

Article

IEMAP—The Italian Energy Materials Acceleration Platform—And Its Application to Cathode Materials for Batteries

Sergio Ferlito ¹, Massimo Celino ^{2,*}, Marco Catillo ³, Serena D'Onofrio ⁴, Simone Giusepponi ³, Sara Marchio ³, Francesco Buonocore ^{3,*} and Giovanni Ponti ²

¹ Division for the Smart Sector Integration and Distributed Generation from RES (SSI), Department for Energy Technologies and Renewable Sources (TERIN), ENEA—Italian National Agency for New Technologies, Energy and Sustainable Economic Development, C.R. Portici, P.le E. Fermi 1, 80055 Portici, Italy; sergio.ferlito@enea.it

² Division for the Development of Systems for Informatics and ICT (ICT), Department for Energy Technologies and Renewable Sources (TERIN), ENEA—Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Sede Centrale, Lungotevere Thaon di Revel, 76, 00196 Rome, Italy; giovanni.ponti@enea.it

³ Division for the Development of Systems for Informatics and ICT (ICT), Department for Energy Technologies and Renewable Sources (TERIN), ENEA—Italian National Agency for New Technologies, Energy and Sustainable Economic Development, C.R. Casaccia, Via Anguillarese 301, 00123 Rome, Italy; marco.catillo@enea.it (M.C.); simone.giusepponi@enea.it (S.G.); sara.marchio@enea.it (S.M.)

⁴ Division for the Development of Systems for Informatics and ICT (ICT), Department for Energy Technologies and Renewable Sources (TERIN), ENEA—Italian National Agency for New Technologies, Energy and Sustainable Economic Development, C.R. Bologna, Via dei Mille 21, 40121 Bologna, Italy; serena.donofrio@enea.it

* Correspondence: massimo.celino@enea.it (M.C.); francesco.buonocore@enea.it (F.B.)

Abstract

The discovery of energy materials increasingly couples high-throughput computation with machine learning, but the data and models behind a given study are rarely left in a state that lets others, or the original authors months later, re-run or extend the work. IEMAP, Italian Energy Materials Acceleration Platform, was built within the Italian Mission Innovation programme to support the management, analysis and reuse of heterogeneous experimental and computational data on energy materials, with electrochemical storage as its primary use case. Data are organized as projects, each described by structured metadata kept in MongoDB and linked to raw files held in a Ceph object store by content hash. The platform offers interactive access through a web interface and programmatic access through an open-source Python client (iemap-mi) that lets a researcher script ingestion, query and inference together. A trained graph neural network (geoCGNN) is served as an endpoint returning the formation energy and redox potential of a candidate crystal, so the model runs against the same store that holds its inputs. We report a component-level FAIR (Findable, Accessible, Interoperable, Reusable) assessment of the platform, publish the domain ontology that underpins its interoperability, and outline a roadmap towards native semantic querying and knowledge-graph integration. As a demonstration, we reproduce a study of Ni/Ti-doped P2-NaMnO₂ sodium-ion cathodes as a single, openly accessible IEMAP workflow that screens a large compositional space, identifies promising chemistries, and stores all intermediate and final results for reuse. The contribution is the infrastructure that turns a one-off computation-and-ML study into a reusable resource for energy-materials research.



Academic Editors: Roberto Bubbico and Kah Chun Lau

Received: 1 July 2026

Revised: 26 August 2026

Accepted: 8 September 2026

Published: 10 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: materials acceleration platform; research data infrastructure; FAIR data; open data; sodium-ion batteries; cathode materials; machine learning; graph neural networks; IEMAP

1. Introduction

The transition to a sustainable energy system critically depends on the development of affordable, efficient and durable electrochemical energy storage technologies that can buffer the increasing penetration of intermittent renewable sources [1–3]. Lithium-ion batteries (LIBs), while technologically mature, face constraints related to raw material availability, cost and sustainability, which motivate the research into alternative chemistries such as sodium-ion batteries (SIBs) based on more abundant elements, including a broad class of layered oxide and composite cathodes [2–8]. At the same time, life-cycle and sustainability assessments impose additional constraints on the choice of active and inactive materials, further enlarging an already wide design space for next-generation batteries [9].

Discovering and optimizing new electrode and electrolyte materials is intrinsically a high-dimensional problem that couples atomic-scale structure, chemistry, processing conditions and operating environment in a non-trivial manner [5,10,11]. Over the past decade, both experimental campaigns and electronic-structure simulations have generated rapidly growing volumes of heterogeneous data; for example, it has been estimated that large established collections now comprise tens of thousands of high-quality measurements and that the battery and solar-cell literature alone encompasses well over one million publications [12]. Yet, the lack of standardized data models, interoperable repositories and robust workflows linking simulations, experiments and machine-learning models still limits the systematic exploitation of this information for materials design.

In response to these challenges, the concept of Materials Acceleration Platforms (MAPs) has emerged as a unifying paradigm for integrating high-throughput experimentation, automated computation, data infrastructures and artificial intelligence (AI) to accelerate materials discovery and optimization [13]. International initiatives such as the Materials Genome Initiative in the United States [14], the Advanced Materials Research Facility in Canada [15], BIG-MAP in Europe [16] and related efforts have demonstrated the potential of MAPs to reduce iteration cycles by combining high-performance computing (HPC), robotics and data-driven models. More recently, the COST Action EU-MACE has been launched to build a coordinated European ecosystem of MAPs for energy materials, emphasizing shared data standards, infrastructure, and best practices [17].

A key requirement for making these platforms effective and sustainable is the adoption of FAIR (Findable, Accessible, Interoperable, Reusable) data principles for both computational and experimental data, including provenance, workflows and model artefacts [18]. Efforts such as the development of the OPTIMADE API for interoperable access to materials databases, the NOMAD Laboratory and the Materials Cloud platform provide essential building blocks, alongside emerging domain ontologies that formalize materials, processes, parameters, properties and provenance in a machine-readable form [19–21]. Electrochemical energy storage poses specific challenges: relevant datasets span multiple length and time scales, from crystal structures and defect chemistries to cycling protocols and device-level performance metrics, and must be linked through consistent metadata that remain faithful to FAIR principles [9].

Within this international context, the Italian Energy Materials Acceleration Platform (IEMAP) has been developed as a national digital research infrastructure to support accelerated discovery and optimization of materials for photovoltaics, electrolyzers and,

crucially, electrochemical storage systems [12,22]. IEMAP combines a web-based front-end, a RESTful API layer, a NoSQL database for structured metadata and an object store for raw data, all deployed on the ENEAGRID/CRESCO high-performance computing infrastructure and designed to interface with external computational and experimental workflows. (ENEAGRID/CRESCO is an integrated computational infrastructure that offers a production-quality, service-oriented system for high-performance and/or high-throughput computing. Distributed over six ENEA centres (four of them equipped with HPC facilities), it provides a unified user environment and a homogeneous access method for all users, regardless of their location. The computational resources are provided by CRESCO Clusters based on different architectures (x86_64 and PowerPC), varying computing power, and different sites (Portici, Frascati, Casaccia, and Brindisi). The main cluster is CRESCO8 (9 PFlops) ranked 228th position in the June 2025 Top500 list. Other clusters are: XCRESO (1.9 PFlops), CRESCO6 (1.4 PFlops), CRESCO7 (0.5 PFlops), CRESCO4F (20 Tflops) and CRESCO4C (10 Tflops CPU).) The platform implements a common metadata model that captures material, process, parameter, property and provenance entities across both experimental and computational domains, which is formalized by a dedicated ontology and aligned with FAIR principles inspired by existing efforts in the community [18].

Beyond data management, IEMAP provides AI and machine-learning services, including a geometric graph neural network (geoCGNN) model [23] for predicting formation energies and redox potentials directly from crystallographic information files (CIFs), which has been trained and validated on large datasets from established materials databases and deployed as a microservice. Programmatic access through an open-source Python module (iemap-mi) enables researchers to automate the ingestion of ab initio and laboratory data, orchestrate high-throughput studies, and integrate IEMAP into existing simulation and analysis pipelines. Together, these components align IEMAP as a domain-specific, FAIR-oriented MAP for energy materials, working alongside international infrastructures such as the Materials Project, NOMAD and Materials Cloud [14,20,21].

IEMAP differs from existing infrastructures in scope and integration model rather than in scale. General-purpose repositories such as the Materials Project [14], NOMAD [20] and Materials Cloud [21] curate large, primarily computational datasets with global coverage; BIG-MAP [16] and the EU-MACE COST Action [17] coordinate European research on autonomous discovery workflows. IEMAP is, to our knowledge, the first national-scale platform that combines (i) a single metadata model spanning both experimental and computational records, formalized in a published OWL (Web Ontology Language) ontology [22]; (ii) co-location of a served ML model (geoCGNN) with the store holding its training data, inputs and predictions, so that inference results are written back as first-class, provenance-tracked projects; and (iii) open, unauthenticated programmatic read access through a versioned REST API and a PyPI-distributed client (iemap-mi). Its role is therefore complementary: a FAIR-oriented national backbone for reusable energy-materials workflows, designed to interoperate with, not replace, the international repositories above.

Electrochemical energy storage has been a primary driver for both the conceptual design and the first scientific use cases of IEMAP, with a particular focus on cathode materials for sodium-ion batteries [6]. In previous works, machine-learning models trained on ab initio data have been applied to doped P2-NaMnO₂-based cathodes to screen large combinatorial spaces of compositions and cation distributions, identifying promising Na–Mn–Ni–Ti chemistries that were subsequently validated by density functional theory calculations [5,6,11]. Embedding these predictive models, together with the underlying computational and experimental data, into a FAIR-aligned platform such as IEMAP enables systematic reuse, extension and cross-comparison of results, and opens the way to more complex workflows where AI guides both simulation campaigns and laboratory efforts.

In this work, we present IEMAP not simply as a materials database or as a software platform, but as a FAIR-oriented Materials Acceleration Platform designed to make energy-materials workflows reproducible, extensible and reusable. This infrastructure integrates structured metadata, persistent raw-data storage, programmatic access, ontological representation and AI inference services into a single operational infrastructure. We demonstrate this approach through a sodium-ion battery case study, where DFT data and a geoCGNN model for formation-energy and redox-potential prediction are embedded into a workflow that leverages the IEMAP framework to screen P2-NaMnO₂-derived cathode compositions. In this way, a one-off high-throughput computation-and-machine-learning study is transformed into a platform-hosted resource that can be queried, reproduced, extended and benchmarked against future candidate materials.

2. The IEMAP Digital Platform

2.1. Overview and Access Model

IEMAP is a modular, microservice-oriented digital platform that couples a web user interface, a RESTful application layer and data services deployed on ENEA's ENEA-GRID/CRESCO infrastructure [22] (Figure 1).

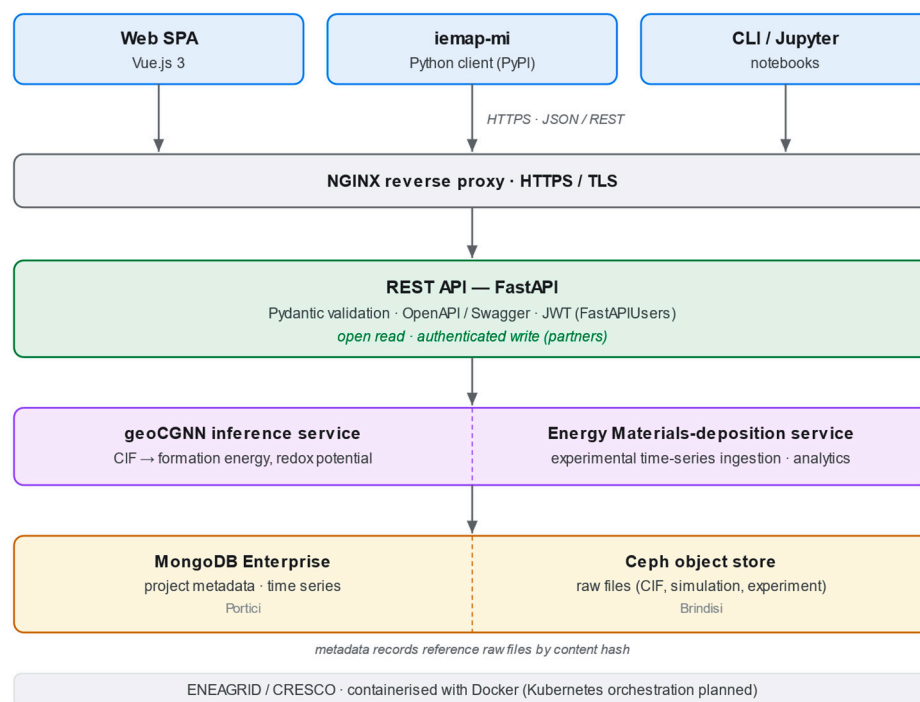


Figure 1. The layered architecture of IEMAP. Web and programmatic clients reach a FastAPI REST layer behind a NGINX reverse proxy; the API serves a geoCGNN inference service and persists project metadata in MongoDB and raw files in a Ceph object store (Brindisi), linked by content hash. The Energy Materials-deposition service handles experimental measurement streams from instrument workstations, which are stored as time series and consumed through the same REST layer (Section 2.3).

It is designed to ingest, curate and expose heterogeneous experimental and computational data on energy materials, and to provide integrated AI services and programmatic interfaces that fit into existing battery-materials workflows. Data are organized around the notion of a project: a self-contained set of metadata describing a material together with the process that produced or measured it, the parameters of that process, the resulting properties, and the agent and provenance of the record. Read access to project metadata and files is open and requires no account; creating projects and uploading data requires an

authenticated account, granted to project partners, following the principle of being as open as possible and as closed as necessary. The platform supports the three founding research domains of the Italian Mission Innovation programme (materials for batteries, electrolyzers and photovoltaics) and is operated across its four partner institutes (ENEA, CNR, RSE and IIT); as of May 2026 the platform held 782 project records, with the current count publicly accessible at <https://iemap.enea.it>.

2.2. Web and API Layers

At the presentation layer, IEMAP is a single-page application (SPA) developed in Vue.js 3, with client-side routing through Vue Router and dedicated libraries for state management and data visualization, giving responsive access to search, upload and inspection. The SPA communicates with the backend only through JSON-based REST calls, which keeps the user interface separate from the application logic and lets other clients, such as command-line tools and Jupyter notebooks, use the same endpoints. The application layer is written in Python 3.10.12 with FastAPI 0.110.0, which combines asynchronous request handling, automatic OpenAPI (Swagger) documentation, and Pydantic data validation that yields typed, self-documented interfaces. Authentication and authorisation use JSON Web Tokens (JWTs) managed through FastAPIUsers, with HttpOnly cookies in the web client; access policies separate public read-only endpoints from authenticated, write-enabled operations reserved to partners. The endpoints cover project and file management, AI-model invocation, statistics and health monitoring, and are exposed through one versioned API namespace shared by the web front-end and the Python client.

2.3. Data Layer

Persistence is handled by an on-premise MongoDB Enterprise 6.0.18 deployment, hosted on the ENEAGRID/CRESCO's Portici data centre, which stores project metadata as JSON documents and, where needed, experimental measurements as time series in dedicated collections suited to high-frequency sensor data. Raw artefacts (experimental files; simulation outputs; and crystallographic information files, CIFs) are kept in a distributed Ceph 17.2.9 (Quincy) object store at ENEAGRID/CRESCO's Brindisi data centre and referenced from MongoDB by cryptographic hash together with descriptive fields, which preserves data integrity and keeps the document store light and quick to query. At the schema level the metadata model encodes the core entities of the platform's ontology (Project, Material, Process, Parameter, Property, Agent and Provenance), as depicted in Figure 2, giving a concrete backbone for FAIR-aligned description and discovery across experimental and computational studies.

Experimental measurement streams follow a separate path, shown as the Energy Materials-deposition service in Figure 1. An acquisition script running on the instrument workstation streams measurements to a dedicated time-series collection as they are produced, and a containerised analytics application consumes them through the versioned REST API rather than by direct database access, so that experimental monitoring reuses the same authenticated, documented interface as every other client; the pipeline currently serves perovskite cell deposition and is at the prototype stage. The storage layer is therefore deliberately two-tier: project metadata are held in ordinary document collections, because they are richly nested and require versioning and partial updates, whereas measurement streams are held in native MongoDB time-series collections, because they are append-dominated, time-indexed and far larger in record count. The two shapes are not interchangeable, and the experimental series are consequently not expressed in the project data model of Figure 2. Ingesting electrochemical cycling series, such as capacity per cycle and impedance spectra, is a direct extension of this path, since the acquisition side is

domain-agnostic; linking them to atomistic predictions requires a bridge between the two tiers, for which the natural mechanism is the metadata field of the time-series collection, recording there the iemapid of the corresponding project and, where a structure exists, the content hash of its CIF. That bridge is not yet in place and is listed in the roadmap in Section 5.

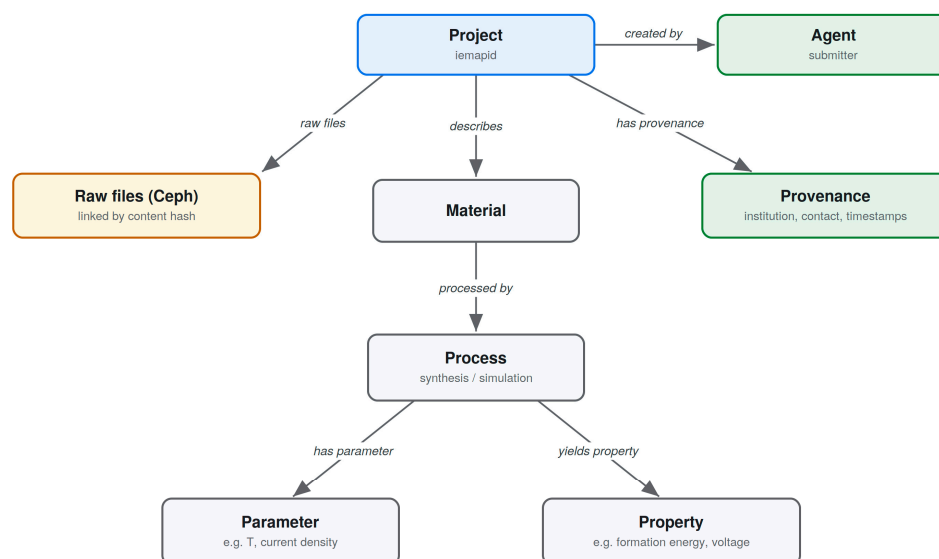


Figure 2. The IEMAP project data model. Each project (identified by an iemapid) describes a material and the process that produced or measured it, along with its parameters and resulting properties; an agent and provenance record document its source, and raw files in object storage are linked to the project via the content hash.

2.4. Deployment and Operations

The services are containerised with Docker 28.1.1 and run on a dedicated virtual machine behind a NGINX 1.18.0 reverse proxy, which terminates HTTPS connections, manages TLS certificates and routes traffic to the individual microservices. This lets the main components be deployed and scaled independently: the core REST API, the AI microservice that exposes the geoCGNN model, a domain-specific service for Energy Materials-deposition data, and the service layer behind the Python client. Backups, firewalling and intrusion detection on the ENEAGRID/CRESCO infrastructure complement the application-level controls, so the platform can host research data that is sensitive yet shareable under institutional policy. The architecture is prepared for Docker and Kubernetes orchestration in distributed or cloud-native settings, which is not yet deployed.

2.5. Programmatic Client: iemap-mi

For advanced users the platform provides an open-source Python package, iemap-mi, distributed on the Python Package Index, which wraps the REST API in an asynchronous client with Pydantic-validated models. The library offers high-level operations for building and validating metadata payloads, uploading and downloading associated files, paginating over large project collections, and invoking the integrated AI models, including geoCGNN, directly from scripts and notebooks. Because the same client both populates and reads the store, it supports workflows that couple ab initio calculations, data-driven screening and large parametric studies of cathode materials within one FAIR-aligned infrastructure, and a study written against iemap-mi can be re-run by anyone with access to the package and the platform.

2.6. Machine-Learning Service: geoCGNN

IEMAP exposes a trained graph neural network, geoCGNN, as an inference endpoint. The model follows the geometric-information-enhanced crystal graph network of [23] and was further trained and validated by ENEAGRID/CRESCO computational facilities using data from the Materials Project [14] (about 150,000 entries for formation energy and roughly 4000 for redox potential, with noble elements excluded). Given a crystallographic information file (CIF) that describes the candidate in its discharged state, the endpoint returns its predicted formation energy and redox potential, reachable through both the REST API and iemap-mi. The inference service runs on Python 3.10 with FASTAPI 0.115.5 and PyTorch 2.5.1 (CPU build), with pymatgen 2024.11.13 used for structure parsing; the client iemap-mi is version 0.2.6. The served model implements the single-crystal direct prediction of [6], which infers the redox potential from the discharged structure alone; it is the same variant used for the validation in Section 4.4, so the predictions reported there are those the platform returns. Two checkpoints are served, one per property: formation energy and redox potential. Both follow the architecture of [23] with five convolution blocks, an 8 Å cut-off and at most twelve neighbours per node and count 1,911,456 and 1,002,144 trainable parameters; the hyperparameters are listed in Supplementary Table S3. As of version 1.1 the inference response carries this provenance inline, so that every prediction served through the REST API and iemap-mi identifies the model that produced it. Because inference runs against the same store that holds the input structures and their provenance, a prediction can be written back as a project; this co-location of a served model with the data it consumes is the platform's main point of integration and the basis for the demonstration in Section 4.

3. Data FAIRness

3.1. FAIR by Design

The design of IEMAP follows the FAIR principles (Findable, Accessible, Interoperable, Reusable) [18], with the aim of making experimental and computational datasets on energy materials discoverable and reusable across groups and application domains. For electrochemical storage this matters because the relevant information ranges from atomistic structures and simulation inputs to cycling protocols, sensor time series and device-level performance, and these must be linked through consistent, machine-readable metadata for comparisons and data-driven modelling to be meaningful.

3.2. Findable

Each project is assigned a persistent internal identifier (iemapid), and a structured metadata schema exposes searchable fields for material composition, process type, parameters and properties. These metadata can be queried from both the web front-end and the REST API and are documented through automatically generated OpenAPI (Swagger) descriptions that support programmatic discovery. The same core concepts (Project, Process, Material, Parameter, Property, Agent and Provenance) are aligned with an OWL ontology derived from the MongoDB schema, which provides a formal semantic layer that other initiatives in materials science can reference.

3.3. Accessible

Access uses HTTPS endpoints, with JWT-based authentication for write operations and open, unauthenticated access for read-only project search and inspection. Metadata are served in JSON and tabular or time-series data in standard formats, so clients can retrieve and process records with common scientific libraries, supported by user and developer documentation and the Swagger-generated API reference. Institutional backup and security

policies adopted by ENEAGRID/CRESCO contribute to the long-term availability of the stored data.

3.4. Interoperable

Interoperability is pursued by mapping the internal data model to the domain ontology, which is exported in OWL, Turtle and JSON-LD and published on Zenodo [22]. MongoDB does not natively support RDF or SPARQL, so the semantics captured in the ontology currently act as an external descriptor; they nonetheless prepare future integration with knowledge-graph backends and with external infrastructures such as OPTIMADE-compliant databases and platforms like the Materials Project, NOMAD and Materials Cloud [14,19–21]. For electrochemical systems the schema separates parameters (for example temperature, current density and cut-off voltages) from properties (such as formation energy, capacity and Coulombic efficiency), giving a structured bridge between atomistic simulations and cell-level performance.

3.5. Reusable

Reusability rests on versioning of project documents in MongoDB, explicit licencing (including an open-source licence for the iemap-mi client), and documented provenance. Each record stores the submitting institution, a contact address, creation and update timestamps, and the hash link between metadata and raw files, which supports traceability and reproducibility. Because trained models such as geoCGNN are served with defined inputs and outputs, users can reuse both data and models in new studies, for instance by applying the same prediction workflow to novel sodium-ion cathode candidates or by comparing machine-learning predictions with new ab initio calculations.

3.6. FAIRness Assessment

The level of FAIRness was assessed with a semi-quantitative scheme that scores each component (the MongoDB metadata store, the REST API and the web interface) against detailed criteria (Table 1). Each of the four principles was decomposed into three criteria scored on a 0–2 scale (Table S1), so a principle reaches at most six points and a component at most twenty-four. The per-criterion score of each of the three components is reported in Table S2. In practice, this is a structured internal audit, in the spirit of the RDA FAIR Data Maturity Model [24], carried out by inspection; it was preferred to automated assessors such as F-UJI [25], which test externally resolvable resources and cannot reach internal components such as the database schema. As a worked example, the MongoDB store scores 5 out of 6 on Findable: 2 for the persistent internal identifier (iemapid), implemented natively for every project; 2 for metadata held in a document model whose primary function is indexing and complex querying; and 1 for discovery documentation, since Swagger is a native feature of the REST API rather than of the store, which therefore benefits from it only indirectly. An automated cross-check of the public API is not reported here: F-UJI assesses an individual data object addressed by a globally resolvable persistent identifier, whereas IEMAP project records currently carry an internal identifier and the unit of assessment in Table 1 is a platform component rather than a data object. Minting externally resolvable identifiers, a precondition for a meaningful automated assessment, is listed in the roadmap in Section 5.

Findability and accessibility score highest across the three components. The REST API is the most FAIR-aligned (87.5%): as a machine-to-machine interface, it returns JSON with provenance metadata, which supports interoperability and programmatic reuse. The web interface scores lowest (66.7%), as expected for a layer built for human interaction. Interoperability is the weakest axis (3 to 4 out of 6), and the score reflects the state of the semantic strategy rather than its absence: the ontology has been formalized and published,

but it is not yet wired into the database's validation or query, and the platform exposes no native RDF or SPARQL endpoint. Closing this gap, by integrating the ontology into validation and query and by adding RDF/SPARQL access and links to external vocabularies for units and properties, is the main item on the platform's FAIR roadmap (see Section 5).

Table 1. A component-level FAIR assessment of IEMAP. Each FAIR principle (Findable, Accessible, Interoperable, Reusable) is scored from three criteria on a 0–2 scale, giving at most 6 points per principle and 24 per component; the final column reports the overall FAIRness as a percentage of the maximum score.

Component	Findable	Accessible	Interoperable	Reusable	Total	FAIRness
MongoDB store	5	6	3	5	19	79.2%
REST API	5	6	4	6	21	87.5%
Web interface	4	5	3	4	16	66.7%

4. Case Study: Cathode Materials for Batteries

4.1. Cathode-Materials Design Problem

We used IEMAP in an electrochemical energy storage application as a demonstration of the platform's capabilities. We focused on designing enhanced layered oxide cathodes for sodium-ion batteries (SIBs), which are considered promising alternatives to lithium-ion batteries due to the abundance and low cost of sodium and the possibility of reusing much of the existing lithium-ion batteries (LIBs) manufacturing infrastructures [4,7–9]. P2-type NaMnO_2 and its derivatives have attracted considerable attention owing to their combination of relatively high operating voltage, competitive capacity and good cycling stability when appropriately engineered [26]. A strategy to enhance the performance of layered oxides as cathode materials is using high-throughput DFT calculations for designing crystals by substituting a fraction of the transition metals in the base crystal (e.g., Mn) with redox-active cations such as Ni and Fe as well as redox-inactive stabilizers such as Ti and Li to optimize redox voltages and diffusion barriers. However, the combinatorial growth in the number of possible compositions, dopant species and cation distributions in the transition-metal layers makes a purely quantum-mechanical exploration of the design space prohibitively expensive, even on large-scale HPC facilities [5]. However, the use of the geoCGNN ML model implemented within IEMAP helped to perform systematic and fast screening of over 500,000 derivatives of NaMnO_2 , in a time of approximately 10^{-3} s to predict a single formation energy and redox potential, on V100 NVIDIA GPU upon network weights upload [6]. The structure of NaMnO_2 is shown in Figure 3, and its derivatives are obtained by replacing the Mn atoms with Ni and Ti atoms, resulting in the general system: $\text{NaMn}_x\text{Ni}_y\text{Ti}_z\text{O}_2$. Furthermore, IEMAP, through the interplay between the geoCGNN inference service and the MongoDB database, where the results of DFT calculations of formation energy and average voltage for some Mn–Ni–Ti compositions were stored, enabled DFT corrections on the ML predictions that facilitated the screening and selection of the best amounts of Mn, Ni, and Ti in our general system with respect to stability and voltage. Further crystallographic information and the quantitative details on the ML performances are provided in [6].

4.2. Reference DFT Dataset and IEMAP Ingestion

Within this context, a workflow has been established in which density functional theory (DFT) calculations are used to compute formation energies E_f and redox potentials V_{redox} for a curated set of Na–Mn–Ni–Ti compositions and cation arrangements derived from a P2- NaMnO_2 base crystal [6] (Figure 3). The former gives indications related to the structural stability of the composition and then to the cycle life; the latter is associated

with the intercalation reaction that brings a discharged material to the charged one or vice versa, and it is related to energy density. The resulting dataset is ingested into IEMAP through the iemap-mi Python client, which ensures that each configuration is described by consistent metadata specifying material formula, structural prototype, computational method, convergence parameters and target properties, together with full provenance information [27]. These curated data serve both as a reference corpus for subsequent analyses and as training and validation input for machine-learning models deployed in the platform.

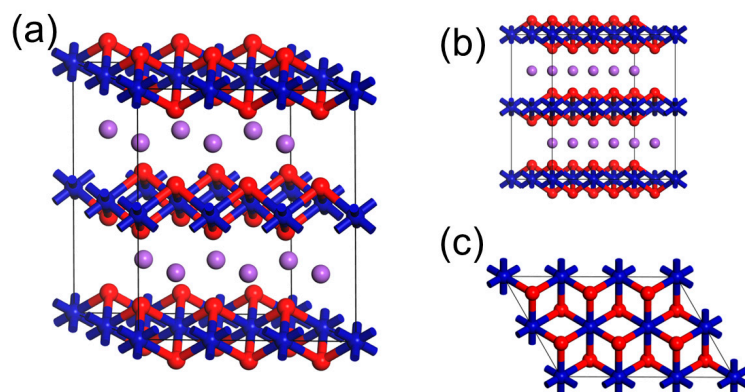


Figure 3. (a) Atomistic model of P2-layered oxide NaMnO₂ supercell with (b) side and (c) top views. In blue, purple, and red are the Mn, Na, and O atoms, respectively.

4.3. geoCGNN Inference Workflow

To accelerate exploration beyond the explicitly computed structures, a geometric graph neural network (geoCGNN) ML model is available in the IEMAP framework as a dedicated microservice for predicting formation energies and redox potentials directly from crystallographic information files (CIFs) describing the discharged state of candidate cathode materials [6]. A CIF of a crystal contains information about atomic positions and cell parameters. The ML model, trained on a large crystalline materials dataset provided by the Materials Project [14], can be invoked through iemap-mi by submitting CIFs generated in high-throughput workflows able to enumerate the possible compositions given the substituents, thereby providing near-instant estimates of key thermodynamic and electrochemical descriptors. In this way, the costly DFT calculations are reserved for a strategically selected subset of candidates, while the AI model guides the search towards regions of composition space that are most likely to exhibit favourable stability and voltage profiles.

4.4. Validation Against DFT and Multi-Objective Screening

To assess the ML approach, a representative set of configurations with different combinations of Mn, Ti and Ni atoms was selected. In Figure 4, the predicted values for E_f (Figure 4a) and V_{redox} (Figure 4b) are compared with the corresponding values calculated by the DFT method and hosted in the IEMAP dataset. The comparison indicates satisfactory agreement between the predictions and DFT calculations, particularly for formation energy, where we obtain MAE = 0.041 eV and RMSE = 0.062 eV, while for V_{redox} we have MAE = 0.465 V and RMSE = 0.617 V. In both cases, we observe the alignment along the $x = y$ reference line, denoted by the red dashed line. The better accuracy for E_f compared to V_{redox} is due to the different sizes of the two datasets. The MP database contains more than 150 thousand values for formation energy but two orders of magnitude fewer values for the redox potential (less than 4500 values). Training on a larger dataset yields more accurate predictions for supervised ML methods, as confirmed by the smaller error bars for

the former physical quantity inspected compared to the latter. The error bars are obtained by adopting the k-fold cross validation technique with $k = 15$ as indicated in ref. [6], during training. This provides 15 sets of weights for the networks and hence 15 possible predictions for a given quantity. In this way, the user can estimate mean value for E_f and V_{redox} which can be compared with the DFT calculations, through for instance the definition of a revised predictor as described in ref. [6], to check whether geoCGNN returns accurate predictions and to perform the proper correction, which is particularly beneficial for V_{redox} .

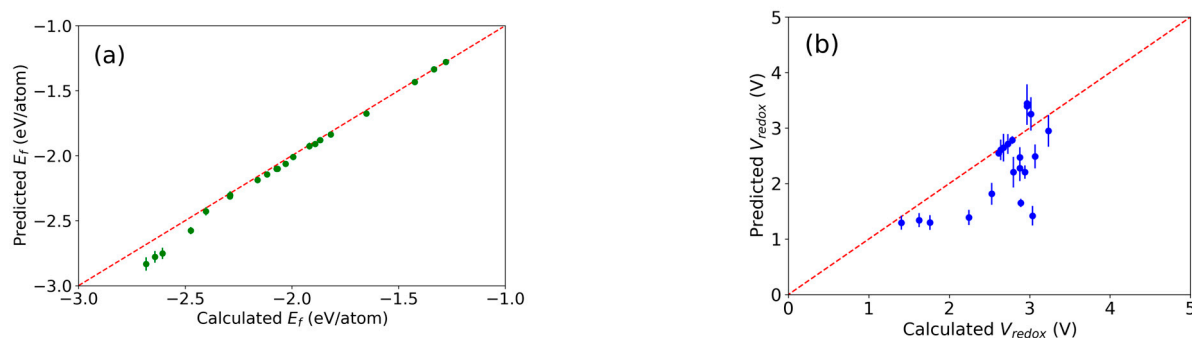


Figure 4. Formation energy E_f ((a), green symbols) and redox potential V_{redox} ((b), blue symbols) predictions values vs. DFT calculated values for structures of selected configurations with different combinations of Mn, Ti and Ni atoms. Figures are adapted from ref. [6].

Having validated the ML model against DFT calculations, we extended its application to the entire compositional space of Mn–Ni–Ti combinations, predicting for each composition p the values of formation energy E_f and redox potential V_{redox} . Since a good candidate for this type of material should combine the two characteristics described above, namely, cyclability (low E_f) and energy density (high V_{redox}), a multi-objective function MO was defined to integrate these two physical quantities.

$$MO(p) = -E_f/e + V_{redox} \quad (1)$$

where e is the electron charge set to fix the correct units. The multi-objective function is used as a relative screening score within the explored Mn–Ni–Ti composition space, rather than as a universal thermodynamic descriptor. In Figure 5, it is reported in the form of a ternary plot, the normalized multi-objective function MO , namely $MO \rightarrow (MO - \min_p(MO))/(\max_p(MO) - \min_p(MO))$, in order to be in the range $[0, 1]$, where the min and max of MO are obtained over its predicted values. Values closer to unity indicate more favourable material compositions since they correspond to a combination of low (more negative) E_f and high V_{redox} .

Analysis of the IEMAP-hosted dataset, supported by machine-learning models trained on the curated features, indicates that compositions with comparable fractions of Ni and Ti and no Mn in the transition-metal layer tend to exhibit superior predicted performance, a trend subsequently confirmed by targeted ab initio calculations [6]. All intermediate and final results, including input structures, computed properties, metadata and model predictions, are stored and versioned within IEMAP, enabling reproducible benchmarking of future Na-ion cathode candidates against this reference landscape [22].

4.5. Reusability and Extension of the Workflow

The novelty of the present work with respect to previous ML/DFT studies on Na-ion cathode materials, does not lie primarily in proposing a new cathode chemistry or in redefining the underlying formation-energy and redox-potential descriptors. Those aspects were introduced and validated in earlier studies on P2-NaMnO₂-derived systems.

Here, the original contribution is that the same type of computation-and-machine-learning workflow that leverages the IEMAP infrastructure can serve as a template for other new Materials Acceleration Platform workflows. The input crystal structures, DFT reference data, model predictions, metadata, provenance information and screening indicators are no longer treated as isolated outputs of a single study, but as structured, queryable and extensible digital objects connected through the IEMAP data model, storage layer, API, Python client and AI inference service. This allows the cathode-screening workflow to be reproduced, benchmarked against new calculations, extended to new compositions and reused as a reference resource for future battery-materials discovery.

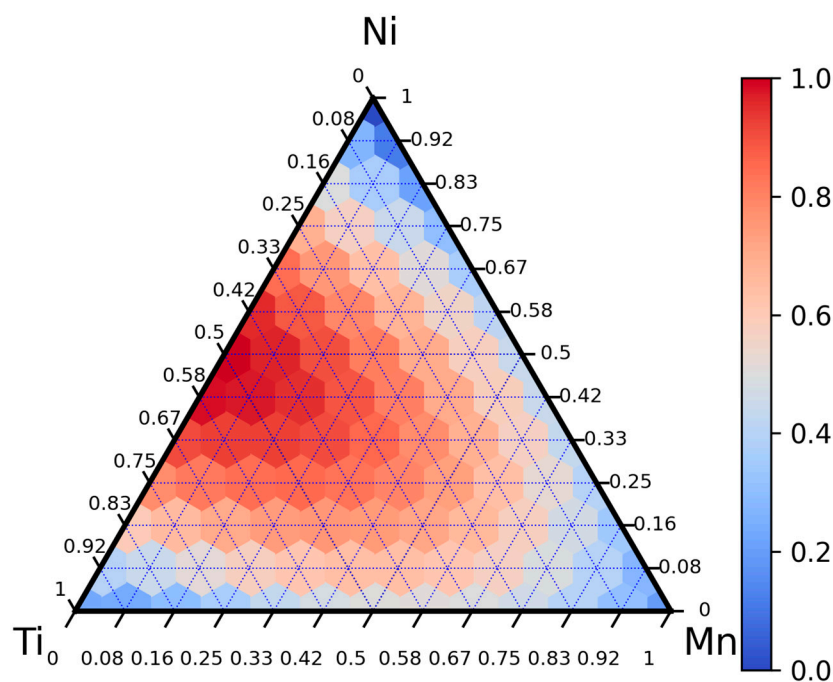


Figure 5. The ternary plot obtained from the ML predictions for the multi-objective MO function (Equation (1)) that is normalized in the $[0, 1]$ range. All the possible Mn–Ni–Ti compositions p are considered.

Beyond this specific chemistry, the same IEMAP-based workflow can be generalized to other families of cathode materials, including polyanionic compounds and prussian blue analogues, by adapting the metadata descriptors and retraining or fine-tuning the AI models on appropriate training sets. The combination of FAIR-oriented data management, programmatic access through *iemap-mi* and integrated AI inference services positions IEMAP as a practical backbone for iterative and increasingly automated battery-materials design workflows, where hypotheses for new cathode formulations can be generated, evaluated, refined and compared in a systematic and scalable manner.

5. Conclusions

IEMAP has been conceived and implemented as a domain-specific Materials Acceleration Platform that integrates web technologies, REST APIs, scalable data services and AI models to support the end-to-end life cycle of data and workflows for energy-materials research, with a particular focus on electrochemical energy storage [22]. By combining an SPA front-end, a FastAPI application layer, a MongoDB/Ceph data backend and an open-source Python client, the platform provides both interactive and programmatic access to FAIR-oriented repositories of experimental and computational data, lowering the

barrier for cross-group collaboration and for the deployment of advanced data-driven methodologies in battery science.

From the perspective of FAIRness, IEMAP shows strong adherence to the Findable and Accessible dimensions and a solid foundation for Reusable and Interoperable data, thanks to a structured metadata model, persistent identifiers, explicit provenance tracking, open documentation and an OWL ontology that formalizes core concepts such as Project, Material, Process, Parameter, Property and Provenance [18]. The current limitations are well characterized: chiefly the absence of native semantic-query capabilities and only partial alignment with external vocabularies for units and properties. A clear roadmap addresses them, through integration of the published ontology into validation and query, mapping to established vocabularies, and experimentation with knowledge-graph backends. Two further items complete the roadmap: minting externally resolvable persistent identifiers for project records, which would in addition make the platform assessable by automated FAIR evaluators; and bridging the time-series and metadata tiers of the data layer, so that experimental measurement series and atomistic predictions for the same composition can be resolved through a single query.

Regarding scalability, the platform's microservice architecture allows each component—REST API, inference service, and data services—to be scaled independently; the services are containerised with Docker, and migration to Kubernetes orchestration is prepared for deployment in distributed or cloud-native settings. On the data side, the MongoDB/Ceph split keeps the metadata store light regardless of raw-data volume, and the underlying ENEAGRID/CRESCO infrastructure provides headroom for both storage growth and compute-intensive inference. As of May 2026 the platform holds 782 project records across the three Mission Innovation domains and four partner institutes.

The case study on Na-ion cathode materials shows how IEMAP can host and orchestrate complex, multiscale workflows in which DFT calculations, curated metadata and the geoCGNN model are combined to explore large compositional and configurational spaces. In this example IEMAP supports systematic ingestion and versioning of ab initio data, rapid AI-based screening of doped NaMnO₂ derivatives, and targeted quantum-mechanical validation of the most promising candidates, reproducing design rules that correlate cation composition and spatial distribution with predicted stability and voltage. Because the entire workflow, from structural generation to model inference and data analysis, can be reproduced and extended through the iemap-mi client, the platform acts as a reusable digital backbone rather than a single-project repository [27].

Looking forward, the same architectural patterns and FAIR-data mechanisms that underpin the Na-ion cathode use case can be transferred to other battery-relevant materials classes, including anodes, solid and liquid electrolytes, interfaces and high-entropy or disordered phases. As further datasets and AI models are integrated, potentially including generative design frameworks [28], autonomous workflow engines and semantic search over heterogeneous data, IEMAP can contribute to the European ecosystem of Materials Acceleration Platforms by providing a national, FAIR-oriented digital backbone for reusable, AI-assisted energy-materials workflows, while future developments will focus on deeper semantic querying, stronger interoperability with external vocabularies and more automated workflow orchestration.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/batteries12090355/s1>, Table S1: Scoring levels; Table S2: Criteria and per-component scores; Table S3: The metadata and hyperparameters of the geoCGNN model in the endpoint. Reference [23] is cited in Supplementary Materials.

Author Contributions: Conceptualization, M.C. (Massimo Celino), F.B., G.P. and S.F.; methodology, M.C. (Massimo Celino), F.B., M.C. (Marco Catillo) and S.F.; software, S.F.; validation, M.C. (Marco Catillo), S.D., S.G. and S.M.; formal analysis, M.C. (Marco Catillo), F.B. and S.G.; investigation, M.C. (Marco Catillo), S.D., S.G., S.M. and F.B.; resources, G.P. and M.C. (Massimo Celino); data curation, S.F. and M.C. (Marco Catillo); writing—original draft preparation, M.C. (Massimo Celino), F.B. and S.F.; writing—review and editing, S.F., M.C. (Massimo Celino), M.C. (Marco Catillo), S.D., S.G., S.M., F.B. and G.P.; visualization, M.C. (Marco Catillo) and S.F.; supervision, M.C. (Massimo Celino), F.B. and G.P.; project administration, M.C. (Massimo Celino); funding acquisition, M.C. (Massimo Celino). All authors have read and agreed to the published version of the manuscript.

Funding: The Project “Italian Energy Materials Acceleration Platform” (IEMAP)—CUP I62C210003800 01 in the Mission Innovation programme is acknowledged for financial support. This work was supported by the European Union through the Next Generation EU funds through the Italian MUR National Recovery and Resilience Plan, Mission 4 Component 2—Investment 1.4—National Center for HPC, Big Data and Quantum Computing—CUP I33C22001270007.

Data Availability Statement: The IEMAP digital platform is publicly accessible at <https://iemap.enea.it>. Project metadata can be queried without authentication through the web interface and the REST API; the iemap-mi Python client is available on PyPI at <https://pypi.org/project/iemap-mi/> (accessed on 15 June 2026). The IEMAP ontology underpinning the metadata model is openly archived on Zenodo at <https://doi.org/10.5281/zenodo.15396693>. The dataset used in the case study is available at <https://iemap.enea.it/> (accessed on 15 June 2026).

Acknowledgments: The authors thank the ENEAGRID/CRESCO HPC infrastructure team at ENEA Portici for technical support, and the IEMAP project partners (CNR, RSE and IIT) for ongoing collaboration on data models and use cases. During the preparation of this manuscript, the authors used Claude (Anthropic), accessed via the claude.ai web interface in June 2026, for editorial assistance, specifically, for cross-checking the bibliography against the cited sources, identifying internal citation inconsistencies, and producing a tracked-changes revision of the manuscript draft. All technical content, scientific claims, computational results and figures were produced and validated by the authors, who reviewed and edited every AI-generated suggestion and take full responsibility for the content of this publication.

Conflicts of Interest: F.B. and M.C. (Massimo Celino) are Guest Editors of the Special Issue “Advances in Data-Driven and Learning Methods Applied to Battery Technologies” of the journal Batteries, in which this manuscript is submitted. They had no role in the peer-review or editorial decision process concerning this manuscript, which was handled independently by the Editorial Office and other members of the Editorial Board. All other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

IEMAP	Italian Energy Materials Acceleration Platform
FAIR	Findable, Accessible, Interoperable, Reusable
MAP	Materials Acceleration Platform
geoCGNN	Geometric-information-enhanced Crystal Graph Neural Network
DFT	Density Functional Theory
ML	Machine Learning
AI	Artificial Intelligence
SIB	Sodium-Ion Battery
LIB	Lithium-Ion Battery
CIF	Crystallographic Information File
API	Application Programming Interface
REST	Representational State Transfer

SPA	Single-Page Application
HPC	High-Performance Computing
JWT	JSON Web Token
OPTIMADE	Open Databases Integration for Materials Design
RDA	Research Data Alliance
CRedit	Contributor Roles Taxonomy

References

- König, A.; Nicoletti, L.; Schröder, D.; Wolff, S.; Waclaw, A.; Lienkamp, M. An overview of parameter and cost for battery electric vehicles. *World Electr. Veh. J.* **2021**, *12*, 21. [\[CrossRef\]](#)
- Duffner, F.; Wentker, M.; Greenwood, M.; Leker, J. Battery cost modeling: A review and directions for future research. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109872. [\[CrossRef\]](#)
- Turcheniuk, K.; Bondarev, D.; Amatucci, G.G.; Yushin, G. Battery materials for low-cost electric transportation. *Mater. Today* **2021**, *42*, 57–72. [\[CrossRef\]](#)
- Lee, D.H.; Xu, J.; Meng, Y.S. An advanced cathode for Na-ion batteries with high rate and excellent structural stability. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3304–3312. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ronchetti, C.; Marchio, S.; Buonocore, F.; Giusepponi, S.; Ferlito, S.; Celino, M. Study of cathode materials for Na-ion batteries: Comparison between machine learning predictions and density functional theory calculations. *Batteries* **2024**, *10*, 431. [\[CrossRef\]](#)
- Catillo, M.; Buonocore, F.; D'Onofrio, S.; Giusepponi, S.; Ferlito, S.; Marchio, S.; Celino, M. Machine learning for redox potential prediction and application on a Na-ion battery system. *Solid State Sci.* **2026**, *177*, 108336. [\[CrossRef\]](#)
- Kim, S.-W.; Seo, D.-H.; Ma, X.; Ceder, G.; Kang, K. Electrode materials for rechargeable sodium-ion batteries: Potential alternatives to current lithium-ion batteries. *Adv. Energy Mater.* **2012**, *2*, 710–721. [\[CrossRef\]](#)
- Rostami, H.; Valio, J.; Suominen, P.; Tynjälä, P.; Lassi, U. Advancements in cathode technology, recycling strategies, and market dynamics: A comprehensive review of sodium ion batteries. *Chem. Eng. J.* **2024**, *495*, 153471. [\[CrossRef\]](#)
- Kraus, P.; Bainglass, E.; Ramirez, F.F.; Svaluto-Ferro, E.; Ercole, L.; Kunz, B.; Huber, S.; Plainpan, N.; Marzari, N.; Battaglia, C.; et al. A bridge between trust and control: Computational workflows meet automated battery cycling. *J. Mater. Chem. A* **2024**, *12*, 10773–10783. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Luna, P.; Wei, J.; Bengio, Y.; Aspuru-Guzik, A.; Sargent, E.H. Use machine learning to find energy materials. *Nature* **2017**, *552*, 23–25. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ronchetti, C.; Puccini, M.; Ferlito, S.; Giusepponi, S.; Palombi, F.; Buonocore, F.; Celino, M. Machine learning techniques for data analysis in materials science. In Proceedings of the 2022 AEIT International Annual Conference (AEIT), Rome, Italy, 3–5 October 2022; pp. 1–6. [\[CrossRef\]](#)
- Celino, M.; Ponti, G. IA e scienza dei materiali: Rivoluzione per il futuro del settore energetico. *Energ. Ambiente Innov.* **2024**, *3*, 69–71. [\[CrossRef\]](#)
- Aspuru-Guzik, A.; Persson, K. Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence. Mission Innovation, Innovation Challenge 6, 2018. Available online: <https://dash.harvard.edu/handle/1/35164974/> (accessed on 9 June 2026).
- Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002. [\[CrossRef\]](#)
- Advanced Materials Research Facility. National Research Council Canada. Available online: <https://nrc.canada.ca/en/research-development/nrc-facilities/advanced-materials-research-facility> (accessed on 15 June 2026).
- Amici, J.; Asinari, P.; Ayerbe, E.; Barboux, P.; Bayle-Guillemaud, P.; Behm, R.J.; Bercibar, M.; Berg, E.; Bhowmik, A.; Bodoardo, S.; et al. A Roadmap for Transforming Research to Invent the Batteries of the Future Designed within the European Large Scale Research Initiative BATTERY 2030+. *Adv. Energy Mater.* **2022**, *12*, 2102785. [\[CrossRef\]](#)
- CA22123 COST Action European Materials Acceleration Center for Energy (EU-MACE). Available online: <https://www.cost.eu/actions/CA22123> (accessed on 15 June 2026).
- Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [\[CrossRef\]](#) [\[PubMed\]](#)
- Andersen, C.W.; Armiento, R.; Blokhin, E.; Conduit, G.J.; Dwaraknath, S.; Evans, M.L.; Fekete, Á.; Gopakumar, A.; Gražulis, S.; Merkys, A.; et al. OPTIMADE, an API for exchanging materials data. *Sci. Data* **2021**, *8*, 217. [\[CrossRef\]](#) [\[PubMed\]](#)
- Draxl, C.; Scheffler, M. The NOMAD Laboratory: From data sharing to artificial intelligence. *J. Phys. Mater.* **2019**, *2*, 036001. [\[CrossRef\]](#)
- Talirz, L.; Kumbhar, S.; Passaro, E.; Yakutovich, A.V.; Granata, V.; Gargiulo, F.; Borelli, M.; Uhrin, M.; Huber, S.P.; Zoupanos, S.; et al. Materials Cloud, a platform for open computational science. *Sci. Data* **2020**, *7*, 299. [\[CrossRef\]](#) [\[PubMed\]](#)

22. ENEA. IEMAP Ontology for Experimental and Computational Materials Metadata, Version v1. *Zenodo* **2025**. [[CrossRef](#)]
23. Cheng, J.; Zhang, C.; Dong, L. A geometric-information-enhanced crystal graph network for predicting properties of materials. *Commun. Mater.* **2021**, *2*, 92. [[CrossRef](#)]
24. RDA FAIR Data Maturity Model Working Group. *FAIR Data Maturity Model: Specification and Guidelines*; Research Data Alliance: Oxfordshire, UK, 2020. [[CrossRef](#)]
25. Devaraju, A.; Huber, R. An automated solution for measuring the progress toward FAIR research data. *Patterns* **2021**, *2*, 100370. [[CrossRef](#)] [[PubMed](#)]
26. Wang, H.; Yang, B.; Liao, X.-Z.; Xu, J.; Yang, D.; He, Y.-S.; Ma, Z.-F. Electrochemical properties of $\text{P2-Na}_{2/3}[\text{Ni}_{1/3}\text{Mn}_{2/3}]\text{O}_2$ cathode material for sodium ion batteries when cycled in different voltage ranges. *Electrochim. Acta* **2013**, *113*, 200–204. [[CrossRef](#)]
27. iemap-mi: Italian Energy Materials Acceleration Platform Python Client. PyPI Project Page. Available online: <https://pypi.org/project/iemap-mi/> (accessed on 15 June 2026).
28. Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Wang, Z.; Shysheya, A.; Crabbé, J.; Ueda, S.; et al. A generative model for inorganic materials design. *Nature* **2025**, *639*, 624–632. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.