

HORIZON2020

Deliverable D2.6
Pipelines to/from external repositories



D2.6

Pipelines to/from external repositories

Flaviano J. dos Santos, Francesco Benedetto, Claudio Rosati, Matteo Bertocchi,
Francesco Aquilante, Arrigo Calzolari, and Nicola Marzari

Document information

Project acronym:	INTERSECT
Project full title:	Interoperable Material-to-Device simulation box for disruptive electronics
Research Action Project type:	Accelerating the uptake of materials modelling software (IA)
EC Grant agreement no.:	814487
Project starting / end date:	1 st January 2019 (M1) / 31 st January 2022 (M37)
Website:	www.intersect-project.eu
Final version:	28/07/2021
Deliverable No.:	D2.6
Responsible participant:	EPFL (participant number 2)
Contributing Consortium members:	FRA, CNR, AMAT
Due date of deliverable:	31/07/2021
Actual submission date:	29/09/2021
Dissemination level:	PU - Public

Authors: Flaviano J. dos Santos, Francesco Benedetto, Claudio Rosati, Matteo Bertocchi, Francesco Aquilante, Arrigo Calzolari, and Nicola Marzari

To be cited as: F. J. dos Santos, F. Benedetto, C. Rosati, M. Bertocchi, F. Aquilante, A. Calzolari, and N. Marzari (2021): Pipelines to/from external repositories. Deliverable D2.6 of the H2020 project INTERSECT (final version as of 28/07/2021). EC grant agreement no: 814487, EPFL, Lausanne, Switzerland.

Disclaimer:

This document's contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

Versioning and Contribution History

Version	Date	Modified by	Modification reason
v.01	05/07/2021	Flaviano J. dos Santos	First Version
v.02	17/07/2021	Claudio Rosati	Revision
v.03	27/07/2021	Arrigo Calzolari	Final Revision

Content

1. Executive Summary	5
1.1 About this document	5
2. Introduction: IM2D overview	5
3. IM2D data scheme	7
3.1 Data pipeline	8
3.2 Sharing the data	10
4. OPTIMADE	10
4.1 REST API	11
4.2 Materials Cloud OPTIMADE client	11
5. Ginestra-AiiDA plugin and GUI	13
5.1 AiiDA connection	14
5.2 Materials Project Connection	15
5.3 OPTIMADE Integration	17
6. Conclusions	19
References	20
ACRONYMS ¹	21

¹ Acronyms are marked in purple in the text and defined at the end of the document.

1. Executive Summary

The Interoperable Material to Device (**IM2D**) simulation box is the core outcome of the INTERSECT project. It is meant to provide material properties on-demand through robust and automatized workflows for advanced device modelling. In the various stages of the IM2D simulation box, data are received, processed, and stored (see also Data Management Plans D4.2 and D4.6).

In this document, we provide an overview of the recent developments on the acquisition of required data for the various simulation steps, in particular from external sources. We describe how the automated infrastructure is able to handle, store, and share data, which is the core of the IM2D *data hub*. We start with an introduction to the components of the IM2D box and how they interact, then we give a complete description of the data pipelines.

An essential piece of information required to perform Density Functional Theory (**DFT**) simulations on a material is its atomistic crystal structure. This contains information on the periodicity of the lattice, its chemical composition, and the arrangement of the atoms inside the unit cell. Up to now, we have implemented data pipelines to give access to the IM2D users materials available to almost all the major open-access materials databases, such as the Materials Cloud [1], Materials Project [2], Crystallography Open Database (**COD**) [3], and many more. An important step to make this access as much universal as possible, has been the development of the **OPTIMADE** (Open Databases Integration for Materials Design) [4] **REST** (REpresentational State Transfer) **API** (Application Programming Interface), which unifies the way various databases expose their data on the web.

We present our effort towards the (**OPTIMADE**) and the tools we help to develop to make it user-friendly and open access. Furthermore, we demonstrate how we implement queries to the OPTIMADE API within the IM2D framework.

1.1 About this document

This document represents the deliverable D2.6 of the INTERSECT project and it is prepared under the Task 2.5 *Data hub: Upscaling Ginestra database* of Work Package 2. The content of this document is intended to describe all the developments on the data pipelines for the IM2D simulation box.

2. Introduction: IM2D overview

To leverage the European leadership in materials' modelling, INTERSECT aims at developing the IM2D simulation box, which integrates some of the most used materials and devices modelling codes to offer turn-key solutions for developments on disruptive electronics. The main components are:

Ginestra™: a software to simulate the operation and electrical characteristics of modern logic (e.g., FinFET, FeFET) and memory (e.g., Flash, RRAM, CBRAM, FeRAM, PCM) devices, with a particular focus on dielectric-related phenomena and novel materials. The software is based on a multiscale modelling approach and relies on a comprehensive physics-based description of charge/ion transport, charge trapping, and more [5].

AiiDA (Automated Interactive Infrastructure and Database for Computational Science): an open-source Python infrastructure to help researchers with automating, managing, persisting, sharing, and reproducing the complex workflows associated with modern computational science and all associated data. AiiDA is built to support and streamline the four core pillars of the ADES (Automation, Data, Environment, and Sharing) model [6,7].

SimPhoNy (Simulation framework for multi-scale phenomena in micro- and nanosystems): an infrastructure to integrate multiscale modelling environment for *in Silico* discovery and design of nano-enabled systems and materials. The main concept of the SimPhoNy framework is to augment existing open-source and commercial simulation tools and supplement them with sophisticated interface software libraries that allow for a smooth flow of information from one component to the other and from one scale to another. The integrated tools range from those describing the electronic structure and atomistic scales up to those modelling mesoscopic and macroscopic device level scales [8].

Quantum ESPRESSO: an integrated suite of Open-Source computer codes for electronic-structure calculations and materials modelling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials [9].

SIESTA: is both a method and its computer program implementation, to perform efficient electronic structure calculations and ab initio molecular dynamics simulations of molecules and solids. SIESTA's efficiency stems from the use of a basis set of strictly-localized atomic orbitals. A very important feature of the code is that its accuracy and cost can be tuned in a wide range of features, from quick exploratory calculations to highly accurate simulations matching the quality of other approaches, such as plane-wave methods [10].

In the following, we present an overview of the IM2D infrastructure, its components, and their interactions.

Users have access to IM2D through the graphical front-end of the Ginestra-AiiDA plugin, which is the core of the IM2D Graphical User Interface (GUI). More info about the current version of the plugin can be found in D2.4, D2.5, and Section 5 of this deliverable. It is a Java application that can run from within the Ginestra™ main code or as a stand-alone application, see Fig. 1. After receiving the user's input, the Ginestra-AiiDA plugin then communicates with AiiDA via two possible routes. The first is a direct communication with AiiDA through the AiiDA-POST plugin (developed entirely within the INTERSECT project, see D2.1). The AiiDA-POST enhances the REST API of AiiDA allowing us to perform specific activities for the IM2D box, most

importantly, the submission and control of calculations and workflows through a web interface. The second route of communication is through SimPhoNy (see D2.5), which works as an interoperability layer and provides a semantic metadata schema based and compliant with the European Materials Modelling Ontology (EMMO) standards [11]. Most of the communication functionalities with the two routes are equivalent, but the interoperability layer through Simphony also allows the specification of the user profile (*persona*, D1.1), which dynamically changes the interface to the users, what is required and exposed to them.

Starting from the request from the user, AiiDA is responsible to store the relevant data related to a given operation and to generate a provenance of all data created in the process. AiiDA then executes the operation, launching specific workflows. A possible operation is running a DFT code. AiiDA, through the respective plugins, prepares the input files for the code, transports the information, monitors the process, retrieves the results, and stores the outputs and its provenance. AiiDA does not send data actively to the Ginestra-AiiDA plugin, which needs to inquire AiiDA through the AiiDA-POST REST API and to consult the result of previously submitted operations.

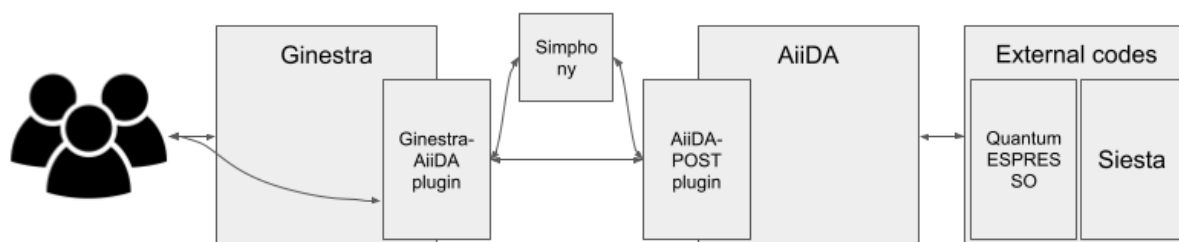


Figure 1: Principle components of the IM2D simulation box. The user interacts with the Ginestra-AiiDA plugin (from Ginestra™ or the stand-alone applet). The communication between Ginestra™ and AiiDA is done both through SimPhoNy and the AiiDA-POST REST API. AiiDA then coordinates the calculations, stores the provenances, and communicates with external DFT codes.

The IM2D box allows for very versatile deployments. For example, Ginestra™ can run on the local computer of the user, but the stand-alone version of the Ginestra-AiiDA plugin can be deployed both in a local machine and in a cloud-based server (e.g., Marketplace). SimPhoNy and AiiDA can be deployed either in the user's machine or in a dedicated server (virtual or not). Furthermore, AiiDA allows running the DFT codes on different servers, such as remote clusters and supercomputers.

In the next section, we will focus on the data flow with the IM2D architecture.

3. IM2D data scheme

The INTERSECT project plans to make the research data produced within the project Findable, Accessible, Interoperable, and Reusable (FAIR). Currently, the AiiDA infrastructure and the

Materials Cloud [1] dissemination platform provide the most comprehensive capability in this respect for calculations and data in computational science. However, we know that industrial users need to keep and protect their data within the company's firewall (D4.2, D4.6). Research data produced during the operation of the IM2D, which is integrated through AiiDA with all codes of the simulation hub, can be exposed in full using the Materials Cloud architecture that we describe in the following.

Materials Cloud is an open platform for computational materials science. It is built to enable the seamless sharing and dissemination of resources in computational materials science, offering educational, research, and archiving tools; simulation software and services; and curated and raw data. These underpin published results and empower data-based discovery, compliant with data management plans and the FAIR principles. On the Materials Cloud, it is possible to access cloud or re-deployable simulation services (such as the AiiDALab turnkey solutions, the Quantum Mobile virtual machine, and the AiiDA registry of plugins and workflows); educational material, tutorials, and lectures; and to share scientific results. The Materials Cloud is developed and maintained mainly by EPFL, and part of its resources are also powered by AiiDA.

In the INTERSECT IM2D simulation box, we exploit the capabilities of AiiDA and of the Materials Cloud to achieve our goals of generating, storing, and sharing FAIR data. The full provenance model in AiiDA implies that all the inputs and outputs of all calculations are preserved together with the links between calculations and data in complex workflows, encoded in the form of AiiDA graphs. Internally, AiiDA relies on the highly-standardized PostgreSQL and provides a platform-independent export format. Universally-Unique IDentifiers (UUIDs) ensure that each node in an AiiDA graph is globally unique.

3.1 Data pipeline

We here describe the data flow within the IM2D infrastructure. In order to do that, let's assume a user needs the band gap of a given semiconductor to power their electronic device simulation with Ginestra™.

The user opens the Ginestra-AiiDA plugin through the GUI. Next, the user needs to specify the materials of interest, by providing the material crystal structure in a CIF (Crystallographic Information File). This file is uploaded in the Ginestra-AiiDA plugin interface, which is sent to AiiDA through the AiiDA-POST REST API (or via SimPhoNy), see Fig. 2. AiiDA then stores that structure in the internal database and assigns to it an UUID. Again, in the Ginestra-AiiDA plugin, the user has the chance of selecting any of the structures in the AiiDA internal database, and may request the calculation of the band gap. This request is then sent to AiiDA, which will prepare the input card, submit the calculation, monitor the execution, and finally parse and store the results while tracking the whole process. The band gap is now stored in the AiiDA internal database. The Ginestra-AiiDA plugin has features that allow the user to monitor the

progress of the calculation. Once the calculation is finished, the band gap (or any other calculated properties) will be displayed in association with the structure of the material. That is, shall the user select that structure in the Ginestra-AiiDA plugin again, they will be presented with all available properties for that material.

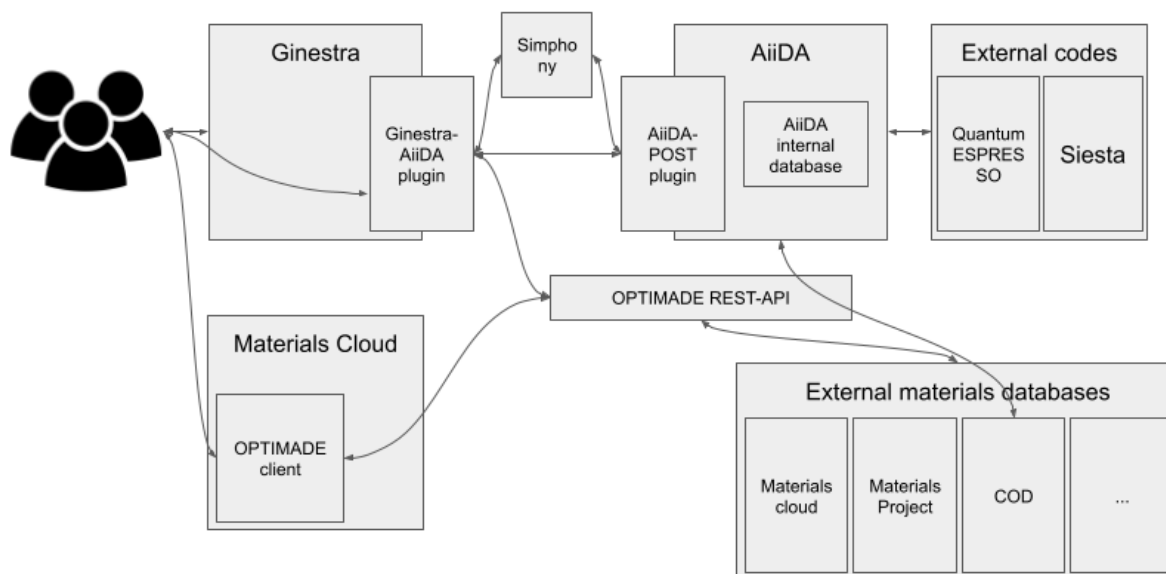


Figure 2: IM2D simulation box and its data pipelines. Users have access through extensive materials databases through the OPTIMADE client available on the Materials Cloud or directly through the IM2D simulation box.

So far, we assumed that the user had the information on the material crystal structure. As this might not always be the case, we offer to the user the opportunity to search for the crystal structure of the material of interest both in the AiiDA internal database and in external sources. This can be done by specifying the material's chemical formula, for example. So far, we have implemented the search on external sources, such as the Materials Project, through its REST API [2] and the COD [3] through the AiiDA-codtools plugin [12]. The implementation of the OPTIMADE REST API is in progress. At its completion, the OPTIMADE route will provide a unified access to the Materials Project, COD, and further sources, such as databases in the Materials Cloud, the Open Quantum Materials Database (OQMD), the AFLOW database (Aflowlib), the Novel Materials Discovery (NOMAD) Laboratory database. The search is performed in the selected data source and the result is stored in the AiiDA internal database making these newly added structures available for the computation of their electronic properties.

Alternatively, the access to the OPTIMADE search capabilities is already available in the Materials Cloud by means of the OPTIMADE Client released in the first semester of 2021 (D4.6). In the DoA we planned to realize the data interconnection of IM2D *to* and *from* and the external databases. However, at the moment the OPTIMADE REST API only offers the possibility of getting data from the external databases, but no data can be sent to them, as most of these

databases do not support such a functionality. Thus, at present the option “to OPTIMADE” is unfeasible, and hence it has not been implemented in IM2D.

3.2 Sharing the data

The IM2D interoperability hub (based on AiiDA and SimPhoNy) allows us to store the results of calculations, and furthermore to automatically record their entire full provenance with rich metadata in a standardised format convertible to other formats with dictionaries. Thus, with AiiDA it is very easy to share all the required data, e.g., for the reproduction of a calculator. The user can export a set of nodes from the AiiDA internal database and upload it, for example, to the Materials Cloud. This, being an open data dissemination platform, allows to expose to the community at large the raw data (Materials Cloud Archive), the curated data (Materials Cloud Discover), and the full calculation workflows (Materials Cloud Explore); these can be automatically uploaded through the AiiDA infrastructure. First, to an entry in the Materials Cloud is assigned a permanent DOI (Digital Object Identifier), then the entire archive is catalogued in re3data [13] and FAIRsharing [14], and most importantly it is linked to the Discover and Explore sections of the Materials Cloud where all data and workflows become fully available. Of course, this happens only if the users willingly upload their AiiDA databases to the Materials Cloud. For safety reasons, industrial users and consultants are expected to keep their databases within the company firewalls.

On the Materials Cloud, data is accessible through an interactive web interface for browsing databases. The data is furthermore accessible through a REST API, providing an interface for automated queries. IM2D users have the option to download *entire* AiiDA databases for import and reuse in their personal AiiDA instances. Industrial users may decide to protect their data for exploitation and commercialization purposes by keeping their data in the AiiDA internal database. Long-term storage of the data is provided by the Materials Cloud Archive, a non-profit research data repository, which provides its own guidelines and templates for good data management practices.

The user also has the option of serving their personal database online via the AiiDA core REST API. In this way, the results of research can be made available to the community through a web interface.

Next, we discuss the developments on the OPTIMADE REST API and OPTIMADE Client at the Materials Cloud that made possible the implementation of the above-mentioned data flow.

4. OPTIMADE

As part of the INTERSECT project, the realization of the IM2D infrastructure includes the development of efficient tools that can be used to access and/or populate public databases for materials modelling. The access to databases and repositories is meant to provide structural

and electronic properties of functional materials, either for direct analysis or as input for computations in both materials and device cycles (*simulation hub*). Conversely, DFT simulations produce the necessary data to populate the databases for exploitation from third parties. Such tasks are managed by means of the AiiDA infrastructure, as the latter guarantees extreme reliability and efficiency in all steps from the generation of the simulation results, to the inspection of data and their provenance. Very importantly, AiiDA is committed to provide a Uniform Resource Identifier ([URI](#)) syntax compliant with the standards defined in the OPTIMADE protocol of common syntaxes on heterogeneous databases. Finally, the AiiDA REST API is used as an on-line service that exposes a repository of data persisted in AiiDA.

4.1 REST API

The best-known materials databases (e.g., AFLOW, COD, Materials Cloud, The Materials Project, NOMAD) are continuously upgraded in their wealth of physically relevant data by means of massive high throughput computer simulations. The OPTIMADE has been created to foster an optimal interoperability of data among these databases. The goal is to provide a universal API to make materials databases accessible and interoperable. The providers of several materials databases joined forces to design a common API specification to enable seamless access and interoperability across materials databases. Initiatives at the level of the semantic tools have been pursued in this respect, and include the CIF dictionary, the Pauling vocabulary, and EMMO. Altogether, they constitute the common semantic ground for the IM2D implementation (see WP1 activity and D1.1, D1.3, D1.4, as well as D2.4 for the description of the automatic procedures of importing COD and/or CIF structures into Ginestra™ through AiiDA). Each provider serves online a REST API that complies with OPTIMADE specifications. Thus, different clients can access all databases through the same interface learning a single way to query and a single way to interpret the returned data.

The IM2D users have the option of serving (online access) to their AiiDA internal database thanks to an OPTIMADE API gateway, compliant with the JSON API 1.0 specification. The actual access *via* AiiDA to the OPTIMADE gateway involves the use of *aiida-optimade*, freely available at <https://github.com/aidata-team/aiida-optimade> and integrated within AiiDA as StructureData nodes importer, in compliance with the REST API technology. AiiDA-optimade is meant to guarantee a full compliance with the OPTIMADE specifications related to the allowed data types, the general API requirements and conventions, the response formats, the API endpoints and filters, and the entry lists. Additional information on the OPTIMADE specifications can be found at www.optimade.org/optimade. AiiDA-optimade is, for example, used to server the databases in the Materials Cloud.

4.2 Materials Cloud OPTIMADE client

A REST API query is usually made through a construction of a web address. Something that may

look like:

```
http://my_server:5000/v1/structures?page_limit=10&page_offset=10&filter=elements HAS ANY "C", "Si", "Ge", "Sn", "Pb".
```

using keywords and the right syntax, the user queries and filters the database. Obviously, the knowledge of the OPTIMADE syntax is needed to perform this task. Should a user lack this expertise, they can use the OPTIMADE client we built: a user-friendly web graphical application suitable for inexperienced users, see Fig. 3. This application is publicly available on the Materials Cloud (<https://www.materialscloud.org/work/tools/optimadeclient>). With the help of the aiida-optimade client, it becomes extremely easy to browse and search materials data from several providers (e.g., materialscloud, AFLOW, open materials database, etc.), and databases (e.g., standard pseudopotentials, 2d structures, 3D crystal database, etc.), through chemical and/or crystallographic specification of the desired system.

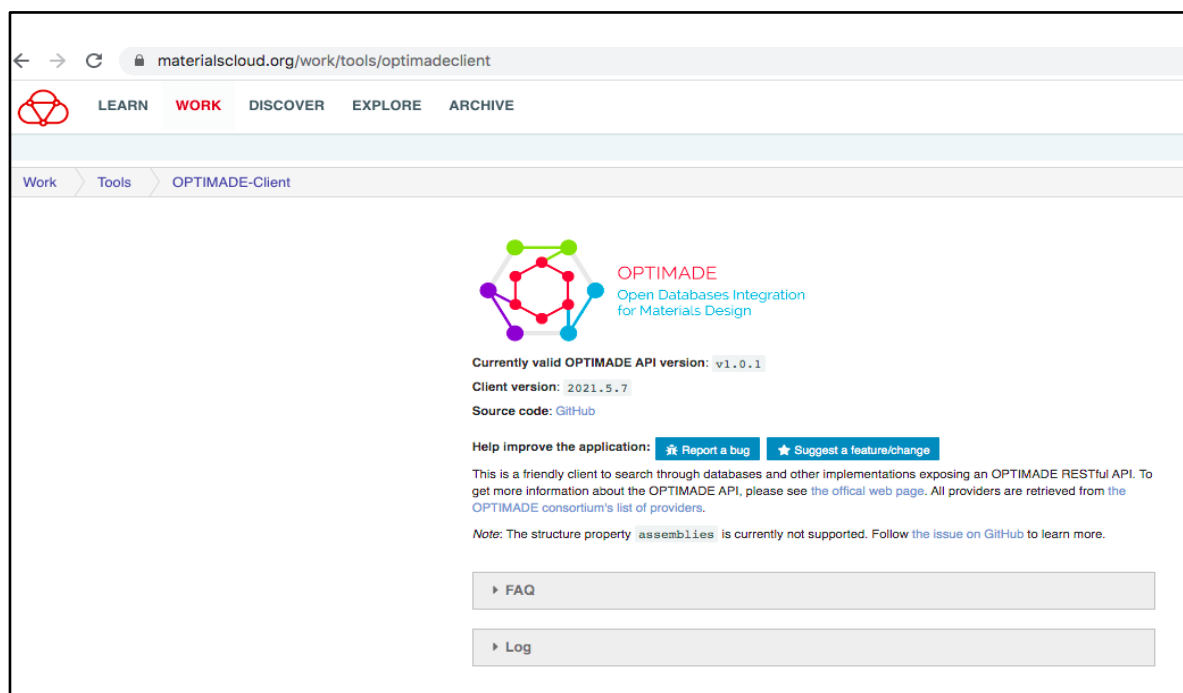


Figure 3: Screenshot of the webpage on the Materials Cloud from which the OPTIMADE client can be accessed.

From the web graphical interface of the OPTIMADE client, the user can choose different available databases. Then, they can specify the elements that compose the material of interest in a periodic table, see Fig. 4 left panel, and some chemical species can be filtered out at this stage. Pressing *Search*, a list of all matching results is given. Results can be organized by attributes of the materials such as the chemical formula, number of sites, number of elements, etc. (see Fig. 4 right panel). The selection of a result triggers the interactive crystal structure visualization. Other properties are also displayed.

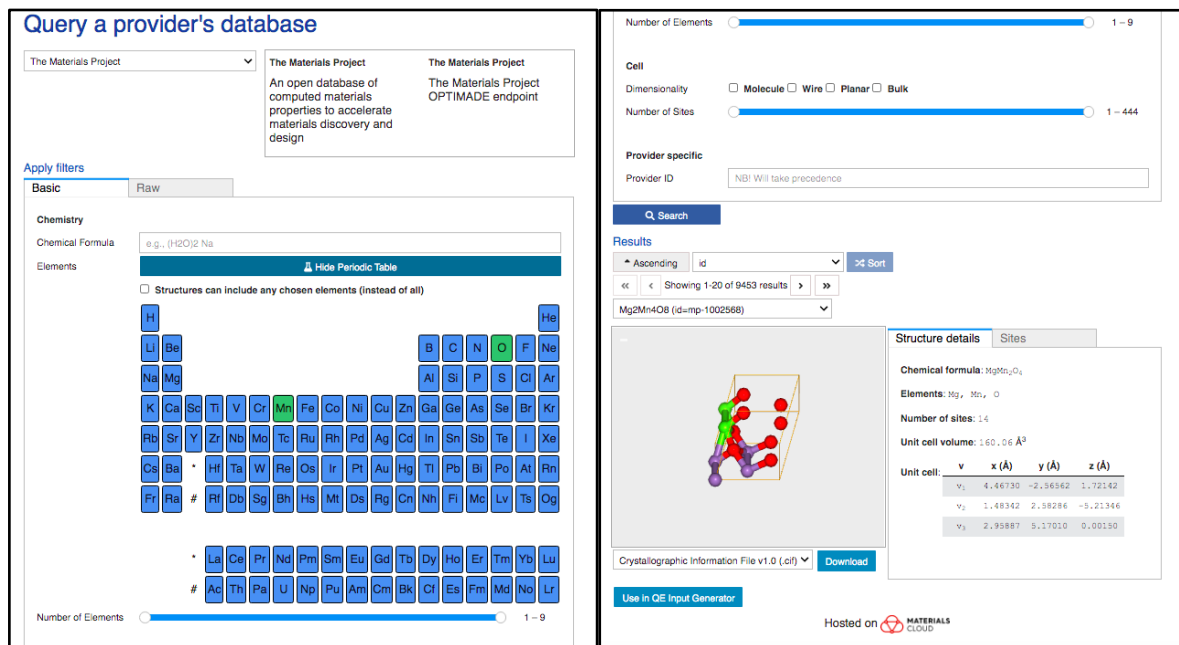


Figure 4: Screenshot of the webpage on the Materials Cloud to access the OPTIMADE client. Left panel shows the dropdown button with the list of available databases to inquire, and the periodic table to specify the chemical composition of the material of interest. Right panel shows the visualization panel of the query results, including an interactive crystal structure visualizer.

Next, we discuss the developments in the Ginestra-AiiDA plugin with respect to the data pipelines.

5 Ginestra-AiiDA plugin and GUI

The Ginestra-AiiDA plugin is a Java application that can be deployed from within Ginestra™ main code or can be run as a stand-alone applet. The latter also offers the non-Ginestra™ users the chance of exploiting its capabilities: searching for materials in external and internal databases, submitting calculations and workflows to calculate materials properties, etc. The stand-alone version can also be integrated in a web interface such as Materials MarketPlace.

Recently, we evolved the Ginestra-AiiDA plugin interface to communicate to AiiDA through SimPhoNy and provide the users with access to public databases. The goal is to keep developing the Ginestra-AiiDA plugin and using the final front end as the IM2D GUI. The interface can now connect directly to AiiDA, SimPhoNy, The Materials Project, and public databases such as NOMAD, AFLOW, OQMD, and others, through the OPTIMADE API. In Fig. 5, a representation of the data pipeline is given. In this section, we describe how the plugin retrieves data from the AiiDA internal database, the Materials Project and other public databases.

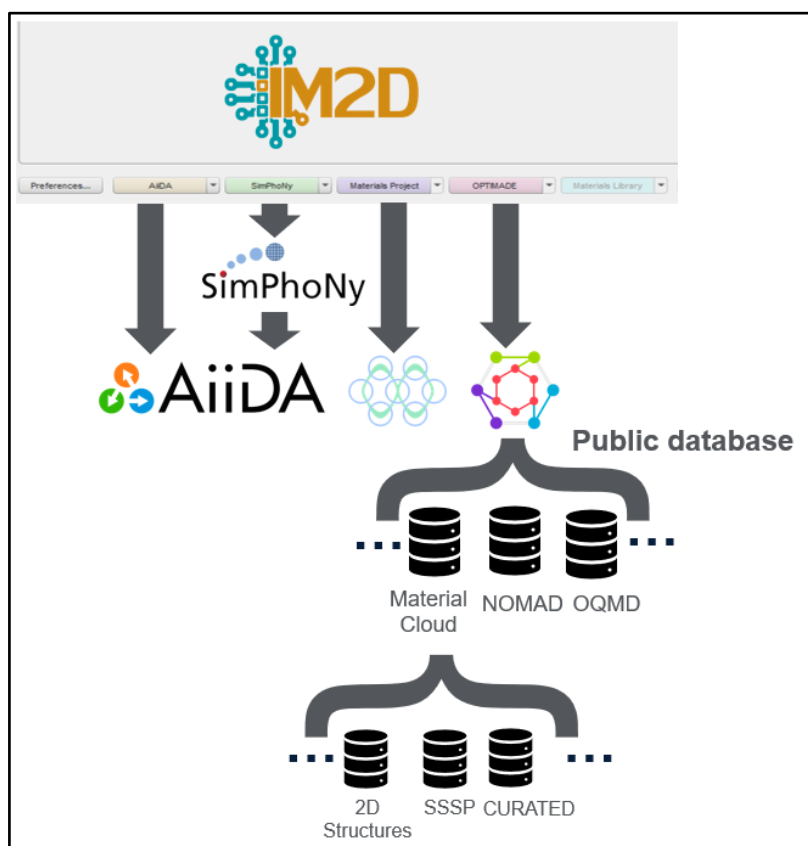


Figure 5: Ginestra-AiiDA Plugin Data Pipeline Scheme. The Ginestra-AiiDA GUI interacts with AiiDA directly or through SimPhoNy. It also accesses external databases of all major providers either through their specific REST API (in the case of Materials Cloud), or through the OPTMADE REST API, which offers a unified REST API for various complaint databases.

5.1 AiiDA connection

The Ginestra-AiiDA plugin interface connects Ginestra™ to the AiiDA database either in a direct manner or through Symphony. Figure 6 schematizes a GET request to the AiiDA internal database, which is realized through the AiiDA-POST REST API. In this example, we request AiiDA the information about all properties that can be computed for a given material. The REST API address reads

```
http://localhost:5000/api/v4/intersect/properties.
```

AiiDA then replies with a dictionary containing the requested information. In this example, a list with properties such as band gap, band structure, dielectric constant, effective mass, formation energy, relaxed structure, etc., is received back.

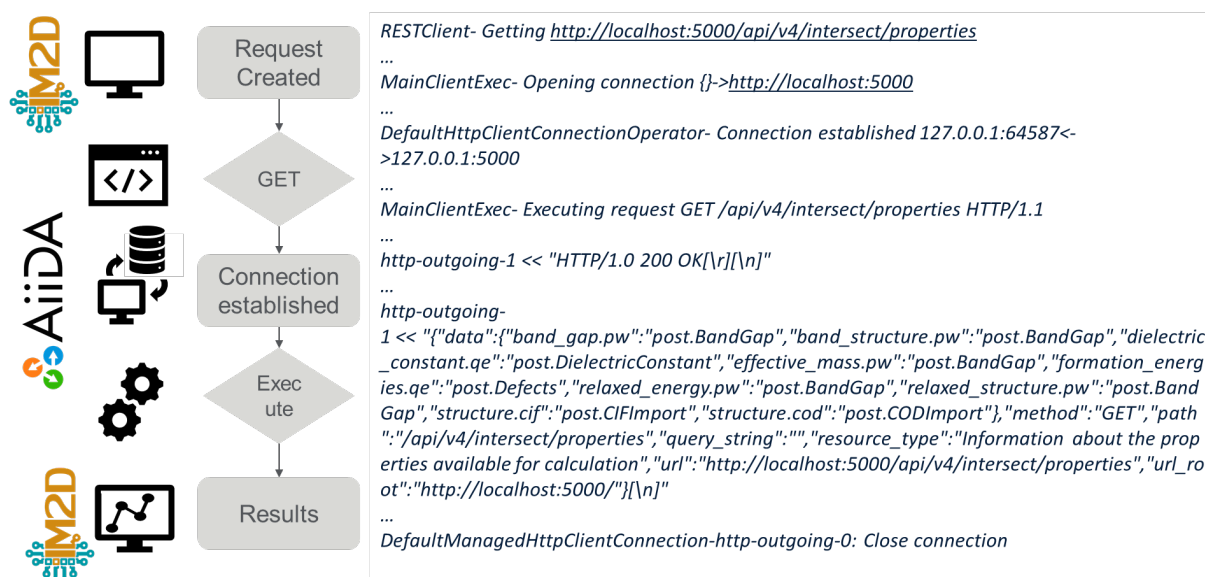


Figure 6: Scheme of the connection framework to AiiDA. The figure shows an example of a query to the AiiDA internal database and stack trace.

5.2 Materials Project Connection

Through the Ginestra-AiiDA plugin it is also possible to retrieve data directly from the Materials Project database, i.e., without using the OPTIMADE REST API, but the one specific to the database. We exploited this as a pilot case to test the interface infrastructure before integrating the OPTIMADE API, which will allow access to all databases of the OPTIMADE consortium.

Figure 7 shows the details and stack trace of a GET request to the Materials Project database. In the following example, we query for all structures containing "Si" in the chemical formula, i.e., all compounds containing silicon. The corresponding REST API address is

https://www.materialsproject.org/rest/v2/materials/*Si*/vasp .

Figure 8 shows a screenshot of the Ginestra-AiiDA plugin displaying the query result, which includes the list of materials and of their relevant properties (such as space point symmetry, chemical formula, formation energy, band gap, volume, density, etc.). In addition, the details of all listed structures in a dedicated page are visualized, see Fig. 9. In this case, another GET request is made to the Materials Project database:

https://www.materialsproject.org/rest/v2/materials/mp-34/vasp/final_structure .

Deliverable D2.6
Pipelines to/from external repositories

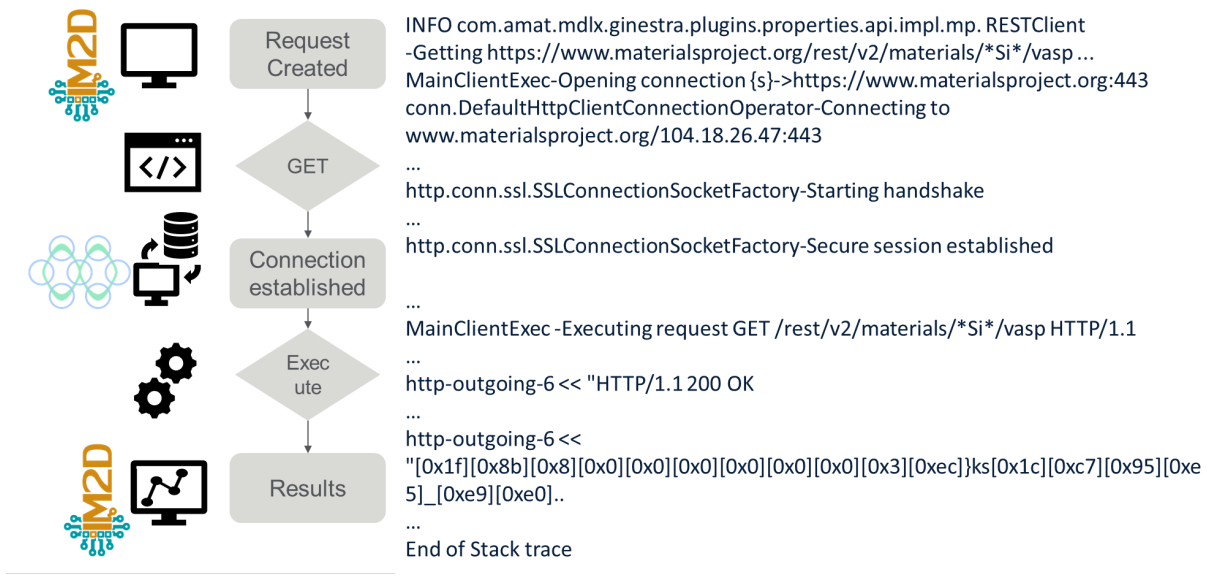


Figure 7: Scheme of the connection framework to the Materials Project database. The figure shows an example of a query for all structures whose formula contains silicon.

ID	Formula	Spacegroup	Formation Energy	E Above Hull	Band Gap	Volume	Unit Cell Sites	Density
mp-571520	Si8	R-3	0.161	0.161	0	146.068	8	2.554
mp-1001113	Si4	Cmcm	0.49	0.49	0	61.423	4	3.037
mp-1120747	Si6	P6_3/mmc	0.454	0.454	0	99.244	6	2.82
mp-34	Si1	P6/mmm	0.305	0.305	0	15.058	1	3.097
mp-168	Si8	Ia-3	0.16	0.16	0	147.415	8	2.531
mp-92	Si2	I4_1/amd	0.294	0.294	0	30.667	2	3.042
mp-999200	Si4	I4/mmm	0.103	0.103	0.452	86.428	4	2.158
mp-1079649	Si8	Cmce	0.428	0.428	0	116.893	8	3.192
mp-1079297	Si8	C2/m	0.074	0.074	0.288	169.804	8	2.197
mp-644693	Si8	P1	0.415	0.415	0	174.108	8	2.143
mp-109	Si4	Fm3m	0.296	0.296	0	61.257	4	3.045
mp-1120447	Si8	P1	0.373	0.373	0	168.439	8	2.215
mp-676011	Si8	P1	0.447	0.447	0	159.131	8	2.345
mp-1268191	Si6	P6/m	0.36	0.36	0	101.229	6	2.764
mp-10649	Si2	P6_3/mmc	0.513	0.513	0	28.743	2	3.245
mp-1094057	Si3	R-3m	0.518	0.518	0	42.938	3	3.258
mp-149	Si2	Fd-3m	0	0	0.853	40.888	2	2.281

Figure 8: Screenshot of the Materials Project interface showing the results of a query for all structures with silicon.

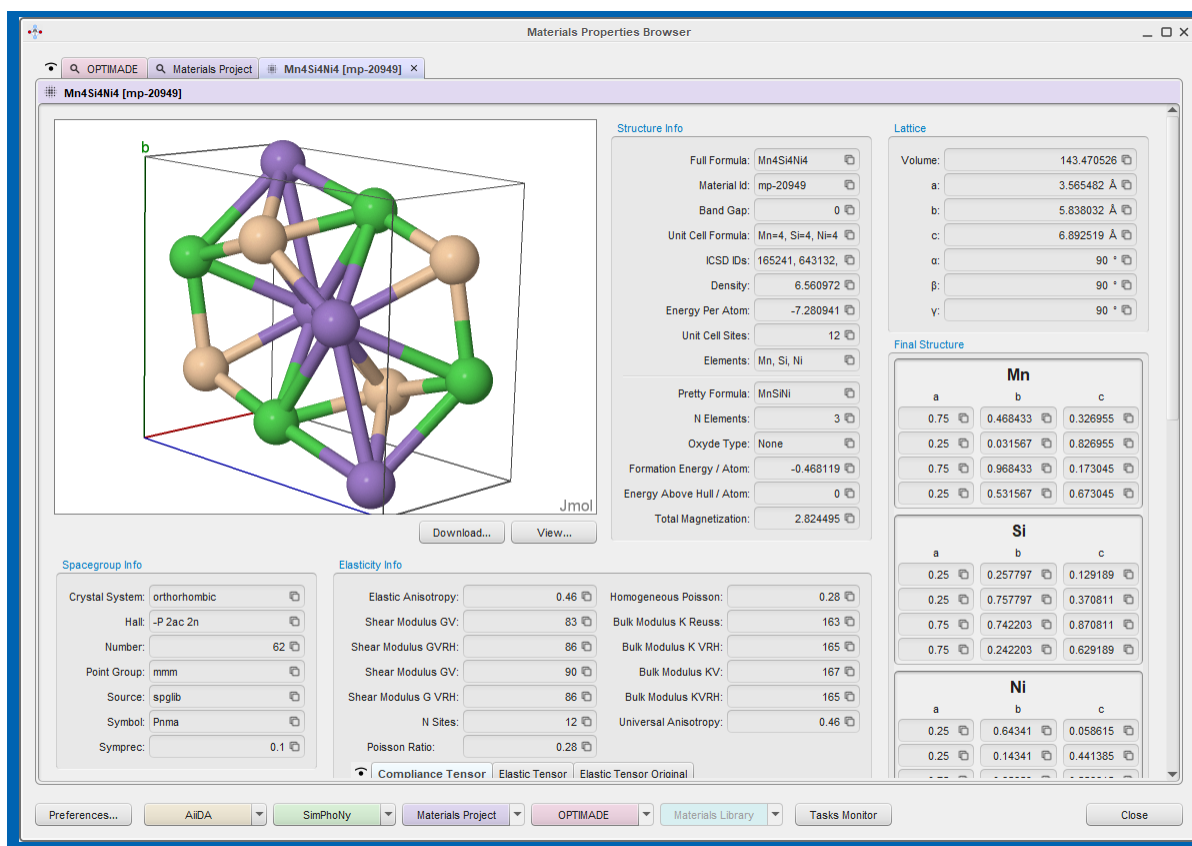


Figure 9: Screenshot of the structure visualization page, which is part of the Ginestra-AiiDA plugin interface with the Materials Project database. After selecting one of the entries in the structure query, all available information for the selected material, including the crystal structure visualization, are retrieved and displayed.

5.3 OPTIMADE Integration

The major development done for the data pipeline in the Ginestra-AiiDA plugin was the implementation of the OPTIMADE REST API. We designed and implemented a *tab* in the plugin interface for this connection. Thus, it is now possible to retrieve data from publicly available databases such as the Materials Project, OQMD, Aflowlib, NOMAD, and especially the Materials Cloud. All these databases are accessed through a universal OPTIMADE REST API.

Opening the Ginestra-AiiDA plugin, the OPTIMADE tab may be opened and the database to be queried may be chosen, see a screenshot in Fig. 10. For example, the user picks the Materials Cloud, that serves various databases accessible to the user (see Fig. 11) and the field called "Provider Child". The user can choose one database and then perform the query via the chemical formula. The REST API address reads:

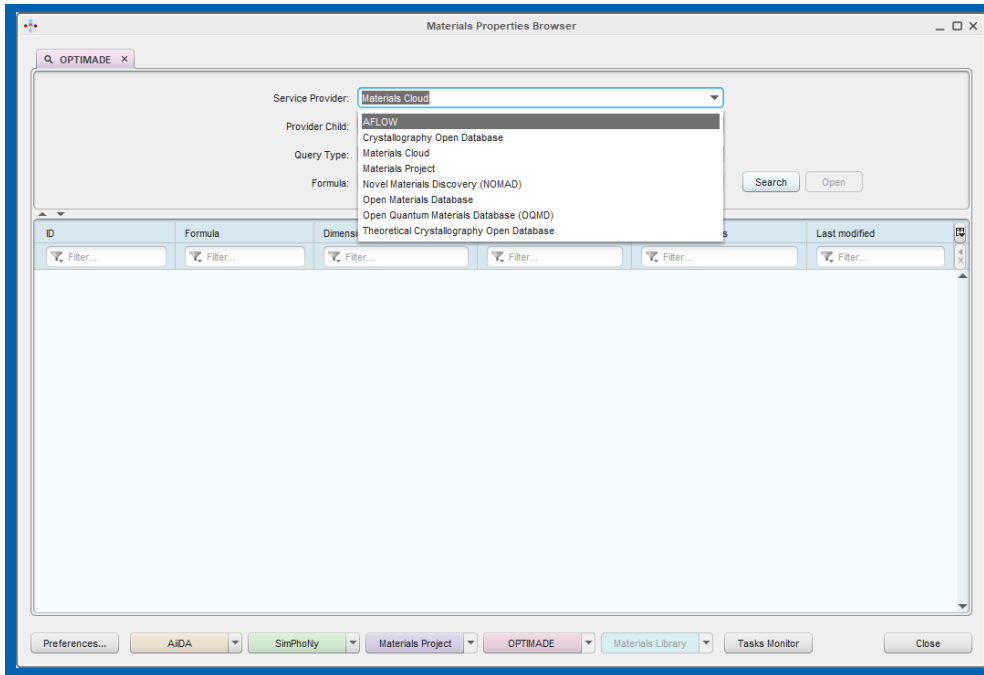


Figure 10: Screenshot of the OPTIMADE interface. The user can choose among many database providers in the OPTIMADE consortium.

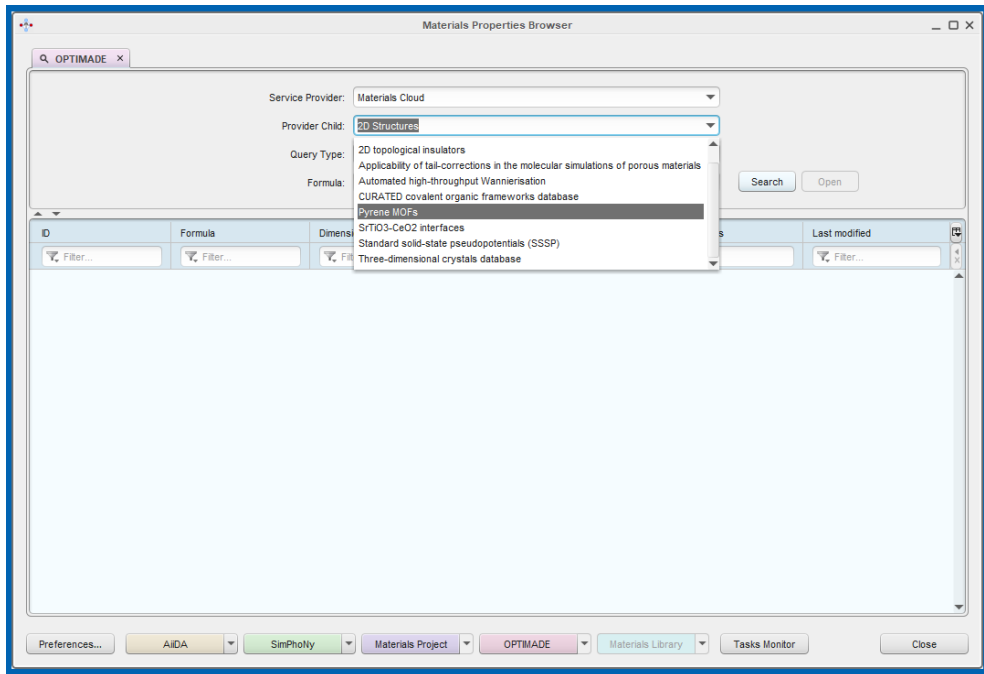
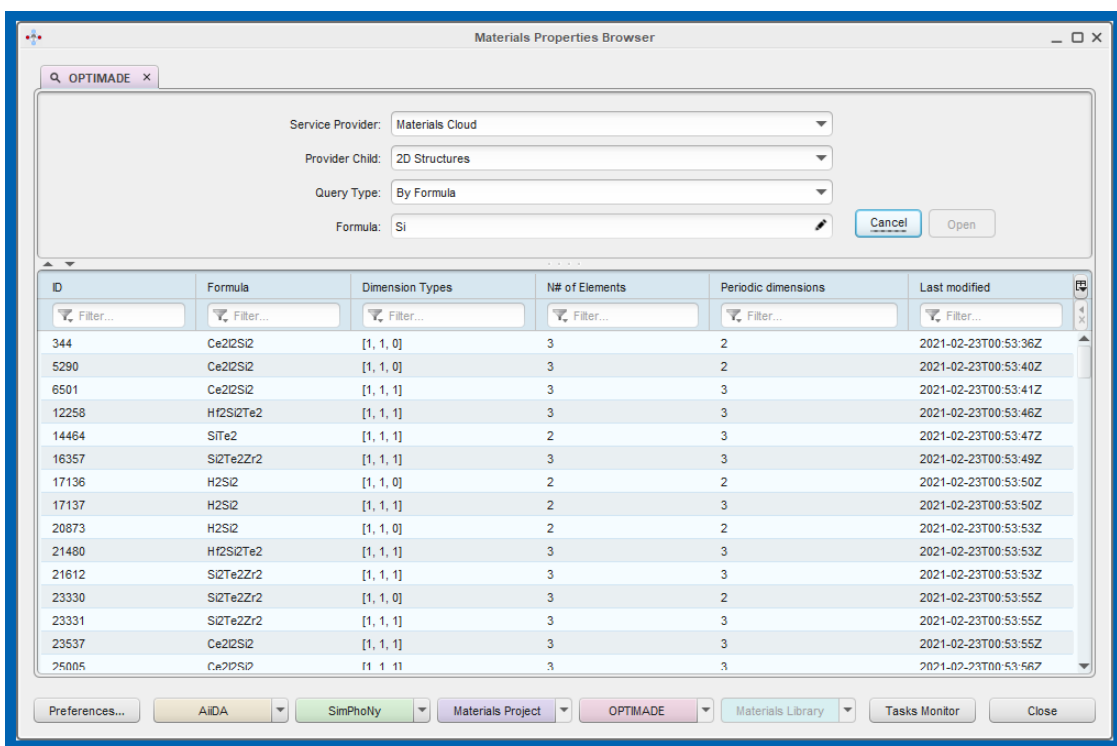


Figure 11: Screenshot of the OPTIMADE interface. Some providers, such as the Materials Cloud, serve several databases, which are displayed to the user.

https://aiida.materialscloud.org/2dstructures/optimade/structures?filter=chemical_formula_descriptiveCONTAINS%22ANDelementsHAS%22Si%22.

In this example, we searched for compounds with Si in the "2D Structure Database" of the Materials Cloud provider. The result is reported in Fig. 12, a screenshot of the Ginestra-AiiDA plugin interface for the OPTIMADE API.



ID	Formula	Dimension Types	# of Elements	Periodic dimensions	Last modified
344	Ce2QSi2	[1, 1, 0]	3	2	2021-02-23T00:53:36Z
5290	Ce2QSi2	[1, 1, 0]	3	2	2021-02-23T00:53:40Z
6501	Ce2QSi2	[1, 1, 1]	3	3	2021-02-23T00:53:41Z
12258	Hf2Si2Te2	[1, 1, 1]	3	3	2021-02-23T00:53:46Z
14464	SiTe2	[1, 1, 1]	2	3	2021-02-23T00:53:47Z
16357	Si2Te2Zr2	[1, 1, 1]	3	3	2021-02-23T00:53:49Z
17136	H2Si2	[1, 1, 0]	2	2	2021-02-23T00:53:50Z
17137	H2Si2	[1, 1, 1]	2	3	2021-02-23T00:53:50Z
20873	H2Si2	[1, 1, 0]	2	2	2021-02-23T00:53:53Z
21480	Hf2Si2Te2	[1, 1, 1]	3	3	2021-02-23T00:53:53Z
21612	Si2Te2Zr2	[1, 1, 1]	3	3	2021-02-23T00:53:53Z
23330	Si2Te2Zr2	[1, 1, 0]	3	2	2021-02-23T00:53:55Z
23331	Si2Te2Zr2	[1, 1, 1]	3	3	2021-02-23T00:53:55Z
23537	Ce2QSi2	[1, 1, 1]	3	3	2021-02-23T00:53:55Z
25005	Ce2QSi2	[1, 1, 1]	3	3	2021-02-23T00:53:56Z

Figure 12: Screenshot of the OPTIMADE tab in the Ginestra-AiiDA plugin. The figure shows the result of a query for all structures containing silicon in the "2D Structures" database in the Materials Cloud.

6. Conclusions

This document summarizes the main implementation work to interconnect the IM2D platform with material databases. The implementation is conceived to follow the OPTIMADE standards, in order to allow the user to access all databases in the OPTIMADE consortium. We presented the several technical steps that realize the traceable data pipeline within IM2D. The queries to databases start from the graphical front-end of the Ginestra-AiiDA plugin, the GUI of the overall IM2D platform, and exploits all the interoperable functionalities of both AiiDA and SimPhoNy infrastructure.

References

- [1] <https://www.materialscloud.org>
- [2] <https://materialsproject.org>
- [3] <http://www.crystallography.net/cod>
- [4] Casper W. Anderson et al, arXiv:2103.02068v1 (2021); <https://www.optimade.org>
- [5] <https://www.appliedmaterials.com/products/applied-mdlx-ginestra-simulation-software>
- [6] <https://www.aiida.net>
- [7] <https://arxiv.org/abs/1504.01163>
- [8] <https://www.simphony-project.eu>
- [9] <https://www.quantum-espresso.org>
- [10] <https://departments.icmab.es/leem/siesta>
- [11] <https://emmc.info/emmo-info>
- [12] <https://github.com/aiidateam/aiida-codtools>
- [13] <http://www.re3data.org>
- [14] <https://fairsharing.org>

ACRONYMS

ADES - Automation, Data, Environment, and Sharing

API - Application Programming Interface

COD - Crystallography Open Database

CIF - Crystallographic Information File

DFT - Density Functional Theory

DOI - Digital Object Identifier

EMMO - European Materials Modelling Ontology

FAIR - Findable, Accessible, Interoperable, and Reusable

GUI - Graphical User Interface

IM2D - Interoperable Materials-To-Device

NOMAD - Novel Materials Discovery

OPTIMADE - Open Databases Integration for Materials Design

OQMD - Open Quantum Materials Database

REST - Representational State Transfer

URI - Uniform Resource Identifier

UUID - Universally Unique Identifier