



Energy Modelling
Platform for Europe

EMP-E 2021: Re-Energising Sustainable Transitions in Europe

Energy System Modelling, Methods & Results to
support the European Green Deal

26th to 28th October • online

Open data for energy system modeling: data architectures, semantics, metadata, systems, buses, and licensing

Robbie Morrison • open energy modeling community
28 October 2021 • 14:00–15:30 CEST • 10 minute presentation
Parallel session 9 • Data and model transparency

Release: 01 • Status: Presentation version • Date: 28 October 2021



Opening remark

My generation has left our planet in a terrible state and very few of our underpinning technologies, systems, and policies are remotely fit for purpose.

The IPCC First Assessment Report in 1990 was entirely clear on the need for deep and rapid decarbonization.

Let's hope we can now understand, articulate, and transition sufficiently quickly to a world that will provide those who follow with a viable future.

The first Fridays for Future school strike held 14 December 2018, Berlin, Germany

Photograph: Robbie Morrison



Abstract

In an ideal world, energy system modeling undertaken in support of public policy would be transparent and reproducible by independent researchers. The data used would be legally open, of known provenance and quality, and examinable by anyone. The data licenses applied would not create data silos and would permit contributors to be named, should they so wish. This paper accordingly reviews the open licensing of legitimately published non-personal data, relative to legislation in the United States, Germany, and the United Kingdom. The data itself would form part of a common pool under community curation. Data semantics would move from a proliferation of localized schemas to an agreed formal ontology. Metadata would likewise be standardized. New model-agnostic "data systems" offering coherent and complete datasets covering particular geographic regions would be widely shared. Data portals would be augmented by a graph of semantic triples relationships, some referencing individual datasets, to realize linked open data. Smart "data buses" would interface between analysts and these web-based resources. Scenarios would be managed collectively and the outputs from different models compared and dissected. Indeed many of the changes indicated are underway. This paper examines recent developments, their contexts, and the many challenges that lie ahead. □

Some context

Non-personal open data

- discussion limited to **non-personal data** that can be or has been made public legitimately
- interest for openness driven equally by transparency and opportunity
- but the journey will be long and arduous .. open data trails behind open modeling by quite some margin and requires substantially higher levels of cooperation
- data may be seen as a "social" construct

A touchstone definition for open data

Recital 16 from the 2019 open data directive:

"Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose."

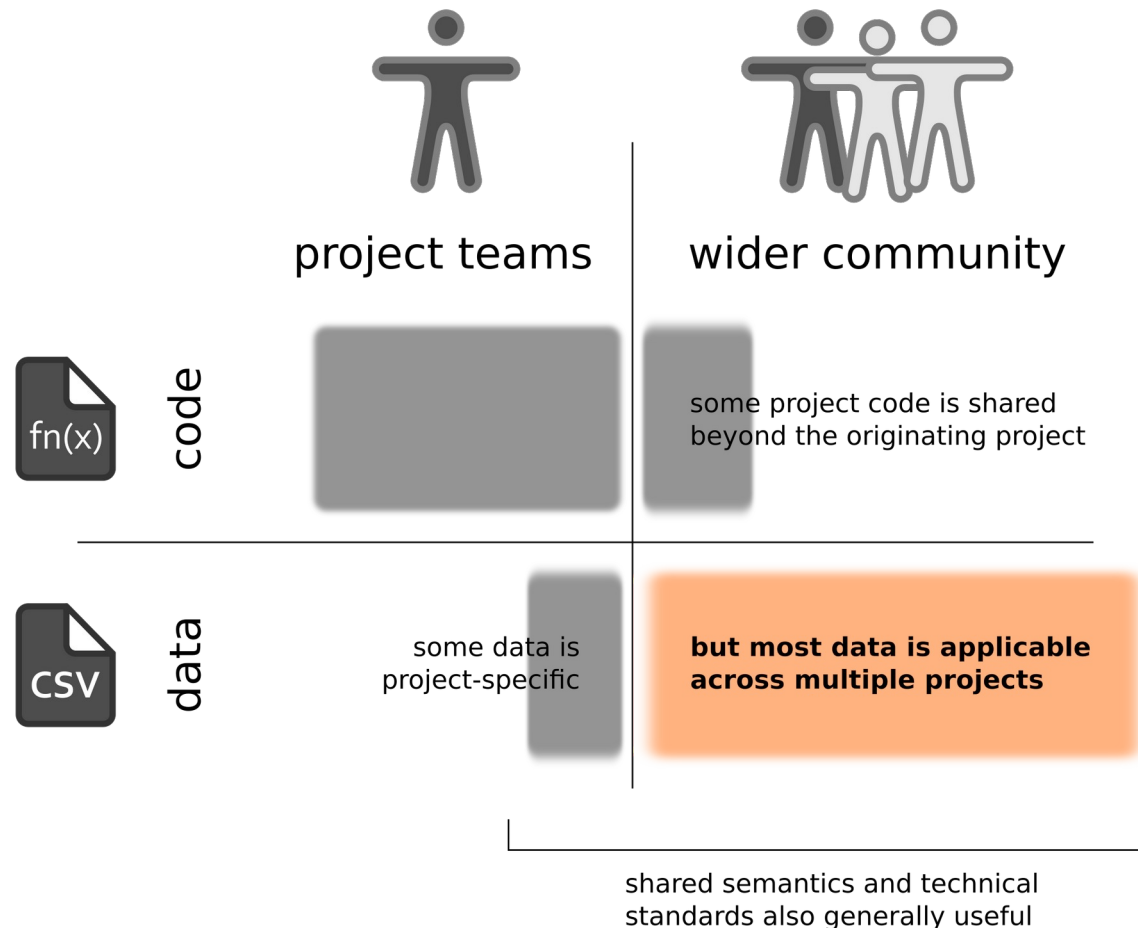
Source: **European Commission** (26 June 2019).

"[Directive \(EU\) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information — PE/28/2019/REV/1](#)"

Official Journal of the European Union. **L 172**: 56–83. The directive entered into force on 16 July 2019. **Recital 16** (page 58) quoted above.

Motivation

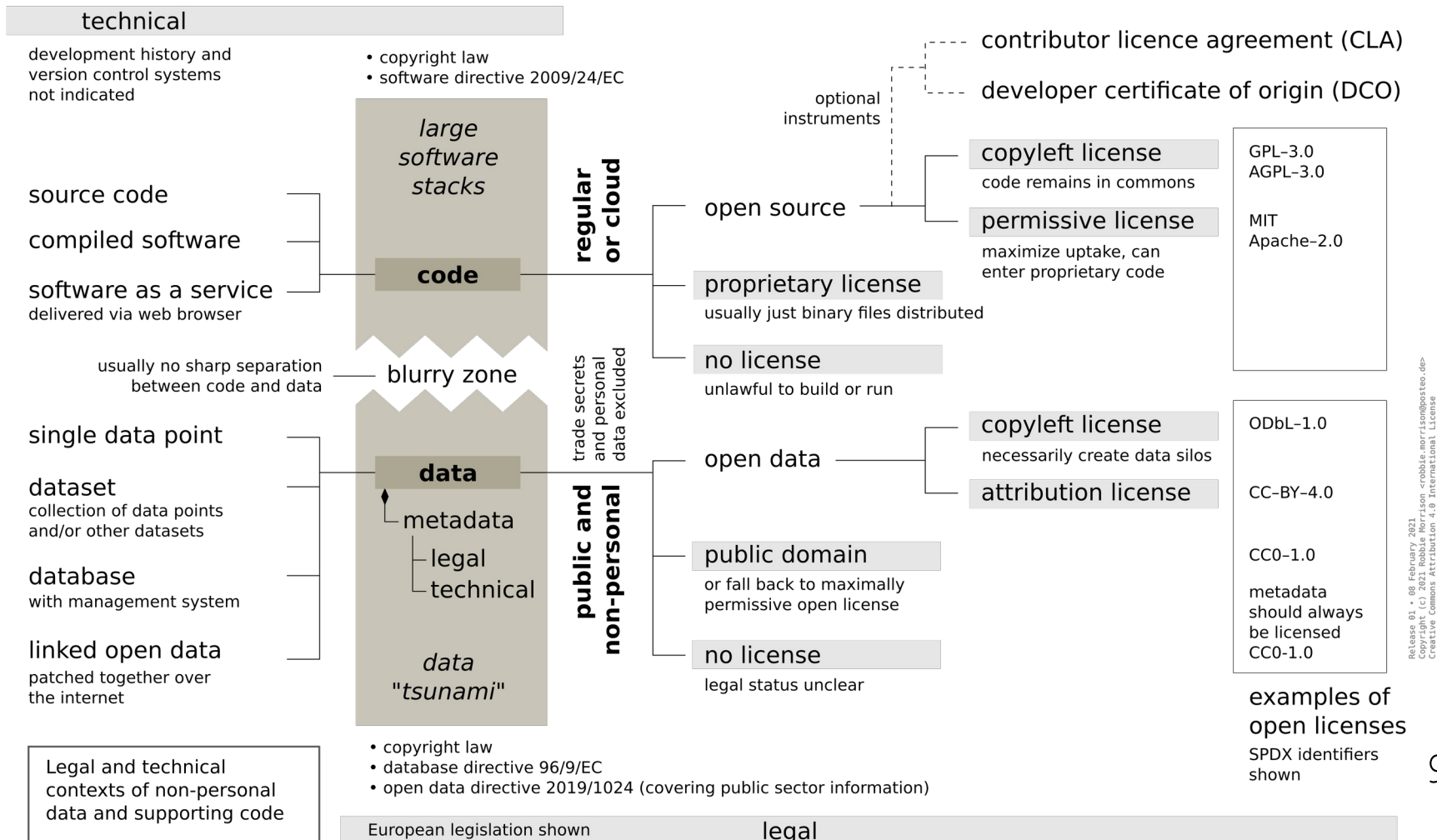
For tackling open data as a wider community



Some themes

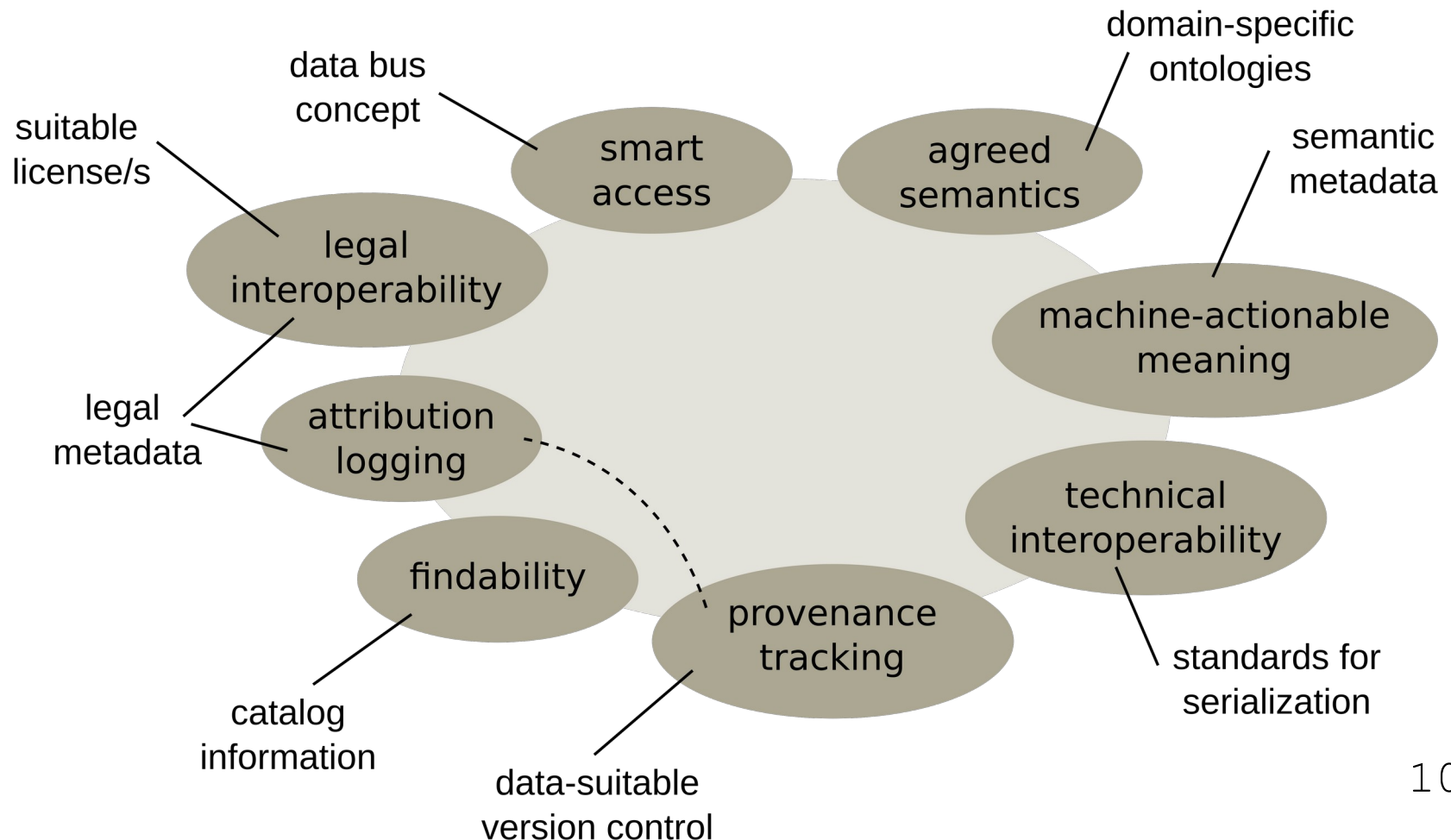
Code/data landscape

From a project viewpoint



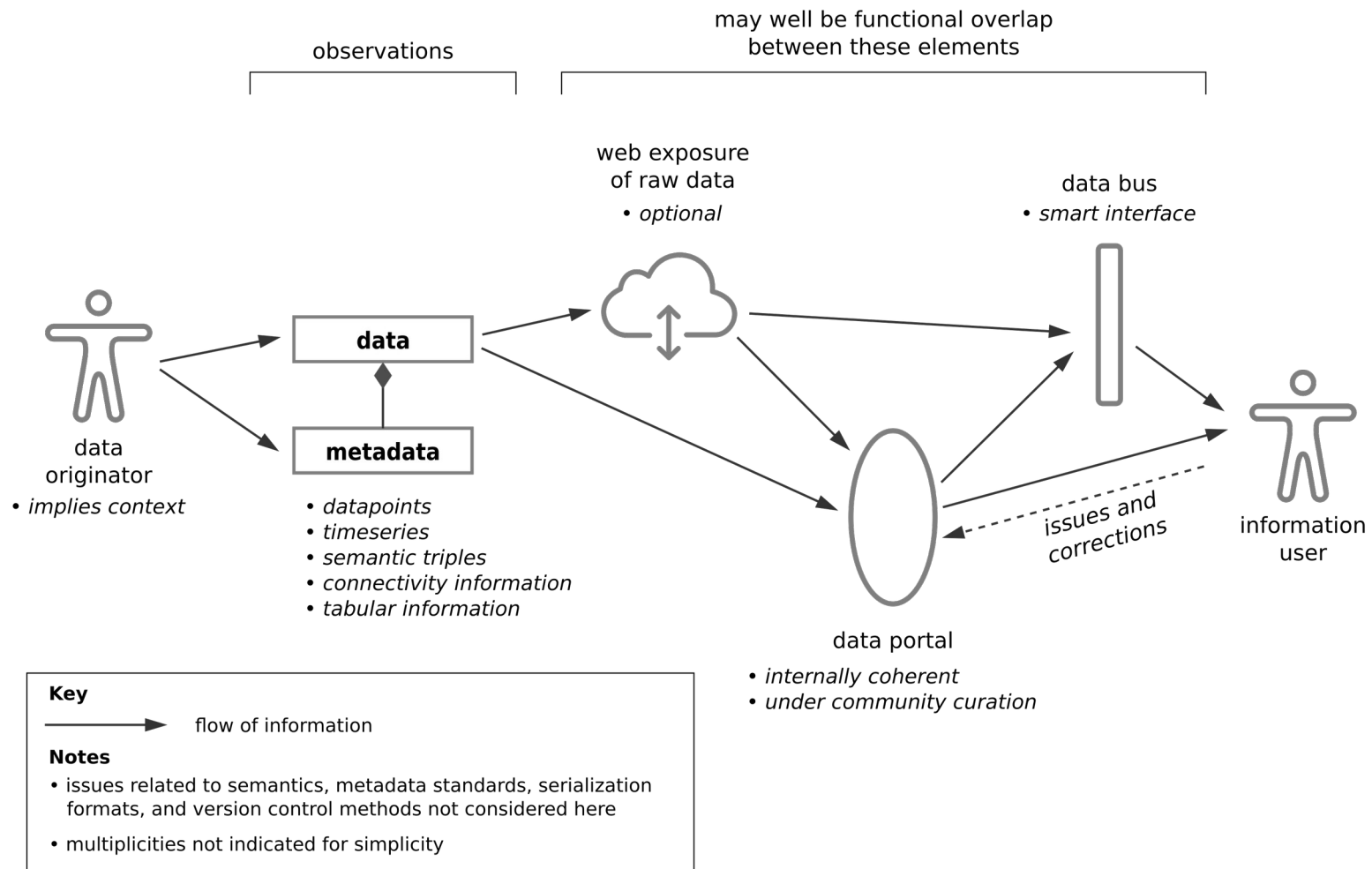
A bundle of interacting themes

From a community perspective



Data flow / architecture

An intentionally simplified schematic



Linked open data

Uncertain role (to me at least)

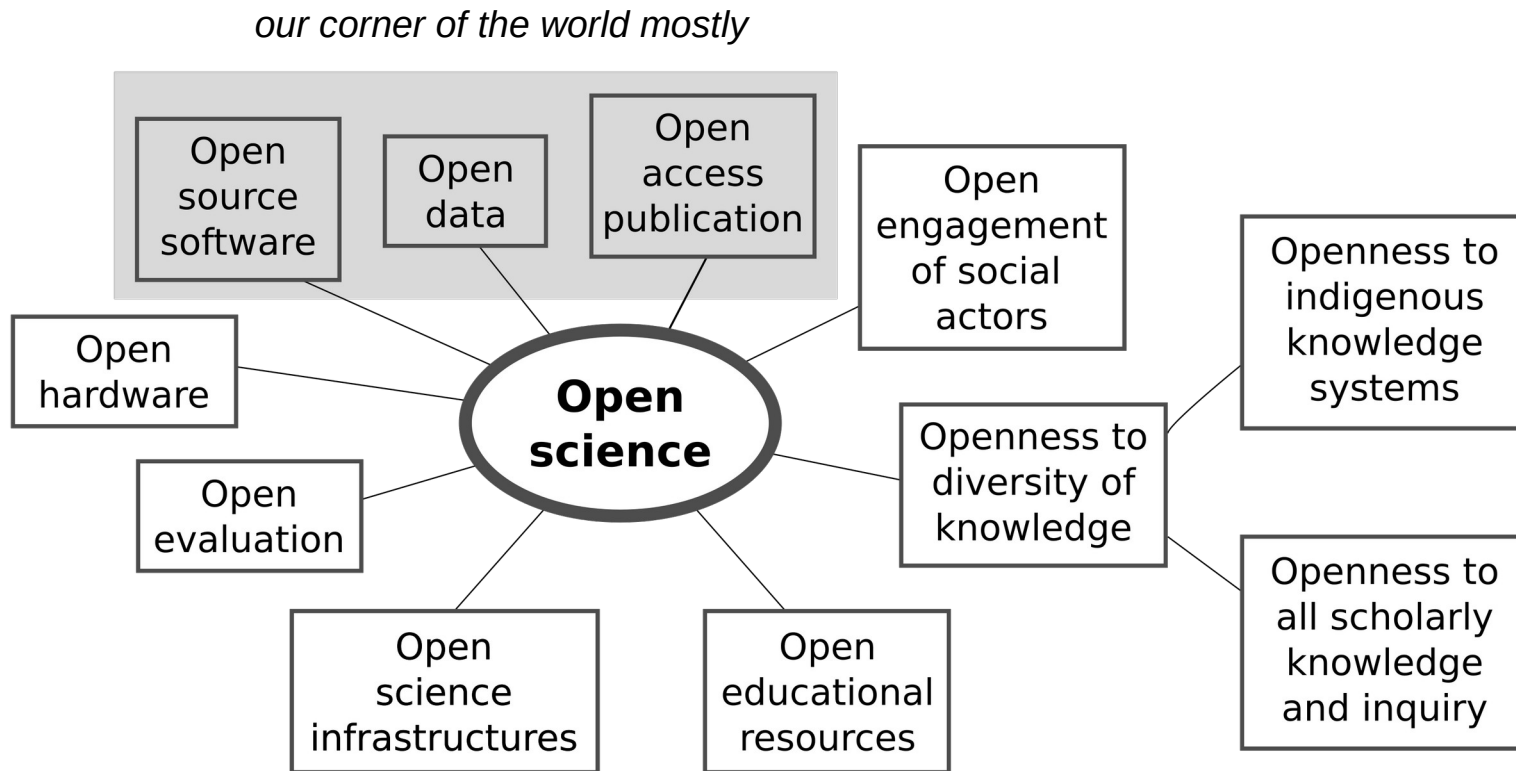
- a central idea is **semantic triples**: two entities joined by a relationship
- these naturally create dynamic **knowledge graphs**
- impact of LOD on energy system analysis is unclear

Refer: Kelly, Jack (18 August 2020). [Linked data for the energy system](#). *Open Climate Fix*. London, United Kingdom. Blog.

Refer: Hitzler, Pascal (February 2021). ["A review of the semantic web field"](#). *Communications of the ACM*. **64** (2): 76–83. ISSN 0001-0782. doi:[10.1145/3397512](#). PDF download available.

Open science

Providing some context



Adapted from: UNESCO draft open science recommendation presented at the Open Science Conference 2021 on 17 February 2021

Official French policy on open science and open data

A recent French plan for open science (referenced below) argues for:

- the development of **high-level open data management plans**
- the establishment of **key canonical datasets**

On the second point (p23):

"Support dataset harmonisation and interoperability to build **large reference datasets for each field**. This involves encouraging dialogue between the scientific domains to facilitate integrated ecosystem approaches. For example, this can be useful when studying interactions between the ocean and the atmosphere, or the continental surface and the atmosphere."

Source: Ministry of Higher Education, Research and Innovation (July 2021).
[Second French Plan for Open Science: Generalising open science in France 2021–2024](#). Paris, France: Ministry of Higher Education, Research and Innovation. Publication date refers to the French language version.

Data semantics

Issues related to meaning

- data semantics are domain-specific
- data semantics are often communicated using formal **ontologies** that cover naming conventions, definitions, and relationships
- earlier in this event, a Joint Research Center (JRC) representative urged that energy system analysts align with **Eurostat** (European Statistical Office) practices

Technical standards

Challenging but this community is technical

- data serialization standards to govern interchange
- metadata standards and cataloging
- version control, provenance tracking, and attribution logging

Sharing of scenarios

Including those covering official policy targets

- should ease some of the modeling burden
- could promote standardized reporting
- could facilitate cross-model comparisons
- should improve confidence in results and interpretations

Proposed EU data act

Speculative comments on my part

- no draft legislation as of yet
- a key objective is improved public interest **business-to-government** (B2G) information flow
- the 2016 trade secrets directive may need to be modified
- unclear to me how B2G ► community data flows might work?
- but with suitable anonymization, could provide a valuable source of representative cost information?

Statutory support for open data is needed

Dedicated legislation would certainly assist



- to my knowledge, absolutely no direct statutory support for open source software and open data anywhere in the world
- open source software and open data sectors based entirely on license texts drafted by civil society
- there is some limited academic work on the theme of statutory support
- I recently argued that the proposed EU data act should traverse this issue:

Refer: Morrison, Robbie (3 September 2021).

Submission on a proposed Data Act for the European Union from the perspective of energy system analysis / 2 — Release 02.
doi:[10.5281/zenodo.5471077](https://doi.org/10.5281/zenodo.5471077). Berlin, Germany. Creative Commons CC-BY-4.0 license.

Global South uptake

Natural technology transfer through open models and open data

- a number of examples now in the modeling area
- regarding data support, the energydata.info project uses CKAN software from Open Knowledge Foundation (OKF)

Some representative projects

OPSD project

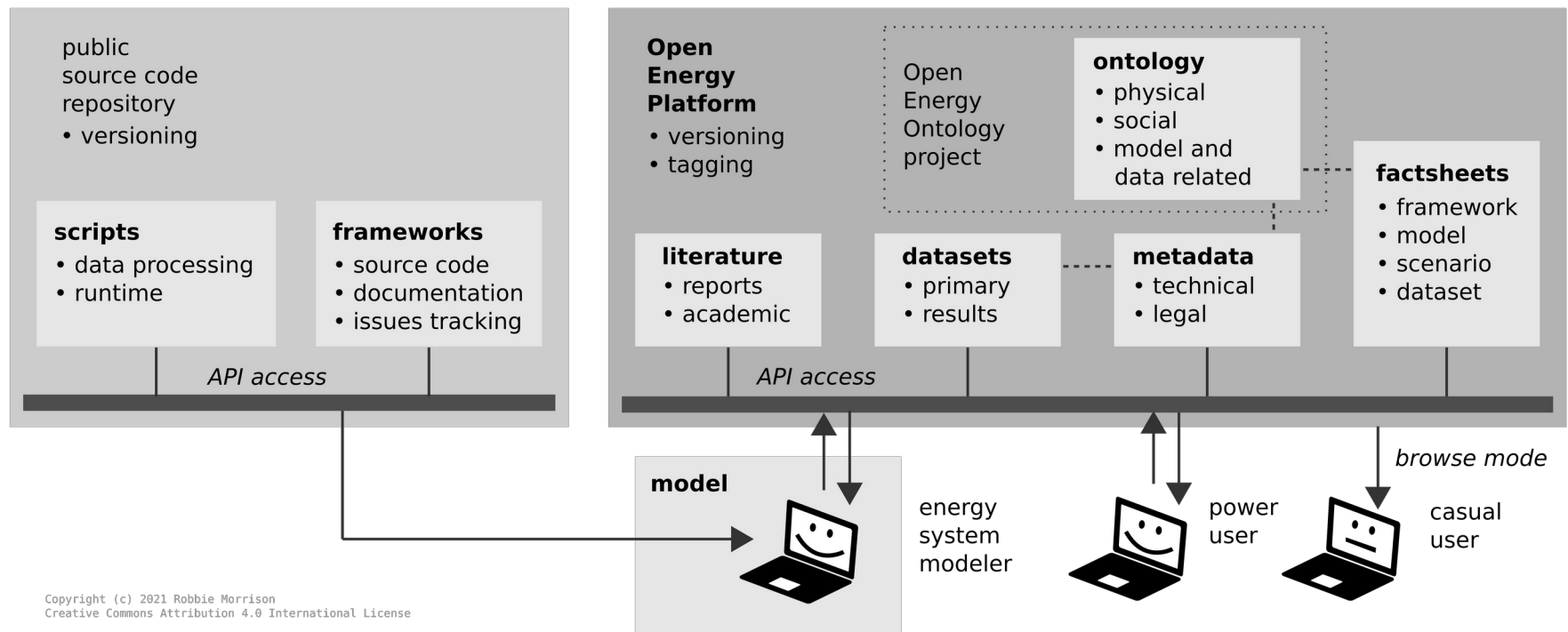
Open Power System Data portal

- scope
 - Germany and further European countries
 - power plant fleets: conventional and renewable
 - time-series: production and load
 - weather data
- process
 - portal launched October 2016
 - extensive curation, all processing scripts on GitHub
 - serves OKF **frictionless data packages** with CSV tabular data and JSON metadata
 - issues remain with open licensing

OKF = Open Knowledge Foundation
CSV = comma-separated variables
JSON = JavaScript object notation

SzenarienDB project

An Open Energy Family initiative



Open Energy Ontology

An open energy modeling community project

- early-stage project
- licensed CC0-1.0
- OvGU, Öko-Institut, Fraunhofer, RLI, FZJ, DLR, Leipzig University
- anchored in computer science
- foundation is the basic formal ontology (BFO) which parents everything

Refer: Booshehri, Meisam, Lukas Emele, Simon Flügel, Hannah Förster, Johannes Frey, Ulrich Frey, Martin Glauer, Janna Hastings, Christian Hofmann, Carsten Hoyer-Klick, Ludwig Hülk, Anna Kleinau, Kevin Knosala, Leander Kotzur, Patrick Kuckertz, Till Mossakowski, Christoph Muschner, Fabian Neuhaus, Michaja Pehl, Martin Robinius, Vera Sehn, and Mirjam Stappel (1 September 2021). "Introducing the Open Energy Ontology: enhancing data interpretation and interfacing in energy systems analysis". *Energy and AI*. 5: 100074. ISSN 2666-5468. doi:[10.1016/j.egyai.2021.100074](https://doi.org/10.1016/j.egyai.2021.100074). Open access.

EERAdata metadata project

Project not limited to open data

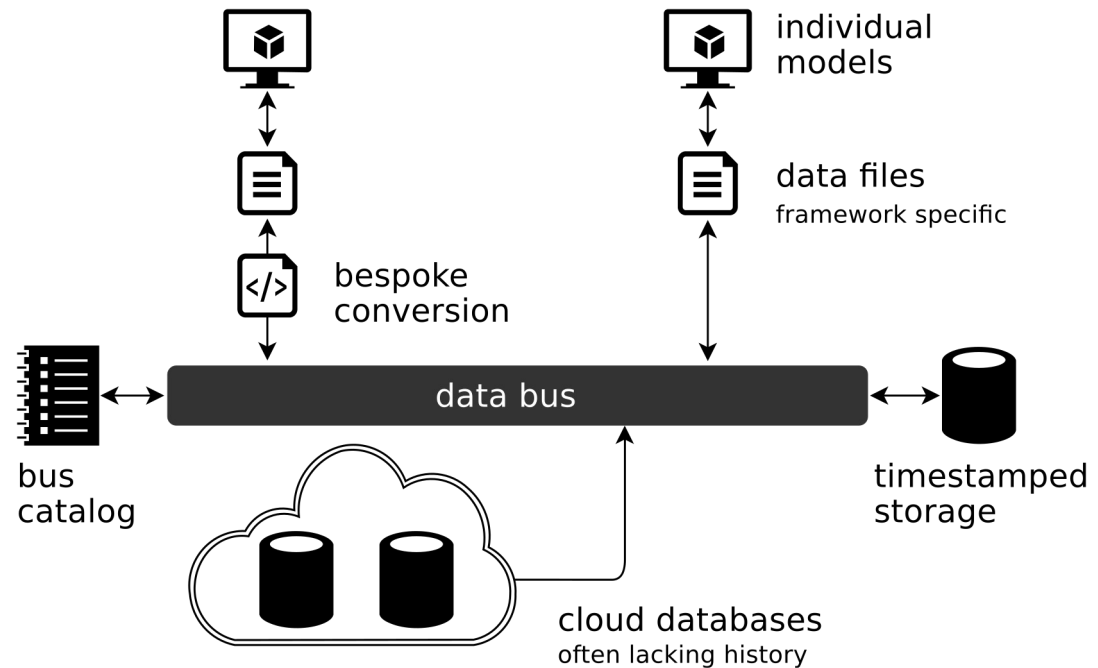
- early-stage project
- EERA is the European Energy Research Alliance
- metadata
 - informed by FAIR data principles
 - machine-actionable

Refer: Wierling, August, Valeria Jana Schwanitz, Sebnem Altinci, Maria Balazińska, Michael J Barber, Mehmet Efe Biresselioglu, Christopher Burger-Scheidlin, Massimo Celino, Muhittin Hakan Demir, Richard Dennis, Nicolas Dintzner, Adel el Gammal, Carlos M Fernández-Peruchena, Winston Gilcrease, Pawel Gladysz, Carsten Hoyer-Klick, Kevin Josho, Mariusz Kruczek, David Lacroix, Malgorzata Markowska, Rafael Mayo-García, Robbie Morrison, Manfred Paier, Giuseppe Peronato, and Mahendranath Ramakrishnan (15 October 2021). "Advancing FAIR metadata standards for low carbon energy research". *Energies*. 4 (20): 6692. ISSN 1996-1073. doi:[10.3390/en14206692](https://doi.org/10.3390/en14206692). Open access.

DBpedia Databus project

Smart interface to cloud databases

project provides secure
and smart gateway to
changeable cloud
databases



Refer: Hellmann, Sebastian (29 September 2019).

DBpedia's Databus and strategic initiative to facilitate "1 billion derived knowledge graphs by and for consumers" until 2025

Leipzig, Germany: DBpedia. PDF.

Closure

My apologies

- I will almost certainly have missed some projects, particularly at the European level ..
- so if under Creative Commons CC-BY-4.0 licensing ..
- I would be grateful to be emailed the details:

`robbie.morrison@posteo.de`

Concluding thoughts

- dedicated EU **open data policy** centered on open science would assist
- **legislative support** for open data would assist
- regarding **public licensing**:
 - data under **CC-BY-4.0**
 - metadata under **CC0-1.0**
- much work for the open energy modeling community to do



Some legal matters

Open data licenses

My personal picks

Creative Commons CC-BY-4.0



- introduced 25 November 2013
- first data-capable license (that deals with the EU Database Directive 96/9/EC)
- requires **attribution and attribution tracking**
- material may be modified or mixed and licensed under more restrictive terms but the attribution requirement must remain

Creative Commons CC0-1.0



- public domain dedication
- falls back to maximally permissive license in civil law jurisdictions like those in Europe
- no legal obligations for users
- metadata should always be licensed CC0-1.0

in most case, open data licenses do not provide **permission** but rather offer **certainty**

Legal metadata for data

Another raft of challenges

- the **Dublin Core** for metadata includes fields for the **creator** and for the **rights** held, the DCMI extension includes a **provenance** field, and all are defined as Resource Description Framework (RDF) properties
- that may work for relatively static works like manuscripts but not for datasets that are widely combined and continually and subtly altered
- some data-specific **challenges** therefore:
 - data versioning
 - data license compliance and attribution tracking
 - the recording of provenance more generally
- these issues have been traversed for software but the solutions developed for code are unlikely to transfer readily to data

Joint authorship

Legal treatment unclear for both data and code

- commonly believed that each contributor retains the copyright to their individually identifiable contributions
- for alternative view based on United States law: Chestek, Pamela S (2017). ["A theory of joint authorship for free and open source software projects"](#). *Colorado Technology Law Journal*. **16**: 285–326. Open access.
- the issue is material in relation to license compliance