Workshop Summary: Data Schema Approaches for Grid Planning Model Interoperability

April 9, 2026 | Convened by GPST, BE-GRIDS (OpenGrid), and Sequoia Climate Foundation (on behalf of Global Grids Catalyst) - Virtual

Context

More than 40 participants from NLR, Princeton, TU Berlin, Encoord, ENTSO-E, TransitionZero, Open Energy Transition, Nodal Tools, Sylvan, Google Tapestry, LF Energy, EPRI, and philanthropy organizations (BE-GRIDS, Sequoia Climate Foundation, Rockefeller Foundation, and Global Energy Alliance) joined this workshop to build shared understanding of data schema approaches for grid planning model interoperability. Seven schema teams presented (CIM/CGMES, PyPSA, CESAME, SAInt, Sienna/GridDB, GenX, and GDM), followed by a cross-schema synthesis and a 60-minute facilitated discussion on a Miro whiteboard. Key findings from the discussion are summarized below.

Introductory remarks described the concepts of *"vertical"* (across modeling scales and domains) and *"horizontal"* (among tools for the same scale/domain) interoperability, with general agreement that the goal is to enable both. They also introduced the concept of *3 layers of interoperability*: Level 1: Flat database with relations, Level 2: Additional language-specific structure, and Level 3: tool-specific data representations. The later (v1) discussion was prompted to focus on Level 1 (Flat database with relations), since it seems the most likely level for broader interoperability. Finally, the concept of *a family of schemas* was described as potentially avoiding the pitfalls of trying to fit everything into a single schema. Such a family could potentially split by infrastructure scale (e.g. distribution, bulk grid, multi-energy) while being inclusive of relevant analysis types within each (i.e. supporting investment planning, operational simulation, and engineering assessment). This approach was referenced, sometimes indirectly, during the later discussion including the idea of separating distribution and bulk schemas, among others.

What Does "Good Enough for v1" Look Like?

Design Principles

Participants converged on several design principles for a v1 common schema:

- **Language-agnostic:** This includes the core storage (e.g. potentially a relational database), interfaces for access, and any serialization format;
- **Grounded in real-world datasets**, not academic test cases: In particular, initial datasets need to include examples at realistic (large) scales and be capable of capturing the quirks and complexities found in actual systems. This includes validation profiles tailored to specific use cases;

- **Able to manage multiple types of data:** This includes structural/stable data (including network topology and connected assets), time series data, and scenario data—and to do so in appropriate structures/formats for each;
- **Independent of modeling approach:** e.g. able to support optimization, simulation, AI, or otherwise.

These principles represent an aspirational ideal; in practice, some trade-offs and compromises may be necessary to achieve a workable v1 implementation.

Essential Initial Scope

The core initial schema effort would focus on the bulk grid with a need to support capacity expansion, production cost modeling (both nodal and zonal), and AC balanced (positive sequence) power flow; with some interest also in resource adequacy. The schema would enable seamless translation of network assets across these domains. Since this implies a base dataset at the nodal level, this implies support for aggregation to zonal levels. In parallel, there could be a separate effort around distribution data.

While it was remarked that "capturing network assets (e.g. generators, lines, etc.) would be a really good start" a number of additional needs emerged in discussion, including: scenario management; ability to capture planning/policy constraints, unit startups, operating reserves; ability to capture HVDC including multi-terminal; and representation of structural links between market, grid, demand, and climate data. Of particular importance was the ability to manage time series data well, including support for aligned formats like Parquet; multi-dimensional support to capture forecasts explicitly and enable alignment between weather, load, and resources; and support for time series relations such as planning year connections to operations timeseries. Similarly, the discussion highlighted that the ability to capture so many technical parameters may also vary as a function of time, such as different heat rate/efficiencies for different seasons/weather.

Items deferred to future phases included dynamics, integration between distribution and transmission, and GUI tooling. Items that were identified as explicitly out of scope were: gas network physics and EMT-scale inverter control logic (due to proprietary IP concerns). There was also agreement that components could be added through a governed extension process.

Challenges to Overcome

Three challenges emerged: (1) managing data resolution and aggregation (capacity expansion models use zonal resolution while many production cost efforts and all power flow requires nodal resolution); (2) recognizing that how a user expresses their modeling intention (in current datasets) may differ from their modeling choices (for example, the use of proxies for HVDC or synchronous condensers), and (3) engaging data providers, whose participation would significantly boost adoption. These challenges will be important for the schema initiative to navigate.

Key Open Questions

The topic of multi-sector coupling generated productive discussion with varying perspectives. Some participants stressed the importance of addressing it in v1 due to its fundamental structural implications, while others advocated for a streamlined initial release with extensibility designed in from the start. The group identified a promising approach, aligned with a family of schemas, where the v1 effort may focus on the (bulk) electric power grid but supporting flexible and universally applicable transformation processes for sector coupling with the goal to build more sector-specific schema elements later.

Another open question centered around the approach for managing units. There was general acceptance of including a year when defining monetary values. For other parameters, some advocated for storing the units alongside the data, while others proposed a common, schema-specified unit for the storing in the database, with conversion happening externally (e.g. as part of tool interfaces).

Adoption Requirements

Participants identified critical factors for successful adoption:

- Completeness is non-negotiable: 70% coverage is a non-starter for power systems that require precision.
- The schema must span capacity expansion, production cost modeling, and AC power flow. Dynamics were flagged though noted as possible future phase requirement.
- Governance must be clear, responsive, and formalized, with consensus mechanisms, highlighting the need for governance structures that ensure continuity independent of any single funder.
- Developer and end-user experience must be a priority; complexity lowers adoption.
- The schema must capture leading proprietary tools (PLEXOS, PowerFactory), at least through conversion support, to provide adoption incentive.
- Much real-world work is raw data to model, not model-to-model conversions, so the schema must address practical workflows.
- The schema must explicitly address how common modeling proxies (e.g., HVDC as proxy objects, synchronous condensers as negative generators) will translate, as these workarounds are prevalent in current practice and their incompatibility could hinder adoption.

Participation and Broader Value

Every participant expressed willingness to contribute and agreed that a common data schema is high-value effort that would ease data sharing, enable model switching and integrated workflows, provide open certification, increase tool competition, and eliminate redundant effort.