MATERIALS CLOUD

The Materials Cloud

Giovanni Pizzi

https://www.materialscloud.org

MARVEL
NATIONAL CENTRE OF COMPETENCE IN RESEARCH

MAX DRIVING THE EXASCALE TRANSITION

swissuniversities

FNSNF
SWISS NATIONAL SCIENCE FOUNDATION

INTERSECT

MarketPlace

EPFL

The Open Science Platform - Vision

Computational Materials Science Challenges

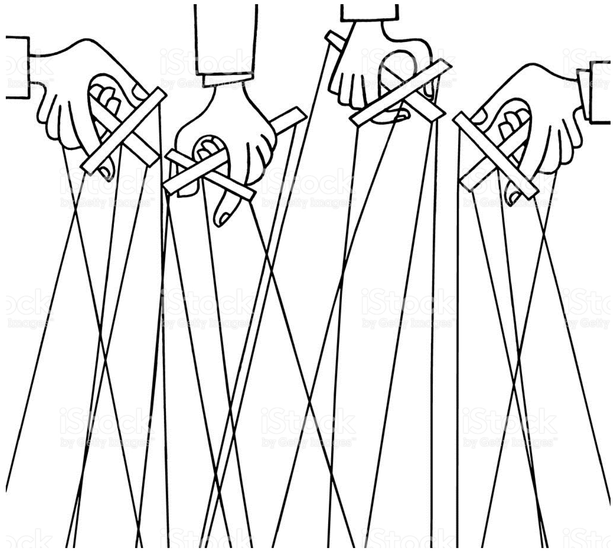
High-Throughput

Reproducibility

Open Data

Knowledge Transfer

- Organize large numbers of calculations
- Deal with corner cases (theory, code, infrastructure)
- Many strings to pull



Source: istockphoto.com

AiiDA

The Open Science Platform - Vision

Giovanni Pizzi

Computational Materials Science Challenges

- High-Throughput
- **Reproducibility**
- Open Data
- Knowledge Transfer

- Keep track of what you calculate
- Keep track of how you did it
- *Within* a research group: Can Alice reproduce what Bob computed 1 year ago?

Source: academiccoachingandwriting.org

The Open Science Platform - Vision

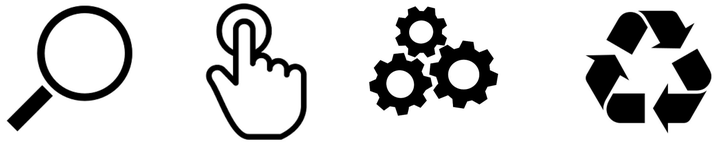
Giovanni Pizzi

Computational Materials Science Challenges

- High-Throughput
- Reproducibility
- **Open Data**
- Knowledge Transfer

- Open Data
  - ~~Supporting Information~~
  - ~~Just upload everything~~
  - FAIR data
- Making data FAIR is hard, can we make it easier?

Findable **A**ccessible **I**nteroperable **R**eusable



Source: Wikipedia

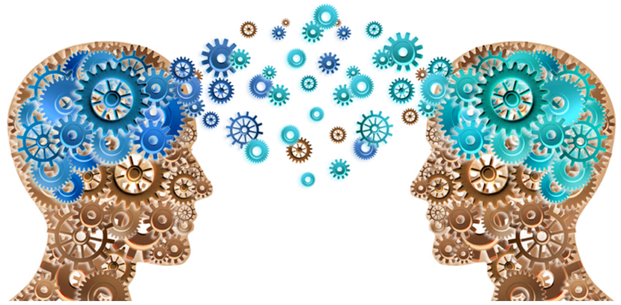
AiiDA + MATERIALS CLOUD

The Open Science Platform - Vision

Computational Materials Science Challenges

- High-Throughput
- Reproducibility
- Open Data
- **Knowledge Transfer**

- Share a workflow for your code with a collaborator or company
- Share the environment needed to run the workflow
- Reduce email traffic by sharing access to live dashboards with simulation results

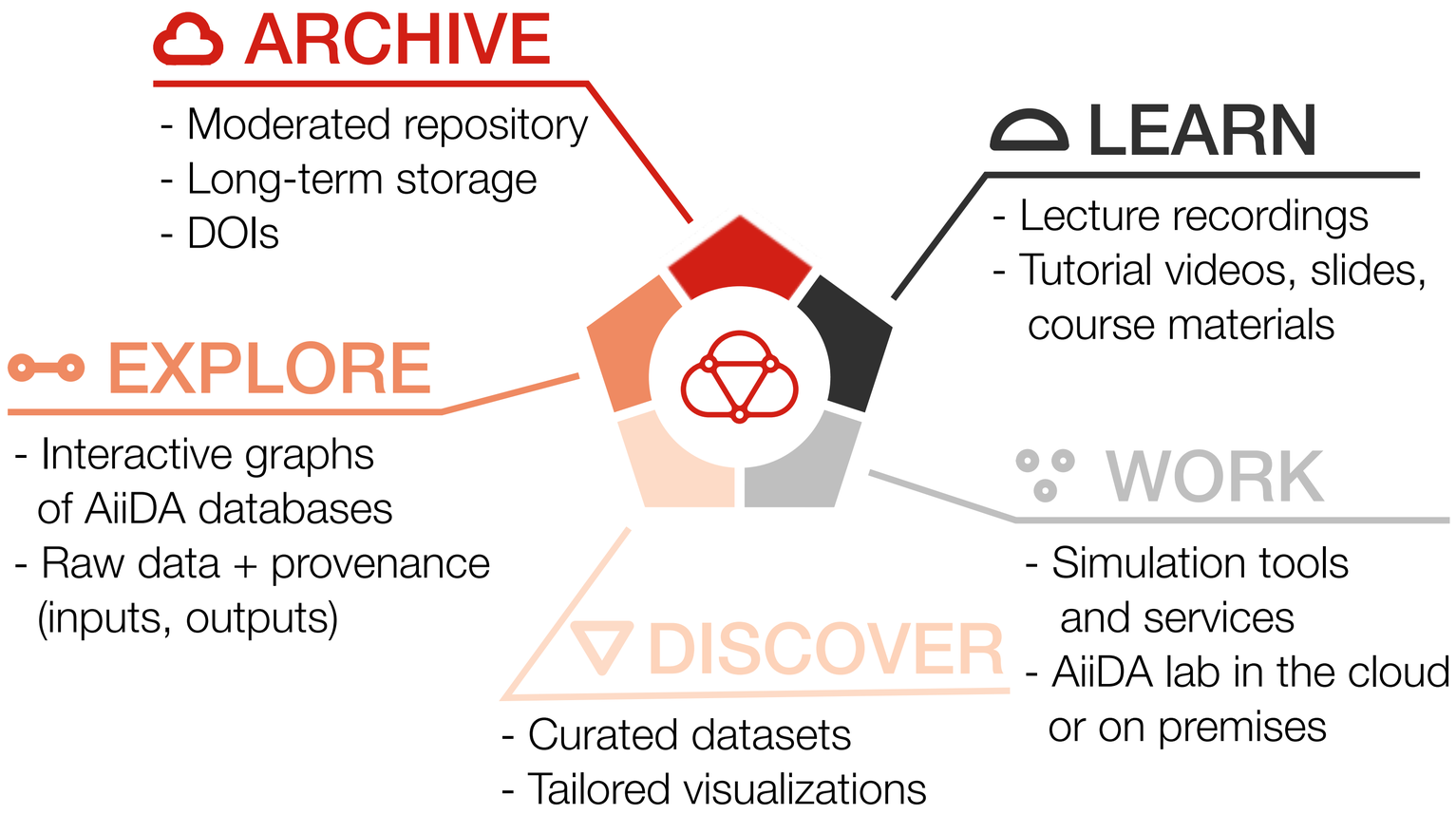


Source: quote.ucsd.edu

AiiDAlab

Materials Cloud

- **AiiDA** is the 'engine', like **Git -** used in production *since 2015*
- **Materials Cloud** is the dissemination platform (like **GitHub**) **and more** (cloud computing and data generation platform) - online since *Dec 2017*



Materials Cloud

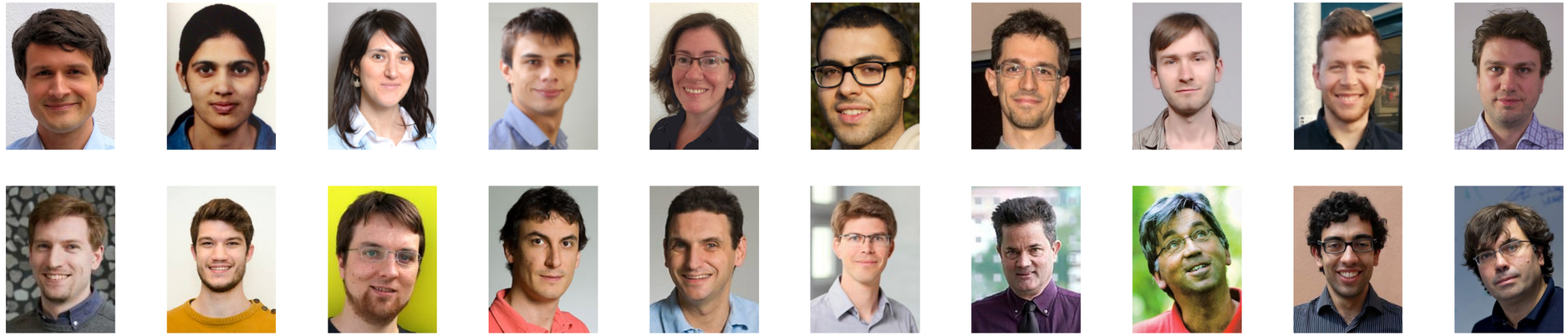
Materials Cloud, a platform for open computational science

Leopold Talirz, Snehal Kumbhar, Elsa Passaro, Aliaksandr V. Yakutovich, Valeria Granata, Fernando Gargiulo, Marco Borelli, Martin Uhrin, Sebastiaan P. Huber, Spyros Zoupanos, Carl S. Adorf, Casper W. Andersen, Ole Schütt, Carlo A. Pignedoli, Daniele Passerone, Joost VandeVondele, Thomas C. Schulthess, Berend Smit, Giovanni Pizzi, Nicola Marzari

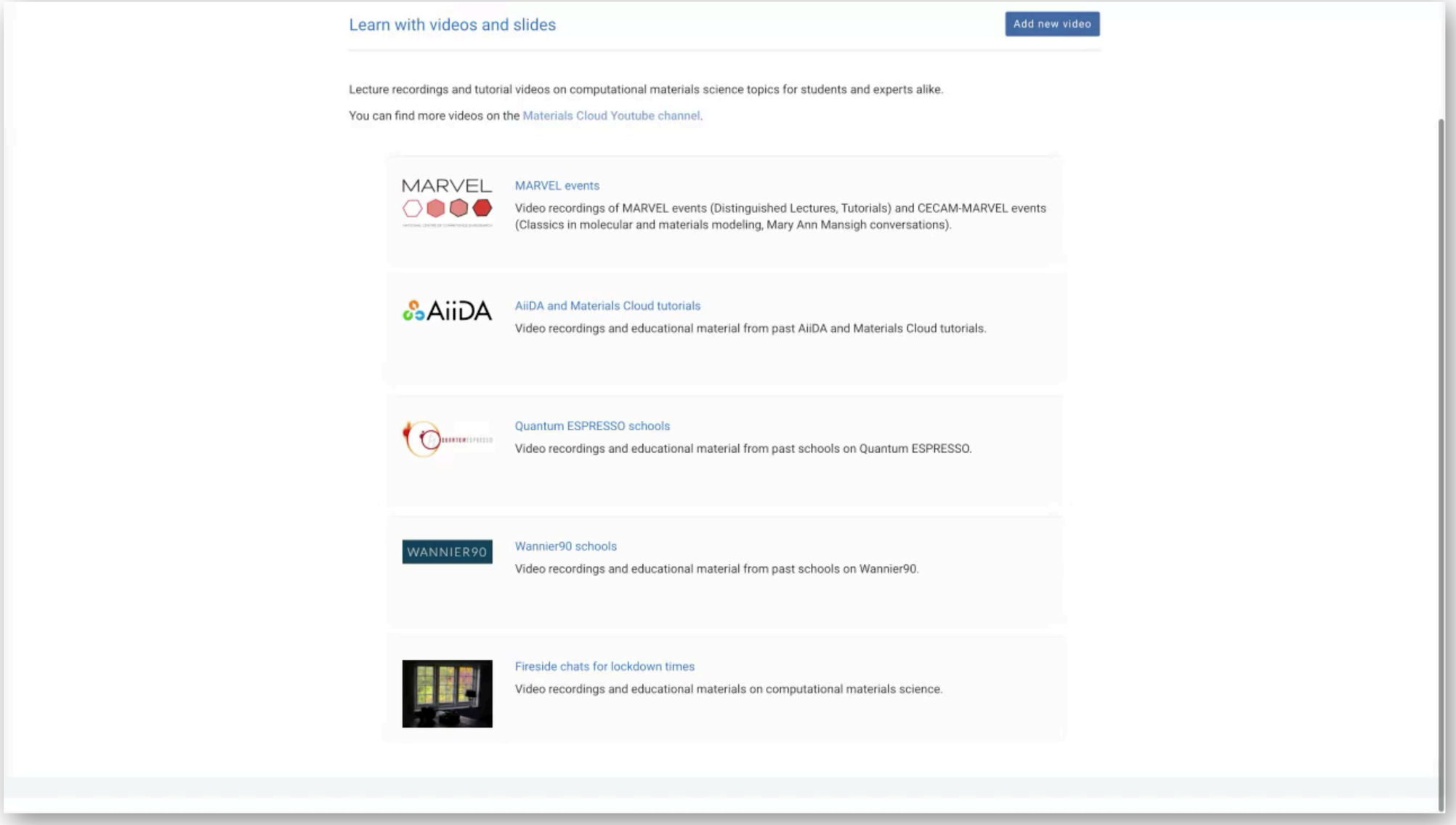
Materials Cloud is a platform designed to enable open and seamless sharing of resources for computational science, driven by applications in materials modelling. It hosts 1) archival and dissemination services for raw and curated data, together with their provenance graph, 2) modelling services and virtual machines, 3) tools for data analytics, and pre-/post-processing, and 4) educational materials. Data is citable and archived persistently, providing a comprehensive embodiment of the FAIR principles that extends to computational workflows. Materials Cloud leverages the AiiDA framework to record the provenance of entire simulation pipelines (calculations performed, codes used, data generated) in the form of graphs that allow to retrace and reproduce any computed result. When an AiiDA database is shared on Materials Cloud, peers can browse the interconnected record of simulations, download individual files or the full database, and start their research from the results of the original authors. The infrastructure is agnostic to the specific simulation codes used and can support diverse applications in computational science that transcend its initial materials domain.

Comments: 19 pages, 8 figures
Subjects: **Materials Science (cond-mat.mtrl-sci)**; Computational Physics (physics.comp-ph)
ACM classes: J.2; I.6; H.4
Cite as: arXiv:2003.12510 **[cond-mat.mtrl-sci]**

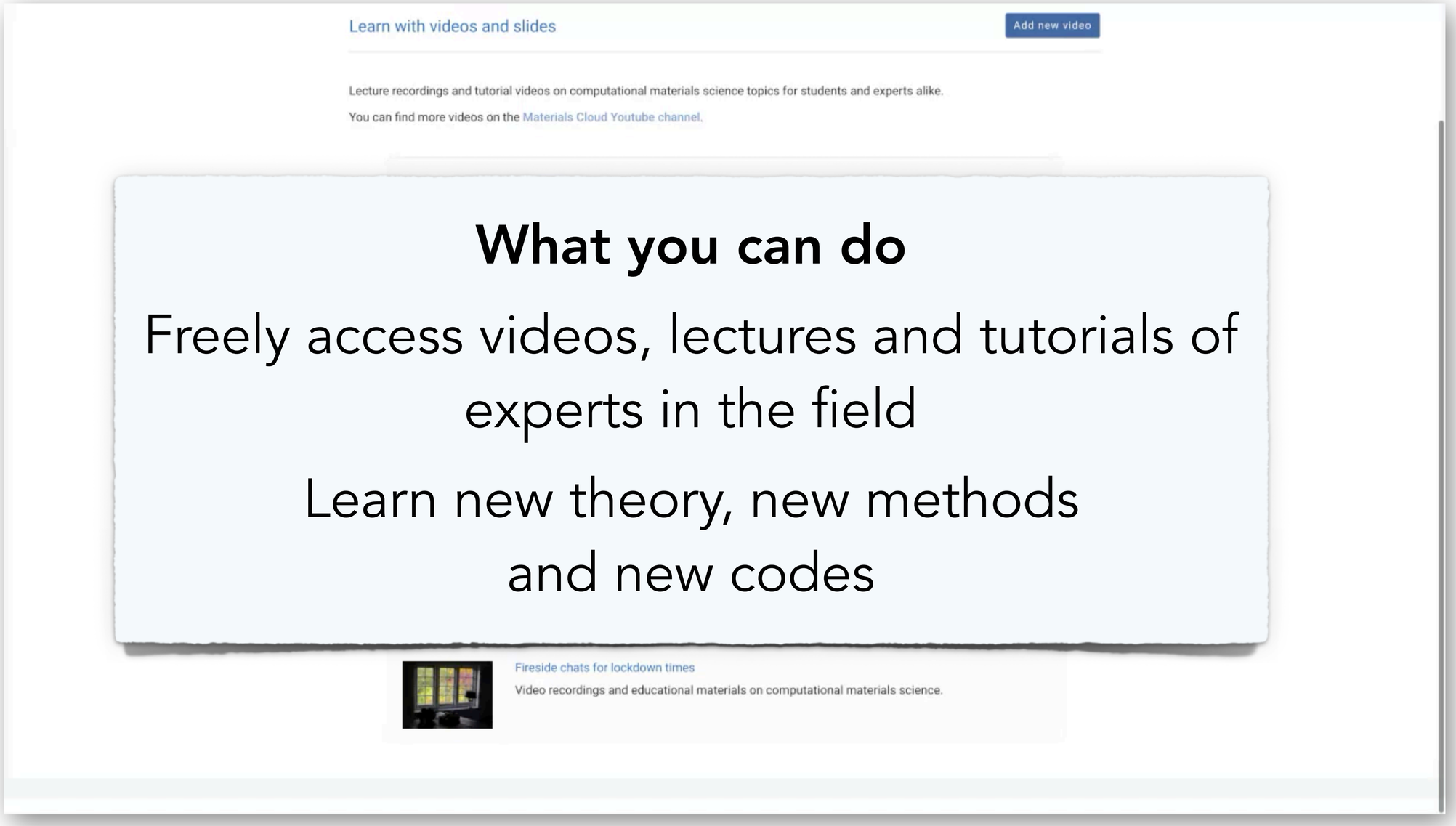
L. Talirz *et al.*, Materials Cloud, a platform for open computational science, arXiv:2003.12510 (2020)



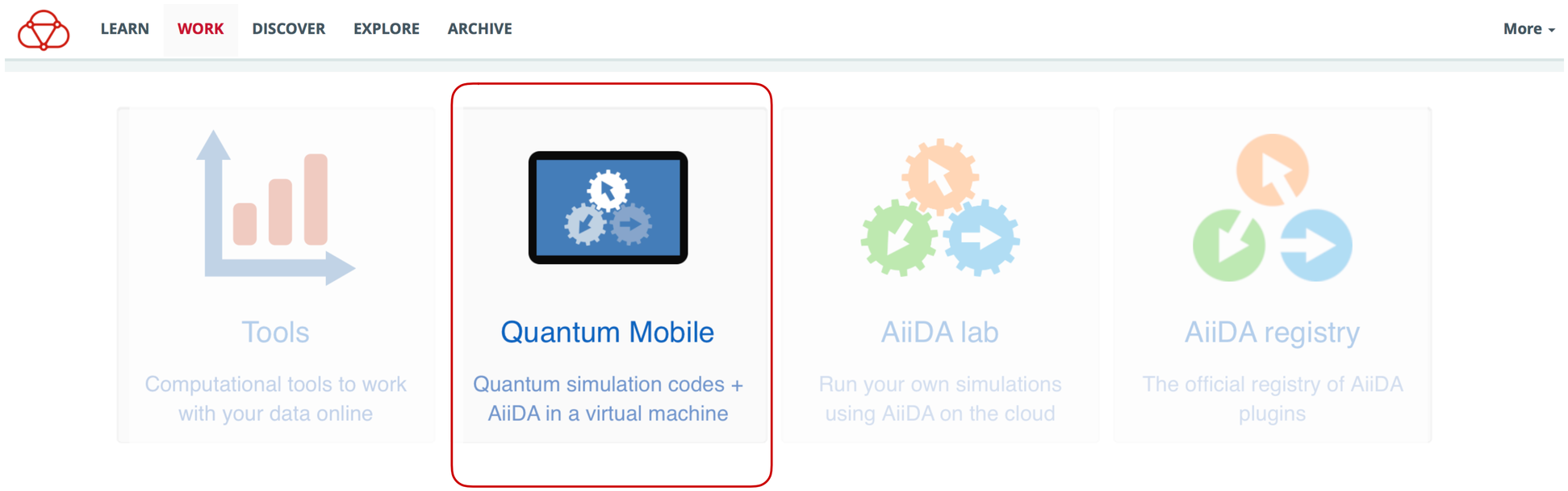
LEARN section



LEARN section



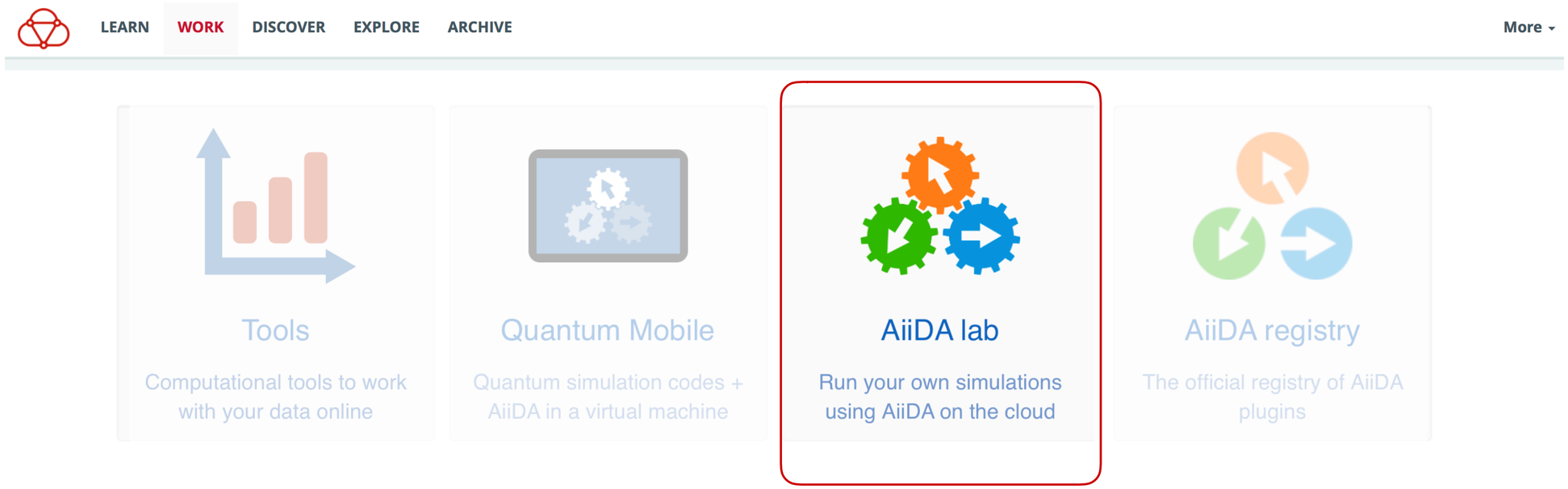
Data generation: Materials Cloud WORK



Quantum Mobile

- **Downloadable VM** with preinstalled **AiiDA and codes** like QE, Yambo, Fleur, Siesta, CP2K, …
- Also features a Cloud version on Amazon AWS, Google Compute Engine, Azure, …
- Ideal for **education** (e.g., Cloud and Desktop versions used for this AiiDA virtual tutorial)

Data generation: Materials Cloud WORK

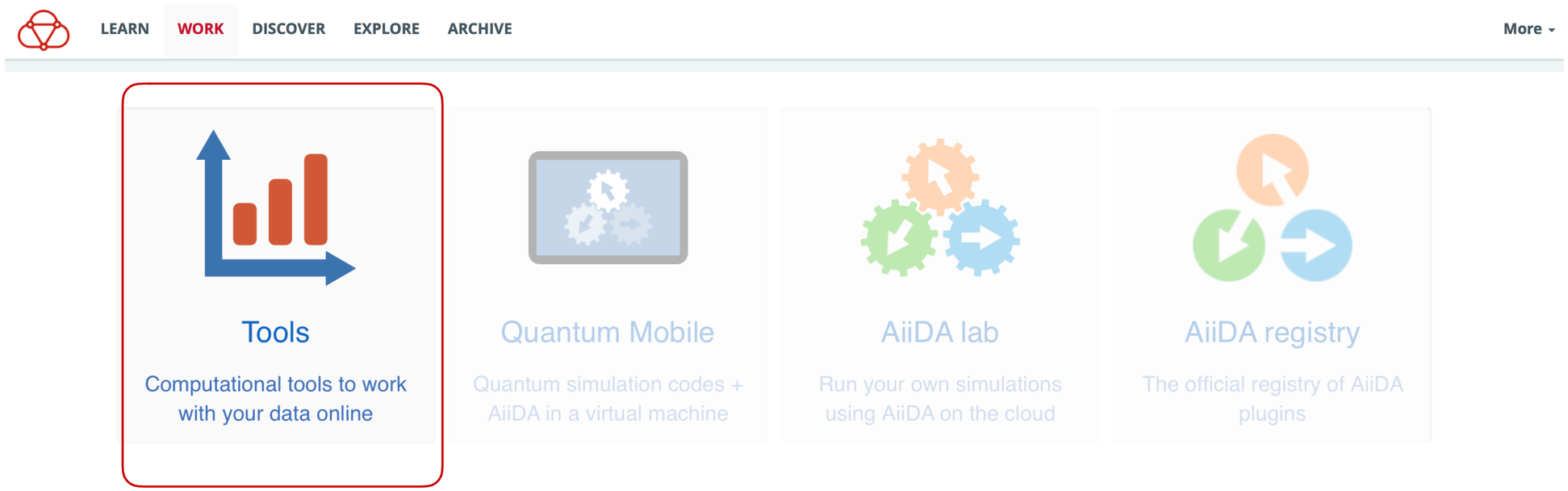


AiiDA Lab

- Comes with a preconfigured AiiDA setup, **ideal interface for GUIs for easy turn-key workflows**
- Custom **appmode** extension to make notebooks look&feel like real web apps knowing only python
- Using JupyterHub + DockerSpawner or Kubernetes

See presentation on AiiDA lab by Aliaksandr Yakutovich

Data generation: Materials Cloud WORK

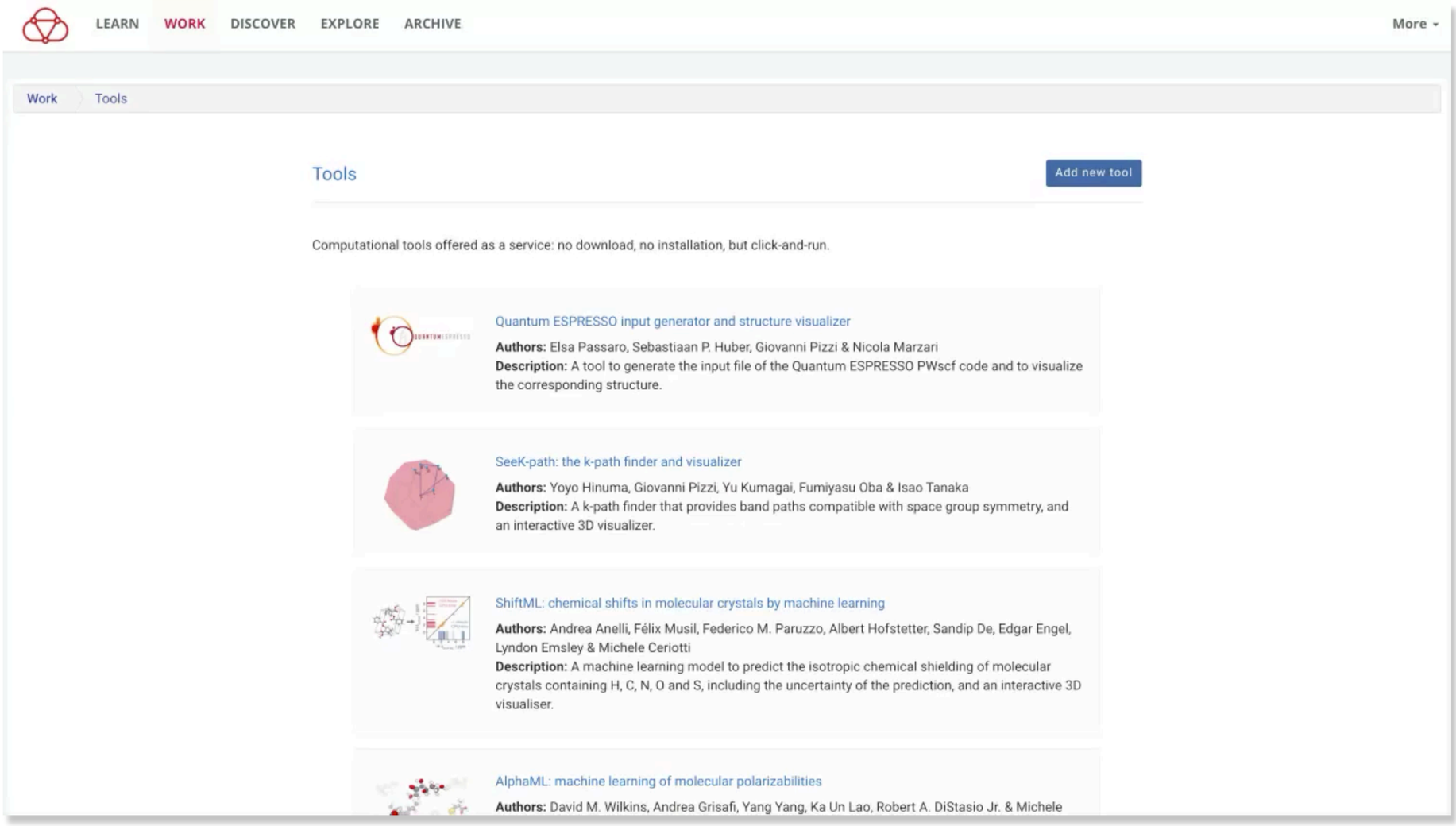


Tools

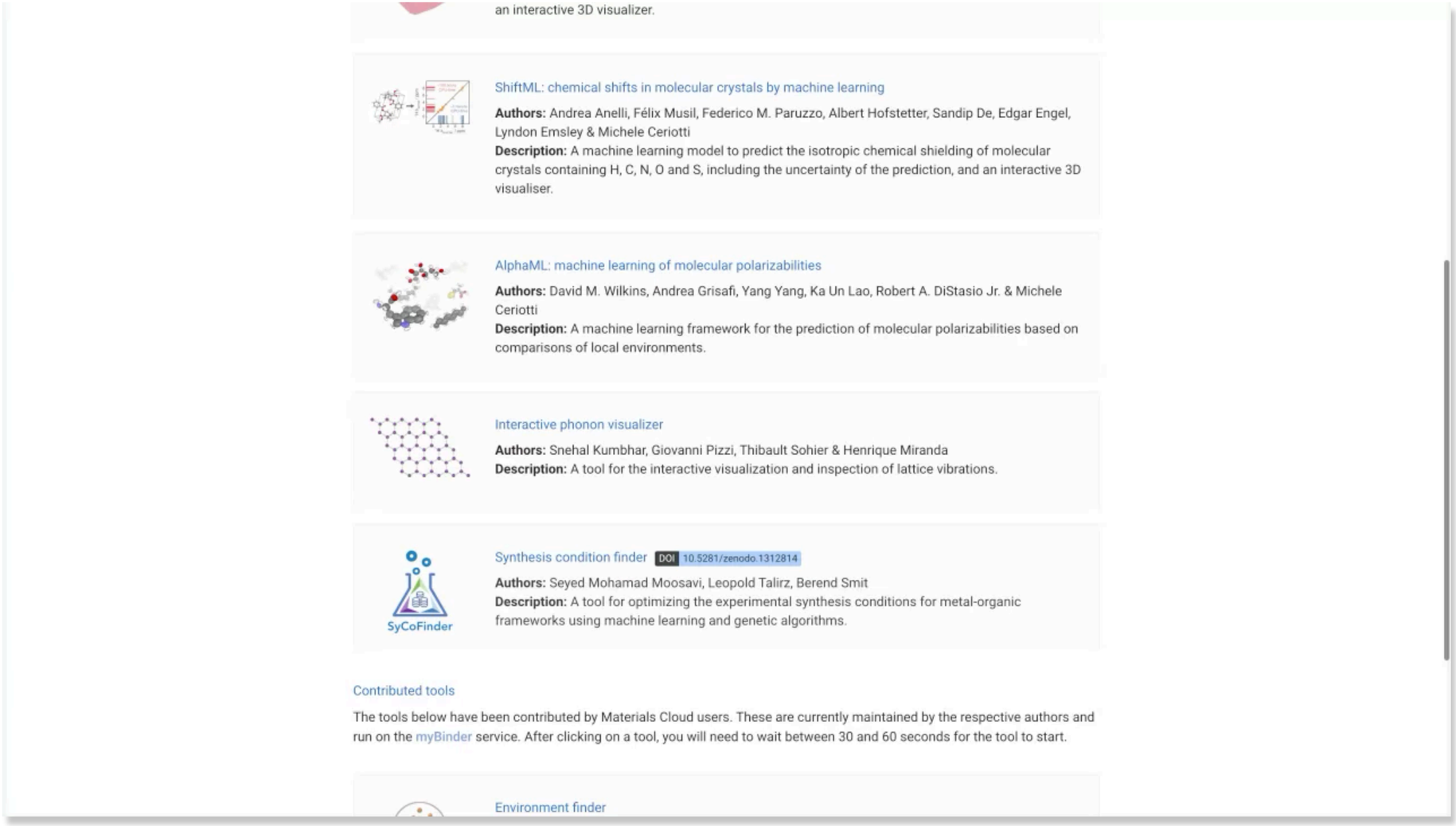
- Simple web-based applications that allow to run simulations within a few seconds (~<10-60 sec)
- Useful to provide interactive visualisation and fast data processing on the web

Giovanni Pizzi

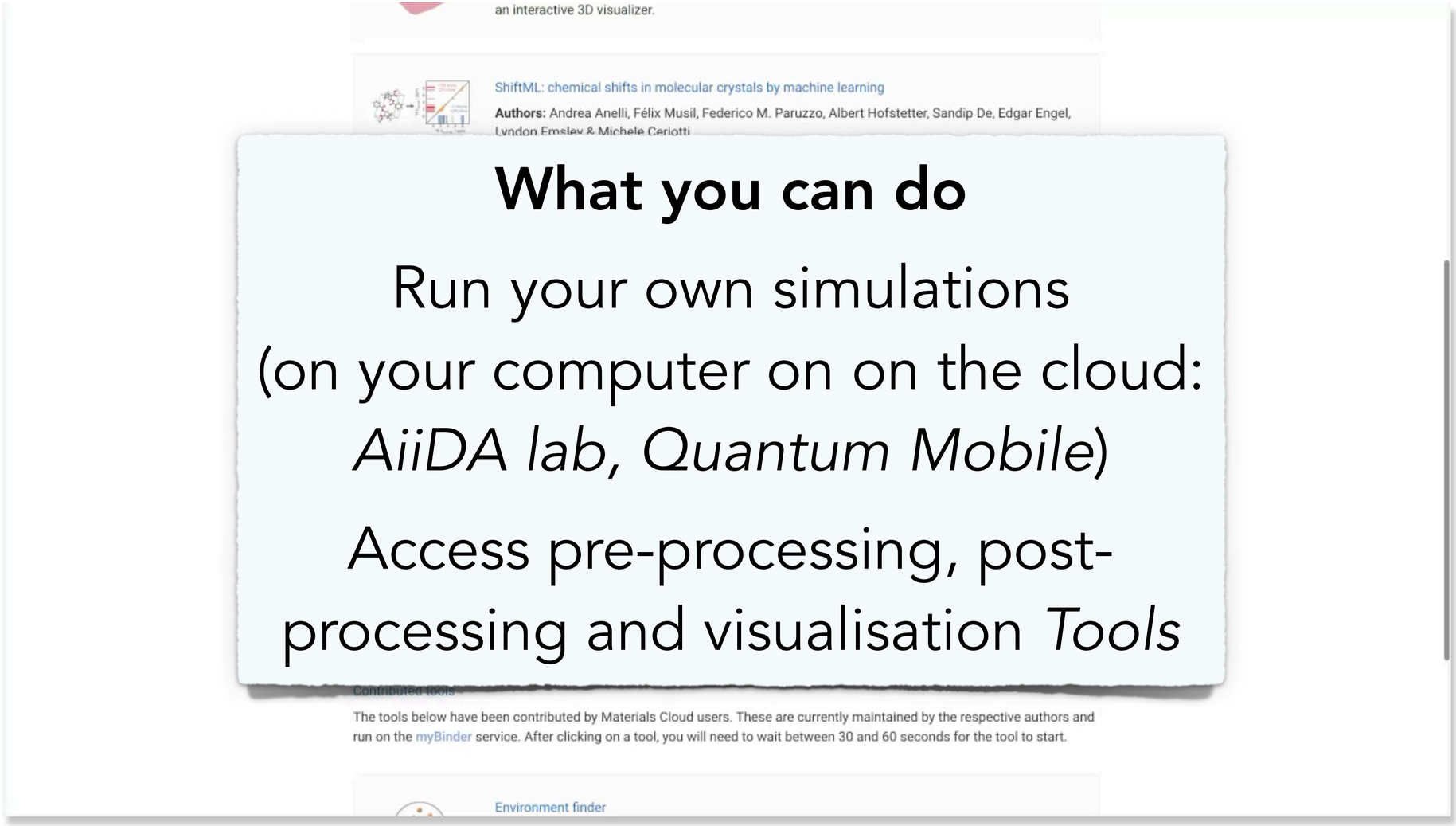
WORK Tools example: seekpath



WORK Tools example: phonon visualizer



WORK Tools example: phonon visualizer



FAIR data sharing: **ARCHIVE**, DISCOVER, EXPLORE

SCIENTIFIC DATA · re3data.org Materials Cloud Archive · FAIRsharing.org · GO FAIR · B2FIND · EUDAT · Google

materialscloud:2017.0008/v3

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

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano E. Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1], 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, Sauletekio 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	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

FAIR data sharing: **ARCHIVE**, DISCOVER, EXPLORE

DOIs assigned

SCIENTIFIC DATA · re3data.org · FAIRsharing.org · GO FAIR · B2FIND · EUDAT · Google

materialscloud:2017.0008/v3

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

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano E. Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1], 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, Sauletekio 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	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

FAIR data sharing: **ARCHIVE**, DISCOVER, EXPLORE

DOIs assigned

SCIENTIFIC DATA · re3data.org · FAIRsharing.org standards, databases, policies · GO FAIR · B2FIND · EUDAT · Google

materialscloud:2017.0008/v3

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

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano E. Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1], 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, Sauletekio 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	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

FAIRsharing.org
re3data.org
b2find.eudat.eu

+

Recommended data repository by Nature's journal *Scientific Data*

+

GO FAIR implementation network

+

Indexed by Google Dataset Search

FAIR data sharing: **ARCHIVE**, DISCOVER, EXPLORE

DOIs assigned

Direct links to DISCOVER & EXPLORE

materialscloud:2017.0008/v3

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

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano E. Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1], 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, Sauletekio 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	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

FAIRsharing.org
re3data.org
b2find.eudat.eu

+

Recommended data repository by Nature's journal *Scientific Data*

+

GO FAIR implementation network

+

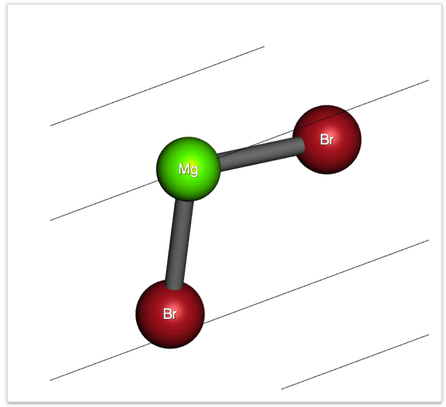
Indexed by Google Dataset Search

DISCOVER (curated data) & EXPLORE (raw data)

Giovanni Pizzi

DISCOVER

Compound: $MgBr_2$



Info and properties

See definitions...

Formula: $MgBr_2$
Spacegroup: P-3m1
Pointgroup: -3m
Prototype: CdI2
Band gap [eV]: 4.8

Magnetic properties:

Magnetic State: non-magnetic
Tot. Magnetization [μB/cell]: -
Abs. Magnetization [μB/cell]: -

Binding Energies:

DF2-C09 Binding energy [meV/Å^2]: 10.2
(From parent COD 9009107)
rVV10 Binding energy [meV/Å^2]: 15.3
(From parent COD 9009107)

Delta in interlayer distance (vdW vs revPBE):

Δ_{DF2} [%]: 17.1 (From parent COD 9009107)
Δ_{rVV10} [%]: 18.3 (From parent COD 9009107)

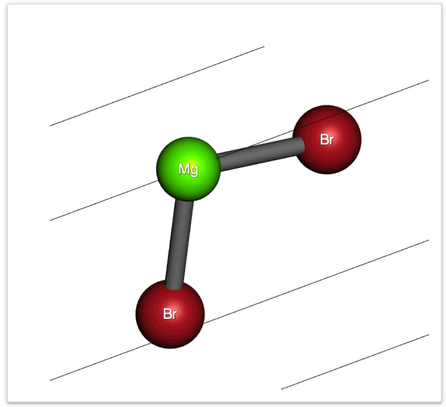
Band structure

DISCOVER (curated data) & EXPLORE (raw data)

Giovanni Pizzi

DISCOVER

Compound: $MgBr_2$



Info and properties

See definitions...

Formula: $MgBr_2$
Spacegroup: P-3m1
Pointgroup: -3m
Prototype: CdI2
Band gap [eV]: 4.8

Magnetic properties:

- Magnetic State: non-magnetic
- Tot. Magnetization [μB/cell]: -
- Abs. Magnetization [μB/cell]: -

Binding Energies:

- DF2-C09 Binding energy [meV/Å^2]: 10.2 (From parent COD 9009107)
- rVV10 Binding energy [meV/Å^2]: 15.3 (From parent COD 9009107)

Delta in interlayer distance (vdW vs revPBE):

- Δ_{DF2} [%]: 17.1 (From parent COD 9009107)
- Δ_{rVV10} [%]: 18.3 (From parent COD 9009107)

Band structure

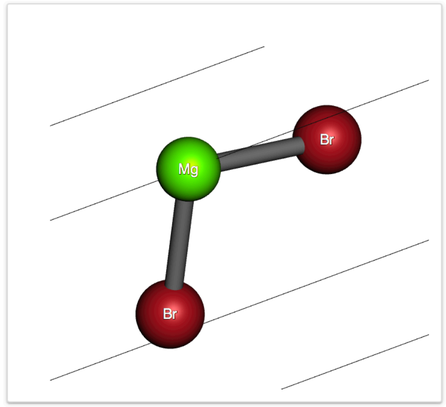
UUID links to jump to the provenance graph in the EXPLORE section

DISCOVER (curated data) & **EXPLORE** (raw data)

Giovanni Pizzi

DISCOVER

Compound: $MgBr_2$



Info and properties

See definitions...

Formula: $MgBr_2$
Spacegroup: P-3m1
Pointgroup: -3m
Prototype: CdI2
Band gap [eV]: 4.8

Magnetic properties:

Magnetic State: non-magnetic
Tot. Magnetization [μB/cell]: -
Abs. Magnetization [μB/cell]: -

Binding Energies:

DF2-C09 Binding energy [meV/Å^2]: 10.2
(From parent COD 9009107)
rVV10 Binding energy [meV/Å^2]: 15.3
(From parent COD 9009107)

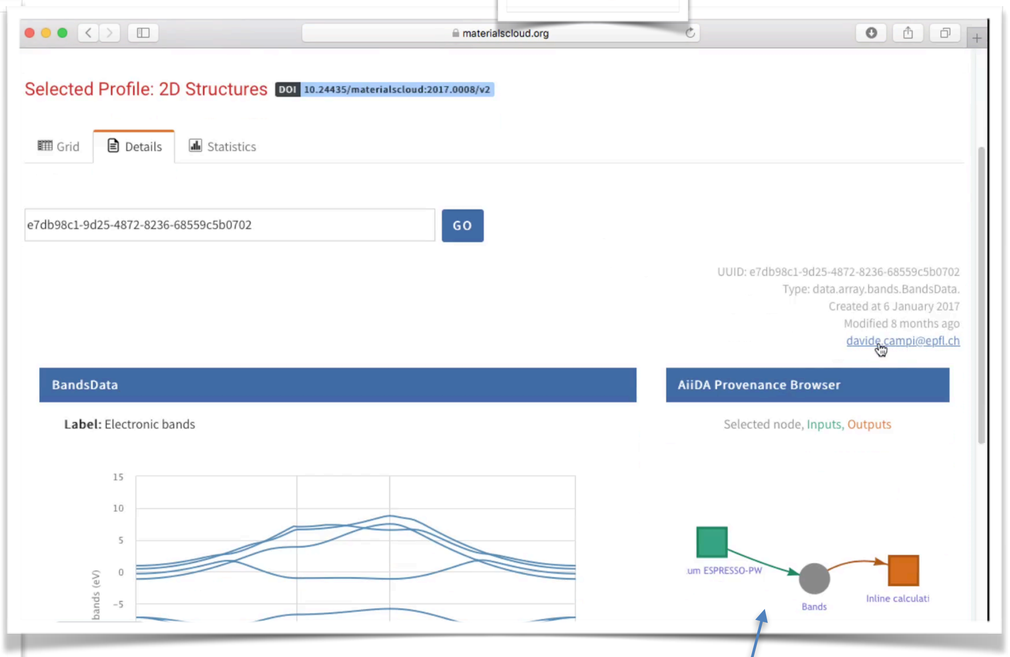
Delta in interlayer distance (vdW vs revPBE):

Δ_{DF2} [%]: 17.1 (From parent COD 9009107)
Δ_{rVV10} [%]: 18.3 (From parent COD 9009107)

Band structure

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

EXPLORE



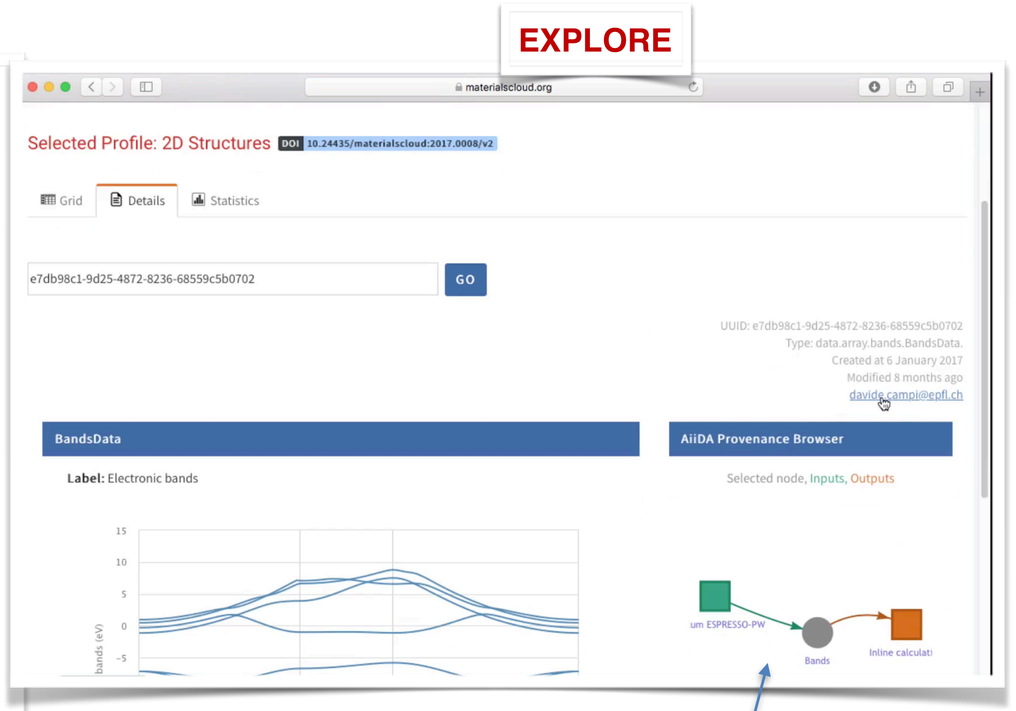
Browse the full AiiDA provenance graph (inputs, outputs, …) at any level

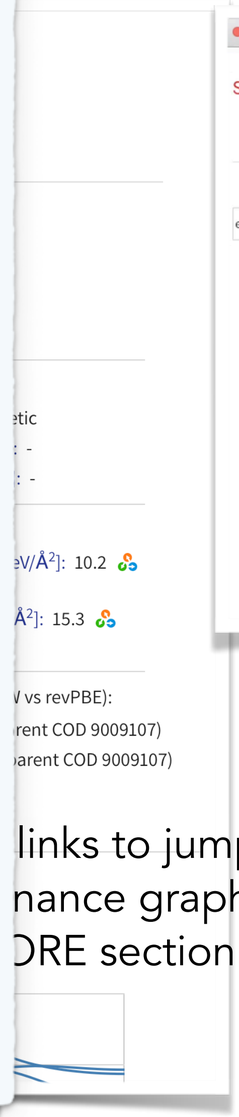
DISCOVER (curated data) & **EXPLORE** (raw data)

What you can do

ACCESS

Freely access hundreds of datasets of other researchers:
(input and output files,
full provenance, …)
and reproduce their research





links to jump to the
nance graph in the
ORE section

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

DISCOVER (curated data) & **EXPLORE** (raw data)

Giovanni Pizzi

What you can do

ACCESS

Freely access hundreds of datasets of other researchers:
(input and output files,
full provenance, …)
and reproduce their research

What you can do

CONTRIBUTE AND SHARE

Freely upload your data to the ARCHIVE (with or without AiiDA), get a DOI with availability guarantee of 10+ years

Make your research easy reproduce by others, share your curated results, make your AiiDA graphs easy to browse

Data Management Plans (DMPs) 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, reproduce in the AiiDA Lab thanks to *full provenance*

Data Management Plans (DMPs) and FAIR principles

Giovanni Pizzi

- 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, reproduce in the AiiDA Lab thanks to *full provenance*

- We provide DMP templates for researchers using Materials Cloud

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

Data Management Plans (DMPs) and FAIR principles

- Combination of **AiiDA + Materials Cloud** (Discover, E
  FAIR-compliant sharing

- **Findable**: DOIs with standardized metadata
- **Accessible**: web interface to browse data, calculations and provenance, curated data in Discover section

- **Interope**
  directed
  betweer
- **Reusabl**
  open (C
  thanks t

- We provide DMP templates for researchers u

Funding Body	DMP template (using AiiDA)
SNF	.docx .odt .pdf
H2020	.docx .odt .pdf

SNF DMP Template

This is a data management plan (DMP) template for a SNF project in computational materials science using the Materials Cloud Archive as a data repository and AiiDA to collect the full provenance of the calculations performed.

Further SNF guidelines on DMPs can be found here.

Please adapt the template to your project!

1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the FAIR Data Principles F2, I3, R1 & R1.2.)

The *ab initio* calculations performed in this project are expected to generate large volumes of data, most of it is in the form of intermediate binary files, which can be straightforwardly regenerated from the original inputs.

In this project, calculations will be performed through the *Automated Interactive Infrastructure and Database for Computational Science* (AiiDA), a python framework for high throughput calculations and provenance tracking (www.aiida.net). AiiDA automatically stores all information required to reproduce the result of each calculation. On the output side, AiiDA is designed to strike a balance between the cost of storage and the cost of recomputing a piece of data. For example, by default, total energies, electronic band structures and log files are stored, while Kohn-Sham wave functions and charge densities are not.

1.2 How will the data be collected, observed or generated?

Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures. (This relates to the FAIR Data Principle R1.)

Before submitting a calculation through AiiDA, all inputs (+ metadata, see 1.3) are automatically stored in a local database. AiiDA then transmits the necessary information to the target computer, which can be a remote supercomputer, a local cluster or the workstation of the researcher. AiiDA adds the calculation to the computer's job queue, monitors the status, and retrieves the results once the calculation is finished.

In this model, all data is generated on the target computer, while only the data intended for preservation is transferred back to the workstation of the researcher.

AiiDA records this data in the form of directed acyclic graphs. For a full specification of the data (+ metadata) stored, see the documentation of the ORM API: https://aiida-core.readthedocs.io/en/stable/apidoc/aiida.orm.html

1.3 What documentation and metadata will you provide with the data?

Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data.

Data Management Plans (DMPs) and FAIR principles

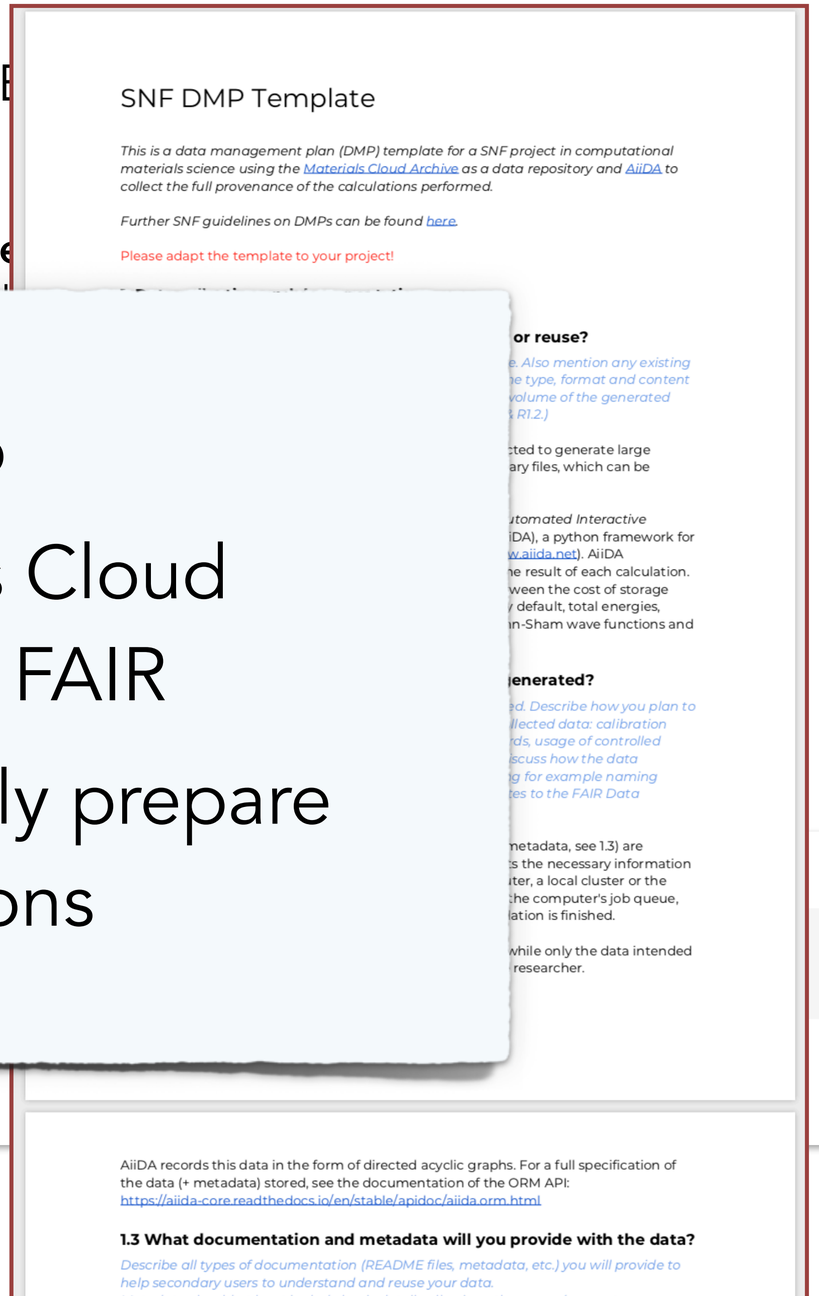
- Combination of **AiiDA + Materials Cloud** (Discover, E
  FAIR-compliant sharing

- **Findable**: DOIs with standardized metadata
- **Acces**
  calcul
  Disco
- **Interope**
- W

What you can do

Use AiiDA and Materials Cloud
to make your research FAIR

Get DMP templates to easily prepare
your grant applications


SNF DMP Template

This is a data management plan (DMP) template for a SNF project in computational materials science using the Materials Cloud Archive as a data repository and AiiDA to collect the full provenance of the calculations performed.

Further SNF guidelines on DMPs can be found here.

Please adapt the template to your project!

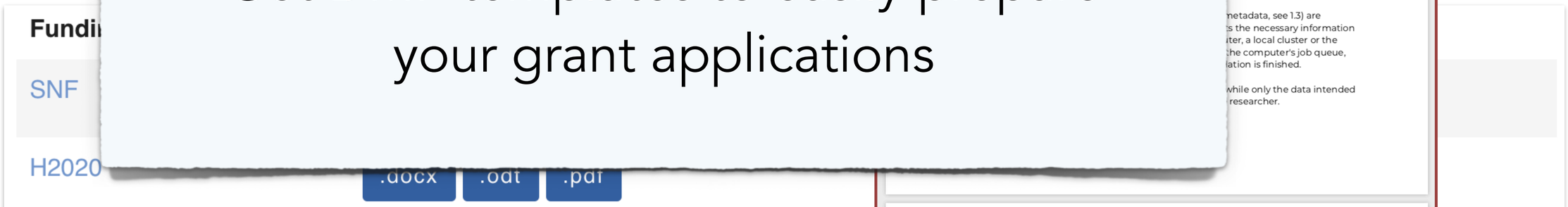
or reuse?

enerated?

AiiDA records this data in the form of directed acyclic graphs. For a full specification of the data (+ metadata) stored, see the documentation of the ORM API:
https://aiida-core.readthedocs.io/en/stable/apidoc/aiida.orm.html

1.3 What documentation and metadata will you provide with the data?

Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data.



ARCHIVE - FAQ

Giovanni Pizzi

- Q: **How much data can I upload per record?**

  A: *Everyone:* 5GB regular, 50GB AiiDA.

  Sponsor & partner projects: for larger data sets just contact us

- Q: **How long will it be stored?**

  A: Guaranteed at least 10 years after submission.

- Q: **Can I update the content of my record?**

  A: Yes, but only references & keyword (otherwise: create new version)

- Q: **How to make a new version?**

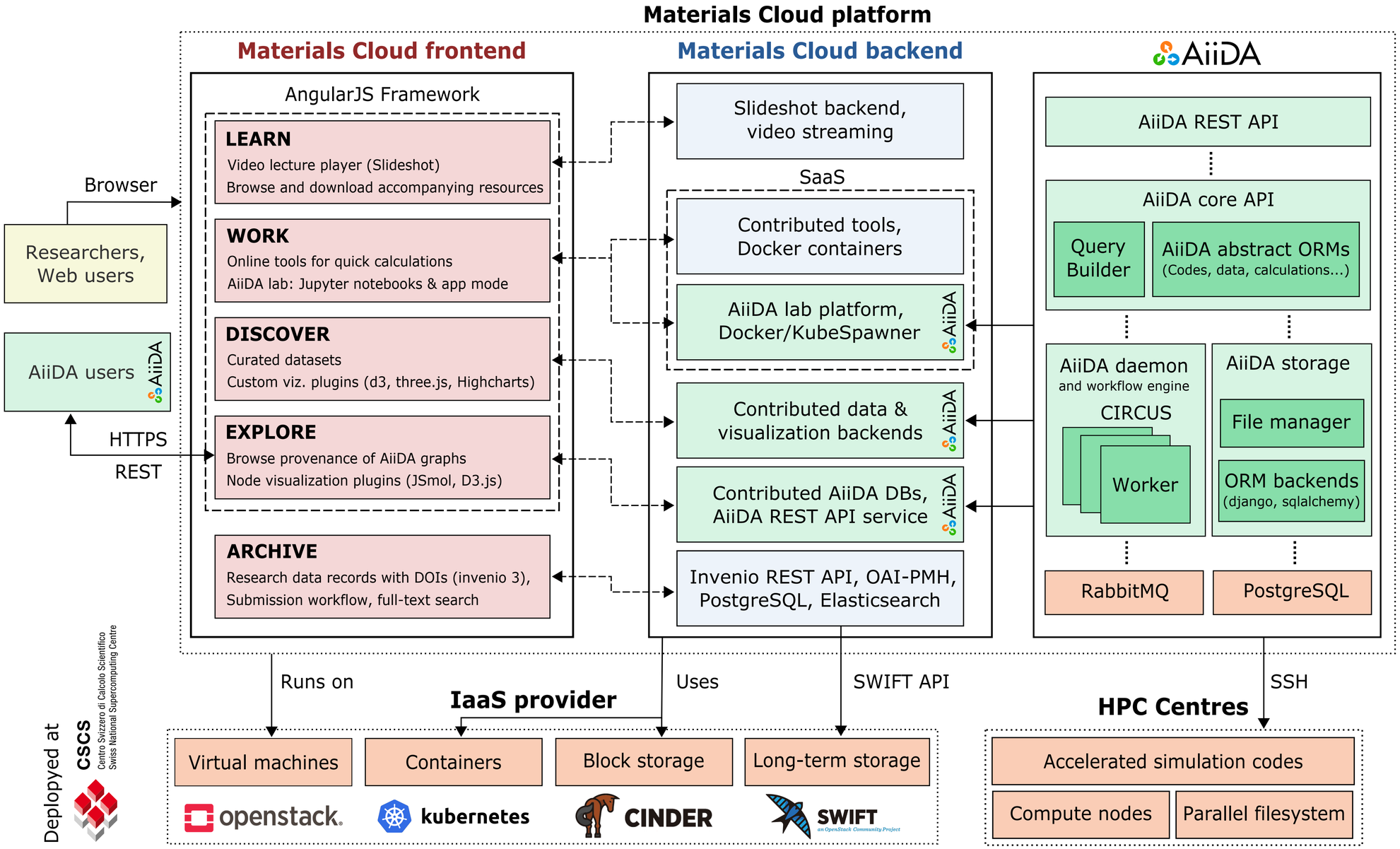
  A: Edit existing submission & request publication again

- Q: **Can I *reserve* a DOI before publication?**

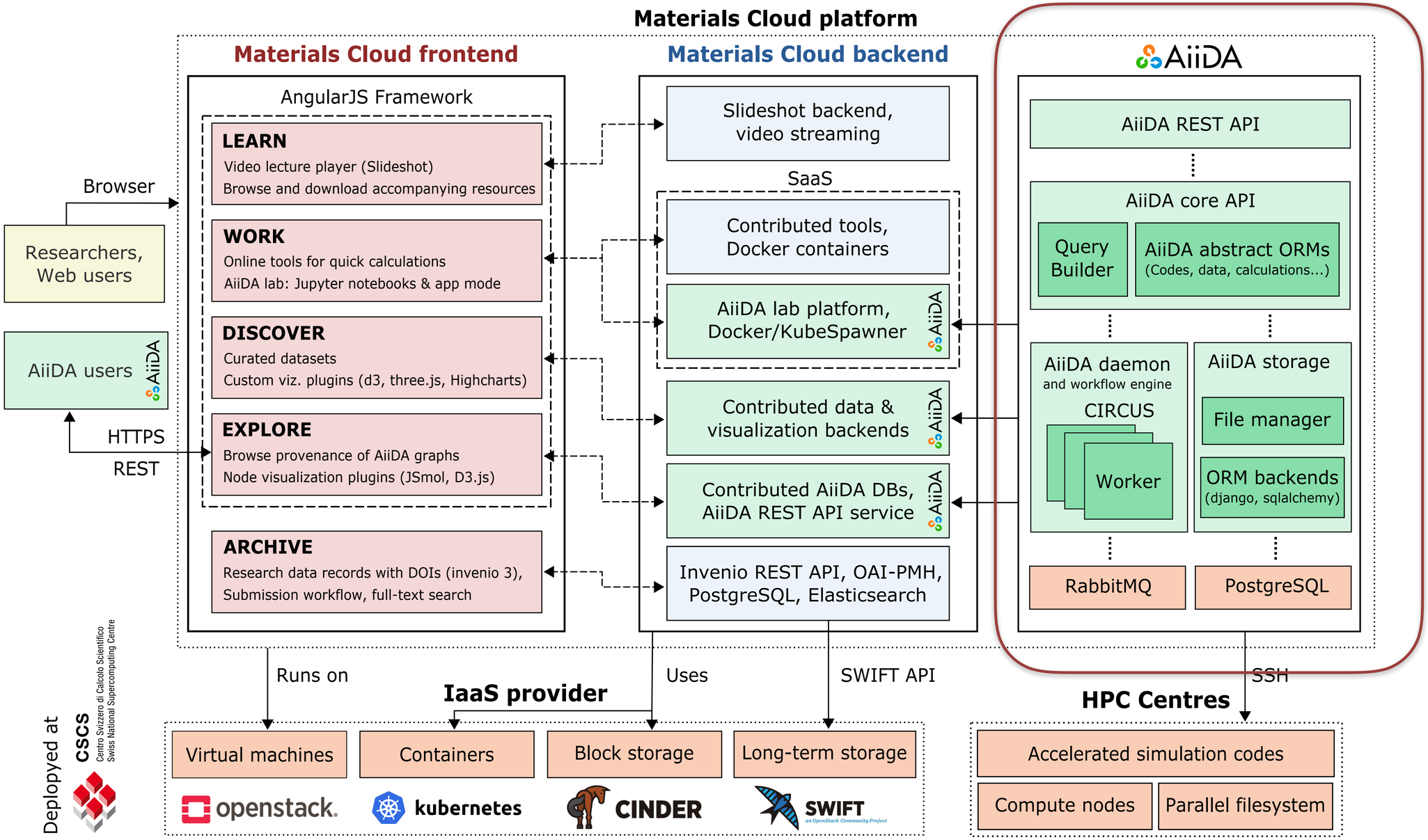
  A: Yes. It will resolve after (moderation and) publication

materialscloud.org/policies

Full platform architecture



Full platform architecture



Future plans

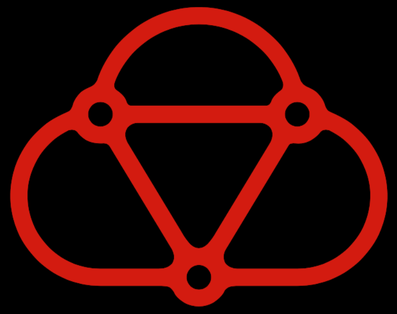
- **Automate and streamline creation of tools/discover contributed sections**
  - Identify suitable technology (or technologies) for app development (tools-barebone, jupyter+voila, plotly dash, bokeh panel, …)
  - Facilitate testing and deployment without interaction with administrators (dokku, …)

- **Automate creation of Explore sections for submitted AiiDA databases**
  - Simplify and speed up integration of new contributed AiiDA databases without the need of an administrator

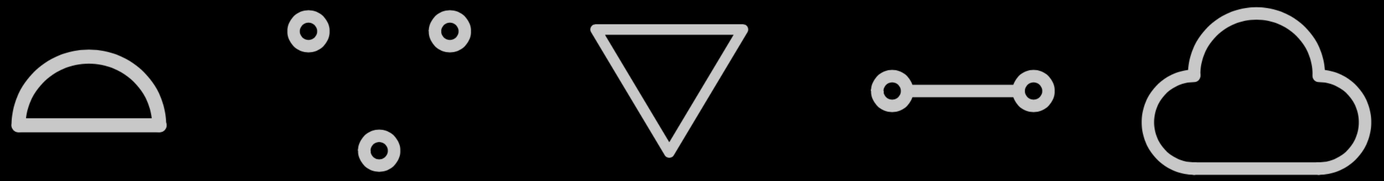
Funding

- **SNSF MARVEL NCCR** (I: 2014-18, II:2018-22, EPFL Lead house) for the Open Science Platform
- **H2020 MaX Centre of Excellence** (I: 2015-18, II: 2018-21) for convergence of HPC, HTC and HPDA via AiiDA
- **H2020 MarketPlace** (2018-22, EPFL co-PI) for providing materials simulation services and data
- **H2020 Intersect** (2019-21, EU; CNR lead PI, EPFL and ICN2 co-PI) for automated modelling of complex devices via AiiDA
- Private collaboration with a **major European company** (2019-2020) for AiiDA-powered materials discovery for Li-ion batteries
- **Swissuniversities P-5 Materials Cloud** (2019-20) for transitioning Materials Cloud to self-sustaining service
- **EPFL Open Science Fund "OSSCAR"** (2019-21) for creating an educational hub for research and teaching

15+ FTEs for the next years



MATERIALSCLOUD



LEARN WORK DISCOVER EXPLORE ARCHIVE

https://www.materialscloud.org