D7.5 – Battery Interface Ontology published according to standards in the European modelling community

VERSION

VERSION	DATE
V0.1	2024-02-27

PROJECT INFORMATION

GRANT AGREEMENT NUMBER	957189
PROJECT FULL TITLE	Battery Interface Genome - Materials Acceleration Platform
PROJECT ACRONYM	BIG-MAP
START DATE OF THE PROJECT	1/9-2020
DURATION	3 years
CALL IDENTIFIER	H2020-LC-BAT-2020-3
PROJECT WEBSITE	big-map.eu

DELIVERABLE INFORMATION

WP NO.	7
WP LEADER	SINTEF
CONTRIBUTING PARTNERS	EPFL, DTU, WUT
NATURE	Report
AUTHORS	Simon Clark
CONTRIBUTORS	Eibar Flores
CONTRACTUAL DEADLINE	2024-02-29
DELIVERY DATE TO EC	2024-02-27
DISSEMINATION LEVEL (PU/CO)	PU

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. The project is part of BATTERY 2030+, the large-scale European research initiative for inventing the sustainable batteries of the future.

ABSTRACT

This report outlines the process of publishing the Battery Interface Ontology (BattINFO) in accordance with the standards set forth by the Elementary Multiperspective Material Ontology (EMMO). BattINFO aims to provide a comprehensive framework for describing battery interfaces, facilitating interoperability and knowledge sharing in battery research. BattINFO ensures compatibility with other ontologies by adhering to EMMO standards and promotes a unified approach to materials science ontology development. This report discusses the design principles, structure, and implementation details of BattINFO, highlighting its alignment with EMMO guidelines.

TABLE OF CONTENTS

1. Introduction

Battery research and innovation face a persistent challenge: the absence of consistent terminology and standardized frameworks for describing battery data, materials, cells, and interfaces. To facilitate collaboration across disciplines and industries, a clear and universally understood language is essential for effective communication and data interoperability.

BattINFO—the Battery Interface Ontology—addresses this challenge by establishing a common vocabulary and structure for characterizing battery interfaces. It aims to bridge the gap between diverse research communities and facilitate seamless data interoperability in the battery and materials science domain. Aligned with the Elementary Multiperspective Material Ontology (EMMO), BattINFO seeks compatibility and coherence with other material ontologies, promoting a unified approach to materials science ontology development.

BattINFO is available as an open-source software resource on GitHub: https://github.com/BIG-MAP/BattINFO. It can also be accessed from Python using the EMMOntoPy[1] package developed to work with EMMO and associated domain ontologies.

This report discusses the process of publishing BattINFO according to EMMO standards, highlighting the significance of clear, grounded language in enabling collaboration, innovation, and knowledge exchange in battery research and development.

2. Design principles of BattINFO

BattINFO's design principles are rooted in the necessity for clarity, flexibility, interoperability, and adherence to Linked Data principles within the domain of battery interface description. Clarity ensures that concepts are well-defined and easily understood by humans and machines. Flexibility allows BattINFO to accommodate diverse battery specifications and use cases without imposing unnecessary constraints. Interoperability ensures that BattINFO can seamlessly integrate with other ontologies and data sources, enabling cross-disciplinary research and knowledge sharing.

By adhering to Linked Data principles, BattINFO further enhances its interoperability by utilizing URIs to uniquely identify concepts and leveraging RDF to represent data in a structured and interconnected manner. This approach facilitates the seamless exchange and integration of battery data with other Linked Data resources, fostering a more comprehensive understanding of battery materials and enabling new insights through data linkage and analysis across different domains.

To achieve these design principles, BattINFO follows a modular approach, breaking down complex battery interface descriptions into manageable components. Each component represents a distinct aspect of the battery, such as material composition, electrochemical properties, or geometric structure. By modularizing the ontology, BattINFO promotes reusability and extensibility, allowing researchers to combine and extend existing concepts to suit their specific needs.

[1] https://github.com/emmo-repo/EMMOntoPy

Furthermore, BattINFO leverages standardized terminology and ontological patterns to enhance clarity and interoperability. Terms and definitions are carefully curated to align with established conventions in battery research, reducing ambiguity and promoting consistency across different relationships between entities in a semantically rich and computationally tractable manner.

In line with EMMO standards, BattINFO adopts a multiperspective approach to ontology design, capturing various viewpoints and scales of battery interface description. This multiperspective framework allows researchers to model battery interfaces at different levels of abstraction, from macroscopic phenomena to atomic-scale interactions. By accommodating diverse perspectives, BattINFO enables holistic modeling of battery interfaces, fostering deeper insights into their structure, function, and behavior.

The key features of BattINFO are: Persistent Identifiers, Standardized Nomenclature, Integration with the EMMO Universe, and Compliance with World Wide Web Consortium (W3C) Standards and Best Practices. These are described in detail below.

Persistent identifiers

BattINFO assigns persistent machine-readable identifiers to concepts from the battery domain. These identifiers are called uniform resource identifiers (URIs), facilitating data exchange and interoperability for humans and machines. The ontology uses the w3id.org Permanent Uniform Resource Locator (PURL) service to ensure that the identifiers are robust and permanent. For example, the URI for the concept of a BatteryCell is:

https://w3id.org/emmo/domain/battery#battery_68ed592a_7924_45d0_a108_94d6275d57f0

We can see that this URI is defined within the w3id.org namespace for EMMO (https://w3id.org/emmo) under the ‘domain’ ‘battery ‘and has the globally unique identifier ‘battery_68ed592a_7924_45d0_a108_94d6275d57f0’. Try clicking the link above. Users who request this URI from an internet browser are automatically re-directed to the reference documentation hosted under the EMMO organization. On the other hand, if it is requested from an application, it is re-directed to the machine-readable .ttl ontology file.

Standardized nomenclature

The ontology builds on standardized battery nomenclature, relying on recognized authorities, including IUPAC[2] and the IEC[3]. IUPAC is the universally recognized authority on chemical nomenclature and terminology. IEC is the world's leading organization that prepares and publishes International Standards for all electrical, electronic, and related technologies. This consistency in naming conventions enhances collaboration and data sharing.

Integration with the EMMO universe

BattINFO is defined under the recommendations of the EMMO[4]. The EMMO provides a top-level ontology describing knowledge related to the natural sciences and engineering. It hosts a variety of

[2] https://iupac.org/what-we-do/nomenclature/
[3] https://www.electropedia.org/
[4] http://www.emmo.info/

domain ontologies on a wide range of topics, including characterization methodology, crystallography, electrochemistry, and more. By defining BattINFO according to the recommendations of the EMMO, we can establish interoperability with all other resources in the EMMO Universe.

Compliance with W3C standards and best practices

The World Wide Web Consortium is the organization responsible for maintaining the standards that form the foundation for the Web. To streamline the exchange and interpretation of Web-based data, they have created a set of standards that form the basis for the Semantic Web: an extension of the World Wide Web that is centred around data and designed to be navigated by machines. The Resource Description Framework (RDF) is one of the cornerstone standards for the Semantic Web that provides a standard model for data exchange.

BattINFO is defined within these W3C recommendations to allow for easy integration of data into the Semantic Web. Furthermore, we provide examples of data annotation in some of the most common RDF formats, like JSON-LD.

3. Publishing BattINFO according to EMMO and W3C standards

BattINFO's compliance with EMMO standards is assessed based on its adherence to EMMO's conceptual framework, categories, relations, and guidelines for ontology development. EMMO defines a multiperspective approach to material ontology, encompassing various levels of description ranging from macroscopic properties to atomic-scale phenomena. BattINFO aligns with this multiperspective framework by providing a comprehensive and systematic representation of battery interfaces, capturing diverse viewpoints and scales of analysis.

One of the key guidelines for ontology development is the re-use of existing terms and ontologies. To facilitate that, EMMO recommends a modular system of domain ontologies that can be imported by others. In addition to the top-level concepts defined in EMMO itself, a battery ontology requires information about not only batteries but also electrochemistry and chemical substances. One could define all of that information in the battery ontology itself, but that would make re-use difficult. For example, if someone wanted to re-use the electrochemistry terms from the battery ontology in a fuel cell ontology, they would need to import the entire battery domain, which could create some practical challenges. Therefore, the battery interface ontology is broken down into constituent domain ontologies defined under the EMMO umbrella and imported together into one framework.

Table 1 shows an overview of the EMMO domain ontologies that comprise BattINFO. At the top level, EMMO itself provides the general framework for the ontology logic and universal terms. Below that, we have developed a domain ontology for chemical substances to describe specific chemical substances. The initial focus is substances that are useful to batteries and electrochemistry but can grow to eventually support other domains. The next level down, we have developed a domain ontology for electrochemistry that could also eventually be used to support similar activities on fuel cells, supercaps, or corrosion. Finally, at the bottom level is the domain ontology for batteries. Publishing the ontology in this way allows users to benefit from the interoperability of domains within the EMMO universe.

Table 1. Overview of EMMO domain ontologies that comprise the Battery Interface Ontology.

Imported EMMO domains	URL
EMMO	https://github.com/emmo-repo/EMMO
domain-chemicalsubstance	https://github.com/emmo-repo/domain-chemicalsubstance
domain-electrochemistry	https://github.com/emmo-repo/domain-electrochemistry
domain-battery	https://github.com/emmo-repo/domain-battery
BattINFO	https://github.com/BIG-MAP/BattINFO

BattINFO also incorporates EMMO's recommendations for ontology documentation, including metadata, versioning, and licensing information. This documentation ensures that BattINFO is well-documented and easily accessible to the broader materials science community, facilitating its adoption and use in battery research. This is discussed in detail in the following section.

Overall, BattINFO's compliance with EMMO standards demonstrates its compatibility with other material ontologies developed within the EMMO framework, promoting interoperability, consistency, and knowledge sharing in the field of battery research.

4. Implementation and documentation of BattINFO

The implementation of BattINFO involves the translation of its conceptual design into a machine-readable format using ontology representation languages and tools. One commonly used language for ontology development is the Web Ontology Language (OWL), which provides a rich set of constructs for defining classes, properties, and relationships within an ontology.

Implementation

BattINFO's implementation in OWL follows best practices for ontology engineering, including modularization, reuse, and documentation. The ontology is divided into separate modules corresponding to different aspects of the battery interface, each containing a coherent set of classes, properties, and axioms. This modular structure facilitates maintenance and updates, allowing researchers to extend and refine BattINFO over time without disrupting existing applications.

In addition to OWL, BattINFO supports other RDF technologies such as JSON-LD (Java Script Object Notation for Linked Data) and SPARQL (SPARQL Protocol and RDF Query Language) for semantic data representation and querying. RDF provides a standardized format for representing data as subject-predicate-object triples. JSON-LD is an RDF serialization format that uses the well-known JSON structure to define RDF triples in a friendly and human-readable way. While JSON-LD allows us to define semantic data triples, the query language SPARQL enables expressive queries over RDF datasets, allowing researchers to retrieve and analyze battery interface information stored in BattINFO.

To enhance interoperability, BattINFO incorporates semantic annotations and metadata following established vocabularies and standards such as Dublin Core and Schema.org, describing bibliographic data and web resources, respectively. These annotations provide additional context and semantics to BattINFO entities, making them more interpretable and discoverable by both humans and machines.In particular, the use of Schema.org markup renders data discoverable by popular search engines such as Google, Microsoft, Yandex and Yahoo!.

Documentation

The reference documentation for BattINFO serves as a comprehensive resource for understanding the ontology's structure, concepts, and usage. Hosted on the website https://big-map.github.io/BattINFO/index.html , this documentation provides researchers, developers, and stakeholders with access to detailed information about BattINFO's design principles, terminology, and implementation (Figure 1).

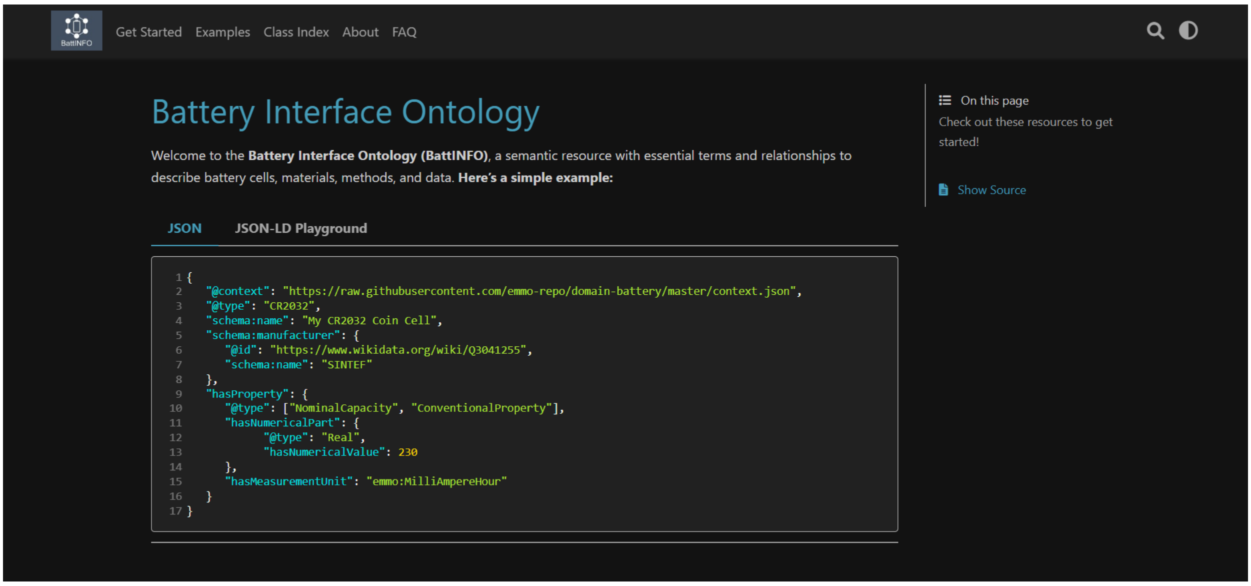

Figure 1. Screenshot from the BattINFO Reference Documentation.

The reference documentation includes many examples of BattINFO usage, two of which are summarized in this report.

Figure 2 shows a simple example of how to use BattINFO to create a JSON-LD description of a CR2032 coin cell. An interactive version of this example is available in the JSON-LD playground. The context file is hosted under the EMMO organization and includes mappings between the labels and URIs for terms in both EMMO and BattINFO. This allows researchers to create JSON-LD files using human-readable terms that can be translated into machine-readable URIs.

For example, when we make the statement “@type”: “CR2032”, JSON-LD parsers will retrieve the context file and see that the label “CR2032” corresponds to the URI: http://emmo.info/battery#battery_b61b96ac_f2f4_4b74_82d5_565fe3a2d88b

Because that term is then linked to material definitions in the ontology, it can infer that this is a battery cell with a lithium negative electrode and a manganese dioxide positive electrode. Furthermore, by importing EMMO definitions for physical quantities and units, we can re-use existing terms that allow us to express properties in a way that can be understood and queried within the EMMO universe. Finally, by linking to other recognized RDF vocabularies like schema.org and knowledge graphs like wikidata, we can integrate our data and retrieve additional information from other web-based sources.


```
{
   "@context":     "https://raw.githubusercontent.com/emmo-repo/domain-battery/master/context.json",
   "@type": "CR2032",
   "schema:name": "My CR2032 Coin Cell",
   "schema:manufacturer": {
      "@id": "https://www.wikidata.org/wiki/Q3041255",
      "schema:name": "SINTEF"
   },
   "hasProperty": {
      "@type": ["NominalCapacity", "ConventionalProperty"],
      "hasNumericalPart": {
            "@type": "Real",
            "hasNumericalValue": 230
      },
      "hasMeasurementUnit": "emmo:MilliAmpereHour"
   }
}
```


Figure 2. Example metadata description of a CR2032 coin cell using BattINFO and JSON-LD.

While JSON-LD is appropriate for representing knowledge that can be expressed as triples, it generally struggles to describe large time-series datasets. For those types of datasets, it is best to leave the raw data in a suitable format (e.g. as csv, tsv, excel, parquet, etc.) and instead use JSON-LD and ontology terms to describe the metadata.

Figure 3 shows an example of how to combine BattINFO with the csvw vocabulary to describe tabular time-series data from a cell discharge test. The csvw vocabulary is a W3C recommended resource for describing tabular data on the web[5]. Basically, it allows developers to create a JSON-LD schema that describes a table with ontology terms. The exact format of the table is not critical, and the same approach can be used for data in csv, tsv, excel or other files. In this example, we follow the W3C recommendations for expressing the general structure of the tabular data using csvw terms (e.g. csvw:tableSchema, csvw:columns, csvw:propertyUrl, etc) and use BattINFO terms to describe the physical quantities expressed in the table (e.g. Time, CellVoltage, CellCurrent, etc.). A complete notebook of this example is available in the BattINFO Reference Documentation[6].

[5] https://www.w3.org/TR/2015/REC-tabular-metadata-20151217/
[6] https://big-map.github.io/BattINFO/examples.html

```
{
            "@context":        "https://raw.githubusercontent.com/emmo-
repo/domain-battery/master/context.json",
            "@type": "MeasurementResult",
            "csvw:url":         "https://raw.githubusercontent.com/BIG-
MAP/BattINFO/master/sphinx/assets/data/discharge_timeseries.csv",
            "dc:title": "Synthetic Discharge Curve Demo",
            "dcat:keyword": ["battery", "li-ion", "discharge"],
            "dc:license":                                       {"@id":
"http://opendefinition.org/licenses/cc-by/"},
            "dc:modified":    {"@value":     "2024-01-10",     "@type":
"xsd:date"},
            "csvw:tableSchema": {
                "csvw:columns": [{
                  "csvw:name": "time",
                  "csvw:titles": "Time / s",
                  "csvw:propertyUrl": {
                    "@type": "Time"
                  },
                  "hasMeasurementUnit": {
                    "@type": "Second"
                  },
                  "csvw:datatype": "number",
                  "csvw:required": "true"
                }, {
                  "csvw:name": "voltage",
                  "csvw:titles": "Voltage / V",
                  "csvw:propertyUrl": {
                    "@type": "CellVoltage"
                  },
                  "hasMeasurementUnit": {
                    "@type": "Volt"
                  },
                  "csvw:datatype": "number"
                }, {
                  "csvw:name": "current",
                  "csvw:titles": "Current / A",
                  "csvw:propertyUrl": {
                    "@type": "CellCurrent"
                  },
                  "hasMeasurementUnit": {
                    "@type": "Ampere"
                  },
                  "csvw:datatype": "number"
                }],
                "csvw:primaryKey": "time",
                "csvw:aboutUrl": "#time-{time}"
              }
        }
```


Figure 3. An example showing how to combine BattINFO with external RDF vocabularies like csvw to describe tabular data like a battery cell discharge curve.

Users not familiar with serialization format can still annotate their datasets via easy-to-use web apps. We aim to build an ecosystem of user-friendly applications where datasets can be annotated interactively, reducing the technical barrier for users to create linked data descriptions based on BattINFO. For example, we have recently developed a prototype metadata annotator for CSV files, where users drag and drop their csv file and interactively link columns to BattINFO concepts. After annotations, users can simply download the JSON-LD description of their csv file. The app has been available open source and deployed in the cloud so it can be used without needing installation.

Overall, the implementation of BattINFO involves a combination of ontology engineering principles, semantic web technologies, and best practices for data interoperability, ensuring that BattINFO is both machine-readable and semantically rich, facilitating its integration with existing data sources and applications in battery research.

5. Summary

The publication of BattINFO according to EMMO standards represents a significant milestone in the field of battery research. By providing a comprehensive and systematic framework for describing battery interfaces, BattINFO enables researchers to capture diverse perspectives and scales of analysis within a unified ontology. The modular structure, clarity, and interoperability of BattINFO facilitate its adoption and use in a wide range of applications, from materials design and characterization to battery performance modelling and optimization.

Moving forward, it is essential to continue refining and expanding BattINFO in line with evolving EMMO guidelines and emerging developments in battery technology. This includes incorporating new concepts, properties, and relationships to address emerging research challenges and applications. Furthermore, efforts to promote the adoption and dissemination of BattINFO within the materials science community will be crucial for maximizing its impact and utility in advancing battery research and development.

In summary, the publication of BattINFO according to EMMO standards underscores the importance of interoperability, consistency, and collaboration in ontology development. By embracing a multi-perspective approach and aligning with established standards, BattINFO contributes to a broader ecosystem of material ontologies, facilitating interdisciplinary research, innovation, and knowledge exchange in the field of materials science and engineering.