 of pymatgen has been released! Check out the change log at <https://pypi.python.org/pypi/pymatgen> for information on the latest changes.

# Materials API – Data As A Service



# Sample JSON output



GET <https://www.materialsproject.org/rest/v1/materials/Fe2O3/vasp/energy>

```
{  
    "created_at": "2013-03-17T09:14:58.158081",  
    "valid_response": true,  
    "version": {  
        "pymatgen": "2.5.4",  
        "db": "2013.02.25",  
        "rest": "1.0"  
    },  
    "response": [  
        {"energy": -132.33005625,  
         "material_id": 542309},  
        {"energy": -66.62512425,  
         "material_id": 24972}  
    ],  
    "copyright": "Copyright 2012, The Materials Project"
```

}

Just the energy and the  
id of the material

# Takeaways



- **Bigdata X-omics:** Compute lots of results and look for solutions in the resultant data set
- **Schema Last Data:** Using a MongoDB NoSQL database for science data to allow for evolving schemas
- **Data as a service:** Make data available through web interfaces and REST APIs to reduce the barrier for end user scientists

# Thanks

- **Materials Project**
  - <http://materialsproject.org>
- **Our code**
  - <https://github.com/materialsproject/pymatgen>
  - <https://github.com/materialsproject/fireworks>
- **Contact:**
  - Shreyas Cholia: [scholia@lbl.gov](mailto:scholia@lbl.gov)
- **Questions?**