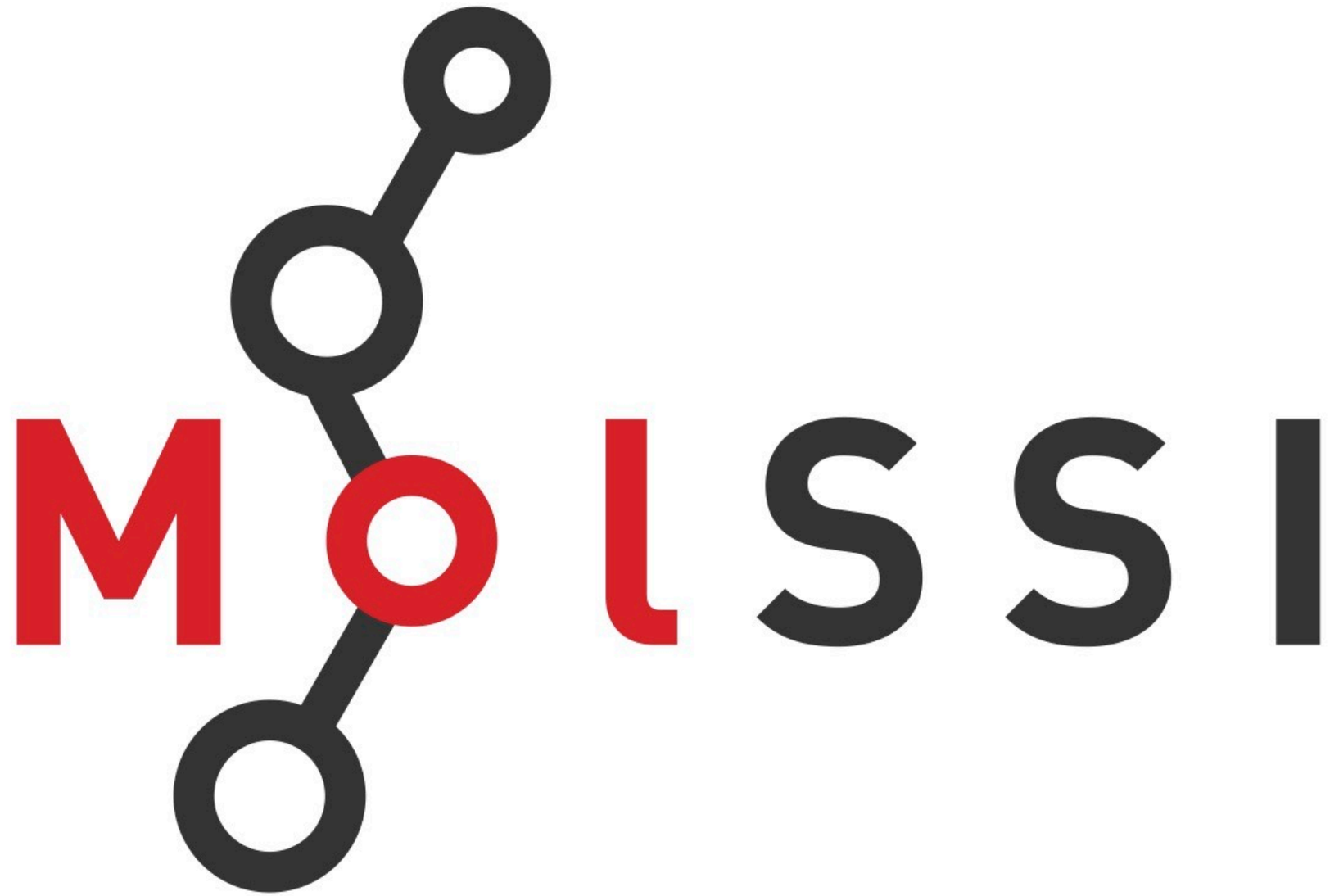


The Molecular Sciences Software Institute

... a nexus for science, education, and cooperation for the global computational molecular sciences community.

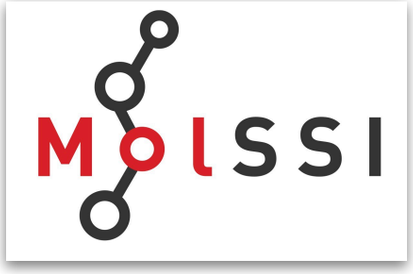
What is the MolSSI?

- Launched August 1st, 2016, funded by the National Science Foundation.
- Collaborative effort by Virginia Tech (TDC), Rice U. (C. Clementi), Stony Brook U. (R. Harrison), U.C. Berkeley (T. Head-Gordon), Stanford U. (V. Pande), Rutgers U. (S. Jha), U. Southern California (A. Krylov), and Iowa State U (T. Windus).
- Part of the NSF's commitment to the White House's National Strategic Computing Initiative (NSCI).
- Total budget of $19.42M for five years, potentially renewable to ten years.
- Joint support from numerous NSF divisions: Advanced Cyberinfrastructure (ACI), Chemistry (CHE), and Division of Materials Research (DMR)
- Designed to **serve** and **enhance** the software development efforts of the broad field of computational molecular science.

MolSSI

Code Complexity and Historical Legacy

- CMS programs contain **millions of lines of hand-written code** and require **hundreds of programmers** to develop and maintain.
- Incredible **language diversity**: F77, F90, F95, HPF, C, C++, C++11, C++14, C++17, Python, perl, Javascript, etc.
- Incredible **algorithmic diversity**: structured and unstructured grids, dense and sparse linear algebra, graph traversal, fast Fourier transforms, MapReduce, and more.
- The packages have evolved in an ad hoc manner **over decades** because of the intricacy of the scientific problems they are designed to solve.



Rapidly Evolving Computing Hardware

- **Multi- and many-core architectures** are the norm, but many CMS codes are developed with limited view to parallel task management.
- Reduced-power solutions will also require improved **error recovery and checkpointing** at the software level – capabilities absent in nearly all CMS codes.
- Anticipated architectural innovations will yield **even greater hardware complexity** – more advanced accelerators, specialized computing cores, reconfigurable logic…
- Many CMS codes (especially for quantum chemistry) are limited to shared-memory paradigms and cannot yet take advantage of GPUs or **large-scale distributed-memory systems**.

MolSSI

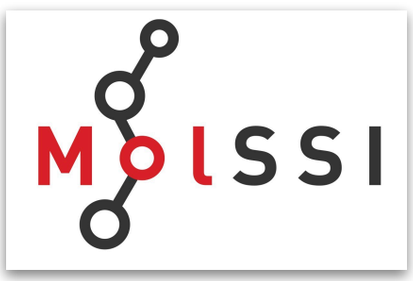
Inertia in the Scientific Education Culture

- Undergraduate programs in chemistry and physics typically require **no training in software development** or programming.
- Graduate programs in these areas require **minimal coursework** between the bachelor and Ph.D.
- Most computer science students lack the underlying knowledge of the scientific domains to help develop creative software solutions.
- **Due credit for software development** is elusive due to a culture that judges productivity based on citations of peer-reviewed papers.
- Thus, a **"just get the physics working"** approach pervades much of CMS software development.

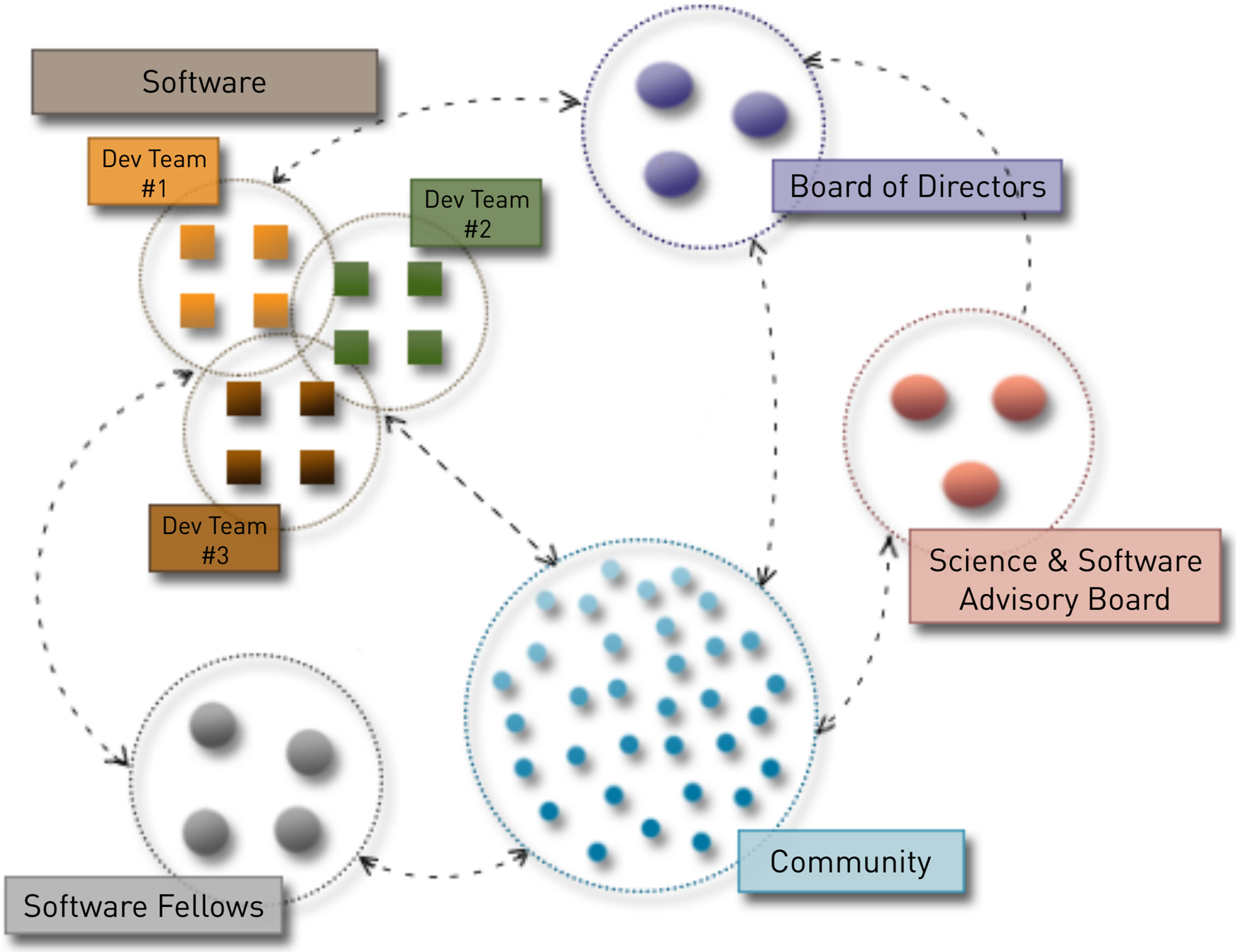
MolSSI

MolSSI Goals

- To provide software expertise and infrastructure
  - Current software projects, filling gaps
- To provide education and training
  - Summer school, best practices
- To provide community engagement and leadership
  - Working groups, standards



The Molecular Sciences Software Institute



MolSSI Software Scientists (MSSs)

- A team of ~12 software engineering experts, drawn both from newly minted Ph.D.s and established researchers in molecular sciences, computer science, and applied mathematics.
- Dedicated to multiple responsibilities:
  - Developing software infrastructure and frameworks;
  - Interacting with CMS research groups and community code developers;
  - Providing forums for standards development and resource curation;
  - Serving as mentors to MolSSI Software Fellows;
  - Working with industrial, national laboratory, and international partners;

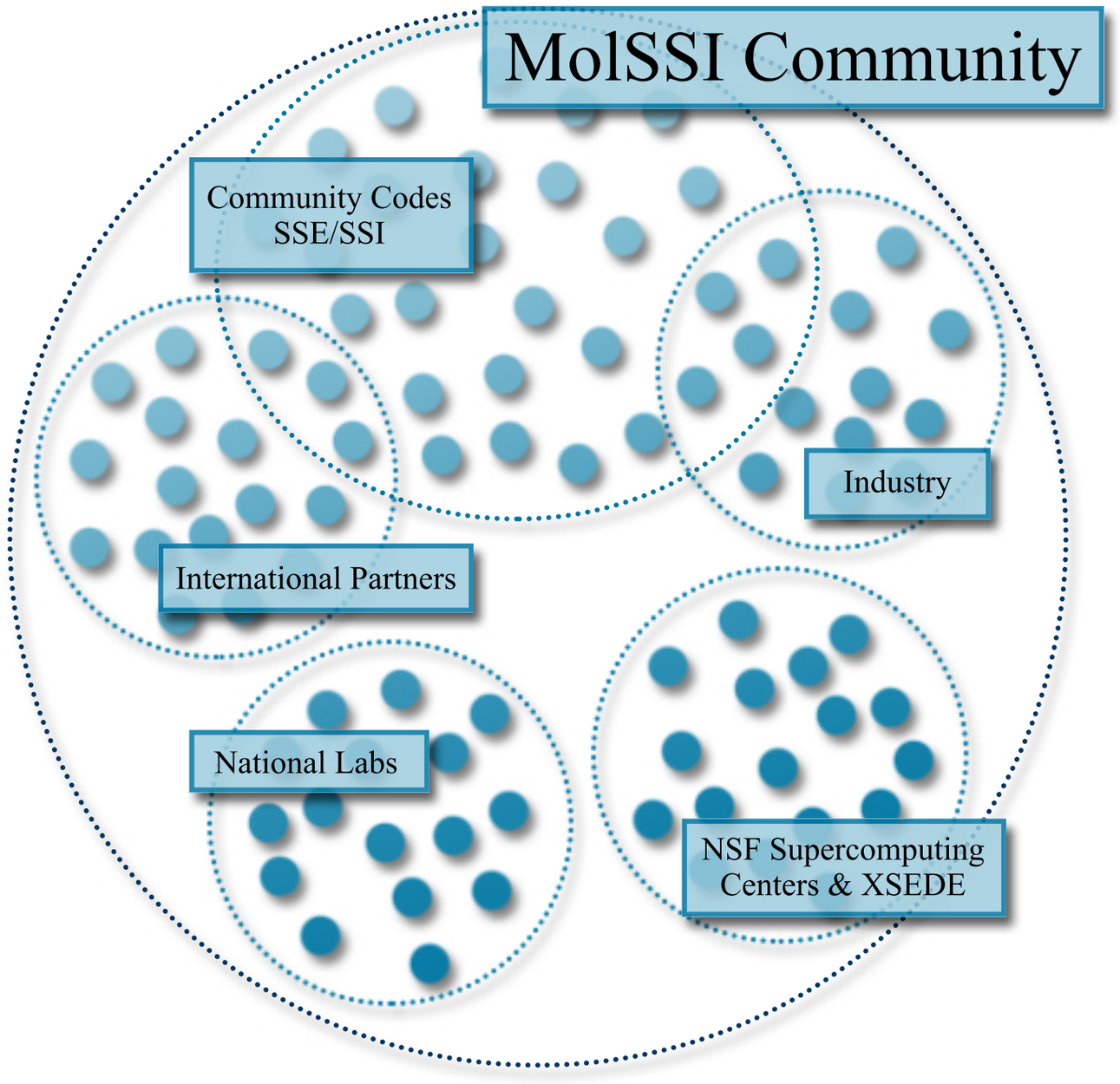
Currently 7 MSS at MolSSI, 2 more accepted

MolSSI Software Fellows (MSFs)

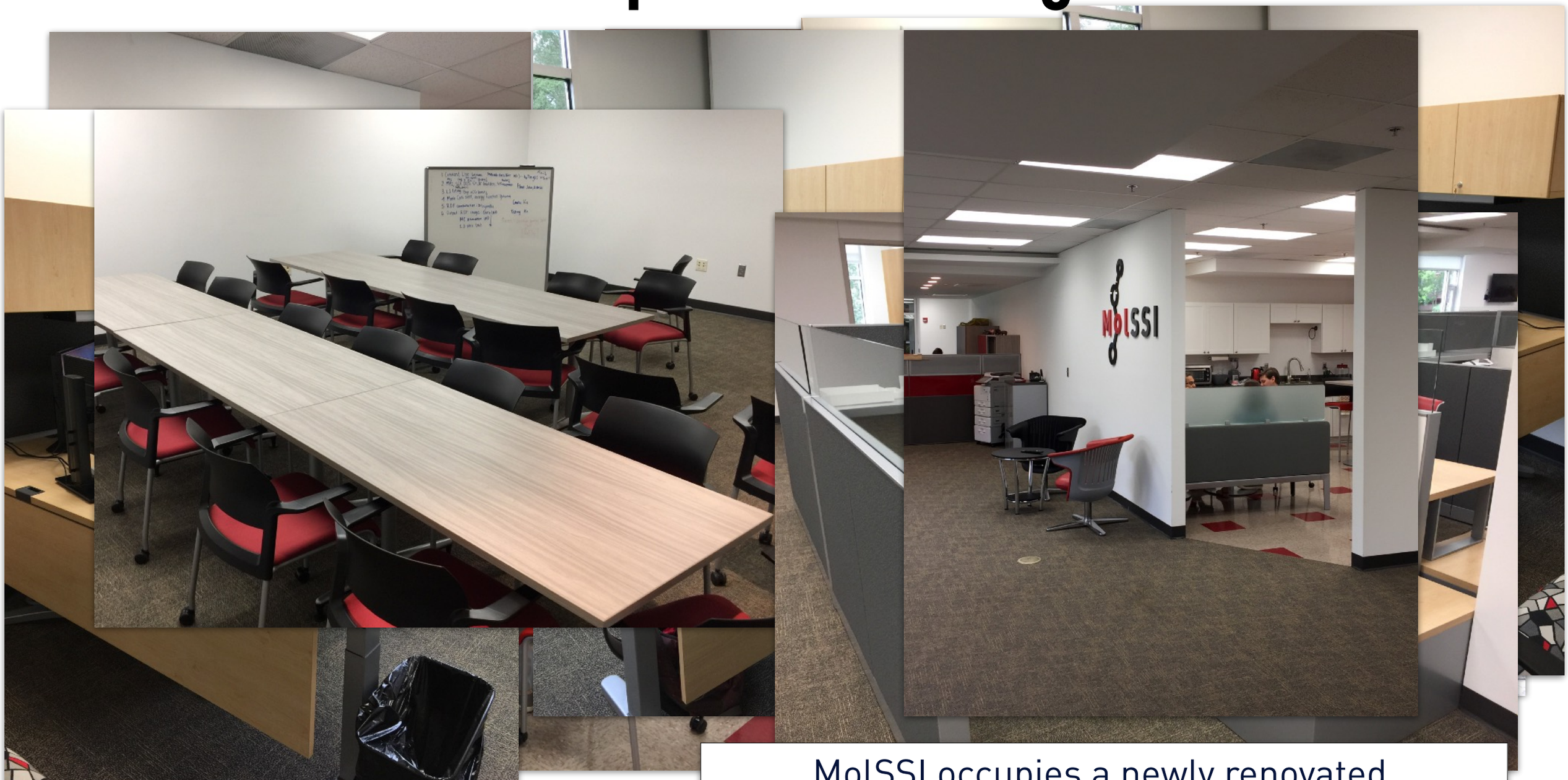
- A cohort of ~20 Fellows supported simultaneously – graduate students and postdocs selected by the Science and Software Advisory Board from research groups across the U.S.
- Fellows work directly with both the Software Scientists and the MolSSI Directors, thus providing a conduit between the Institute and the CMS community itself.
- Fellows work on their own projects, as well as contribute to the MolSSI development efforts, and they will engage in outreach and education activities under the Institute guidance.
- Funding for MolSSI Software Fellows follows a flexible, two-phase structure, providing up to two years of support.

MolSSI

The MolSSI Community



MolSSI Headquarters @ Virginia Tech



MolSSI occupies a newly renovated,
6,900 sq. ft. facility adjacent to campus.

MolSSI

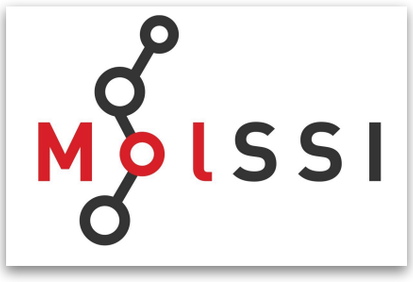
MolSSI Integral Reference Project

https://github.com/MolSSI/mirp

- Reference implementation and values
- Utilizes arbitrary-precision interval arithmetic (ball arithmetic)
- **Very slow,** but relatively simple implementation

```
4.785065404705502970263665171263153090347776322991832463900955205746500551584592749047052813525448252 6 +/- 4.63e-101
```

“Exact” double precision: `0x1.323e82f79b97dp+2`



Basis Set Exchange

EMSL Office of Science

BASIS SET EXCHANGE

Username: bennyp Password: ••••••••••••

Login | Become a Contributor

Basis Set Exchange: v1.2.2

Feedback | About | ReleaseNotes | Help

All

2ZaPa-NR
2ZaPa-NR_CV
3-21++G
3-21++G*
3-21G
3-21G*
3-21G* Polarization
3-21GSP
3ZaPa-NR
3ZaPa-NR_CV
4-22GSP
4-31G
4ZaPa-NR
4ZaPa-NR_CV
5ZaPa-NR

Search Basis Set Name

Total: 601 published basis sets

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Uun	Uuu	Uub	Uut	Uuq	Uup	Uuh	Uus	Uuo
			Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	
			Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr	

Format: NWChem ☑ Optimized General Contractions Get Basis Set

"2ZaPa-NR" Basis Set Information

Summary: Ranasinghe-Petersson 2ZaPa-NR basis set
Primary Developer: George A. Petersson
Last Modified: Mon, 11 Jul 2016 17:05:01 GMT
Contributor: Jan M.L. Martin
Curation Status: published

More information...
User annotations...

When publishing results obtained from use of the Basis Set Exchange (BSE) software and the EMSL Basis Set Library, please cite:

The Role of Databases in Support of Computational Chemistry Calculations
Feller, D., J. Comp. Chem., 17(13), 1571-1586, 1996.

Basis Set Exchange: A Community Database for Computational Sciences
Schuchardt, K.L., Didier, B.T., Elsethagen, T., Sun, L., Gurumoorthi, V., Chase, J., Li, J., and Windus, T.L.
J. Chem. Inf. Model., 47(3), 1045-1052, 2007, doi:10.1021/ci600510j.

Security and Privacy | Disclaimer

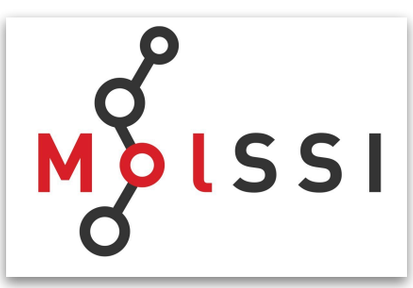
nECS Knowledge Environment for Collaborative Science — Scientific Annotation Middleware — powered by Chef — Jetspeed

KnECS v1.0 | SAM v2.1.4b8 | CHEF v1.1.01 [build #307231] | Jetspeed v1.4b2[cvs08oct2002p]

MolSSI

Current BSE

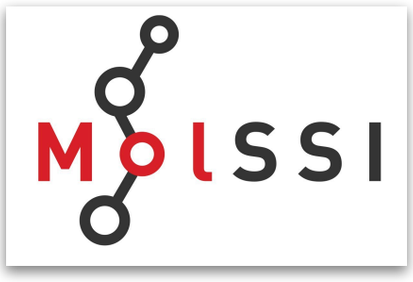
- Recognized as a central source
- Interface is generally liked
- Needs some improvements
  - “Select All” button
  - **Slow** and hard to maintain (due to backend structure)
  - Some mistakes in the data
  - Could use some alternative ways of accessing data programmatically



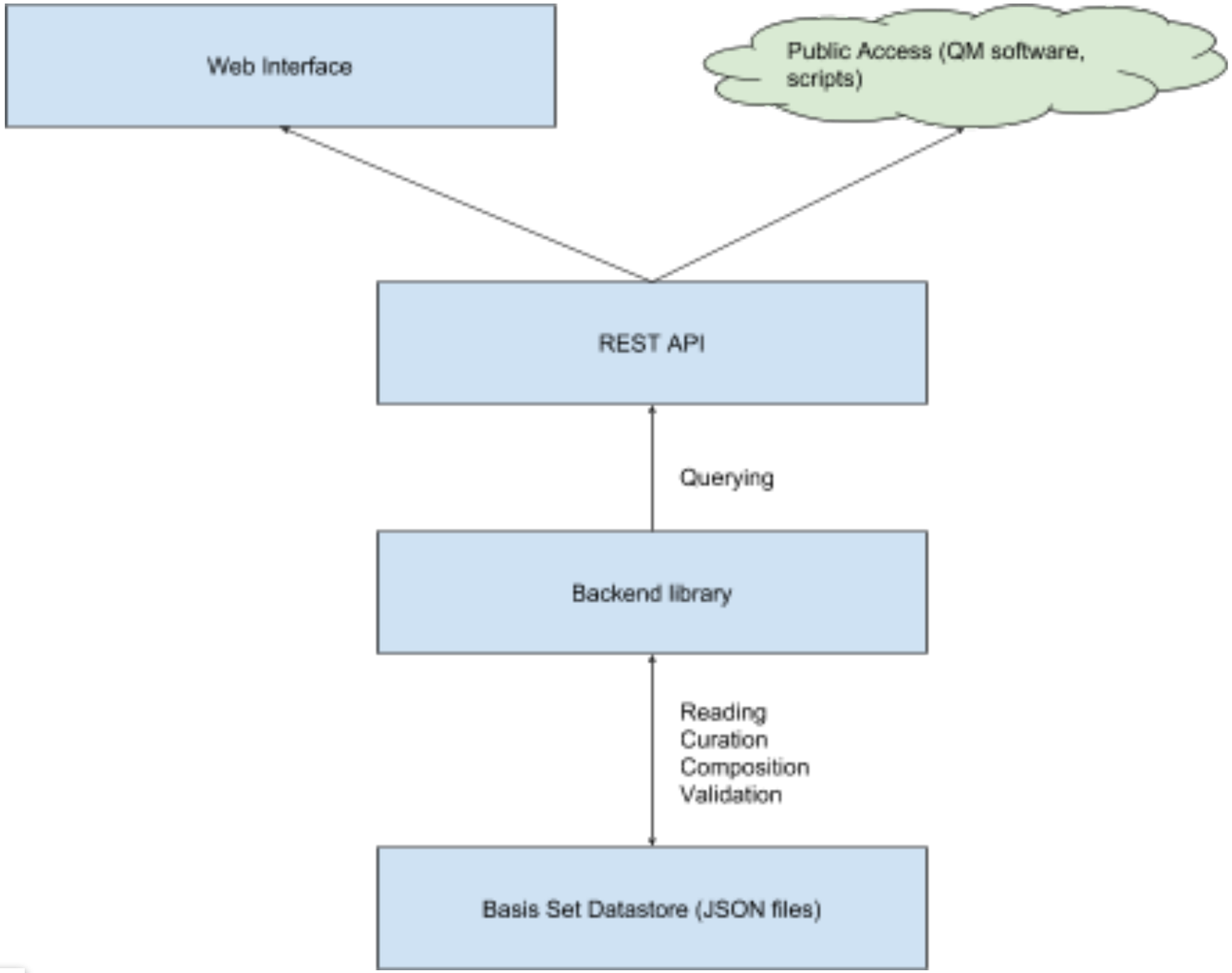
Basis Set Exchange v2

- Newer formats and languages (Python + JSON)
- Separate functionality into modules
  - Data + Library
  - Web frontend (Doaa)
  - Curate data, fixing references and errors
  - Develop unique identifiers (including versioning)
  - Collaboration with PNNL and others

https://github.com/MolSSI-BSE/basis_set_exchange



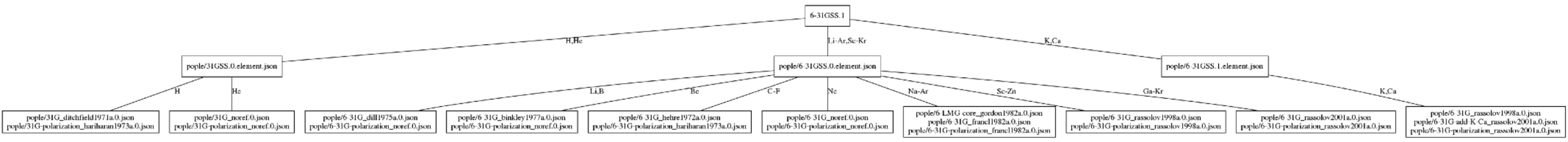
Basis Set Exchange v2



Basis Set Curation

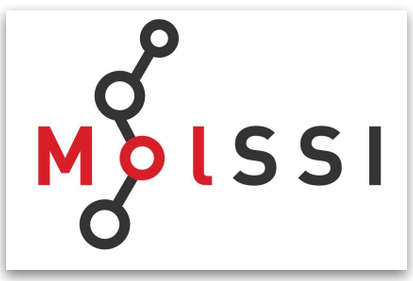
Basis sets can be complicated

- Decimal places
- Additions & corrections
- Multiple descendants
- Differing opinions on scaling factors, etc
- Unknown provenance



BSE Command Line

```
>>> import bse
>>> print(bse.get_basis("6-31G**", elements=[1,6], fmt="nwchem"))
# Basis set: 6-31G**
BASIS "ao basis" PRINT
#BASIS SET: (4s,1p) -> [2s,1p]
H    S
     18.731137              0.0334946
      2.8253944             0.2347269
      0.6401217             0.8137573
H    S
      0.1612778             1.0000000
H    P
      1.1000000             1.0000000
#BASIS SET: (10s,4p,1d) -> [3s,2p,1d]
C    S
   3047.5249000             0.0018347
    457.3695100             0.0140373
    103.9486900             0.0688426
     29.2101550             0.2321844
      9.2866630             0.4679413
      3.1639270             0.3623120
C    SP
      7.8682724            -0.1193324             0.0689991
      1.8812885            -0.1608542             0.3164240
      0.5442493             1.1434564             0.7443083
C    SP
      0.1687144             1.0000000             1.0000000
C    D
      0.8000000             1.0000000
END
```



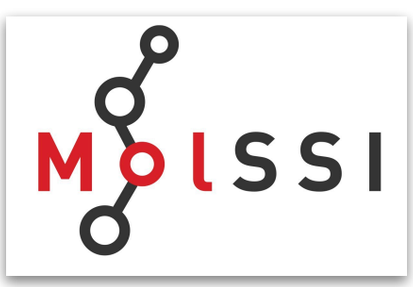
BSE Command Line

```
>>> print(bse.get_references("6-31G**", elements=[1,6], fmt="txt"))
H
    R. Ditchfield, W. J. Hehre, J. A. Pople
    J. Chem. Phys., 54, 724-728 (1971)
    10.1063/1.1674902

    P. C. Hariharan, J. A. Pople
    Theor. Chim. Acta, 28, 213-222 (1973)
    10.1007/bf00533485

C
    P. C. Hariharan, J. A. Pople
    Theor. Chim. Acta, 28, 213-222 (1973)
    10.1007/bf00533485

    W. J. Hehre, R. Ditchfield, J. A. Pople
    J. Chem. Phys., 56, 2257-2261 (1972)
    10.1063/1.1677527
```



Basis Set Exchange v2

BASIS SET EXCHANGE ver. 2.0 Release Notes Feedback Help

All

3-21G
4-31G
5-21G
6-21G
6-31++G
6-31++G*
6-31++G**
6-31+G
6-31+G*
6-31+G**
6-311G
6-311G*
6-311G**
6-31G
6-31G*
6-31G**
cc-pV5Z
cc-pV6Z
cc-pVDZ
cc-pVQZ
cc-pVTZ
CRENBL ECP

search basis sets...

Total found:

Select All Reset Selection

1 H																	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba		72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra		104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og

57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Basis Set: 6-311G

Description: VTZ Valence Triple Zeta: 3 Funct.'s/Valence AO
Last Updated:
Latest Version: 0

More Information Citations

Download basis set

Format Gaussian94 Optimize General Contractions Get Basis set

Citation

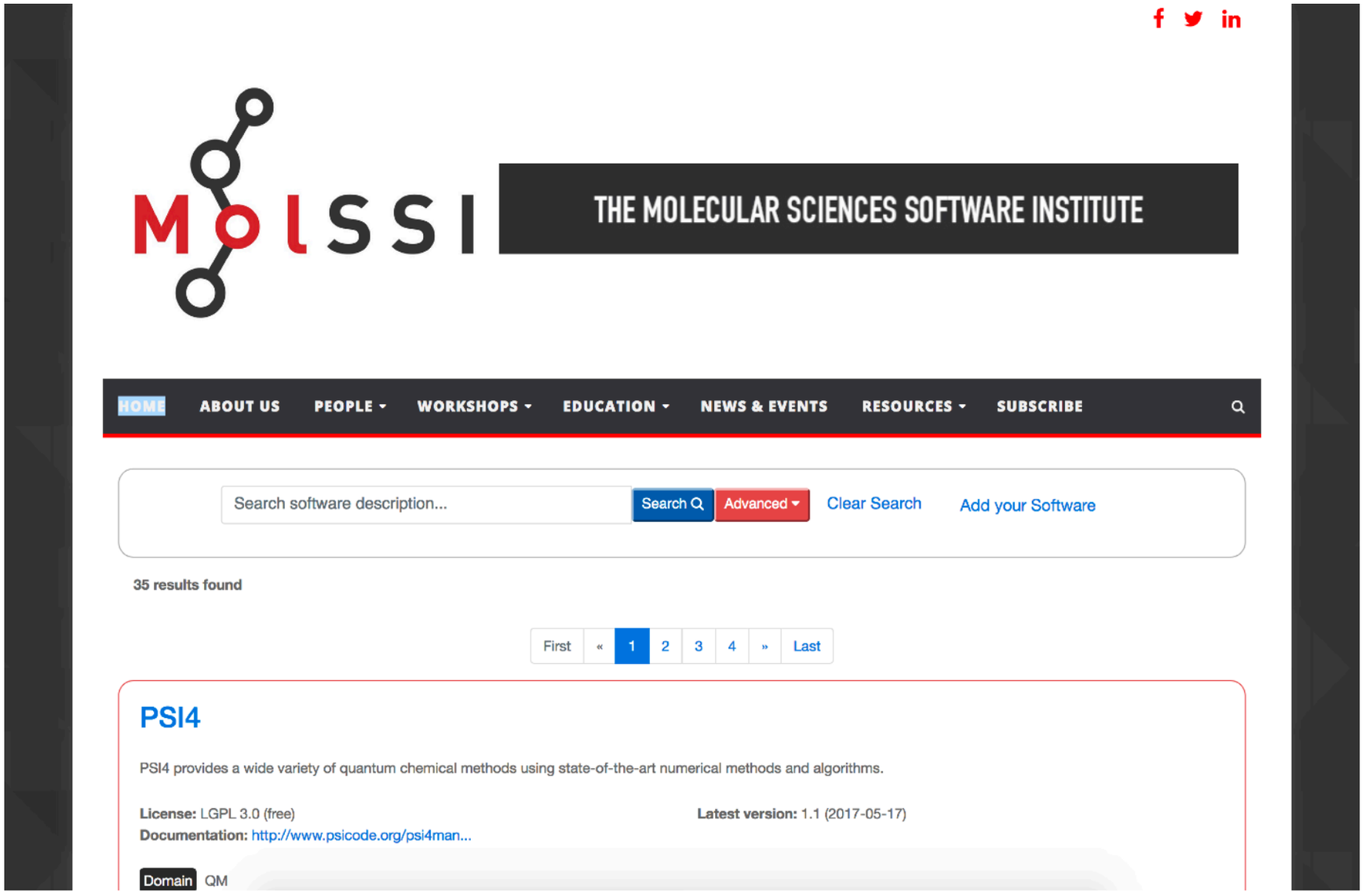
When publishing results obtained from use of the Basis Set Exchange (BSE) software and the EMSL Basis Set Library, please cite:

- *The Role of Databases in Support of Computational Chemistry Calculations*. Feller, D., J. Comp. Chem., 17(13), 1571-1586, 1996.
- *Basis Set Exchange: A Community Database for Computational Sciences*. Schuchardt, K.L., Didier, B.T., Elsethagen, T., Sun, L., Gurumoorthi, V., Chase, J., Li, J., and Windus, T.L. J. Chem. Inf. Model., 47(3), 1045-1052, 2007, doi:10.1021/ci600510j.

MolSSI

MolSSI Code Database

Convenient and up-to-date information on CMS community codes



MolSSI

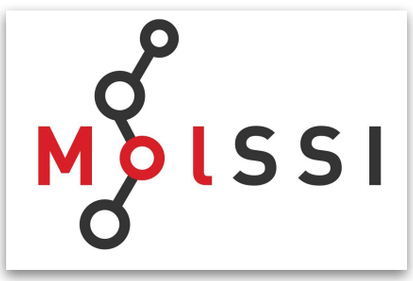
http://molssi.org/software-search/

Quantum Chemistry Schema

- MolSSI QM Schema – a JSON-based standard for common data to enable more complex workflows among quantum chemistry codes
- Just released v1

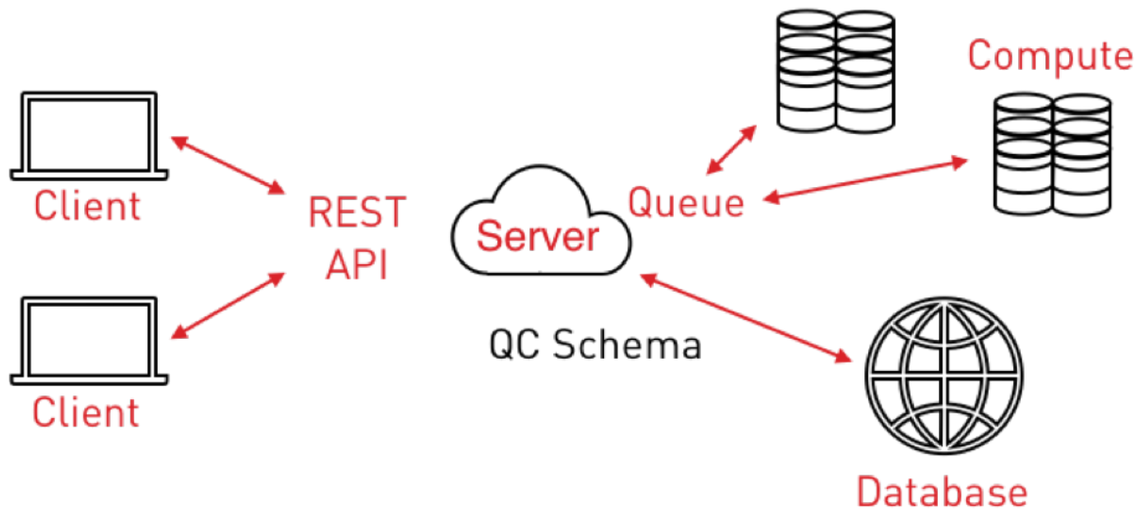
https://github.com/MolSSI/QC_JSON_Schema/

http://molssi-qc-schema.readthedocs.io/en/latest/index.html



MolSSI QC Database

Goal: Provide an open, community-wide quantum chemistry database to 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.



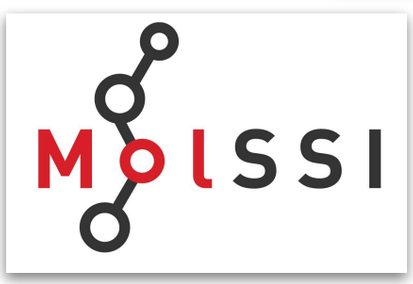
MolSSI QC Database

Features:

- General hybrid compute and data manipulation tools
- Deployability at scale by MolSSI or locally by research groups
- Interoperates with any QM program who adheres to the schema
- Distributed computing technology baked in
- Intuitive data organization layers
- Built on a completely open-source software stack







MolSSI QC Database

Force fields:

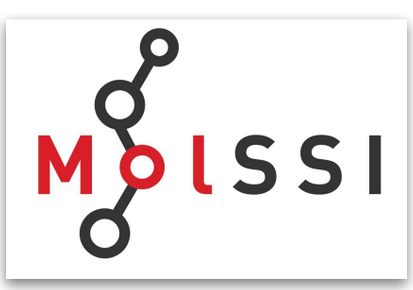
- Democratizes the enormous computational burden of high-level quantum chemical computations required to construct advanced forcefields to many stakeholders and beneficiaries

Supply reference computations:

- Provide uniform access to both the current and future quantum chemistry reference datasets in addition to standard sets of more approximate methods

Satisfy the data needs of machine learning:

- Central database that holds all computational results of other projects to assist chemistry in harnessing the data revolution.



Other MolSSI Software Infrastructure Projects

- MolSSI Framework – a light-weight Python-based plugin structure for interoperability of CMS codes for new scientific calculations;
- MolSSI QM/MM Driver – an API and communication layer for a control code using QM and MM codes as clients for QM/MM and other similar calculations;
- MolSSI Energy Expression Exchange – to allow translations of forcefields between molecular dynamics codes;

Acknowledgments

- Cecilia Clementi, Robert Harrison, Teresa Head-Gordon, Shantenu Jha, Anna Krylov, Vijay Pande, Theresa Windus;
- The dozens of members of the CMS community who helped to develop the vision for the Institute over the last five years;
- NSF ACI-1547580.

Watch molssi.org for the latest information!

MOLSSI