# Integration of TCOD (Theoretical Crystallography Open Database) and AiiDA (Automated Interactive Infrastructure and Database for Atomistic simulations)

Andrius Merkys[a,b], Giovanni Pizzi[a], Andrea Cepellotti[a], Nicolas Mounet[a], Saulius Gražulis[b] and Nicola Marzari[a]

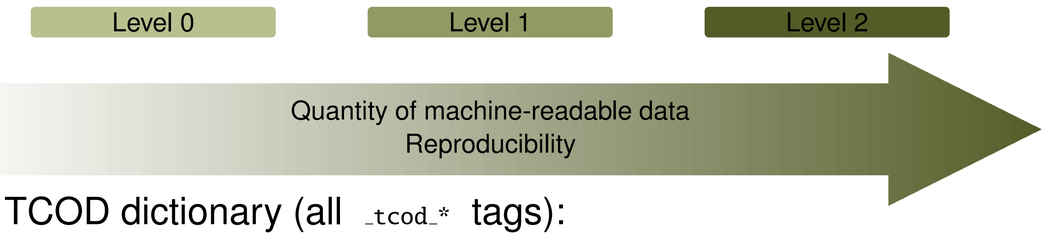
[a]Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; [b]Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, V. A. Graičiūno 8, LT-02241 Vilnius, Lithuania

## Abstract

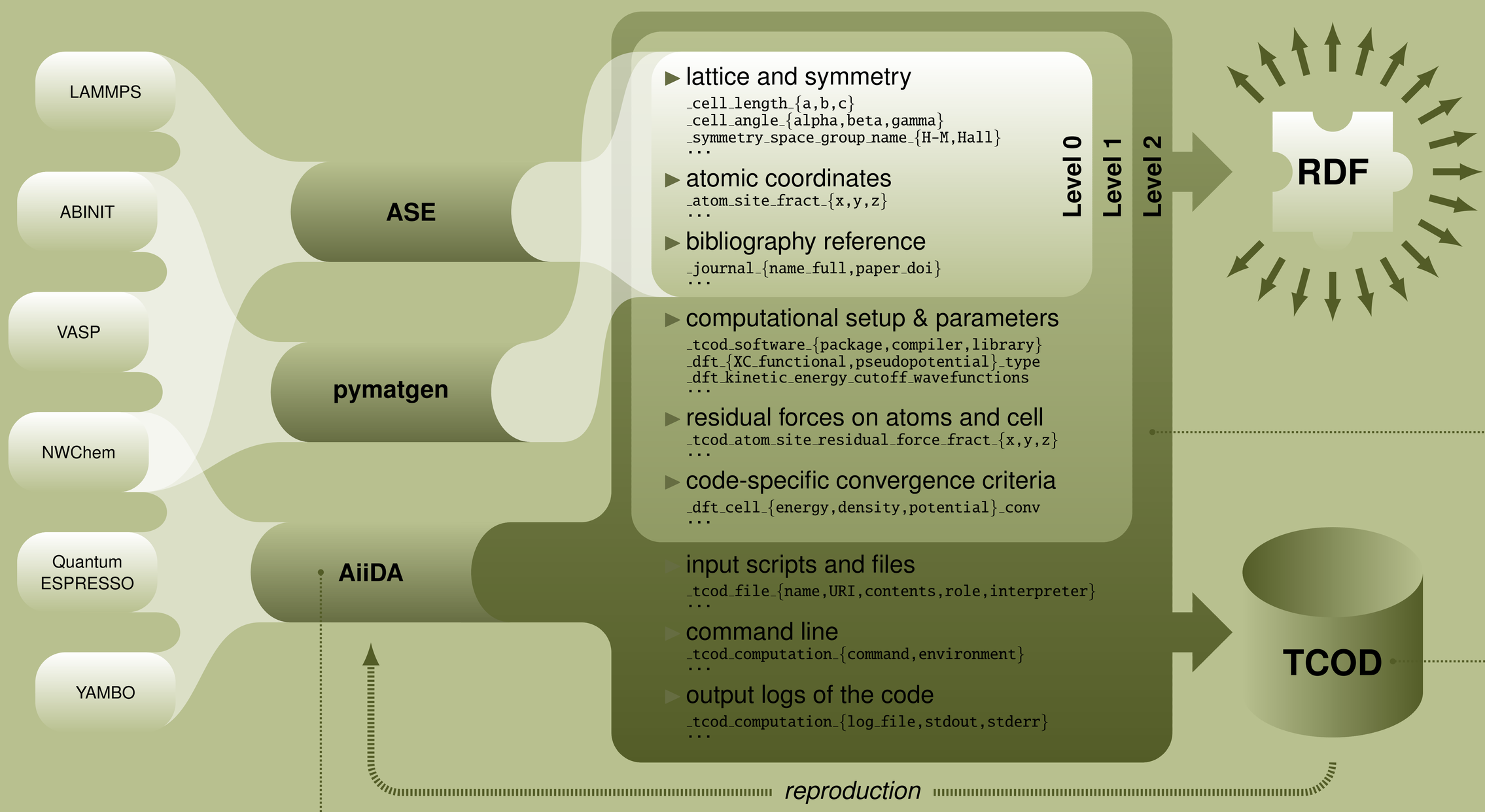
The Theoretical Crystallography Open Database (TCOD) has been launched in order to collect the results of calculations performed by many groups using various modern theoretical approaches (DFT, post-HF, QM/MM, etc.), into an open-access resource. TCOD, based on the architecture of the Crystallography Open Database (COD) [1], has adopted the best practice of using the CIF format [2], defining data validation criteria for automated checks and devising approach-specific dictionaries to homogenize the data in the fields of theoretical crystallography, which is expanding quickly due to unprecedented developments of electronic structure methods. Furthermore, TCOD puts an emphasis on the provenance and reproducibility of the results by devising a specific dictionary for the related metadata. We have employed TCOD CIF dictionaries as an interface for the integration of TCOD with AiiDA framework [3] (`http://www.aiida.net`), a materials' informatics infrastructure that provides a high-level research environment to automate the execution of computations, automatically store inputs and outputs in a tailored graph database (with particular care in keeping track of the full provenance of data) and share the results. Such an integration of automation framework and database storage allows for the deposition of simulation results with automatically recorded metadata, guaranteeing the reproducibility of each calculation as well as the full data provenance for each item in the database. Moreover, data from the TCOD database can be easily retrieved and imported back into AiiDA as an input for further calculations and analysis.

## CIF dictionaries

- Offer ontologies for data description;
- Aim at automated checks for convergence, computational quality and reproducibility;
- Enable automated deposition and data mining;
- Can be transformed to Resource Description Framework (RDF) schemas.

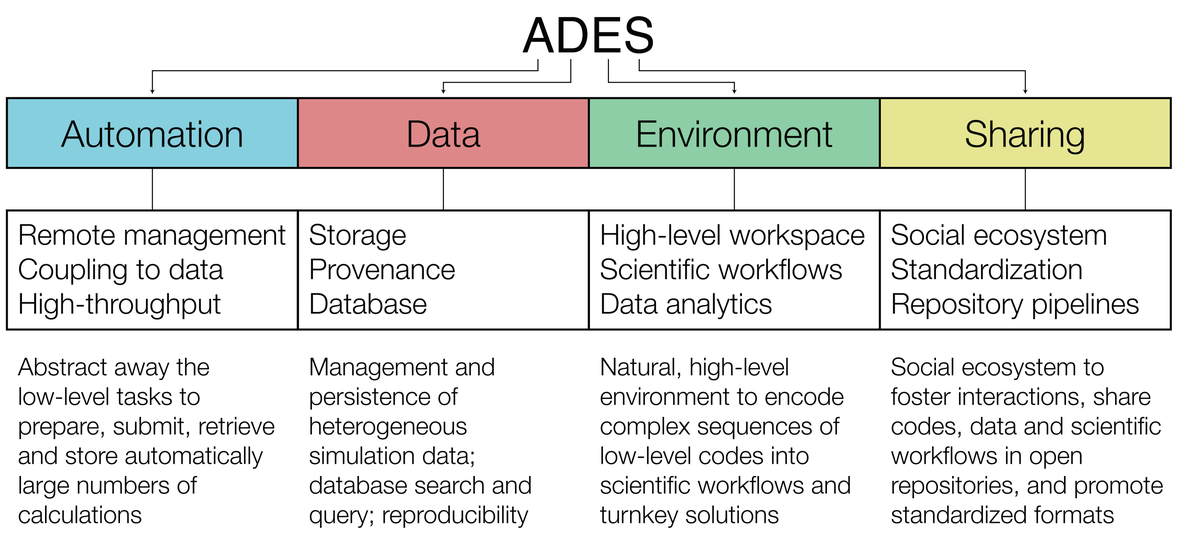



- Main TCOD dictionary (all `_tcod_*` tags):
  - `http://www.crystallography.net/tcod/cif/dictionaries/cif_tcod.dic`
  - `svn://www.crystallography.net/tcod/cif/dictionaries/cif_tcod.dic`
- DFT dictionary (all `_dft_*` tags):
  - `http://www.crystallography.net/tcod/cif/dictionaries/cif_dft.dic`
  - `svn://www.crystallography.net/tcod/cif/dictionaries/cif_dft.dic`
- Open mailing list for discussions:
  - `http://lists.crystallography.net/cgi-bin/mailman/listinfo/tcod`




## AiiDA

- AiiDA – Automated Interactive Infrastructure and Database for Atomistic simulations, `http://www.aiida.net`
- An engine for automation of computations and storage of full data provenance;
- Employs a high-level plugin interface;
- Support extendable to all command line interface-based codes;
- Four pillars of AiiDA infrastructure:




## TCOD

- Accessible at `http://www.crystallography.net/tcod/`
- An open-access resource of theoretical computation results;
- Based on the infrastructure of COD;
- Stores supplementary material of published research as well as prepublication and personal communication material;
- Aims to store the metadata for the full reproducibility of computation results.

**Theoretical Crystallography Open Database**

## Bibliography

[1] S. Gražulis et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*, 40(D1):D420–D427, Jan 2012.

[2] S. R. Hall et al. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A*, 47(6):655–685, Nov 1991.

[3] G. Pizzi et al. AiiDA: Automated Interactive Infrastructure and Database for Computational Science. arXiv:1504.01163.

This research is funded by the SCIEX Fellowship grant No. 13.169

Online version of the poster: `http://j.mp/1K110ro`



andrius.merkys@gmail.com