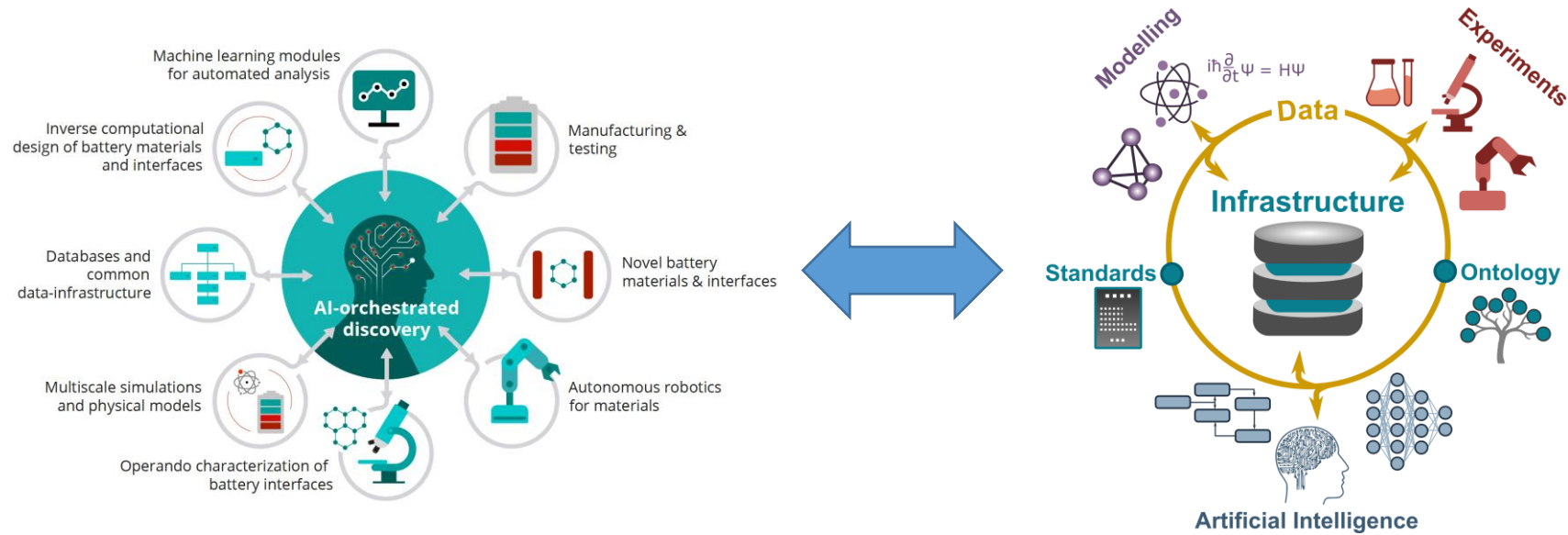




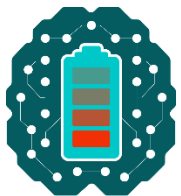
BIG-MAP

Roadmap for Research Data Management



Ivano E. Castelli

Technical University of Denmark

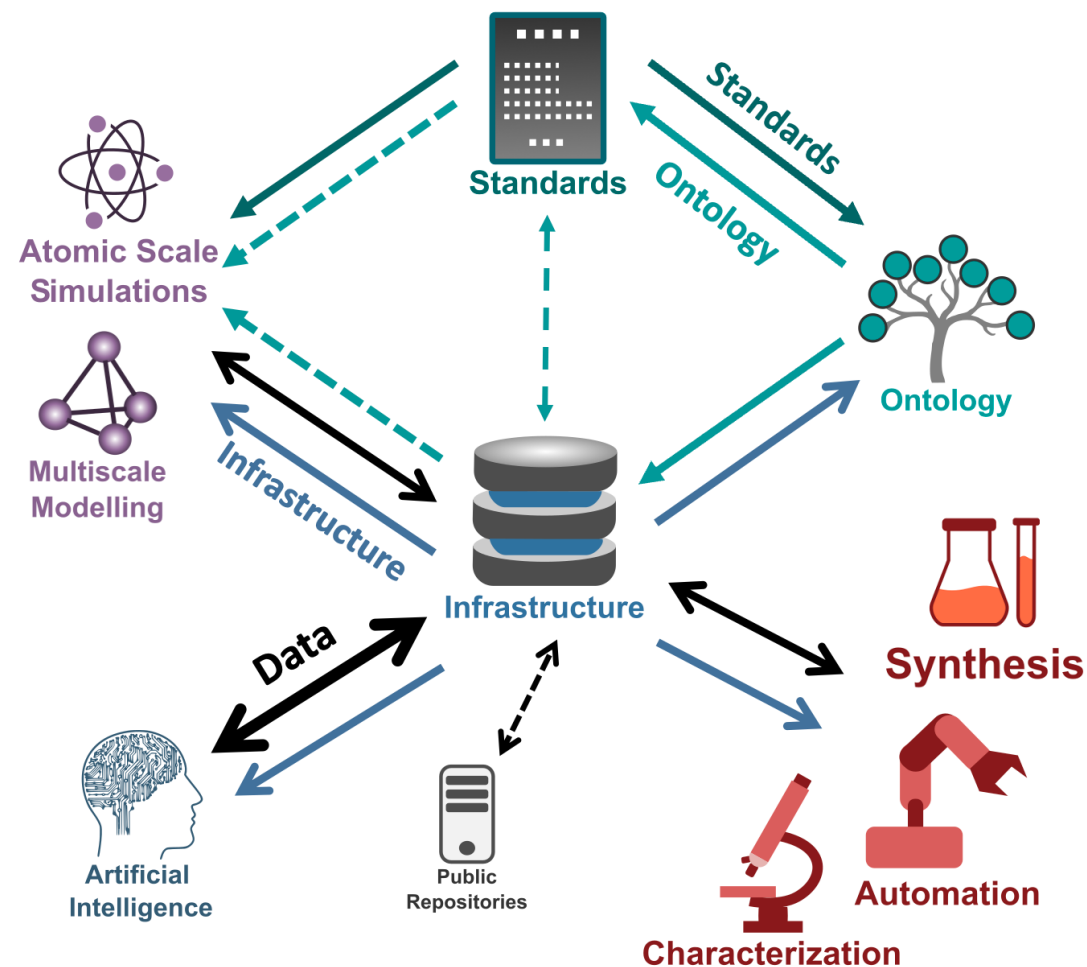


Data is the Key

BIG-MAP

Research **DATA** is far more than numbers (results)

- Documents (text, Word), spreadsheets
- Laboratory notebooks, photographs, films
- Samples
- Models, algorithms, scripts
- Methodologies and workflows
- Standard operating procedures and protocols
- Contents of an application (input, output, log files for analysis software, simulation software ..)
- ...



Batteries & Supercaps | Concepts | doi.org/10.1002/batt.202100117 | Chemistry Europe | European Chemical Societies Publishing

Very Important Paper

Data Management Plans: the Importance of Data Management in the BIG-MAP Project[]**

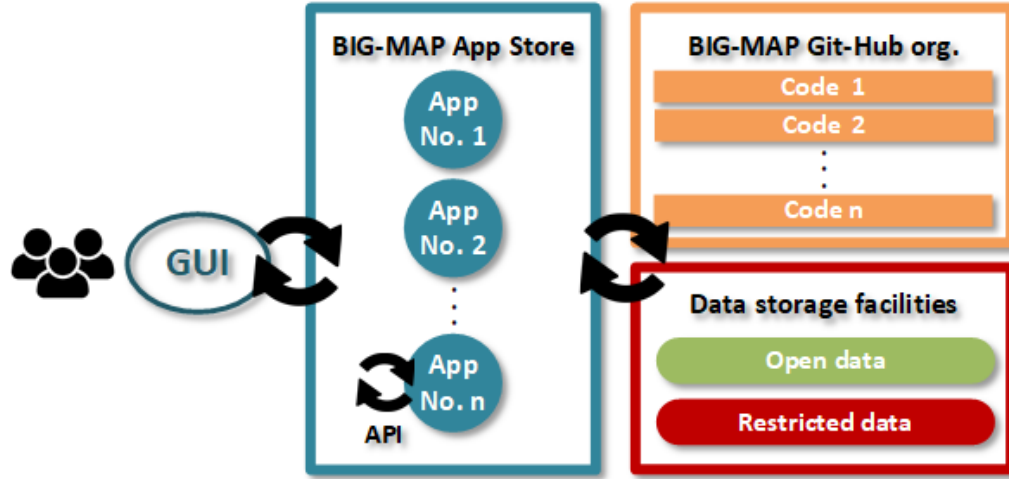
Ivano E. Castelli,^[a] Daniel J. Arismendi-Arrieta,^[b] Arghya Bhowmik,^[a] Isidora Cekic-Laskovic,^[c] Simon Clark,^[d] Robert Dominko,^[e] Eibar Flores,^[a] Jackson Flowers,^[f] Karina Ulvskov Frederiksen,^[a] Jesper Friis,^[g] Alexis Grimaud,^[h, i] Karin Vels Hansen,^[a] Laurence J. Hardwick,^[j] Kersti Hermansson,^[b] Lukas Königer,^[k] Hanne Lauritzen,^[a] Frédéric Le Cras,^[l] Hongjiao Li,^[m] Sandrine Lyonnard,^[n] Henning Lormann,^[o] Nicola Marzari,^[p] Leszek Niedzicki,^[q] Giovanni Pizzi,^[r] Fuzhan Rahmanian,^[s] Helge Stein,^[t, v] Martin Uhrin,^[a] Wolfrano Wenzel,^[n] Martin Winter,^[c] Christian Wölke,^[c] and Teic Venna^[a]

Castelli *et al.*, *Battery & Supercaps* **4**, 1803 (2021)

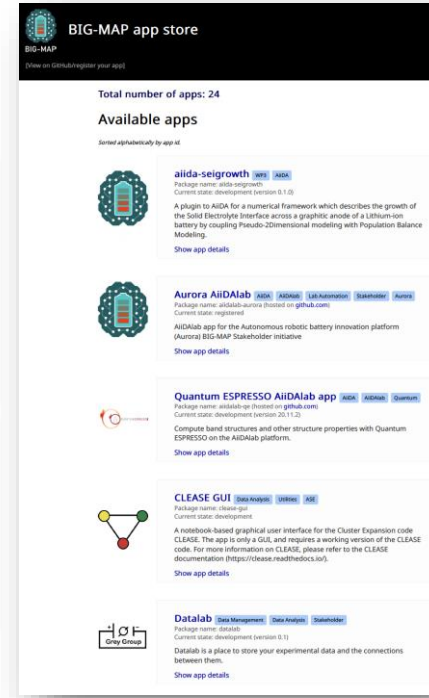


BIG-MAP

The BIG-MAP Infrastructure



<https://archive.materialscloud.org/>
<https://big-map.github.io/big-map-registry/>

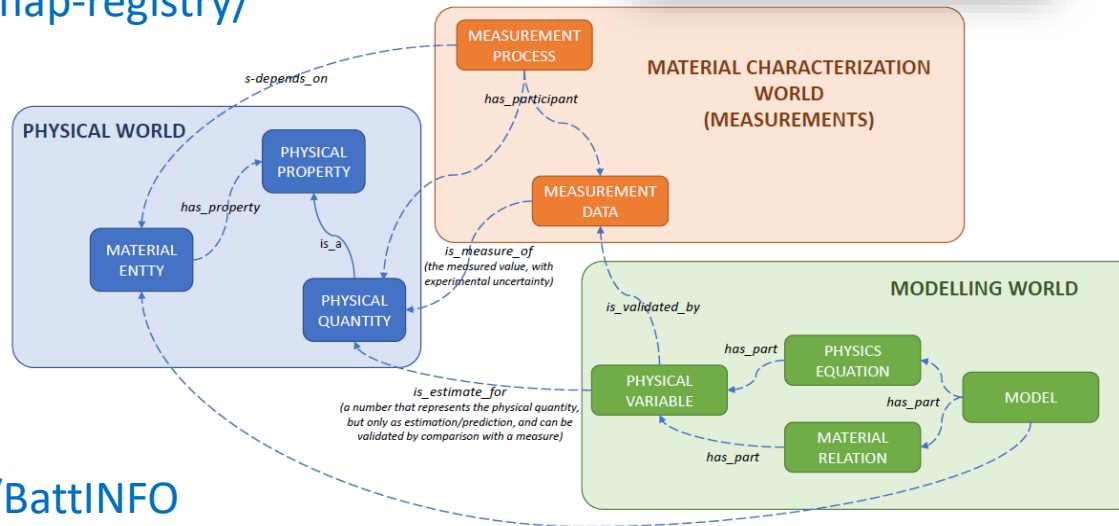


BIG-MAP NOTEBOOK

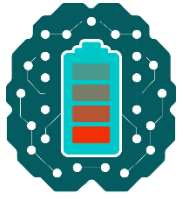


New Characterization Entry	Search Characterization Entry
New Electrochemistry Entry	Search Electrochemistry Entry
New Simulation/Modeling Entry	Search Simulation/Modeling Entry
New Synthesis/Coatings Entry	Search Synthesis/Coatings Entry
New SOP File	Search SOP File
Ontology Explorer	

<http://big-map-notebook.eu/>

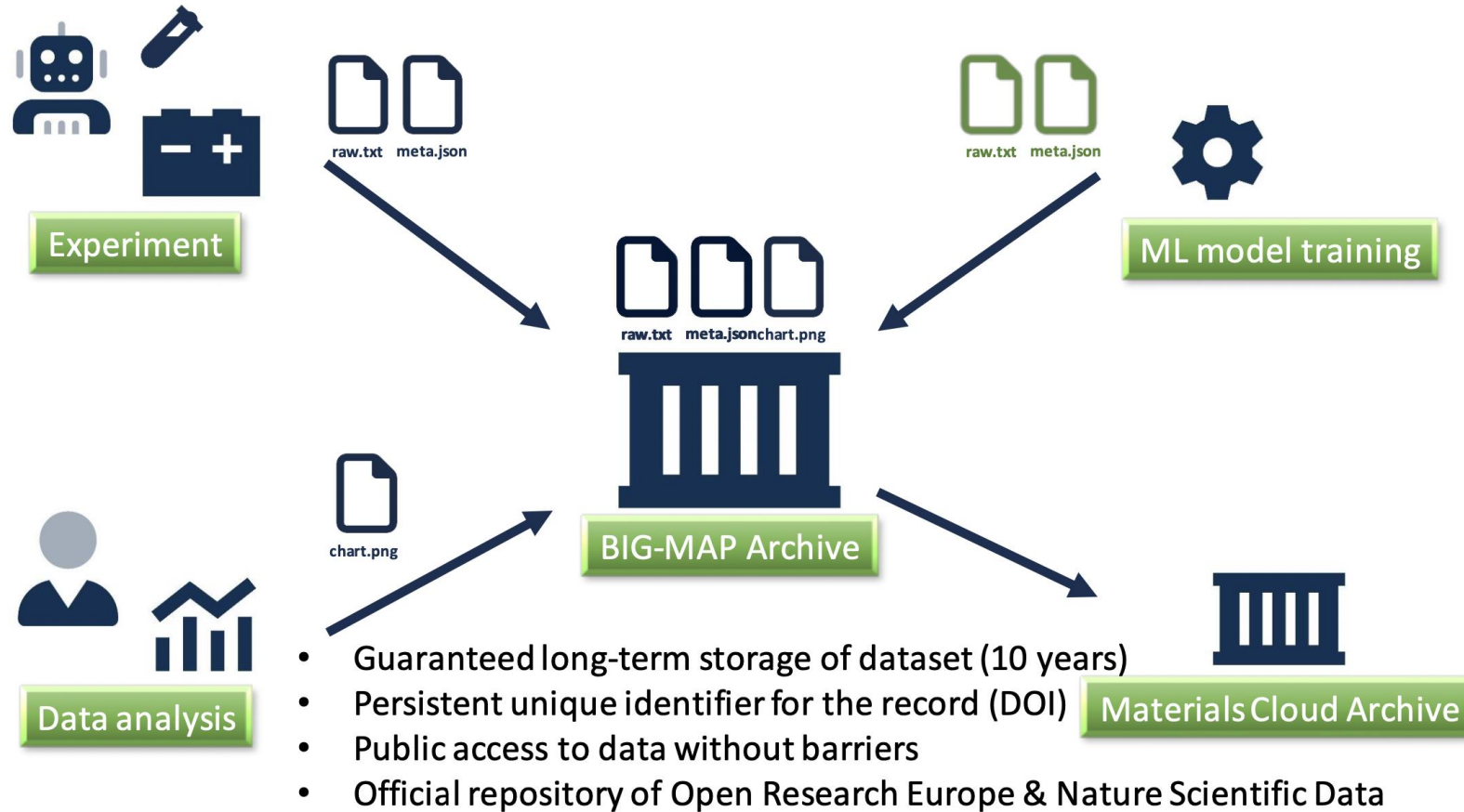


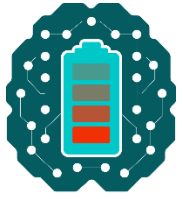
<https://github.com/BIG-MAP/BattINFO>



BIG-MAP

BIG-MAP Data Archive





BIG-MAP Data Archive



BIG-MAP

Welcome to BIG-MAP Archive

Search publications



A Repository for Sharing Files within the BIG-MAP Community



Upload

Upload files up to 100 GB in size & of any format.



Publish

Make your files available to other BIG-MAP members.



Search

Search for records with ease.



Download

Download files of published records.



Leverage API

Exchange files with the archive from your own client applications.



Update

Create new versions of your published records.

Data needs metadata to become FAIR

- stored in a restricted-access storage (e.g., for subsequent analysis)
- shared within the BIG-MAP/BATTERY 2030+ community (e.g., for training a ML model)
- uploaded to an open-access repository (e.g., manuscript submission)

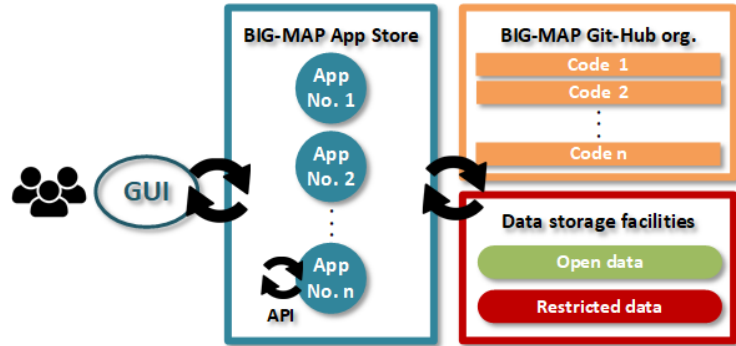
BIG-MAP Archive:

- is a research data repository (restricted to the BIG-MAP community, open to BATTERY 2030+, fully open)
- offers data storage & data sharing (manually via GUI, programmatically via API)
- is based on a third-party open-source software (InvenioRDM developed by CERN & 10 other institutions), and customised



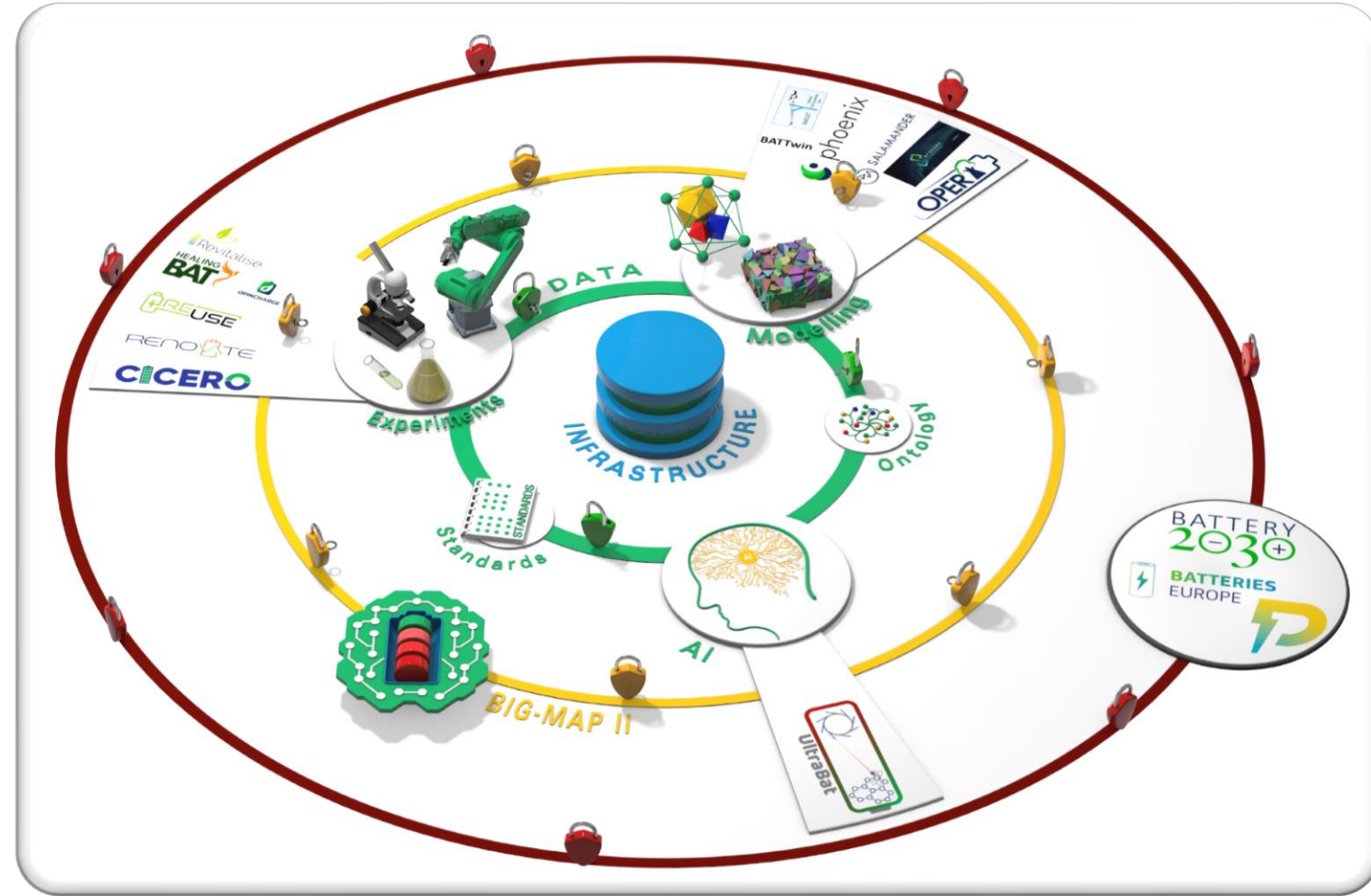
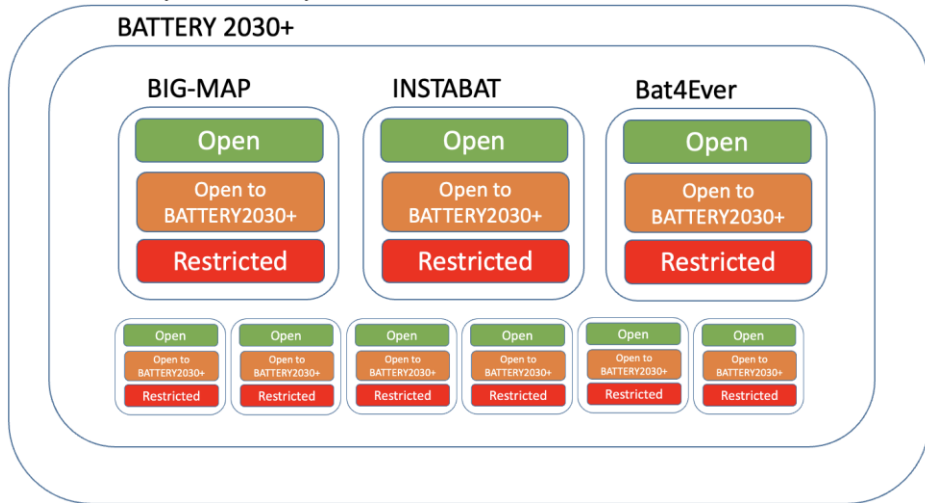
Connect Data between BATTERY 2030+ Projects

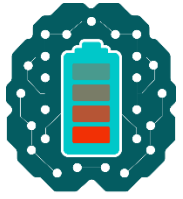
BIG-MAP



Battery Community

BATTERY 2030+





BIG-MAP

Connect Data between BATTERY 2030+ Projects



BIG-MAP community



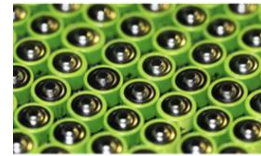
BIG-MAP
member



BATTERY2030+ community



BIG-MAP
PROJECT



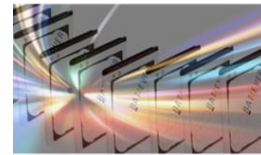
SENSIBAT
PROJECT



INSTABAT
PROJECT



HIDDEN
PROJECT



BAT4EVER
PROJECT



SPARTACUS
PROJECT

BATTERY 2030+





BIG-MAP

New Data Standards for the EU Battery Community



BIG-MAP community



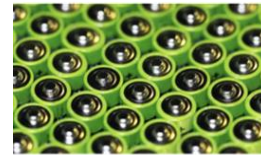
BIG-MAP member



BATTERY2030+ community



BIG-MAP
PROJECT



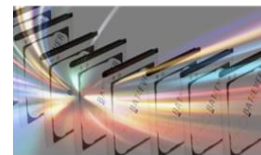
SENSIBAT
PROJECT



INSTABAT
PROJECT



HIDDEN
PROJECT



BAT4EVER
PROJECT



SPARTACUS
PROJECT



NEW PROJECTS

RUNNING PROJECTS

BatCat

BATTwin

Cicero

Renovate

ReUse

Revitalise

Streams

Big-Map

HealingBat

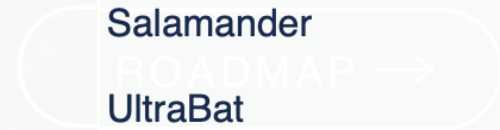
Opera

Opincharge

Phoenix

Salamander

UltraBat





BIG-MAP

Data repositories: BIG MAP Archive (<https://archive.big-map.eu/>)



BATTERY 2030+ Shared records My records New record Log in

Welcome to BIG-MAP Archive

Search shared records

A Research Data Repository for the BATTERY2030+ Initiative

- Upload**
Upload up to 100 GB per dataset.
- Share**
Privately share datasets with communities.
- Search**
Effectively search over dataset metadata.
- Download**
Download datasets in a single click.
- Leverage API**
Upload and download from custom client apps.
- Update**
Create new versions of datasets.

Tutorial Test on demo

Live since Nov 22

InvenioRDM v12
released May 24

Currently:
263 records
348 users
1.6 TB of data



Welcome to BIG-MAP Archive



A Research Data Repository for the BATTERY2030+ Initiative



Upload

Upload up to 100 GB
per dataset.



Share

Privately share datasets
with communities.



Search

Effectively search over
dataset metadata.



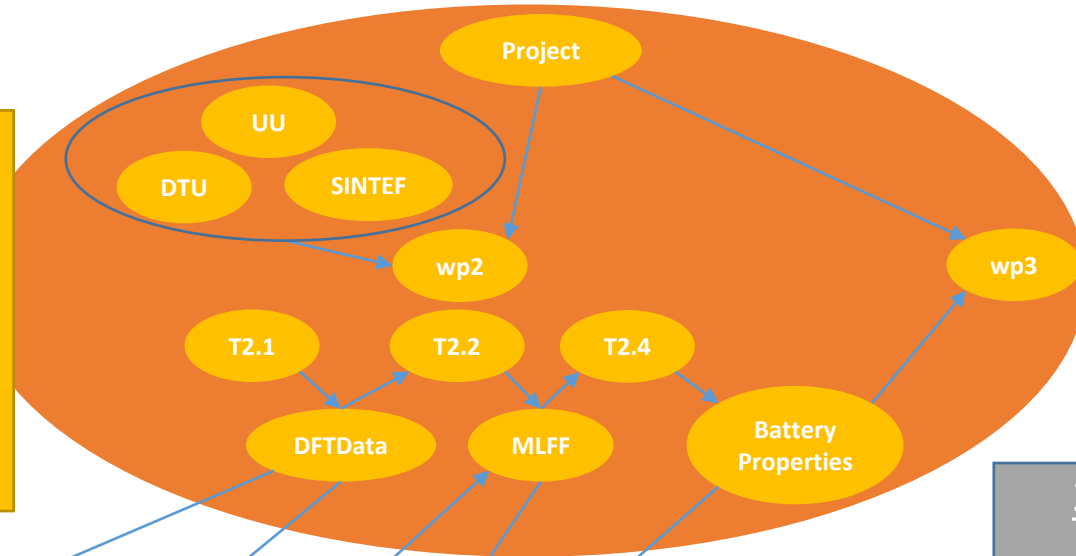


Ontology for Data – Towards a Dynamic DMP

BIG-MAP

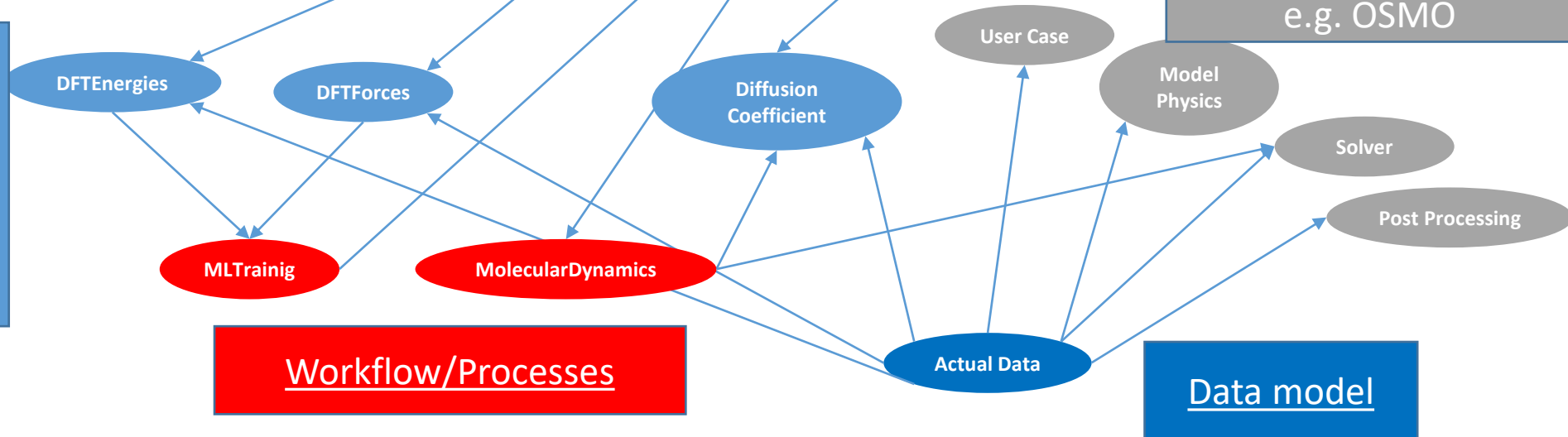
1. DMP Ontology

- WPs' and Tasks' objectives and participants
- Data types generated by each WP/Tasks
- Data types/information flow within and among WPs



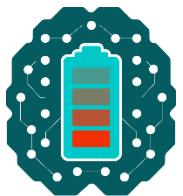
2. Domain Ontology

- Add description of the data using available ontologies



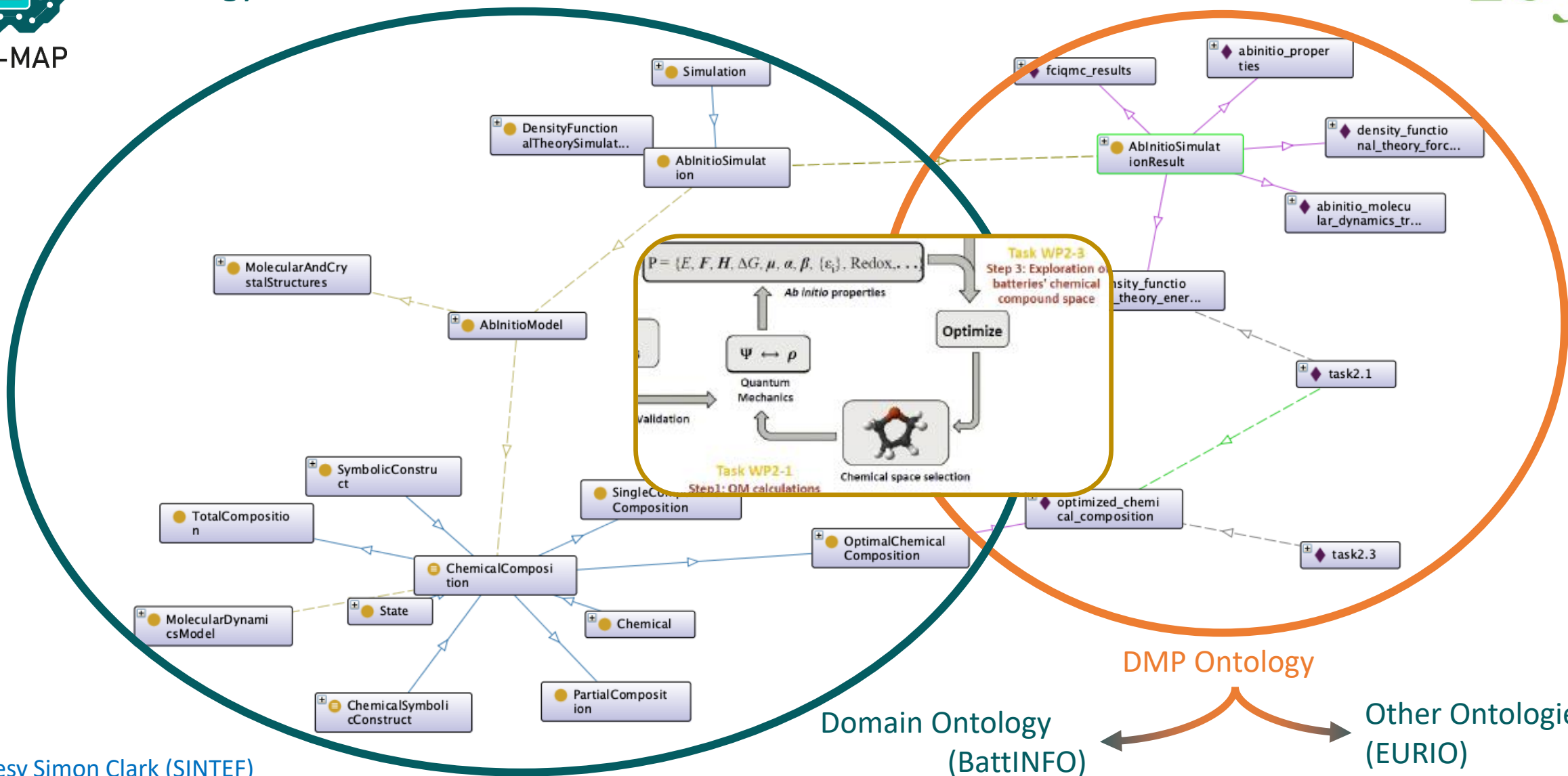
3. MODA/CHADA
Ontology
e.g. OSMO

Courtesy Simon Clark (SINTEF)



Ontology for Data – Data Flow

BIG-MAP



Courtesy Simon Clark (SINTEF)

Domain Ontology
(BattINFO)

DMP Ontology

Other Ontologies
(EURIO)

Monday, January 29, 2024

- Free arrival** – Optional activities/excursions
Official check-in time 16:00 (earlier if room available)
- 17:00 **Opening session: BIG-MAP and Battery 2030+**
Tejs Vegge • Kristina Edström
- 17:20 **Opening session: all project leads/representatives**
Quentin Bigouraux • Ivano Castelli • Gerhard Domann • Simon Clark • Joris de Hoog • Martin Thomas Horsch • Samson Yuxiu Lai • Kourosh Malek
- 18:20 **Presentations by BIG-MAP prize winners**
Eibar Flores • Fuzhan Rahmanian • Laura Hannemose Rieger • Tushar Thakur • Monika Vogler
- 19:30 **End of talks**
- 19:45 **Dinner**

Tuesday, January 30, 2024

- 08:00 **Breakfast**
- 09:00 **Data management plan for BIG-MAP**
Ivano Castelli
- 09:30 **Demo session on BIG-MAP Archive**
Valeria Granata
- 10:00 **Demo session on ELNs and semantic data annotations**
Fernando Caro • Lukas Gold • Simon Stier
- 10:30 **Coffee break**
- 11:00 **Demo session on FINALES and remote experiments**
Johan Carlsson • Matthias Albert Popp • Simon Steensen • Simon Stier • Monika Vogler
- 12:00 **Follow-up/wrap-up Q&A**
- 12:30 **Lunch**

- 13:30 **Demo desk: hands-on sessions (in parallel)**
- Hands-on session: uploading your data to the BIG-MAP Archive
Valeria Granata
 - Hands-on session: FINALES and remote experiments
Johan Carlsson • Matthias Albert Popp • Simon Steensen • Simon Stier • Monika Vogler
 - Hands-on session: Sharing ontologized scientific data through the BIG-MAP notebook
Fernando Caro • Lukas Gold

- 15:00 **Big picture session: panel discussion**
What are the special data requirements, formats, and quantities. What do we generate, what will we need in terms of data? Digital Materials and AI for Science in EU
Ivano Castelli • Kristina Edström • Gerhard Goldbeck • Nicola Marzari • Giovanni Pizzi • Gian-Marco Rignanesi • Tejs Vegge

16:00 **Coffee break**

- 16:30 **Demo session on BIG-MAP App Store**

Tushar Thakur

- 16:45 **Demo session on BIG-MAP Apps: AiIDAlab Quantum ESPRESSO app**
Xing Wang

- 17:15 **Demo on PANOSC search API in large-scale facilities**

Majid Ounsy

17:45 **Follow-up/wrap-up Q&A**

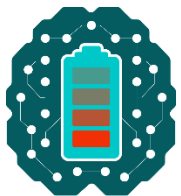
18:15 **End of talks**

19:00 **Dinner**

20:30 **Socializing evening**

Wednesday, January 31, 2024

- 08:00 **Breakfast**
- 09:00 **Big picture session: Introduction to EMMO, EMMC, Optimade**
Emanuele Ghedini • Gerhard Goldbeck • Gian-Marco Rignanesi
- 10:30 **Coffee break**
- 11:00 **Demo session on integration with BATTINFO**
Simon Clark
- 11:30 **Demo session on AURORA**
Edan Bainglass • Corsin Battaglia
- 12:00 **Conclusions**
- 12:30 **Lunch**



Acknowledgements



BIG-MAP

Batteries & Supercaps

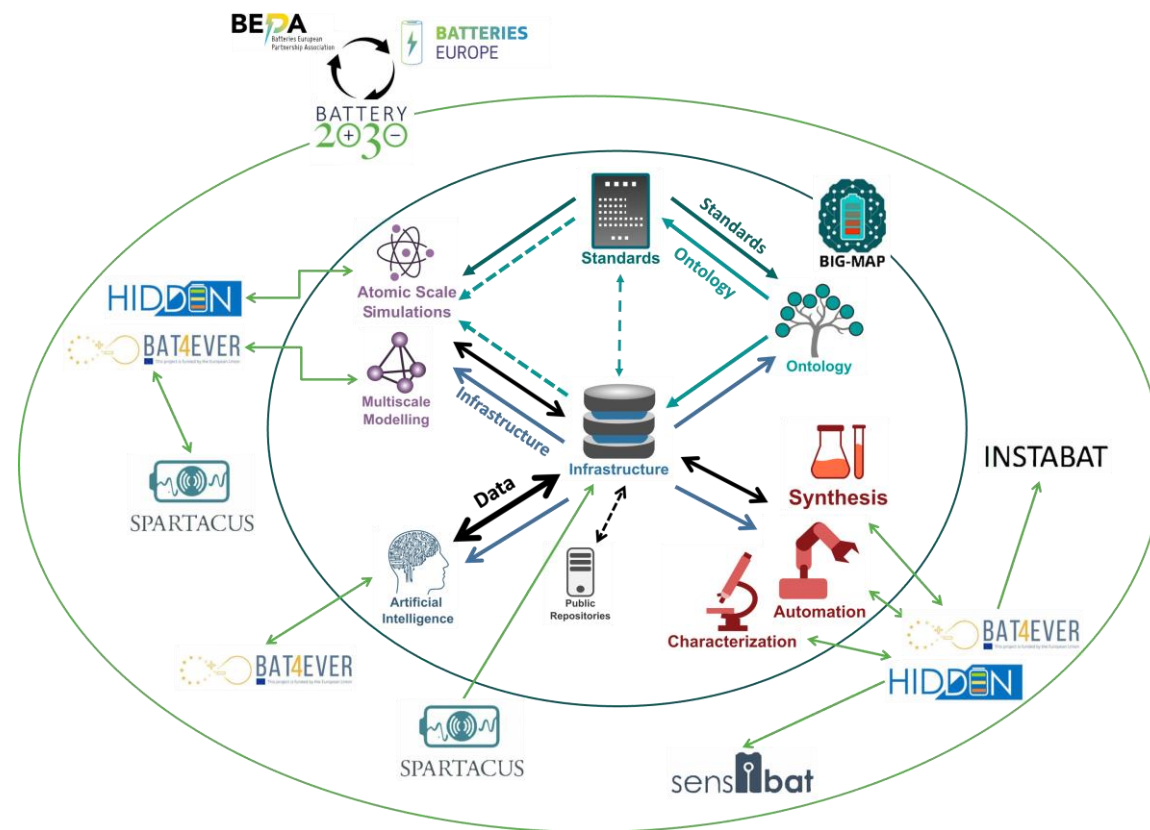
Concepts
doi.org/10.1002/batt.202100117

Chemistry Europe
European Chemical Societies Publishing

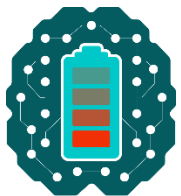
VIP Very Important Paper

Data Management Plans: the Importance of Data Management in the BIG-MAP Project[**]

Ivano E. Castelli,^[a] Daniel J. Arismendi-Arrieta,^[b] Arghya Bhowmik,^[a] Isidora Cekic-Laskovic,^[c] Simon Clark,^[d] Robert Dominko,^[e] Eibar Flores,^[a] Jackson Flowers,^[f] Karina Ulvskov Frederiksen,^[a] Jesper Friis,^[g] Alexis Grimaud,^[h, i] Karin Vels Hansen,^[a] Laurence J. Hardwick,^[j] Kersti Hermansson,^[b] Lukas Königer,^[k] Hanne Lauritzen,^[a] Frédéric Le Cras,^[l] Hongjiao Li,^[m] Sandrine Lyonnard,^[n] Henning Lormann,^[o] Nicola Marzari,^[p] Leszek Niedzicki,^[q] Giovanni Pizzi,^[p] Fuzhan Rahmanian,^[f] Helge Stein,^[f, r] Martin Uhrin,^[a] Wolfaana Wenzel,^[n] Martin Winter,^[c] Christian Wölke,^[c] and Teis Veage^[a]



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

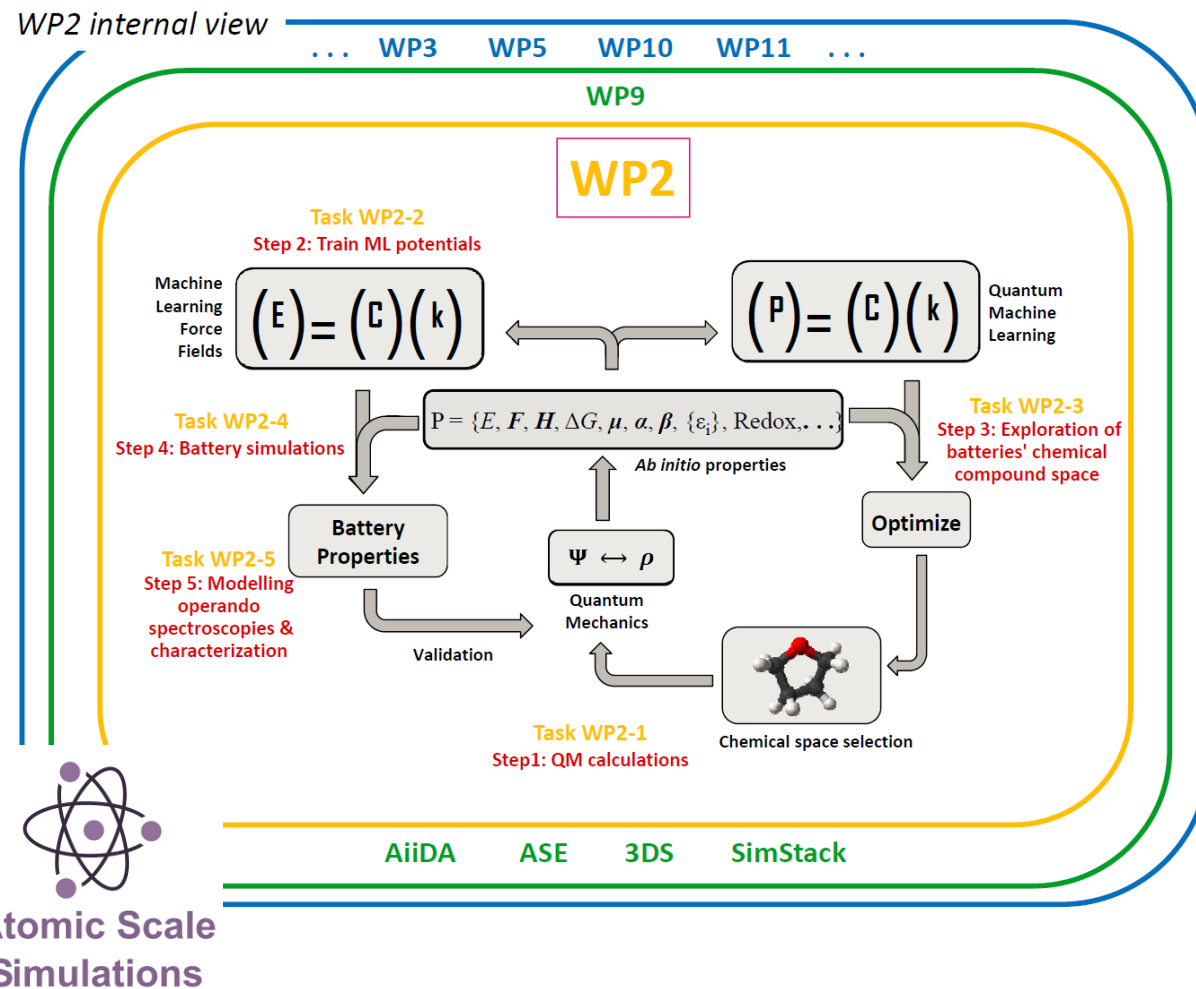


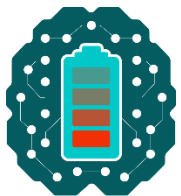
The Data Tables

BIG-MAP

Data generated in WP2

Datatype	Description	Data sets	Type	Format	Size
Electronic Structure: WFT, DFT, QMC	Structures, energy-related data, wave functions & electronic properties, ab-initio molecular dynamics (AIMD) trajectories, different types of spectra	Data generated by different tools: Engines (molecular): GAUSSIAN, ORCA, MOLPRO, TURBOMOLE, NWChem, QChem, ADF, PSI4, MRCC, NECI Engines (periodic): CP2K, VASP, QUANTUM ESPRESSO, Yambo, Castep, GPAW, QuantumATK, Crystal, NECI	Tarball files can be created from the calculation folder, including relevant inputs and output raw data	.tar.gz (an archive of input and output text, XML, netctdf, hdf5, or any other machine-readable file)	TB





The Data Tables

BIG-MAP

Data generated in WP2


Datatype	Description	Data sets	Type	Format	Size
Electronic Structure: WFT, DFT, QMC	Structures, energy-related data, wave functions & electronic properties, ab-initio molecular dynamics (AIMD) trajectories, different types of spectra	Data generated by different tools: Engines (molecular): GAUSSIAN, ORCA, MOLPRO, TURBOMOLE, NWChem, QChem, ADF, PSI4, MRCC, NECI Engines (periodic): CP2K, VASP, QUANTUM ESPRESSO, Yambo, Castep, GPAW, QuantumATK, Crystal, NECI	Tarball files can be created from the calculation folder, including relevant inputs and output raw data	.tar.gz (an archive of input and output text, XML, netcdf, hdf5, or any other machine-readable file)	TB




Atomic Scale Simulations

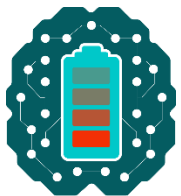
Castelli *et al.*, *Battery & Supercaps* **4**, 1803 (2021)

Data collected in WP2

WP	What	To be used for	Suggested		
			type	format	size
From WP5  Characterization	Parameters from structural and chemical characterizations at the local scale, including e.g., spectra (vibrational, absorption, ...)	Model refinement, cross-analysis, validation, and design of simulated experiments	Tarball files can be created with post-processed data and parameters of interest	.tar.gz (with databases and reproducible data sheets)	MB

Data delivered from WP2

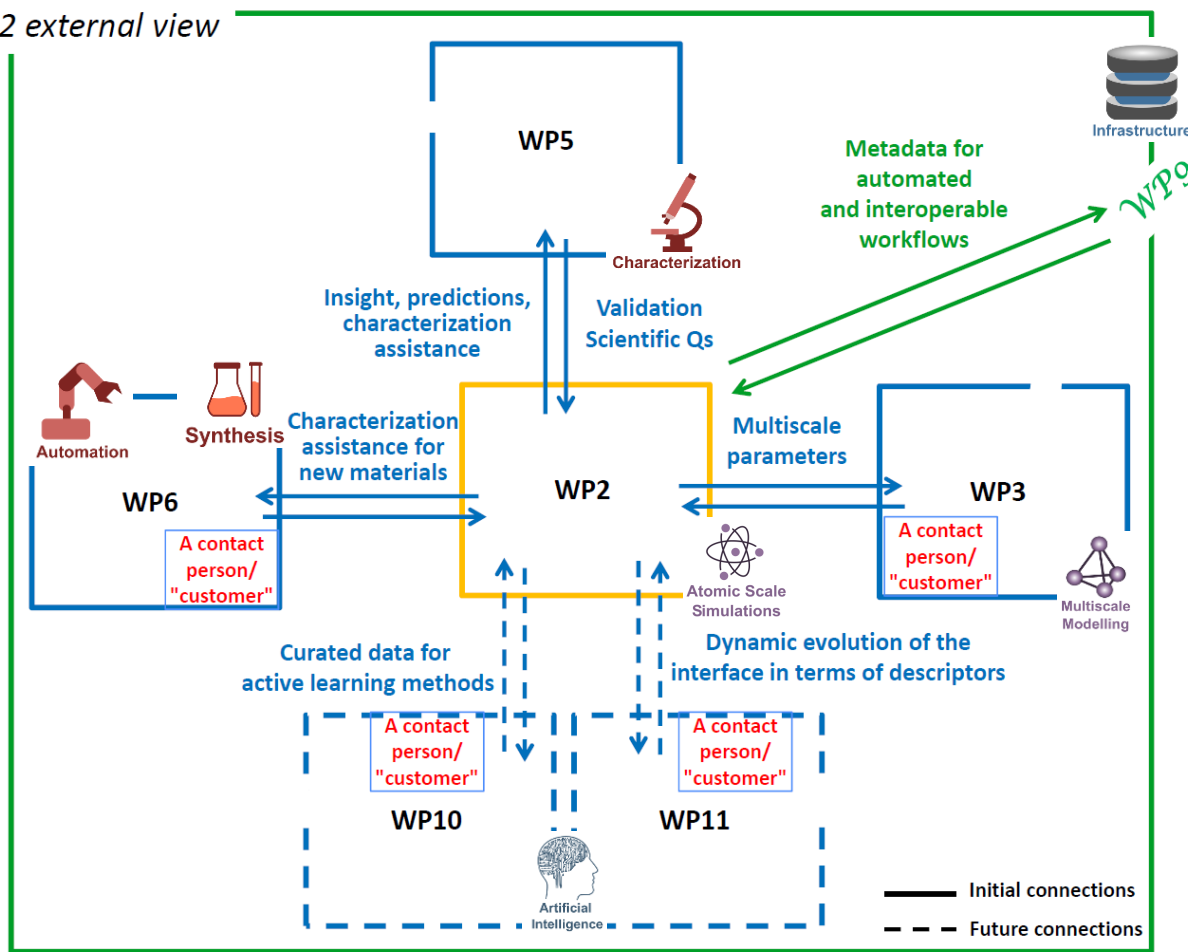
WP	What	Usable for	Suggested		
			type	format	size
To WP5  Characterization	Computational predictions for bulk/interfacial structures, "chemical environments", spectra, transport properties, or any experimental characterization requiring atomistic or electronic interpretation	Guiding the characterization effort and supporting the interpretation of the results	Tarball files can be created with post-processed data and parameters of interest	.tar.gz (with databases and reproducible data sheets)	TB




The Data Tables

BIG-MAP


WP2 external view



Data collected in WP2

WP	What	To be used for	Suggested		
			type	format	size
From WP5 	Parameters from structural and chemical characterizations at the local scale, including e.g., spectra (vibrational, absorption, ...)	Model refinement, cross-analysis, validation, and design of simulated experiments	Tarball files can be created with post-processed data and parameters of interest	.tar.gz (with databases and reproducibility data sheets)	MB

Data delivered from WP2

WP	What	Usable for	Suggested		
			type	format	size
To WP5 	Computational predictions for bulk/interfacial structures, "chemical environments", spectra, transport properties, or any experimental characterization requiring atomistic or electronic interpretation	Guiding the characterization effort and supporting the interpretation of the results	Tarball files can be created with post-processed data and parameters of interest	.tar.gz (with databases and reproducibility data sheets)	TB



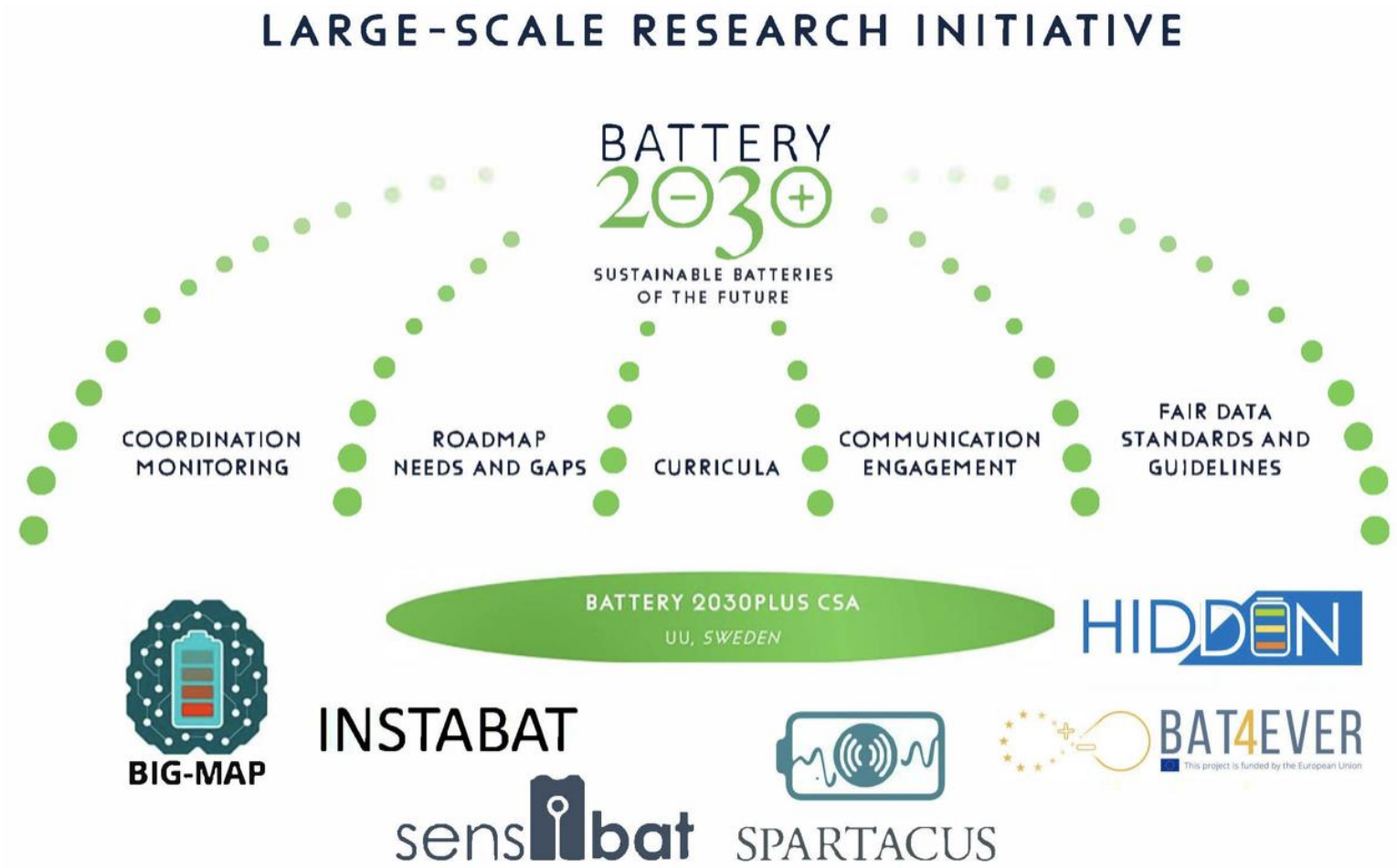
BIG-MAP DMP and the BATTERY 2030+ Consortium



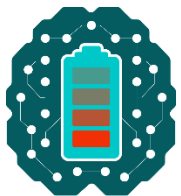
BIG-MAP

LARGE-SCALE RESEARCH INITIATIVE

AMBITION: connect data across the entire European Battery Landscape



Endorse the Memorandum:
<https://battery2030.eu/research/research-data-management-/>



BIG-MAP

Connect Data between BATTERY 2030+ Projects



BIG-MAP

Type	Description	Data sets	Topic	Scale	Component
Electronic Structure: WFT, DFT, QMC	Structures, energy-related data, wave functions & electronic properties, ab-initio molecular dynamics (AIMD) trajectories, different types of spectra	Data generated by different tools: Engines (molecular): GAUSSIAN, ORCA, MOLPRO, TURBOMOLE, NWChem, QChem, ADF, PSI4, MRCC, NECI Engines (periodic): CP2K, VASP, QUANTUM ESPRESSO, Yambo, Castep, GPAW, QuantumATK, Crystal, NECI	Modelling	Atomistic	Electrode, Electrolyte, Interface
ML-FF (representation & regression)	SOAP/ACE + GAP; Parameters	Engines: QUIP & GAP codes	AI		
Alchemical Exploration and Optimization	Property relationships & Compound space search	Engines: QML & APDFT codes	Model		
Atomistic simulations	Atomic trajectories, and associated transport, spectral etc. properties	Engines: LAMMPS, GROMACS, QUIP	Model		
Electrochemical data	Redox properties of electrolytes	Linear sweep voltammetry (LSV), cyclic voltammetry (CV)	Electrolyte		
Spectroscopic data	Local chemical environment	FTIR, Raman	Characterization		
Structural data	Relation between phase transformations of the crystal structures in the electrodes, and the electrochemical activity, i.e. characterising the cells in real time	X-ray diffraction, neutron diffraction	Characterization		

SPARTACUS

Datatype	Description	Data sets	Topic
Specification	Specification for sensors, CMS/BMS and Validation		Specification
Specification	Specifications for Data Preprocessing		Specification
Specification	Specifications for CMS/BMS		Specification
Specification	Specifications for Validation		Specification
Specification	Methodologies for LCA		Specification
Specification	Sensor Specification		Specification
Design	Design layouts		Design
Specification	Sensor Integration document (ORP-EIS)		Specification
Specification	Process Flow Table		Specification
Characterization	Impedance Data		Characterization
Specification	Text Matrix for Aging campaign		Specification