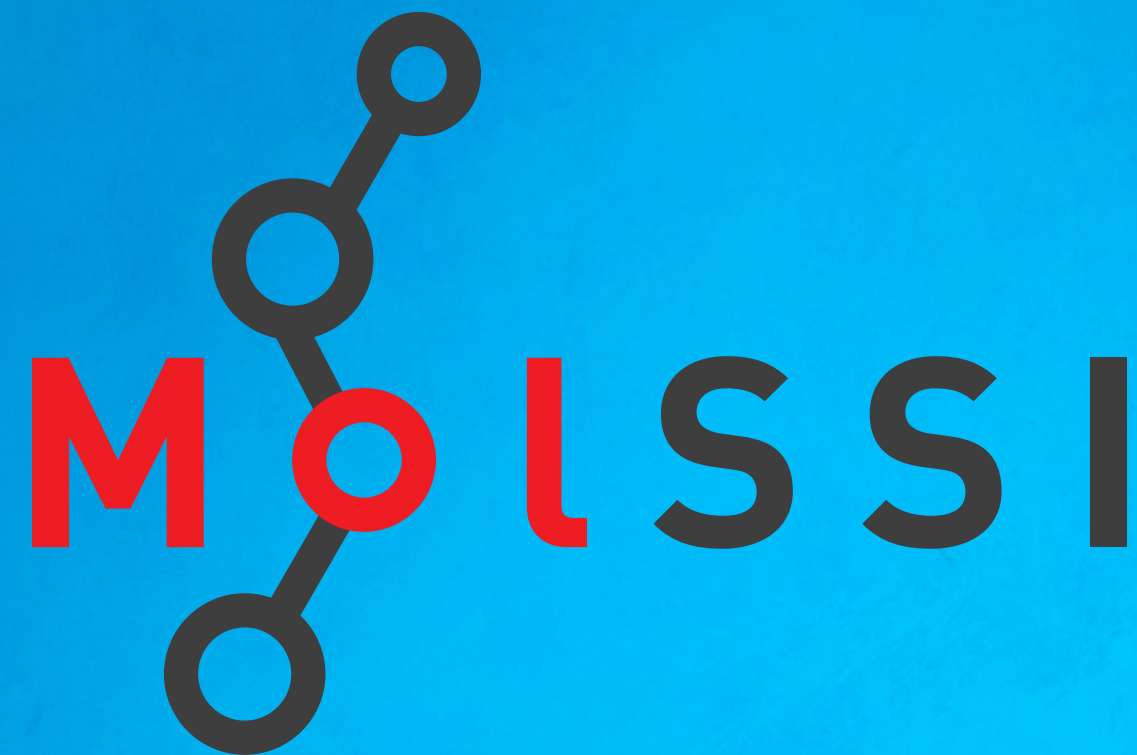# Tools for trajectory file sharing

Daniel G. A. Smith

The Molecular Sciences Software Institute

@dga_smith

molssi.org

The Molecular Sciences Software Institute (MoISSI)

- Designed to *serve* and *enhance* the software development efforts of the broad field of computational molecular science (CMS).

- Funded at the institute level under the Software Infrastructure for Sustained Innovation (SI^2) initiative by the National Science Foundation.

- Launched August 1st, 2016

- Collaborative effort by:
  - Virginia Tech
  - Rice U.
  - Stony Brook U.
  - U.C. Berkeley
  - Stanford U.
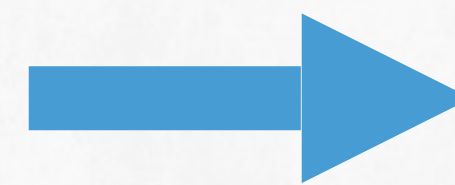 - Rutgers U.
  - U. Southern California
  - Iowa State U.

QC Schema

https://github.com/MolSSI/QCSchema

- Communication channel between all piece of the ecosystem.

- *Community* project useful for many aspects of quantum chemistry.

- Not only JSON, but any key/value/array language (BSON/HDF5/XML/YAML)

- Molecule
- Input
- Output
- Optimization Trajectory

```
{
    "molecule": {
        "geometry": [0, 0, 0, 0, 0, 1],
        "atoms": ["He", "He"]
    },
    "driver": "energy",
    "model": {
        "method": "SCF",
        "basis": "sto-3g",
    },
    "keywords": {},
}
```
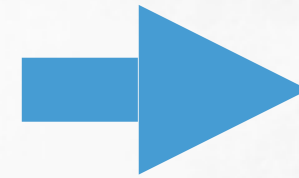
```
{
    ...Input
    "provenance": {
        "creator": "My QM Program",
        "version": "1.1rc1",
        ...
    },
    "properties": {
        "scf_n_iterations": 2.0,
        "scf_total_energy": -5.43319881443323,
        "nuclear_repulsion_energy": 2.11670883436,
        "one_electron_energy": -11.67399006298957,
        ...
    },
    "error": "",
    "success": true,
    "raw_output": "Output storing was not requested."
}
```
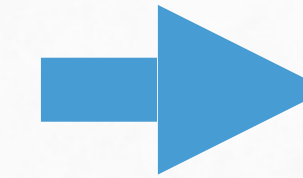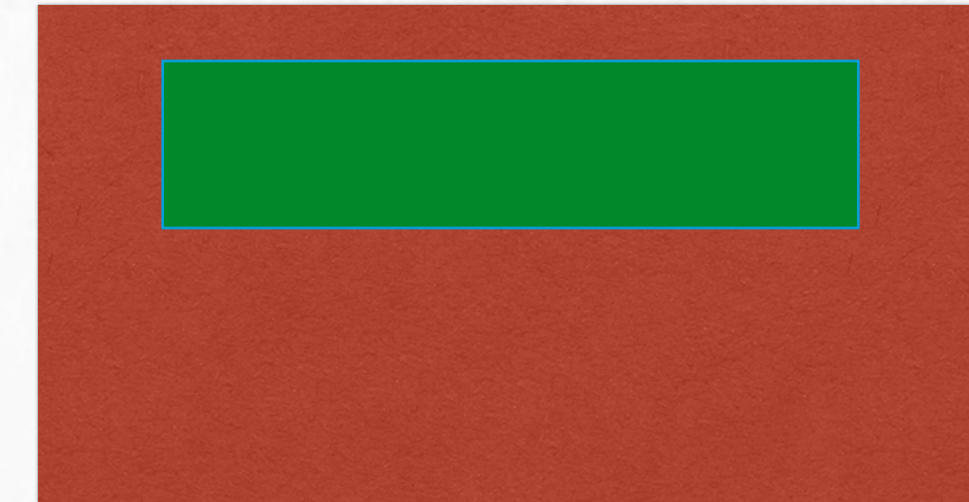
QC Schema

Molecule

```json
{
    "molecule": {
        "geometry": [0, 0, 0, 0, 0, 1],
        "atoms": ["He", "He"]
    },
}
```
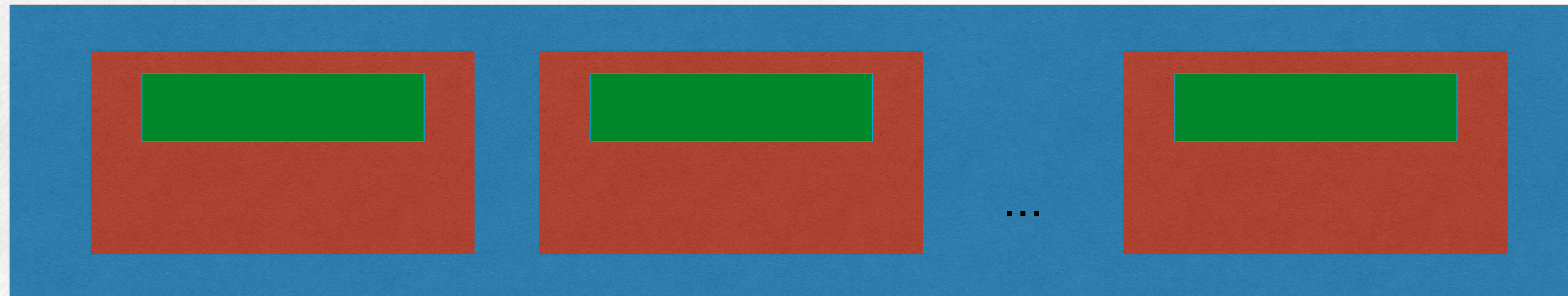
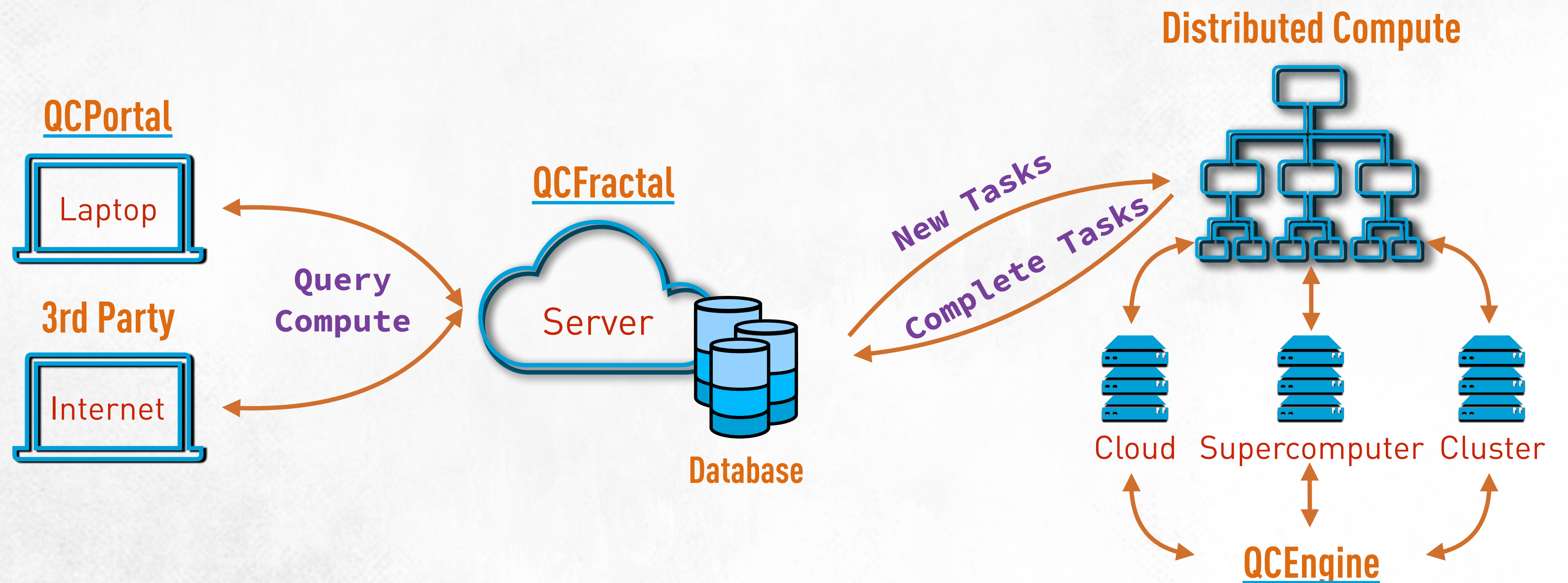### Molecule

Result

Geometry Optimization/Trajectory

...

QCArchive

Provide an open, community-wide quantum chemistry database to both facilitate and capture hundreds of millions of hours of computing time to enable large-scale forcefield construction, physical property prediction, new methodology assessment, and machine learning from data that would otherwise end up siloed or inaccessible.
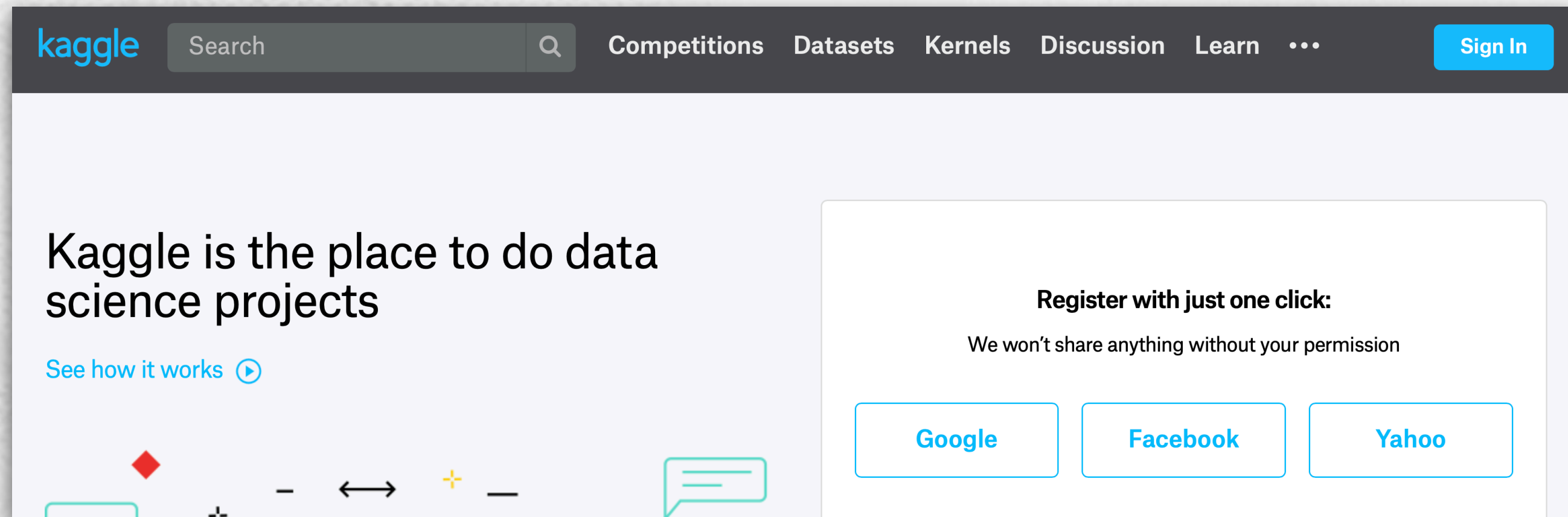
Communities of Practice

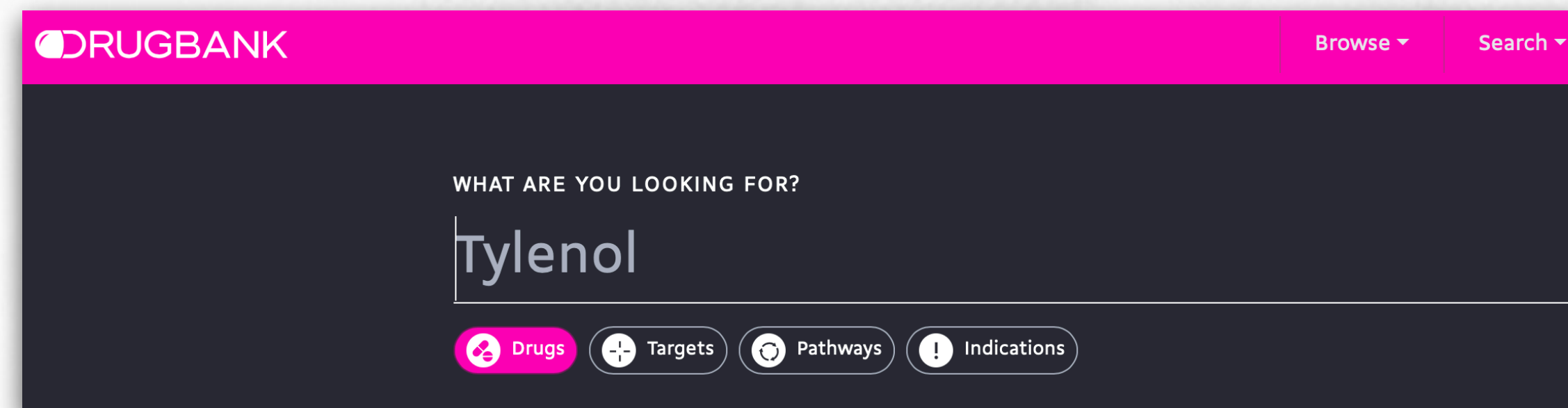"**Best**" is subjective and depends on your own community.

- Depends greatly on:
  - Target audience
  - Size of the project
  - Activity level of the project
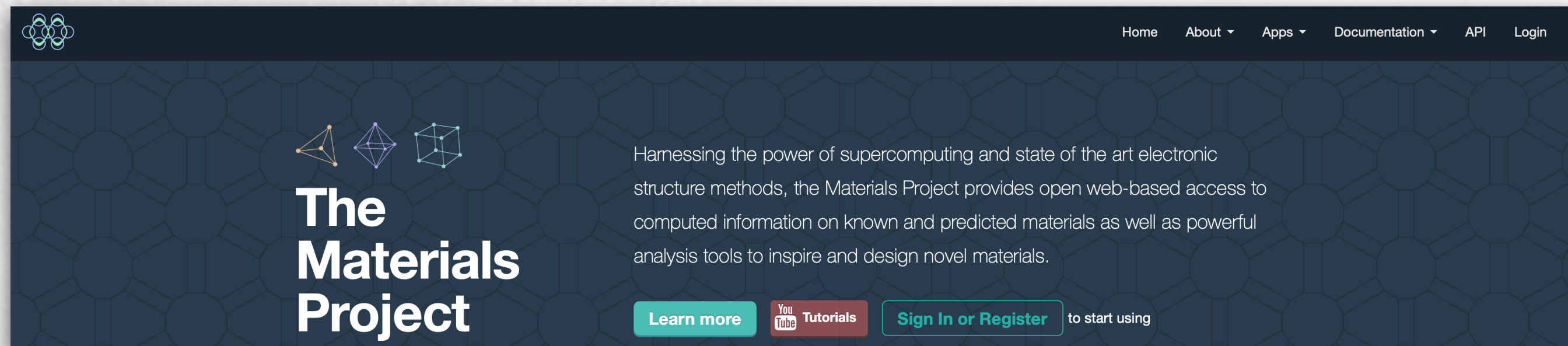- What practices are optimal for your community and what can be derived from many other communities?

Takeaway Ideas

kaggle

Search

Competitions   Datasets   Kernels   Discussion   Learn   ...   Sign In

Kaggle is the place to do data science projects

See how it works ▶

Register with just one click:

We won't share anything without your permission

Google   Facebook   Yahoo

DRUGBANK

Browse ▾   Search ▾

WHAT ARE YOU LOOKING FOR?

Tylenol

Drugs   Targets   Pathways   Indications

The Materials Project

Home   About ▾   Apps ▾   Documentation ▾   API   Login

Harnessing the power of supercomputing and state of the art electronic structure methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.

Learn more   Tutorials   Sign In or Register   to start using

CITRINE INFORMATICS

THE AI PLATFORM FOR MATERIALS DEVELOP

NOMAD

The NOMAD Laboratory
A European Centre of Excellence

PROJECT   INDUSTRY   TEAM   RELATED PROJECTS   NEWS   PRESS KIT   CONTACT US

Enter Search...

NOMAD REPOSITORY   THE ARCHIVE   ENCYCLOPEDIA   BIG-DATA ANALYTICS   ADVANCED GRAPHICS   HPC INFRASTRUCTURE   OUTREACH

Tools for trajectory file sharing

- **Focus recommendations:**
  - Avoid technical challenges for a moment.
  - Enumerate the tools involved.
  - Focus on goals of these projects rather than the tool itself.
  - Concentrate on the social considerations.
  - See if we can get the 90% case.
- **Gotchas:**
  - Databases can cross reference each other (and frequently do).
  - Best to assume trajectory format transformation is possible.
  - Assume hosting is possible.