How to Make Your Battery Data Work for You: A Practical Guide for Lasting Impact

Dr. Simon Clark, SINTEF AS

SAMBA Summer School, DTU, 2025-09-04

Agenda

- Introduction
- Motivation
- Semantic Technology
- 5-Star Battery Data
- Examples!

Good Advice for Presentations

- Stay on time!
- Make the text as big as you think it needs to be… then make it bigger!
- Know your audience!

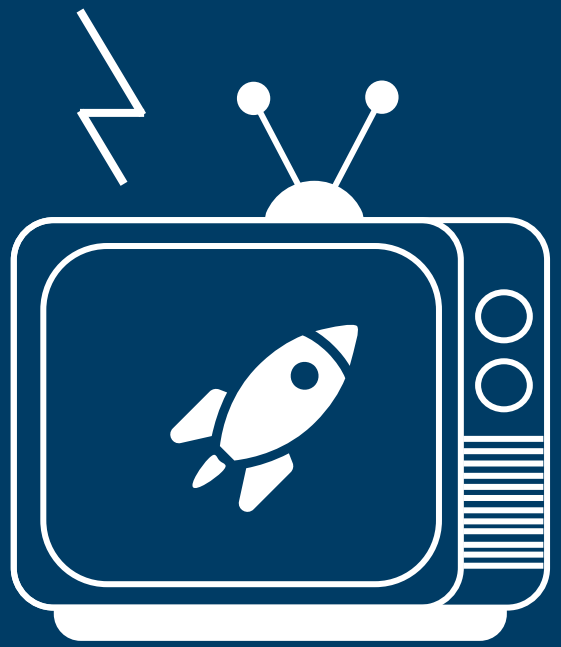
~~Data is available from the corresponding author upon reasonable request~~

Well-described, open datasets are some of the most valuable yet under-utilized resources we have in battery research.

Research articles that include data have higher impact via data re-use and citations.

I. Motivation

*Computer, retrieve **engine data** and plot **new course**!*



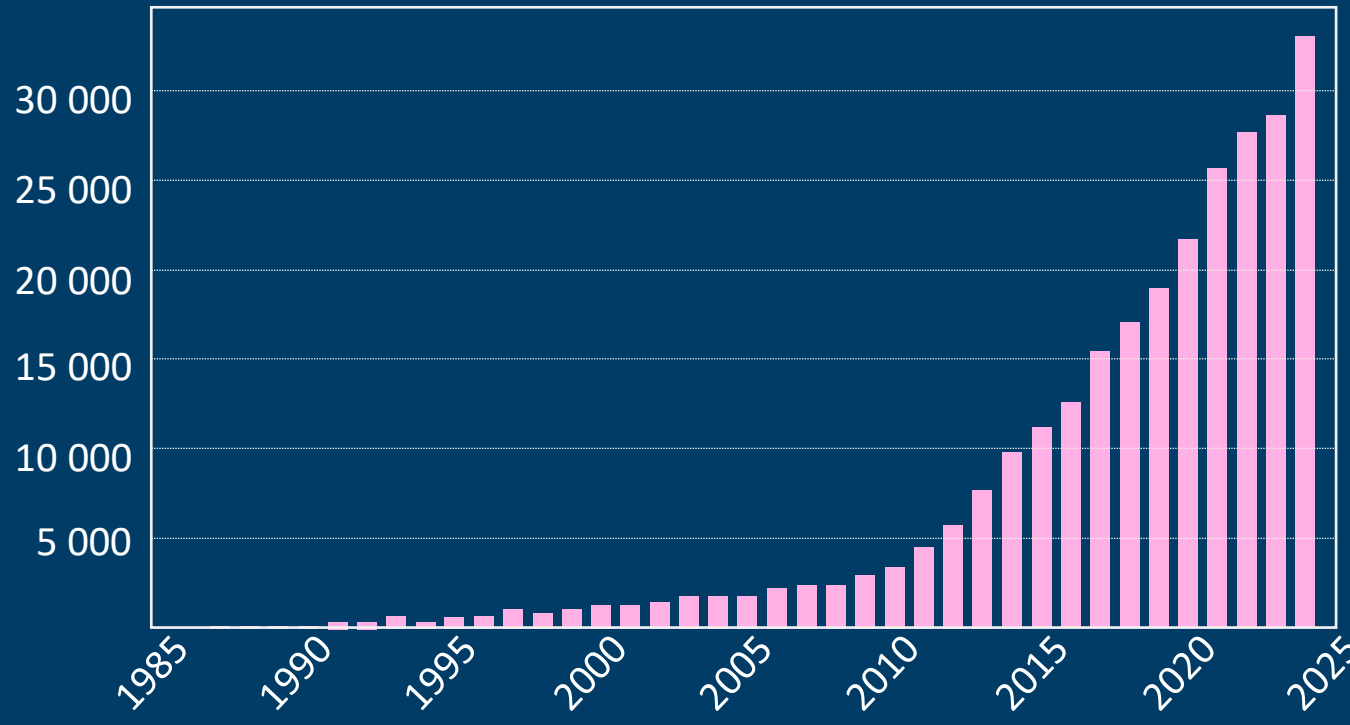
Science Fiction

*Computer, retrieve **the OCP curve for graphite**!*

Science Fact (Almost)

The Knowledge Paradox

There have been more battery articles published in the last 5 years, than the preceding 30-year period

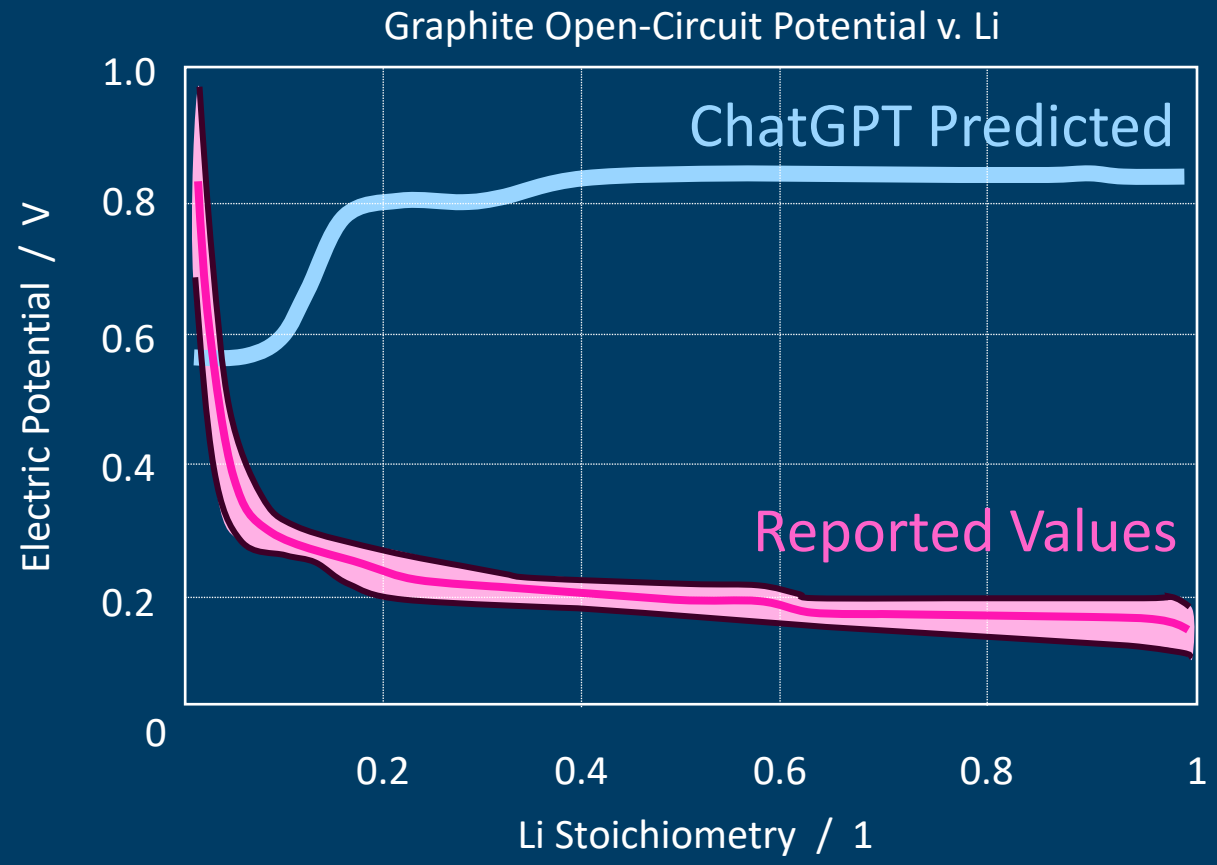


☛ We have access to a wave of new information….but how much structured data is really behind it?

https://doi.org/10.5281/zenodo.14914796

The Knowledge Paradox

Chatbots are excellent responding to queries about **stats data** from sports, finance, and pop culture … but struggle with battery **data** queries. *Why?*



Structured, semantic data.

A cursory literature search* for Graphite OCP data that contains:

- ✓ Raw Data
- ✓ Open Format
- ✓ Permissive License

Yielded **6 results**. We need more.

*credit: Sergi Obrador, IREC

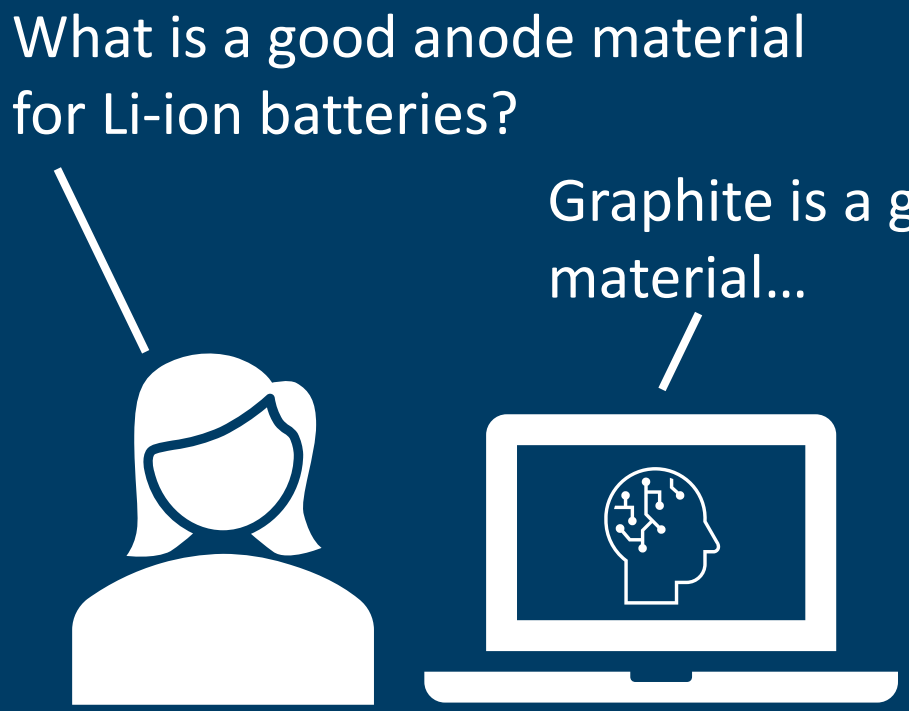
The Knowledge Paradox

The bottleneck is **not data generation**…
it’s **knowledge extraction**.

We need to give data **structure, meaning, and links.**

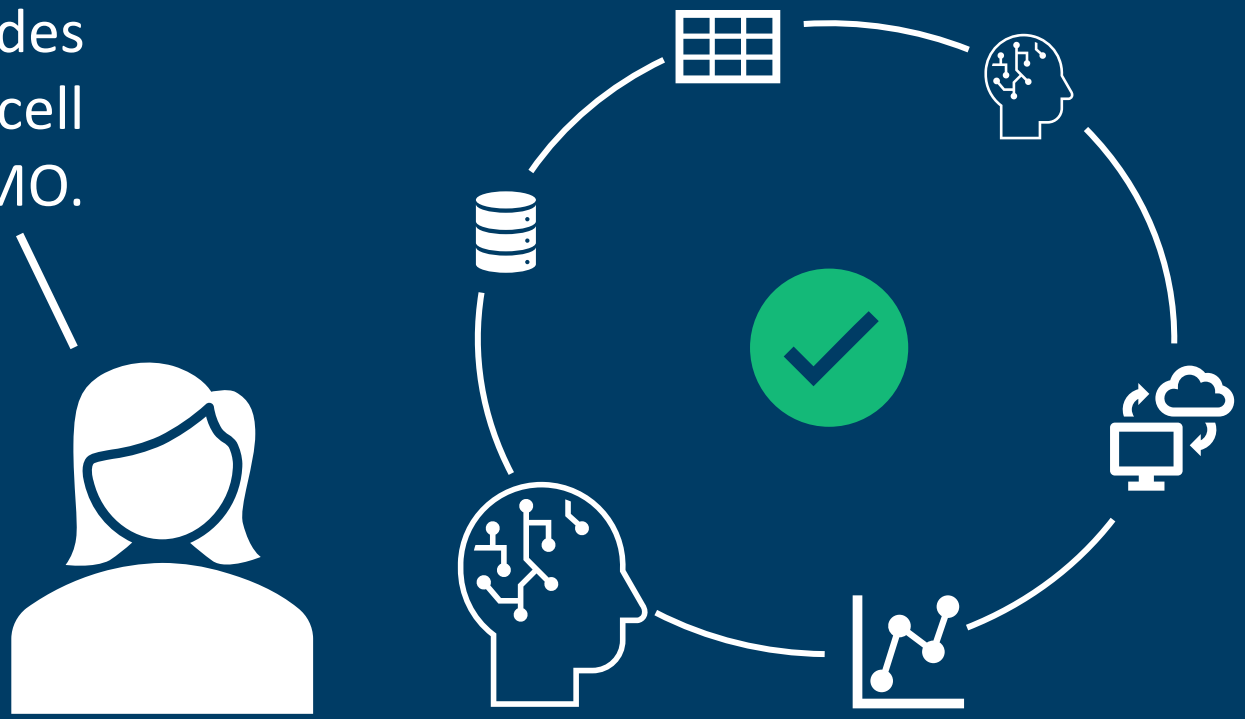
II. Semantic Technology

Chatbots like ChatGPT are reactive and interact with the user through a flat interface



Agentic AI understands goals, makes plans, takes initiative, uses tools, and even interacts with other agents. It’s designed to complete complex tasks .

Evaluate if our Si-Gr anodes are suitable to reach the cell KPIs for Gen 3b with LNMO.



This requires an **awareness and understanding of the different systems** that are available, the **meaning of the data** that they contain, and their **exchange interfaces**

2025

Technology for a better society

Agents in the Semantic Web

Scientific American: Feature Article: The Semantic Web: May 2001

SCIENTIFIC AMERICAN

Main Menu | Interview | Bookmarks | Feedback
Current Issue | Explore! | Ask the Experts | Marketplace | Search the Site

FEATURE ARTICLES

The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

by TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA

SUBTOPICS:
Expressing Meaning
Knowledge Representation
Ontologies
Agents

The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other *local* devices that had a *volume control*. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Agent (n): A software entity that acts autonomously on behalf of a user or another system, using structured, semantically annotated data and web standards to make decisions and coordinate actions.

Agentic AI is powered by the intelligence of AI and the structure of the Semantic Web.

Agents in the Semantic Web


Scientific American: Feature Article: The Semantic Web: May 2001

SCIENTIFIC AMERICAN

Main Menu | Interview | Bookmarks | Feedback
Current Issue | Explore! | Ask the Experts | Marketplace | Search the Site

FEATURE ARTICLES

The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

by TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA

SUBTOPICS:
Expressing Meaning
Knowledge Representation
Ontologies
Agents

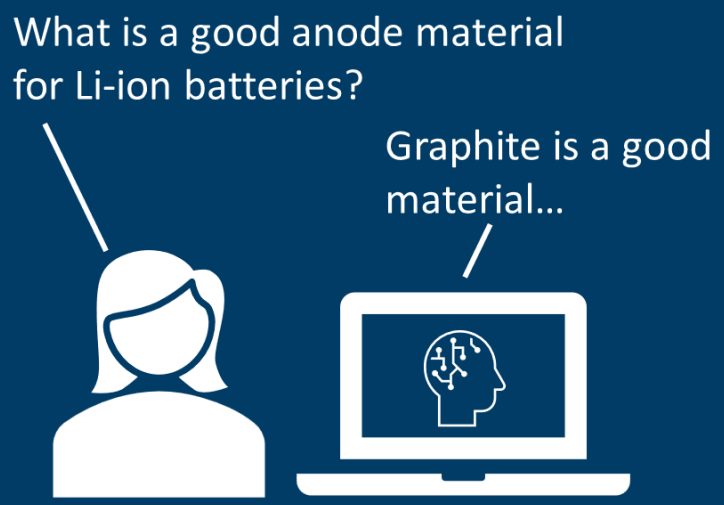
The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other *local* devices that had a *volume control*. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

"Artificial intelligence and Web researchers have co-opted the term for their own jargon, and for them ***an ontology is a document or file that formally defines the relations among terms****. The most typical kind of ontology for the Web* ***has a taxonomy and a set of inference rules.****"*

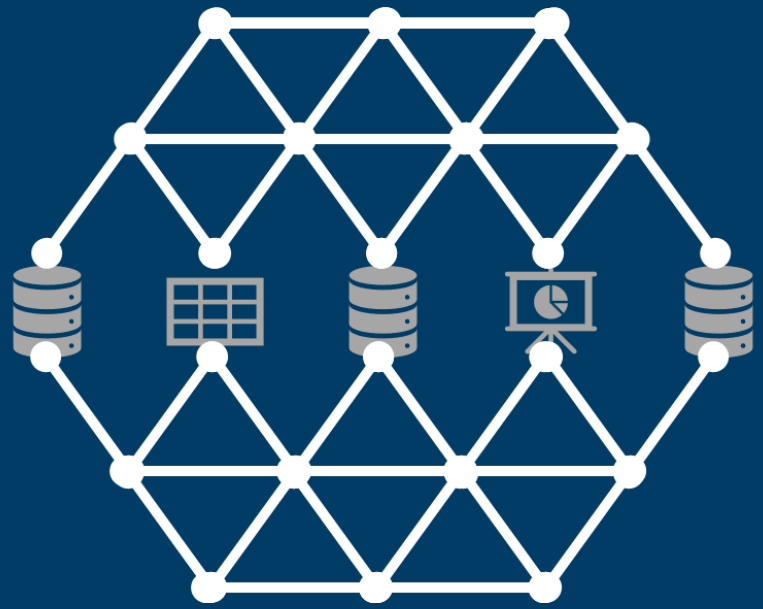
Ontologies have been extremely **well developed for many general knowledge** that enable **search engine optimization**. But they are **slower to reach technical or scientific domains**.

Agents in the “Battery Semantic Web” ?

We have very good tools and infrastructure for humans to interact with machines.

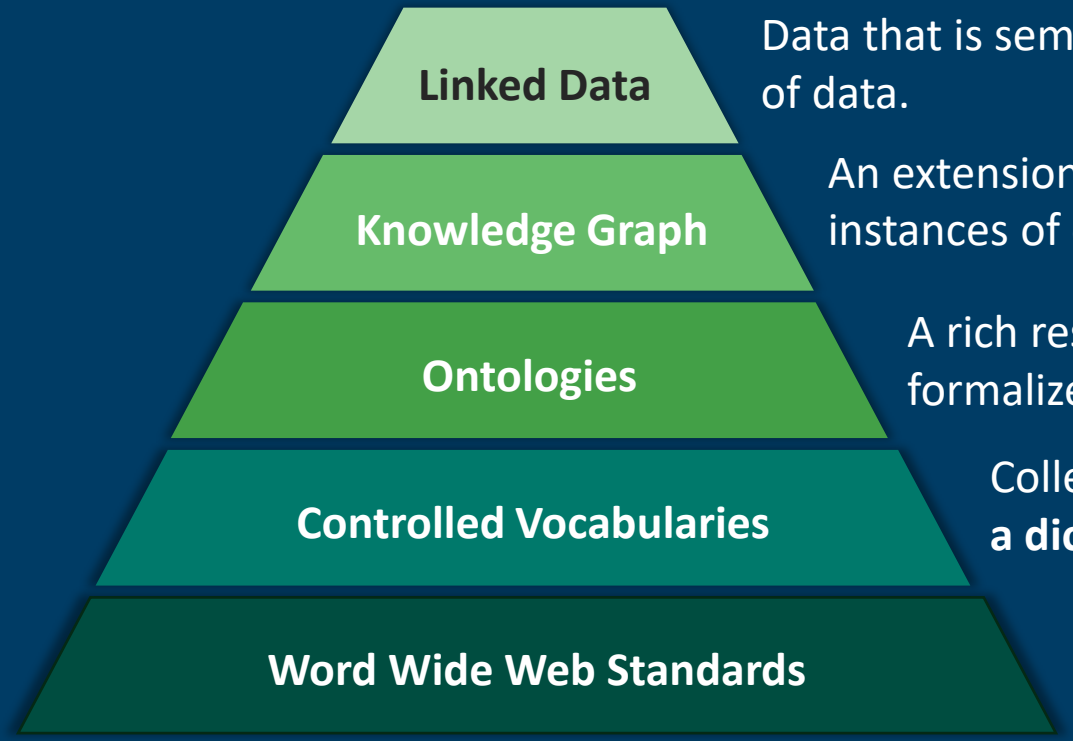


But we are lacking the extensive networks of battery data and tools for agents to work with.



Semantic Technology

Methods and tools that **attach meaning to data.**



Data that is semantically annotated and linked to other pieces of data.

An extension of an ontology that is used to describe actual instances of data and information. **Like a social network.**

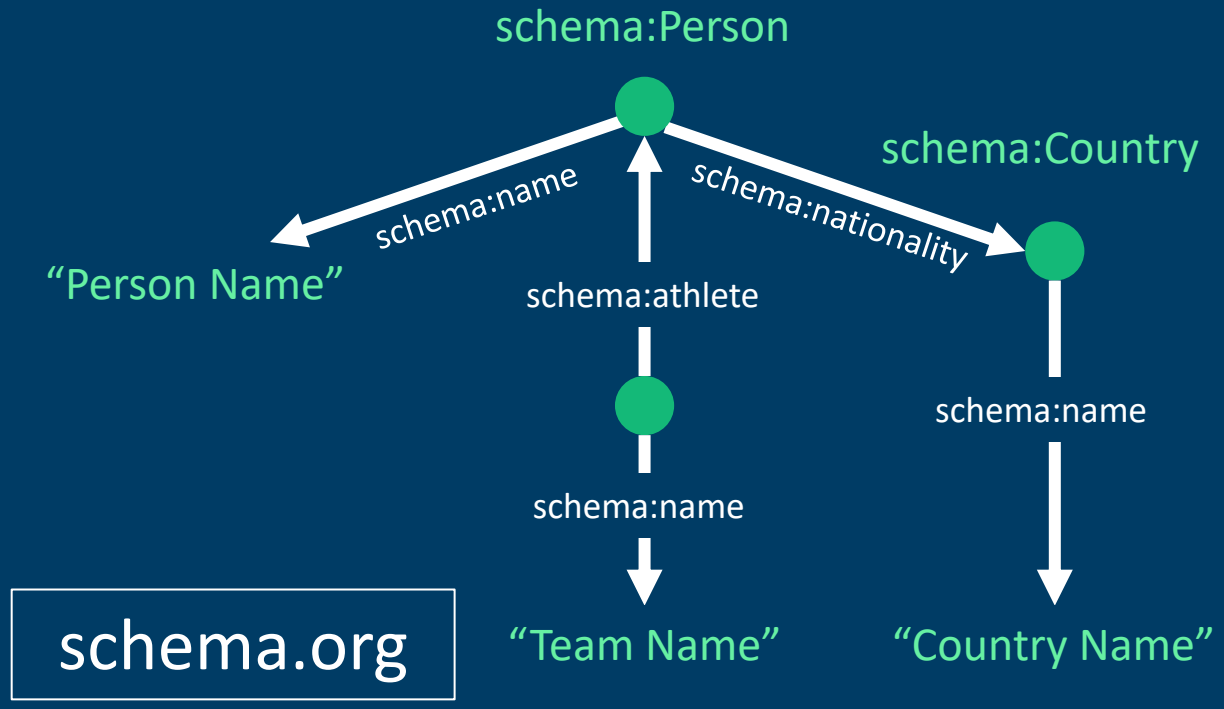
A rich resource that includes relationships between terms to formalize knowledge about a domain. **Like a textbook.**

Collection of terms with a defined and controlled meaning. **Like a dictionary.**

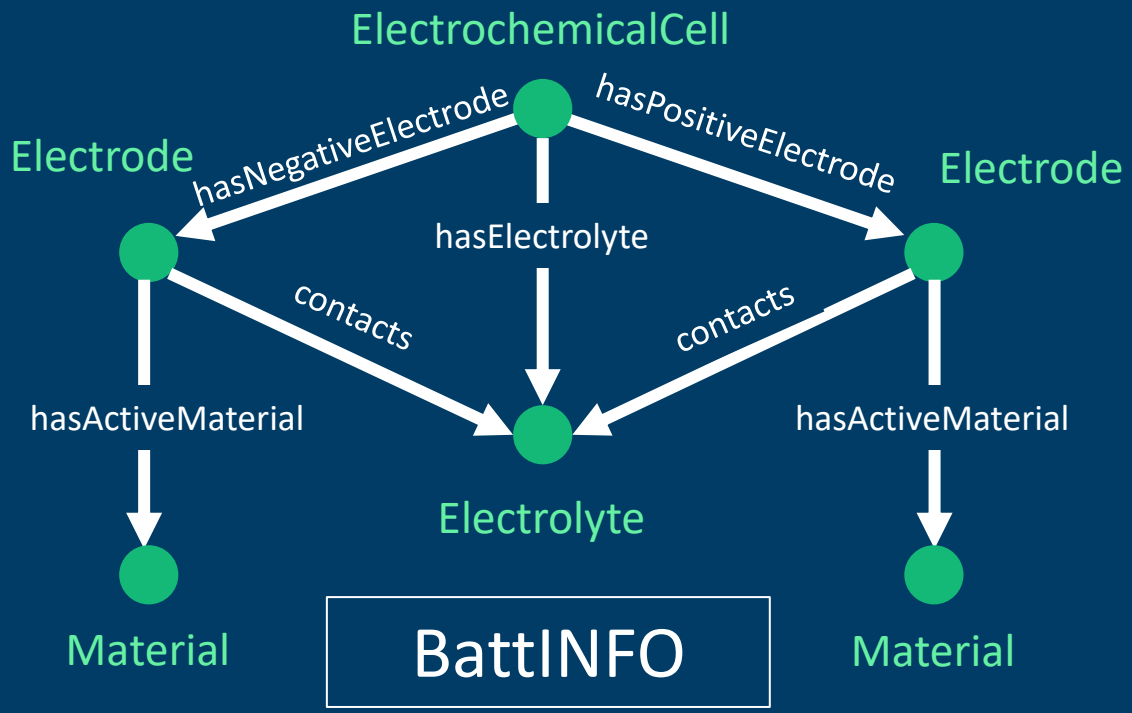
Sets the fundamental rules and grammar. Builds on “triples” like: Simon – isA- Person ; Simon – works at – SINTEF

What is an Ontology?

A formal description of knowledge about a domain that is understandable for both people and machines.



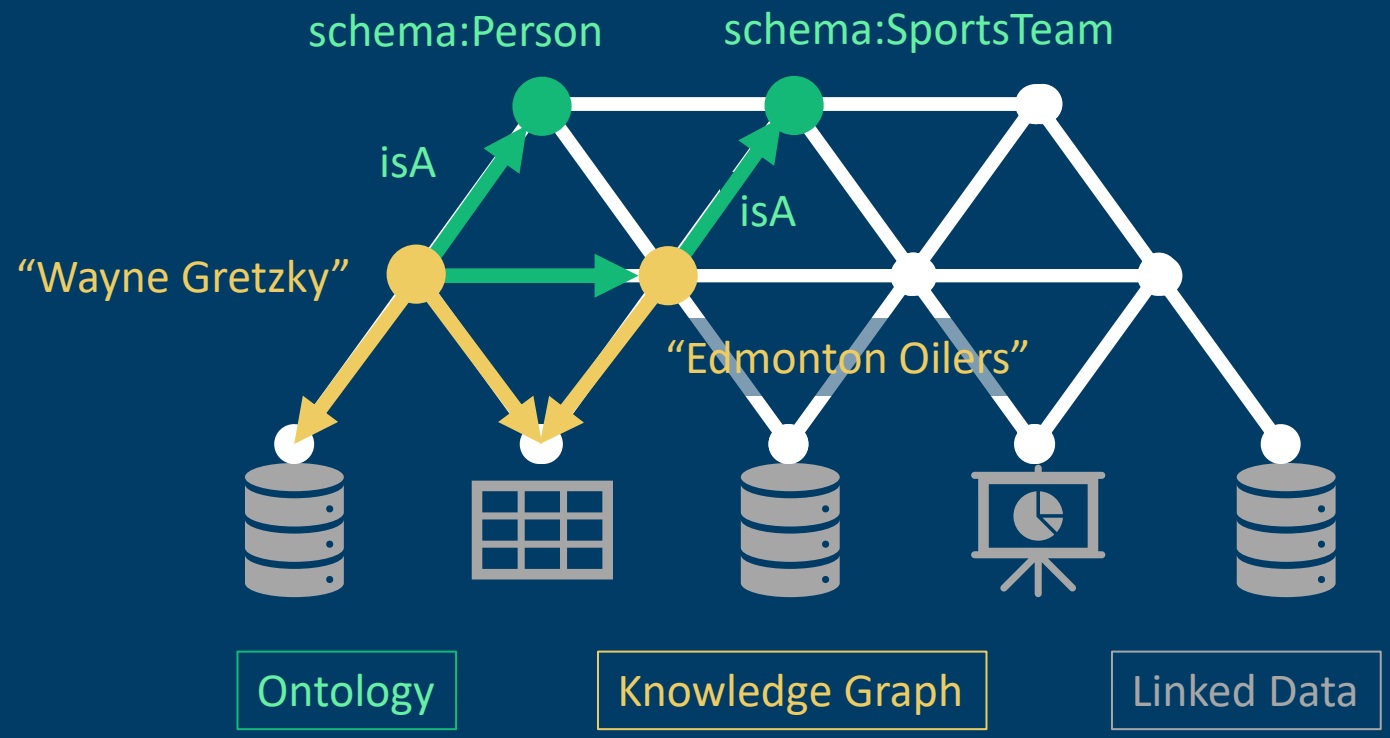
schema.org



BattINFO

The Semantic Web

The Semantic Web is an extension of the World Wide Web, designed to be a web of data, navigated by machines.



Computer, retrieve the results from discharging my CR2032 cell!

UI Layer

Researchers

Endpoint

Ontology

Semantic Layer

Developers

Knowledge Graph

Linked Data

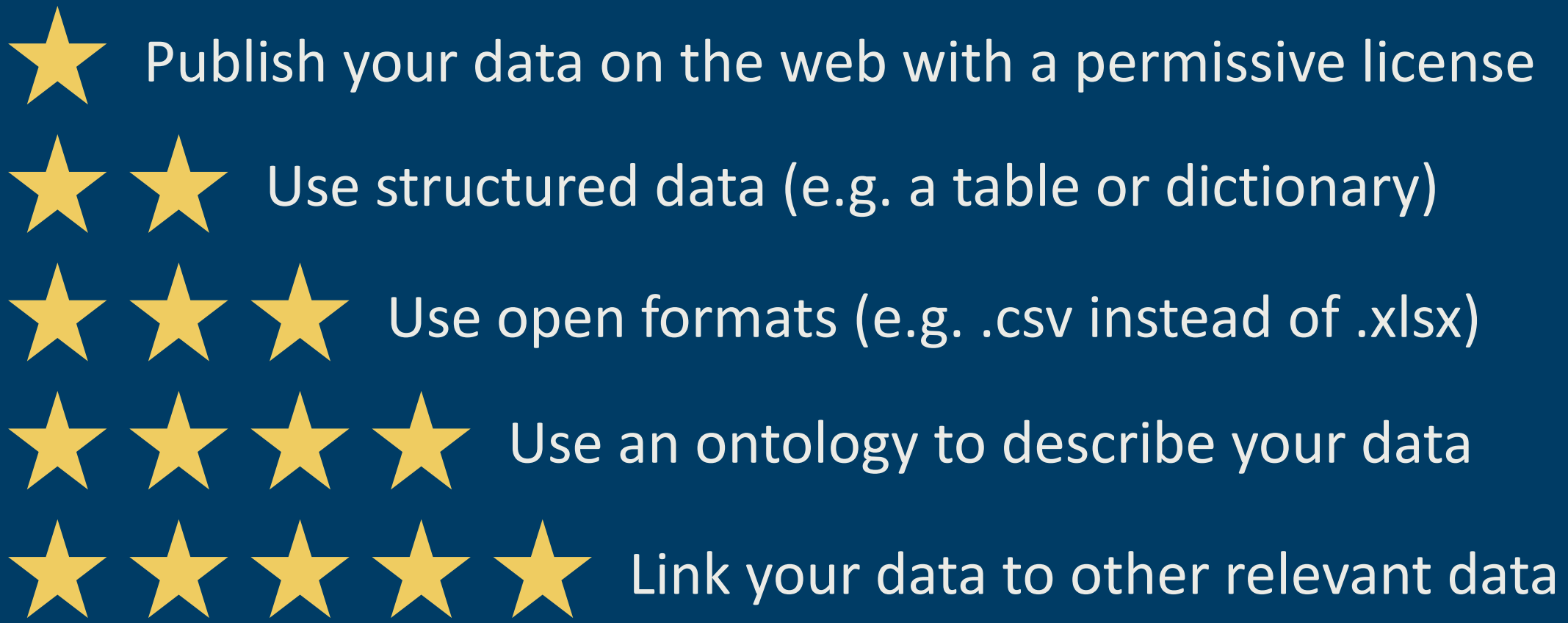
Data Layer

Data Stewards

(do a search example)

III. Five-Star Battery Data

Five-Star Battery Data



★ Publish your data on the web with a permissive license

★★ Use structured data (e.g. a table or dictionary)

★★★ Use open formats (e.g. .csv instead of .xlsx)

★★★★ Use an ontology to describe your data

★★★★★ Link your data to other relevant data

★ Public Repositories and Permissive Licences

Goal: Ensures that battery data is stored in an accessible place and assigned a permissive license.

Public data repositories often offer additional benefits including:

- Managing persistent, unique, and resolvable identifiers (e.g. DOI)
- Supporting version control
- Structuring and embedding bibliographic metadata

Let’s explore an example of a public data repository

(switch to Zenodo)

★★Structured and Machine-Readable Data

Goal: Organize your data so that it has some clear regular pattern that defines the pieces and allows them to be read and processed by machines.

Common structures make it easier to process data.

- **Structured** means that your data is organized in a regular, defined, accessible structure (not just a blob of text)
- **Machine-readable** means that the structure and its content can be read by machines (like code, parsers, or other software tools) in a way that is indexable, unambiguous, and queryable

(switch to BDF)

★★Structured and Machine-Readable Data

The SI and IUPAC style guidelines make some important statements about quantities:

- Quantities are comprised of two parts: **a value** and **a unit**, separated by a space
- Quantities **follow the rules of algebra**

When we instantiate a quantity like voltage:

Voltage = 4.2 V

We can isolate the value by dividing both sides by the unit:

Voltage / V = 4.2

That is why **proper quantity notation in tables and plot axes include a slash** before the unit, and **not parenthesis or square brackets**.

★★★ Open Data Formats

Goal: Provide your data in an open, non-proprietary format that anyone can access.

Battery data is often first serialized in proprietary binary file formats

- Makes it difficult to integrate that data with other systems or processing tools
- Acts as a barrier to sharing and re-use of research data
- Open data formats can make data more accessible and increase its value

SINTEF

★★★ Open Data Formats

A variety of open data formats for different uses. CSV is good when human-readability is important. Parquet is generally best for large, tabular datasets. JSON is suited for metadata. HDF5 supports more complex arrays.

Format	Structure	Human-Readable	Compressed	Typical File Size	Best For
CSV	Tabular	Yes	No	Large	Simple tables, small time series
PARQUET	Tabular	No	Yes	Small	Large time series, structured data
TXT	Unstructured	Yes	No	Large	Logs, raw data
JSON	Hierarchical	Yes	No	Medium	Metadata, structured data
HDF5	Hierarchical	No	Yes	Small	Arrays

★★★ Open Data Formats

If you use .CSV format and you are working in a country that uses the comma (,) as the decimal place indicator, **please be sure to serialize with a dot decimal instead**! It is very **difficult to read files that use a comma for both the decimal and the delimiter!**

✓ YES

Use a dot (.) as the decimal place

Test Time / s, Voltage / V, Current / A
1.00, 4.20, 0.50
2.00, 4.18, 0.50
3.00, 4.16, 0.50

✗ NO

Use a comma (,) as the decimal place

Test Time / s, Voltage / V, Current / A
1,00, 4,20, 0,50
2,00, 4,18, 0,50
3,00, 4,16, 0,50

★★★★ Ontology-Annotated Metadata

Goal: Create machine-readable metadata using controlled, semantic vocabularies and ontologies.

Metadata provides essential information to interpret data and extract its full value.

- Type of battery test being carried out
- Type of battery or material being tested
- Quantities and units used in the data

Ontology-annotated metadata ensures that this information and its meaning is easily available to both humans and machines

DON’T PANIC

It’s really the responsibility of software developers and data stewards to create tools that do this for you

★★★★ Ontology-Annotated Metadata

Unstructured text is the most common way of expressing metadata today.

```
This dataset contains test data obtained from cycling a
lithium-polymer battery at 1C under ambient laboratory
conditions. The result is tabular data, serialized in csv files
with the following columns.
U(V): Cell Voltage
I(A): Cell Current
…
```

Difficult for machines to parse and understand. Structured metadata with controlled vocabularies is better.

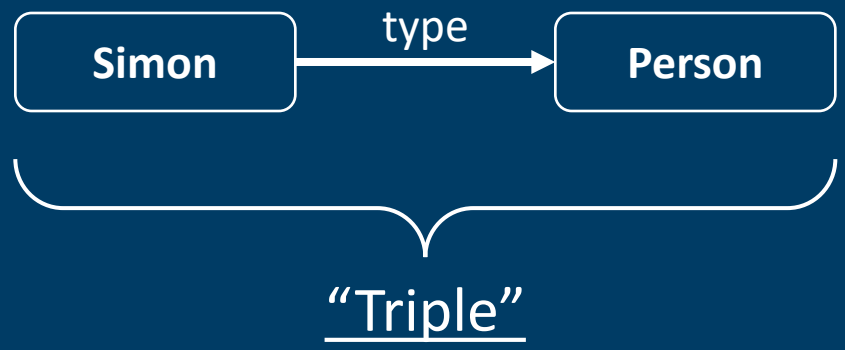
★★★★ Ontology-Annotated Metadata

There are three important aspects to consider when creating ontology-annotated metadata:

1. **Structure**. Machines interpret metadata as a web-like graph of nodes that are linked by edges.
2. **Terms**. Controlled vocabularies provide a dictionary of terms that can be used to create metadata.
3. **Format**. Metadata is stored in file formats that can be read by both people and machines, like JSON-LD.

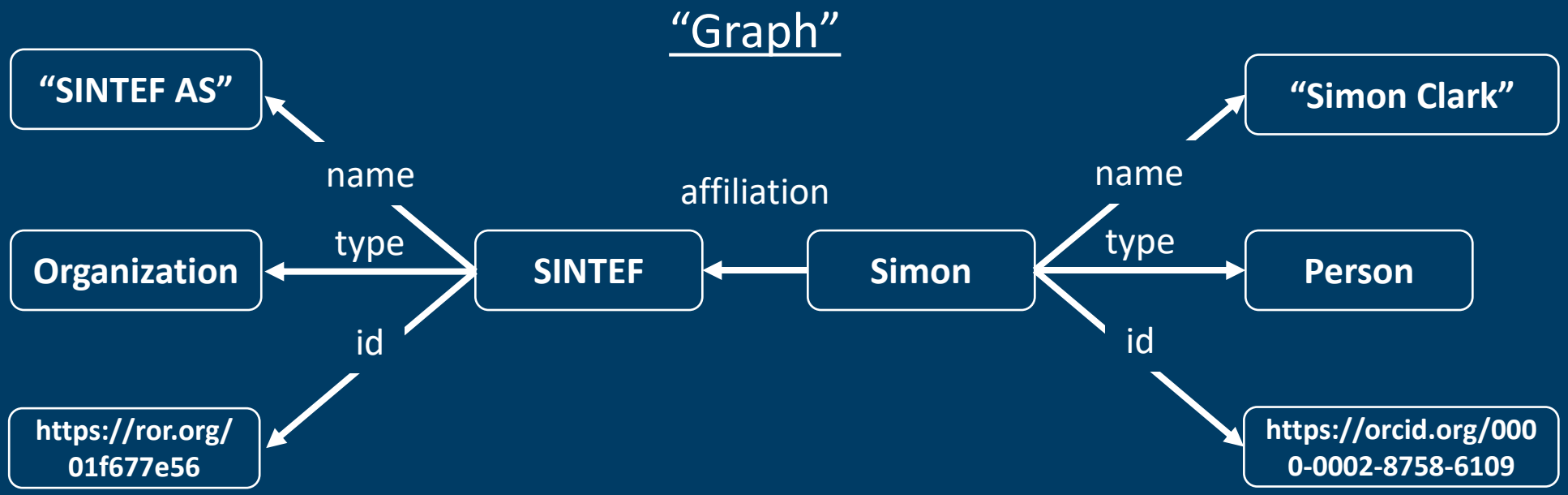
★★★★ Ontology-Annotated Metadata

1. **Structure**. Machines interpret metadata as a web-like graph of nodes that are linked by edges.



★★★★ Ontology-Annotated Metadata

1. **Structure**. Machines interpret metadata as a web-like graph of nodes that are linked by edges.



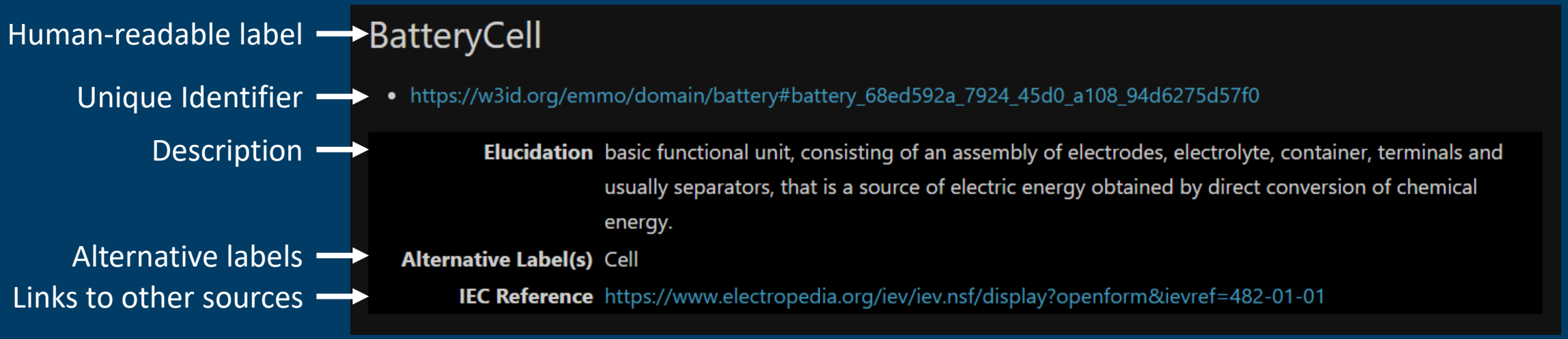
★★★★ Ontology-Annotated Metadata

2. **Terms**. Controlled vocabularies provide a dictionary of terms that can be used to create metadata.

RDF / RDFS	Foundational terms, that set the rules
DCAT	Terms for describing data catalogues
CSVW	Terms for describing tabular data on the web
Schema.org	General purpose terms for search engine optimization
BattINFO	Terms for describing batteries, materials, tests, etc.

★★★★ Ontology-Annotated Metadata

2. **Terms**. Controlled vocabularies provide a dictionary of terms that can be used to create metadata.



★★★★ Ontology-Annotated Metadata

3. **Format**. Metadata is stored in file formats that can be read by both people and machines. JSON-LD is an easy serialization format for expressing RDF graphs in a human-readable way: metadata.jsonld

```
{
    "@context": "https://w3id.org/emmo/domain/battery/context",
```

← Dictionary of terms

```
    "PerTemperatureUnit": "https://w3id.org/emmo#EMMO_02a935c8_c2d4_4a00_bd6f_b89d05aac79e",
    "PerTeslaMetre": "https://w3id.org/emmo#PerTeslaMetre",
    "PerTeslaSecond": "https://w3id.org/emmo#PerTeslaSecond",
    "PerThermalTransmittanceUnit": "https://w3id.org/emmo#EMMO_869e6e4f_a9b5_4db8_a978_8ad050239933",
    "PerTimeMassUnit": "https://w3id.org/emmo#EMMO_52b029aa_e525_4907_95d8_759298b04f97",
    "PerVolumeUnit": "https://w3id.org/emmo#EMMO_95a9bf22_eabc_4a84_863d_9ea398c8a52e",
    "PerWeber": "https://w3id.org/emmo#PerWeber",
    "PercentQuantity": "https://w3id.org/emmo/domain/electrochemistry#electrochemistry_cf386c1e_e62b_43ef_9b0e_e5e58935d63f",
    "PercentUnit": "https://w3id.org/emmo/domain/electrochemistry#electrochemistry_8430be19_6cae_4a15_a570_3c15d0d190eb",
    "Perceptual": "https://w3id.org/emmo#EMMO_649bf97b_4397_4005_90d9_219755d92e34",
    "Perchlorate": "https://w3id.org/emmo/domain/chemical-substance#substance_3286d184_258d_4f80_965b_b9277d57f1b0",
    "PerchloricAcid": "https://w3id.org/emmo/domain/chemical-substance#substance_ecd84623_43e2_457e_8988_1bfc9d1fb818",
    "PerforatedFoil": "https://w3id.org/emmo/domain/electrochemistry#electrochemistry_c84a1d8f_e405_42ba_8c80_bbe3b9c6c66a",
    "PeriodDuration": "https://w3id.org/emmo#EMMO_90e59882_4592_4036_a75d_5fbefb22dc80",
    "PermanentLiquidPhaseSintering": "https://w3id.org/emmo#EMMO_1acb552d_281a_40a4_9d55_5e31b85d4dc1",
    "Permeability": "https://w3id.org/emmo#EMMO_09663630_1b84_4202_91e6_e641104f579e",
    "PermeabilityUnit": "https://w3id.org/emmo#EMMO_fa9c8c56_314f_4a5a_a71d_bae66446b185",
    "Permeance": "https://w3id.org/emmo#EMMO_e0feea8c_318e_4dcf_92f0_751e228ed99d",
    "Permittivity": "https://w3id.org/emmo#EMMO_0ee5779e_d798_4ee5_9bfe_c392d5bea112",
    "PermittivityUnit": "https://w3id.org/emmo#EMMO_5f89cb0c_3171_47ee_b2ab_027a07c34c4b",
    "Persistence": "https://w3id.org/emmo#EMMO_e04884d9_eda6_487e_93d5_7722d7eda96b",
    "Person": "https://schema.org/Person",
    "Perspective": "https://w3id.org/emmo#EMMO_49267eba_5548_4163_8f36_518d65b583f9",
    "Peta": "https://w3id.org/emmo#EMMO_d7c74480_a568_4470_acff_f18b499cc850",
    "PetaCoulomb": "https://w3id.org/emmo#PetaCoulomb",
    "PetaJoule": "https://w3id.org/emmo#PetaJoule",
    "PetaPrefixedUnit": "https://w3id.org/emmo#EMMO_43a6b269_da31_4bb6_a537_c97df4fff32a",
    "PhaseAngle": "https://w3id.org/emmo#EMMO_2a0e5777_348c_475b_adf0_1b1e71a29bc9",
    "PhaseCoefficient": "https://w3id.org/emmo#EMMO_ff1212da_108e_4aaf_a2b0_b691fe53685e",
    "PhaseDifference": "https://w3id.org/emmo#EMMO_7f8ef5cf_7d80_46fa_951c_369014e3a8af",
    "PhaseHeterogeneousMixture": "https://w3id.org/emmo#EMMO_0e030040_98a7_49b2_a871_dced1f3a6131",
    "PhaseHomogeneousMixture": "https://w3id.org/emmo#EMMO_0e6378df_1ce8_4321_b00c_ee9beea60a67",
    "PhaseOfMatter": "https://w3id.org/emmo#EMMO_668fbd5b_6f1b_405c_9c6b_d6067bd0595a",
    "PhaseSpeedOfElectromagneticWaves": "https://w3id.org/emmo#EMMO_9b9e0029_8b16_4382_bd47_571a7ae7d6f6",
    "PhaseVelocity": "https://w3id.org/emmo#EMMO_2b69b940_a4ab_40dc_b565_8daab23e4df2",
    "PhosphoricAcid": "https://w3id.org/emmo/domain/chemical-substance#substance_350eaefc_75e1_49d8_b161_01a634607857",
    "PhosphoricAcidSolution": "https://w3id.org/emmo/domain/electrochemistry#electrochemistry_78f75a64_55b6_4243_a35e_3d279c83209b",
    "PhosphorusAtom": "https://w3id.org/emmo#EMMO_7905373a-cf03-57fe-a627-bfdb1db9369c",
    "PhosphorusSymbol": "https://w3id.org/emmo#EMMO_6bd7b6f4-db3b-58da-b9a0-58f44bb3ed37",
    "PhotochemicalProcesses": "https://w3id.org/emmo#EMMO_00f2dc2d_2f64_468a_a77c_d70841b0b5f0",
    "PhotoelectrolyticCell": "https://w3id.org/emmo/domain/electrochemistry#electrochemistry_7760b241_775f_4be1_b827_59f9bde9e5b2",
    "PhotoluminescenceMicroscopy": "https://w3id.org/emmo/domain/characterisation-methodology/chameo#PhotoluminescenceMicroscopy",
    "PhysicalBasedSimulationSoftware": "https://w3id.org/emmo#EMMO_8d4962d7_9608_44f7_a2f1_82a4bb173f4a",
    "PhysicalConstant": "https://w3id.org/emmo#EMMO_b953f2b1_c8d1_4dd9_b630_d3ef6580c2bb",
```

★★★★ Ontology-Annotated Metadata

3. **Format**. Metadata is stored in file formats that can be read by both people and machines. JSON-LD is an easy serialization format for expressing RDF graphs in a human-readable way: metadata.jsonld

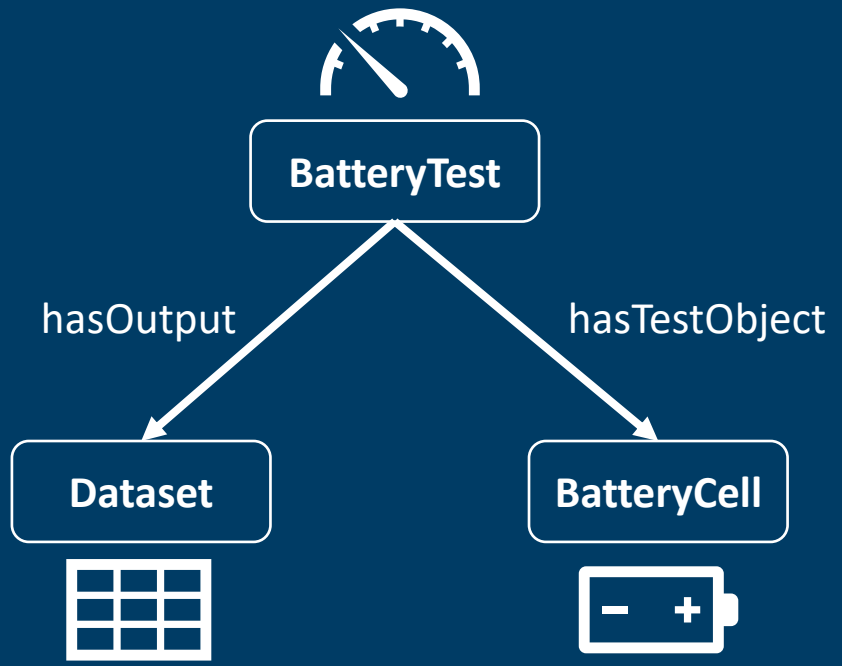
```
{
    "@context": "https://w3id.org/emmo/domain/battery/context",
    "@type": "schema:Person",
    "@id": "https://orcid.org/0000-0002-8758-6109",
    "schema:name": "Simon Clark",
    "schema:affiliation": {
        "@type": “schema:Organization",
        "@id": "https://ror.org/01f677e56",
        "schema:name": "SINTEF AS"
    }
}
```

Dictionary of terms
Type of thing
ID of specific thing
Additional properties

Learn more at: https://json-ld.org/

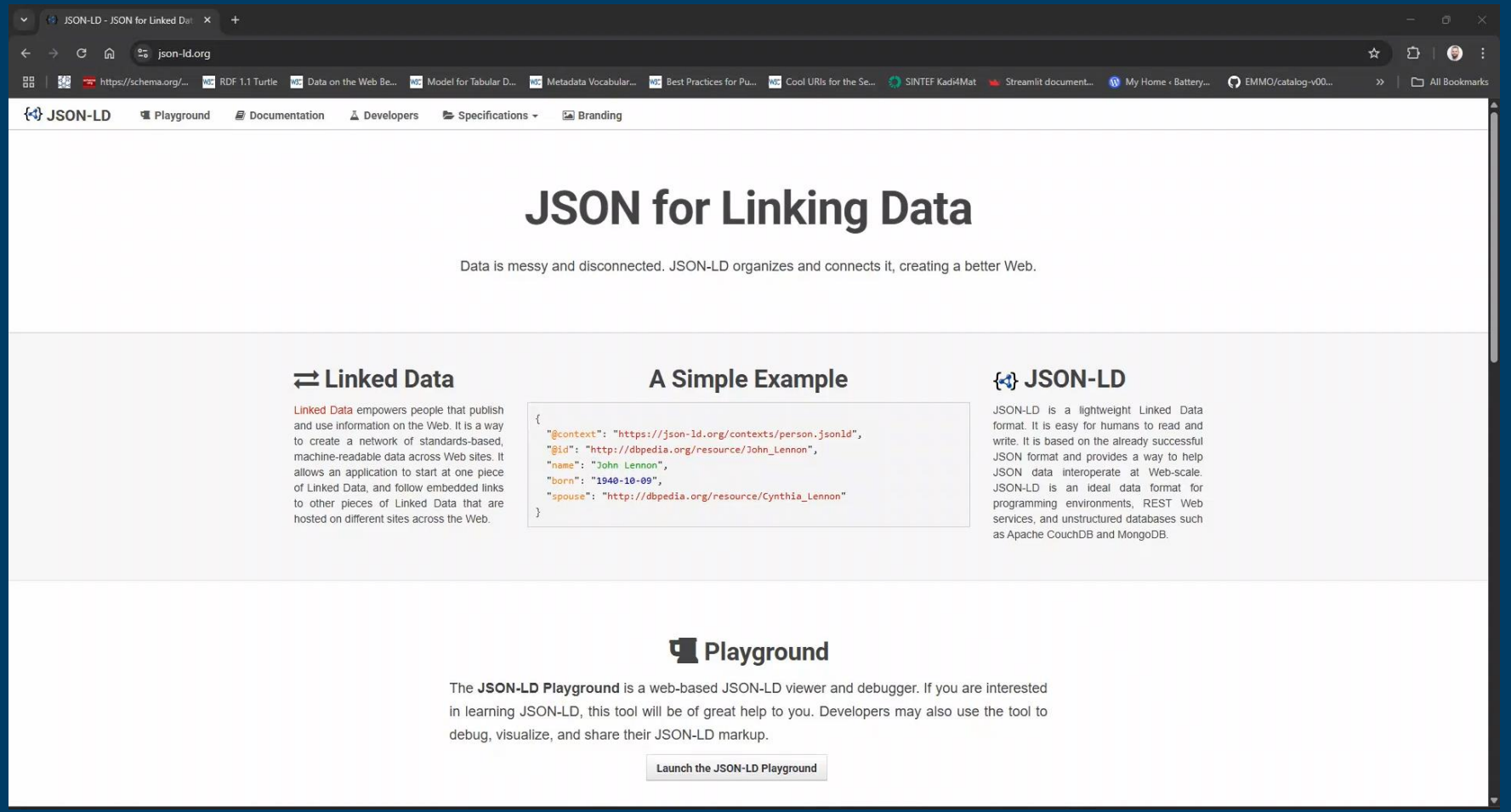
★★★★ Ontology-Annotated Metadata


```
{
    "@context": "https://w3id.org/emmo/domain/battery/context",
    "@type": "BatteryTest",
    "@id": "urn:uri:bbdc6a42-550b-4038-a0a0-9236048b125f",
    "hasTestObject": {
        "@type": "BatteryCell",
        "@id": "urn:uri:89afd7f5-b62f-4749-bba8-70fe4cffaad8"
    },
    "hasOutput": {
        "@type": ["dcat:Dataset", "BatteryTestResult"],
        "@id": "urn:uri:0d8197a5-6801-47e8-a45c-95f3d7b248dc",
        "dcat:distribution": {
            "@type": "dcat:Distribution",
            "@id": "urn:uri:2d610bd8-7a65-4ba8-8433-4c2ac23f96df",
            "dcat:mediaType": "text/csv",
            "dcat:downloadURL":
"https://zenodo.org/records/15069341/files/sintef__cr2032__discharging_11_mA__20240216.bdf.csv",
            "csvw:tableSchema": "https://w3id.org/battery-data-alliance/ontology/battery-data-
format/schema"
        }
    }
}
```

SINTEF

★★★★ Ontology-Annotated Metadata



★★★★★ Linked Data

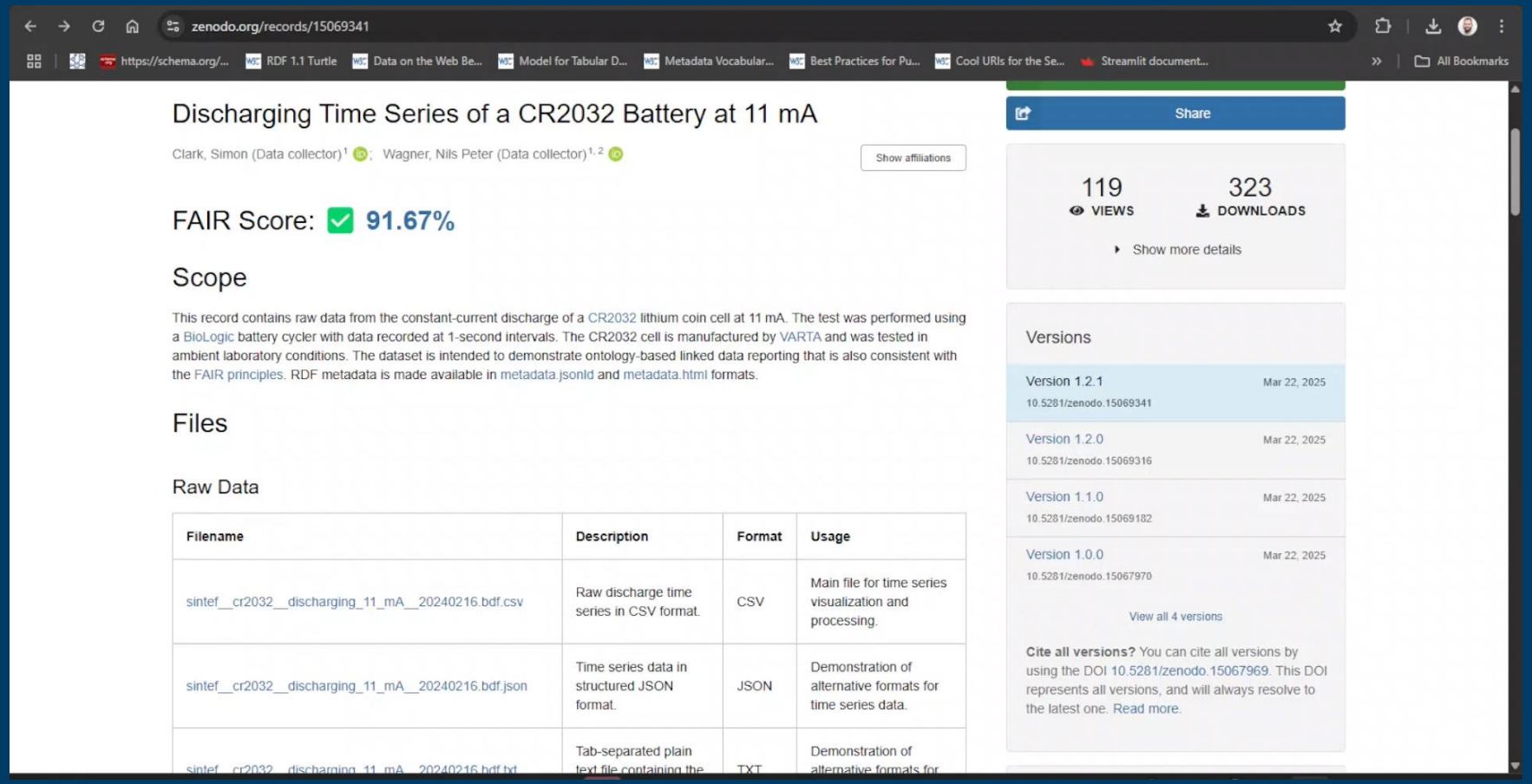
Three tips for implementation

1. **Use BattINFO terms and structures in your metadata.**
   - Many contain built-in mappings to other common data sources like Wikidata, Pubchem, etc.
2. **Submit your dataset to the Battery Knowledge Base community on Zenodo**
   - We can scrape the metadata from datasets in the community to index them in the Battery Knowledge Base and link them with other datasets.
3. **Make explicit links in your metadata to other datasets that are related**
   - `schema:isBasedOn`, `schema:hasPart`, `schema:isPartOf`
   - Include human-readable links in the description

SINTEF

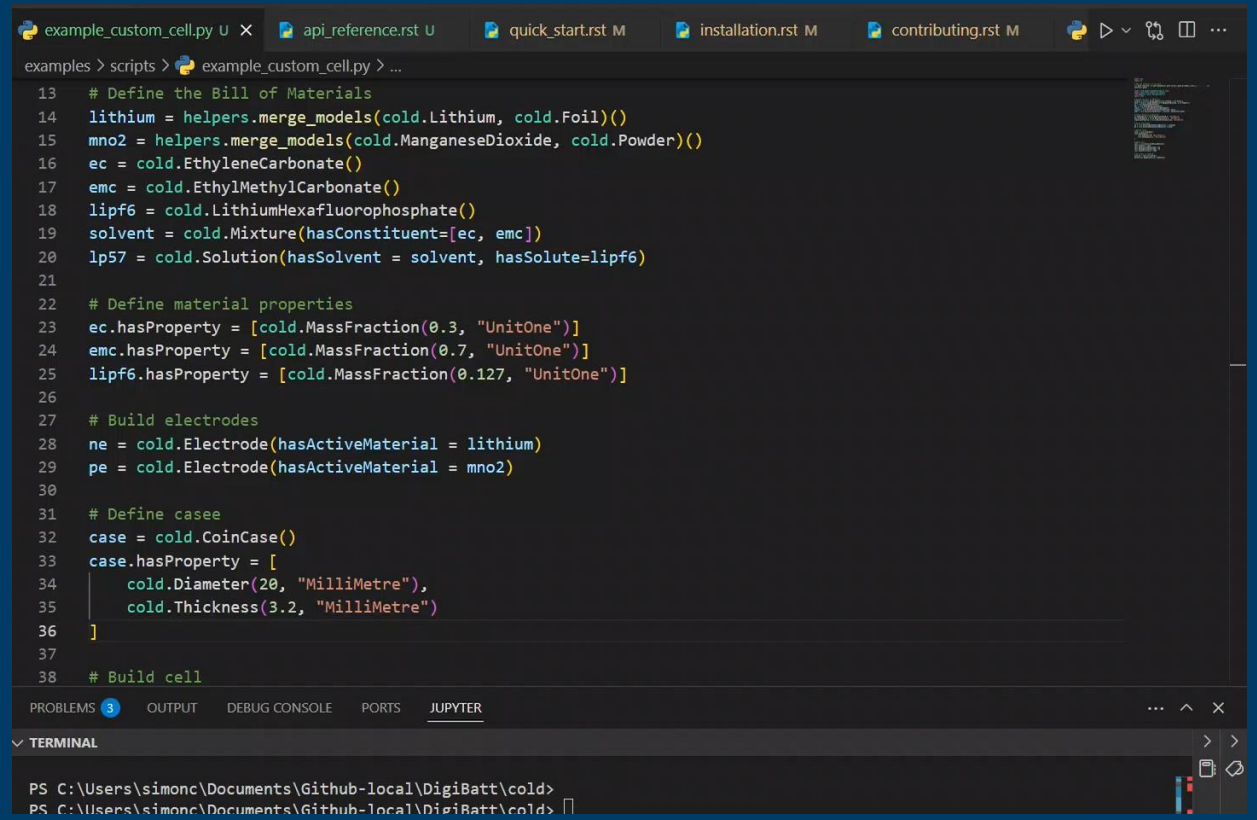
★★★★★Linked Data

It’s also for humans!



Python Packages

Validation is like salt…not too little, not too much.

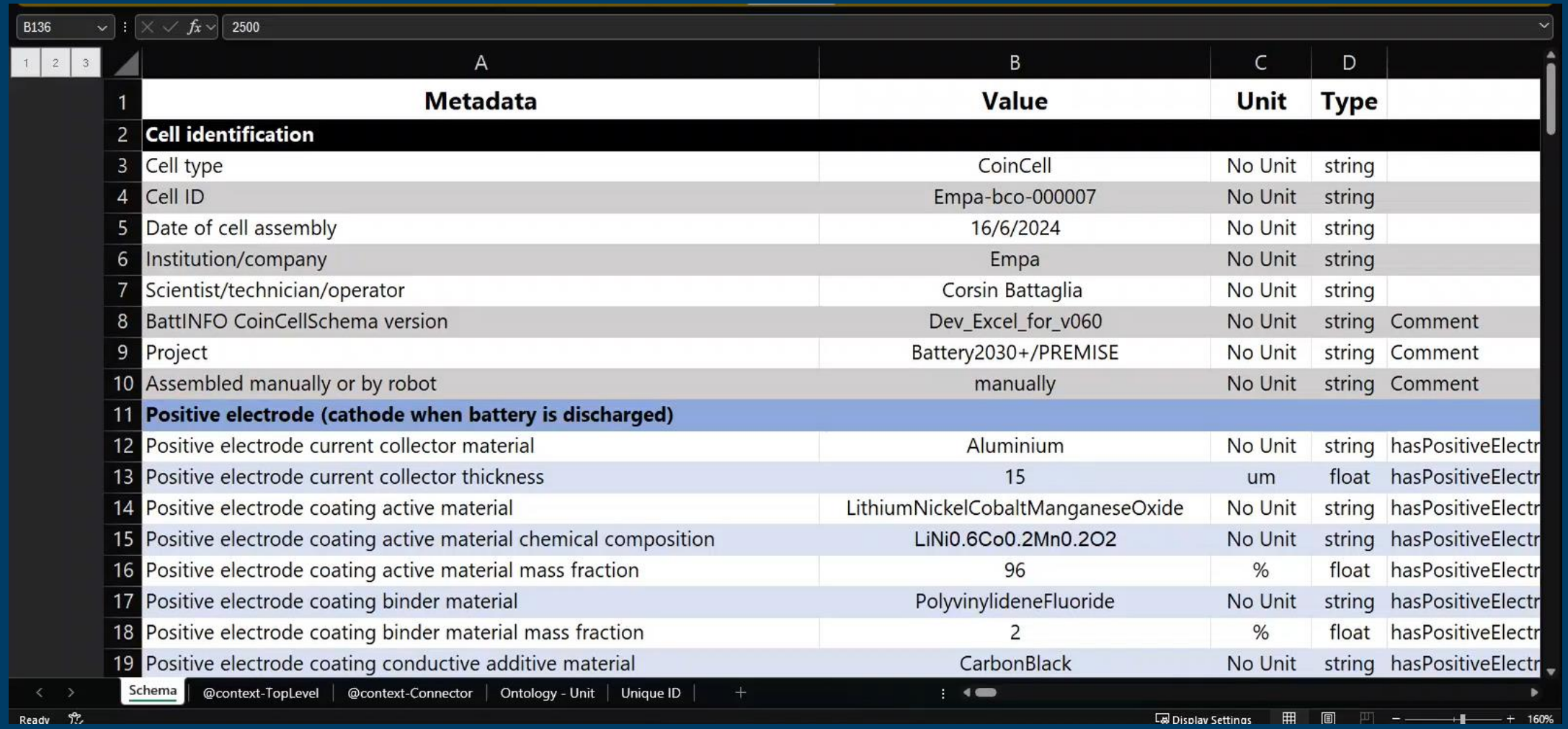


- ✓ **Flexible definitions** to support describing complex objects
- ✓ Built-in validation for **units**
  - **Hard validation** that enforces correct unit types
- ✓ Built-in validation for **ontology semantics**
  - **Soft validation** that flags questionable usage
- ✓ **Maintains a single-source of knowledge** in the ontology

COLD

digibatt

Excel-Based Interfaces: BattINFO Converter



B136 | 2500

	A	B	C	D	
1	**Metadata**	**Value**	**Unit**	**Type**	
2	**Cell identification**				
3	Cell type	CoinCell	No Unit	string	
4	Cell ID	Empa-bco-000007	No Unit	string	
5	Date of cell assembly	16/6/2024	No Unit	string	
6	Institution/company	Empa	No Unit	string	
7	Scientist/technician/operator	Corsin Battaglia	No Unit	string	
8	BattINFO CoinCellSchema version	Dev_Excel_for_v060	No Unit	string	Comment
9	Project	Battery2030+/PREMISE	No Unit	string	Comment
10	Assembled manually or by robot	manually	No Unit	string	Comment
11	**Positive electrode (cathode when battery is discharged)**				
12	Positive electrode current collector material	Aluminium	No Unit	string	hasPositiveElectr
13	Positive electrode current collector thickness	15	um	float	hasPositiveElectr
14	Positive electrode coating active material	LithiumNickelCobaltManganeseOxide	No Unit	string	hasPositiveElectr
15	Positive electrode coating active material chemical composition	LiNi0.6Co0.2Mn0.2O2	No Unit	string	hasPositiveElectr
16	Positive electrode coating active material mass fraction	96	%	float	hasPositiveElectr
17	Positive electrode coating binder material	PolyvinylideneFluoride	No Unit	string	hasPositiveElectr
18	Positive electrode coating binder material mass fraction	2	%	float	hasPositiveElectr
19	Positive electrode coating conductive additive material	CarbonBlack	No Unit	string	hasPositiveElectr

Schema | @context-TopLevel | @context-Connector | Ontology - Unit | Unique ID

Ready | Display Settings | 160%

Technology for a Better Soceity

N. Plainpan, S. Clark, C. Battaglia, BattINFO converter: An automated tool for semantic annotation of battery cell metadata, submitted

Community Building

Battery Semantic Technology is a community resource.

- Continue to develop the Battery Knowledge Base (BKB) as a human-readable source of knowledge
- Monthly community “office hours” to answer questions and discuss improvements
- Widen the formal governance structure to cover a wider group of stakeholders

Teamwork makes the dream work!

Battery Semantic Technology

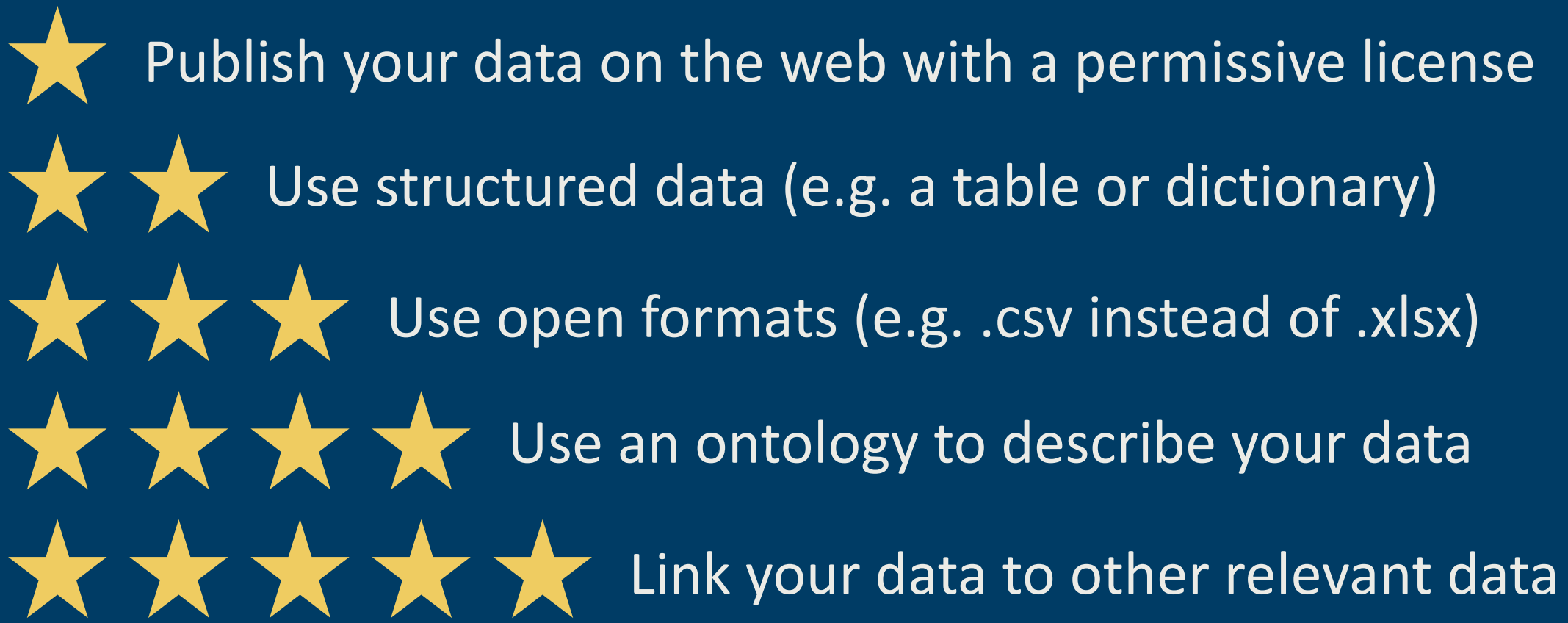
Never before have we had access to so much battery data, and the tools that we need to process it.

Semantic technology allows us to combine the power of machine processing with human intuition, for a new generation of battery innovation!

Summary

- Well-described, open data is a very valuable resource that can extend the impact of your work
- Use the Five-Star Battery Data guidelines when publishing your data
- Join our community discussions to help shape the future of battery data

Five-Star Battery Data



★ Publish your data on the web with a permissive license

★★ Use structured data (e.g. a table or dictionary)

★★★ Use open formats (e.g. .csv instead of .xlsx)

★★★★ Use an ontology to describe your data

★★★★★ Link your data to other relevant data

Thank You

SINTEF:

- Eibar Flores
- Sridevi Krishnamurthi
- Jesper Friis
- Francesca L. Blekken
- Casper W. Andersen

DTU:

- Tejs Vegge
- Ivano E. Castelli

EMPA:

- Corsin Battaglia
- Nukorn Plainpan

KIT:

- Christian Punckt

Fraunhofer ISC:

- Simon Stier
- Lukas Gold

BATTERY 2030+

BIG-MAP

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

digibatt

Teknologi for et bedre samfunn