Open energy system modelers call for CC-BY-4.0 data licensing wherever possible — our world needs comprehensive usable and reusable data to transition to net-zero

18 January 2023

To whom it may concern

Audience

This open letter is directed toward organizations that variously collect, generate, collate, curate, revise, package, summarize, catalog, and publish or republish structured information describing energy systems — with a particular focus on those datasets suitable for **integrated system modeling** and allied numerical analysis.

Such **organizations** would include public bodies, multilateral agencies, civil society organizations, universities and research institutes, commercial entities, open source modeling projects, and citizen science projects.

The measures presented are not tied to particular **national jurisdictions**, although some legal systems offer less natural protection for data than other systems — with the United States and the United Kingdom providing respective examples along the spectrum from lax to strict.

Context

National energy transitions require a sequence of substantial, robust, and timely investment decisions. Good systems planning can therefore assist. It can help allocate or exclude resources, identify the more rapid and feasible pathways, provide cost estimates, and engage the interested public. The modeling software is only part of that story — far more important is **good quality systems information**.

At present, much of the data needed is either retained, paywalled, public but not licensed for reuse, or offered under bespoke terms which necessarily create legal silos. The upshot is that much of the information needed for independent analysis remains necessarily piecemeal and incomplete and **wide-ranging curation and uptake** is prevented because the various terms-of-use provided are either too restrictive or legally immiscible.

Moreover, some civil society and public organizations publish their datasets under **non-commercial terms** to allow allied data sales. Such terms can be problematic for university researchers in receipt of third-party funding who may not clearly class as non-commercial in this context. The associated **private data sales** also prevent consultancies and corporations from sharing their populated systems models for wider benefit and cooperation.

The relatively easy and proven solution to these problems is **open data licensing**. This letter asks that the Creative Commons CC-BY-4.0 license be applied, wherever possible, to **primary data of public interest** within the energy domain.

This particular solution mostly requires no funding (although some sales revenues may need compensation), no change in legislation, and no third-party consent (presuming the data provider generated the requisite information) — just simply a commitment to ideal of **common information** and a desire to share and receive. And despite some costs, this solution is likely to provide **very substantial benefits** to systems analysts and beyond as nations and regions race to decarbonize.

The following section — signed mostly by independent energy systems analysts — fleshes out the details of and rationale for the solution being advocated.

Our requests

We, the undersigned, request that, wherever possible, **structured data** made public for energy system analysis be licensed using the Creative Commons CC-BY-4.0 license. With the accompanying metadata — which captures context, provenance, and legal status — released instead under the Creative Commons CC0-1.0 public domain waiver to facilitate cataloging and discovery.

Both the aforementioned legal instruments provide for full use and reuse in original or modified form, but CC-BY-4.0 additionally requires that **attribution be recorded and tracked**. Attribution aids provenance and should not be considered a burden for robust and repeatable numerical analysis.

We additionally ask that any overarching formal or informal **data standards** that apply — which may include vocabularies, ontologies, data models, reference architectures, and database schemas — be placed under any of the Creative Commons open licenses, including CC-BY-SA-4.0, so that the concepts and specifications provided can be applied to both numerical information and system modeling frameworks without the legal risk of less-open or unlicensed intellectual property rights transferring across to duly informed datasets and codebases.

This letter applies only to material that has been or can be **made public** — thereby taking personally identifiable information (PII), trade secrets, and other forms of personal and commercial confidential information off the table.

Two fundamental drivers lie behind these recommendations. First, applying such licenses will provide **legal certainty** to the users and reusers of such data, as the legal status of the information in question is often far from clear. And second, the information so licensed will be **widely legally miscible** and thus able to contribute to the development of a knowledge commons that spans the energy systems domain.

A user-led knowledge commons is especially important because systems modelers need and seek **coherent and complete structured data**. And open data standards can play a key role in ensuring some level of **semantic and technical consistency** across both models and data.

Other public licenses and waivers that are **inbound compatible** with CC-BY-4.0 are also acceptable — with the proviso that robust independent analysis supports such claims.

More information can be found on the Open Energy Modeling Initiative discussion forum under tag "ccyby2023": https://forum.openmod.org/tag/ccby2023

The signatories to this open letter are attached as a table, together with their institutional affiliations when also provided.

Yours sincerely, the undersigned.

☐

filename: open-energy-modeling-open-data-letter-early-2023.05.md|pdf