



Public Utility Data Liberation

Towards open-source U.S.
energy data

Christina Gosnell
Catalyst Cooperative

catalyst.coop
data.catalyst.coop
hello@catalyst.coop

What is Catalyst Cooperative?

- Data engineerings & analysts
- Support folks with the data they need to do their work
- Creators & maintainers of **PUDL**: free & open-source analysis-ready energy data



The Public Utility Data Liberation Project

Loads of energy data is **public** but is not always **accessible**.

We support **researchers, advocates, journalists and clean energy practitioners** with the data they need to do their work.

We aim for **clean, standardized, granular, connected** and **complete** data about the US energy system.



What data does PUDL process?

- **FERC** - Form 1, Form 714
- **EIA** - 860, 923, 861, 930, 176, AEO
- **EPA** - Continuous Emissions Monitoring System (CEMS)
- **NREL** - Annual Technology Baseline (ATB)
- **SEC** - 10-K Parent-Subsidiary Ownership
- **GridPath** - Resource Adequacy Toolkit
- **Vibrant Clean Energy** - Resource Adequacy Renewable Energy
- **PHMSA** - Gas Distribution Pipelines

Downstream Projects:

- PowerGenome (ZEROLab, Princeton)
- Open Grid Emissions (Singularity Energy)
- Engage&Act + Utility Transition Hub (RMI)
- PY-PSA USA
- Many others!



How we transform the data

- Standard format(s)
- Long time series
- Standardize units & codes
- Linkages *within* a dataset (FERC plants, EIA units, etc.)
- Linkages *between* datasets (FERC 1-EIA, EIA-EPA, SEC-EIA, etc.)
- Removal of egregious data errors
- Tidy, well-normalized tables
- Imputing missing values
- General-use analysis

Least opinionated



Most opinionated



How do we do it?

- **Raw data archiving:** automated versioned archive creation on Zenodo
- **Data processing pipeline:** Grabs raw archives and EtLT using a pipeline orchestrator
- **Extensive data validation:** ensure baseline data quality
- **Nightly builds:** ensure up-to-date outputs, maintain reliability
- **Stable monthly releases:** Integration of new data. Stable, versioned releases



Many Modes of Access

<u>Platform</u>	<u>Format</u>	<u>Version</u>	User Types
<u>Data Viewer</u> data.catalyst.coop	Web, CSV & Parquet	nightly	Data Explorer, Spreadsheet Analyst
<u>Notebooks</u>	SQLite & Parquet	nightly	Data Scientist, Data Analyst, Jupyter Notebook User
<u>Cloud Storage</u>	SQLite & Parquet	nightly & stable	Data Scientist, Analytics Engineer, Data Engineer, Cloud Developer
<u>Zenodo Archives</u>	SQLite & Parquet	stable	Researcher, Publisher, Archivist

Tooling philosophy

- **Meet users where they are.**
- **Choose Boring Tools.** But do adopt new tools and best practices.
- **Automate everything you can.** But maintain a human element and test extensively.
- **Maintainability.** Its hard but worth it.
- **Faster Feedback Loops.** Its hard but worth it.



Why Open Source?

- ... It's **public** data
- Reducing Toil
- Building Blocks
- Transparency (!= Black Box)
- Information Asymmetry



How can we collaborate?

- **Use PUDL Data!**: Access PUDL open data! Do good stuff with it.
- **Tell us about it!**: We love hearing about what PUDL data is doing in the world. ([github](#), [slack](#), [newsletter](#))
- **Collaborate on funding!**: Need a data partner on a project?
- **Become a PUDL Sustainer!**: Be key part of keeping PUDL free and regularly updated!
- **Hire Us!**: Need data analysis or engineering? Want a new dataset in PUDL? Need help with an internal data process?



Thank you!

Get in touch!

hello@catalyst.coop

Follow our work:

catalyst.coop

github.com/catalyst-cooperative/pudl

On the socials:



catalyst.coop



catalyst-cooperative



ALFRED P. SLOAN
FOUNDATION



RMI
ENERGY. TRANSFORMED.

GridLAB



ZERO LAB
PRINCETON UNIVERSITY



Climate Change AI



Singularity

Down
Ballot
Climate
Partners

PSE