D7.3 – First stable release of the battery interface ontology

VERSION

VERSION	DATE
1.0	25.02.2022

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	DTU, WUT, NIC, EPFL,
NATURE	Report
AUTHORS	Simon Clark
CONTRIBUTORS	J. Friis, F.L. Blekken, C. Andersen
CONTRACTUAL DEADLINE	28.02.2022
DELIVERY DATE TO EC	25.02.2022
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

The Battery Interface Ontology (BattINFO) is a domain ontology for batteries and electrochemistry, which aims to provide a basis for creating machine-readable semantic linked battery data. By mapping data to entities in BattINFO, users can ensure that their data is easily interoperable with other BattINFO-compliant datasets. Furthermore, the relationships defined in BattINFO can help elucidate connections between pieces of data which may not be readily apparent. BattINFO has been in development since Autumn 2020, and the repository is already open to the public. The purpose of this deliverable is to announce the release the first stable version, which will allow users to begin integrating BattINFO into their workflows.

TABLE OF CONTENTS

1. Introduction

Battery data plays an essential role in accelerating the development of new materials, cell designs, models, and much more. As large-scale initiatives in both industry and research start to gain momentum, laboratories and Gigafactories around the world are generating vast amounts of battery data. Battery designers and researchers today have access to an unprecedented amount of information. This is an important development that enables novel approaches such as artificial intelligence and laboratory automation.

However, battery data is generated by many different groups around the world, each of whom can use different structures, metadata schemas, or descriptive terms. To address this challenge, the community needs a universal way of describing and sharing battery data, based on a common conceptualization. This conceptualization can be embodied in a machine-readable battery language, containing both terms and relations needed to describe batteries and their data. Using this language to annotate battery data with standardized terms will allow data generated by one laboratory or instrument to be seamlessly understood and used by a different laboratory or instrument. Creating machine-readable linked battery data can greatly accelerate the pace of research and development.

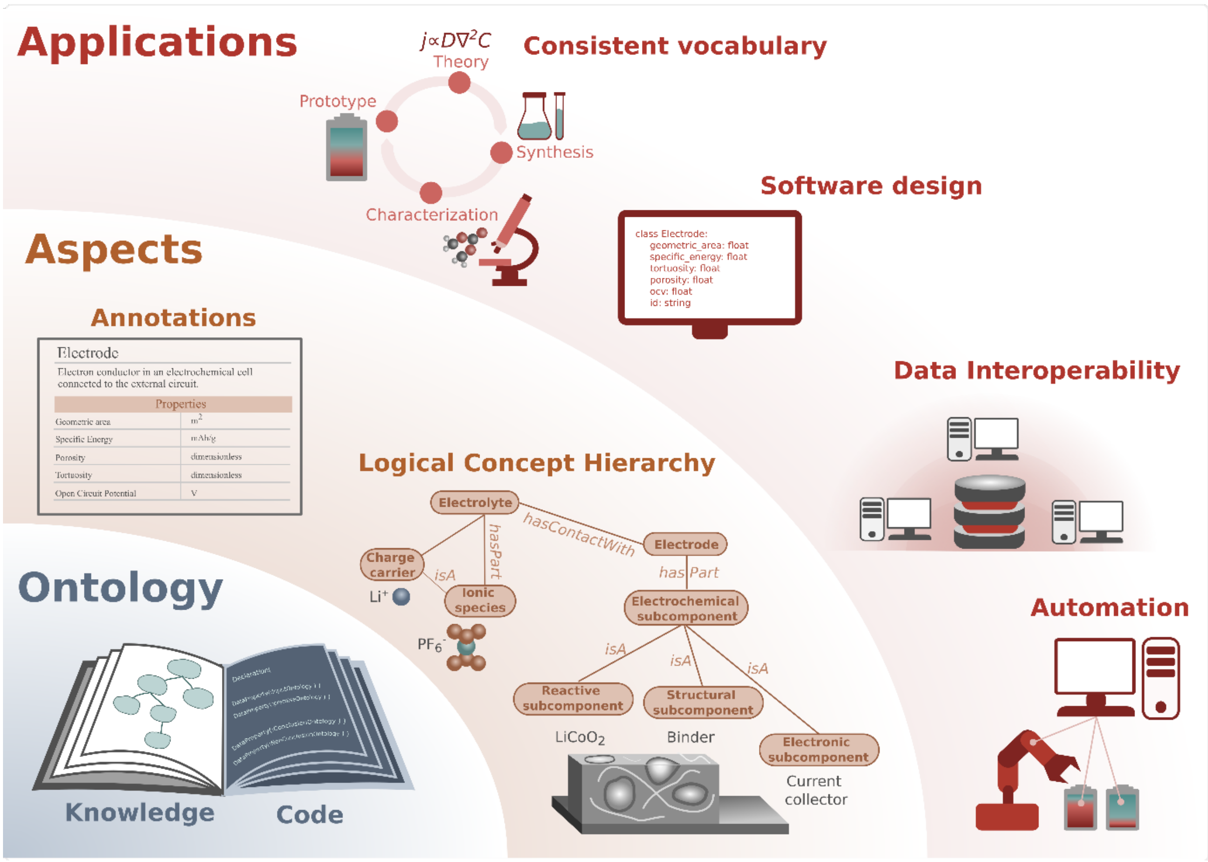
The FAIR Guiding Principles for scientific data were created as a first step to help address this need. FAIR aims to enhance the ability of both humans and machines to (automatically) find and reuse data, and it stipulates that scientific data management should adhere to four core principles: Findability, Accessibility, Interoperability, and Reusability. The FAIR principles are not a formal standard and they do not establish specific technical requirements, rather they define guidelines.

Among the criteria for FAIR data sharing are the requirements that:

- (meta)data use vocabularies that follow FAIR principles,
- (meta)data include qualified references to other (meta)data, and
- (meta)data meet domain-relevant standards.

The development of community-driven ontologies offers a solution well-suited to address these needs.

An ontology provides a common conceptualization of a domain, which is represented as a map of concepts and the relationships between the concepts. Often taking the form an annotated machine-readable knowledge graph, an ontology ascribes meaning to data, provides links to other pieces of data, and allows for machine reasoning according to that meaning. An overview of the aspects and applications of an ontology is shown in Figure 1. More introductory information about ontologies and their role in data interoperability can be found in our publications [1].



Conceptual overview of an ontology, its aspects, and applications. The core of an ontology is knowledge, which can be expressed as a map of concepts and relations, expressed in machine-readable code. Aspects of an ontology include annotations, which provide context and elucidations of concepts and relations that support the arrangement of concepts in a logical hierarchy. An ontology defines a consistent vocabulary and common conceptualization, which has applications that include software design, data interoperability, and automation. Reprinted with permission from Ref. [1]**.**

Until now, the ontologies needed to fulfil the FAIR criteria and enable the large-scale interoperability of battery data simply did not exist. The BIG-MAP project seeks to address this need with the development of the Battery Interface Ontology (BattINFO).

The purpose of this deliverable is to confirm that the first stable release of BattINFO has been made. BattINFO has been developed since the Autumn of 2020 and the repository has been open to the public since February 2021. Ontologies are living projects and are constantly evolving to meet the needs of the community. The first stable release provides users with the main terms and relations they need to start annotating their data and provides consistent namespace and IRIs that will be maintained throughout the life of the ontology.

The ontology itself contains the information and tools resulting from this deliverable that will be of greatest interest to the BIG-MAP project and the battery community. Readers are encouraged to download and explore the ontology directly. The GitHub repository also allows users to communicate directly with the development team.

The remainder of this report provides an overview of BattINFO with instructions for how to obtain and explore the ontology. The ontology itself contains hundreds of terms and relations to support annotating battery data. Some of the main highlights are presented and discussed in this report. Finally, the plans for the continuing development of the ontology are discussed.

2. Battery Interface Ontology

2.1 Obtaining the ontology

BattINFO is freely available from the BIG-MAP organization page on GitHub (https://github.com/BIG-MAP/BattINFO), and it is shared under the permissive Creative Commons Attribution 4.0 International (CC-BY-4.0) license.

There are two main ways users are encouraged to interact with the ontology:

- **For new users**, the free ontology editor Protégé (https://protege.stanford.edu/) offers an easy-to-use interface to explore and learn more about BattINFO.
- **For more advanced users** who would like to interact with the ontology programmatically, the package EMMOntoPy (https://github.com/emmo-repo/EMMO-python) is designed to provide an intuitive representation of EMMO-based ontologies in Python.

BattINFO builds on a pre-inferred version of the Elementary Multiperspective Material Ontology (EMMO) top and middle ontologies. The correct path to the inferred version of EMMO is specified in the catalog file, catalog-v001.xml.

To obtain BattINFO, the first step is to clone the repository:

```
git clone https://github.com/BIG-MAP/BattINFO.git
```

When opening BattINFO.ttl in Protégé, the correct version of emmo-inferred will be downloaded and imported.

In EMMOntoPy, correct import is obtained with:

```
from ontopy import get_ontology

# Loading from local repository
battinfo =
get_ontology('/path/to/BattINFO/battinfo.ttl').load(url_from_
catalog=True)

# Loading from web
battinfo =
get_ontology('https://raw.githubusercontent.com/BIG-
MAP/BattINFO/master/battinfo.ttl').load()
```

2.2 Import structure of the ontology

The Battery Interface Ontology (BattINFO) is defined as a domain ontology of the Elementary Multiperspective Material Ontology (EMMO). The EMMO is a multidisciplinary top- and middle level ontological framework for applied sciences and engineering. EMMO is designed to address the needs for a semantic description which is deeply rooted in the physical sciences, incorporating:

i. description of materials from a rigorous physics perspective;
ii. formal relations between granularity levels to facilitate multiscale materials description;
iii. definition of material processes to capture the changing and evolution of materials as chain of different states.

Defining BattINFO as an EMMO domain ontology allows BattINFO to take advantage of the extensive ontological framework already developed in EMMO. Furthermore, it facilitates integration with other related domain ontologies such as characterization, modelling, etc. For practical purposes, this means that BattINFO takes on a hierarchical structure, shown in Figure 2.

The top-level imports the inferred version of the current EMMO release. The next level down, we have defined an ontology for generic concepts, which are needed in BattINFO but are technically outside of the scope. Grouping these terms in generic concepts allows them to be reviewed for integration into EMMO or other domain ontologies. Next, BattINFO imports four ontologies with the necessary domain terms: electrochemical quantities, electrochemistry, battery quantities, and battery. The battery domain is closely coupled to electrochemistry. However, some other domains like corrosion might also wish to import electrochemistry without needing the entire battery ontology as well. For this reason, we have separated electrochemistry as a standalone domain ontology. Furthermore, the associated quantities for the domains are also separated into their own ontologies to facilitate re-use.

To view BattINFO in Protégé, the user need only open the file BattINFO.ttl. The importing of the constituent ontologies is performed automatically.

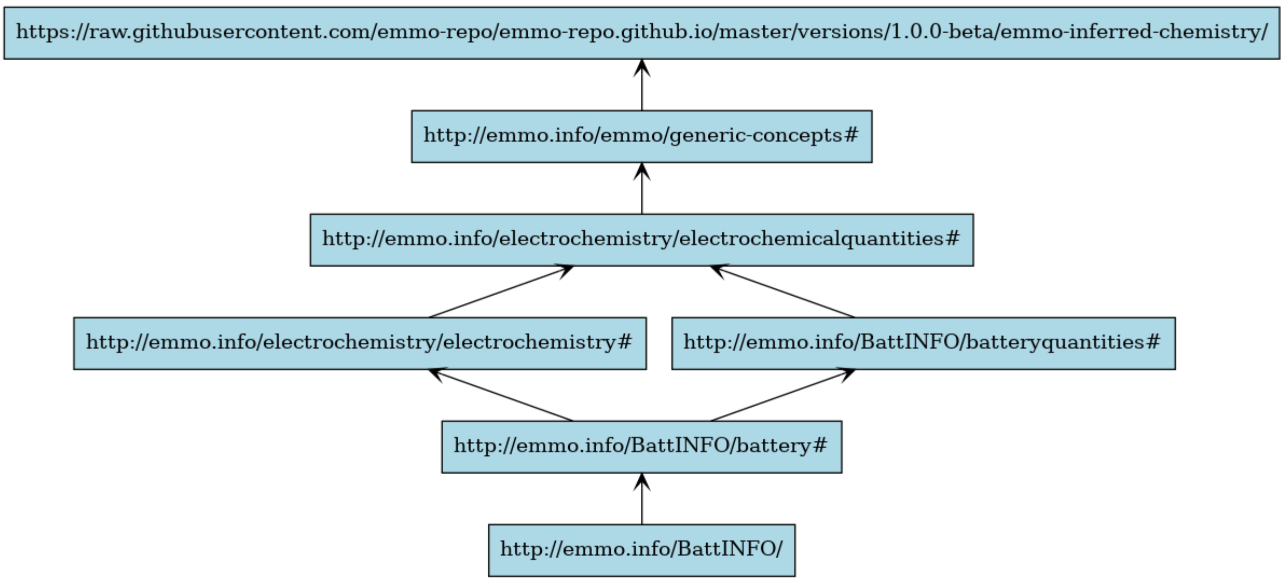


Figure 2. Overview of the BattINFO import structure.

2.3 Highlights from the ontology

BattINFO contains hundreds terms, complete with their own annotations and relationships to other terms. The best way to explore the ontology is to view it in Protégé. Alternatively, users can consult the documentation (https://big-map.github.io/BattINFO/). In this report, we provide an overview of some highlights.

Figure 3 shows an example of a term from BattINFO viewed in the Protégé environment [1]. The top row shows that the preferred label for the entity (skos:prefLabel) is 'Battery', with two alternative labels (skos:altLabel) 'ElectricBattery' or the German word 'Batterie' that could also be used. Although the primary development of BattINFO is in English, there are planned extensions to support labelling in other languages. The annotations provide links to the dbpedia and Wikipedia entries for the term, and provides the elucidation "One or more electrochemical cells fitted with devices necessary for use, for example case, terminals, marking and protective devices" taken from the IEC standard 60050-482 for electrotechnical vocabulary.

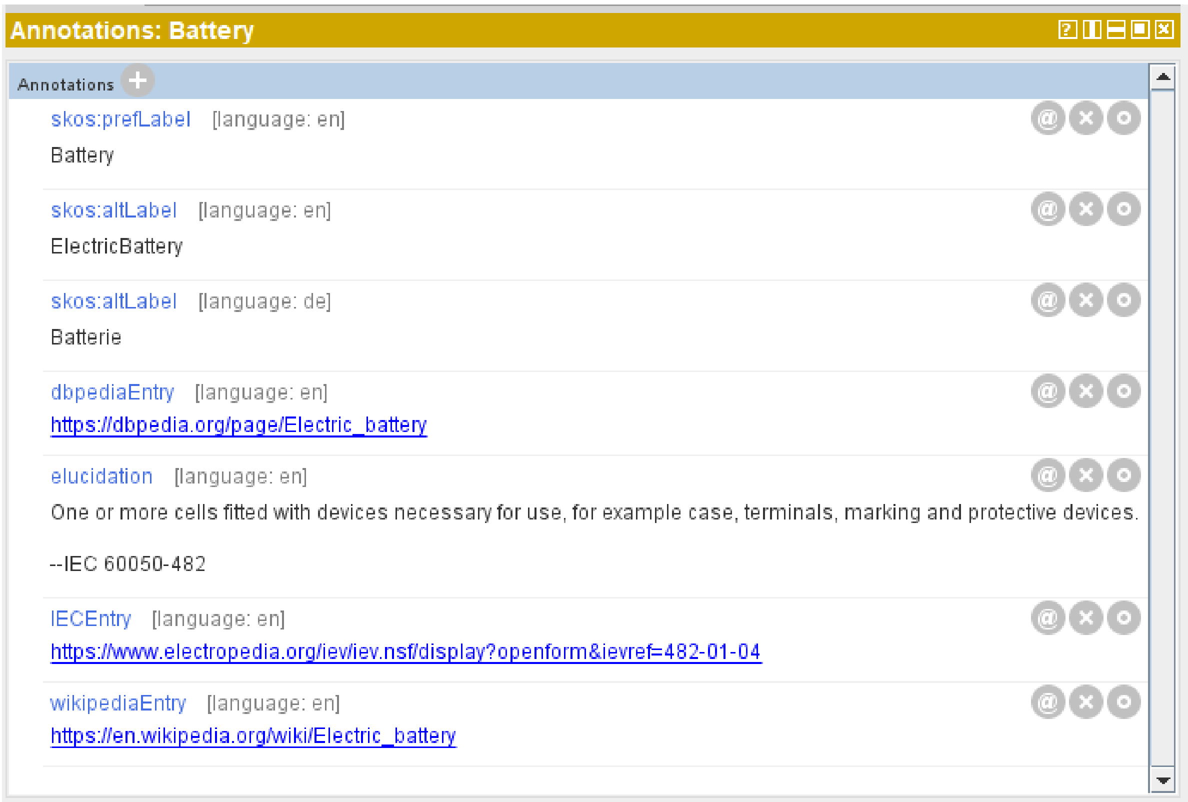


Figure 3. Example of annotations for an entity in BattINFO as visualized in Protégé, reprinted with permission from Ref. [1].

Each term defined in BattINFO contains annotations like those listed above. In this way, any piece of data that points back to the IRI of a term in BattINFO can automatically have access to not only the extra information provided in the annotations but also to the relations that exist from that entity to others. In the following sub-sections, we provide some examples using terms defined in BattINFO.

2.3.1 Quantities

From a data interoperability point of view, accurate and extensive descriptions of quantities are among the more important entities defined in an ontology. EMMO-based ontologies rely on standards like ISO 80000 and International Vocabulary of Metrology (VIM) for ontologizing quantities and their relationships to measurements. In EMMO, a `Quantity` is elucidated as:

> "A symbolic that has parts a reference unit and a numerical object separated by a space expressing the value of a quantitative property (expressed as the product of the numerical and the unit)."

For example, a quantity can be expressed as a symbolic construct such as, "6.8 m" or "6 MeV". A Quantity can be further classified as an `OrdinalQuantity`, `PhysicalQuantity`, and/or a `QuantitativeProperty`. An `OrdinalQuantity` is a quantity like hardness or resilience, for which the magnitude of the quantity can be compared with quantities of the same kind, but no algebraic expressions among those quantities exist. For the purpose of the battery and electrochemistry domains, most relevant quantities fall under the `PhysicalQuantity` term.

EMMO elucidates a `PhysicalQuantity` as:

> "A 'Mathematical' entity that is made of a 'Numeral' and a 'MeasurementUnit' defined by a physical law, connected to a physical entity through a model perspective. Measurement is done according to the same model."

Figure 4 shows a screenshot of the `PhysicalQuantity` entity description. A `PhysicalQuantity` inherits the structure of a `Quantity`, and adds the relation that it `hasReferenceUnit only MeasurementUnit`. A `MeasurementUnit` is a type of `ReferenceUnit` that `hasPhysicalDimension exactly 1 PhysicalDimension`. This allows the physical dimension of each physical quantity to be defined independently of the actual unit.

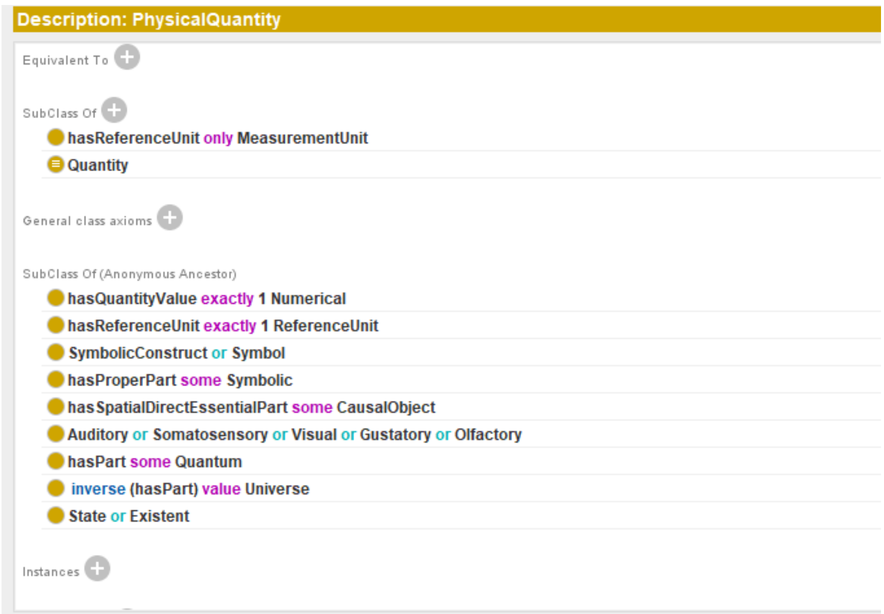


Figure 4. Screenshot of the description of PhysicalQuantity in EMMO.

Physical dimensions are expressed as a metrological symbol (or, for the ontologist, a string) with the form: Ta Lb Mc Id Θe Nf Jg, where a, b, c, d, e, f and g are 0 or signed integers and T, L, M, I, Θ, N, and J are time, length, mass, electric current, temperature, amount, and luminous intensity. For example, the physical dimension of:

- a volume is expressed as T0 L-3 M0 I0 Θ0 N0 J0,
- an amount is expressed as T0 L0 M0 I0 Θ0 N+1 J0,
- an amount concentration (i.e. amount per volume) is expressed as T0 L-3 M0 I0 Θ0 N+1 J0

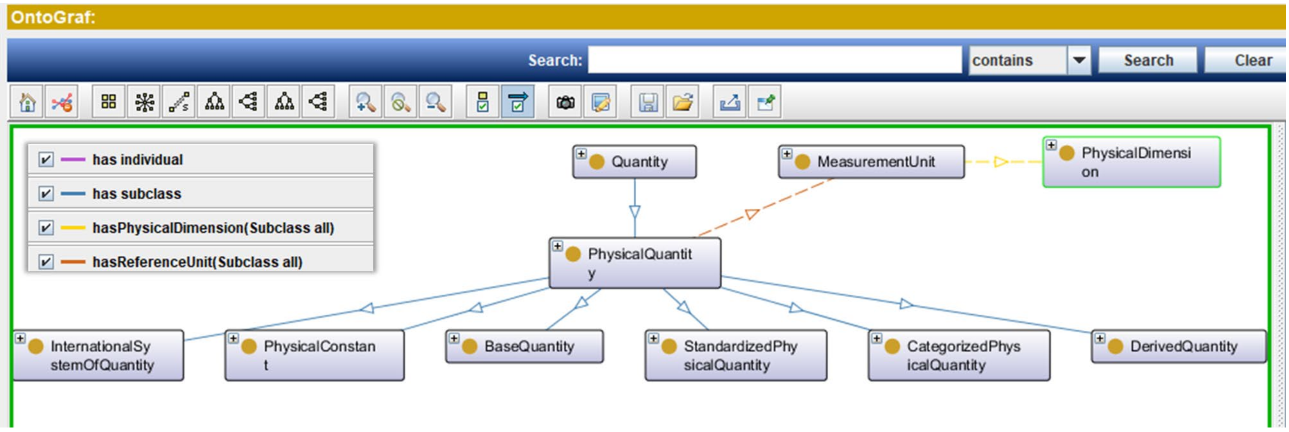


Figure 5. Overview of PhysicalQuantity relationships, visualized using the OntoGraf plugin in Protégé.

Figure 5 provides an overview of the relationships around `PhysicalQuantity`. A `PhysicalQuantity` can be further characterized as a `BaseQuantity`, `CategorizedPhysicalQuantity`, `DerivedQuantity`, `InternationalSystemOfQuantity`, `PhysicalConstant`, and `StandardizedPhysicalQuantity`. For more information about these quantities, readers are encouraged to explore the annotations and descriptions in the ontology itself.

ISO 80000 is an international standard that introduces the International System of Quantities (ISQ). As of February 2022, EMMO contains a selection of the most important quantities from ISO 80000, but it does not offer universal coverage. To be sure that BattINFO users have access to all of the quantities needed in batteries and electrochemistry, the BattINFO team has added terms for the missing ISO 80000 quantities in the generic-concepts ontology. These are marked for eventual integration into EMMO itself and will retain their IRIs after the move.

2.3.2 Electrochemistry

An electrochemistry domain ontology is necessary to support data interoperability not only for batteries, but also fuel cells, electrolysers, corrosion, etc. For this reason, the electrochemistry ontology is designed to facilitate re-use in other EMMO domains.

The electrochemistry vocabulary is implemented in accordance with the recommendations from both the International Electrotechnical Commission (IEC) and the International Union of Physical and Applied Chemistry (IUPAC). These recommendations are summarized in IEC 60050 [2] and the IUPAC Recommendations for Terminology of Electrochemical Methods of Analysis [3]. In addition to the standard vocabulary for entities, we have introduced some foundational relations for the electrochemistry ontology.

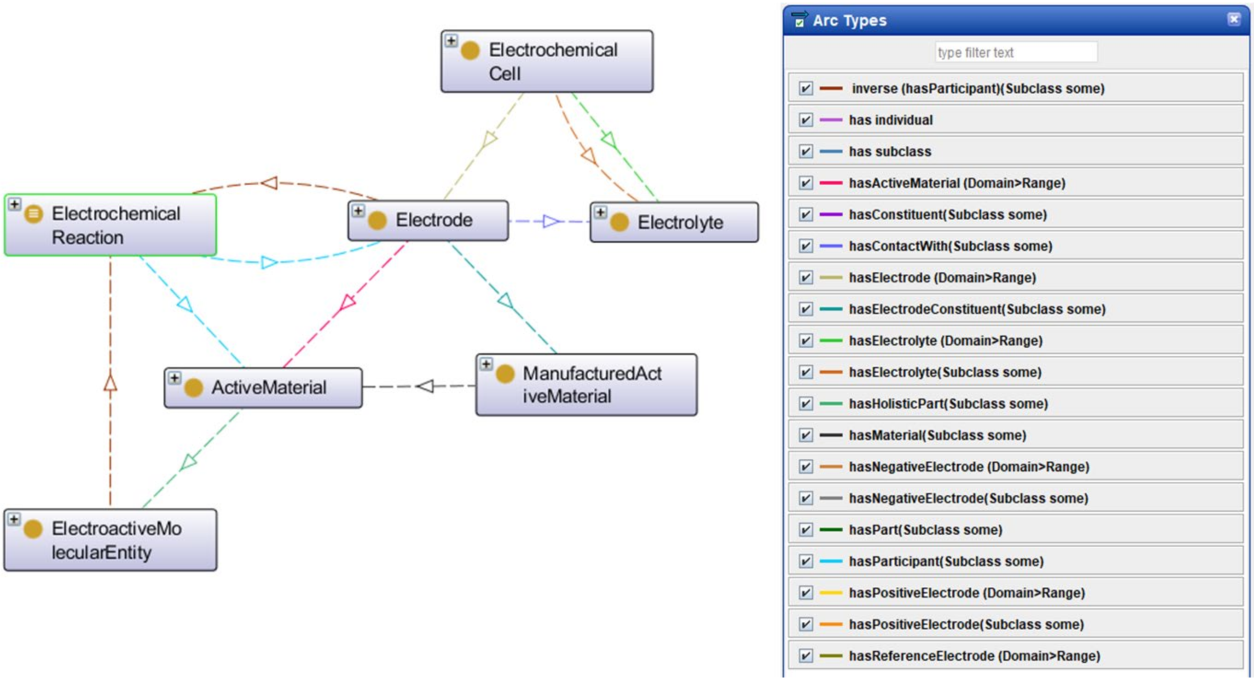


Figure 6. An overview of the foundational holistic systems in the electrochemistry domain ontology, visualized using the OntoGraf plugin in Protégé.

Figure 6 provides an overview of the holistic systems defined in the electrochemistry domain ontology. One of the main holistic systems in electrochemistry is the electrochemical cell, which comprises (at least) two electrodes and an electrolyte. Metadata describing the type of electrochemical cell in question, will nearly always relate back to these components. Therefore, we have defined the relationships `hasElectrode` (`domain: ElectrochemicalCell`, `range: Electrode`) and `hasElectrolyte` (`domain: ElectrochemicalCell`, `range: Electrolyte`). These relationships are defined in EMMO as a type of hasConstituent, which is designed to relate an object and one of its holistic parts that contributes to the object under some spatial-based criteria. If desired, the `hasElectrode` relationship can then be further refined as `hasPositiveElectrode`, `hasNegativeElectrode`, and `hasReferenceElectrode`.

Electrodes themselves can also be viewed as holistic systems, which can comprise one or more constituents. The typical example is a Li-ion battery electrode, which usually contains active material, binder, and conducting additives. To allow users to make these specifications in an efficient way, we have introduced the relation `hasElectrodeConstituent` (`domain: Electrode`, `range: ElectrodeConstituent`). All electrodes must have an active material, and the relation `hasActiveMaterial` (`domain: Electrode`, `range: ActiveMaterial`) is defined as a special case of `hasElectrodeConstituent`. In this notation, some typical electrodes might be ontologized as:

```
LithiumGraphiteElectrode hasActiveMaterial some GraphiteMaterial
LithiumGraphiteElectrode hasElectrodeConstituent some CMCMaterial
```

Or

```
ZincElectrode hasActiveMaterial some ManufacturedZincMaterial
```

Furthermore, to begin to link electrodes to the electrochemical reactions that take place there, we define that an active material must contain some molecular entity that is a participant in an electrochemical reaction. This is expressed in the ontology using the relations:

```
ActiveMaterial hasHolisticPart some ElectroactiveMolecularEntity
```

And

```
ElectroactiveMolecularEntity inverse(hasParticipant) some
ElectrochemicalReaction
```

In addition to the standard vocabulary from IUPAC and IEC and the relationships described above, the electrochemistry domain ontology includes a few examples for common types of materials, electrodes, and electrochemical cells to which users can refer.

2.3.3 Battery

The battery domain ontology extends the electrochemistry domain with terms and relations that are unique to batteries. The main holistic systems in the battery domain are shown in Figure 7. A `BatteryCell` is defied to comprise an `ElectrochemicalCell` and a `BatteryCellContainer`. The form factor of the battery is determined by the type of container. BattINFO includes terms to describe pouch cells, coin cells, cylindrical cells, Swagelok cells, and prismatic cells.

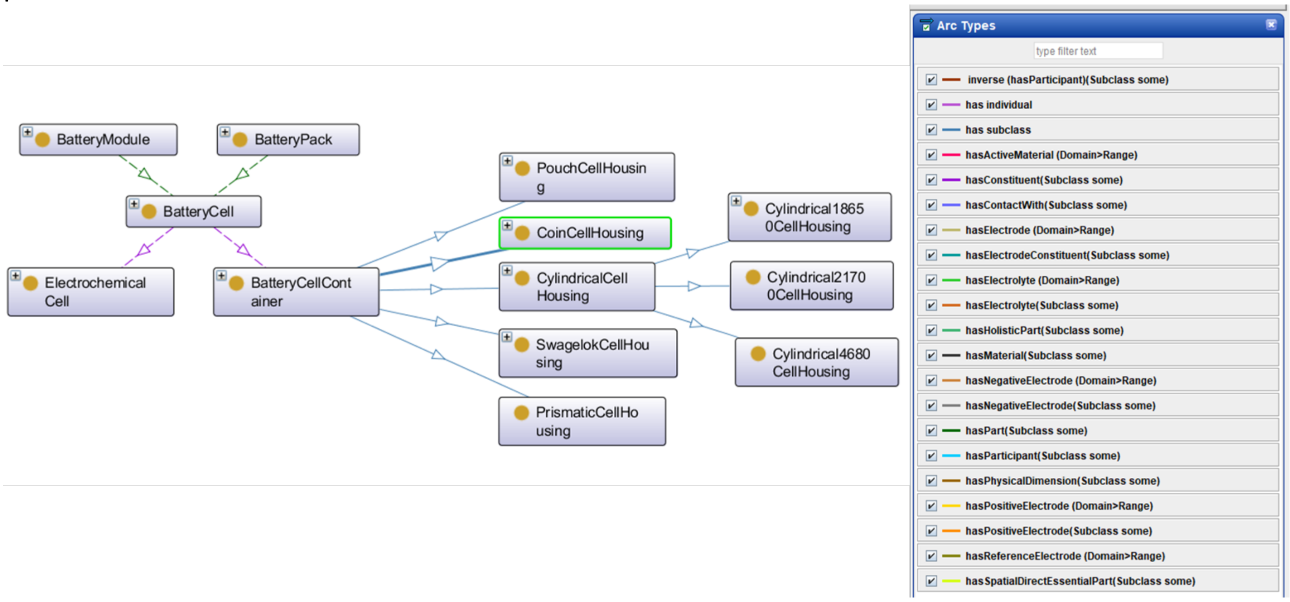


Figure 7. An overview of the foundational holistic system in the battery domain ontology, visualized using the OntoGraf plugin in Protégé.

BattINFO includes pre-defined electrochemical cells, which cover many of the most widely used cell chemistries in industry and research today. An overview of the types of electrochemical cells included is shown in Figure 8.

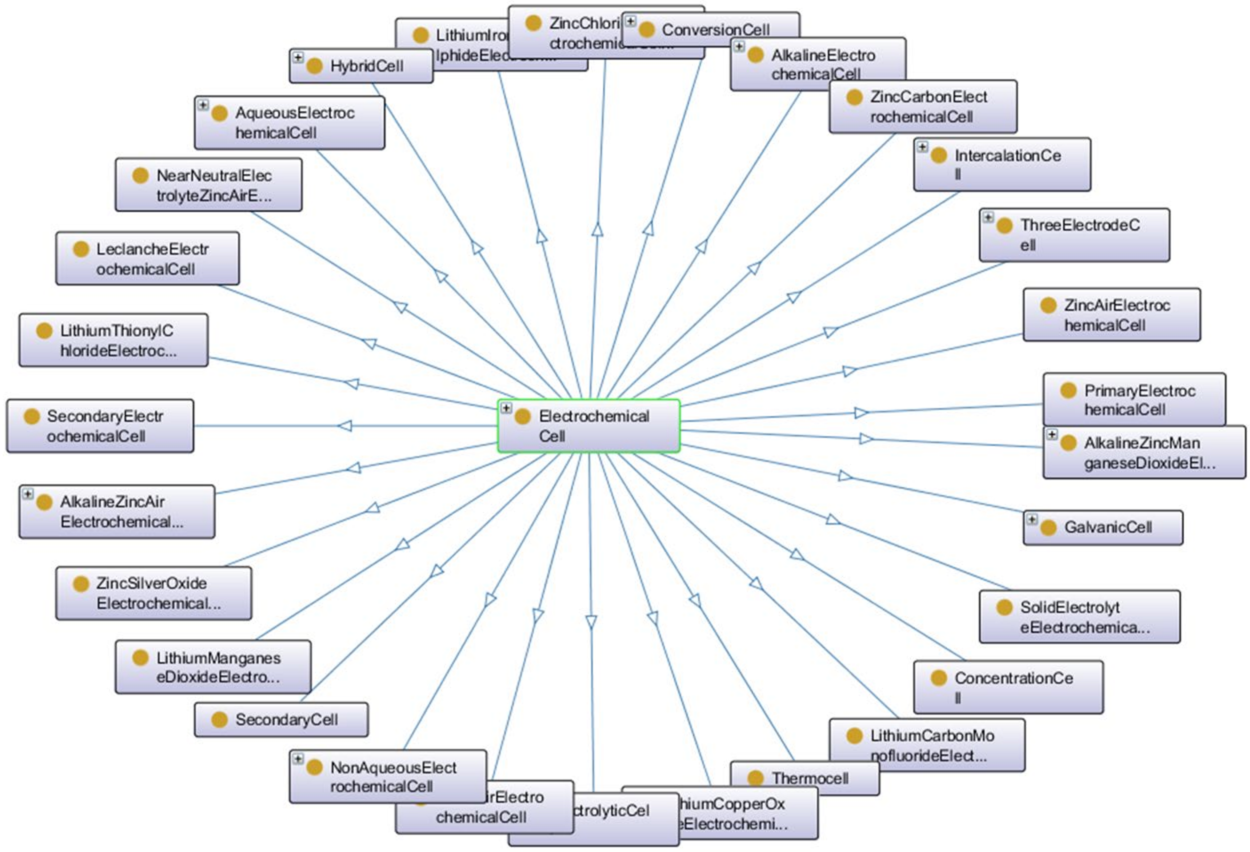


Figure 8. An overview of pre-defined types of electrochemical cells available in the battery ontology, visualized using the OntoGraf plugin in Protégé.

As an example, Figure 9 shows a detailed expansion of one type of electrochemical cell defined in BattINFO: an alkaline zinc-air electrochemical cell. The cell is ontologized using the `hasNegativeElectrode` and `hasPositiveElectrode` to distinguish the zinc and air electrodes, respectively. It is also specified that it `hasElectrolyte some AlkalineElectrolyte`. The exact type of electrolyte is not defined, because an embodiment of an alkaline zinc-air electrochemical cell could use many different types of electrolytes (e.g. KOH, NaOH, etc.). This would then need to be specified on the individual-level for the cell in question. The overview also shows the preferred electrochemical reactions that occur at the respective electrodes. The forward reaction is defined to take place during discharging and the back reaction during charging.

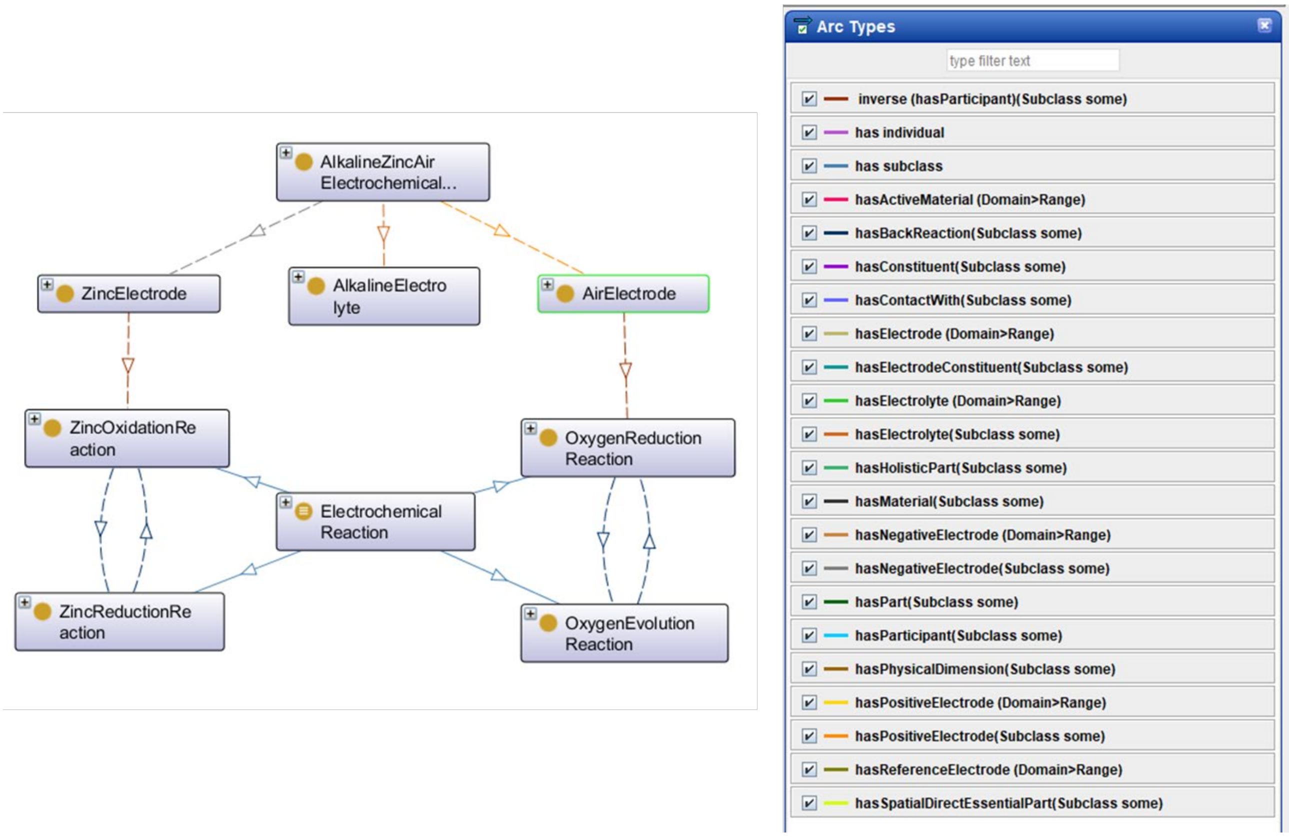


Figure 9. Overview of an alkaline zinc-air electrochemical cell, visualized using the OntoGraf plugin in Protégé.

3. Summary and Outlook

BattINFO provides battery researchers with semantically rich, machine-readable terms to annotate their data. In this first stable release, the ontology includes all of the vocabulary from common standards like ISO 80000 and IEC 60050 together with domain-specific recommendations from IUPAC. Together with a defined namespace and consistent IRIs, this provides researchers in BIG-MAP and beyond all the basic tools they need to start annotating their data and reaping the benefits of a linked open battery data environment.

Ontologies are living projects. They are constantly growing and evolving in order to respond to the needs of the community. This will also be the case with BattINFO. Over the course of its development, BattINFO will rely on a semantic versioning system together with git version control to keep track of changes and expansions in the ontology. EMMO, the top- and middle-level ontologies that BattINFO is based on, are currently in the beta stage of development. As changes are introduced into EMMO, BattINFO will be adapted as needed. Care will be taken by the developers that terms are properly depreciated or modified to ensure compatibility with previous versions.

With the stable release of BattINFO in place, the focus over the coming months will shift towards developing training materials and demonstrators to help researchers become more comfortable and familiar with the process of annotating data. This will be integrated first into the electronic lab notebook, currently in development in the BIG-MAP project. Afterwards, the approach will be expanded to all Battery2030+ participating projects and beyond.

4. References

[1] S. Clark *et al.*, “Toward a Unified Description of Battery Data,” *Adv. Energy Mater.*, vol. 2102702, 2021.

[2] International Electrotechnical Commission, *IEC 60050 - International Electrotechnical Vocabulary*. International Electrotechnical Commission, 2021.

[3] J. M. Pingarrón *et al.*, “Terminology of electrochemical methods of analysis (IUPAC Recommendations 2019),” *Pure Appl. Chem.*, vol. 92, no. 4, pp. 641–694, 2020.