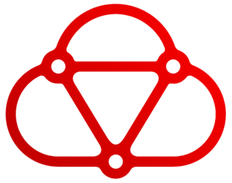
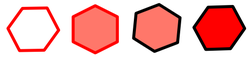
Open science and open access in publications and data

Using Materials Cloud Archive to deposit research data

Giovanni Pizzi
THEOS, EPFL and MARVEL (OSP)

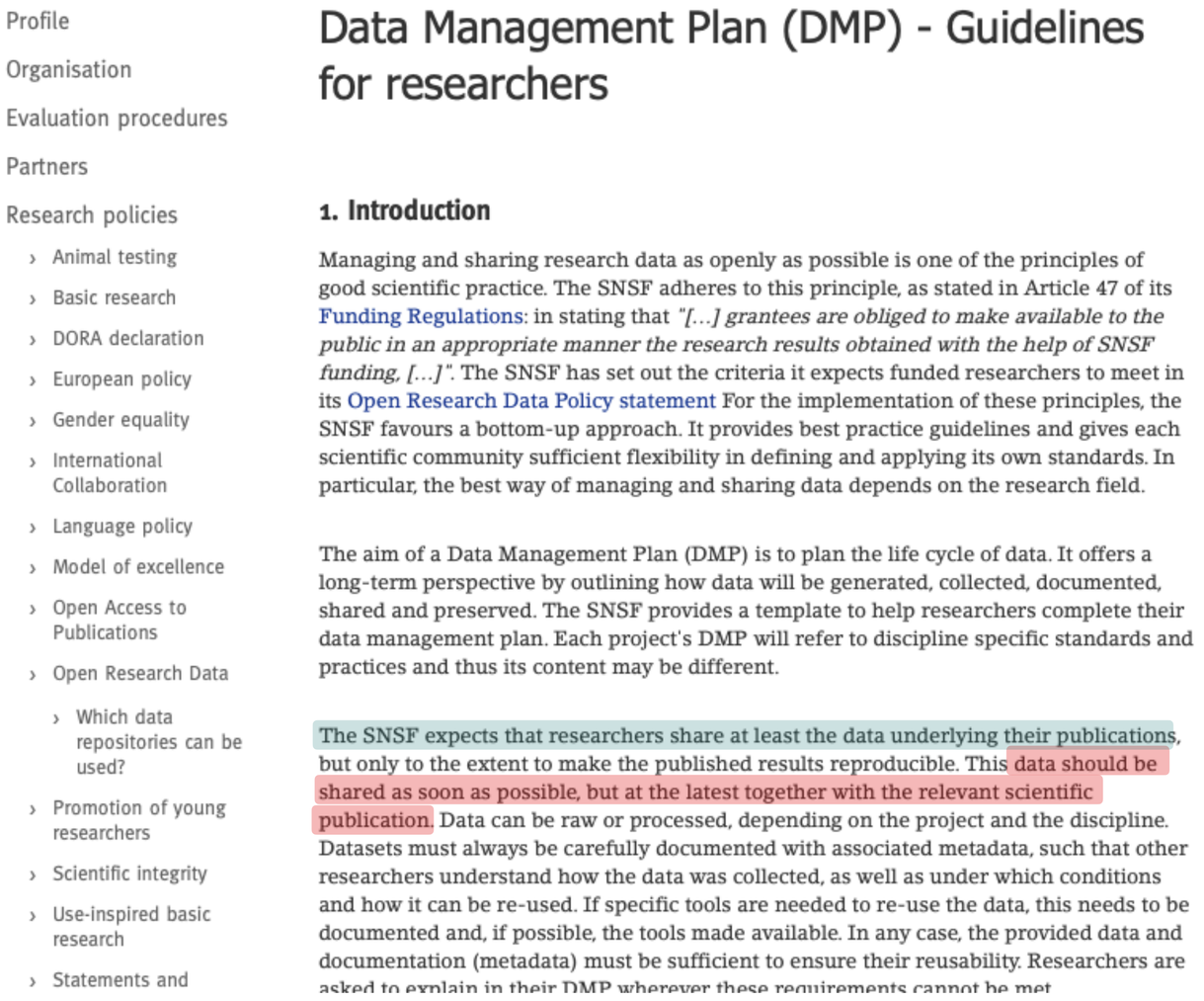


MATERIALSCLOUD



SNSF requirements on Open Research Data

- SNSF **requires that all data** needed to reproduce funded publications must be made **openly available at the latest by the time of publication**



Profile
Organisation
Evaluation procedures
Partners
Research policies
› Animal testing
› Basic research
› DORA declaration
› European policy
› Gender equality
› International Collaboration
› Language policy
› Model of excellence
› Open Access to Publications
› Open Research Data
› Which data repositories can be used?
› Promotion of young researchers
› Scientific integrity
› Use-inspired basic research
› Statements and

Data Management Plan (DMP) - Guidelines for researchers

1. Introduction

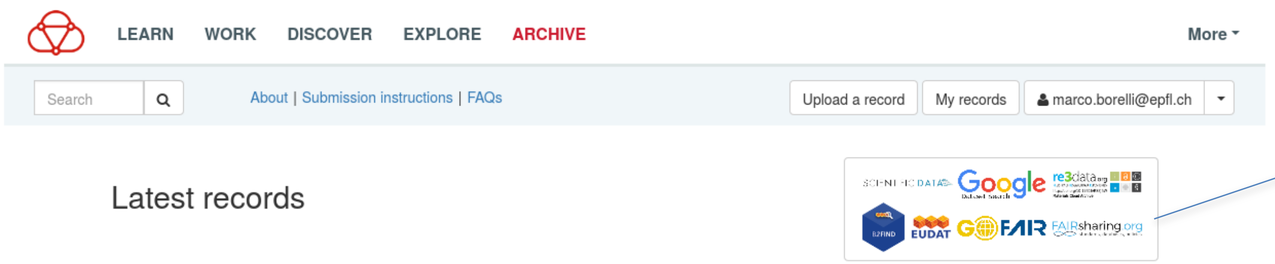
Managing and sharing research data as openly as possible is one of the principles of good scientific practice. The SNSF adheres to this principle, as stated in Article 47 of its Funding Regulations: in stating that *"[…] grantees are obliged to make available to the public in an appropriate manner the research results obtained with the help of SNSF funding, […]"*. The SNSF has set out the criteria it expects funded researchers to meet in its Open Research Data Policy statement For the implementation of these principles, the SNSF favours a bottom-up approach. It provides best practice guidelines and gives each scientific community sufficient flexibility in defining and applying its own standards. In particular, the best way of managing and sharing data depends on the research field.

The aim of a Data Management Plan (DMP) is to plan the life cycle of data. It offers a long-term perspective by outlining how data will be generated, collected, documented, shared and preserved. The SNSF provides a template to help researchers complete their data management plan. Each project's DMP will refer to discipline specific standards and practices and thus its content may be different.

The SNSF expects that researchers share at least the data underlying their publications, but only to the extent to make the published results reproducible. This data should be shared as soon as possible, but at the latest together with the relevant scientific publication. Data can be raw or processed, depending on the project and the discipline. Datasets must always be carefully documented with associated metadata, such that other researchers understand how the data was collected, as well as under which conditions and how it can be re-used. If specific tools are needed to re-use the data, this needs to be documented and, if possible, the tools made available. In any case, the provided data and documentation (metadata) must be sufficient to ensure their reusability. Researchers are asked to explain in their DMP wherever these requirements cannot be met

http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx

Materials Cloud Archive: https://archive.materialscloud.org



Recommended data repository by Nature's journal Scientific Data

Indexed by Google Dataset Search and by EUDAT/EOSC's B2FIND

Registered on FAIRsharing.org and re3data.org

New! Recommended by the new "Open Research Europe" journal

European Commission

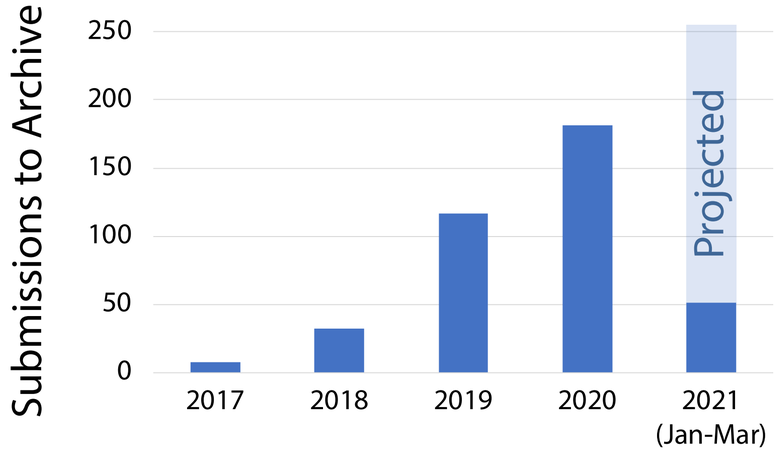
Research and Innovation

Open Research Europe

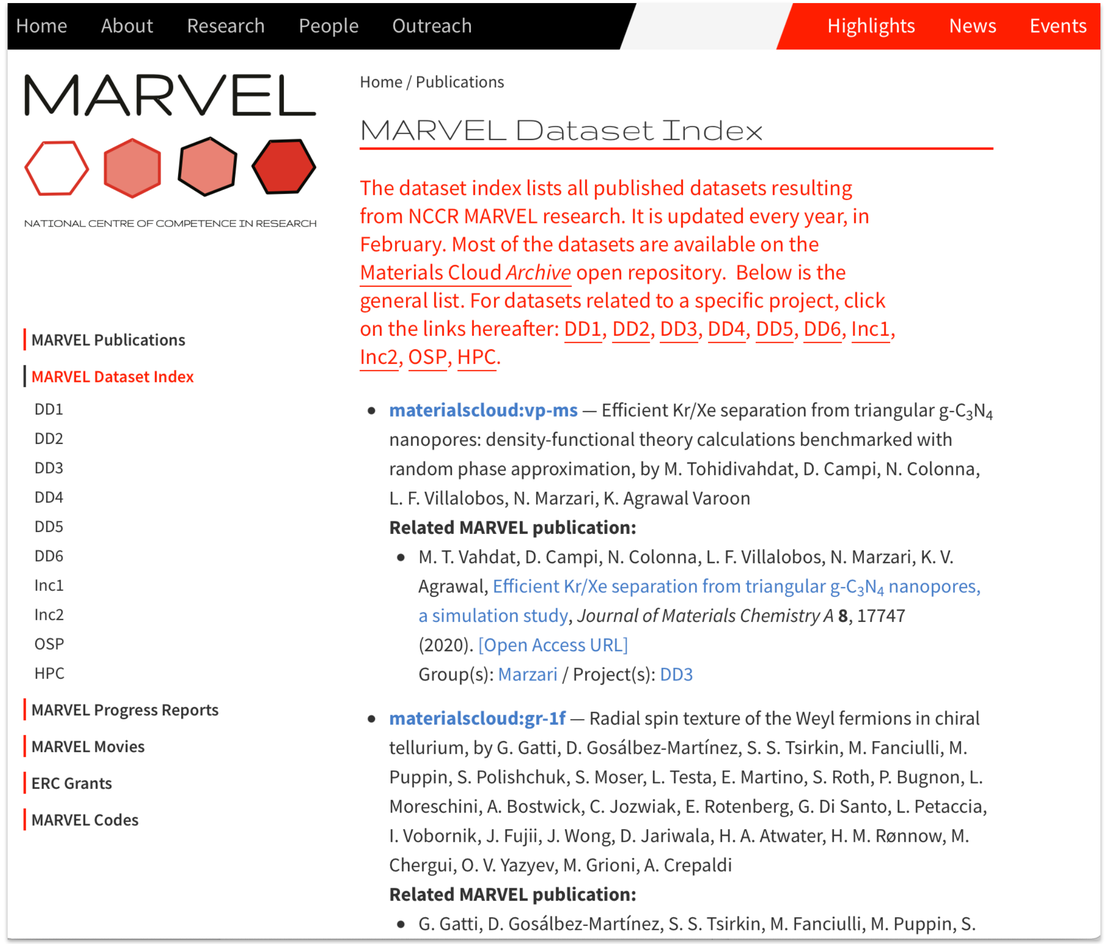
https://open-research-europe.ec.europa.eu/for-authors/data-guidelines

Materials Cloud Archive

- Host MARVEL's data (but also open to the world)
- Limit: 5GB (50GB if using AiiDA); but limits lifted within MARVEL
-



- *Statistics (March 2022)*: 600GB files hosted: 22'000'000+ crystal structures, 7'500'000+ DFT structures, 670'000+ reproducible calculations, ...
- **~13'000 file downloads/month, ~5-10TB/month downloaded (in the past year)**

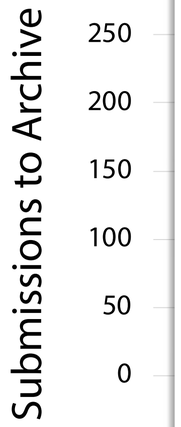


MARVEL Dataset Index automatically generated

https://nccr-marvel.ch/publications/dataset-index

Materials Cloud Archive

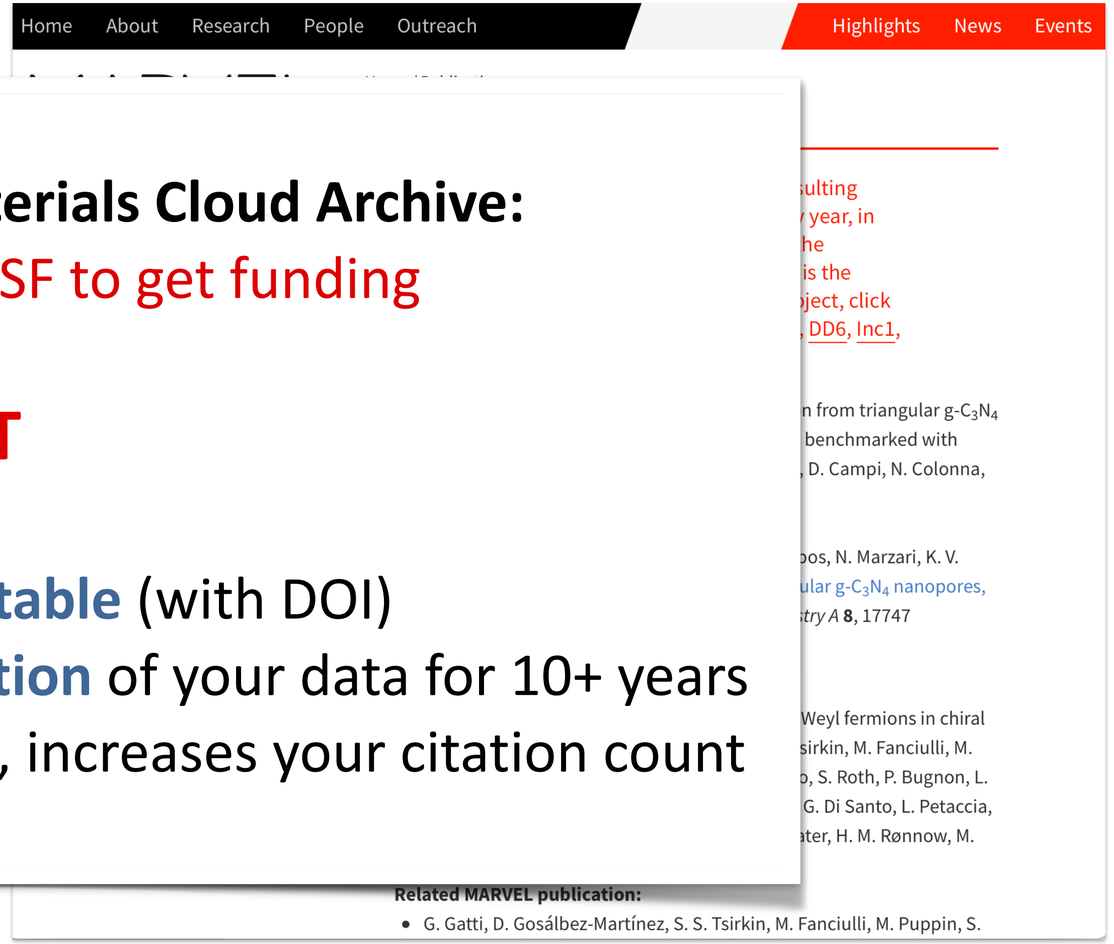
- Host MARVEL's data (but also open to the world)
- Limit: 5GB
but limits
-
- *Statistics (Mar*
22'000'000+ crystal structures, 7'500'000+ DFT structures,
670'000+ reproducible calculations, ...
- **~13'000 file downloads/month,**
~5-10TB/month downloaded (in the past year)



Hosting files on the Materials Cloud Archive:
a requirement of SNSF to get funding

BUT

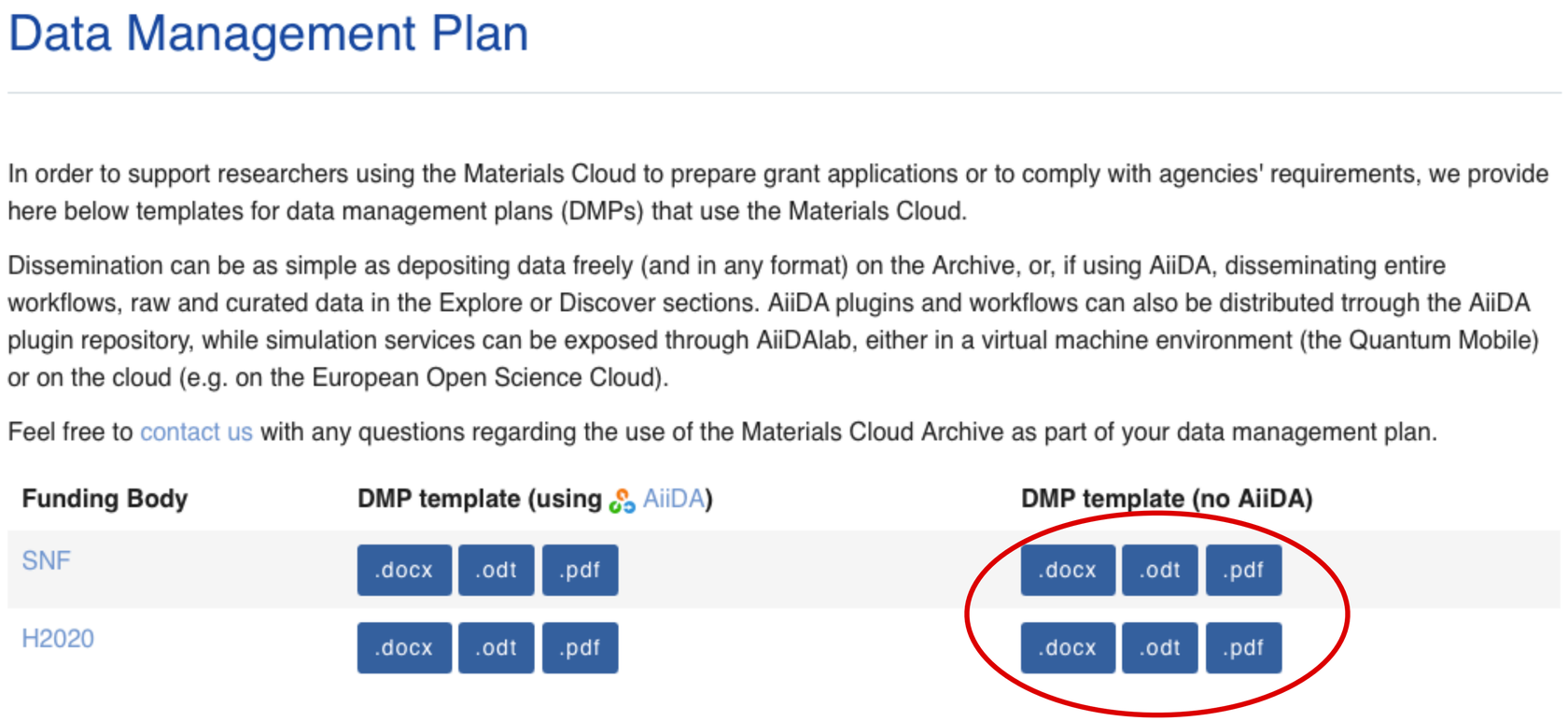
Makes your data **citable** (with DOI)
Takes care of **long-term preservation** of your data for 10+ years
Boosts visibility of your research, increases your citation count



MARVEL Dataset Index automatically generated
https://nccr-marvel.ch/publications/dataset-index

DMP templates

- In Materials Cloud we provide templates for your DMPs, if you use the Materials Cloud Archive (that guarantees long-term storage for 10+ years)



Data Management Plan

In order to support researchers using the Materials Cloud to prepare grant applications or to comply with agencies' requirements, we provide here below templates for data management plans (DMPs) that use the Materials Cloud.

Dissemination can be as simple as depositing data freely (and in any format) on the Archive, or, if using AiiDA, disseminating entire workflows, raw and curated data in the Explore or Discover sections. AiiDA plugins and workflows can also be distributed trrough the AiiDA plugin repository, while simulation services can be exposed through AiiDAlab, either in a virtual machine environment (the Quantum Mobile) or on the cloud (e.g. on the European Open Science Cloud).

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

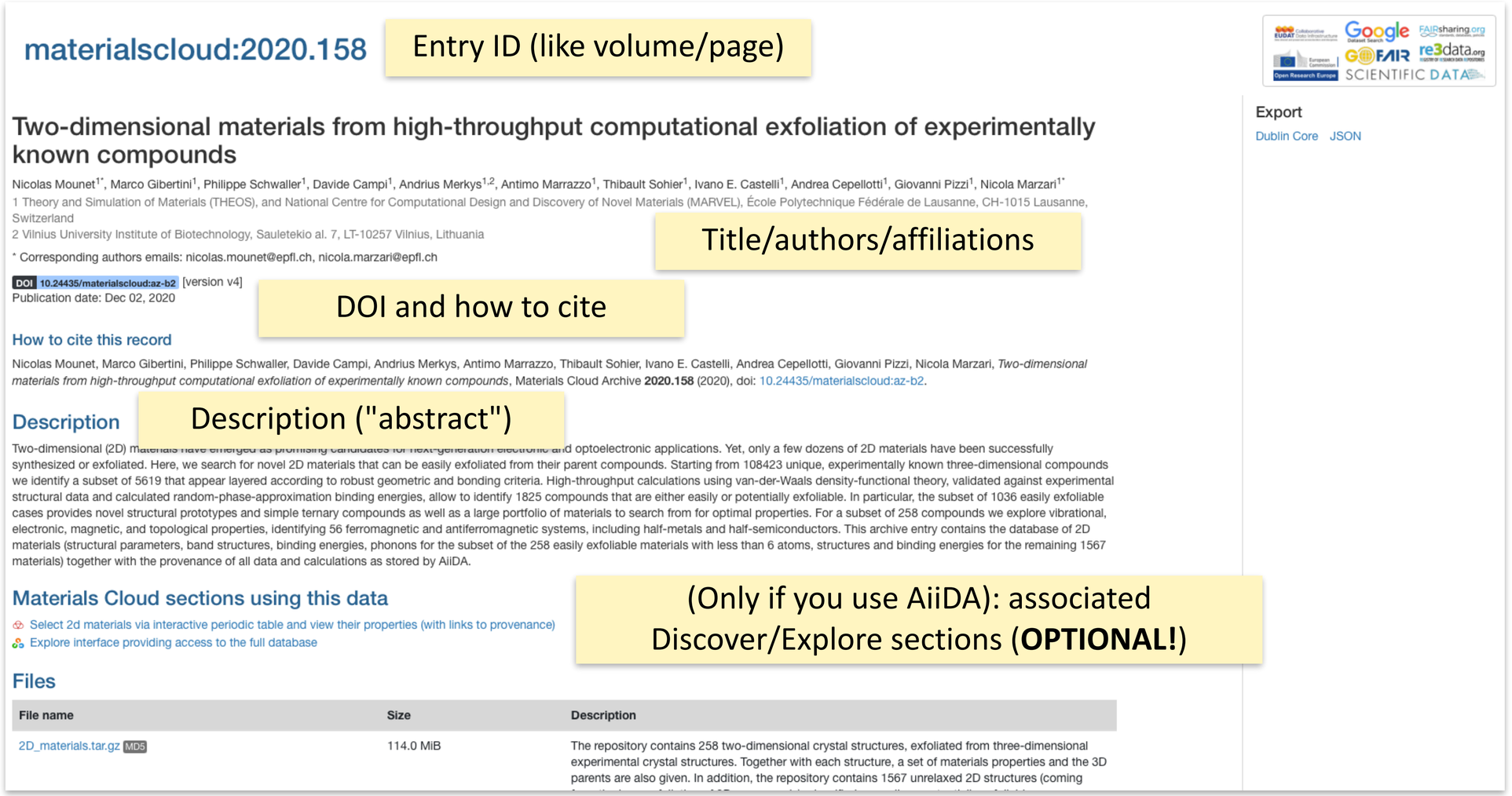
Funding Body	DMP template (using AiiDA)	DMP template (no AiiDA)
SNF	.docx .odt .pdf	.docx .odt .pdf
H2020	.docx .odt .pdf	.docx .odt .pdf

Please also note the resources provided by EPFL, including extensive DMP templates for many different project types (SNSF, ERC, H2020, ...).

https://www.materialscloud.org/dmp

A quick-start guide for the Materials Cloud Archive

An example entry: https://doi.org/10.24435/materialscloud:az-b2



materialscloud:2020.158

Entry ID (like volume/page)

Export

Dublin Core JSON

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

Title/authors/affiliations

DOI 10.24435/materialscloud:az-b2 [version v4]

Publication date: Dec 02, 2020

DOI and how to cite

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 **2020.158** (2020), doi: 10.24435/materialscloud:az-b2.

Description

Description ("abstract")

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)

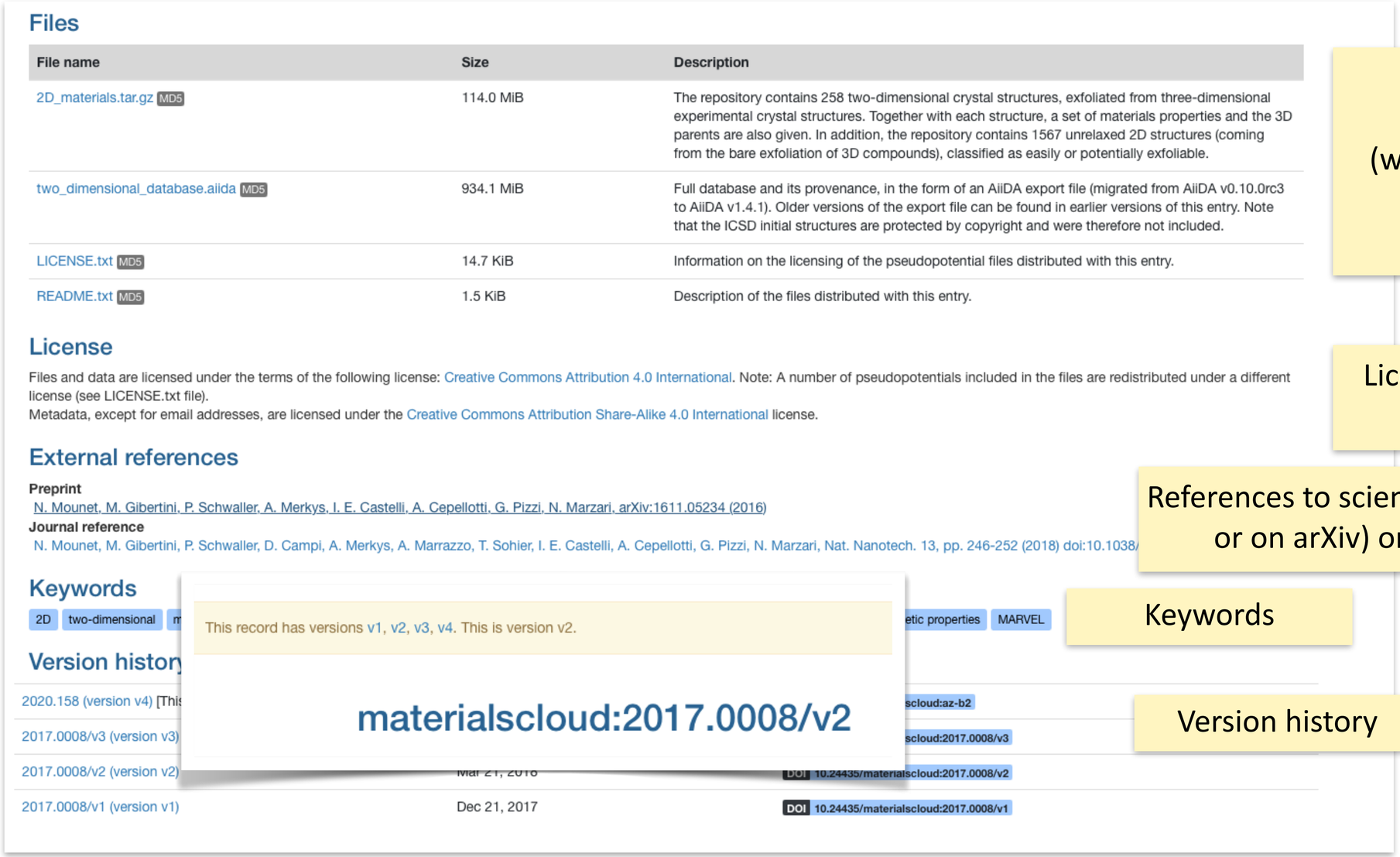
Explore interface providing access to the full database

(Only if you use AiiDA): associated Discover/Explore sections (**OPTIONAL!**)

Files

File name	Size	Description
2D_materials.tar.gz MD5	114.0 MiB	The repository contains 258 two-dimensional crystal structures, exfoliated from three-dimensional experimental crystal structures. Together with each structure, a set of materials properties and the 3D parents are also given. In addition, the repository contains 1567 unrelaxed 2D structures (coming

An example entry (continued)



Files

File name	Size	Description
2D_materials.tar.gz MD5	114.0 MiB	The repository contains 258 two-dimensional crystal structures, exfoliated from three-dimensional experimental crystal structures. Together with each structure, a set of materials properties and the 3D parents are also given. In addition, the repository contains 1567 unrelaxed 2D structures (coming from the bare exfoliation of 3D compounds), classified as easily or potentially exfoliable.
two_dimensional_database.aiida MD5	934.1 MiB	Full database and its provenance, in the form of an AiiDA export file (migrated from AiiDA v0.10.0rc3 to AiiDA v1.4.1). Older versions of the export file can be found in earlier versions of this entry. Note that the ICSD initial structures are protected by copyright and were therefore not included.
LICENSE.txt MD5	14.7 KiB	Information on the licensing of the pseudopotential files distributed with this entry.
README.txt MD5	1.5 KiB	Description of the files distributed with this entry.

License

Files and data are licensed under the terms of the following license: Creative Commons Attribution 4.0 International. Note: A number of pseudopotentials included in the files are redistributed under a different license (see LICENSE.txt file).
Metadata, except for email addresses, are licensed under the Creative Commons Attribution Share-Alike 4.0 International license.

External references

Preprint
N. Mounet, M. Gibertini, P. Schwaller, A. Merkys, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, arXiv:1611.05234 (2016)
Journal reference
N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, Nat. Nanotech. 13, pp. 246-252 (2018) doi:10.1038/

Keywords

2D | two-dimensional | ... | etic properties | MARVEL

Version history

This record has versions v1, v2, v3, v4. This is version v2.

materialscloud:2017.0008/v2

2020.158 (version v4) [This ... scloud:az-b2
2017.0008/v3 (version v3) ... scloud:2017.0008/v3
2017.0008/v2 (version v2) ... DOI 10.24435/materialscloud:2017.0008/v2
2017.0008/v1 (version v1) Dec 21, 2017 DOI 10.24435/materialscloud:2017.0008/v1

Files
(with file description, MD5, ...)

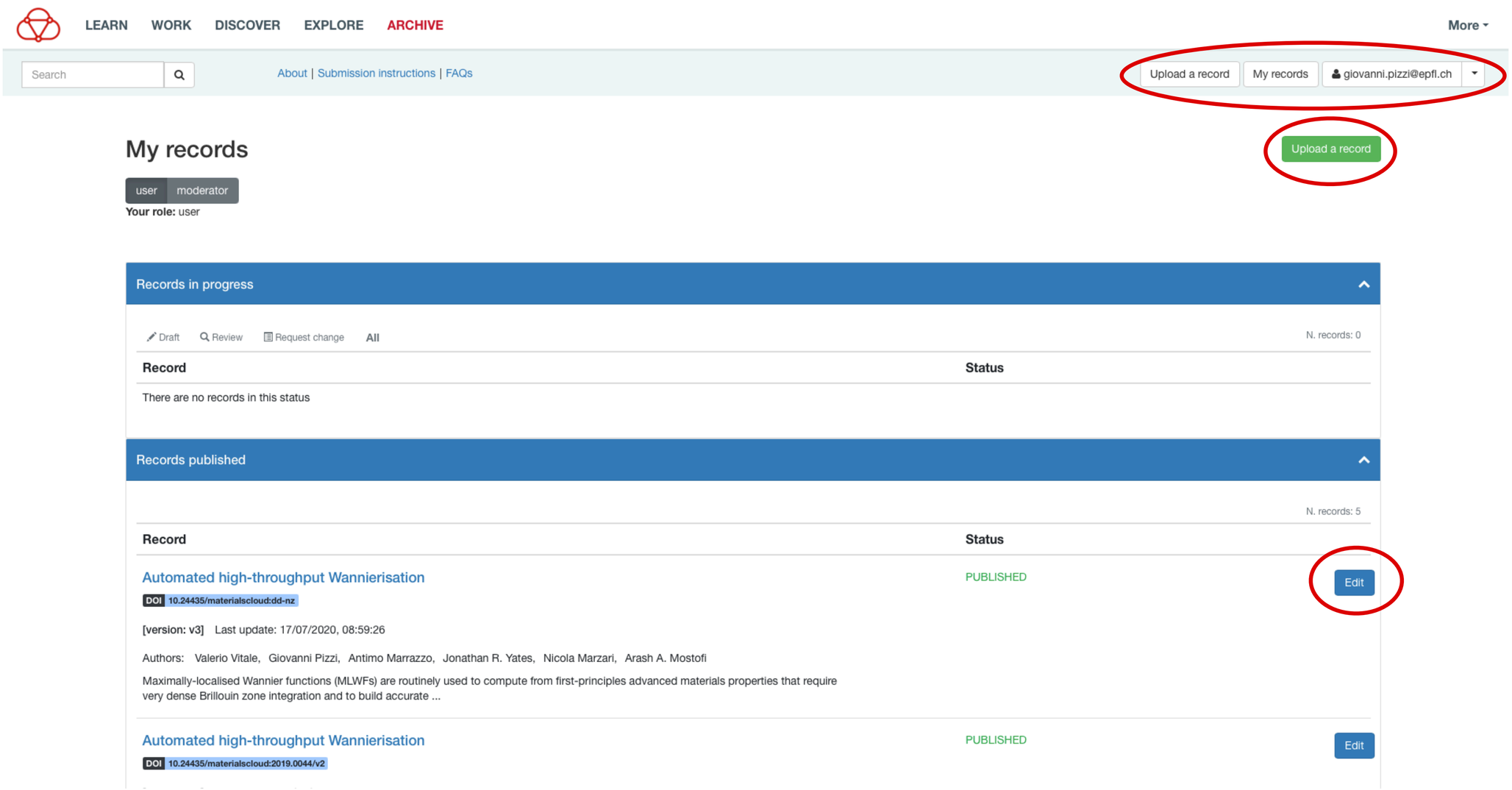
License (pick default if unsure)

References to scientific paper (published or on arXiv) or other resources

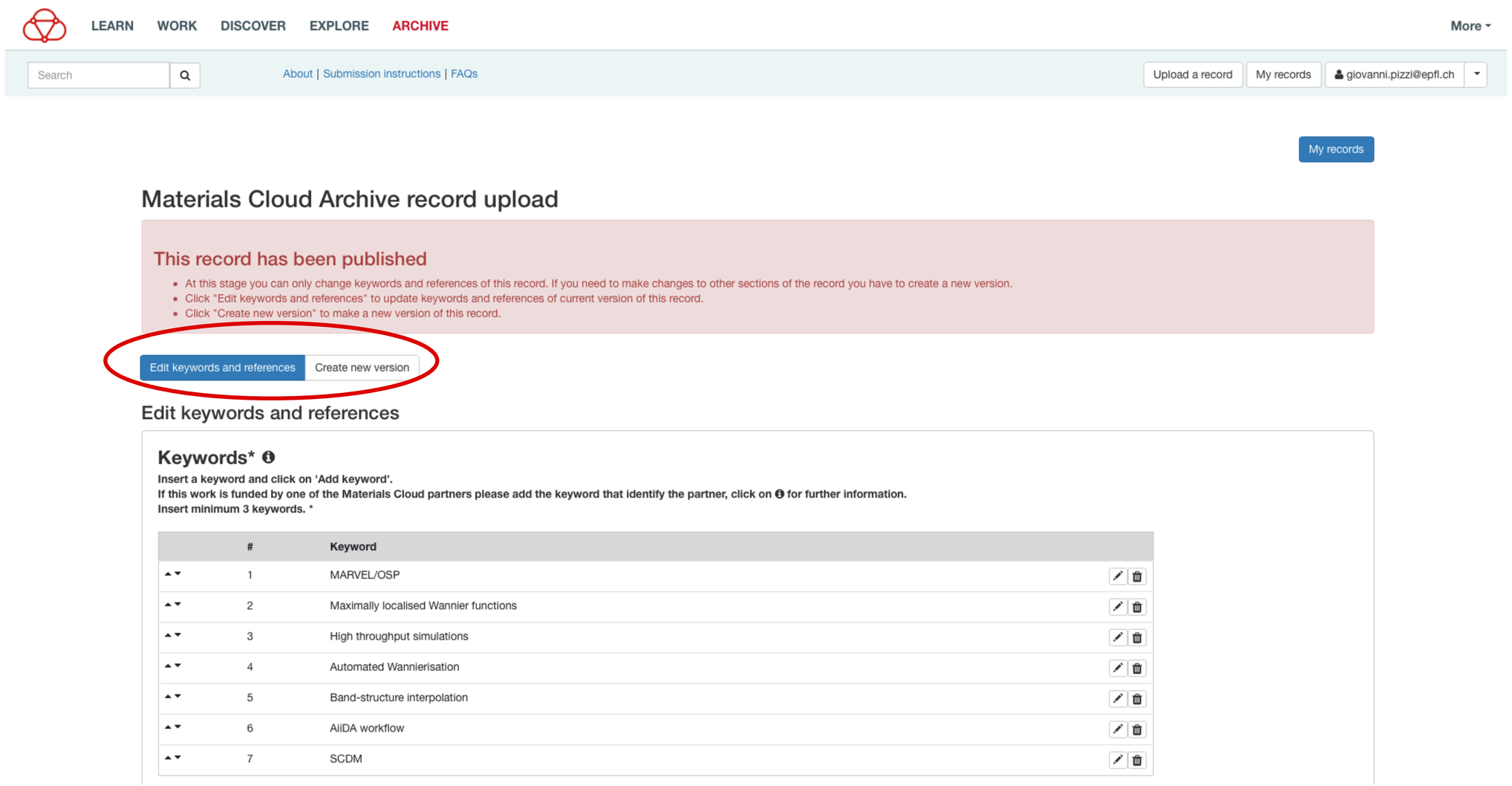
Keywords

Version history

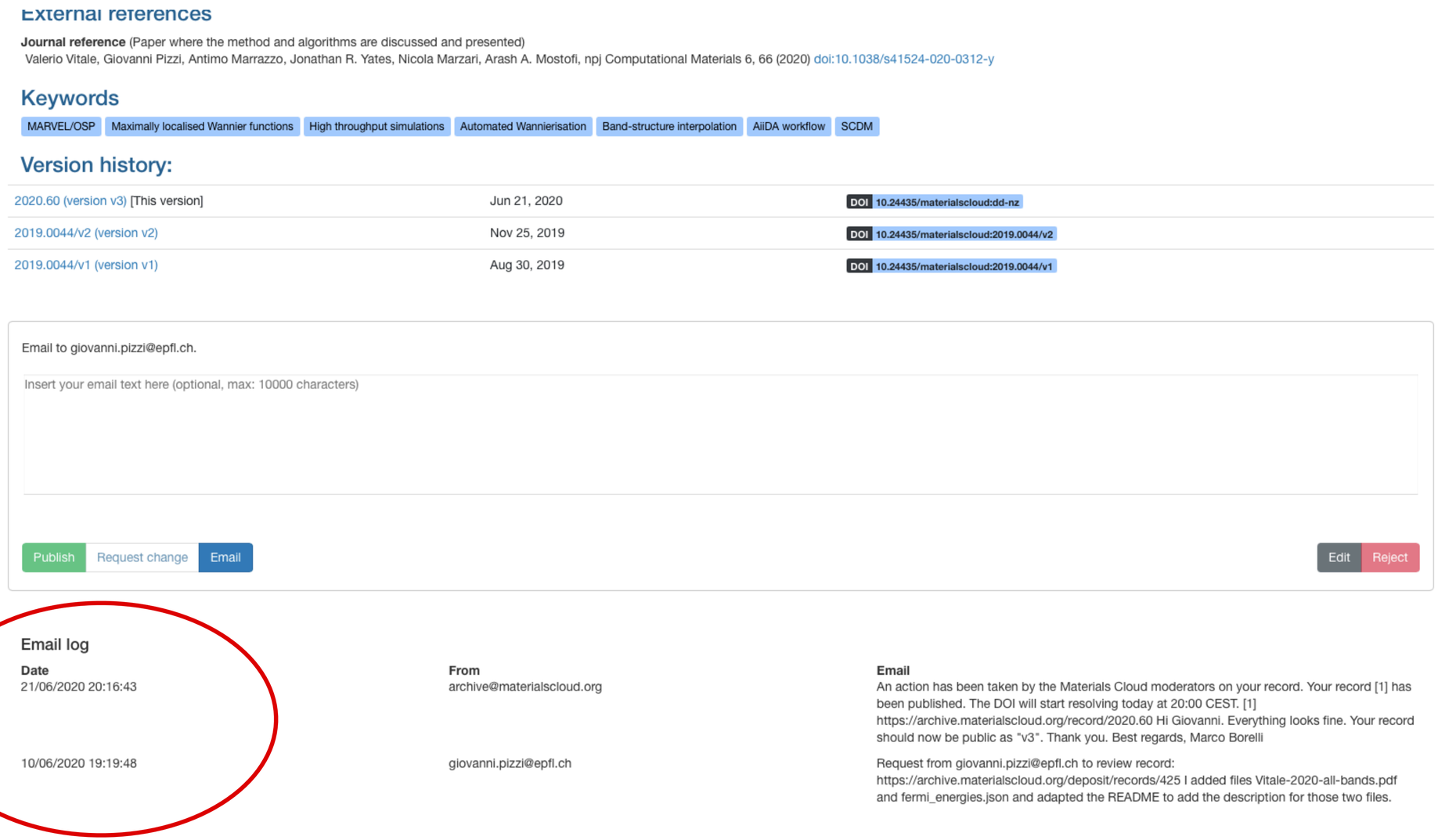
User view on your entries (after login)



Editing references, or creating a new version



Private (pre)view of an entry, and log of messages with moderators



External references

Journal reference (Paper where the method and algorithms are discussed and presented)
Valerio Vitale, Giovanni Pizzi, Antimo Marrazzo, Jonathan R. Yates, Nicola Marzari, Arash A. Mostofi, npj Computational Materials 6, 66 (2020) doi:10.1038/s41524-020-0312-y

Keywords

MARVEL/OSP | Maximally localised Wannier functions | High throughput simulations | Automated Wannierisation | Band-structure interpolation | AiiDA workflow | SCDM

Version history:

2020.60 (version v3) [This version]	Jun 21, 2020	DOI 10.24435/materialscloud:dd-nz
2019.0044/v2 (version v2)	Nov 25, 2019	DOI 10.24435/materialscloud:2019.0044/v2
2019.0044/v1 (version v1)	Aug 30, 2019	DOI 10.24435/materialscloud:2019.0044/v1

Email to giovanni.pizzi@epfl.ch.

Insert your email text here (optional, max: 10000 characters)

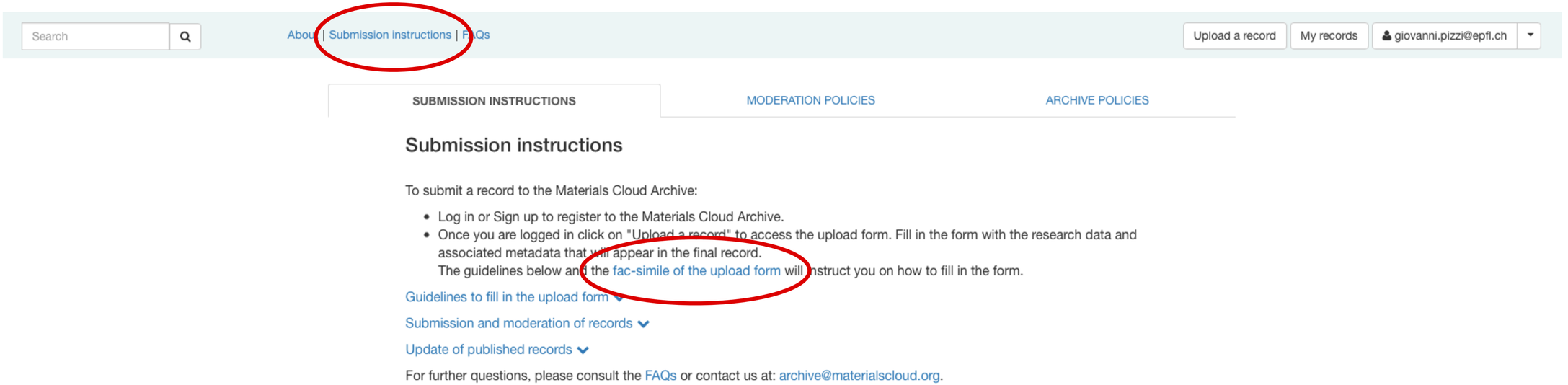
Publish | Request change | Email

Edit | Reject

Email log

Date	From	Email
21/06/2020 20:16:43	archive@materialscloud.org	An action has been taken by the Materials Cloud moderators on your record. Your record [1] has been published. The DOI will start resolving today at 20:00 CEST. [1] https://archive.materialscloud.org/record/2020.60 Hi Giovanni. Everything looks fine. Your record should now be public as "v3". Thank you. Best regards, Marco Borelli
10/06/2020 19:19:48	giovanni.pizzi@epfl.ch	Request from giovanni.pizzi@epfl.ch to review record: https://archive.materialscloud.org/deposit/records/425 I added files Vitale-2020-all-bands.pdf and fermi_energies.json and adapted the README to add the description for those two files.

Submission instructions and policies



Submission form

Materials Cloud Archive record upload

Submission procedure

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.

WARNING: For submissions above 5 GB please contact us. For more information on the files limits click on ℹ in the section Files below.

Title*

The title should have only the first letter of the first word in capital letters, not all words (except where English grammar requires it, e.g. for proper nouns).

Insert title of record *

Description* ℹ

Insert description of record *

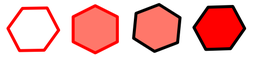
License type*

Choose a license for the data associated with your record (see the spdx license list for details).
The metadata will anyway be released under the Creative Commons Attribution Share-Alike 4.0 International license, except for email addresses.

Select a license *

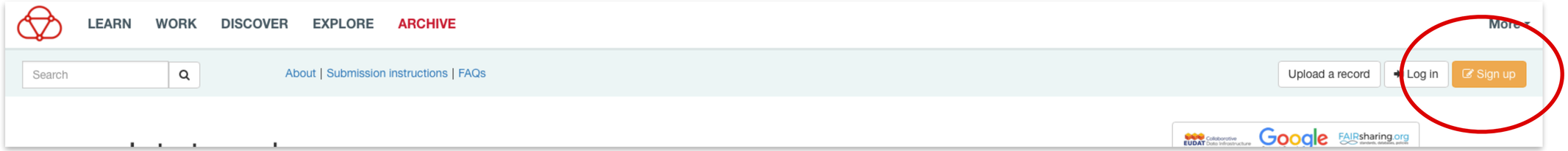
Additional non-binding comments

Further comments regarding the license (optional)



What should you do?

- **Sign up** and create an account (if you don't have one) on https://archive.materialscloud.org



- *For every paper* you are a co-author of:
  - **Create files** with the data needed to reproduce your paper + README + ...
  - **Fill in the form**, if needed answer the moderators' suggestions, get a DOI (DOI available immediately, but active after "acceptance" and publication of the entry, within a week)
  - **Cite your data entry in your paper**. See "How to cite this record" in each entry; e.g.:
    The data and the scripts used to create all results and images in this work are available on the Materials Cloud Archive [XXX] A. Author *et al.*, *My fancy title*, Materials Cloud Archive 2022.01 (2022), doi: 10.24435/materialscloud:9f-2d
  - (After publication of your paper): **update the reference to your scientific paper from your page** (so you get more citations!)

A few notes on filling the form

- **In case of doubt: read the guidelines (or write to archive@materialscloud.org)**
- Use title and description (=abstract) as in a paper
- Use at least three keywords
  - **For MARVEL papers**: add "MARVEL" as a keyword, as well as "MARVEL/XXX" (XXX=DD1, DD2, OSP, ...): you will get then template/bibtex for the yearly MARVEL report pre-filled!! **(less work for you!)**
- Favour open formats, organise your files in a way another person (or yourself in 1 year!) can understand and reuse (and, thus, cite you)
- If you have many files: put them in a single .zip/.tar.gz
- Keep an "outer" README text file: people can read that first before downloading your 1GB+ zip files!
  (see e.g. https://archive.materialscloud.org/record/2020.158, https://archive.materialscloud.org/record/2020.60 or https://archive.materialscloud.org/record/2021.73 for some entries of myself with README files)