Deliverable D2.6 - Tools for simulation and modelling workflows

Responsible Partner:	EPFL	24 Nov 2024
Contributor(s):	Nicola Marzari	24 Nov 2024
Reviewer(s):	Bjørn Tore Løvfall (SINTEF), Silvia Chiacchiera (UKRI)	28 Nov 2024
Coordinator:	CMCL Innovations	29 Nov 2024
Dissemination Level:	Public	
Due Date:	M48 (Dec. 2024)	
Submission Date:	29.11.2024	

Project Profile

Programme	**Horizon 2020**
Call	**H2020-NMBP-TO-IND-2020-twostage**
Topic	**DT-NMBP-40-2020** **Creating an open marketplace for industrial data (RIA)**
Project number	**953163**
Acronym	**DOME 4.0**
Title	**Digital Open Marketplace Ecosystem 4.0**
Start Date	**December 1st, 2020**
Duration	**48 months**

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 953163. It is the property of the DOME 4.0 consortium and do not necessarily reflect the views of the European Commission.

Document History

Version	Date	Author	Remarks
V1.0	24 Nov 2024	Nicola Marzari	First draft
V1.1	26 Nov 2024	Adham Hashibon	Review
V1.2	27 Nov 2024	Silvia Chiacchiera	Review
V2.0	28 Nov 2024	Nicola Marzari	Revision
V2.1	29 Nov 2024	Amit Bhave	Final review
V3.0	29 Nov 2024	Willem van Dorp	Finalization

Publishable Summary

The DOME 4.0 platform gives its users the capability to interrogate some external databases, and to perform simulations on these structures. At the time of writing, the DOME 4.0 platform is connected to 14 external data sources, out of which two implement the OPTIMADE API specification (namely, Materials Project and Crystallography). This takes place through the API specification registered in DOME 4.0, and compliance of a data provider with it is shown as part of the high-level provider metadata that a user is presented with when clicking on a simple search result (via the "Conforms to standard" property).

Also, the DOME 4.0 platform allows to perform automatically and reliably state-of-the-art atomistic simulations of materials and of their properties. This is achieved through the online deployment of graphical user interfaces (GUIs) as front-ends to complex and sophisticated back-end capabilities. The front-end is managed by AiiDAlab, a JupyterLab-like interface to AiiDA, the workflow engine that is able to orchestrate complex protocols for materials simulations.

The capabilities that are exposed are those of data handshaking between all the computational repositories abiding to the specifications of the universal OPTIMADE API (59,521,800 structures in 29 databases), and the simulation capabilities using density functional theory of materials properties – from geometry optimization to electronic band structures and density of states, phonon calculations and infrared and Raman spectra, and x-ray spectroscopies such as XPS and XANES.

Executive Summary

The DOME 4.0 platform gives its users the capability to interrogate external databases, and to perform simulations on these structures. At the time of writing, the DOME 4.0 platform is connected to 14 external data sources, out of which two implement the OPTIMADE API specification (namely, Materials Project and Crystallography). This takes place through the API specification registered in DOME 4.0, and compliance of a data provider with it is shown as part of the high-level provider metadata that a user is presented with when clicking on a simple search result (via the "Conforms to standard" property).

Also, the DOME 4.0 platform allows to perform automatically and reliably state-of-the-art atomistic simulations of materials and of their properties. This is achieved through the online deployment of graphical user interfaces (GUIs) as front-ends to complex and sophisticated back-end capabilities. The front-end is managed by AiiDAlab, a JupyterLab-like interface to AiiDA, the workflow engine that is able to orchestrate complex protocols for materials simulations.

The capabilities that are exposed are those of data handshaking between all the computational repositories abiding to the specifications of the universal OPTIMADE API (59,521,800 structures in 29 databases), and the simulation capabilities using density functional theory of materials properties – from geometry optimization to electronic band structures and density of states, phonon calculations and infrared and Raman spectra, and x-ray spectroscopies such as XPS and XANES.

The software design specification and how to use the tools and services template developed are already covered in D3.7 and are not repeated here; the tools for simulation and modelling workflows are connected to the DOME 4.0 platform by creating a wrapper for the tool using the templated outlined in D3.7, and registering it on the DOME 4.0 platform, creating an API specification and linking this to the relevant connectors. This is covered extensively in D2.5.

Table of Contents

List of Figures

List of Tables

No table of figures entries found.

1. The AiiDALab Frontend

AiiDAlab (https://www.aiidalab.net/) is an open-source web platform designed to facilitate the development, execution, and sharing of scientific workflows, particularly in computational materials science. It integrates the AiiDA workflow manager with the Jupyter environment, enabling users to create and interact with computational workflows through user-friendly graphical interfaces.

Key Features of AiiDAlab:

- **Workflow Automation and Management:** AiiDAlab leverages AiiDA's capabilities to automate complex computational workflows, ensuring efficient execution and management of tasks.
- **Provenance Tracking:** The platform provides comprehensive tracking of data provenance, allowing users to trace the origin and evolution of data throughout the computational process, thereby enhancing reproducibility and transparency.
- **App Store and Customization:** AiiDAlab features an app store where users can install, update, and remove various applications tailored to specific computational tasks. Developers can also create and share their own apps, fostering a collaborative environment.
- **Integration with Simulation Codes:** Through AiiDA's plugin ecosystem, AiiDAlab supports integration with a wide range of simulation codes, providing a unified interface for diverse computational tools.
- **Deployment Options:** Users can deploy AiiDAlab locally or access official cloud-backed deployments maintained for researchers affiliated with specific projects. For instance, the Materials Cloud deployment is available to researchers associated with Materials Cloud partners.
- **Getting Started with AiiDAlab:** To begin using AiiDAlab, users can refer to the official documentation, which provides comprehensive guides on installation, usage, and app development. The platform is released under the MIT license and is actively maintained, with a supportive community available for discussions and support through the Discourse forum.

By combining workflow automation, provenance tracking, and a customizable app ecosystem, AiiDAlab serves as a powerful tool for researchers aiming to streamline and share their computational workflows in a reproducible and transparent manner.

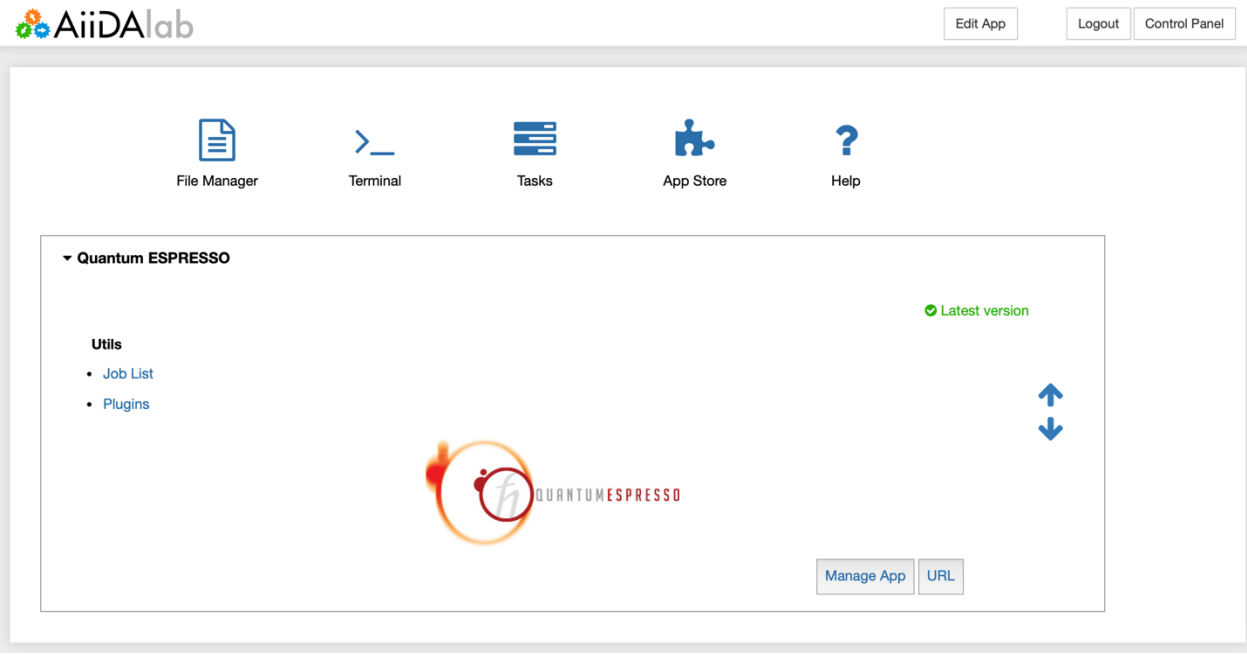


Figure 1: Landing page for AiiDAlab.

2. The AiiDA back-end

AiiDA (Automated Interactive Infrastructure and Database for Computational Science; https://aiida.net/) is an open-source, Python-based platform designed to facilitate and manage computational workflows, particularly in the field of computational materials science and chemistry. Its robust framework enables automation, data management, and reproducibility of complex scientific computations. Below is an extended and detailed overview of AiiDA, covering its architecture, features, ecosystem, and applications. Core Components of AiiDA are:

1. Workflow Management

AiiDA provides a flexible framework for the automation and management of workflows, enabling the orchestration of simulations and computations.

WorkChain Framework: Users can define complex workflows, known as WorkChains, which automate sequences of computations. These are reusable and support conditional branching, error handling, and dynamic inputs.

Automatic Error Handling: AiiDA workflows include mechanisms for catching and recovering from errors in simulations, such as retrying a task or adjusting parameters.

2. Data Provenance and Reproducibility

One of AiiDA's standout features is its focus on ensuring reproducibility through detailed tracking of the origin and evolution of data.

Provenance Graph: A directed acyclic graph (DAG) is used to represent the relationships between inputs, outputs, and processes. Every computation and dataset is linked to its source, allowing users to trace the entire history of a result.

Database Backend: AiiDA stores all metadata and provenance information in a relational database (PostgreSQL by default). This structure ensures scalability and efficient querying of data.

3. Plugin Ecosystem

AiiDA's modularity is achieved through its plugin system, which allows seamless integration with external simulation codes.

Code Plugins: AiiDA supports a wide range of scientific codes (e.g., Quantum ESPRESSO, VASP, CP2K, Gaussian) via community-developed plugins. These plugins standardize the submission, monitoring, and parsing of jobs across different computational resources.

Pip Installable: Plugins are Python packages distributed via PyPI, making it easy to install and update them.

4. Computational Resource Management

AiiDA simplifies interactions with High-Performance Computing (HPC) resources and cloud infrastructures.

Job Scheduler Abstraction: AiiDA abstracts the specifics of job schedulers (e.g., SLURM, PBS, Torque), allowing users to submit jobs without needing in-depth knowledge of the resource.

Transport Layers: It includes support for SSH and other protocols to facilitate remote job submission, retrieval, and monitoring.

Computer Abstraction: AiiDA maintains a registry of computational resources, including their configurations, ensuring consistency and reproducibility across simulations.

5. Querying and Sharing Data

AiiDA provides tools for querying stored data and sharing it with collaborators.

QueryBuilder API: AiiDA offers a powerful and intuitive API for querying the database. Users can search for workflows, nodes, or specific data attributes using Python syntax.

Export and Import: Data and workflows can be exported to share with collaborators, preserving all provenance information. Importing these datasets is seamless and integrates into the recipient's database.

6. Extensibility and Integration

AiiDA is designed to integrate into broader computational and scientific ecosystems.

REST API: AiiDA includes a RESTful API for interacting with the database programmatically, enabling web-based tools or external software to query AiiDA data.

Visualization Tools: The platform provides tools to visualize provenance graphs and workflow states, aiding in debugging and presentation.

Integration with Jupyter: AiiDA workflows and data can be managed and visualized within Jupyter notebooks, making it ideal for interactive exploration and analysis.

7. Key Advantages

Reproducibility: Through automated provenance tracking, AiiDA ensures every result is reproducible, from input parameters to the computational environment.

Scalability: Designed to handle millions of computational tasks, AiiDA is suited for large-scale high-throughput studies.

Collaboration: The ability to export, share, and import workflows promotes collaboration and transparency among researchers.

Flexibility: AiiDA is domain-agnostic, meaning it can be adapted to various fields beyond materials science, including chemistry and physics.

AiiDA is under active development, with new features, plugins, and integrations regularly released. Its modular design ensures that it can adapt to advancements in computational science, such as integration with machine learning workflows or quantum computing tools.

In summary, AiiDA is a powerful, versatile tool for computational research, offering a comprehensive solution for managing and automating workflows while ensuring data provenance and reproducibility. It empowers researchers to tackle complex scientific challenges more efficiently and transparently.

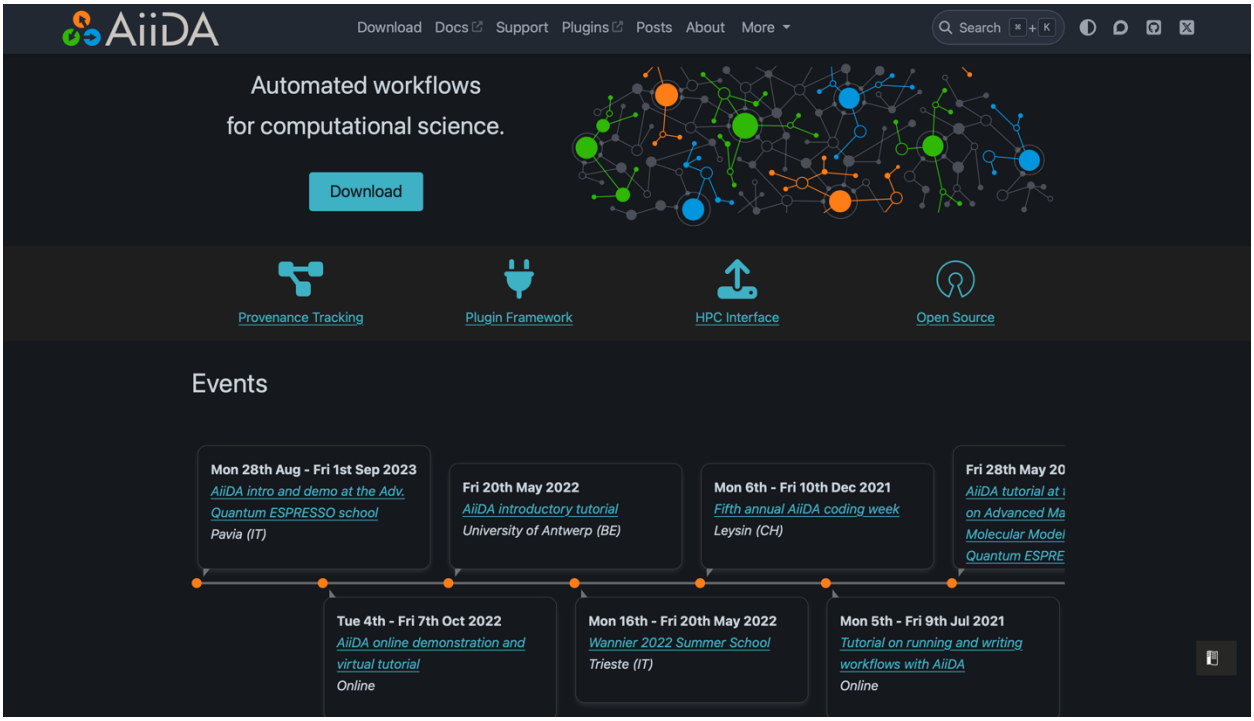


Figure 2: The landing page for AiiDA.

3. Simulation and Modeling Tools: the Optimade universal REST API for data searches

The Open Databases Integration for Materials Design (OPTIMADE; https://www.optimade.org/) consortium has developed a standardized RESTful API to facilitate interoperability among materials databases. This API specification enables uniform access to diverse materials data repositories, promoting data sharing and integration across platforms.

3.1 Key Features of the OPTIMADE REST API

- Standardized Data Access: The API defines a common protocol for querying materials databases, allowing users to retrieve data using consistent methods regardless of the underlying database structure.
- Interoperability: By adhering to the OPTIMADE specification, different materials databases can interoperate seamlessly, enabling combined data analyses and fostering collaborative research.
- Extensibility: The API is designed to be extensible, allowing database providers to include custom properties and data types while maintaining compliance with the core specification.
- Filtering and Querying: OPTIMADE supports complex filtering capabilities, enabling users to perform detailed searches based on various criteria such as chemical composition, structural properties, and more.
- Versioning: The API includes versioning mechanisms to ensure backward compatibility and to facilitate the adoption of new features without disrupting existing implementations.

3.2 Core Components

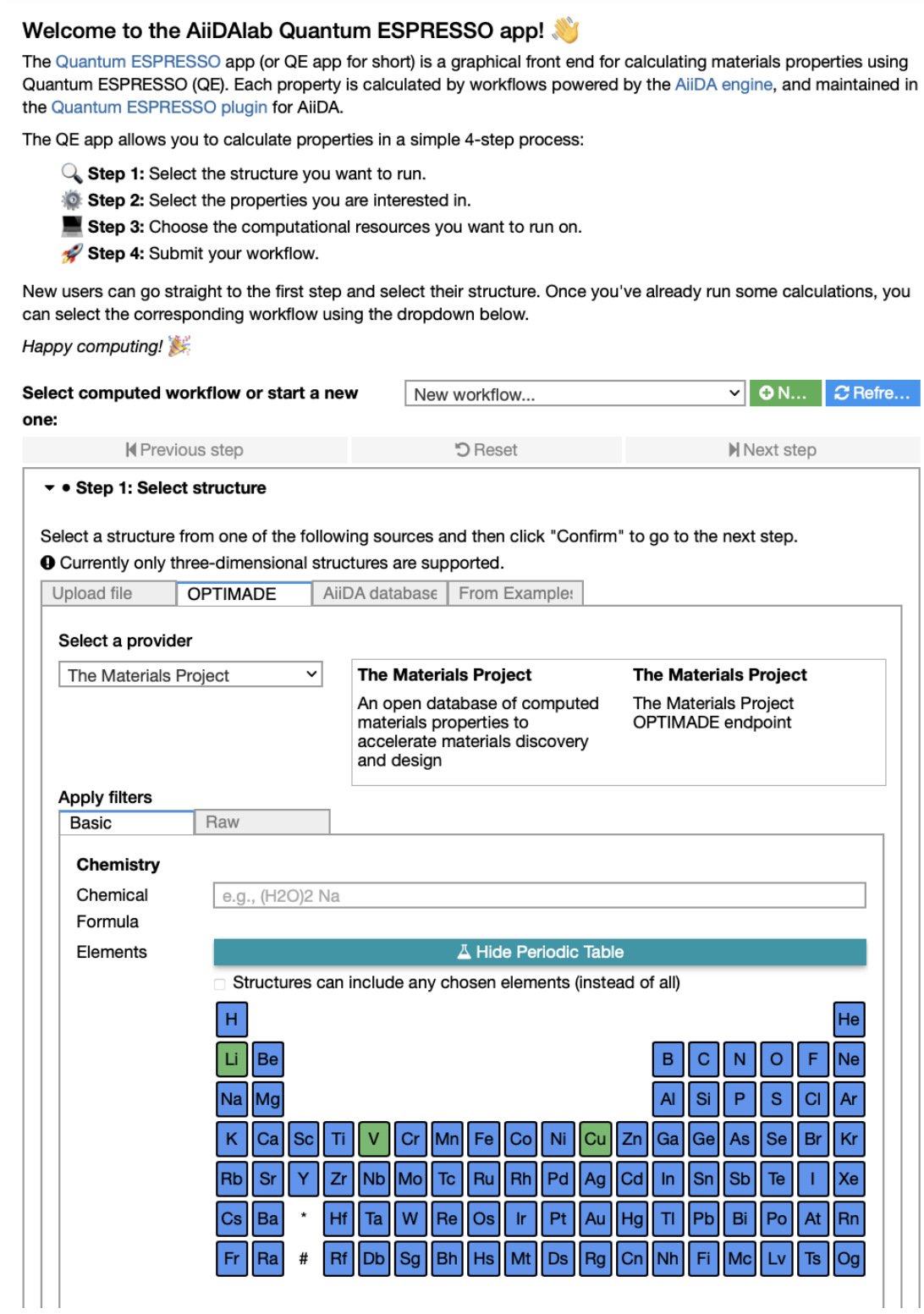
1) Base URL: The API is accessed via a base URL provided by each database, serving as the entry point for all API requests.
2) Endpoints: Standardized endpoints are defined for accessing various types of data, such as structures, calculations, and references. These endpoints provide uniform access points across different databases.
3) Query Parameters: The API utilizes query parameters to refine searches and filter results, allowing users to specify criteria for data retrieval.
4) Response Format: Responses are formatted in JSON, adhering to a standardized schema that ensures consistency and ease of parsing across different implementations.

3.3 Implementation and Adoption

Several materials databases have implemented the OPTIMADE API, enabling researchers to access a wide array of materials data through a unified interface. The consortium maintains a list of providers and their respective implementations, fostering a collaborative ecosystem for materials data sharing.

3.4 Resources and Documentation

The full API specification, including detailed definitions of endpoints, query parameters, and response formats, is available on the official OPTIMADE website; the OPTIMADE consortium's GitHub repository hosts the specification documents, schemas, and related resources for developers and contributors. Last, the consortium encourages community involvement and provides forums for discussion, feedback, and support to facilitate the adoption and evolution of the API.


Welcome to the AiiDAlab Quantum ESPRESSO app!
The Quantum ESPRESSO app (or QE app for short) is a graphical front end for calculating materials properties using Quantum ESPRESSO (QE). Each property is calculated by workflows powered by the AiiDA engine, and maintained in the Quantum ESPRESSO plugin for AiiDA.
The QE app allows you to calculate properties in a simple 4-step process:
Step 1: Select the structure you want to run.
Step 2: Select the properties you are interested in.
Step 3: Choose the computational resources you want to run on.
Step 4: Submit your workflow.
New users can go straight to the first step and select their structure. Once you've already run some calculations, you can select the corresponding workflow using the dropdown below.
Happy computing!
Select computed workflow or start a new one:
New workflow...
N...
Refre...
Previous step
Reset
Next step
Step 1: Select structure
Select a structure from one of the following sources and then click "Confirm" to go to the next step.
Currently only three-dimensional structures are supported.
Upload file
OPTIMADE
AiiDA database
From Example
Select a provider
The Materials Project
The Materials Project
An open database of computed materials properties to accelerate materials discovery and design
The Materials Project
The Materials Project OPTIMADE endpoint
Apply filters
Basic
Raw
Chemistry
Chemical Formula
e.g., (H2O)2 Na
Elements
Hide Periodic Table
Structures can include any chosen elements (instead of all)
H
He
Li
Be
B
C
N
O
F
Ne
Na
Mg
Al
Si
P
S
Cl
Ar
K
Ca
Sc
Ti
V
Cr
Mn
Fe
Co
Ni
Cu
Zn
Ga
Ge
As
Se
Br
Kr
Rb
Sr
Y
Zr
Nb
Mo
Tc
Ru
Rh
Pd
Ag
Cd
In
Sn
Sb
Te
I
Xe
Cs
Ba
*
Hf
Ta
W
Re
Os
Ir
Pt
Au
Hg
Tl
Pb
Bi
Po
At
Rn
Fr
Ra
#
Rf
Db
Sg
Bh
Hs
Mt
Ds
Rg
Cn
Nh
Fi
Mc
Lv
Ts
Og

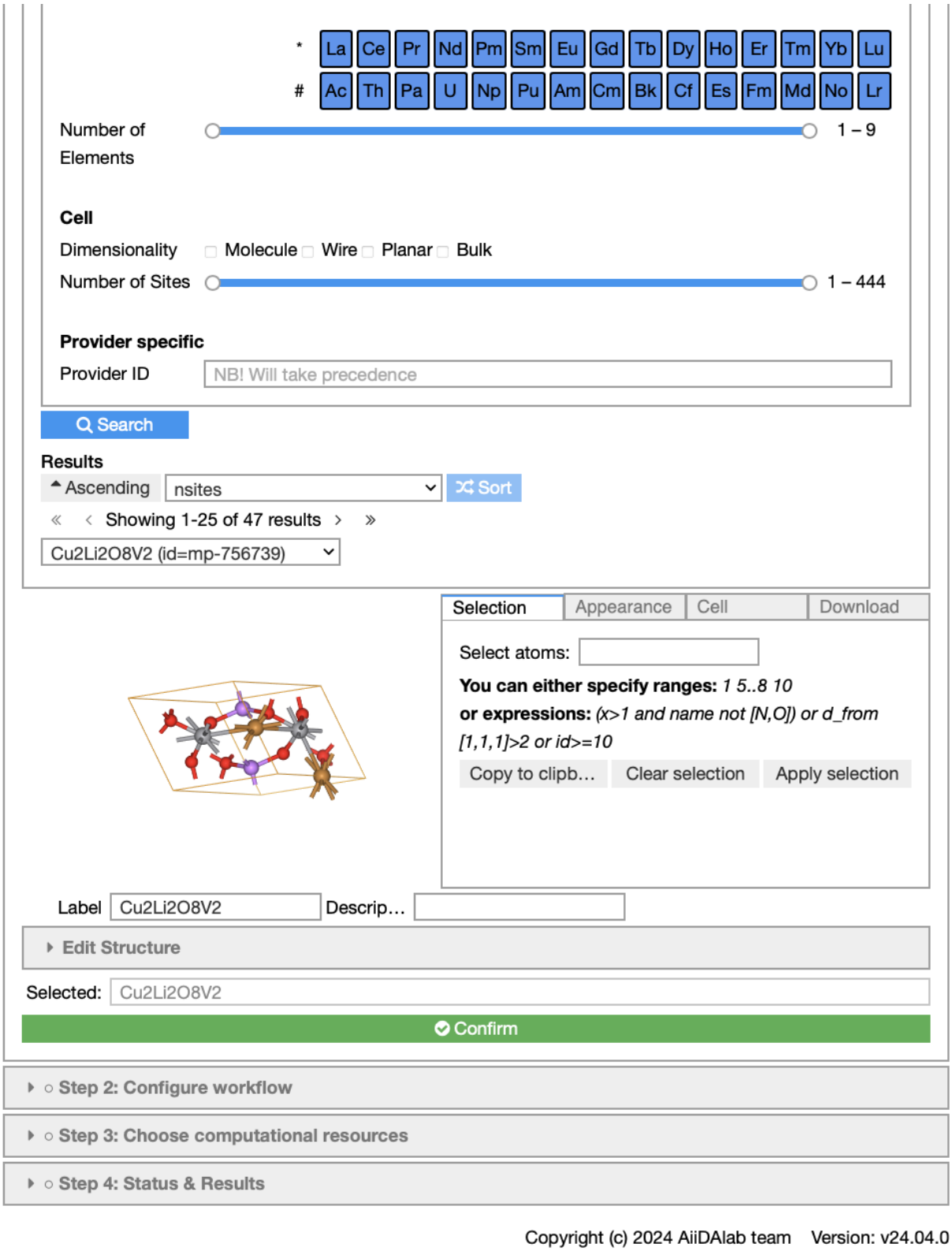


Figure 3: Search of a Li-ion cathode materials through the Optimade universal REST API, having chosen the Materials Project as a structure provider.

4. Simulation and Modeling Tools: the Quantum ESPRESSO App for materials simulations

The Quantum ESPRESSO App (https://aiidalab-qe.readthedocs.io) within AiiDAlab offers a comprehensive, user-friendly interface for conducting a wide array of quantum mechanical simulations using the Quantum ESPRESSO suite. By integrating AiiDA's workflow management capabilities, the app streamlines complex computational tasks, ensuring reproducibility and efficiency.

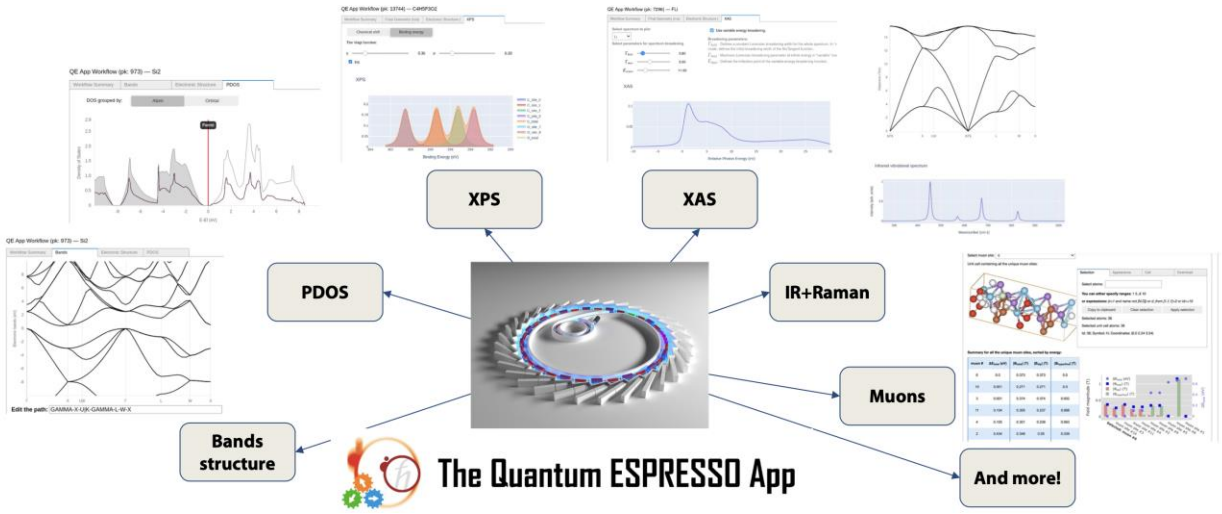


Figure 4: simulation capabilities of the Quantum ESPRESSO App.

Below is an extended overview of its capabilities:

4.1 Graphical User Interface (GUI)

The app provides an intuitive Jupyter-based GUI, allowing users to set up, monitor, and analyze simulations without extensive command-line interactions. This interface simplifies the process of configuring input parameters, selecting computational resources, and visualizing results.

4.2 Workflow Automation

Leveraging AiiDA's powerful workflows, the app automates various computational processes, including:

- Structure Optimization: Automated relaxation of atomic positions and cell parameters to find the ground-state configuration.
- Electronic Structure Calculations: Computation of electronic properties such as band structures and density of states.
- Phonon Calculations: Evaluation of vibrational properties and phonon dispersions.

- Spectroscopy Simulations: Simulation of X-ray absorption (XANES) and photoelectron (XPS) spectra.

4.3 Integration with Quantum ESPRESSO

The app seamlessly interfaces with Quantum ESPRESSO, enabling users to perform:

1) Self-Consistent Field (SCF) Calculations: Determination of ground-state electronic densities.
2) Non-SCF Calculations: Computation of properties requiring fixed electronic densities.
3) Density of States (DOS) and Projected DOS: Analysis of electronic states distribution.
4) Band Structure Calculations: Evaluation of electronic band dispersions.
5) Phonon Dispersion Calculations: Assessment of vibrational properties.
6) XANES and XPS Simulations: Modeling of spectroscopic properties.

4.4 Remote Computational Resource Management

Users can configure the app to submit and manage jobs on remote high-performance computing clusters, facilitating large-scale simulations. The app supports:

- Remote Code Setup: Configuration of computational codes on external resources.
- Job Submission and Monitoring: Automated submission and real-time tracking of computational tasks.
- Data Retrieval: Efficient transfer of results back to the local environment for analysis.

4.5 Advanced Settings and Customization

The app offers advanced configuration options, including:

- Protocol Selection: Choice between predefined protocols (e.g., 'fast', 'moderate') to balance accuracy and computational cost.
- Exchange-Correlation Functional Override: Ability to specify different functionals for calculations.
- K-Point Grid Customization: Adjustment of k-point sampling for Brillouin zone integration.
- Pseudopotential Selection: Choice of appropriate pseudopotentials for elements in the simulation.

4.6 Tutorials and How-To Guides

Comprehensive documentation is available at https://aiidalab-qe.readthedocs.io, including:

- Basic Tutorials: Introduction to the app's features and functionalities.

- Advanced Tutorials: In-depth guides on complex simulations and analyses.
- How-To Guides: Step-by-step instructions for specific tasks, such as setting up remote computations and importing structures.

The app is actively maintained, with regular updates introducing new features and improvements. Users can stay informed about the latest developments through the app's blog and release notes.

In summary, the Quantum ESPRESSO App in AiiDAlab serves as a powerful tool for researchers, combining the computational capabilities of Quantum ESPRESSO with AiiDA's workflow management to provide a streamlined, reproducible, and efficient environment for quantum mechanical simulations.

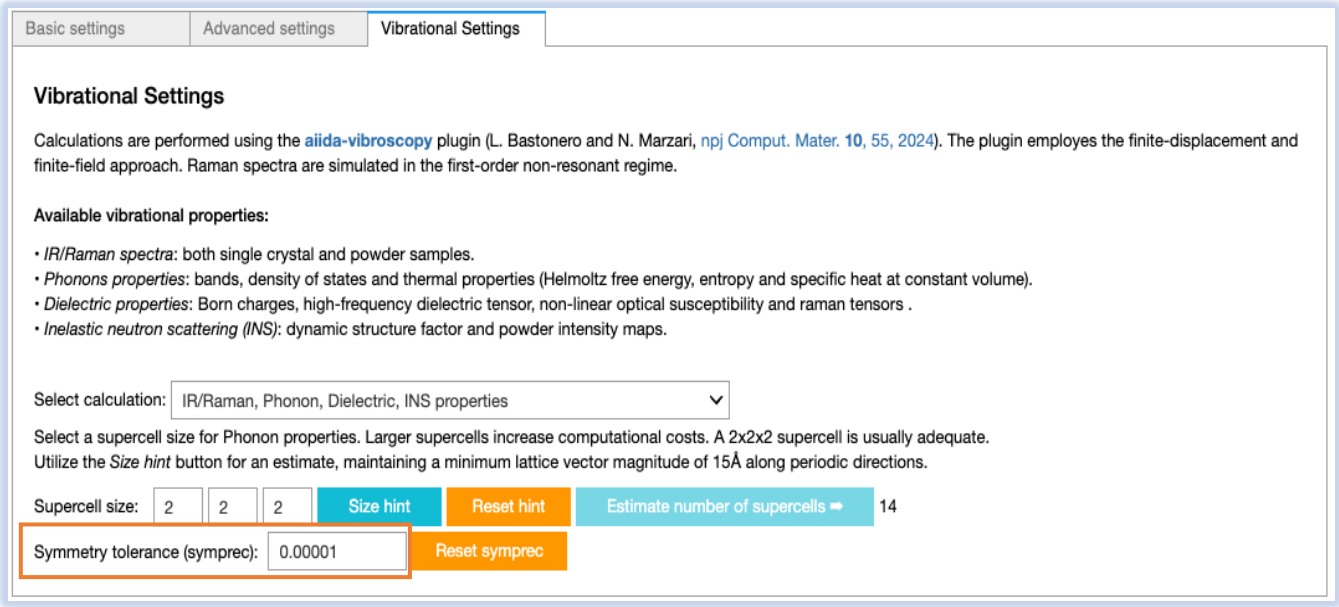


Figure 5: set up for the calculation of infrared and Raman properties.

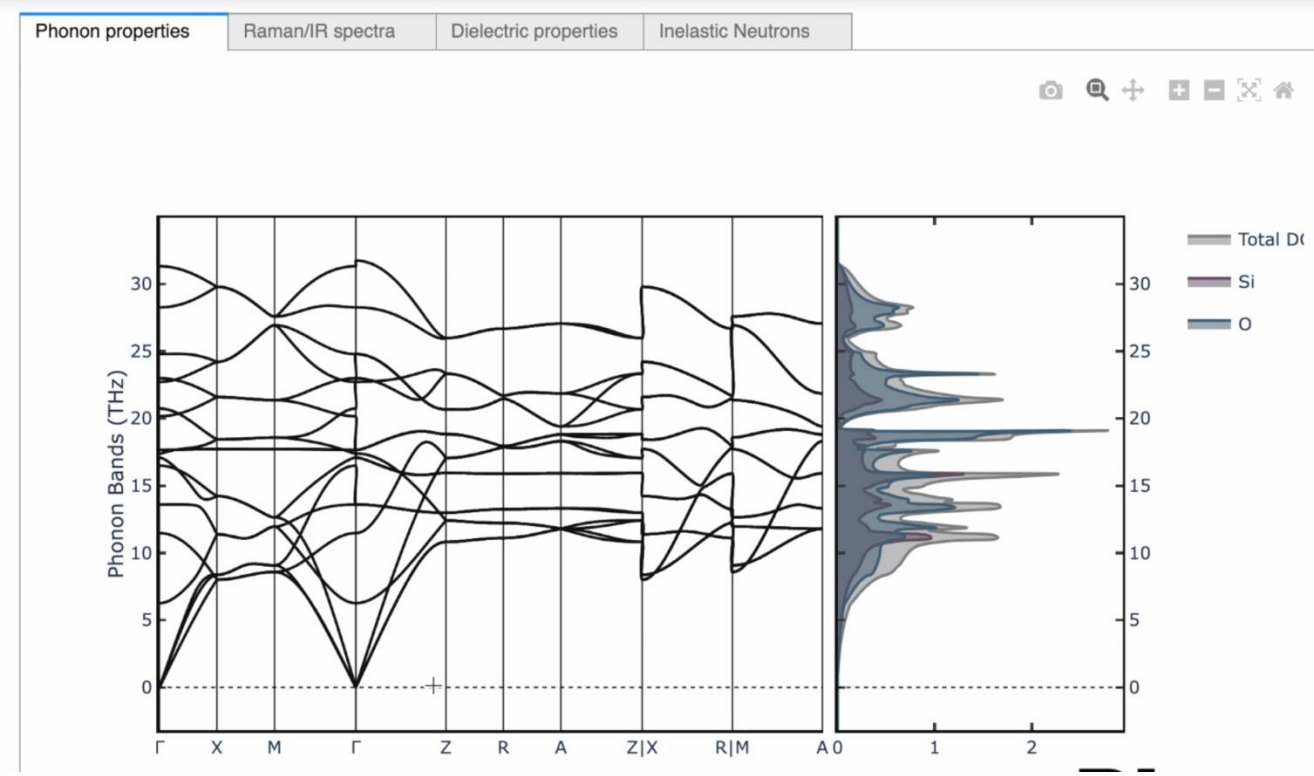


Figure 6: calculation of the phonon dispersions and phonon density of states.

5. Provenance and clearing house system integration

The integration of semantic technology into the DOME 4.0 back end means that every entity, be it a resource of any kind, specifically a dataset, and an actor/agent, such as a data set owner, a tool/service or data prosumer, are connected with in a semantic, meaningful way to all other relevant resources and actions. Provenance, which traces the evolution or history of a resource, is inherently part of the clearing house, as information on who accessed what and when (provenance) and whether they have access rights to (clearing house) are very related. In traditional systems, i.e., those that do not integrate semantics deeply, the business data are stored in different data bases and and additional layer is needed to keep these data in sync and maintain the relations. Having both data in one back end, reduced this complexity and enables seamless integration. The Semantically enriched Clearing House and Provenance system then enables integrated workflows, where the access request to a resource trigers a workflow in which the user can procure access to, say, a dataset, and at the same time document this in a provenance system.

The workflow depicted in Fugure 7 below, represents the interaction of agents, datasets, and activities in a semantically enriched clearinghouse system. Let's suppose that two users, we name them, "David" and "Marie", log into the system, generating timestamped login activities linked to their user profiles. David conducts a search using the system's connector and external platform, producing a search record containing the keyword "Vehicle" and associating it with Dataset1. Marie uploads data to the system, creating an upload record linked to Dataset2. David submits an access request for Dataset1, triggering interactions with the clearinghouse to manage permissions and track provenance. A contract is generated, including details like an NDA, duration, and access rights, followed by a payment from David, which grants him access to Dataset1. This process is documented in a transaction record, and a monetary transaction is processed to enable David's access. Finally, Dataset1 access is granted, associating David with the dataset and completing the workflow. The system ensures semantic integration, linking users, datasets, activities, and provenance for efficient and compliant resource management.

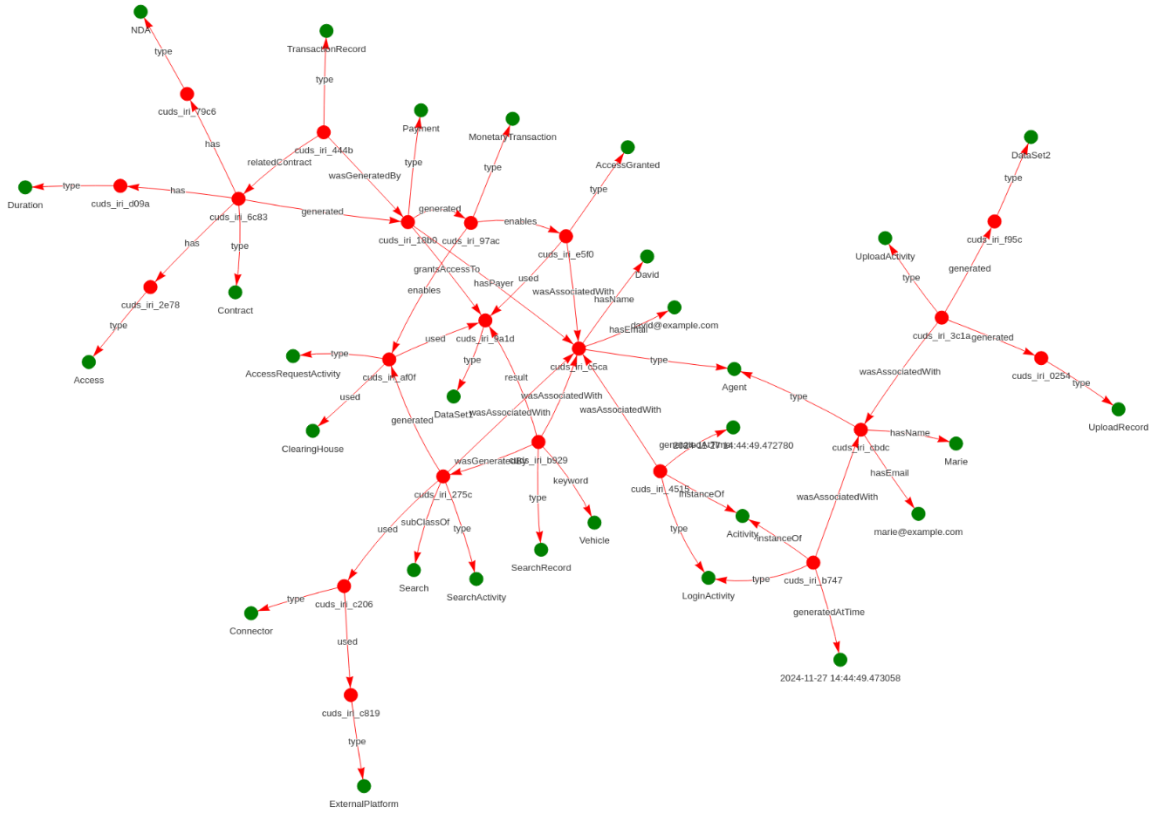


Figure 7:Visualization of Dome40_workflow.html

6. Deviations from Annex 1

There are no deviations from Annex 1

7. Conclusions

The DOME 4.0 platform provides mechanisms connect to external data sources and to link dataset and tools. Here we focus on the possibility to search across computational repositories worldwide for materials structures and materials data, together with simulation capabilities to perform atomistic and electronic-structure modelling using state-of-the-art open-source software. Adherence to a given (standard) API specification has a twofold benefit in DOME 4.0 ecosystem: first, it facilitates the creation of new DOME 4.0 connectors to external data sources, and, secondly, it allows to create a link between datasets and modelling tools. These two general mechanisms are described in detail in DOME4.0 D3.4 and D2.5 (also submitted in Nov 2024); the present document has focused on the benefits and functionalities of the OPTIMADE API and the AiiDALab tool, two major assets in the field of materials data and materials modelling. The capabilities offered by the semantic approach within DOME 4.0, in particular via the provenance and clearing house features are key to systematically track how the users interact with the entities of the DOME 4.0 data sharing ecosystem.

8. Lessons learnt

Intuitive user interfaces, user friendliness, and robust and reliable workflows are key to the success of the platform and to returning users that engage with the tools that are being offered. In this line, an important functionality is a seamless connection of the data to relevant tools, e.g., to visualize, transform or take data as an input of a workflow. By providing these type connections automatically (e.g., as clickable links), DOME enhances both sides (data and tools) and lowers the usability barrier.

When it comes to data sharing across domains or sectors, it is important to understand the exact sources of the data transactions and the way in which data transaction activities evolve. Using a semantic approach within the DOME 4.0 backend, this capability of offered via the two critical digital services, namely, the provenance and clearing house. The uncertainty around the ownership of the resulting datasets is intertwined with the business models of a data ecosystem, and paves the way for further research and development in this area in the future.

9. Acknowledgement

The author(s) would like to thank the partners in the project for their valuable comments on previous drafts and for performing the review.

Project partners:

#	Type	Partner	Partner full name
1	SME	CMCL	Computational Modelling Cambridge Limited
2	Research	FHG	Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung E.V.
3	Research	INTRA	Intrasoft International SA
4	University	UNIBO	Alma Mater Studiorum – Universita di Bologna
5	University	EPFL	Ecole Polytechnique Federale de Lausanne
6	Research	UKRI	United Kingdom Research and Innovation
7	Large Industry	SISW	Siemens Industry Software NV
8	Large Industry	BOSCH	Robert Bosch GmbH
9	SME	UNR	Uniresearch B.V.
10	Research	SINTEF	SINTEF AS
11	SME	CNT	Cambridge Nanomaterials Technology LTD
12	University	UCL	University College London

This document is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 953163. It is the property of the DOME 4.0 consortium and do not necessarily reflect the views of the European Commission.

10. Table of Abbreviations

Abbreviation	Explanation
XPS	X-ray photoelectron spectroscopy
XANES	X-ray absorption near edge structure