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

Robbie Morrison <robbie.morrison@posteo.de>

15 June 2021

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 challenges that lie ahead.

▢