EMI-2020
CODERS: Introducing an open access dataset for decarbonizing Canada's energy system

Hendriks, R.M., Jurasz, J., Cusi, T., Aldana, D., Monroe, J., Kiviluoma, J., McPherson, M.*

Abstract

Data is at the core of energy systems modelling. Energy systems modelling is at the heart of developing decarbonization pathways and ultimately implementing mitigating actions. Inclusive access to transparent datasets that are available to users in a timely way has been a key barrier to modelling and analysis efforts across Canada. This report introduces CODERS, the Canadian Open-source Database for Energy Research and Systems Modelling, to fill this gap. CODERS contains the data that has been collected by the Sustainable Energy Systems Integration and Transitions (SESIT) group to populate our suite of energy systems models that span sectors (power, transport, buildings), scales (municipal, provincial, federal) and vectors (electricity, heat, water). CODERS is currently hosted on a platform that is only available to SESIT collaborators, but will be made open-access when the licensing process and host portal are finalized. When complete, we hope that CODERS will close a significant gap in the Canadian energy systems modelling workflow and contribute to an ongoing national energy transitions dialogue.

* Corresponding author. Email address: mmcpherson@uvic.ca

Table of Contents

List of Tables

List of Figures

Abbreviations

API	-	application programming interface
CCEI	-	Canadian Centre for Energy Information
CEMS	-	Continuous Emissions Monitoring System
CER	-	Canada Energy Regulator
CODERS	-	Canadian Open-source Database for Energy Research and Systems Modelling
DOE	-	US Department of Energy
ECCC	-	Environment and Climate Change Canada
EIA	-	US Energy Information Administration
EPA	-	US Environmental Protection Agency
EU	-	European Union
FERC	-	Federal Energy Regulatory Commission
GUI	-	graphical user interface
ISO	-	independent system operators
LNG	-	liquified natural gas
NGL	-	natural gas liquid
NRCan	-	Natural Resources Canada
NREL	-	National Renewable Energy Laboratory
OEDI	-	Open Energy Data Initiative
PUDL	-	Public Utility Data Liberation Project
REST	-	representational state transfer
RPKM	-	University of Victoria Research Partnerships and Knowledge Mobilization Unit
RPP	-	regulated price plan
RTED	-	Real-time electricity data
SESIT	-	Sustainable Energy Systems Integration & Transitions Group
SQL	-	Structured Query Language
StatCan	-	Statistics Canada
US	-	United States

1 Introduction

Transforming our energy system is a cornerstone of decarbonizing our economy. Implementing this transition spans several systems (power, transportation, buildings), multiple stakeholders (policymakers, utilities, NGOs, researchers, citizens), and many institutional platforms (markets, multi-jurisdictional policy, international trade). Given the complexity of the energy system and these interdependencies, decision-makers must make use of modelling platforms to illuminate trade-offs and navigate choices. The result has been a proliferation of models and modelling approaches aimed at answering important questions related to future energy systems (Foley, Gallachóir, Hur, Baldick, & McKeogh, 2010) (Ringkjøb, Haugan, & Solbrekke, 2018). In particular, energy planning, which is the domain of policy makers, has a long-standing history of being supported by a large variety of models for demand forecasting, expansion planning, demand management, and integrated resources planning (Doukas, Patlitzianas, Kagiannas, & Psarras, 2008). In part, policy makers build their decisions on models to be able to claim that these decisions are not subject to biases but rather result from technical analyses, referred to by Porter (1995) as the "pursuit of objectivity" (MacGillivray B.H & Richards, 2015). However, as explained by Horschig and Thrän (2017) there are no available standards to which the model can be compared and its quality assessed.

In Canada, modelling tools have been used to support decarbonization policy, including the memorandum of understanding between the British Columbia and Canadian governments on the electrification of the gas sector (Government of Canada, 2019a), the Atlantic Canada Clean Energy Growth Strategy (Government of Canada, 2019a), the Regional Electricity Cooperation and Strategic Infrastructure Initiative (RECSI) (Natural Resources Canada, 2021) and the RECSI internal evaluation (Government of Canada, 2019b). The United States Mid-Century Strategy for Deep Decarbonization (The Office of the White House, 2016) relied heavily on analysis of quantitative energy methods, including a 24 model intercomparison study (Clarke, et al., 2014), and the EnergyPATHWAYS modeling tool for deep decarbonization assessment (Williams, Haley, Kahrl, & Moore, 2014); the latter was also used by the State of Washington to develop pathways for strengthening emission limits while growing its economy (Washington Governor Jay Inslee, 2016). In the European Union (EU), a suite of interlinked models supports the European Commission’s impact assessment and analysis of policy options (European Commission, 2021a). For example, the current models have been used in the Commission's climate policy impact assessments (European Commission, 2021b). Furthermore, the EU organizes the Energy Modelling Platform conference where “modellers meet decision-makers” with an objective of narrowing the gap between scientific modellers and policy makers at all levels with the help of Commission representatives (Energy Modelling Platform For Europe, 2020). Initiatives such as these clearly illustrate decision-makers’ interests in energy system planning models and their outputs.

However, barriers across the modelling-decision-maker interface remain, acting as one of the critical impediments to accelerating deep decarbonization. A key barrier pertains to information flow across institutional, disciplinary and regulatory boundaries that has traditionally been slow and opaque. This is no longer acceptable given the urgency to prevent and mitigate climate change. Overcoming this barrier requires far-reaching efforts and processes that define a new approach to navigating the modelling - decision-maker - public interface. An important aspect of this is the fact that the models and their input data lack transparency and accessibility or, in some cases, are simply unavailable. In particular, the accessibility of electricity data in Canada is limited and disjointed when compared to the United States and Europe (Leach, Rivers, & Shaffer, 2020). The Energy Information Administration (EIA) in the US publishes electricity data using standardized metrics at the scale of balancing authorities (of which there are 71 in the US) (U.S. Energy Information Administration, 2021). Similarly, ENTSO collects and distributes supply and demand data real time for each country in the European Union (ENTSOE, 2021). Electricity data in Canada, on the other hand, are published at the provincial level, while the suite of data published and its spatial and temporal formatting are often inconsistent between provinces (Leach, Rivers, & Shaffer, 2020).

The result is substantial data gaps that leave modellers with inadequate resources to perform in-depth and timely analyses of Canada's low-carbon energy transition, which in turn frustrates the efforts of policy-makers and deprives the public of complete information (Leach, Rivers, & Shaffer, 2020). The reasons for such data gaps are manifold, including individual, organizational, commercial and legal requirements that hinder the development of open access databases and models (Pfenninger, DeCarolis, Hirth, Quoilin, & Staffell, 2017). Hirth (2020) elaborates on the legal aspect, explaining how researchers often infringe upon the intellectual rights of data holders, due to the unclear legal status of many energy systems databases, an important issue that we return to later in this report. As a result, we lack a common and standardized energy dataset for Canada. Instead, individual institutions within Canada have developed their own datasets and tools, leading to overlapping and wasted effort as well as significant delays to project and policy implementation timelines. Furthermore, the lack of data and model openness often leads to unnecessary debate, wrongful conclusions, errors repetition and errors propagation (e.g., as observed in the case of land availability for renewable sources in Europe, (Ryberg, Robinius, & Stolten, 2018)).

With the recent shift towards open data and open source tools, institutional barriers preventing effective collaboration are now collapsing. At the same time, the need for effective collaboration tools is becoming apparent. The growing literature regarding the merits of open data and models points to improving the quality of science, enabling collaboration between investigation and policymaking, improving productivity, and fostering societal trust and debate (Pfenninger, DeCarolis, Hirth, Quoilin, & Staffell, 2017). The current project builds on this momentum by creating a standardized suite of data and data processing tools that enable more timely and efficient execution of energy systems modelling and research. The database is built with a standardized and common structure across all 10 Canadian provinces, and is designed to interface with a range of energy systems models. Furthermore, it is structured to be flexible, so that data can be added as it becomes available, removed when no longer relevant, or modified as circumstances change. By assembling such a database and preparing to make it open source, we seek to support the development of *accessible* models and the production of *useful* analyses that depend on high quality, accessible and transparent data inputs.

This report is organized as follows. The remainder of the introduction reviews the data landscape in Canada (Section 1.1) as well as Europe and the US (Section 1.2). This review is synthesized into a list of core attributes that define an effective database (Section 1.3), which then guides our database development work. Next, we turn to explaining the data collection process (Section 2.1), introducing the overall database structure (Section 2.2), and describing the database licensing process (Section 2.3). Our results demonstrate how the database can be accessed within the broader modelling landscape (Section 3.1), followed by a discussion of the database limitations (Section 3.2) and envisioned future work (Section 3.3). Finally, we close with a vision for the database as a key pillar for building a national modelling platform and supporting an ongoing dialogue (Section 4).

1.1 Database landscape in Canada

While there is no single reliable and standardized source for electricity data that would support electricity systems or integrated energy systems modelling in Canada, there are several databases that host data pertaining to various aspects of Canada's energy system. Statistics Canada (StatCan) publishes data on electricity supply, demand and pricing data (Statistic Canada, 2021), while the Canadian Energy Regulator (CER) tracks and publishes electricity import and export data (Canada Energy Regulator, 2021a). Though standardized and high-quality, these data are of insufficient scope and granularity to support existing and evolving electricity systems models necessary to informing grid decarbonization policy imperatives. Data of sufficient scope and granularity is collected and maintained to greater or lesser degrees by provincial utilities and independent electricity system operators (ISOs) across Canada. However, typically there is no strategic or standardized approach to this data collection or provision. For example, hourly electricity demand data are only publicly available at or near real time for BC, Alberta, Ontario, Québec, Nova Scotia and New Brunswick, but not the remaining provinces (Leach, Rivers, &

Shaffer, 2020). Data at the plant-level on the supply side of electricity systems is even more limited and difficult to obtain (Leach, Rivers, & Shaffer, 2020). High-frequency electricity supply data are only available for participating facilities in the electricity markets in Ontario and Alberta, while the other provinces publish supply data with monthly or annual frequency (Leach, Rivers, & Shaffer, 2020). Oftentimes, data is made available by utilities in response to intervener requests during regulatory proceedings or provided for a fee by system operators in response to requests from market participants and stakeholders (AESO, 2021).

In response, several initiatives, led by a variety of institutions, have undertaken efforts to fill gaps in the Canadian energy data landscape. The major efforts of federal departments and agencies are described below.

Canadian Centre for Energy Information (CCEI) - Statistics Canada

The Canadian Centre for Energy Information (CCEI) was recently launched by StatCan with an overall investment of $15 million over 5 years (NRCan, 2020). Developed by Natural Resources Canada (NRCan) and StatCan in consultation with Environment and Climate Change Canada (ECCC) and the Canada Energy Regulator (CER), the CCEI aims to be an independent, one-stop shop for comprehensive energy data and expert analysis, providing Canadians with access to independent and credible information and expertise. The CCEI's initial priority is to create a modern, user-friendly website that collects, summarizes, and synthesizes energy data primarily from StatCan sources (Government of Canada, 2021). This includes both energy production and consumption data, along with analysis from federal departments and agencies, provincial ministries and the private sector. The CCEI intends to expand the scope of its offerings in the medium term with forward-looking forecasts of energy supply and demand, which although shared on the CCEI website, would be external to StatCan and CCEI (Government of Canada, 2021). By consolidating data within the CCEI, StatCan seeks to provide a comprehensive inventory of credible data sources, assist in the development of standardized concepts/metadata and provide a focused lens to better identify data gaps (Government of Canada, 2021).

Real-time electricity data (RTED) – NRCan, CER, Statistics Canada (CCEI)

The real-time electricity data (RTED) dashboard is being developed by NRCan, the CER, and StatCan to provide granular data on near-real time electricity systems operations (Préfontaine, 2021). In its first phase, RTED compiled and displayed all of Canada's available real-time or near real-time high frequency electric system data by province and territory in a single access point (Préfontaine, 2021). In its second phase, RTED aims to create a national statistical framework to provide a consolidated, central, and easily accessible dashboard for the public, policy makers and energy stakeholders (Préfontaine, 2021). The RTED dashboard will include indicators on electricity demand (load and load forecast), generation (by energy source), trade (between jurisdictions) and pricing (Préfontaine, 2021). Such data is already collected and stored by provincial transmission system operators, but only a few operators make their data publicly available. Furthermore, the frequency at which data is released generally does not allow data users to delve into dynamic analyses that rely on greater temporal resolution. Finally, the data is typically dated by the time it is released; in some cases, the data are ~3 months behind. In 2021, the RTED plans to consult with utilities and system operators to obtain access to the currently unavailable data. In filling these gaps, RTED aims to support policy and research efforts, enable assessments of grid reliability and resilience, improve situational awareness, and support a better understanding of international trade in the electricity grid, while improving transparency and trust among stakeholders (Préfontaine, 2021).

Commodity Tracking System - CER

The CER commodity tracking system contains monthly energy trade (imports, exports, volumes, prices) data for natural gas and LNG, crude oil, RPPs, NGLs, and electricity (Canada Energy Regulator, 2021b). There is up to ~3 months lag in the published data. The data by commodity can also be found on the OpenGov Portal.

Reference scenario data – CER

The CER's annual Canada's Energy Future report provides a conceptually consistent "Reference Case" of long-term supply and demand projections that incorporates the current economic outlook, a moderate view of energy prices as well as technological improvements, while considering announced climate and energy policies (Canada Energy Regulator, 2021c). It can be seen as a common baseline reference case, since it: (1) acts as a de facto scenario for several prominent analyses; (2) is produced by a public organization; (3) includes projections with broad coverage; and (4) is freely available, published annually, and supported by a full narrative report that provides context to the quantitative results. Oftentimes, modelers who use the Reference Case require greater granularity than is available in Canada's Energy Future appendices and contact the CER to request access to the additional data.

The following list describes the data available in the Canada's Energy Future report appendices. Historical data is largely drawn from StatCan, with modifications throughout to address data issues and align with certain provincial datasets.

- High-level macroeconomic indicators at the provincial/territorial scale
- Benchmark prices at the national and international scales
- End-use prices, end-use energy demand, and primary energy demand disaggregated by sectoral and fuel type at the national and provincial/territorial scales
- Crude oil and natural gas resources and reserves per basin at the national and provincial/territorial scales
- Crude oil and natural gas production by product type at the national and provincial/territorial scales
- Refinery balances at the national scale
- Natural gas liquids (ethane, propane, butane, pentanes plus) according to supply and disposition categories at the national scale
- Natural gas drilling (by type in some provinces) at the provincial scale

- Coal supply and demand by coal type at the National scale
- Electricity Generation by primary fuel at the National/Prov/Terr
- Electricity Capacity by primary fuel at the National/Prov/Terr
- Electricity Interchange (Interprovincial & international exports/ imports) National/Prov/Terr

1.2 Database landscape in Europe and the US

Several organizations in Europe have sought to make the collaborative energy systems modelling effort more efficient and goal-oriented for all researchers by developing open data and modelling platforms. The resulting databases have been developed with various spatial and temporal resolutions, including at the multi-national (Jensen & Pinson, 2017), and individual city scale (Wallin, 2020). At the level of the European Union, the value of data is evident from the rapid development of data-driven innovation hubs and smart cities. Annually, an Open Data landscaping exercise is performed that analyses and reports on the development of open data activities among European countries. The recent report (Cecconi & & Radu, 2018) categorized countries into four categories regarding their progress (from beginners to trend-setters), indicating the different speeds at which Europe is progressing while highlighting the need for a better transfer of knowledge and the never-ending need to excel at reporting high quality data.

The US Department of Energy (DOE) and its many subsidiary agencies, including the Energy Information Administration (EIA), the Federal Energy Regulatory Commission (FERC), the National Renewable Energy Laboratory (NREL), and the 17 National Laboratories, provide energy system data to the public in many different forms. The EIA, being the statistical agency of the DOE, hosts a number of data repositories including one recently published that contains cleaned hourly electricity demand for electric balancing authorities within the contiguous US (Ruggles, Farnham, Tong, & Caldeira, 2020). The EIA also has application programming interface (API) features for its main repositories allowing energy systems models to query the data (U.S. Energy Information Administration, 2021). FERC has limited organizational capacity in terms of data repositories, but offers many forms of data scattered throughout its system with links to external data sources (Federal Energy Regulatory Commission, 2021). NREL hosts a data catalogue of all its publicly available data (National Renewable Energy Laboratory 2021), including those in the form of databases, but they are not entirely organized into a master database. The DOE itself has an open data catalogue that directs users to data repositories made publicly available by agencies under its umbrella (U.S. Department of Energy, 2021a). Additionally, the DOE manages the Open Energy Data Initiative (OEDI), which is a centralized repository of energy research data that is sourced from private industry, universities, national laboratories, and DOE agencies among others (U.S. Department of Energy, 2021b). Unfortunately, the DOE does not host a public database that can be queried to supply their information and data.

There are a number of other US agencies that host energy system data, including the Department of the Interior, Mine Safety & Health Administration, Army Corps of Engineers, Environmental Protection Agency (EPA) and various independent system operators (ISO), among others. While there is a considerable quantity of data made available by the different organizations, it is distributed and lacks a central database consolidating the information into a standard format. There have been efforts to compile data from all the major sources, such as the Public Utility Data Liberation (PUDL) Project (Catalyst Cooperative, 2021). PUDL currently houses hundreds of gigabytes of information from the EIA, FERC, and the EPA Continuous Emissions Monitoring System (CEMS) (Catalyst Cooperative, 2021); however, it still fails to capture all of the data in a standard format. Overall, the US could benefit from a nationally standardized database system that consolidates data from all federal energy institutions and agencies while reconfiguring it in an acceptable format for use in common energy systems modeling frameworks.

1.3 Database attributes

As we survey the data landscapes in the EU and US, review the academic literature on open data and models, and build on conversations and insights with policy makers, several key attributes critical to an effective database come to the fore. In particular, there are three essential but lacking components to the integration of model-based evidence into the decision-making process in the Canadian context: (a) timeliness, (b) transparency, and (c) inclusiveness.

(a) **Timeliness:** Data collection is often one of the early and time consuming steps in the modelling process. For example, the Sustainable Energy Systems Integration and Transitions (SESIT) research team (sesit.cive.uvic.ca) has spent almost two years collecting data to model Canada's electricity system (pertaining to generation assets, the transmission network, and load). For a policy maker with a defined and limited window to propose policy or implement programs, this significant upfront data collection effort is non-workable. By assembling this data in a 'standing' database that is ready to be leveraged on a moment's notice, the current database contributes to overcoming one of the most difficult and time-consuming aspects of energy systems modelling.

(b) **Transparency:** The trope "garbage in, garbage out" is a common saying in the modelling field, in part because it is so very true. When presenting modelling results, the reaction of many decision makers (as well as other modellers) is often (rightly): *but what were your inputs?* Oftentimes, energy system models leverage thousands of data points which are difficult to communicate explicitly. By developing a common database that is open, shared, version controlled, and standardized, this effort takes a significant and important step in injecting transparency into the modelling workflow.

(c) **Inclusiveness:** For model results to carry the weight required to impact decision making, the process must convene a diverse range of disciplines, perspectives, and stakeholders within the modelling process, specifically in the scenario definition and input definition stages. Outside of Canada, there are examples of this inclusivity that demonstrate its effectiveness. For example, the Energy Modelling Platform for Europe (EMP-E) convenes model teams (funded under Horizon 2020 grants) with policy makers from the European Commission. In their annual workshops, policy makers communicate upcoming policy priorities to modellers and utilize the results from modelling efforts to inform next steps. In addition to policy makers, engaging the public and citizen scientists is important for the energy system transition – i.e., gaining public support from the outset. This open, shared, version controlled, and standardized database makes model inputs accessible to a broad range of stakeholders. Further, by developing data processing tools, additional modelling teams as well as an even broader list of stakeholders will be able to interact and use the database.

The database and data processing tools described below will equip and enable the modelling community by reducing time-consuming and redundant data collection work, while also improving evidence-based decision-making by increasing the timeliness, transparency, inclusivity and ultimately the uptake of energy systems modelling outputs.

2 Method

The tasks within this project pertain to the database development itself, including:

1) assembling the raw data and documenting each data set's key characteristics (Section 2.1)
2) accessing and customizing the Spine Toolbox (Section 2.2), and populating it with the collected data
3) establishing licenses for the database (Section 2.3).

2.1 Data collection

Data collection occurred over the two-year period from 2019 to 2021 and was initially driven by the modelling requirements of the SESIT team and our research collaborators. The dataset, Canadian Open-source Database for Energy Research and Systems-Modelling (CODERS) leverages the existing national and provincial databases made public by utilities, system operators, independent power producers, regulators, government agencies and energy associations. Data contained in CODERS relates to generation facilities, transmission networks, substations and other system assets, as well as to system operations, demand, forecasts, imports, exports and costs. As the database evolves in response to a broader set of modelling and policy requirements, it is anticipated that the scope of data contained in CODERS will also evolve and expand. Table 1 provides a high-level summary of the current status of data collection for CODERS.

Table 1: Data availability status in CODERS

Generation / Storage	Status	Generation	Status	Storage	Status	Transmission	Status	Provincial Annual Demand	Status	Provincial Hourly Demand	Status
Name	Obtained	Heat Rate	Not available – calculated	Technology	Obtained	Circuit ID	Obtained	Historical Peak Capacity	Obtained	Historical Energy	Need
Owner	Obtained	Min. Capacity	Not available – calculated	Duration	Obtained	Owner	Obtained				
Location	Obtained	Max. Capacity	Not available – calculated	Associated Generation	Obtained	Region	Obtained	Historical Annual Energy	Obtained	Interprovincial Transfers	Obtained
Lat/long	Obtained	Min. Up Time	Not available – calculated			Current	Obtained				
Region	Obtained	Min. Down Time	Not available – calculated	Cost	Obtained	Length	Not available – calculated	Forecasted Peak Capacity	Obtained	International Transfers	Obtained
Substation	Obtained	Ramp Rates	Not available – calculated	Outage Rates	Under development	Voltage	Obtained				
Start Year	Obtained	Must Run	Not available – calculated	O&M Costs	Under development	Reactance	Not available – calculated	Forecasted Annual Energy	Obtained	International Prices	Obtained
End Year	Not available – calculated	Outage Rates	Not available – calculated	**Hydro**		Rating	Not available – calculated				
Type	Obtained	Start Up Cost	Not available – calculated	Development Potential	Obtained	Capacity	Not available – calculated	Before DSM	Obtained	**System**	
Capacity	Obtained	Shut Down Cost	Not available – calculated			Start Node	Obtained	After DSM	Obtained	Reserve Reqs.	Obtained
Energy	Obtained	O&M Costs	Under development	Reservoirs	Under development	End Node	Obtained	Imports/Exports	Obtained	System Losses	Obtained

Obtained · Not available – calculated · Under development · Need

Although the data was primarily obtained from national and provincial utility and system operator databases, additional data was collected and streamlined from a disparate range of sources. The data collection entailed systematic review of utility resource and infrastructure plans, capital expenditure plans, regulatory board filings and related documentation, supplemented with written information requests to utilities and system operators seeking additional public data sources. In some instances, where data were unavailable, unsuitable or unassembled, direct measurement techniques or estimations using standardized metrics were employed, as discussed further below. As shown in Table 1, while a few of the desired data remain under development or are still required, much of the database is populated with obtained or calculated data.

Electricity generation by province, generation type and interconnection location were tabulated from utility sources, databases and regulatory filings. Design characteristics, including installed capacity and annual energy were tabulated from reported historical operational performance or calculated from known capacity factors. Effective capacities, important to the determination of resource adequacy, were obtained directly from reports of resource adequacy by utilities or reliability agencies, or calculated from reported effective capacity metrics (e.g., electric load carrying capability). Where available, data is reported on a unit-by-unit basis by generating facility. Additional generation spatial information was obtained from the World Resource Institute *Global Power Plant Database* (World Resources Institute, 2019) and verified to ensure accuracy. Only facilities with installed capacities larger than 1 MW, interconnected to a provincial transmission grid and at least partially available to serve provincial demand are included in the database. Remote, territorial and self-generation facilities are currently excluded. Facilities that have received regulatory approval are included in the database based on their reported installed capacities and anticipated in-service dates.

CODERS includes locational, design and operational data for existing and planned energy storage facilities. Design characteristics, including technology, storage capacity, storage energy and duration were sourced from utilities or independent power producers. Storage facilities charged directly by a generation facility are distinguished from those that are grid-charged. Only grid-connected storage facilities with installed capacities larger than 1 MW are included in CODERS. Facilities located behind the meter are excluded from the database.

Data respecting transmission facilities within CODERS includes locational, dimensional, design and operational characteristics. Data was sourced using a "snowball" technique in which asset identification codes for transmission lines and substations were used to locate lists of additional transmission lines and substations and their associated identification codes. Evolving lists of transmission lines and substations were compared against utility system maps and single line diagrams until all known substations and lines were identified. Locational information, primarily in the form of interconnecting nodes (i.e., substations, generators and junctions) was obtained from utility system regulatory filings, government documents and system maps, and verified using Google Maps. Where transmission line lengths were not located or were not specifically reported by utilities or government agencies, lengths were estimated using available transmission system maps, single line diagrams and Google Maps. For reactance, a comprehensive study of FERC regulated transmission lines (Athari & Zhifang, 2017) was used to estimate the relationship between voltage and reactance. In order to estimate line capacity (flow limits), St. Clair curves along with industry standard information (Glover, Overbye, & Sarma, 2015) were used to determine the relationship between voltage, line length and surge impedance loading (Gutman, Marchenko, & Dunlop, 1979). These data estimates were verified for accuracy using known utility line reactance and capacity data publicly available for select transmission lines. International and interprovincial transfer capacities were obtained from the Canada Energy Regulator (Canada Energy Regulator, 2021a), utilities and electric reliability entities.

Historical and forecasted annual energy (GWh/year) demand and peak capacity (MW) requirements, before and after the effects of demand-side management (DSM), were sourced from utility or system operator resource plans, load forecasts or regulatory filings. Hourly electricity demand was similarly sourced, with most utilities and system operators making hourly internal demand and total demand data (including imports/exports) publicly available, with some also providing interprovincial and international hourly intertie flows and hourly prices.

2.2 Spine toolbox

The Spine Toolbox offers an effective platform to structure, standardize, version control and ultimately share data (Kouveliotis-Lysikatos, Marin, Olason, Amelin, & Söder, 2020). An ongoing 4-year effort led by VTT in Finland and funded under Horizon 2020, the Toolbox uses processing tools to interconnect data and modelling tools. The Toolbox provides a modular and adaptable platform that can be applied to a broad set of problems. The Spine platform is utilized in this study as a tool to create an interface connecting the developed database with a set of energy systems models in order to provide modelling input data.

Specifically, the Spine Toolbox provides an intuitive user-interface that allows developers to build flow-based models of energy systems; the visualization, data storage, and data manipulation tools used in the application exist in a Python execution environment allowing modelers to implement frameworks using a programming language commonly used to build energy systems models. Spine was designed to facilitate quick and flexible model creation.

The visualization features of the Spine Toolbox provide rich illustrations of modelling work-flows offering developers a more complete view of complex model executions. Figure 1 shows the workflow for a theoretical example with two data sources and importers, three databases and four tools.

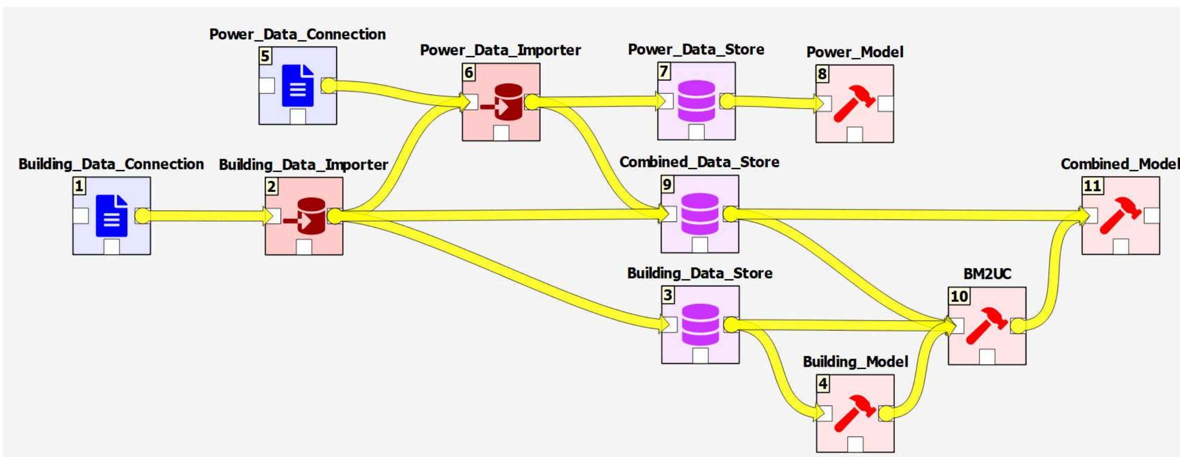


Figure 1: Example Spine Toolbox workflow combining power system and building models

The Spine platform establishes a common data storage structure that uses data processing tools to provide data to energy models of different scope, thus allowing for an efficient modelling workflow for complex interlinked systems. This approach facilitates efficient sharing of resources across modelling tools. More specifically, Spine allows users to:

1) build data processing tools that other users can utilize, avoiding duplication of effort
2) use shared server-based databases that house data in a standardized format
3) implement version control tools in repositories and built-in metadata structures
4) interconnect models to the standardized format
5) use the shared data as a starting point with additional functionality for project-based modifications

6) utilize tools and models developed by others (within and outside of Canada)
7) execute the workflow in a computing cluster or in the cloud
8) provide simplified access and query capabilities for non-technical stakeholders

In the current project, the open source release of the Spine Toolbox[2] is applied to CODERS to enable data querying and selection. A Spine platform is developed here as a testbed to gain access to the database system using API functionality, as described further below. The platform is currently used as the central access point to the database, which allows the development team to isolate activity around the system until a safe and effective solution can be designed to deploy the database online using a URL. The platform is developed into a generic interface that can pull information from the database and supply it to common modelling frameworks as input. The main intent is for the Spine Toolbox to act as a platform for connecting different energy systems models to the novel database, and to serve as a medium to both execute and compare the outcomes of those models using a standard set of input data that is taken from the database.

2.3 Licensing

The developers intend to publish CODERS non-commercially and make it fully available to the public under an open-source license. The database has not yet been licensed but is currently in the licensing process. Working with the University of Victoria, we are undertaking a full review of the data sources while seeking external counsel to ensure appropriate management of the licensing process. The steps that remain before obtaining a license include: 1) reviewing the general publishing constraints of each data source; 2) confirming the status of data taken from utilities as being public; 3) seeking the expertise of a legal copyright expert; and 4) performing an internal review of the final database product with the University of Victoria's Research Partnerships and Knowledge Mobilization unit (RPKM). The RPKM is the main unit within the University with whom the development team is working to license the developed database

[2] https://github.com/Spine-project/Spine-Toolbox

system, and have indicated their willingness to provide proof of legal support for the licensing process. The main goal of the RPKM, in regard to licensing the database, is to limit the liability of the host in relation to use of the database.

3 Results and analysis

The following sections describe how APIs link the data to models (Section 3.1), the database limitations (Section 3.2) as well as envisioned future work (Section 3.3).

3.1 Linkage to the SESIT suite of models

CODERS supports a broad suite of activities both within and beyond the SESIT group. Following its initial development, the database has been linked to a suite of energy systems models developed or used within the SESIT team, as shown in Figure 2. The data that populates each of these models is summarized in Table 2 and the figures that follow (updated from (McPherson M. , 2020)).

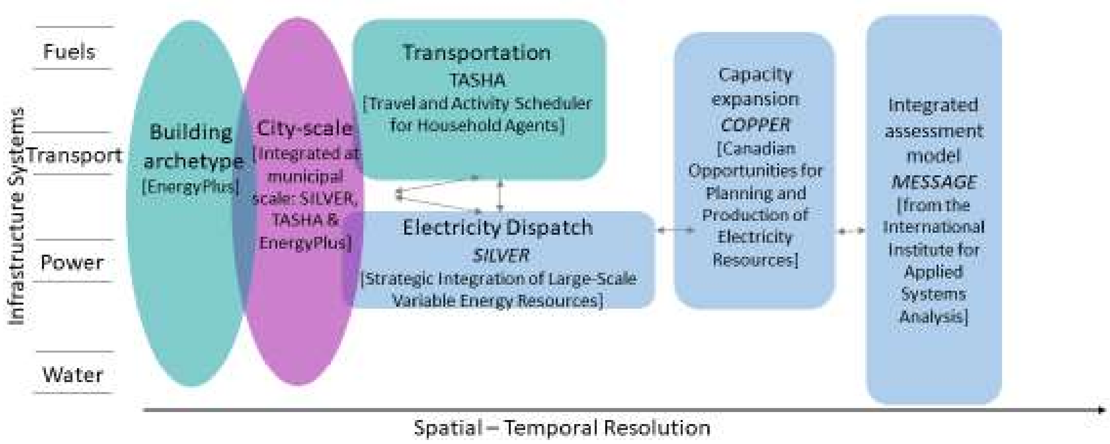


Figure 2: The SESIT modelling suite spanning infrastructure systems and spatial-temporal scales

(adapted from (McPherson & Akhtar, 2019))

Table 2: Summary of data that populates each of the SESIT suite of models

Capacity Expansion
Load - hourly demand data at city scale resolution; annual forecasted growth rate for each province or city; hourly import and export data to the US
Transmission - system (GIS) map; a list of under construction and future transmission lines
Generation – existing, under-construction and future generation unit data; retirement schedules; investment costs by province; locations for wind, solar and hydro development; water resource availability
Policy - emission reduction targets and policies per province (existing and under review)
Other - fuel type price per province; natural gas supply limitations and supply curve per province
Electricity Dispatch
Load - total hourly load; hourly nodal load (per bus)
Transmission - network configuration (buses and their connections through lines); inter-provincial lines and hourly transferred power; line capacity, reactance, length, voltage
Generation - asset type and installed capacity; operational characteristics (ramp, min up/down time, cost curve, etc.); bus location; hydro reservoir (if applicable)
Transport
Network - zone to zone auto and transit travel times/ distances
Vehicle Fleet - future electric vehicle stock predictions; household vehicle ownership level
Travel Behaviour - electric vehicle charging behaviour; individual travel behaviour
Electricity Grid - mapping between travel zone and substation
Infrastructure - work/home/commercial charging availability
Building
Buildings - meter data; building characteristics (lot size, wall-to-window ratio, wall insulation, thermal envelope, equipment, air infiltration, occupancy)
Regional Characteristics - weather data; GIS features
Population Characteristics - demographic information; occupant preferences

Table 3: Data status for the capacity expansion model

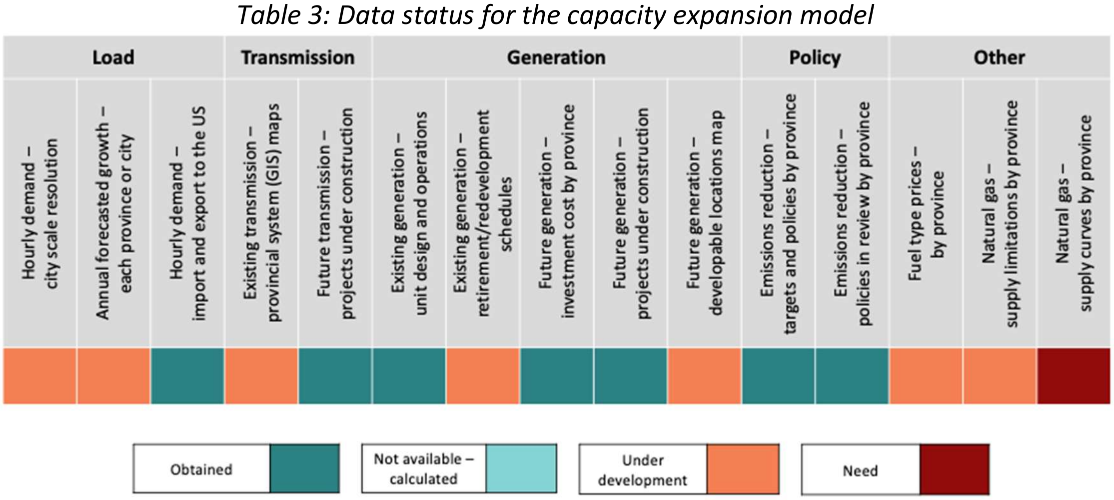


Table 4: Data status for the electricity dispatch model

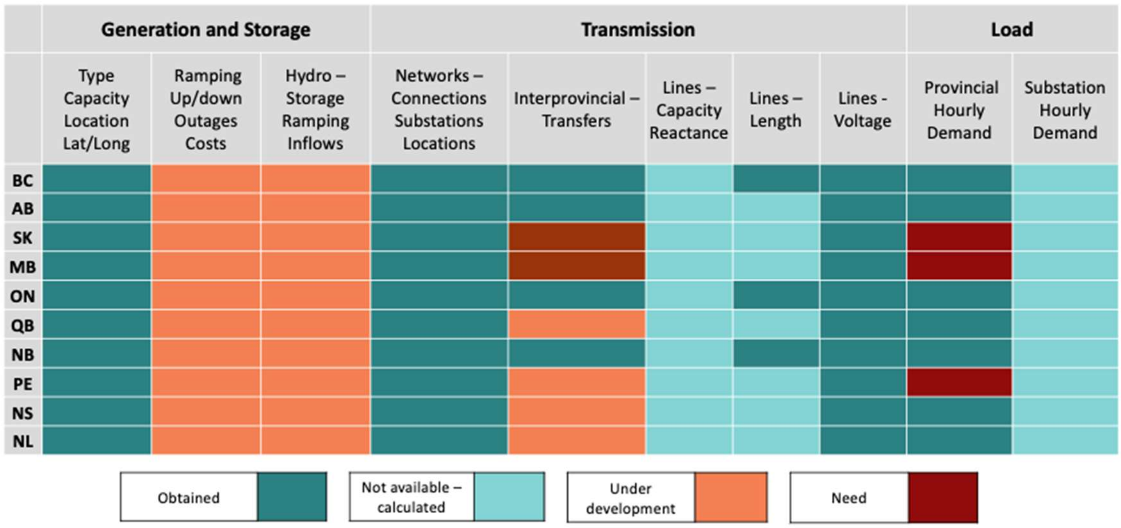


Table 5: Data status for the transport sector model

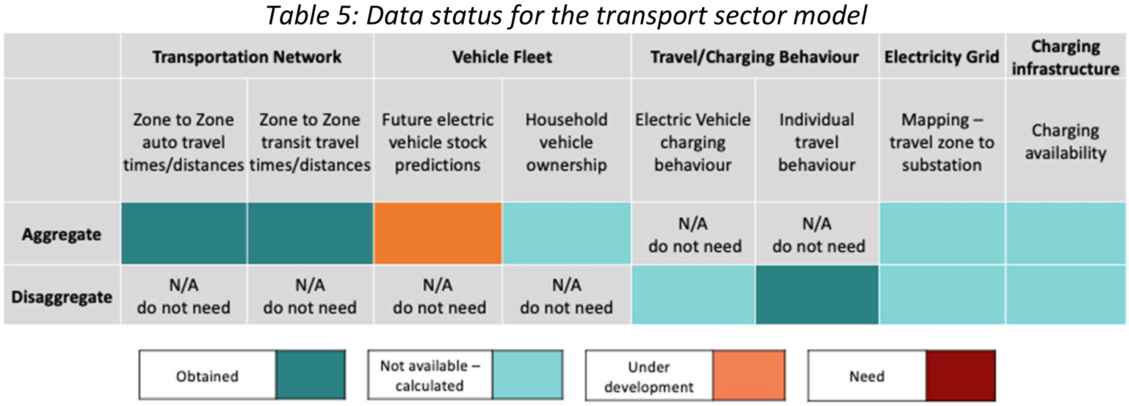


Table 6: Data status for the building system model

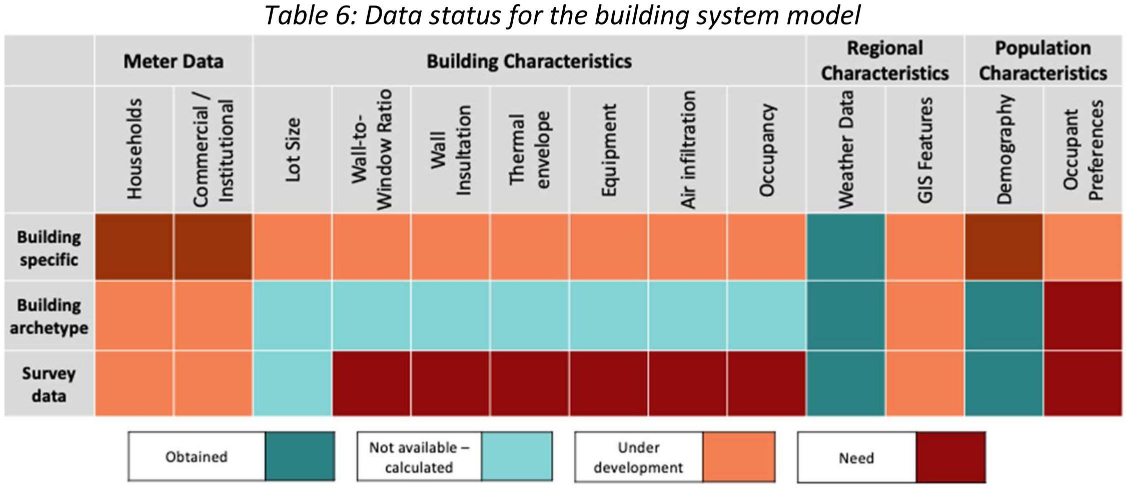
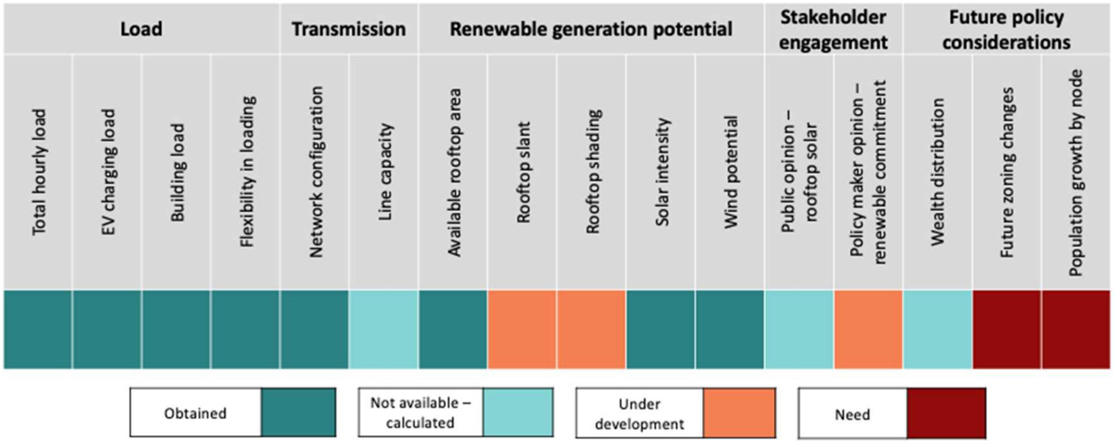


Table 7: Data status for the Integrated city model



Over the long term, our hope is that other modellers will access the database either indirectly (via APIs) or directly (by querying the database) for use in their own work. For now, the database will be accessible to the public via the API, and as the API is the layer of communication between the user and the database, the typical user is actually accessing it indirectly. A direct access to the database to regular users cannot be granted because it poses a security risk and endangers the integrity of the database. The plan is for the API to be open source, meaning others can view source code and potentially even add to it. The API code itself can be expanded by users because of its routes and endpoints. A route is the name used to access an available endpoint, which performs a function on a parameter and returns the resulting data pursuant to that function. In essence, the combination of a route/endpoint converts a user request into a database query. Any query that is required can thus be added as a route. The API can be used to pass data to the model code, pass data to the website or application, as well as to update and make structural changes to the database. In our case, the API will be created by Flask which is a lightweight, micro web framework written in Python. Being lightweight/micro indicates that it does not require third-party libraries or tools. This is useful because it means the API has fewer dependencies making it easier to maintain, scale and adapt into the future. The API itself will be held in a Docker container which is a structure that bundles all the code, libraries, and configurations in one package, simplifying deployment.

3.2 Limitations

The database has notable limitations, which we aim to overcome in time (refer to the *Envisioned future work* section, below).

The first database limitation arises from the data itself. The data procured to develop the database generally comes from sources within the electricity sector. It would be useful to include energy-related data from other sectors, particularly including water and agriculture because of their deep interconnections with the energy system and to support research that is directed towards the food-water-energy nexus. Other sectors would be useful as well, including industry, particular in relation to recent, ongoing and future decarbonization efforts. Appending data from other sectors will enable interdisciplinary research that creates bridges in knowledge between electricity system operations and the behaviours of end-users.

The database system is currently hosted by Compute Canada, which limits access privileges and imposes rules. Compute Canada provides computing, storage and software solutions for Canadian researchers, but a researcher must be verified as a faculty member to be granted access. Faculty members can sponsor a number of different individuals, including graduate students and external collaborators for access privileges to the database. However, the pool of users is limited to those with a connection to academia. The database should eventually be open-sourced on a hosting platform that can provide open access to the data, allowing a broader spectrum of users to interact with the system and fostering creative energy solutions.

3.3 Envisioned future work

Currently, the database can be queried via Structured Query Language (SQL), but there is no graphical user interface (GUI) for accessing the database directly. Public GUIs suffer from security issues that could make the database vulnerable, which means an intervening layer, such as a representational state transfer (REST) API, must be implemented. An API provides a layer of abstraction that helps block malicious users from accessing the database, while a REST API utilizes an HTTP request in order to access and use data. Future work will develop an API and browser-based GUI (website) which will offer easier access to the database and eliminate the need for SQL experience. Furthermore, the database will be migrated to an open-source platform that can host the system for an audience that spans beyond academics and their external collaborators.

Additionally, the database will be integrated with several energy system models and an interactive visualization platform via the Spine Toolbox. This will create a new comprehensive energy system workflow framework that will be used to compare different energy system models. Applying the database to more energy systems models across Canada will identify the need for additional input data, allowing the database to grow according to national and institutional needs. The database can be extended to pull information from other major sources of energy system data, storing it in a common format for use in a national modelling platform.

Finally, the data hosted in the database will require ongoing updates and maintenance.

4 Discussion

The Energy Modelling Initiative (EMI) seeks to convene Canadian energy system stakeholders to facilitate evidence-based decision-making. By doing so, policy and decision makers can leverage a diverse range of perspectives and models, building transparency and confidence. Reference datasets are a key pillar in this larger framework, as they are an essential ingredient to evidence-based decision-making. CODERS aims to address gaps in the energy information landscape to support the integration of model-based evidence into decision-making processes within the Canadian context. By providing standardized and accessible electricity systems information, CODERS addresses several gaps in the Canadian energy information landscape related to timeliness, transparency, and inclusiveness.

(a) **Timeliness**: Many researchers, regulatory interveners, and non-utility stakeholders report having to “fight” for access to usable electricity systems information in Canada, a marked contrast from more open access in other jurisdictions. CODERS addresses this disparity by providing ready access to standardized and accessible data for modellers, policymakers and the general public. By gathering and collating data historically held in disparate utility databases or hidden away in regulatory filings, CODERS provides a single, standardized, accessible and accurate source of electricity systems data for Canada that will enable timely analyses. As an open-source database, CODERS is amenable to continual expansion as additional data becomes available, improving the characterization of the national and provincial electricity systems over time. Information flow across the modeller-decision-maker interface has traditionally been slow, and CODERS substantially improves the timeliness of information availability while avoiding duplication of effort in data collection by multiple modelling teams across academia, industry and government.

(b) **Transparency**: As and open, shared, version controlled, and standardized database, CODERS introduces into the electricity systems modelling effort a higher level of transparency than has been available to date in Canada. Indeed, increasing the transparency of modelling exercises is the core criteria that drives the motivation for this work. By possessing an open access database for use by modelling teams, much of the uncertainty associated with 'black box' data inputs to models is removed. The lack of transparency associated with model data inputs has been a substantial barrier to delivering robust model-based insights. By overcoming the transparency of data inputs, this project takes a significant step in overcoming this issue. Open source models themselves are another key ingredient. Data clarity and standardization within an open-source and accessible database supports comparisons of outputs across modelling platforms allowing for improvements in model design, performance and integration. Greater availability of transparent national electricity systems data is also anticipated to result in new research hitherto deemed infeasible due to the significant time and resource hurdles involved in gathering data to support complex projects. Transparency must be balanced against privacy, confidentiality and security, and though CODERS is designed as an open-source, publicly-accessible database, provisions are made to support additional research involving confidential or proprietary data.

(c) **Inclusiveness**: CODERS can be leveraged to populate energy system models outside those described in this report, providing data in a format that can be used with other common modeling frameworks that simulate energy systems. Through this open-source and accessible database, modelling teams can begin from a consistent starting point before branching out into research-specific, exploratory scenarios. As such, CODERS allows for consistency of input data between machines running the same model, eliminating the drawbacks of using locally hosted data. Further, the data contained in the system can be used to develop national standards for input data for fair comparisons between energy modeling platforms. CODERS can also be used as a tool for both parameterizing the various energy system models and validating their output against known test cases. This supports collaborative and interdisciplinary processes that convene a diverse range of disciplines, perspectives, and stakeholders within the modelling process, particularly in the early scenario definition stages where consensus and collaboration are essential to defining relevant and meaningful analyses to support policy and decision-making.

A national modelling platform

Our hope is that CODERS forms a foundation that can be extended into a broader series of activities that convene, leverage, and showcase Canadian modelling capacity. Such other activities could include a visualization suite, a repository of open-access models, or an ongoing series of modelling forums that convene modelling teams. By publishing an open access database, we take the first step in what we hope is a broader and longer term effort.

Given its ability to provide standardized input into other energy modeling frameworks, the database can be integrated into a national modelling platform to supply input for a number of models used by decision-makers to manage energy infrastructure. In an integrated modelling platform, the database can be used by various models as a transparent reference that gives confidence to stakeholders when making decisions concerning the Canadian energy system. Integrating the database into a national modelling platform brings big advantages to the modelling community. The integrated platform approach will make it possible to explore a range of energy transition scenarios with multiple modeling frameworks and to compare their outputs using a standard of input data developed from the database.

Bibliography

AESO. (2021). *Data Requests.* Retrieved from https://www.aeso.ca/market/market-and-system-reporting/data-requests/

Athari, M., & Zhifang, W. (2017). Interdependence of Transmission Branch Parameters on the Voltage Levels. *Proceedings of the 51st Hawaii International Conference on System Sciences.* Big Island, HI.

Canada Energy Regulator. (2021a). *International Power Lines Dashboard*. Retrieved from https://www.cer-rec.gc.ca/en/data-analysis/facilities-we-regulate/international-power-lines-dashboard/index.html

Canada Energy Regulator. (2021b). *Energy Commodities*. Retrieved from https://www.cer-rec.gc.ca/en/data-analysis/energy-commodities/

Canada Energy Regulator. (2021c). *Canada's Energy Future.* Retrieved from https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/

Catalyst Cooperative. (2021). *The Public Utility Data Liberation Project*. Retrieved from https://catalyst.coop/pudl/

Cecconi, G., & & Radu, C. (2018). *Open data maturity in Europe. Report 2018. European Commission report.* Retrieved from https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

Clarke, L., Fawcett, A., Weyant, J., McFarland, J., Chaturvendi, V., & Zhou, Y. (2014). Technology and US emissions reductions goals: Results of the EMF 24 modeling exercise. *The Energy Journal, 35*(Special Issue).

DeCarolis, J. F., Jaramillo, P., Johnson, J. X., McCollum, D. L., Trutnevyte, E., Daniels, D. C., & ... Zhou, Y. (2020). Leveraging open-source tools for collaborative macro-energy system modeling efforts. *Joule, 4*(12), 2523-2526.

Doukas, H., Patlitzianas, K. D., Kagiannas, A. G., & Psarras, J. (2008). Energy policy making: an old concept or a modern challenge? *Energy Sources, Part B, 3*(4), 362-371.

Energy Modelling Platform For Europe. (2020, October 6-8). *EMP-E 2020 Modelling Climate Neutrality for the European Green Deal.* Retrieved from https://emp-e-2020-modelling-climate.b2match.io/page-121

ENTSOE. (2021). *Transparency Platform*. Retrieved from https://transparency.entsoe.eu

European Commission. (2021a). *Modelling tools for EU analysis*. Retrieved from https://ec.europa.eu/clima/policies/strategies/analysis/models_en

European Commission. (2021b). *EU 2030 Climate and Energy Framework*. Retrieved from https://ec.europa.eu/clima/policies/strategies/2030_en

Federal Energy Regulatory Commission. (2021). *Shaping the Grid of the Future*. Retrieved from https://www.ferc.gov

Foley, A. M., Gallachóir, B. Ó., Hur, J., Baldick, R., & McKeogh, E. J. (2010). A strategic review of electricity systems models. *Energy, 35*(12), 4522-4530.

Glover, J., Overbye, T., & Sarma, M. (2015). *Power System Analysis and Design (Sixth Edition).* Boston: Cengage.

Government of Canada. (2019a). *Atlantic Canada Growth Strategy - Clean Energy*. Retrieved from https://www.canada.ca/en/atlantic-canada-opportunities/news/2019/03/atlantic-growth-strategy--clean-energy.html

Government of Canada. (2019b). *Evaluation of the Regional Electricity Cooperation and Strategic Infrastructure (RECSI) Initiative.* Retrieved from https://www.nrcan.gc.ca/nrcan/plans-performance-reports/strategic-evaluation-division/reports-plans-year/evaluation-reports-2019/evaluation-regional-electricity-cooperation-and-strategic-infrastructure-recsi-initiative/22363

Government of Canada. (2020a). *Memorandum of Understanding between the Government of Canada and the Government of British Columbia on the electrification of the natural gas sector.* Retrieved from https://pm.gc.ca/en/news/backgrounders/2019/08/29/memorandum-understanding-between-government-canada-and-government

Government of Canada. (2020b). *Clean Energy and Electricity Infrastructure*. Retrieved from https://www.nrcan.gc.ca/climate-change/canadas-green-future/clean-energy-and-electricity-infrastructure/21294

Government of Canada. (2021). *Canadian Center for Energy Information*. Retrieved from https://energy-information.canada.ca/en

Gutman, R., Marchenko, P., & Dunlop, R. (1979). *Analytical Development of Loadability Characteristics for EHV and UHV Transmission Lines.* Retrieved from https://ieeexplore.ieee.org/document/4113522

Hirth, L. (2020). Open data for electricity modeling: Legal aspects. *Energy Strategy Reviews, 27*, 100433.

Horschig, T., & Thrän, D. (2017). Are decisions well supported for the energy transition? A review on modeling approaches for renewable energy policy evaluation. *Energy, Sustainability and Society, 7*(1), 5.

Jensen, T. V., & Pinson, P. (2017). RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system. *Scientific Data, 4*(1), 1-18.

Kouveliotis-Lysikatos, I., Marin, M., Olason, J., Amelin, M., & Söder, L. (2020). A Network Aggregation Tool for the Energy System Modelling Framework Spine. *2020 International Conference on Smart Energy Systems and Technologies (SEST).* IEEE.

Leach, A., Rivers, N., & Shaffer, B. (2020). Canadian Electricity Markets during the COVID-19 Pandemic: An Initial Assessment. *Can. Public Policy, 46*(S2), S145–S159.

MacGillivray B.H, .., & Richards, K. (2015). Approaches to evaluating model quality across different regime types in environmental and public health governance. *Glob Environ Change, 33*, 23–31.

McPherson, M. (2020, March 11). Modelling electrification with the SESIT suite: Data sources and gaps. *CCEI – EMI Workshop*. Ottawa.

McPherson, M., & Akhtar, T. (2019). Introducing a Flexible Platform for Modelling Energy Systems Integration. *18th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants*. Dublin, Ireland.

National Renewable Energy Laboratory. (2021). *NREL Data Catalog*. Retrieved from https://data.nrel.gov/submissions

Natural Resources Canada. (2021). *Clean Energy and Electricity Infrastructure.* Retrieved from https://www.nrcan.gc.ca/climate-change/canadas-green-future/clean-energy-and-electricity-infrastructure/21294

NRCan. (2020, October 7). *Government of Canada Launches the Canadian Centre for Energy Information Website*. Retrieved from https://www.canada.ca/en/natural-resources-canada/news/2020/10/government-of-canada-launches-the-canadian-centre-for-energy-information-website.html

Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., & Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind? *Energy Policy, 101*, 211-215.

Porter, T. (1995). *Trust in numbers: the pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Préfontaine, S. (2021, February 11). *Real-Time Electricity Data.* Retrieved from Webinars: https://emi-ime.ca/events/webinars/

Ringkjøb, H. K., Haugan, P. M., & Solbrekke, I. M. (2018). A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews, 96*, pp. 440-459.

Ruggles, T., Farnham, D., Tong, D., & Caldeira, K. (2020). Developing reliable hourly electricity demand data through screening and imputation. *Scientific data, 7*(1), p. 1:14.

Ryberg, D. S., Robinius, M., & Stolten, D. (2018). Evaluating land eligibility constraints of renewable energy sources in Europe. *Energies, 11*(5), 1246.

Statistic Canada. (2021). *Electricity and renewable energy.* Retrieved from https://www150.statcan.gc.ca/n1/en/subjects/energy/electricity_and_renewable_energy

The Office of the White House. (2016, November 4). *United States Mid-Century Strategy for Deep Decarbonization.* Retrieved from https://unfccc.int/files/focus/long-term_strategies/application/pdf/mid_century_strategy_report-final_red.pdf

U.S. Department of Energy. (2021a). *Open Data Catalogue*. (U.S. Department of Energy)

U.S. Department of Energy. (2021b). *OpenEI*. (U.S. Department of Energy) Retrieved from https://openei.org/wiki/Main_Page

U.S. Energy Information Administration. (2021). *EIA*. Retrieved from https://www.eia.gov

Wallin, F. (2020). *World Class energy solutions*. Retrieved from https://www.mdh.se/en/malardalen-university/research/research-projects/worldclass-energy-solutions

Washington Governor Jay Inslee. (2016, December 16). *Deep Decarbonization*. (T. O. Washington, Producer) Retrieved from https://www.governor.wa.gov/issues/issues/energy-environment/deep-decarbonization

Williams, J., Haley, B., Kahrl, F., & Moore, J. (2014). *Pathways to Deep Decarbonization in the United States.* Retrieved from USDDPP Reports: http://usddpp.org/usddpp-reports/

Appendix A – CODERS Database Spec Sheet

Appendix B – CODERS Manual