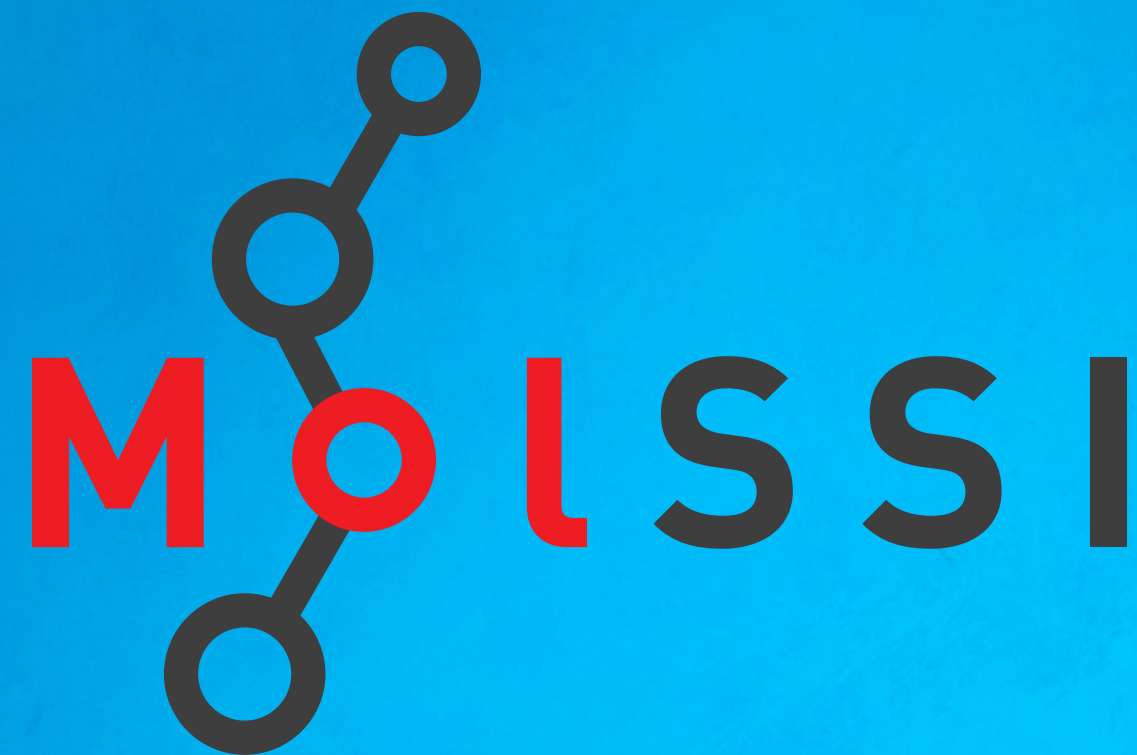


# Tools for trajectory file sharing



Daniel G. A. Smith

The Molecular Sciences Software Institute

@dga\_smith

[molssi.org](http://molssi.org)

# The Molecular Sciences Software Institute (MolSSI)

- 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<sup>2</sup>) 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.



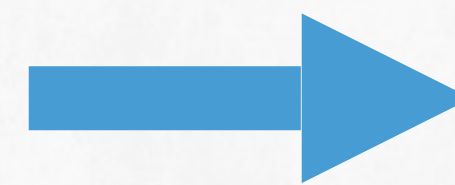
# QC Schema

<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": {},
}
```



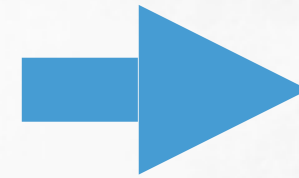
```
{
  ...Input
  "provenance": {
    "creator": "My QM Program",
    "version": "1.1rc1",
    ...
  },
  "properties": {
    "scf_n_iterations": 2.0,
    "scf_total_energy": -5.433191881443323,
    "nuclear_repulsion_energy": 2.11670883436,
    "one_electron_energy": -11.67399006298957,
    ...
  },
  "error": "",
  "success": true,
  "raw_output": "Output storing was not requested."
}
```

# QC Schema

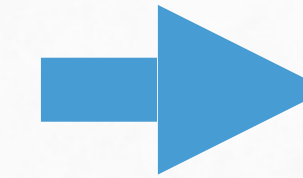
<https://github.com/MolSSI/QCSchema>

Molecule

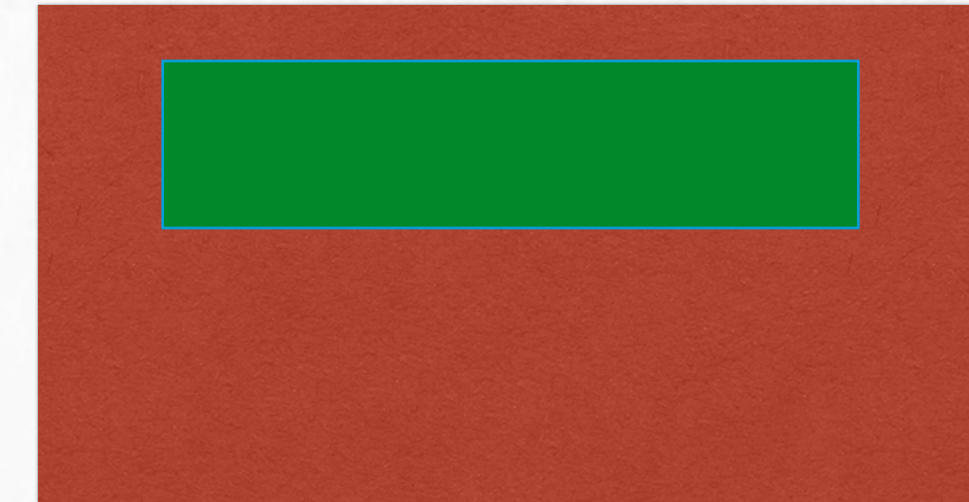
```
{  
  "molecule": {  
    "geometry": [0, 0, 0, 0, 0, 1],  
    "atoms": ["He", "He"]  
  },  
}
```



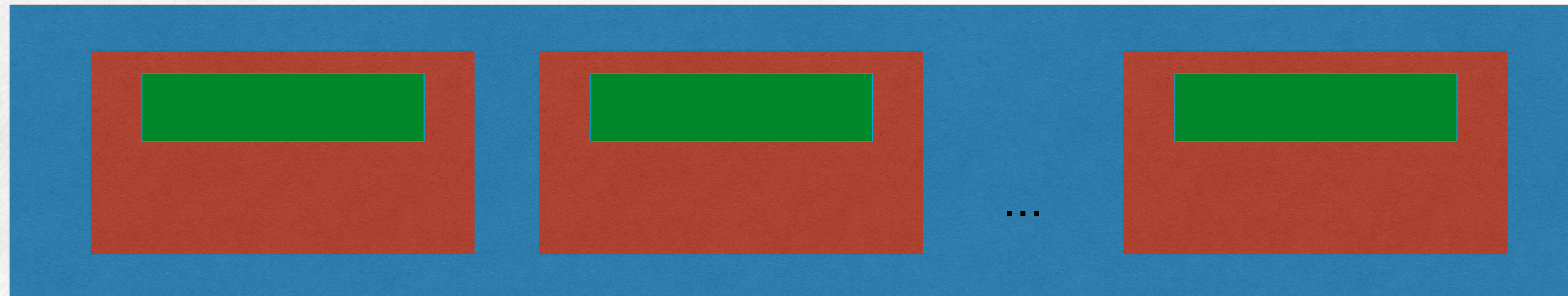
Molecule



Result

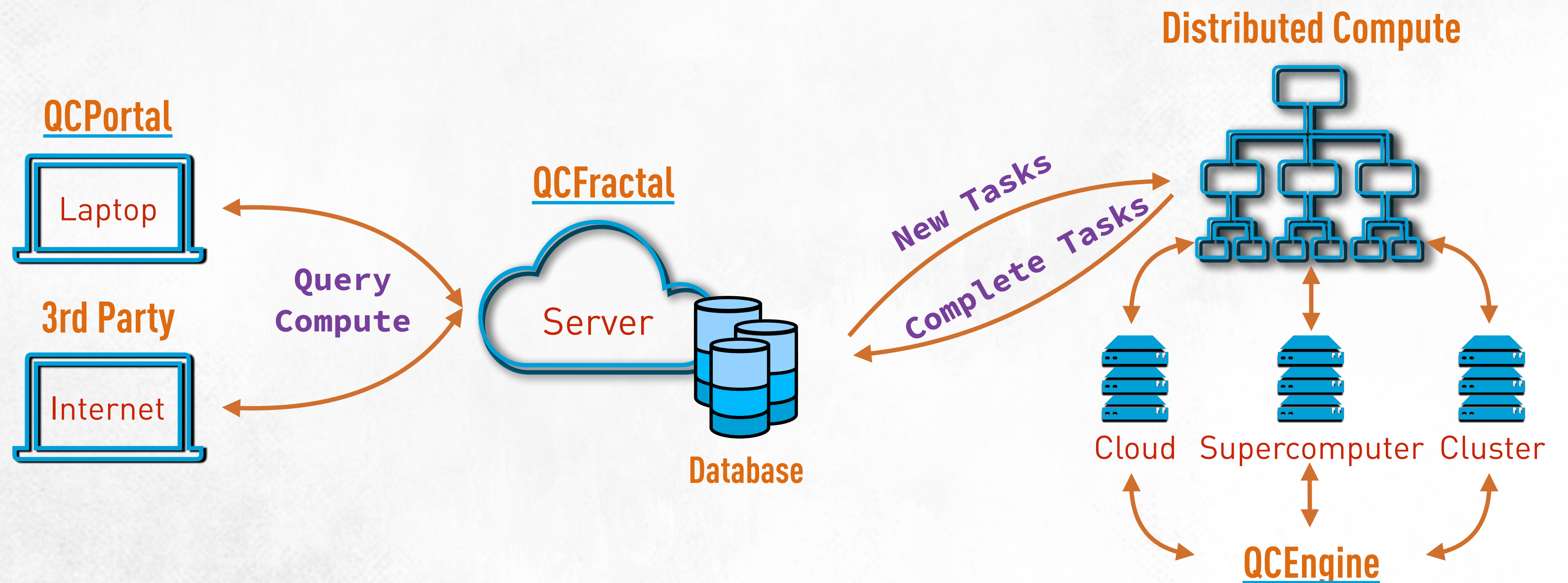


Geometry Optimization/Trajectory



# 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.



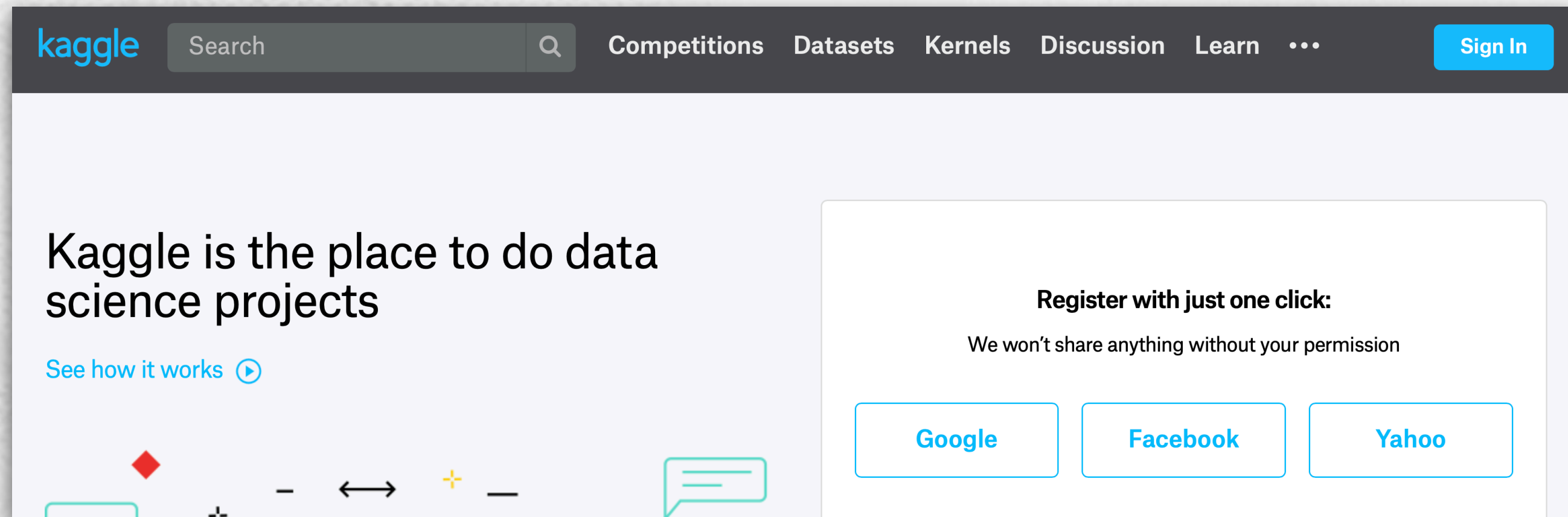
# Communities of Practice

“**Best**” is subjective and depends on your own community.

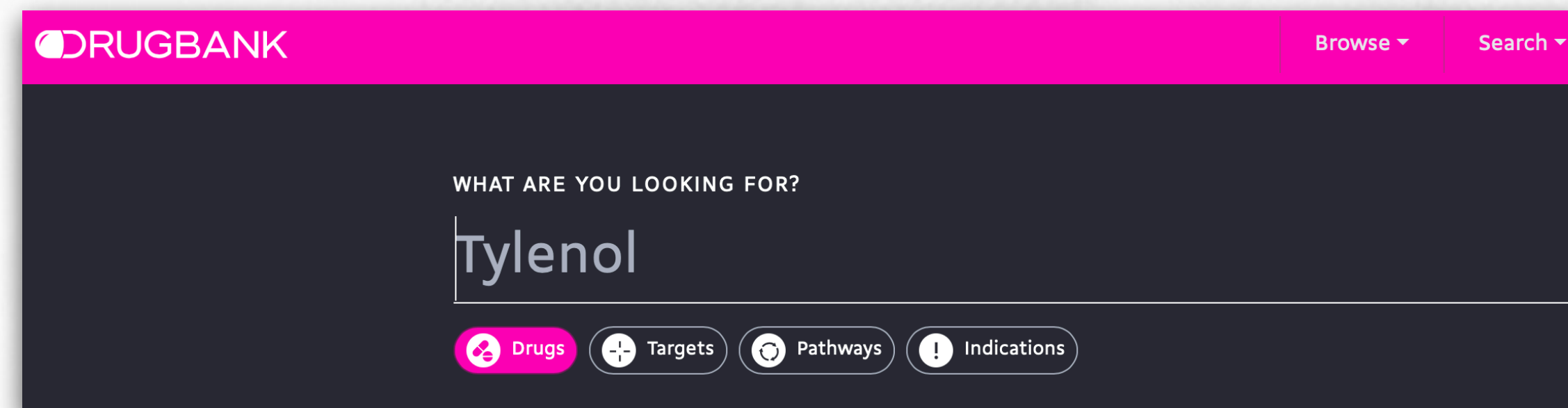
- Depends greatly on:
  - Target audience
  - Size of the project
  - Activity level of the project
- What practices are optimal for your community and what can be derived from many other communities?



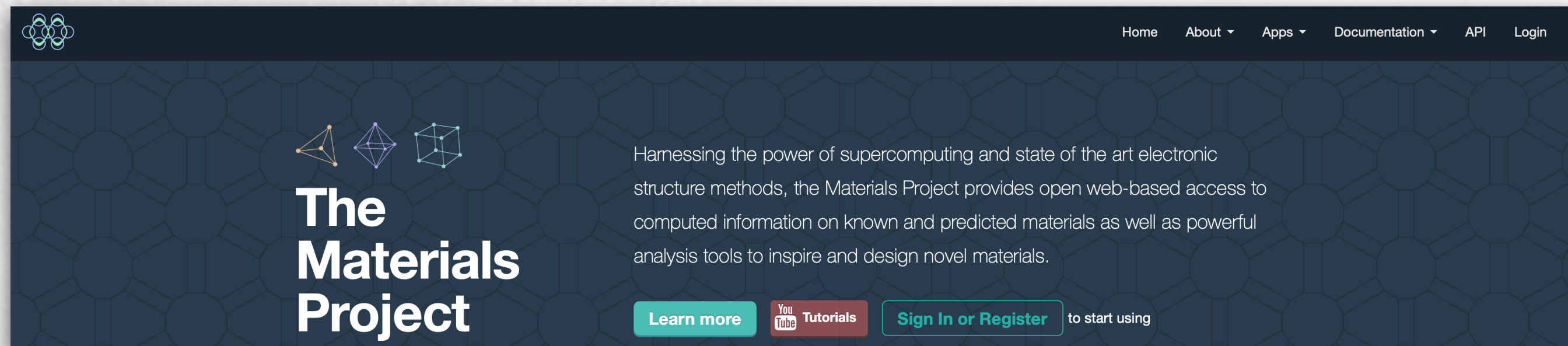
# Takeaway Ideas



The image shows the top section of the Kaggle website. It features a dark navigation bar with the Kaggle logo, a search bar, and links for Competitions, Datasets, Kernels, Discussion, and Learn. A prominent blue 'Sign In' button is on the right. Below the navigation bar, a white box contains the text 'Kaggle is the place to do data science projects' and a 'See how it works' link. To the right, a registration prompt says 'Register with just one click: We won't share anything without your permission' and offers buttons for Google, Facebook, and Yahoo.



The image shows the DrugBank website search interface. It has a bright pink header with the DrugBank logo and navigation links for 'Browse' and 'Search'. The main content area is dark grey and features the text 'WHAT ARE YOU LOOKING FOR?' above a search input field containing 'Tylenol'. Below the search field are four filter buttons: 'Drugs', 'Targets', 'Pathways', and 'Indications'.



The image shows the header of 'The Materials Project' website. It has a dark blue background with a hexagonal pattern. The title 'The Materials Project' is in large white font, accompanied by icons of geometric shapes. A paragraph of text describes the project's mission: '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.' Below the text are buttons for 'Learn more', 'YouTube Tutorials', and 'Sign In or Register to start using'.



The image shows the header of the Citrine Informatics website. It features a light grey header with the 'CITRINE INFORMATICS' logo. Below the header is a dark banner with the text 'THE AI PLATFORM FOR MATERIALS DEVELOPMENT' in white.



The image shows the header of 'The NOMAD Laboratory' website. It has a dark blue background with the 'NOMAD' logo in a stylized font. The text 'The NOMAD Laboratory A European Centre of Excellence' is prominently displayed. Below this is a navigation bar with links for PROJECT, INDUSTRY, TEAM, RELATED PROJECTS, NEWS, PRESS KIT, and CONTACT US, along with a search bar. At the bottom, there is a row of seven colored boxes, each with an icon and a label: NOMAD REPOSITORY (purple), THE ARCHIVE (teal), ENCYCLOPEDIA (orange), BIG-DATA ANALYTICS (green), ADVANCED GRAPHICS (yellow), HPC INFRASTRUCTURE (blue), and OUTREACH (maroon).

# 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.