



SINTEF

# A Linked Data Primer

Eibar Flores  
SINTEF





# Acknowledgements



Simon Clark



Hannah Hansen



Julien Lagarde



Casper Wenzel



Jesper Friis



This project has received funding from the European Union's Horizon 2020 research and innovation initiative under grants agreement No 957189 (BIG-MAP) and No 957213 (BATTERY 2030+).



# Outline

**What are we trying to solve with linked data**

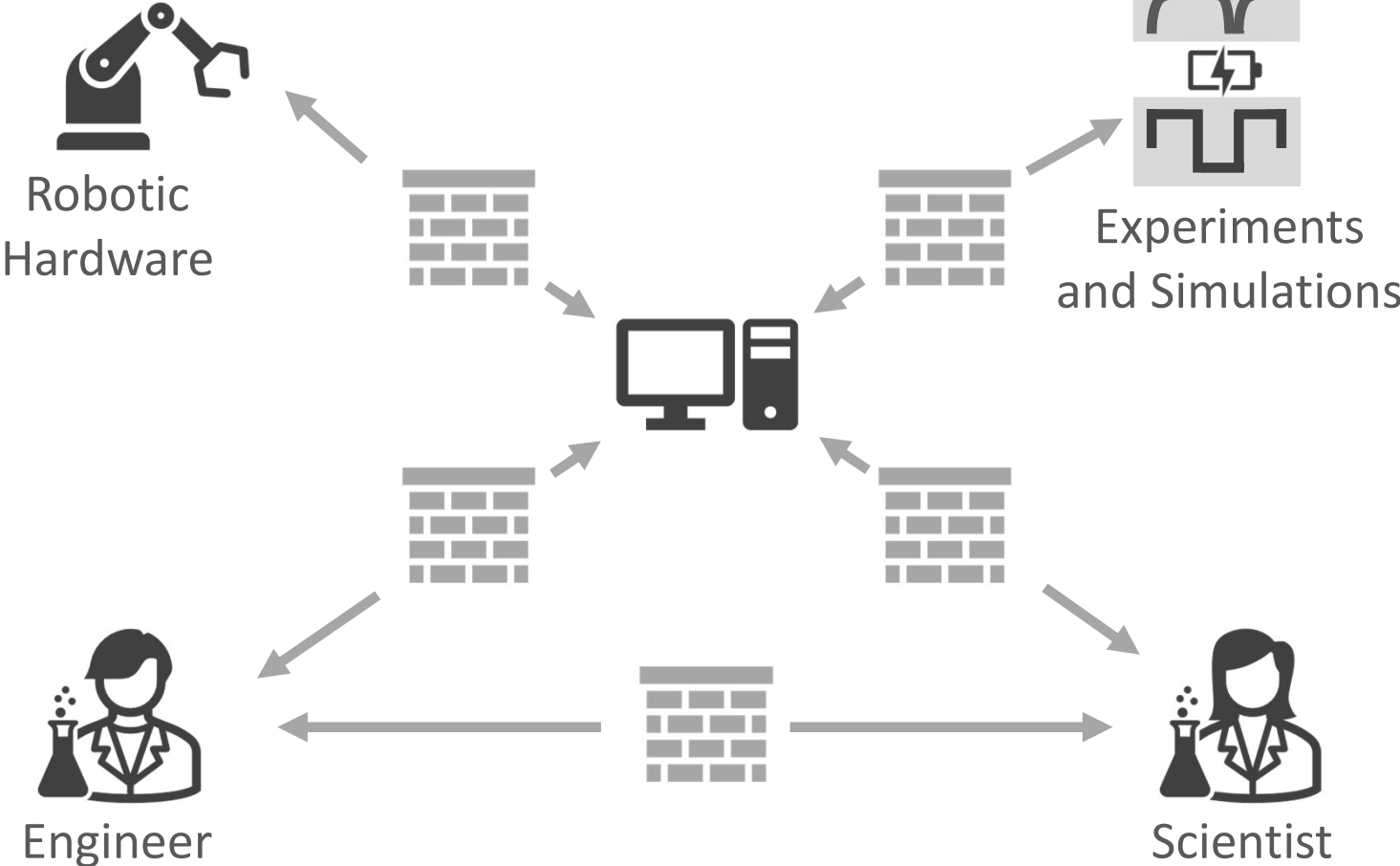
**How to link data**

**What can we do with linked data**



# What are we solving?

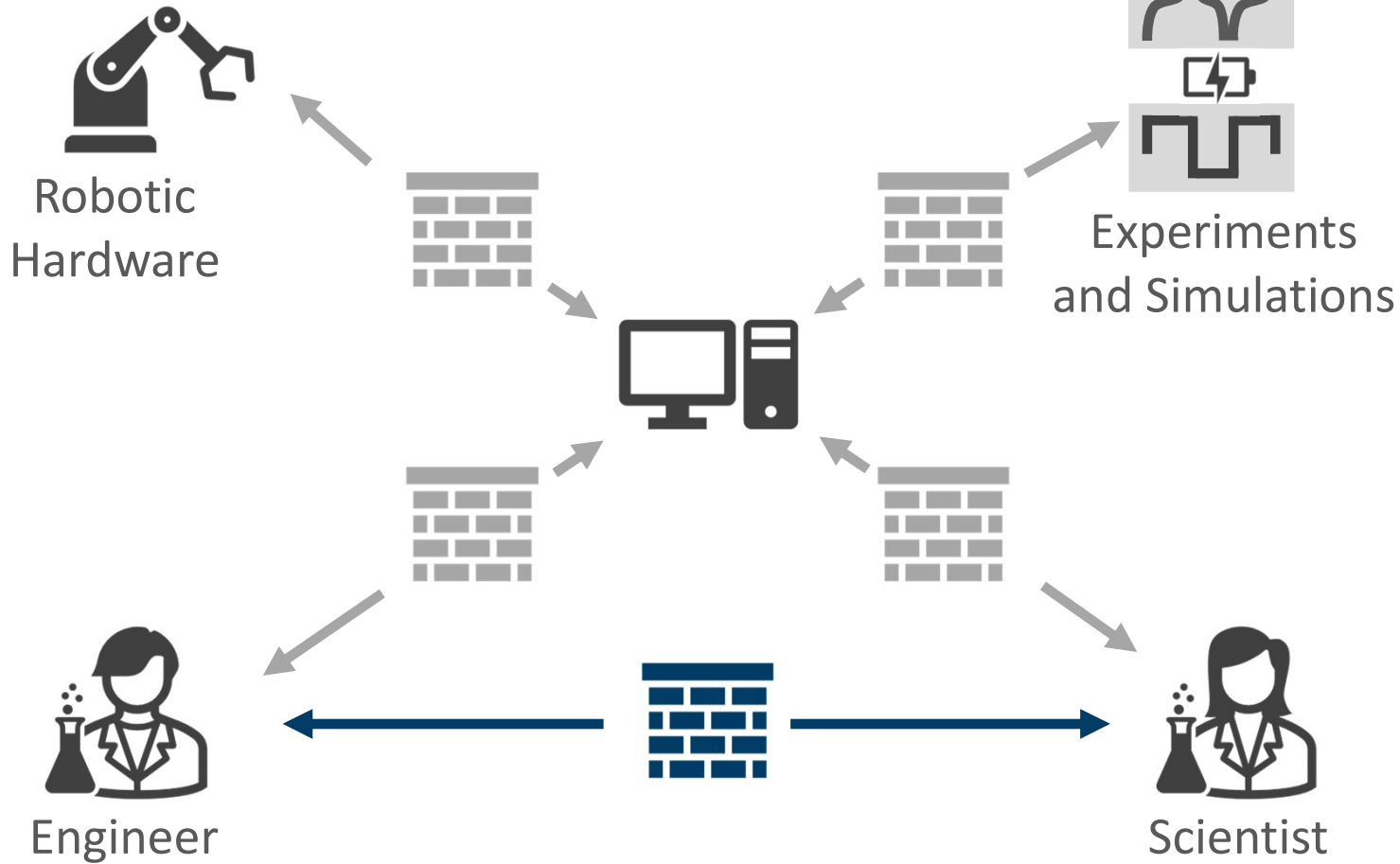
A communication issue





# What are we solving?

A communication issue



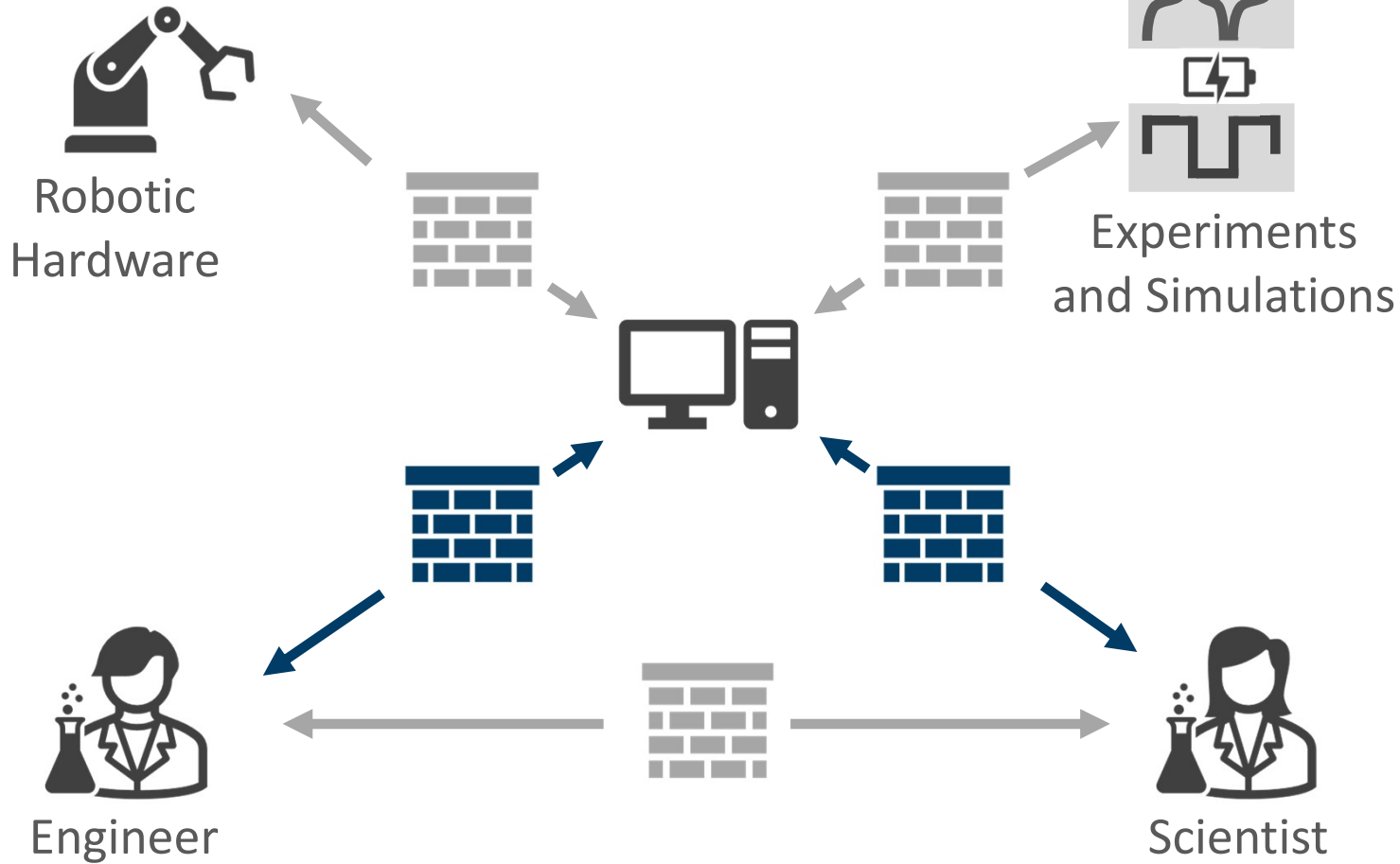
## Interaction between human actors

- Terminology
- Languages
- Requirements
- Technical background



# What are we solving?

A communication issue



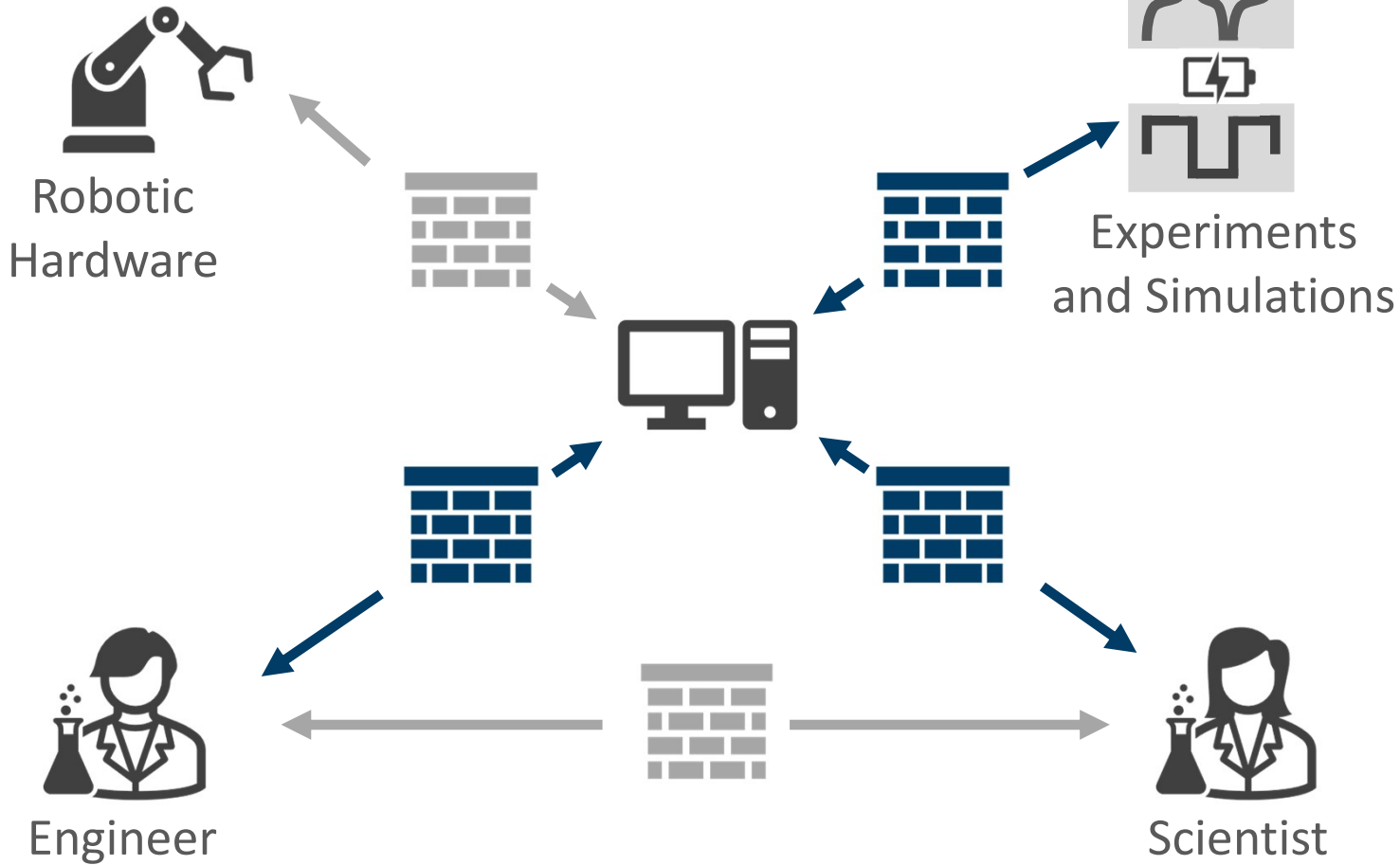
## Exchange of data

- Formats
- Metadata
- Methodology
- Postprocessing
- Access



# What are we solving?

A communication issue



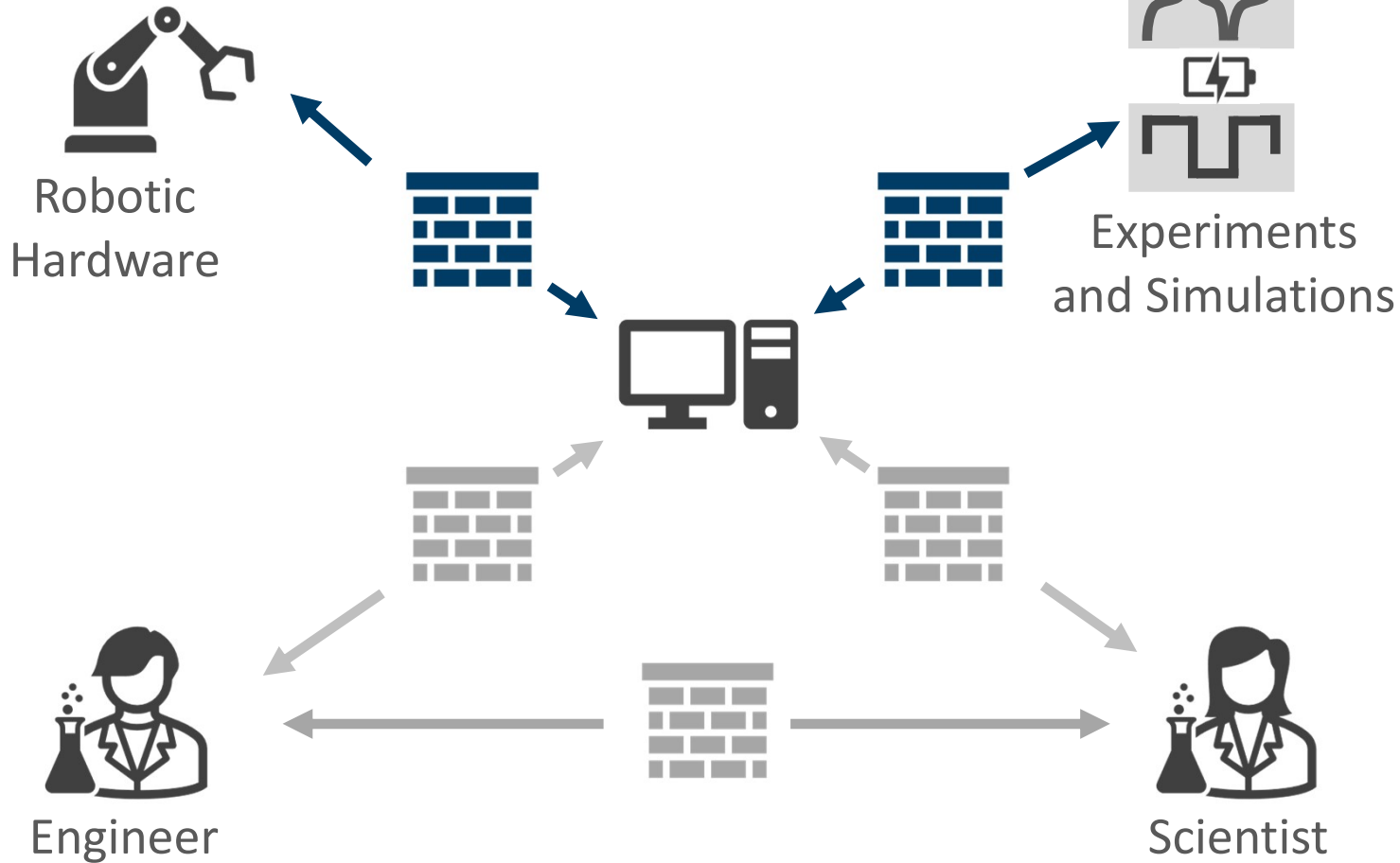
## Reproducibility

- Scope
- Inputs
- Methodology
- Pre-/post-processing
- Comparison
- Provenance



# What are we solving?

A communication issue



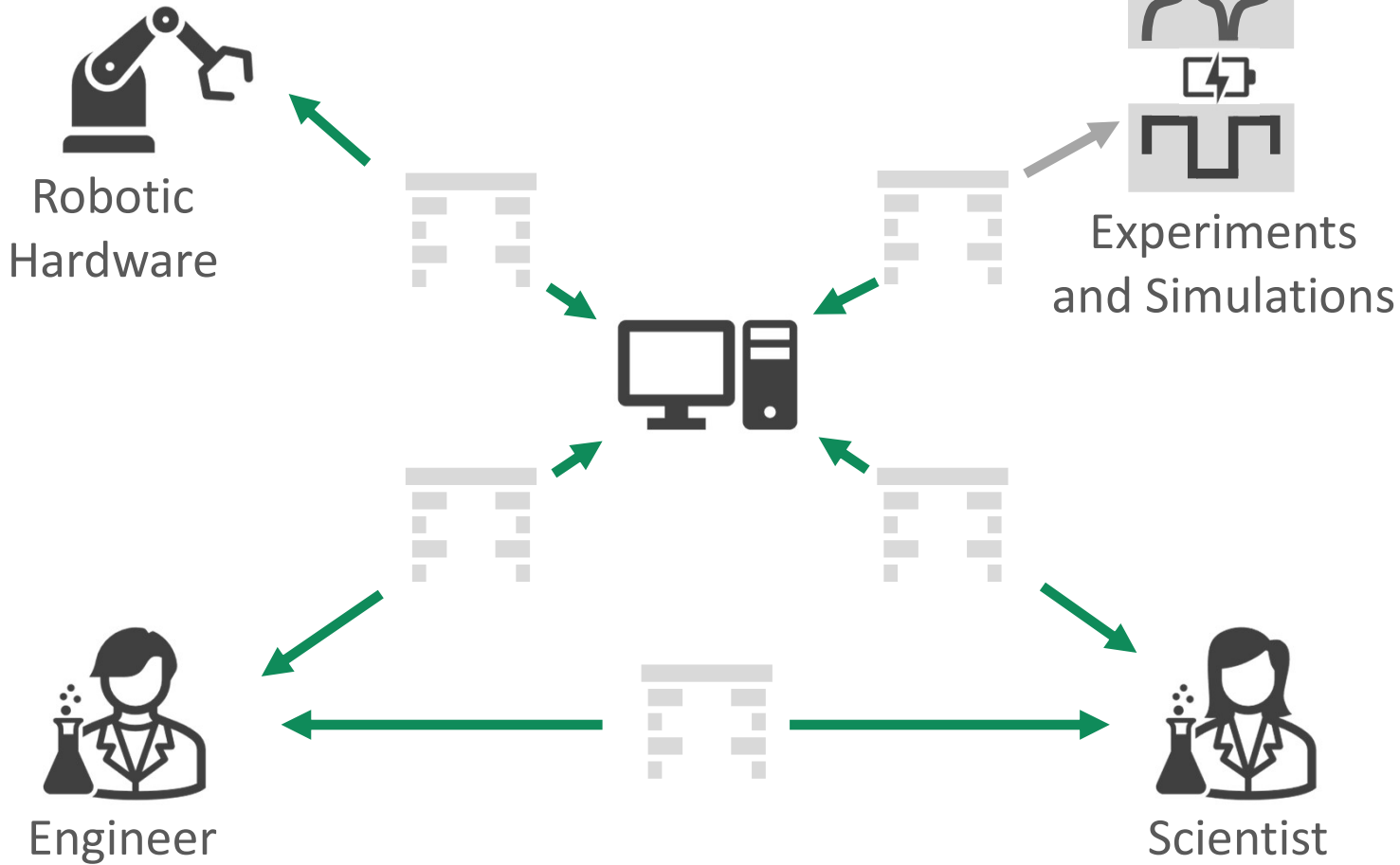
## Interaction with autonomous machines

- Data Exchange format
- Metadata
- Safety and security
- Robustness





# What are we solving?



**Linked data breaks barriers down by ensuring**

- Enough context to understand data
- Enough information to reproduce results
- Streamlined interactions with hardware
- Streamlined exchange and access



SINTEF

# How to link data

In 3 ~~easy~~ steps



# How to link data

## 1. Assign unique identifiers

**Unique identifier:** Sequence of characters uniquely identifying a resource used in the web. IDs must resolve to a document offering further data about the resource.

### Person

<https://orcid.org/0000-0003-2954-1233>

resolve

The screenshot shows the ORCID profile for Eibar Flores. The profile includes the ORCID iD, name, activities, and employment history. The employment history shows a position at SINTEF Industri, Trondheim, NO, from 2022-08-15 to the present, as a Research Scientist (Sustainable Energy Technologies).

### Dataset

<https://archive.big-map.eu/records/vzm2s-m5484>

resolve

The screenshot shows a dataset record for 'FINALES (09/2023) – Electrolyte Optimization for Maximum Conductivity and for Maximum Cycle Life'. The record includes a list of authors, a description of the study, and a 'Show affiliations' button.

### Publication

<https://doi.org/10.1002/batt.202100117>

resolve

The screenshot shows a publication record for 'Data Management Plans: the Importance of Data Management in the BIG-MAP Project'. The record includes the title, authors, publication date, and a graphical abstract.



# How to link data

## 2. Use open, controlled vocabularies

**Controlled vocabulary:** List of concepts of predefined, agreed and curated terms with unambiguous meaning. Concepts themselves must be uniquely identified (URLs).



### **Web vocabularies: digital resources**

Datasets, file formats, images, multimedia, events, locations, persons, software, ...



### **BattINFO: batteries + electrochemistry**

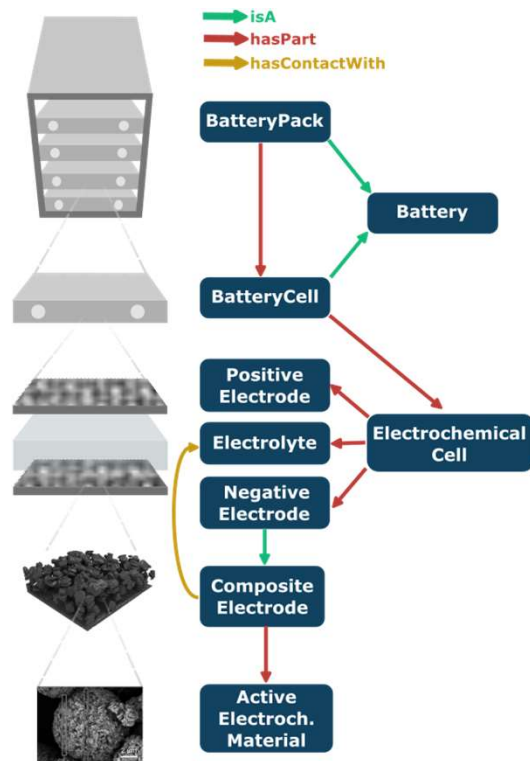
Voltage, current, EIS, CCCV, Battery, Electrode, Active Material, LFP, NCM, SEI, Lithium Plating ...



# How to link data

## 2. Use open, controlled vocabularies

**BattINFO:** a machine-readable description of concepts in batteries and electrochemistry.



Concepts are organized as a network:

**Nodes:** Battery Concepts

**Edges:** relations among concepts

Provides:

- The vocabulary to index battery data
- The connections representing battery knowledge



# How to link data

## 2. Use open, controlled vocabularies

**Schema.org:** a machine-readable description of web resources

The screenshot shows a Google search result for "The Matrix". At the top, there's a search bar with "the matrix" entered. Below it, the movie title "The Matrix" is displayed with a small image and a rating of 13. Navigation tabs include "Overview", "Cast", "Watch movie", "Reviews", "Trailers & clips", and "Quotes". The "Cast" section lists Keanu Reeves (Neo), Carrie-Anne Moss (Trinity), Laurence Fishburne (Morpheus), Hugo Weaving (Agent Smith), Gloria Foster (The Oracle), and Marcus Chong (Tank). The "Watch movie" section offers options to watch on Amazon Prime Video, TV 2 Play, and YouTube. The "About" section shows an IMDb rating of 8.7/10 and a Rotten Tomatoes score of 83%. A Wikipedia link is also present.

Concepts are organized as hierarchy:

**Nodes:** Web Concepts

**Edges:** Relations among concepts

Provides:

- The vocabulary to describe and index web resources
- Interoperability for search engines (e.g. google)



# How to link data

## 2. Use open, controlled vocabularies

**Additional vocabularies:** tailored to some domains.

### CSVW Namespace Vocabulary Terms

W3C Document 06 June 2017



**Describes:** CSV (tabular) files  
**Maintainer:** W3C  
**Prefix:** "csvw"



**Describes:** Characterization of materials  
**Maintainer:** EMMC Task Force  
**Prefix:** "chameo"



**Describes:** Web resources  
**Maintainer:** The Dublin Core™ Metadata Initiative  
**Prefix:** "dc"

European Research Information Ontology (EURIO)



**Describes:** Research projects  
**Maintainer:** European Union Vocabularies  
**Prefix:** "eurio"



# How to link data

## 3. Describe and link data in open format

**JSON-LD:** Method of annotating and linking data with the widely-used, machine-readable JSON format.

Test.csv

Col 1	Col 2	Col 3

Test\_metadata.jsonld

```
"@context": "https://mydomain.com/context.json"
"csvw:tableSchema": {
  "csvw:columns": [
    {
      "csvw:name": "time",
      "csvw:titles": "Time / s",
      "csvw:propertyUrl": {
        "@type": "Time",...},...}
    {
      "csvw:name": "voltage",
      "csvw:titles": "Voltage / V",
      "csvw:propertyUrl": {
        "@type": "CellVoltage"
      },
    },
  ],
  ...
}
```

Annotations become entries in the hierarchy of JSON fields

Annotations might live:

- Within the same data file (e.g. HDF5)
- As a separate accompanying file to the data
- All data, metadata and annotations in single JSON file



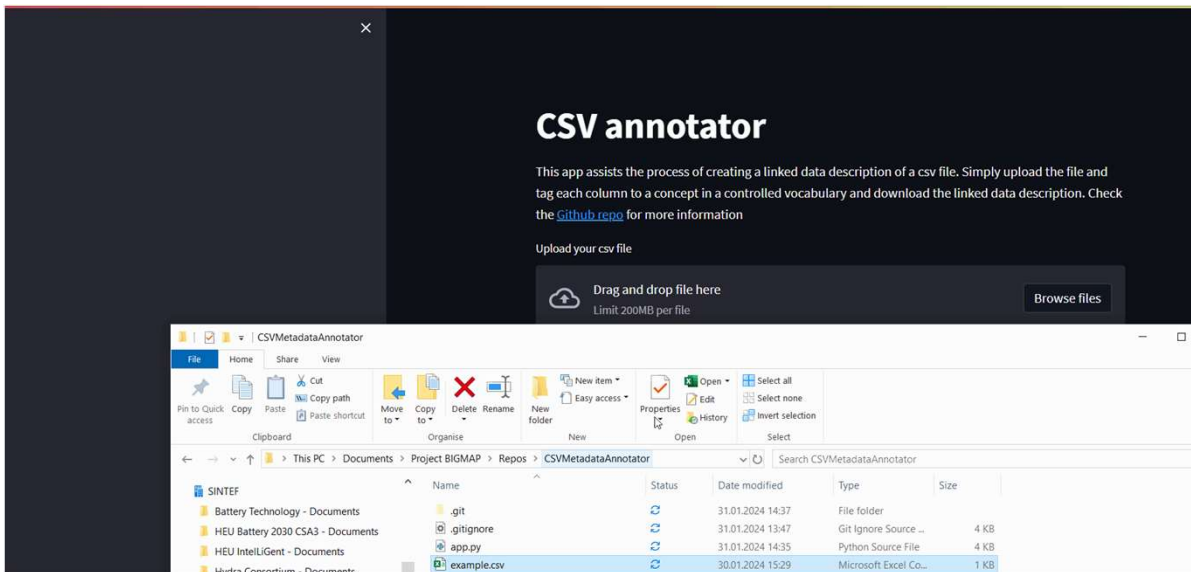


# How to link data

## 3. Describe and link data in open format

Do I have to write all these annotations manually?

**Hopefully NOT:** we build User Interfaces to make annotations user-friendly



Test\_metadata.jsonld

```
"@context": "https://mydomain.com/context.json"
"csvw:tableSchema": {
  "csvw:columns": [
    {
      "csvw:name": "time",
      "csvw:titles": "Time / s",
      "csvw:propertyUrl": {
        "@type": "Time",...}
    },
    {
      "csvw:name": "voltage",
      "csvw:titles": "Voltage / V",
      "csvw:propertyUrl": {
        "@type": "CellVoltage"
      },
    },
    ...
  ]
}
```



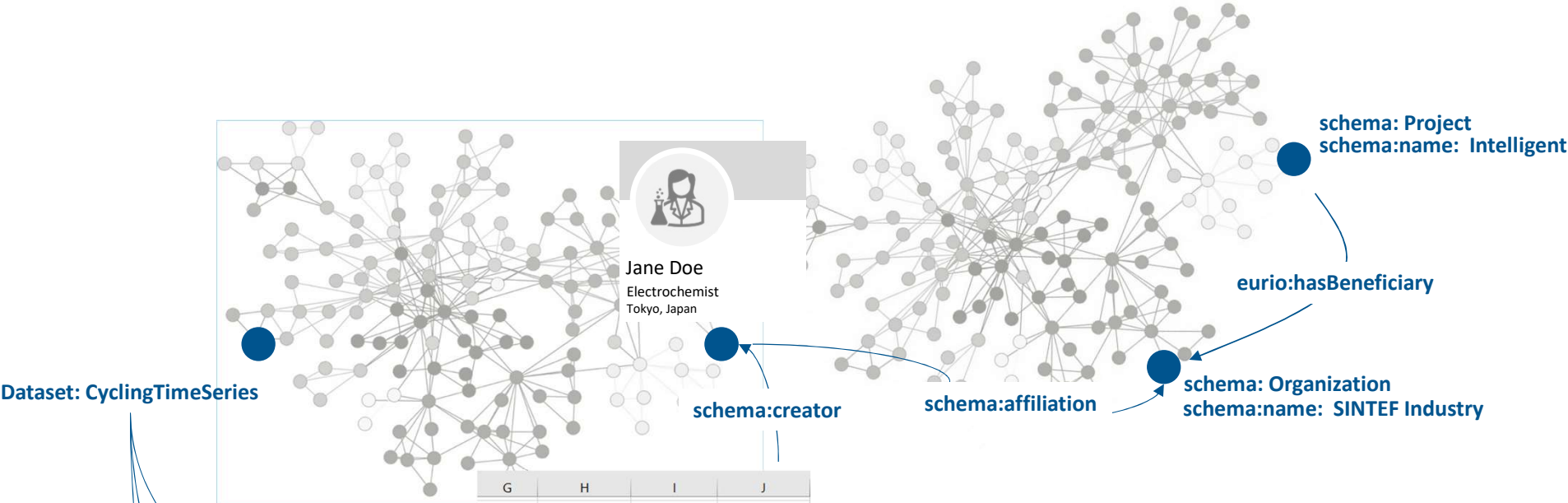
SINTEF

# Linked data: what can we do with it?



# What can we do with Linked Data

## Build a network of battery assets



G	H	I	J
Cap. [Ah]	Ener. [Wh]	Current [A]	Voltage [V]
0	0	0	3,2253
0	0	0	3,2227
0	0	0	3,221
0	0	0,000014	3,223
4E-08	0,00000013	0,00001398	3,2662
8E-08	0,00000025	0,00001398	3,304
1,2E-07	0,00000038	0,00001398	3,3384
1,6E-07	0,00000051	0,00001398	3,3692
1,9E-07	0,00000064	0,00001398	3,3974
2,3E-07	0,00000078	0,00001397	3,4226
2,7E-07	0,00000091	0,00001398	3,4461
3,1E-07	0,00000104	0,00001398	3,4681

Column: Voltage

Column: Capacity

Column: Voltage

**Network of battery assets**

Annotated with understandable attributes  
Uniquely identifiable  
Linked to other data: people, organizations...



# What can we do with Linked Data

## Make network accessible in Knowledgebase

Autonomous experiment platform

URI: <https://institution.com/api/battexp>



Book.pdf

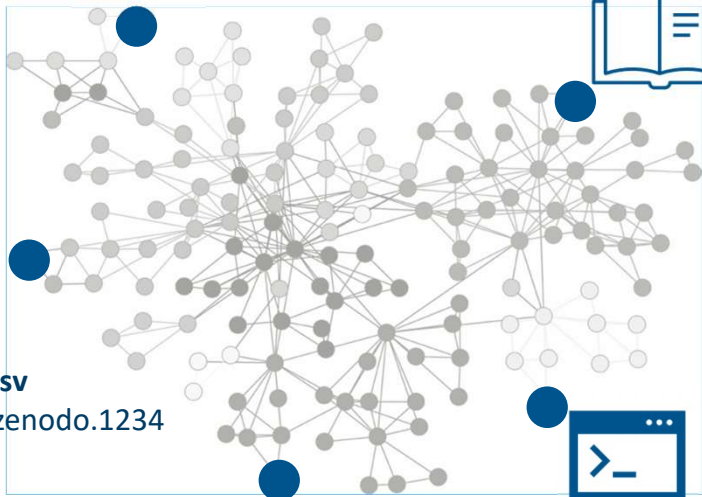
ISBN: 0393866866 5678



Col 1	Col 2	Col 3

Cycling\_data.csv

DOI: 10.5281/zenodo.1234



Battery\_simulation.py

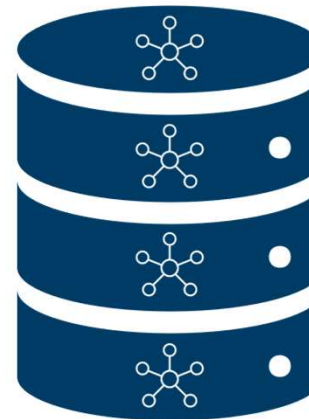
URI: <https://github.com/battsim>



Jane Doe

ORCID: 000-0123-456

Collect jsonld and populate graph DB



Query Response Interface





# Upcoming exciting talks

Autonomous experiment platform

URI: <https://institution.com/api/battexp>

**12:15**  
**Nukorn Plainpan**  
**(EMPA)**



Book.pdf

ISBN: 0393866866 5678



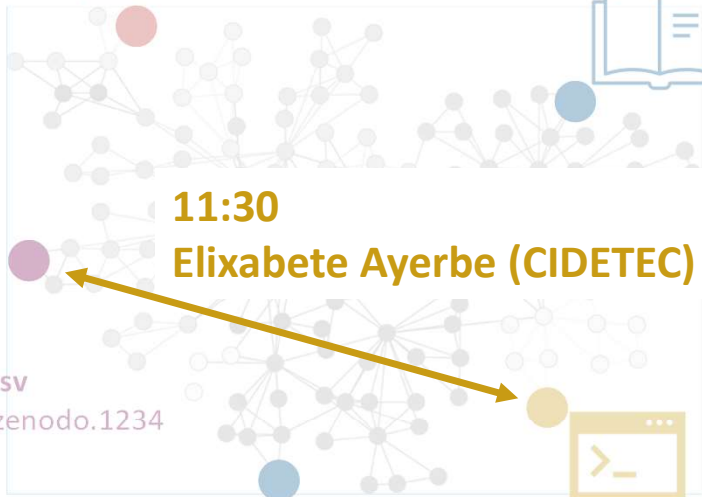
Col1	Col2	Col3

**11:30**

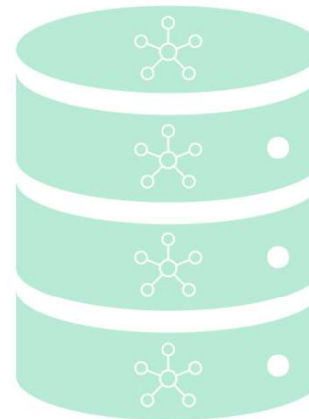
**Elixabete Ayerbe (CIDETEC)**

Cycling\_data.csv

DOI: 10.5281/zenodo.1234



Collect jsonld and  
populate graph DB



Query  
Response  
Interface



**10:45**  
**Brady Planden (Oxford)**



Jane Doe

ORCID: 000-0123-456

Battery\_simulation.py

URI: <https://github.com/battsim>

**14:00**  
**Philipp Veit (KIT)**  
**Lukas Gold (Fraunhofer ISC)**



SINTEF

# Linked data

1. Uniquely identified and referencing other data (URLs)
2. Annotated with controlled vocabularies (Ontologies)
3. In a machine-readable format supporting annotations (JSON-LD)

**Linked data enables reproducing, exchanging and automating battery research.**



SINTEF

# Technology for a better society