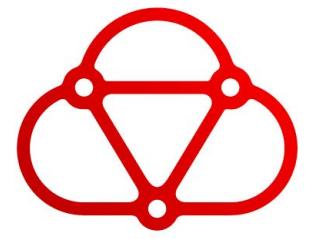




Open-research data repositories and reproducible workflows



MATERIALS CLOUD

Giovanni Pizzi

Paul Scherrer Institute (PSI) and NCCR MARVEL, Switzerland

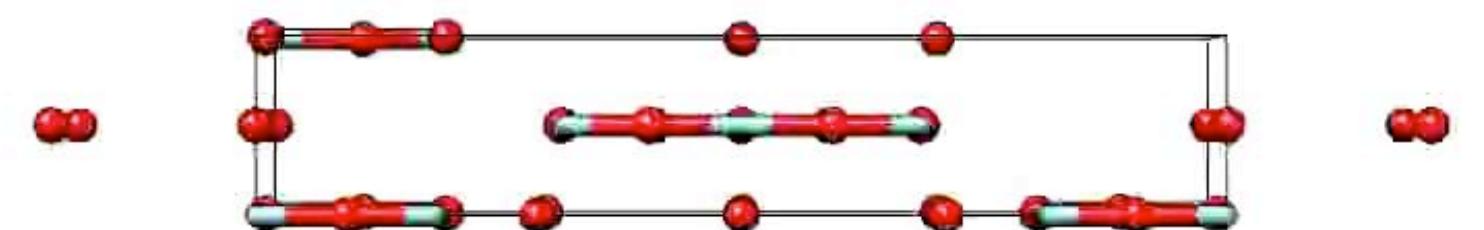
NCCR Network Ventures, Open Research Data events
Basel, 2 June 2022



Outline

- Managing **automated simulations with full reproducibility**
- **Data repositories** for open-research data
- Combining the two: **Open Science**
- Addressing robotic experiments, **towards autonomous labs**

Leverage supercomputers to compute and predict materials properties



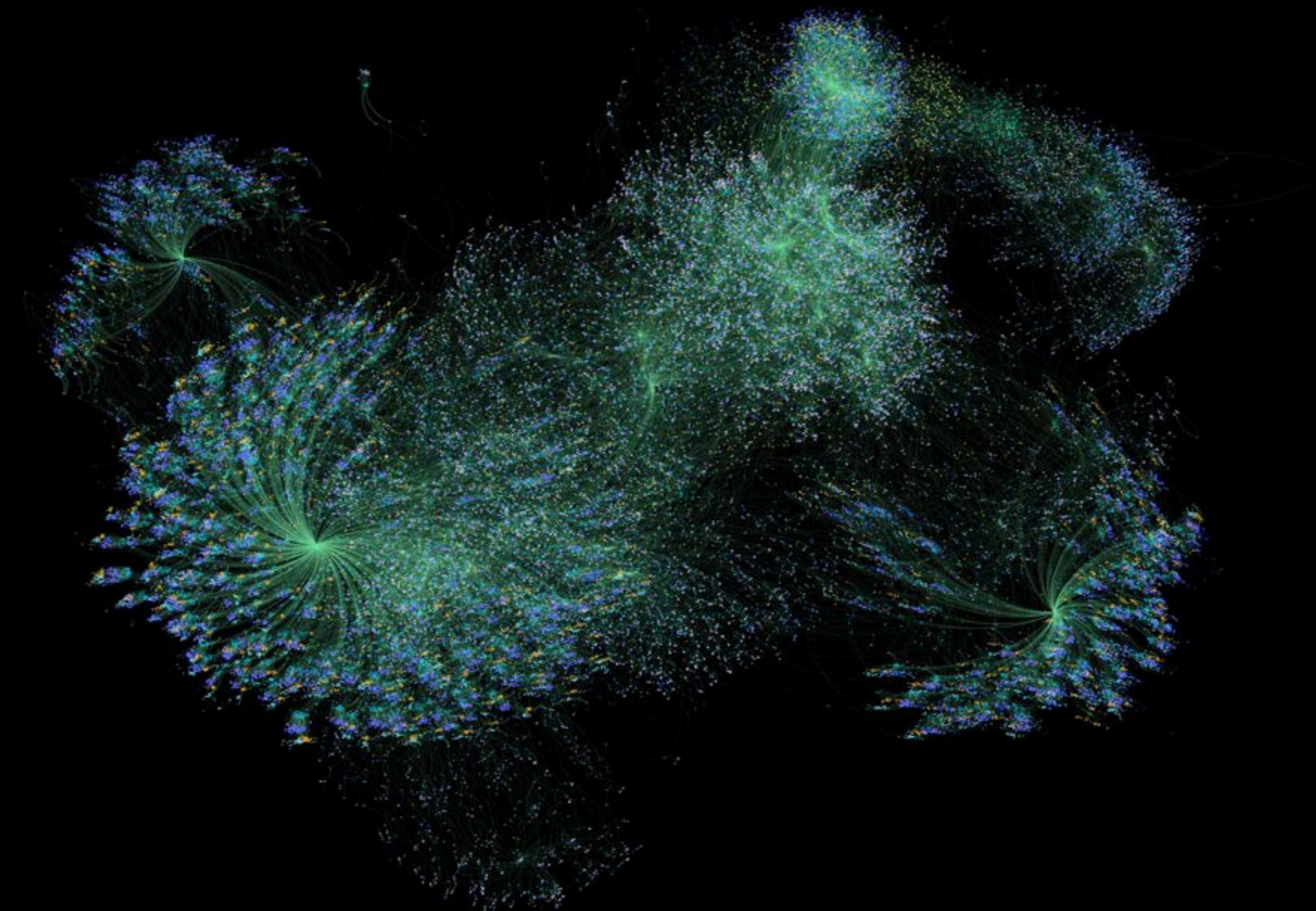
Aim: Compute properties for all of them
(and even new, invented ones)
and **discover novel functional materials**



Challenges in high-throughput HPC

- **Workflow automation**
 - Need tools to define complex workflows with advanced error handling
 - An automated, robust and scalable engine to run the workflows
- **Data management**
 - Data should be stored reliably and efficiently
 - Stored data should be interoperable and queryable
- **Reproducibility**
 - All produced data should be reproducible by storing the full provenance

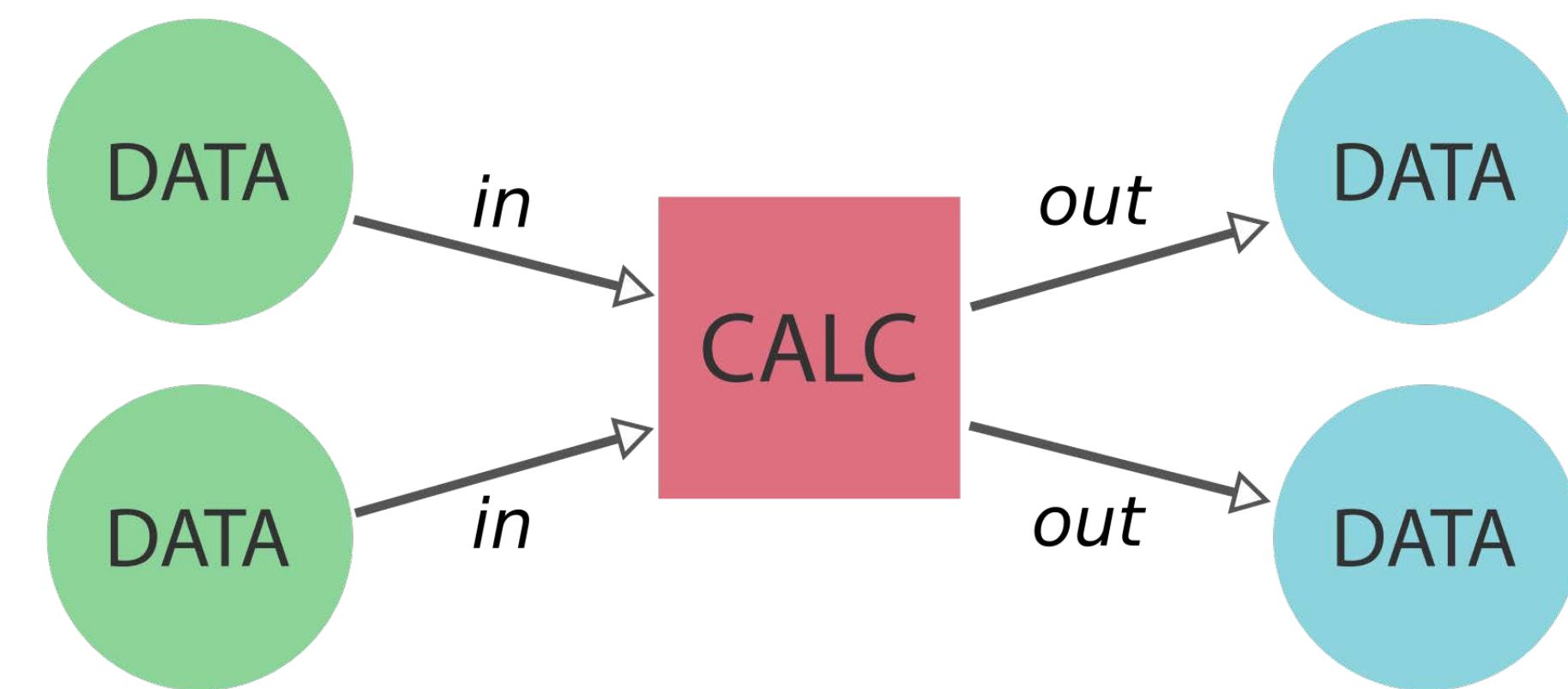
Reproducible simulations and interoperable workflows



Data provenance

Simple recipe

- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**



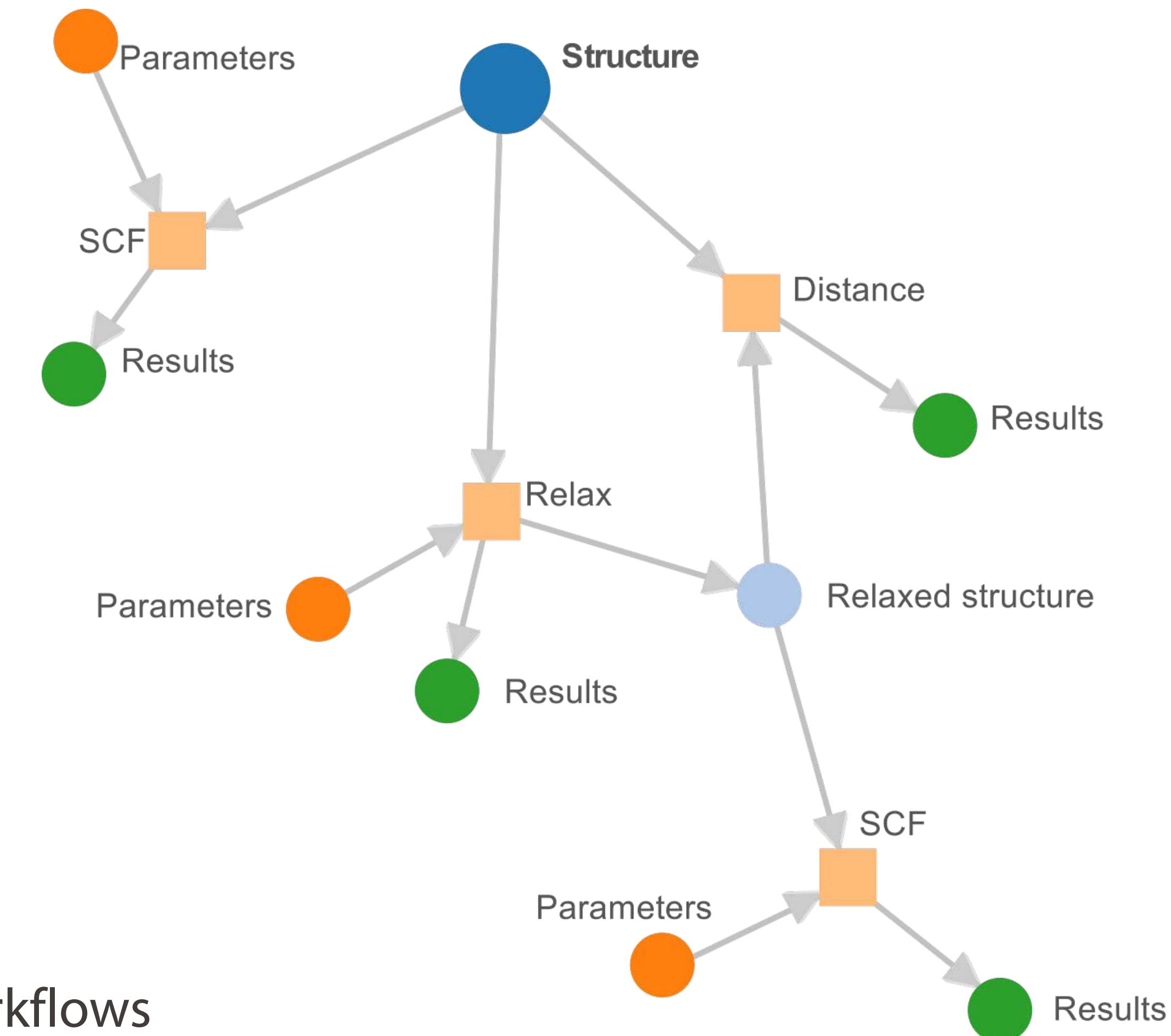
Data provenance

Simple recipe

- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**

Provenance graphs

- When data gets reused, a directed graph is created
- That quickly grow in complexity even for "simple" workflows



Data provenance

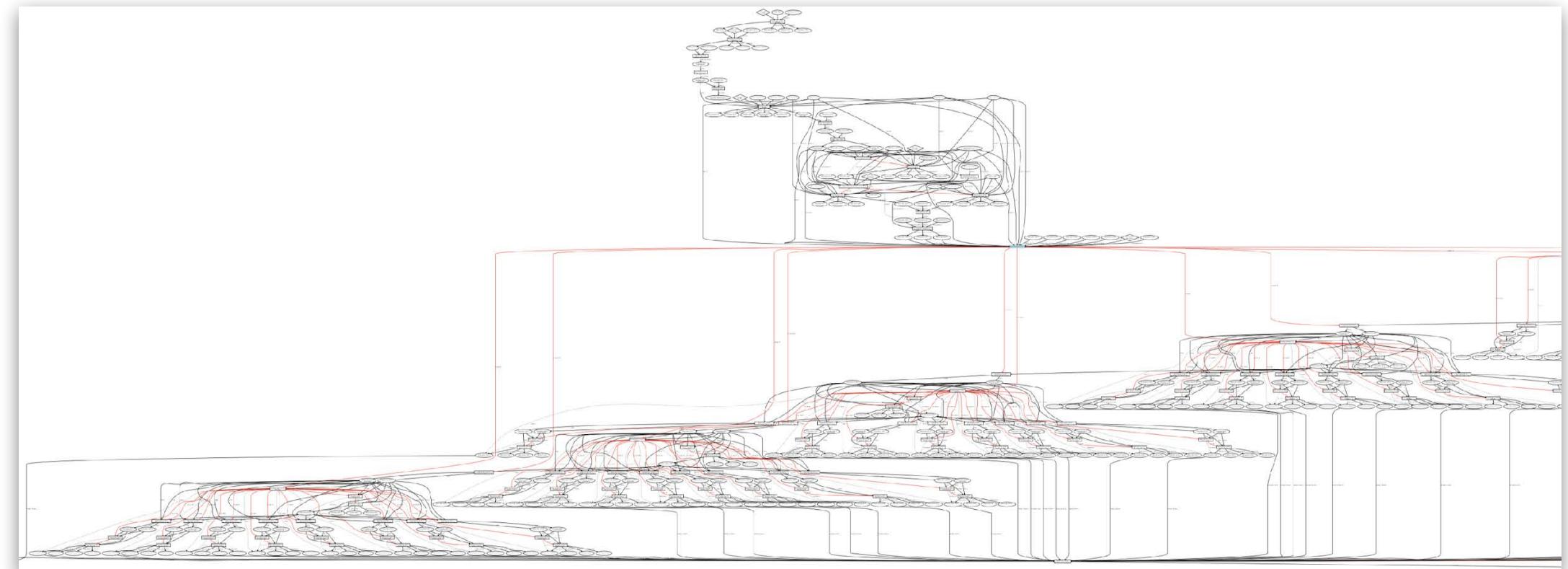
Simple recipe

- Store data transformations or '**calculations**'
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**

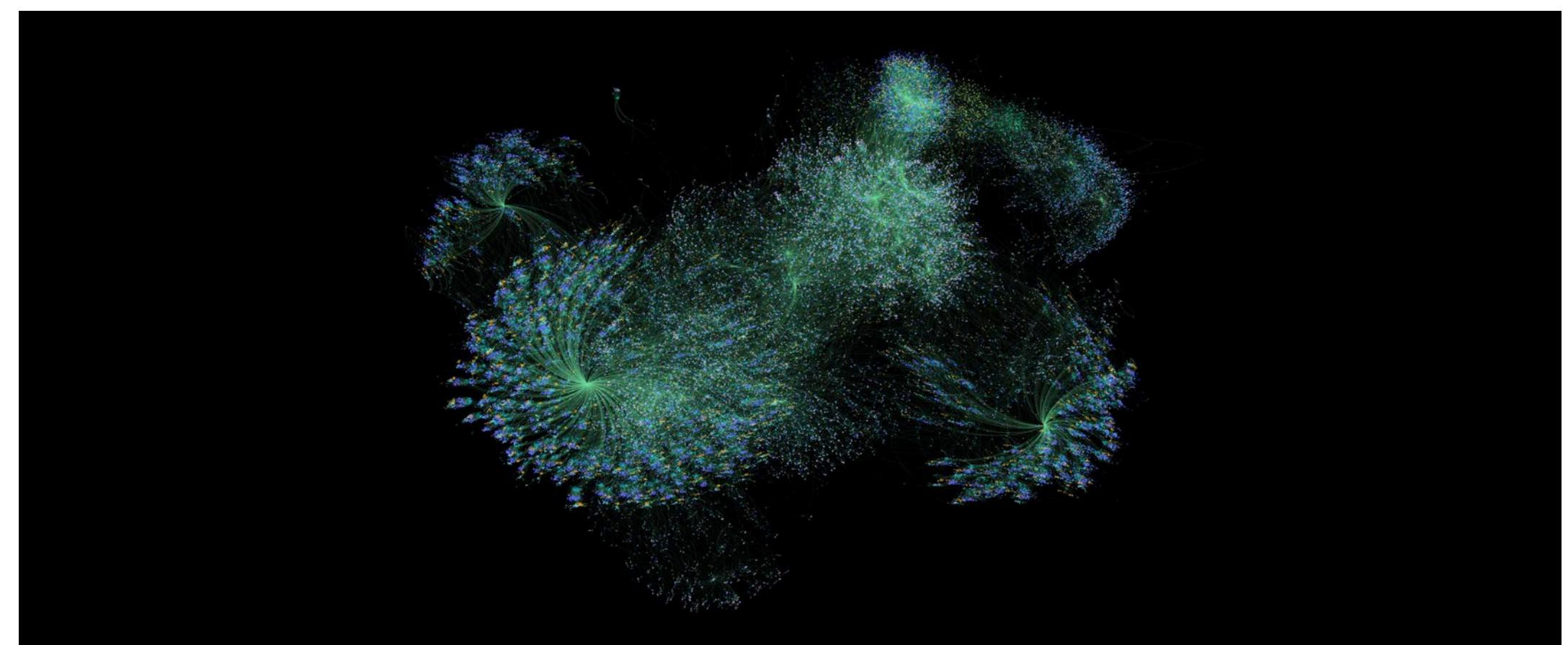
Graph requirements

- Needs to be automated
- Needs to be stored *as data is created*

Complexity grows quickly even for simple workflows and is impossible to reconstruct *a posteriori*



Molecular dynamics study of Lithium in a solid electrolyte



Graphical representation of actual database

- COMPUTATIONAL SCIENCE INFRASTRUCTURE
- FOR HIGH THROUGHPUT WORKFLOWS
- WITH FULL DATA PROVENANCE

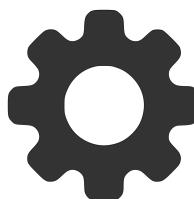


Language: implemented and API in python

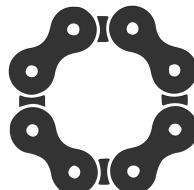


License: MIT open source <http://www.aiida.net/>

Source: <https://github.com/aiidateam/aiida-core>



Scalable workflow engine: *robustness*

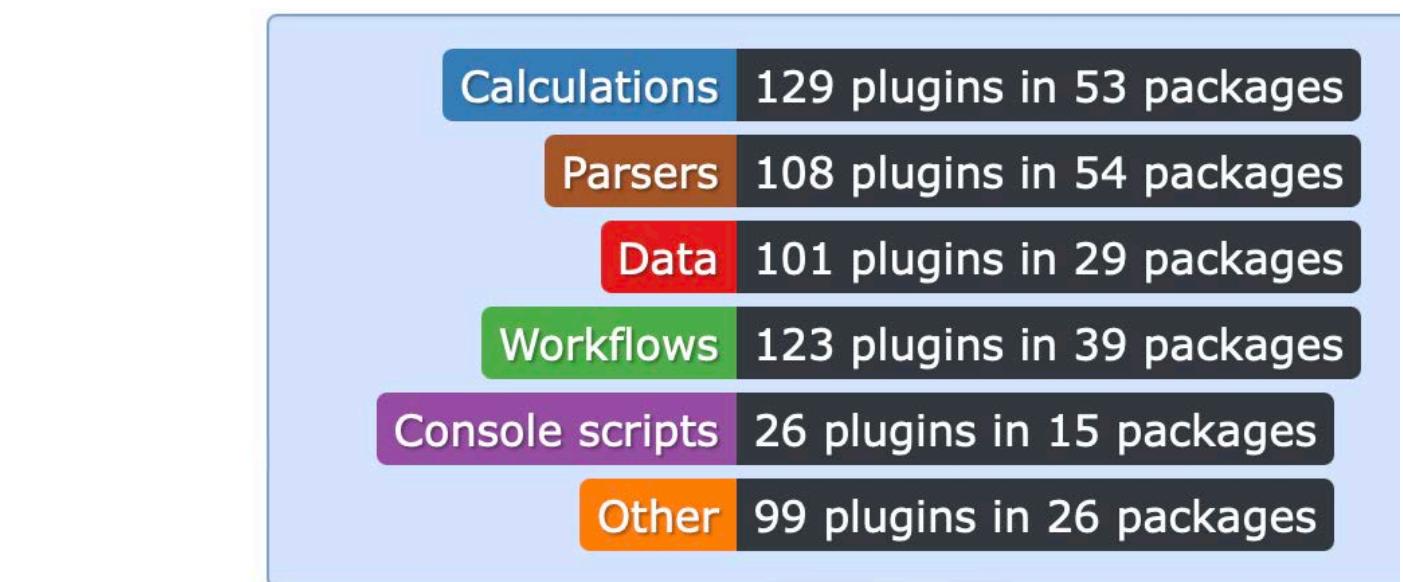


Built-in support for HPC: *performance*



[View on GitHub/register your package]

Registered plugin packages: 89



<https://aiidateam.github.io/aiida-registry/>



Automated full data provenance: *reproducibility*



G. Pizzi *et al.*, Comp. Mat. Sci. 111, 218-230 (2016)
S.P. Huber *et al.*, Scientific Data 7, 300 (2020)

Flexible plugin system: *interoperability*

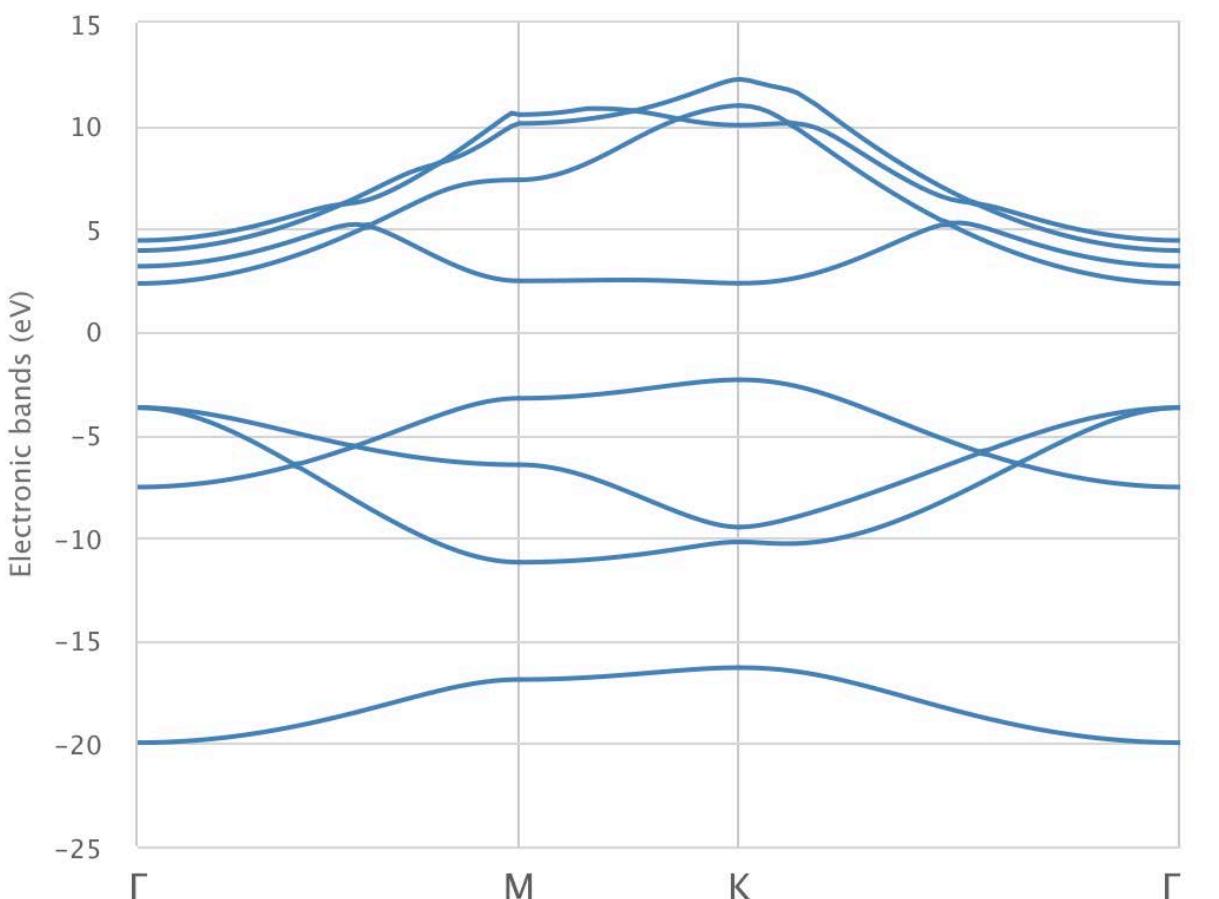
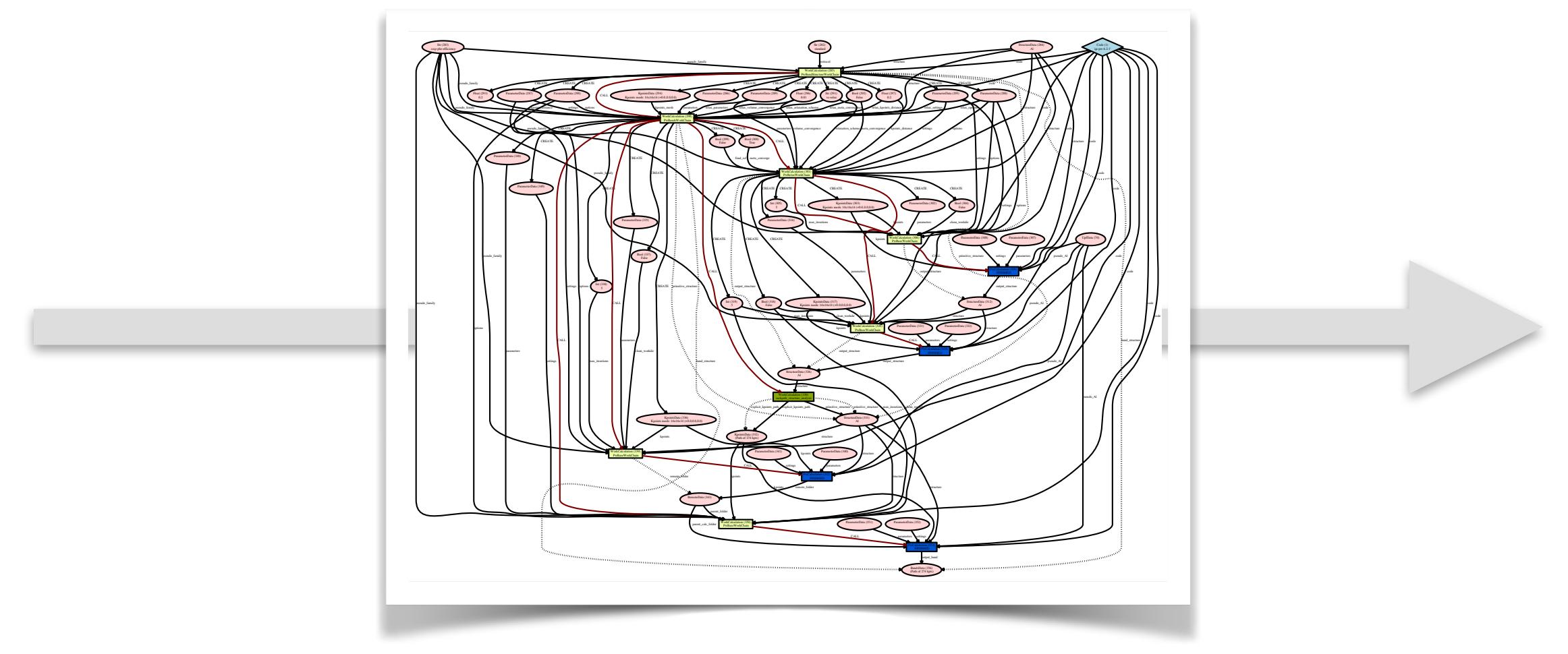
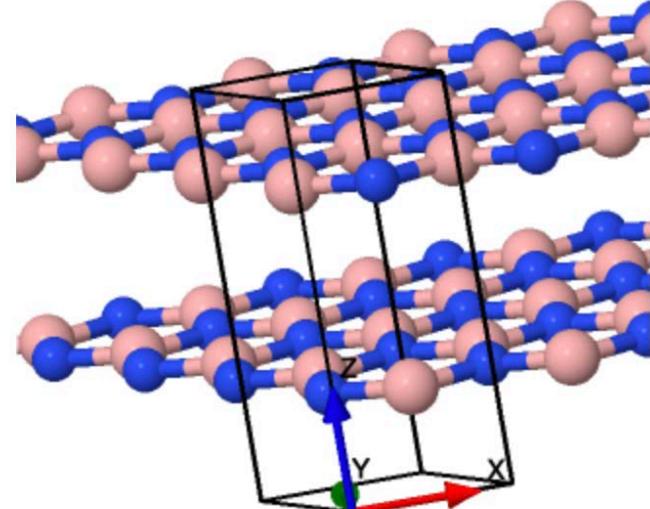


Running at (exa)scale



Workflows to generate data

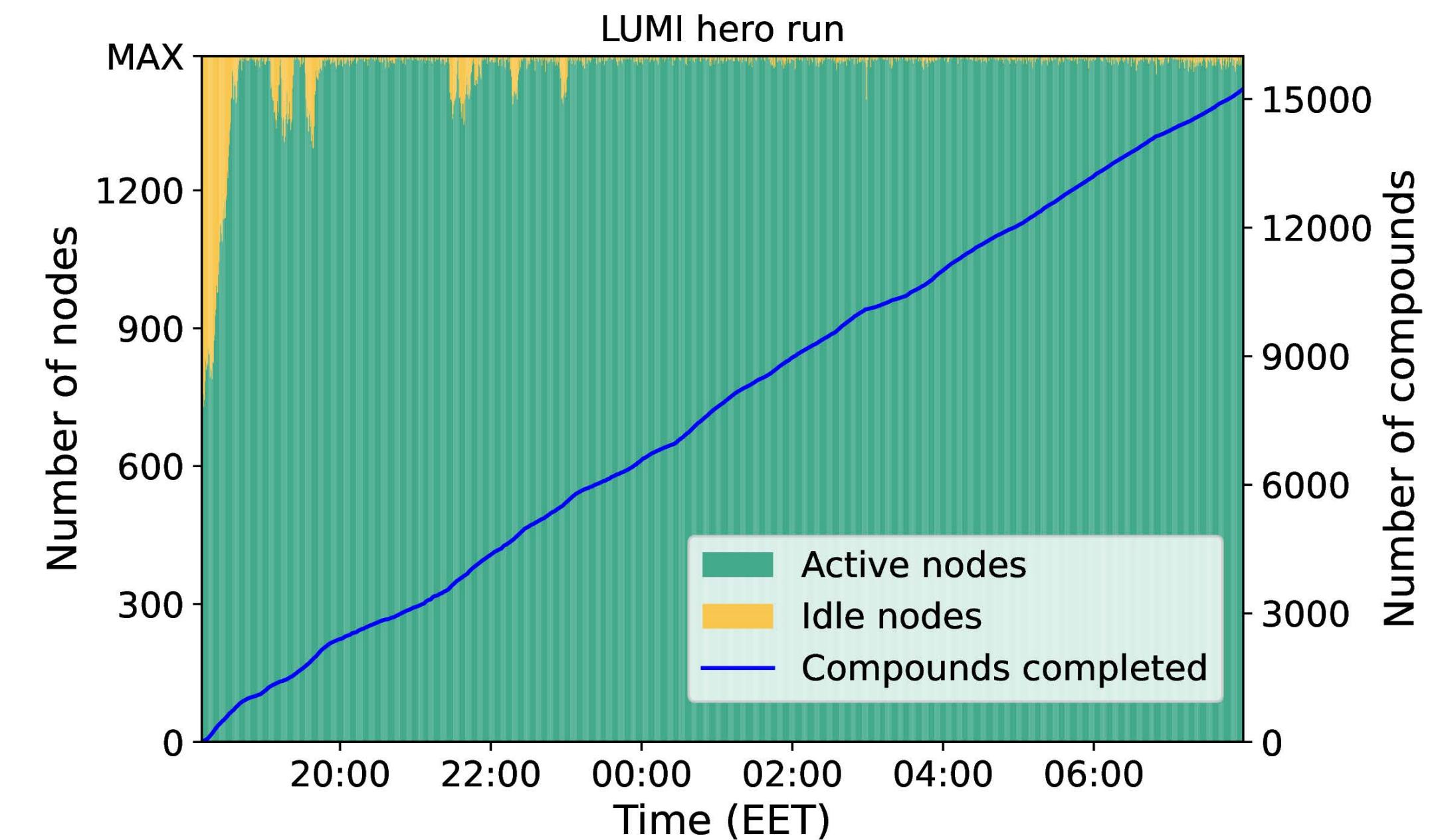
- Given a material, we compute advanced quantities: often non-trivial, result of a complex workflow



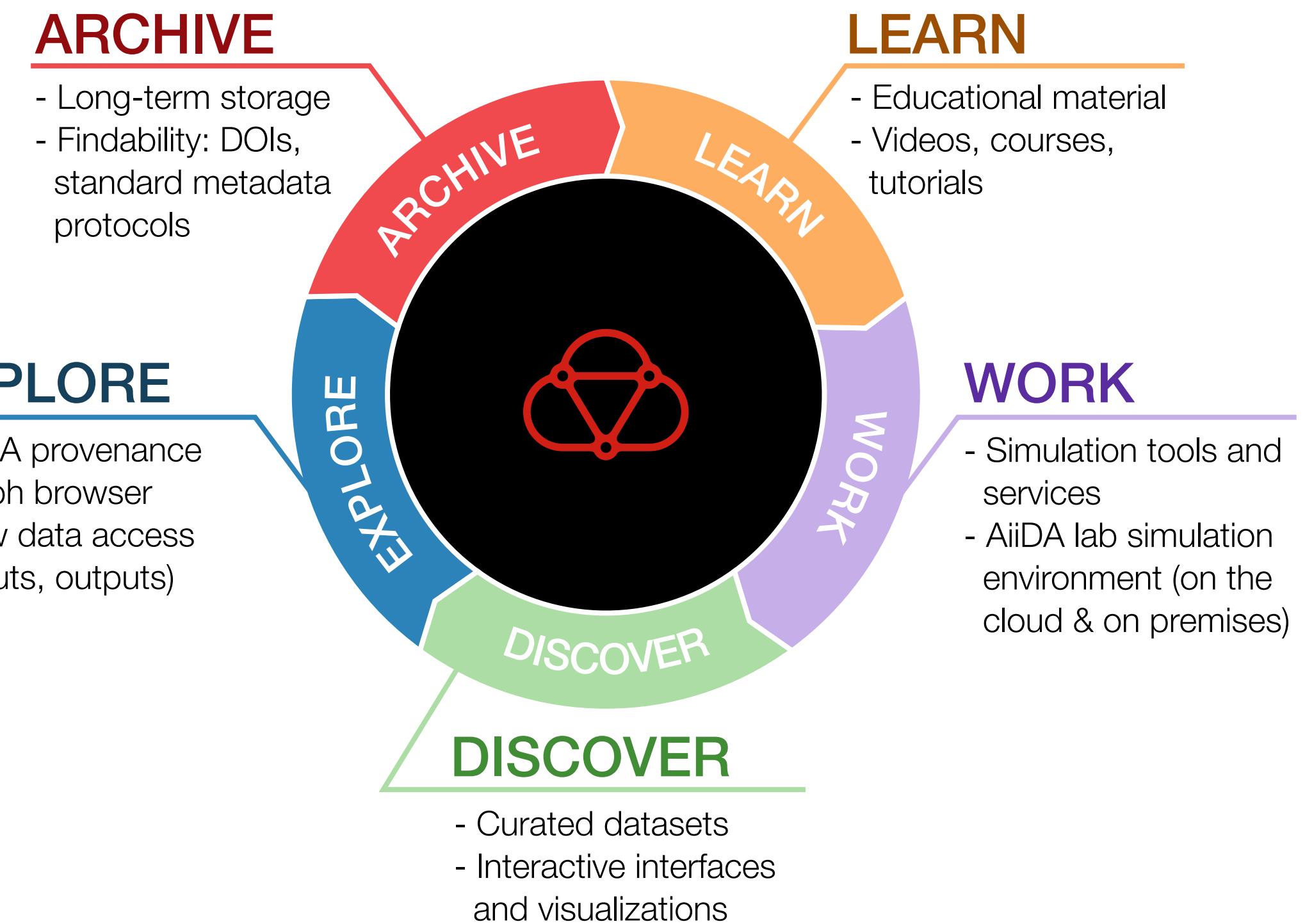
- Provenance graph: *log of “what happened in the past”*
 - reproduce that single specific workflow execution**
 - Workflow engine**: python interface to encode complex scientific steps, parameter choices, ...
 - Added bonus: automated provenance tracking and reproducibility

Toward (pre-)exascale applications, while ensuring reproducibility

- Workflow engine optimised for high-throughput performance, **ready to scale on (pre-)exascale machines**
- Demonstration on 19 Nov 2021: **"hero run" on LUMI-C** (Finland)
- Record utilization of the full machine (196608 AMD Epyc cores)
- AiiDA orchestrating 55,704 DFT calculations, optimising geometry of 15,324 inorganic compounds
- **All fully reproducible and with automatic provenance tracking**



Data repositories



<https://www.materialscloud.org>

Data repositories: Materials Cloud Archive

The screenshot shows the Materials Cloud Archive interface. At the top, there are tabs: LEARN, WORK, DISCOVER, EXPLORE, ARCHIVE (which is highlighted in red), and More. Below the tabs are search and login fields. A banner at the top right features logos for SCIENTIFIC DATA, Google, re3data, B2FIND, EUDAT, GOFAR, and FAIRsharing.org. The main content area displays three research entries:

- Semi-local and hybrid functional DFT data for thermalised snapshots of polymorphs of benzene, succinic acid, and glycine**
DOI: 10.24435/materialscloud:vp-jt
Edgar A. Engel, Venkat Kapil
Structure prediction for molecular crystals is a longstanding challenge, as often minuscule free energy differences between polymorphs are sensitively affected by the description of electronic structure, the statistical mechanics of the nuclei and the cell, and thermal expansion. The importance of these effects has been individually established, but rigorous free energy calculations, which simultaneously account for all terms, have not been computationally viable. Here we reproduce the experimental stabilities of polymorphs of prototypical compounds -- benzene, glycine, and succinic acid -- by computing rigorous first-principles Gibbs free energies, at a fraction of the cost of conventional methods ...
Latest version: v1
Publication date: Mar 26, 2021
- Simulating solvation and acidity in complex mixtures with first-principles accuracy: the case of CH₃SO₃H and H₂O₂ in phenol**
DOI: 10.24435/materialscloud:2x-7x
Kevin Rossi, Veronika Juraskova, Raphael Wischert, Laurent Garel, Clemence Corminboeuf, Michele Ceriotti
Set of inputs to perform the calculations reported in the paper. The i-pi input enables to perform molecular dynamics / metadynamics / REMD / PIMD simulations, with adequate thermostats. The DFTB and LAMMPS input respectively enable to calculate force and energies within the DFTB and Neural Network Forcefield frameworks. The CP2K input files enable to calculate force and energies at PBE and PBE0 level. The latter is used as the reference to train the neural network correction on top of DFTB. Brief description of the work: We present a generally-applicable computational framework for the efficient and accurate characterization of molecular structural patterns and acid properties in explicit solvent using H₂O₂ and CH₃SO₃H in phenol as an example ...
Latest version: v2
Publication date: Mar 26, 2021
- Detecting electron-phonon coupling during photoinduced phase transition**
DOI: 10.24435/materialscloud:c0-q1
Takeshi Suzuki, Yasushi Shinohara, Yangfan Lu, Mari Watanabe, Jiadi Xu, Kenichi L. Ishikawa, Hide Takagi, Minoru Nohara, Naoyuki Katayama, Hiroshi Sawa, Masami Fujisawa, Teruto Kanai, Jiro Itatani, Takashi Mizokawa, Shik Shin, Kozo Okazaki

- Provides DOI for datasets (files with metadata: authors, affiliations, abstract) and guarantees 10 year persistence
- **Focused on one discipline** (materials science)
- It is a **moderated** archive: entries are checked, moderators suggest improvements (*not peer review: rather check of quality, openness, fitting into materials-science domain*)

<https://archive.materialscloud.org>

Interface for users submitting data

Submission form

The screenshot shows the 'Materials Cloud Archive record upload' submission form. At the top, there's a navigation bar with 'LEARN', 'WORK', 'DISCOVER', 'EXPLORE', 'ARCHIVE', and a 'More' dropdown. Below it is a search bar and a user profile for 'marco.borelli@epfl.ch'. The main area starts with a 'Submission procedure' section containing a list of steps: 1. Before submitting your record please read the information on the Materials Cloud Archive policies and submission instructions available [here](#). 2. Click "Save record" to create a draft of your record. 3. Click "Submit record for publication" to submit your record and notify the Archive moderators. 4. Records are reviewed within 7 days from submission. A warning at the bottom states: 'WARNING: you can upload a maximum total files size of 25 GB using this form. Please [contact us](#) if you need to upload more than 25 GB.' Below this are fields for 'Title*' (with a note about capital letters), 'Description*', and 'Keywords*'. A sidebar on the right lists 'MarketPlace' partners: MAR, Marie Curie Fellowship, MARVEL, MARVEL/DD1, MARVEL/DD2, MARVEL/DD3, and MARVEL/DD4. A blue 'Add keyword' button is located at the bottom right of the keyword input field.

User view (own entries)

The screenshot shows the 'My records' section of the interface. At the top, there's a navigation bar with 'LEARN', 'WORK', 'DISCOVER', 'EXPLORE', 'ARCHIVE', and a 'More' dropdown. Below it is a search bar and a user profile for 'valeria.granata@epfl.ch'. The main area is titled 'My records' and contains sections for 'Records in progress' and 'Records published (latest 10 published records)'. The 'Records published' section shows two entries:

Record	Owner	Status
Towards constant potential modeling of CO-CO coupling at liquid water-Cu(100) interfaces DOI: 10.24435/materialscloud:p9-q7	hhk@chem.ku.dk	PUBLISHED
Simulating the ghost: quantum dynamics of the solvated electron DOI: 10.24435/materialscloud:dz-a0	jinggang.lan@uzh.ch	PUBLISHED

Each entry includes a DOI link, version information (v1), last update date, and authors.

Open data sharing: Archive, Discover, Explore

DOIs
assigned

materialscloud:2017.0008/v3

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mounet^{1*}, Marco Gibertini¹, Philippe Schwaller¹, Davide Campi¹, Andrius Merkys^{1,2}, Antimo Marrazzo¹, Thibault Sohier¹, Ivano E. Castelli¹, Andrea Cepellotti¹, Giovanni Pizzi¹, Nicola Marzari^{1*}

¹ 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

² Vilnius University Institute of Biotechnology, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

DOI: 10.24435/materialscloud:2017.0008/v3 [version v3]

Publication date: Apr 03, 2019

How to cite this record

Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Davide Campi, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari, *Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds*, Materials Cloud Archive 2017.0008/v3 (2019), doi: 10.24435/materialscloud:2017.0008/v3.

Description

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozens of 2D materials have been successfully synthesized or exfoliated. Here, we search for novel 2D materials that can be easily exfoliated from their parent compounds. Starting from 108423 unique, experimentally known three-dimensional compounds we identify a subset of 5619 that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van-der-Waals density-functional theory, validated against experimental structural data and calculated random-phase-approximation binding energies, allow to identify 1825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search from for optimal properties. For a subset of 258 compounds we explore vibrational, electronic, magnetic, and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors. This archive entry contains the database of 2D materials (structural parameters, band structures, binding energies, phonons for the subset of the 258 easily exfoliable materials with less than 6 atoms, structures and binding energies for the remaining 1567 materials) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

- Select 2d materials via interactive periodic table and view their properties (with links to provenance)
- Explore interface providing access to the full database

Files

File name	Size
2D_materials.tar.gz MDS	113.0 MiB

Description

We provide 258 two-dimensional crystal structures (lattice vectors, atomic species and positions), exfoliated from three-dimensional experimental crystal structures. The structures were relaxed at the DFT-PBE level. Together with each structure, a set of materials properties is also given (at the DFT-PBE level): chemical formula, spacegroup, structural prototype, magnetic state, magnetization, band-gap, electronic bands, and phonon



Export

Dublin Core JSON

</div

Recommended by Scientific Data and Open Research Europe

SCIENTIFIC DATA

Marine Geosciences Data System	view re3data entry
UNAVCO, Inc.	view re3data entry
Incorporated Research Institutions for Seismology (IRIS)	view re3data entry
OpenTopography	view FAIRsharing entry

Physics ↑

HEPData	view re3data entry
---------	------------------------------------

Materials science ↑

NoMaD Repository	view FAIRsharing entry
Materials Cloud	view FAIRsharing entry

Social sciences ↑

Archaeology Data Service	view re3data entry
Harvard Dataverse	view re3data entry
ICPSR	view re3data entry
Open Science Framework	view FAIRsharing entry
Qualitative Data Repository	view FAIRsharing entry
UK Data Service	view re3data entry

Generalist repositories ↑

<https://www.nature.com/sdata/policies/repositories#materials>



Research and Innovation

Open Research Europe

High Energy Physics	HEPData	Title, DOI
---------------------	---------	------------

Materials Science

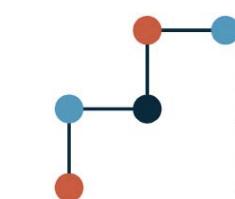
DATA TYPE	WHERE TO SUBMIT	WHAT TO INCLUDE IN THE DATA AVAILABILITY SECTION OF YOUR ARTICLE
<i>Ab initio</i> electronic structures	NOMAD Repository	Title, DOI
Computational, but especially calculations with full provenance	Materials Cloud	Title, DOI

2.3 Add a Data Availability Statement to Your Article

All articles must include a Data Availability statement, even where there is no data associated with the article. This statement should be added to the end of the article *prior to submission*. The Data Availability statement should not refer readers or reviewers to contact an author to obtain the data, but should instead include the applicable details listed below.

No associated or additional data

[https://open-research-europe.ec.europa.eu/
for-authors/data-guidelines](https://open-research-europe.ec.europa.eu/for-authors/data-guidelines)



Swiss National
Science Foundation



What's new Research Funding AI

Harvard Dataverse	Discipline-specific (Social sciences)	http://service.re3data.org/repository/r3d10001051
Image Data Resource (IDR)	Discipline-specific (Life sciences [imaging])	https://www.re3data.org/repository/r3d100012435
Materials Cloud Archive	Discipline-specific (Natural sciences [material sciences])	https://www.re3data.org/repository/r3d100012611
PRoteomics IDENTifications Database (PRIDE)	Discipline-specific (Life sciences [proteomics])	https://www.re3data.org/repository/r3d100010137
RCSB Protein Data Bank	Discipline-specific (Life sciences [protein])	https://www.re3data.org/repository/r3d100010327

[https://www.snf.ch/en/WtezJ6qxuTRnSYgF/
topic/open-research-data-which-data-repositories-can-be-used](https://www.snf.ch/en/WtezJ6qxuTRnSYgF/topic/open-research-data-which-data-repositories-can-be-used)

Recommended by Scientific Data and Open Research Europe

The screenshot shows the homepage of the Scientific Data website. At the top, it says "SCIENTIFIC DATA" with a binary logo. Below that is a search bar. Under "Data types", there are sections for "Marine Geosciences Data System", "UNAVCO, Inc.", "Incorporated Research Institutions for Seismology (IRIS)", and "OpenTopography". Under "Disciplines", there are sections for "Physics" (with HEPData), "Materials science" (with NoMaD Repository and Materials Cloud), "Social sciences" (with Archaeology Data Service, Harvard Dataverse, ICPSR, Open Science Framework, Qualitative Data Repository, and UK Data Service), and "Generalist repositories". Each section has a "view re3data entry" link.

<https://www.nature.com/sdata/policies/repositories#materials>

The screenshot shows the "for-authors/data-guidelines" page of the Open Research Europe website. It features a large central image with the text "These links are a great resource to discover if there are discipline-specific repositories for you!". Below this, there is a table with rows for different disciplines:

Category	Repository	Description	Link
Harvard Dataverse	Discipline-specific (Social sciences)		http://service.re3data.org/repository/r3d10001051
Image Data Resource (IDR)	Discipline-specific (Life sciences [imaging])		https://www.re3data.org/repository/r3d100012435
Materials Cloud Archive	Discipline-specific (Natural sciences [material sciences])		https://www.re3data.org/repository/r3d100012611
PRoteomics IDENTifications Database (PRIDE)	Discipline-specific (Life sciences [proteomics])		https://www.re3data.org/repository/r3d100010137
RCSB Protein Data Bank	Discipline-specific (Life sciences [protein])		https://www.re3data.org/repository/r3d100010327

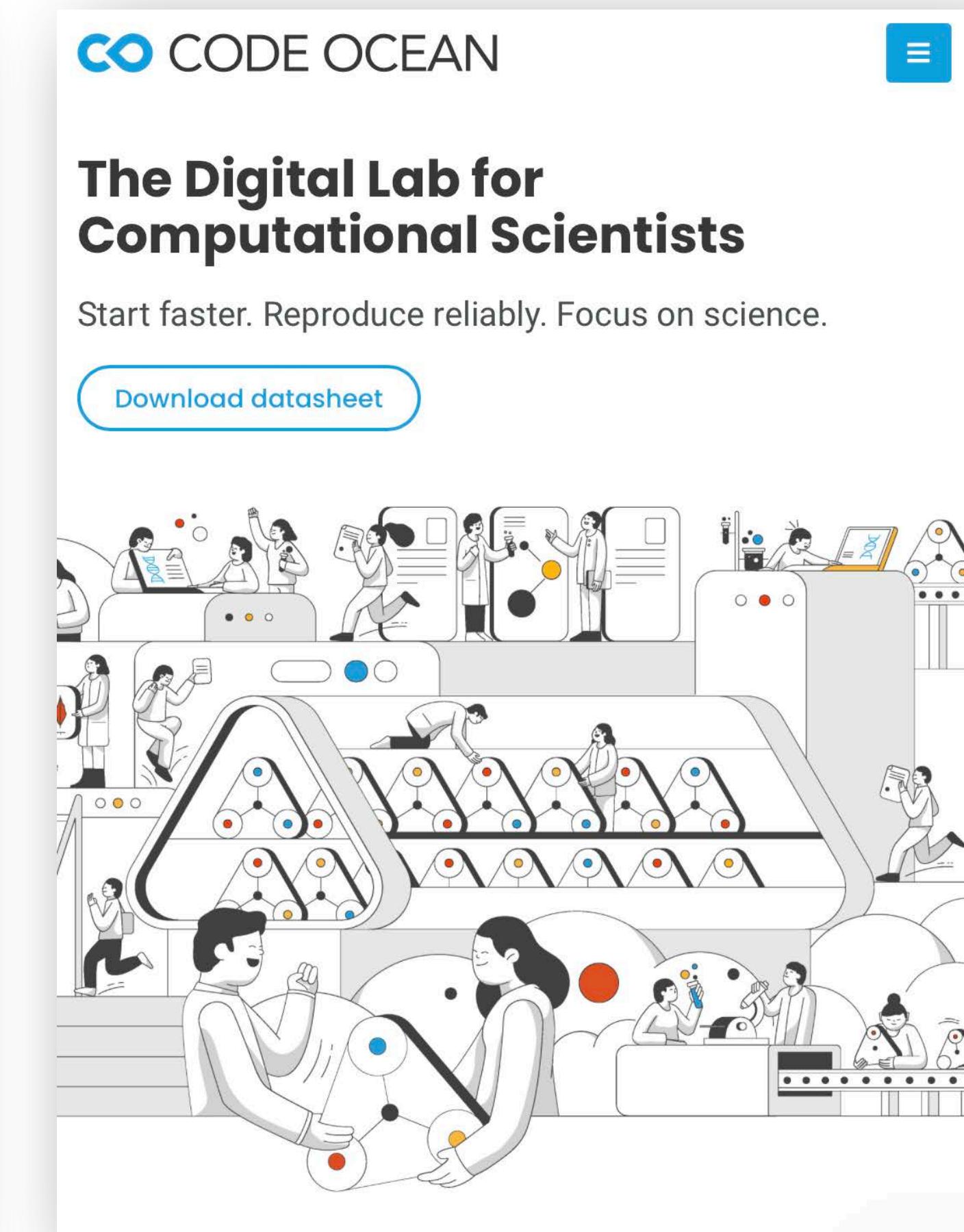
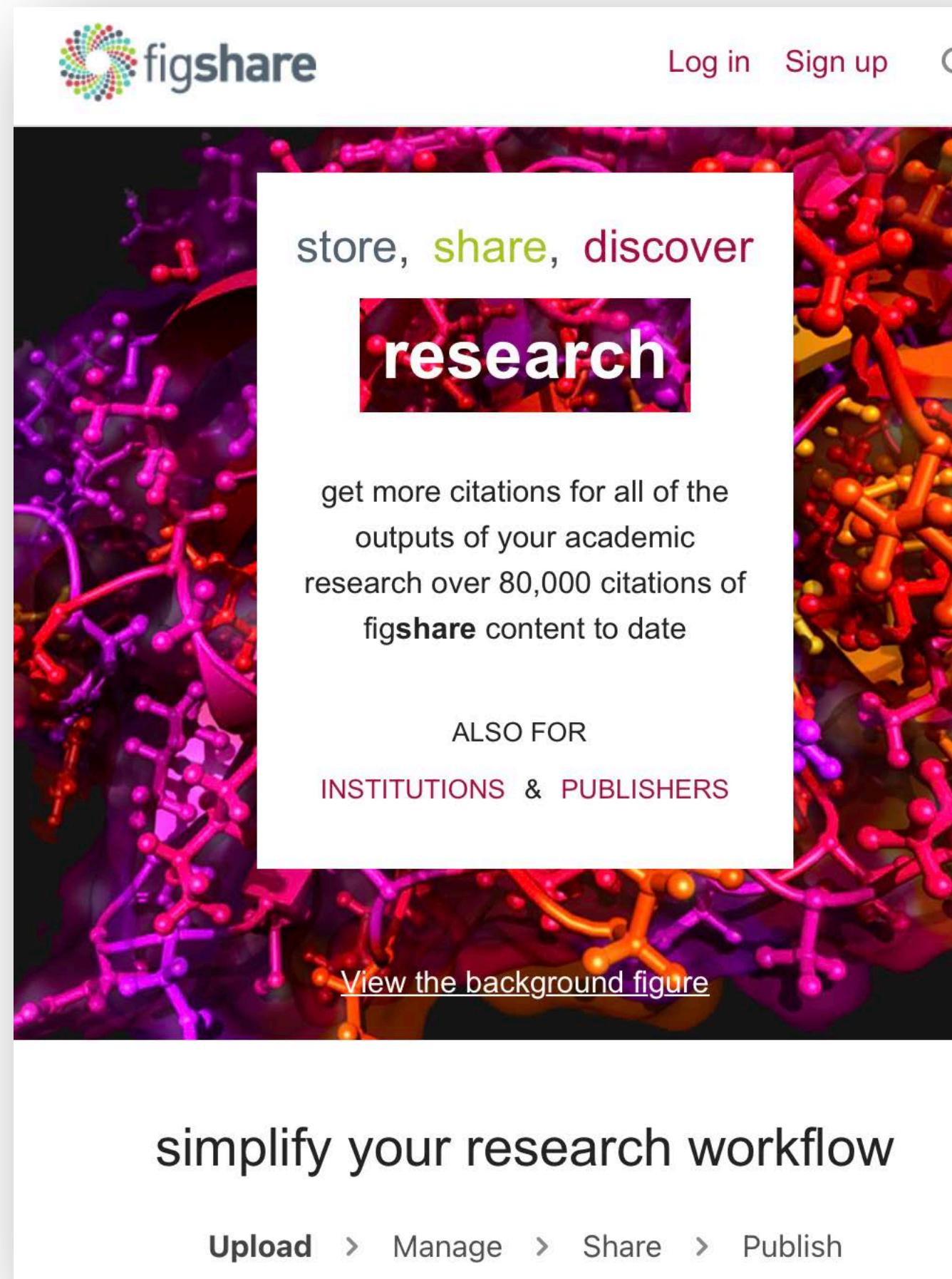
No associated or additional data

<https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>

<https://www.snf.ch/en/WtezJ6qxuTRnSYgF/topic/open-research-data-which-data-repositories-can-be-used>

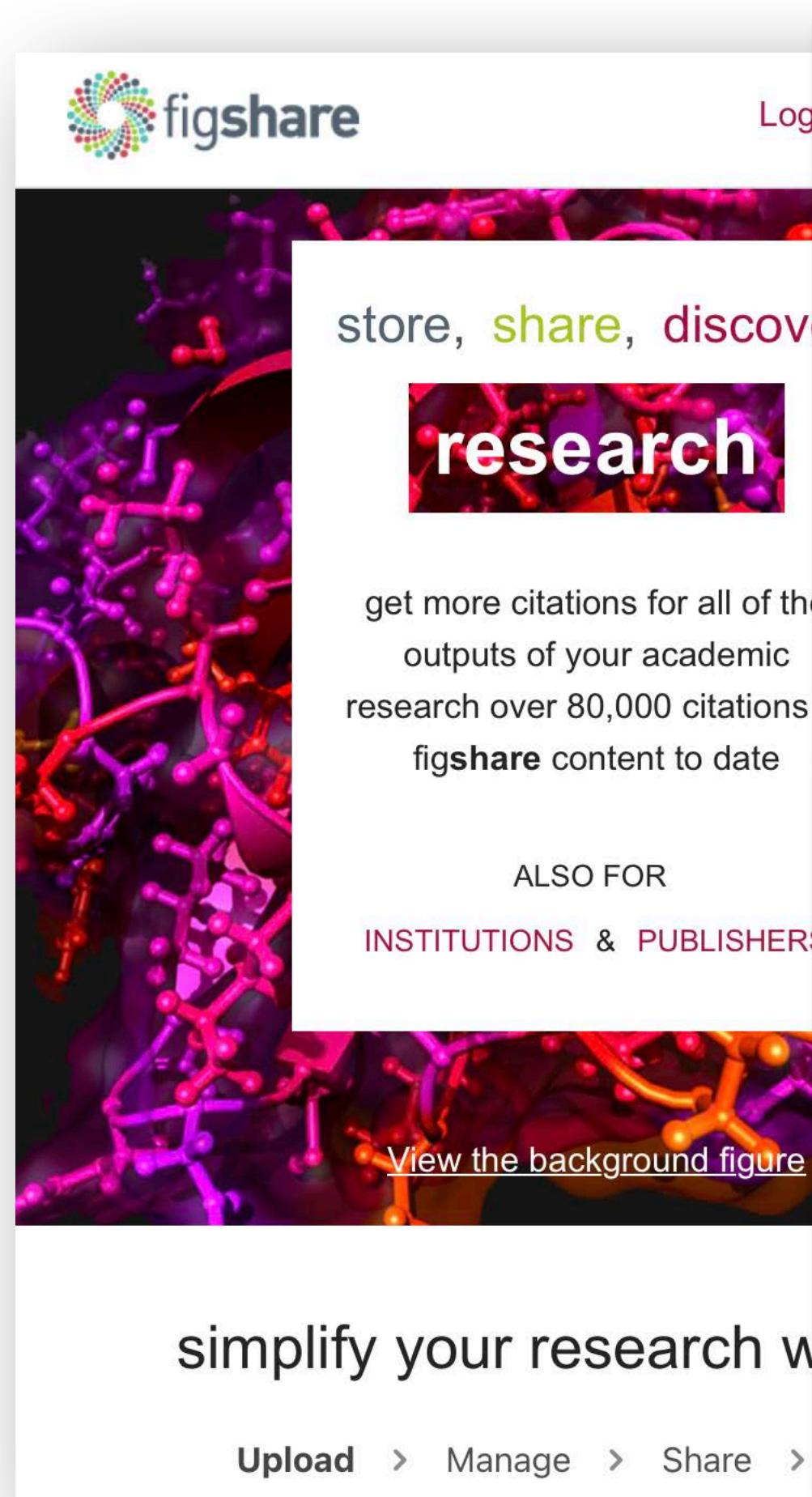
Generalist repositories

- A number exist, see list in previous links



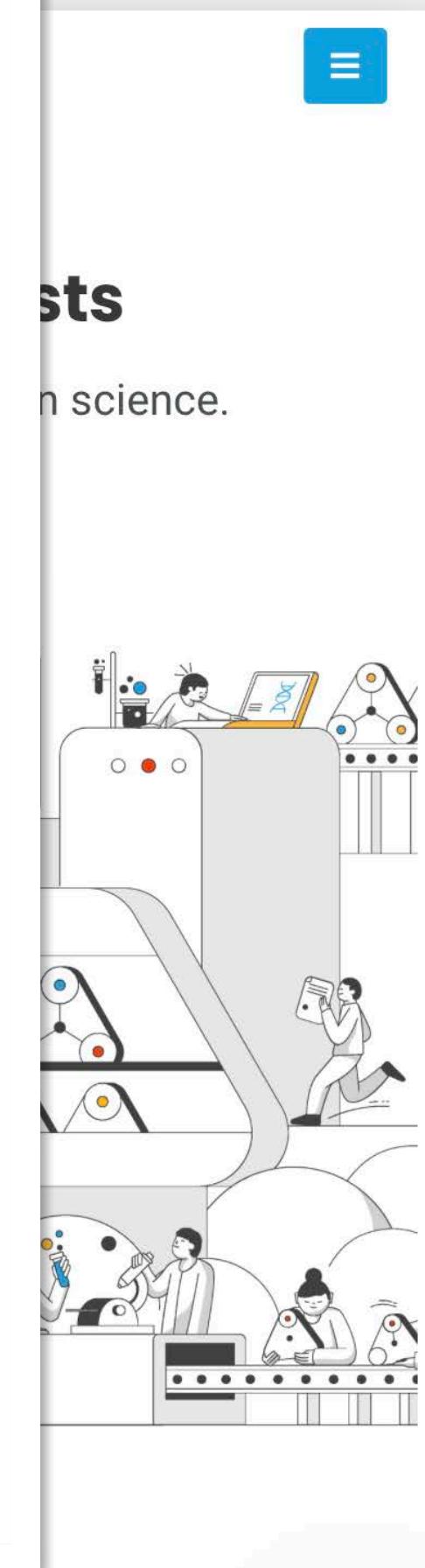
Generalist repositories

- A number exist, see list in previous links

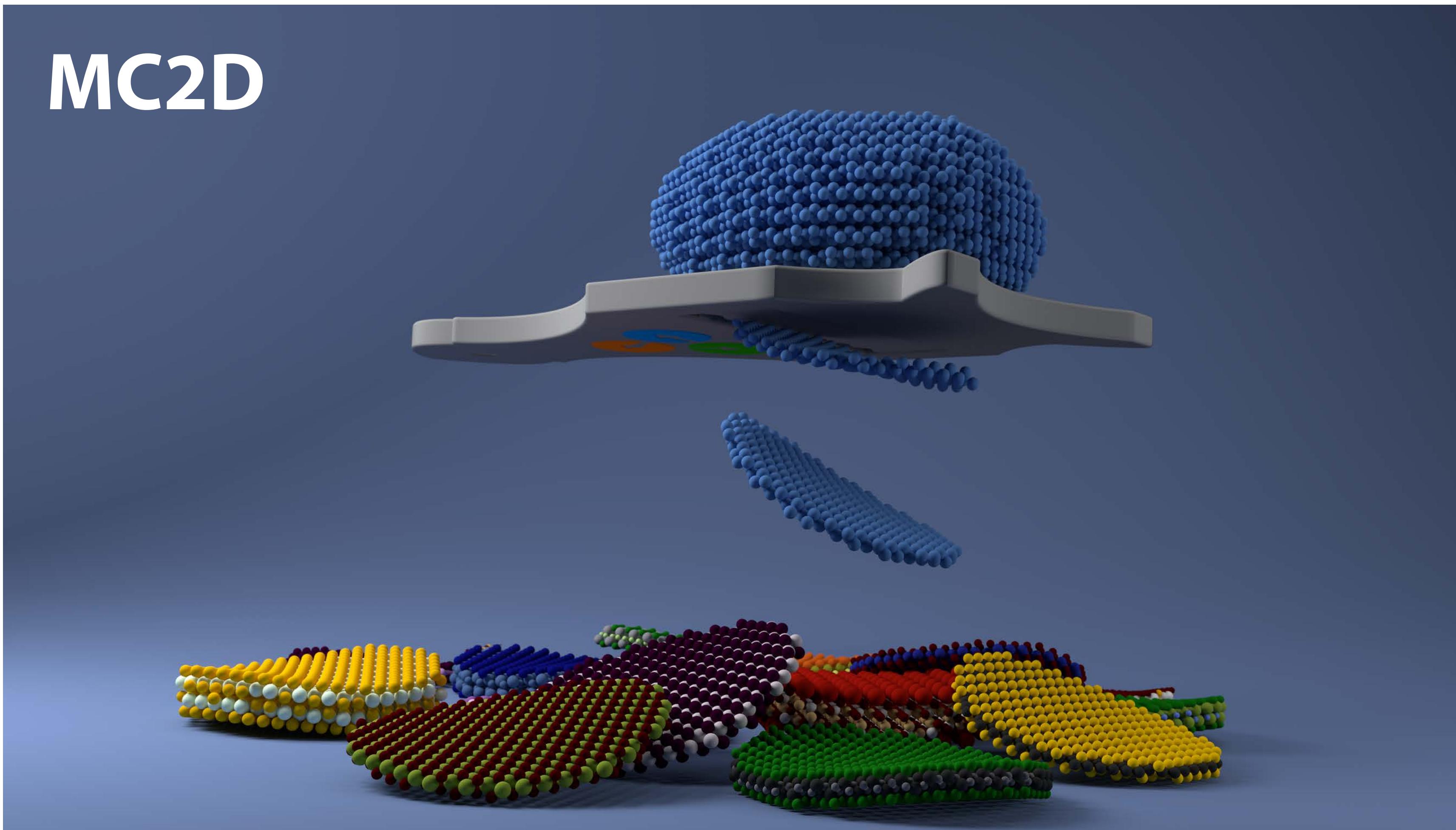


Discipline-specific repository
might provide features
more suited to your research
(and higher visibility among your peers)

Generalistic repos: still good option if you
only need a DOI and long-term preservation
of data associated to your paper



Combining reproducible workflows with data repositories



N. Mounet et al., Nat. Nanotech. 13 246 (2018)

<https://www.materialscloud.org/mc2d>

N. Mounet et al., Materials Cloud Archive 2020.158
(2020), doi: 10.24435/materialscloud:az-b2

FAIR data sharing (MC2D): Materials Cloud Archive, Discover, Explore

ARCHIVE

materialscloud:2017.0008/v3

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mounet¹, Marco Gibertini¹, Philippe Schwaller¹, Davide Campi¹, Andrius Merkys^{1,2}, Antimo Marrazzo¹, Thibault Sohier¹, Ivano E. Castelli¹, Andrea Cepellotti¹, Giovanni Pizzi¹, Nicola Marzari¹

¹ 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

² Vilnius University Institute of Biotechnology, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

DOI: 10.24435/materialscloud:2017.0008/v3 [version v3]

Publication date: Apr 03, 2019

How to cite this record

Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Davide Campi, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds, Materials Cloud Archive 2017.0008/v3 (2019), doi: 10.24435/materialscloud:2017.0008/v3.

Description

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozens of 2D materials have been successfully synthesized or exfoliated. Here, we search for novel 2D materials that can be easily exfoliated from their parent compounds. Starting from 108423 unique, experimentally known three-dimensional compounds we identify a subset of 5619 that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van-der-Waals density-functional theory, validated against experimental structural data and calculated random-phase-approximation binding energies, allow to identify 1825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search for optimal properties. For a subset of 258 compounds we explore vibrational, electronic, magnetic, and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors. This archive entry contains the database of 2D materials (structural parameters, band structures, binding energies, phonons for the subset of the 258 easily exfoliable materials with less than 6 atoms, structures and binding energies for the remaining 1567 materials) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

Selected 2d materials via interactive periodic table and view their properties (with links to provenance)
Explore interface providing access to the full database

FILES

File name	Size	Description
2D_materials.tar.gz [MD5]	113.0 MiB	We provide 258 two-dimensional crystal structures (lattice vectors, atomic species and positions), exfoliated from three-dimensional experimental crystal structures. The structures were relaxed at the DFT-PBE level. Together with each structure, a set of materials properties is also given (at the DFT-PBE level): chemical formula, spacegroup, structural prototype, magnetic state, magnetization, band-gap, electronic bands, and phonon

Direct links to curated data
in DISCOVER section

N. Mounet et al., Materials Cloud
Archive 2020.158 (2020),
doi: 10.24435/materialscloud:az-b2

FAIR data sharing (MC2D): Materials Cloud Archive, Discover, Explore

ARCHIVE

materialscloud:2017.0008/v3

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mounet¹, Marco Gibertini¹, Philippe Schwaller¹, Davide Campi¹, Andrius Merkys^{1,2}, Antimo Marrazzo¹, Thibault Sohier¹, Ivano E. Castelli¹, Andrea Cepellotti¹, Giovanni Pizzi¹, Nicola Marzari¹

¹ 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

² Vilnius University Institute of Biotechnology, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

DOI: 10.24435/materialscloud:2017.0008/v3 [version v3]

Publication date: Apr 03, 2019

How to cite this record

Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Davide Campi, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds, Materials Cloud Archive 2017.0008/v3 (2019), doi: 10.24435/materialscloud:2017.0008/v3.

Description

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozens of 2D materials have been successfully synthesized or exfoliated. Here, we search for novel 2D materials that can be easily exfoliated from their parent compounds. Starting from 108423 unique, experimentally known three-dimensional compounds we identify a subset of 5619 that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van-der-Waals density-functional theory, validated against experimental structural data and calculated random-phase-approximation binding energies, allow to identify 1825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search for optimal properties. For a subset of 258 compounds we explore vibrational, electronic, magnetic, and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors. This archive entry contains the database of 2D materials (structural parameters, band structures, binding energies, phonons for the subset of the 258 easily exfoliable materials with less than 6 atoms, structures and binding energies for the remaining 1567 materials) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

Selected 2d materials via interactive periodic table and view their properties (with links to provenance)
Explore interface providing access to the full database

Files

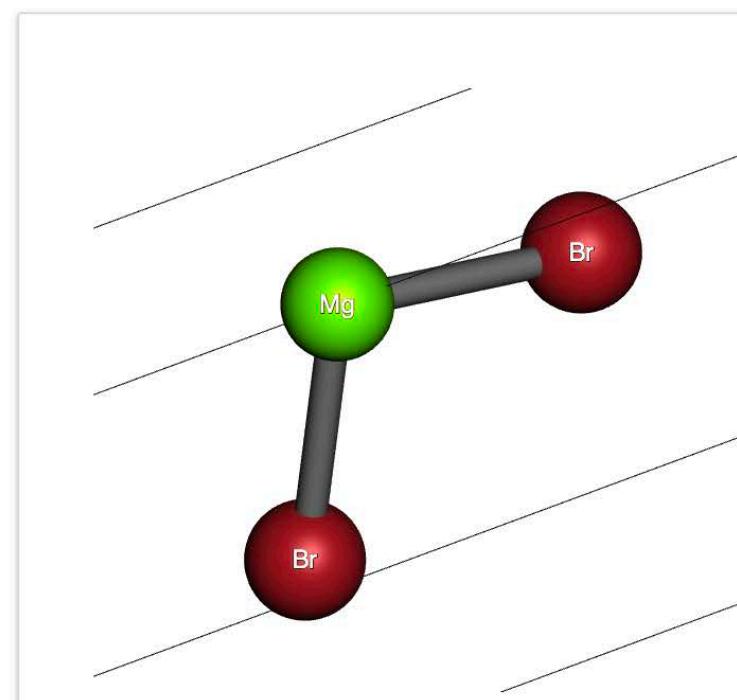
File name	Size	Description
2D_materials.tar.gz	113.0 MiB	We provide 258 two-dimensional crystal structures (lattice vectors, atomic species and positions), exfoliated from three-dimensional experimental crystal structures. The structures were relaxed at the DFT-PBE level. Together with each structure, a set of materials properties is also given (at the DFT-PBE level): chemical formula, spacegroup, structural prototype, magnetic state, magnetization, band-gap, electronic bands, and phonon

Direct links to curated data
in DISCOVER section

N. Mounet et al., Materials Cloud
Archive 2020.158 (2020),
doi: 10.24435/materialscloud:az-b2

DISCOVER

Compound: MgBr₂



Info and properties

See definitions...

Formula: MgBr₂

Spacegroup: P-3m1

Pointgroup: -3m

Prototype: CdI2

Band gap [eV]: 4.8

Magnetic properties:

Magnetic State: non-magnetic

Tot. Magnetization [$\mu\text{B}/\text{cell}$]: -

Abs. Magnetization [$\mu\text{B}/\text{cell}$]: -

Binding Energies:

DF2-C09 Binding energy [meV/ \AA^2]: 10.2

(From parent COD 9009107)

rVV10 Binding energy [meV/ \AA^2]: 15.3

(From parent COD 9009107)

Delta in interlayer distance (vdW vs revPBE):

$\Delta_{\text{DF2}} [\%]$: 17.1

(From parent COD 9009107)

$\Delta_{\text{rVV10}} [\%]$: 18.3

(From parent COD 9009107)

UUID links to jump to the
provenance graph in the
EXPLORE section

<https://www.materialscloud.org/mc2d>

FAIR data sharing (MC2D): Materials Cloud Archive, Discover, Explore

ARCHIVE

materialscloud:2017.0008/v3

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mounet¹, Marco Gibertini¹, Philippe Schwaller¹, Davide Campi¹, Andrius Merkys^{1,2}, Antimo Marrazzo¹, Thibault Sohier¹, Ivano E. Castelli¹, Andrea Cepellotti¹, Giovanni Pizzi¹, Nicola Marzari¹

¹ 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

² Vilnius University Institute of Biotechnology, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

DOI: 10.24435/materialscloud:2017.0008/v3 [version v3]

Publication date: Apr 03, 2019

How to cite this record

Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Davide Campi, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds, Materials Cloud Archive 2017.0008/v3 (2019), doi: 10.24435/materialscloud:2017.0008/v3.

Description

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozens of 2D materials have been successfully synthesized or exfoliated. Here, we search for novel 2D materials that can be easily exfoliated from their parent compounds. Starting from 108423 unique, experimentally known three-dimensional compounds we identify a subset of 5619 that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van-der-Waals density-functional theory, validated against experimental structural data and calculated random-phase-approximation binding energies, allow to identify 1825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search from for optimal properties. For a subset of 258 compounds we explore vibrational, electronic, magnetic, and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors. This archive entry contains the database of 2D materials (structural parameters, band structures, binding energies, phonons for the subset of the 258 easily exfoliable materials with less than 6 atoms, structures and binding energies for the remaining 1567 materials) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

Selected 2d materials via interactive periodic table and view their properties (with links to provenance)

Explore interface providing access to the full database

Files

File name	Size	Description
2D_materials.tar.gz	113.0 MiB	We provide 258 two-dimensional crystal structures (lattice vectors, atomic species and positions), exfoliated from three-dimensional experimental crystal structures. The structures were relaxed at the DFT-PBE level. Together with each structure, a set of materials properties is also given (at the DFT-PBE level): chemical formula, spacegroup, structural prototype, magnetic state, magnetization, band-gap, electronic bands, and phonon

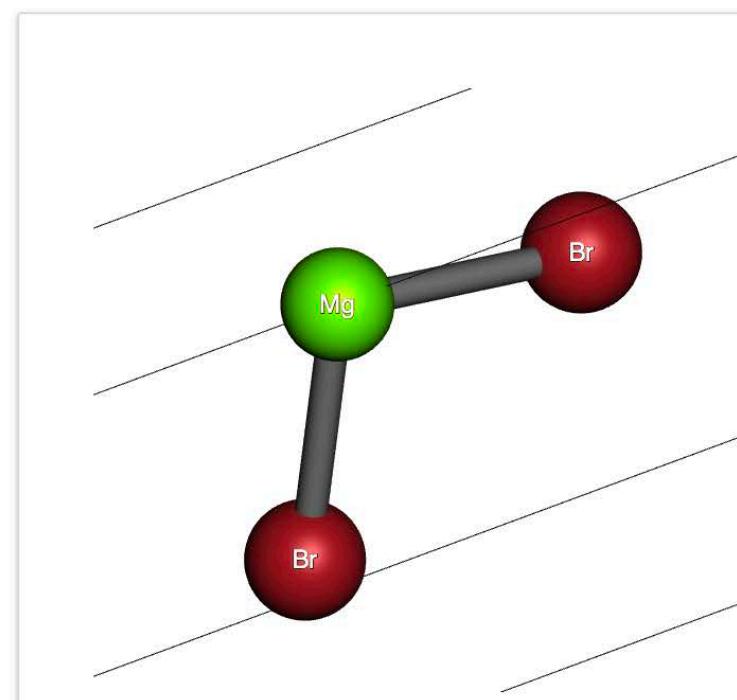
Direct links to curated data
in DISCOVER section

N. Mounet et al., Materials Cloud
Archive 2020.158 (2020),

doi: 10.24435/materialscloud:az-b2

DISCOVER

Compound: MgBr₂



Info and properties
See definitions...

Formula: MgBr₂
Spacegroup: P-3m1
Pointgroup: -3m
Prototype: CdI2
Band gap [eV]: 4.8

Magnetic properties:
Magnetic State: non-
Tot. Magnetization [μ]
Abs. Magnetization [μ]

Binding Energies:

DF2-C09 Binding energy
(From parent COD 9009107)
rVV10 Binding energy
(From parent COD 9009107)
Delta in interlayer distance
 Δ_{DF2} [%]: 17.1
 Δ_{rVV10} [%]: 18.3

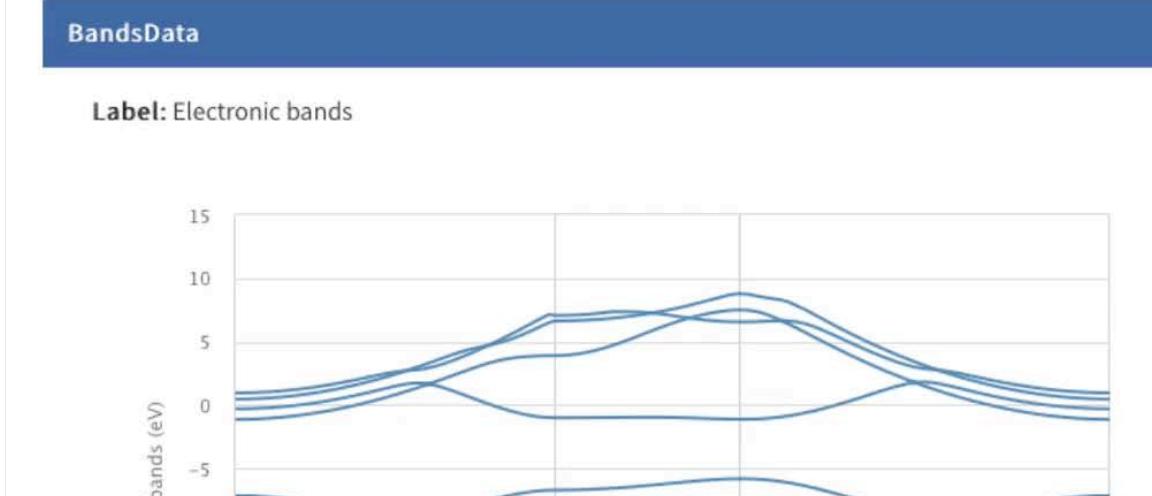
EXPLORE

Selected Profile: 2D Structures DOI: 10.24435/materialscloud:2017.0008/v2

Grid Details Statistics

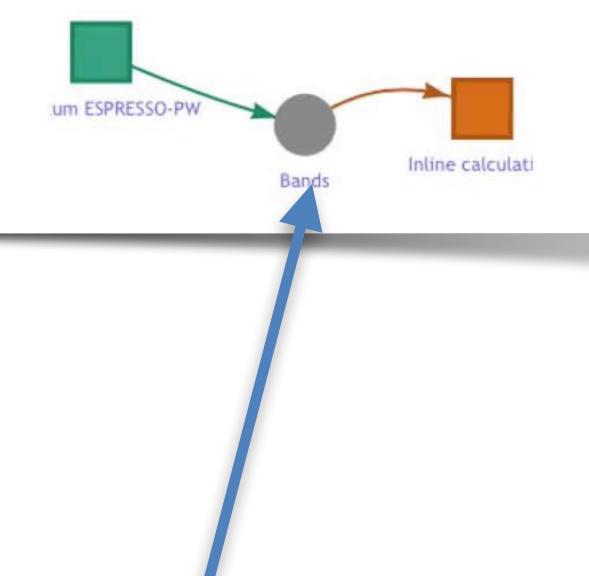
e7db98c1-9d25-4872-8236-68559c5b0702 GO

UUID: e7db98c1-9d25-4872-8236-68559c5b0702
Type: data.array.bands.BandsData.
Created at 6 January 2017
Modified 8 months ago
davide.campi@epfl.ch



AiiDA Provenance Browser

Selected node, Inputs, Outputs



UUID links to jump to the
provenance graph in the
EXPLORE section

<https://www.materialscloud.org/mc2d>

Browse the full AiiDA
provenance graph
(inputs, outputs, ...) at any
level

Data management plans (DMP) and FAIR principles

- Combination of AiiDA + Materials Cloud (Discover, Explore, Archive): **FAIR-compliant sharing**
- **Findable**: DOIs with standardized metadata
- **Accessible**: web interface to browse data, calculations and provenance, curated data in Discover section
- **Interoperable**: data linked via the AiiDA directed graph; data structures reusable between different codes
- **Reusable**: downloadable data, encourage open (CC) licences
- + **reproducible** thanks to *full provenance*

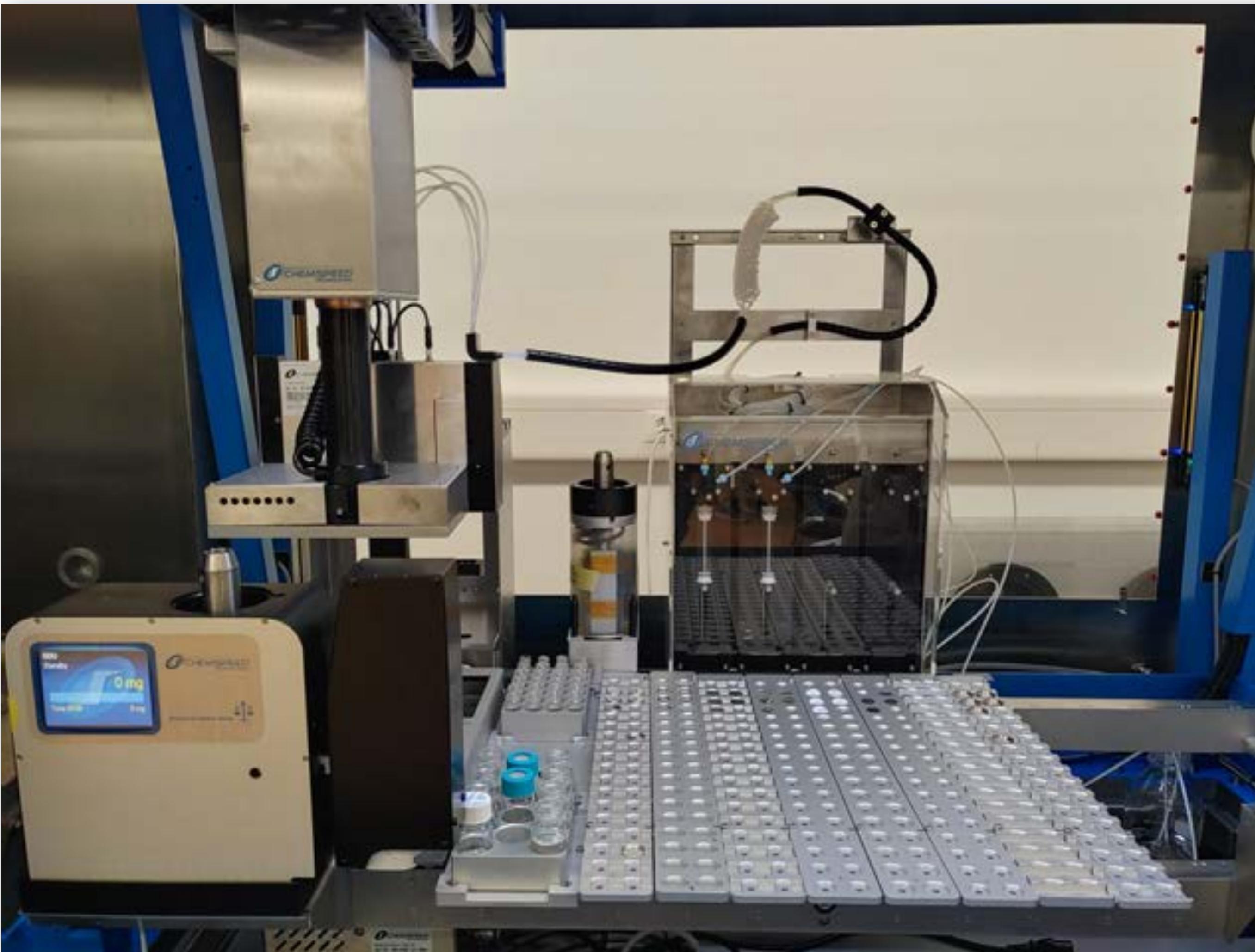
- We provide DMP templates for researchers using Materials Cloud

Below, we provide templates for data management plans using the Materials Cloud Archive (with and without AiiDA).

Feel free to [contact us](#) with any questions regarding the use of the Materials Cloud Archive as part of your data management plan.

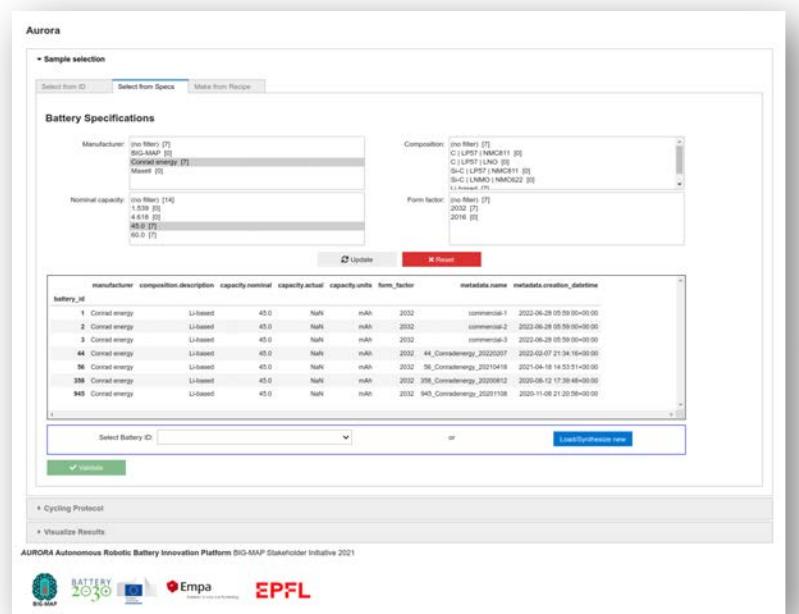
Funding Body	DMP template (using  AiiDA)	DMP template (no AiiDA)
SNF	.docx .odt .pdf	.docx .odt .pdf
H2020	.docx .odt .pdf	.docx .odt .pdf

Getting out of the comfort zone: automated experiments



Workflow engines to drive robotic experiments

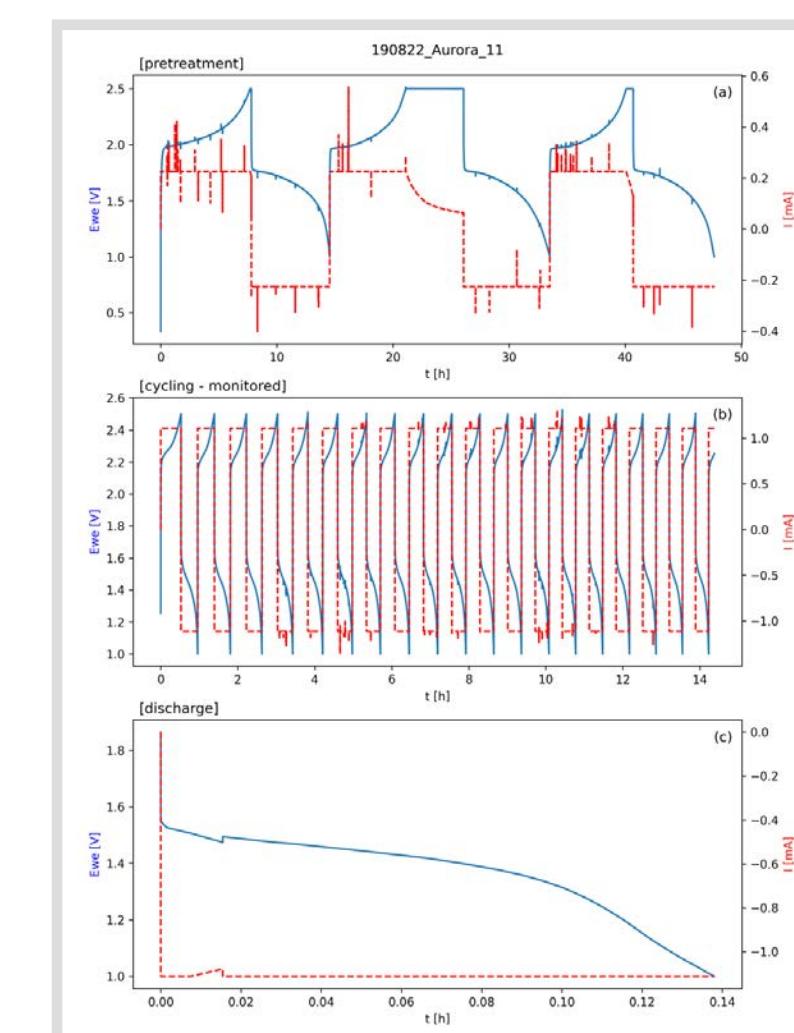
Aurora-AiiDA integration (collaboration with Empa)



AiiDA workflows + GUI



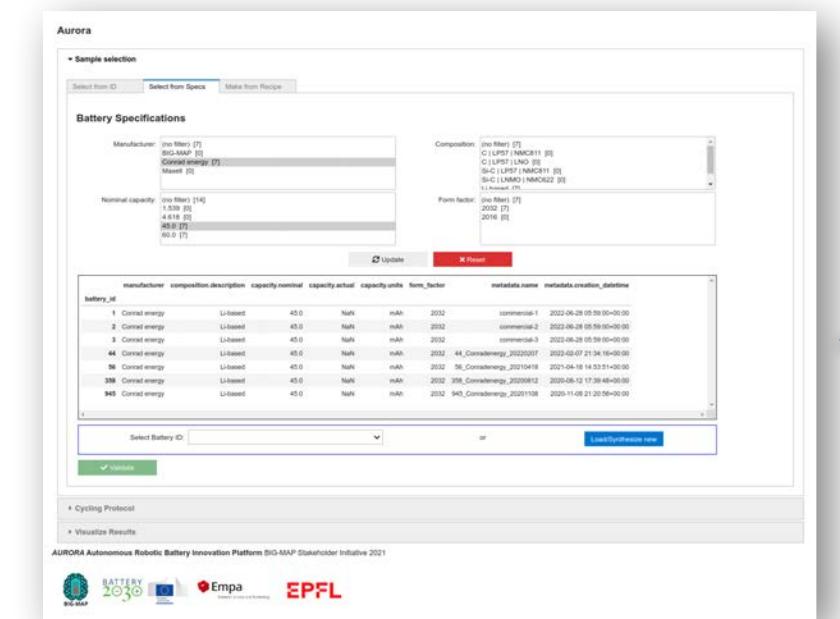
"Aurora" battery assembly robot
+ cell cycling platform



Automated coin-cell assembly
and cycling test with real-time
monitoring and control

Workflow engines to drive robotic experiments

Aurora-AiiDA integration (collaboration with Empa)

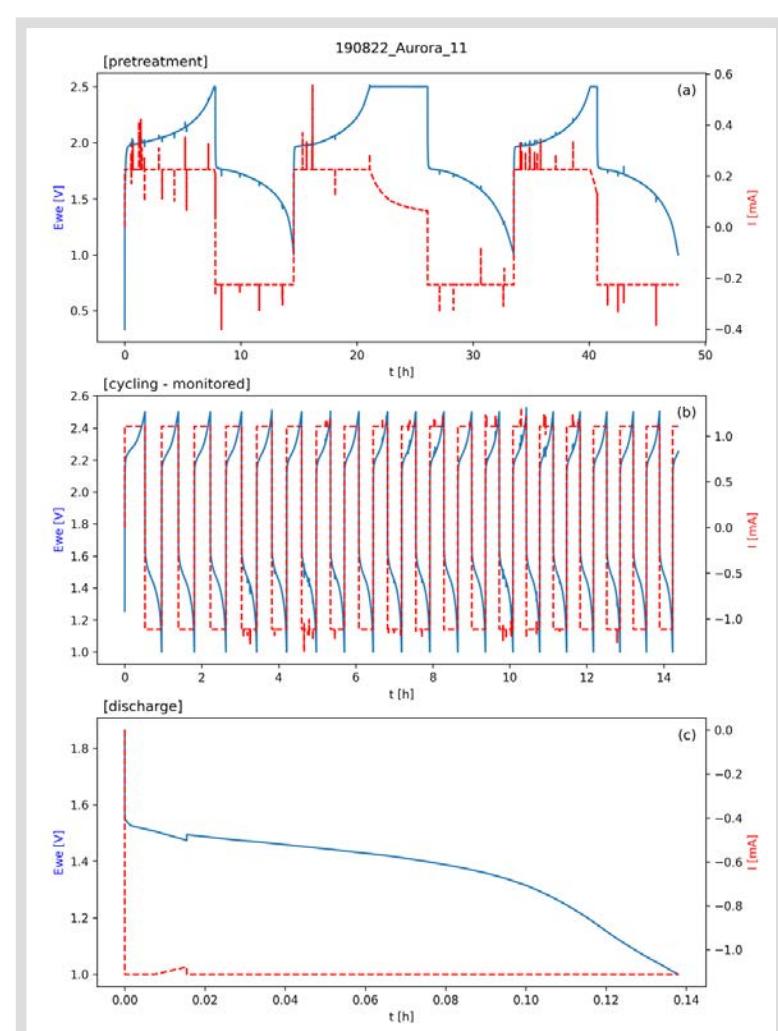


AiiDA workflows + GUI

+

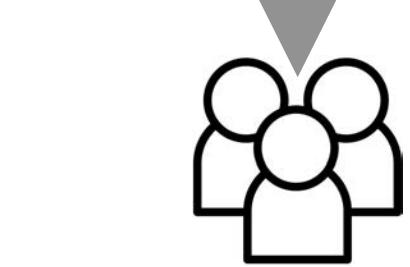
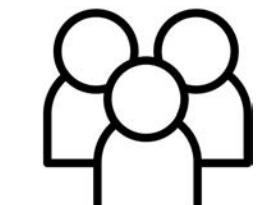


"Aurora" battery assembly robot
+ cell cycling platform



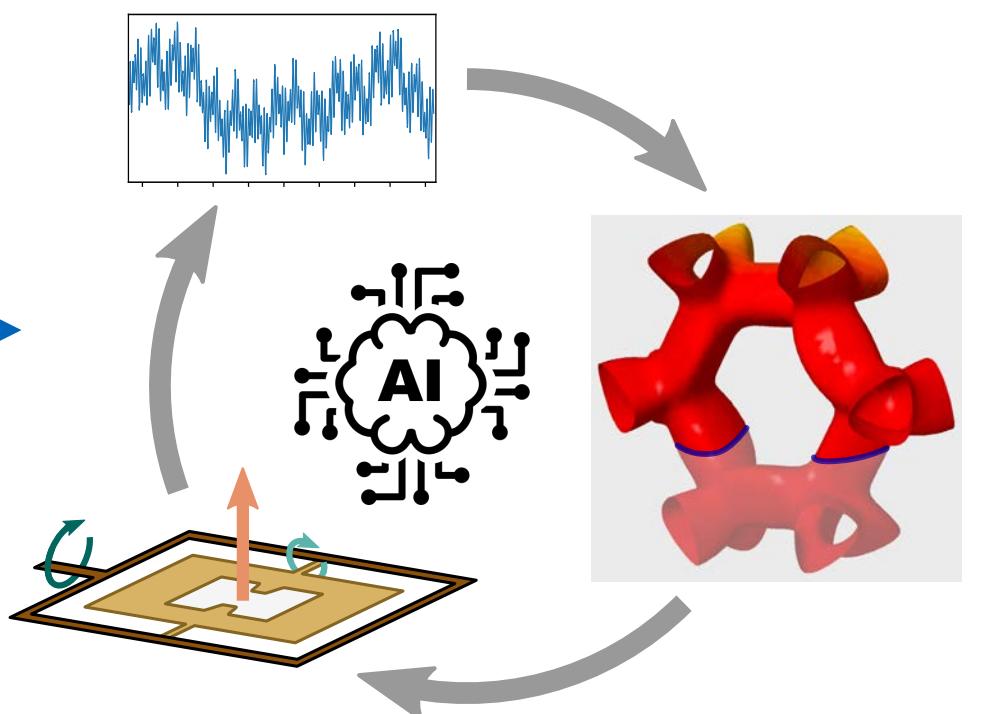
Automated coin-cell assembly
and cycling test with real-time
monitoring and control

Automatic
Search strategy defined
at the start



Crucial first step

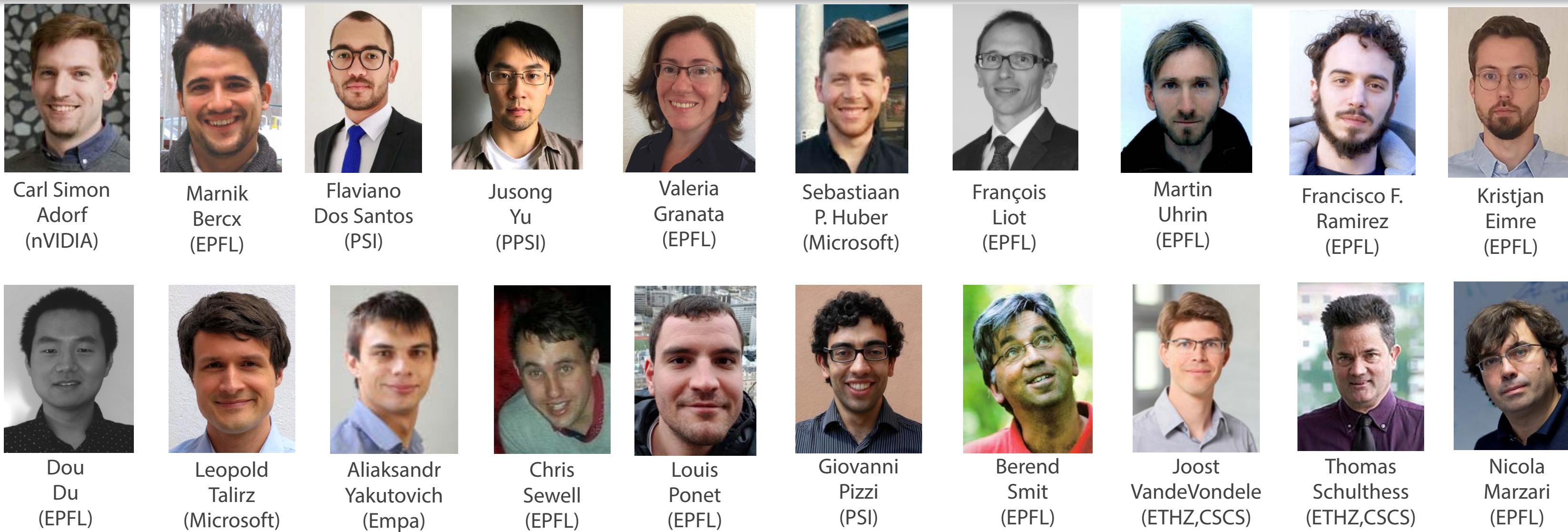
Autonomous
Closed-loop, decisions taken
based on real-time results



Accelerated discovery

AiiDA and Materials Cloud teams

The Materials Cloud And AiiDA teams



Contributors for the 40+ plugins: **Quantum ESPRESSO, Wannier90, CP2K, FLEUR, YAMBO, SIESTA, VASP, CASTEP, CRYSTAL, ...**

Contributors to **aiida-core** and former AiiDA team members —
Oscar Arbelaez, Michael Atambo, Valentin Bersier, Marco Borelli, Jocelyn Boullier, Jens Bröder, Ivano E. Castelli, Andrea Cepellotti, Keija Cui, Vladimir Dikan, Marco Dorigo, Y.-W. Fang, Fernando Gargiulo, Marco Gibertini, Davide Grassano, Dominik Gresch, Conrad Johnston, Rico Häuselmann, Daniel Hollas, Eric Hontz, Jianxing Huang, Christoph Koch, Espen Flage-Larsen, Ian Lee, Daniel Marchand, Antimo Marrazzo, Andrius Merkys, Simon Pintarelli, Nicolas Mounet, Tiziano Müller, Gianluca Prandini, Philip Rüßmann, Riccardo Sabatini, Ole Schütt, Phillippe Schwaller, Andreas Stamminger, Atsushi Togo, Daniele Tomerini, Nicola Varini, Martin Uhrin, Jason Yu, Austin Zadoks, Bonan Zhu, Mario Zic, Spyros Zoupanos

Acknowledgements and funding



SNSF NCCR "MARVEL"

Discovery of new materials via simulations
and dissemination of curated data

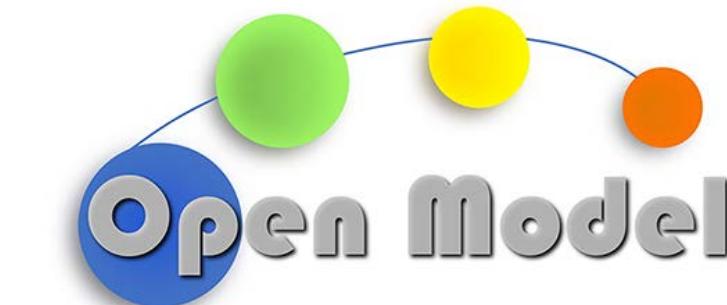
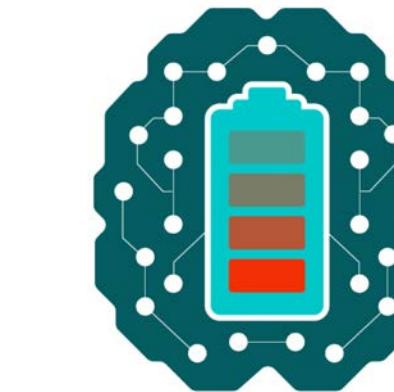
H2020 Centre of Excellence "MaX"

Scaling towards exascale machines and
high-throughput efficiency

Moreover:



swissuniversities



Summary

- Automated provenance tracking guarantees reproducibility for high-throughput HPC simulations
- Generalists repositories exist and are very valuable; discipline-specific repositories provide added value and higher visibility to your research
- Journals and institutions recommend generalists and discipline-specific repositories
- Combining workflow tracking and open repositories: achieving the goals of Open Science for computational simulations (and robotic experiments, towards autonomous labs)