ENERGY RESEARCH AND DEVELOPMENT DIVISION

FINAL PROJECT REPORT

Open Framework for Integrated Data Operations (OpenFIDO)

March 2024 | CEC-500-2024-026

PREPARED BY:

David P. Chassin
Duncan Ragsdale
Alyona I. Teyber
SLAC National Accelerator Laboratory

Matthew Tisdale
Katie Wu
Gridworks

Bo Yang
Yanzhu Yu
Hitachi America Laboratories

Kevin Rohling
Presence Product Group
Primary Authors

Qing Tian, Ph.D., P.E.
Project Manager
California Energy Commission

Agreement Number: EPC-17-047

Rey Gonzalez
Branch Manager
ENERGY SYSTEMS RESEARCH BRANCH

Jonah Steinbuck, Ph.D.
Director
ENERGY RESEARCH AND DEVELOPMENT DIVISION

Drew Bohan
Executive Director

ACKNOWLEDGEMENTS

Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory is operated by Stanford University for the U.S. Department of Energy under Contract DE-AC02-76SF00515.

The authors would like to thank Pedram Jahangiri at National Grid for his contribution to the development of OpenFIDO. Thanks also go to our collaborators at Gridworks and the Hitachi America Laboratories staff including Rehana Aziz, Deborah Shields, Yanzhu Ye, and Panitarn (Joseph) Chongfyanprinya, for their efforts supporting the integration of OpenFIDO into their products and services. The authors would also like to thank the engineers at Pacific Northwest National Laboratory, including Jason Fuller, Tom McDermott, and Frank Tuffner, for their contributions. Thanks also go to the software engineers and designers at Presence Product Group, including Natalie Hansen, Jason Monberg, Nick Polkowski, Kevin Rohling, Dane Summers, and Peuan Thinsan.

The authors would like to recognize the team of engineers, administrators, postdocs, graduate students, and interns at SLAC National Accelerator Laboratory and Stanford University for their support, participation, and contributions to this project. Special recognition goes to Natalie Cramar, Steve Chao, Jonathan Goncalves, Veronika Lubeck, Mayank Malik, Jewel Newman, Anna Peery, Supriya Premkumar, Nani Sarosa, Berk Serbercioglu, Derin Serbercioglu, Sara Borchers, Mohamed Nijad, Palash Goiporia, Jimmy Leu, and Fuhong Xie.

Special thanks to Karen Schooler for formatting and editing this document as well as the administration of student research associate assignments and internships, and to Pamela Wright-Brunache for administering the project subcontracts.

PREFACE

The California Energy Commission's (CEC) Energy Research and Development Division supports energy research and development programs to spur innovation in energy efficiency, renewable energy and advanced clean generation, energy-related environmental protection, energy transmission, and distribution and transportation.

In 2012, the Electric Program Investment Charge (EPIC) was established by the California Public Utilities Commission to fund public investments in research to create and advance new energy solutions, foster regional innovation, and bring ideas from the lab to the marketplace. The EPIC Program is funded by California utility customers under the auspices of the California Public Utilities Commission. The CEC and the state's three largest investor-owned utilities—Pacific Gas and Electric Company, San Diego Gas and Electric Company, and Southern California Edison Company—were selected to administer the EPIC funds and advance novel technologies, tools, and strategies that provide benefits to their electric ratepayers.

The CEC is committed to ensuring public participation in its research and development programs that promote greater reliability, lower costs, and increase safety for the California electric ratepayer and include:

- Providing societal benefits.
- Reducing greenhouse gas emission in the electricity sector at the lowest possible cost.
- Supporting California's loading order to meet energy needs first with energy efficiency and demand response, next with renewable energy (distributed generation and utility scale), and finally with clean, conventional electricity supply.
- Supporting low-emission vehicles and transportation.
- Providing economic development.
- Using ratepayer funds efficiently.

For more information about the Energy Research and Development Division, please visit the CEC's research website (www.energy.ca.gov/research/) or contact the Energy Research and Development Division at ERDD@energy.ca.gov.

ABSTRACT

California utilities, customers, consulting engineers and regulators need to exchange power system data, yet this can be cumbersome and error prone, raising barriers to effective resource integration, curbing growth of these resources, and limiting how quickly California can reach its climate change mitigation goals.

OpenFIDO provides a framework for interorganizational data exchange, data and model synthesis, and system performance analysis across multiple power systems tools. OpenFIDO is used to (1) collect data from a wide variety of sources, (2) transfer model and telemetry data between tools, and (3) enable creation of permanently available reproducible results.

OpenFIDO allows utility planners and grid researchers to move data quickly between applications. OpenFIDO supports emerging user groups including system integrators and aggregators who use multiple tools to analyze distributed energy resource grid impacts, and governments and agencies using models for oversight and planning clean energy deployments.

Project results included (1) identifying data analysis requirements needed to support critical electricity use-cases, (2) delivering a new open-source utility data interoperability and analytics workflow management platform at Southern California Edison, and (3) delivering an open-source analytic tool delivery environment for workstation, on-premises server, and cloud infrastructure to enable key utility use-cases for commercialization by Linux Foundation Energy.

Keywords: data curation framework, interorganizational data exchange, distributed energy resource grid impacts, integration/hosting capacity analysis, electrification, tariff design, resilience analysis

Please use the following citation for this report:

Chassin, David P., Duncan Ragsdale, Alyona I. Teyber, Matthew Tisdale, Katie Wu, Bo Yang, Yanzhu Tu, and Kevin Rohling. 2024. *Open Framework for Integrated Data Operations (OpenFIDO).* California Energy Commission. Publication Number: CEC-500-2024-026.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive Summary

The Open Framework for Integrated Data Operations (OpenFIDO) is a data interchange, model synthesis, and analytics support framework that provides access to open-source data and model exchange and analytics tools for power systems researchers, utility planners, and regulators. OpenFIDO was developed to exchange system models, weather data, load, and system telemetry data between various software products as part of the suite of tools widely used by utilities, distributed resource engineers, and regulators in California.

OpenFIDO identified requirements for four electricity use-cases: integration/hosting capacity analysis, electrification, tariff design, and resilience analysis. The system was delivered open-source through Linux Foundation Energy for multi-organization interoperability that runs on individual workstations, on-premises servers, and cloud infrastructure services.

Background

California utilities, customers, consulting engineers, and regulators need to reliably and consistently exchange power system data and models to help the state achieve its climate goals. For example, utility engineers and analysts specifically are required to (1) collect data from a wide variety of sources, (2) transfer model and telemetry data between tools, and (3) create permanently available, reproducible results. The wide variety of sources and various tools causes this process to be cumbersome and error prone, raising barriers to effective resource integration, curbing growth of these resources, and limiting how quickly California can reach its climate change mitigation goals.

OpenFIDO is designed for utility planners and grid researchers that need a tool to move data and models quickly and accurately from one application to another as part of their engineering, planning, and review activities. OpenFIDO supports emerging user groups such as distributed energy resource integrators and aggregators that use multiple tools to analyze and manage the grid impacts of distributed energy resources, as well as government agencies that use these data and models for both oversight and identifying opportunities for new energy and climate policies.

Project Purpose and Approach

The purpose of OpenFIDO is to minimize the time and effort involved in setting up models, assembling data, and running analyses of various scenarios and use-cases with particular emphasis on renewable energy resource integration and energy system decarbonization.

The specific objectives of the OpenFIDO project included the following:

- Identify data exchange and analysis requirements by working with California's investor-owned utilities, Pacific Gas & Electric, San Diego Gas & Electric, and Southern California Edison, and tool vendors through understanding their planning processes

- Develop and test a platform that can use data and models from investor-owned utilities and convert them to those used by emerging data-intensive analysis and agent-based simulation tools
- Demonstrate data exchange and analysis capabilities with the primary use-cases identified by investor-owned utilities.

OpenFIDO implements a scalable data curation framework capable of (1) ingesting data from various sources, including cloud-hosted data systems, power system simulation tools, and web-based endpoints, (2) running data-intensive analyses and agent-based simulations, and (3) delivering data to external users through modern data-exchange infrastructure.

OpenFIDO builds upon several capabilities introduced in the DOE Visualization and Analytics of Distribution Systems with High Penetrations of Distributed Energy Resources (VADER) system (Sevlian 2017), a scalable data management platform. OpenFIDO has three core components that are oriented toward highly efficient, scalable, and customizable data processing: methods, pipelines, and workflows. Each component builds on the previous in a self-similar architecture that facilitates speed, diverse execution environments, and open-source distribution. Methods are primitive data operations needed to perform basic data processing functions, such as obtaining data from remote sources, manipulating data, and delivering data to endpoints for access by other tools. Pipelines are simple data operations that require zero or more inputs, perform multiple data processing steps, and generate one or more outputs. Pipelines are always composed of zero or more methods that must be executed in series. Workflows are complex data operations that require zero or more inputs, perform multiple data processing steps, and generate one or more outputs. Workflows are always composed of two or more pipelines that can be executed in parallel.

OpenFIDO was developed using the following approach:

1. Conduct workshops and interview utility engineers and potential users to gather use-cases and data requirements: Two workshops were held, and interviews were conducted with investor owned, municipal, and cooperative utilities, reliability organizations, government agency users, vendors, and consultants as well as a select group of people who have served in technical advisory groups. The interviews provided insight into the data handling requirements among various tools used in the process of planning distribution systems.
2. Develop an open source platform to support planning process data exchange: The platform comprises the methods, processing pipelines, and analysis workflows described above. The platform also delivers the data handling tools and services to convert data to and from different power system tools used by the utilities, developers, researchers, and regulators.
3. Gather sample datasets: Sample data and model files were collected from the industry to test and validate the framework.

4. Design and develop data architecture for the data interchange platform: Presence Product Group designed the system architecture to meet the needs of users and developers.
5. Develop and test platform features using real-world use-cases: The platform was tested and validated for four use-cases identified by the project team in collaboration with two other California Energy Commission-funded projects, High Performance Agent-Based Simulation (HiPAS) (EPC-17-046) focusing on speeding up the analyses and GridLAB-D Open Workspace (GLOW) (EPC-17-043) focusing on providing a graphical user interface.
    - Integration/hosting capacity analysis: utilities need to complete system-wide hosting capacity analysis to determine the maximum node-level resource integration limit for a circuit to remain within key power system criteria.
    - End-use electrification: utilities need to study the impact of increasing electrification of fossil-fueled end-use loads, specifically heating, cooking, hot water, clothes drying, and light-duty vehicles.
    - Tariff design: utilities need to study the revenue impacts of emerging technologies and planning scenarios that include high penetration of distributed energy resource and end-use electrification.
    - Resilience analysis: utilities need to study the resilience impacts of emerging technologies that support high penetration of distributed energy resource and end-use electrification in the presence of increasing climate change impacts on system operations and planning.
6. Organize technology workshops and plan the commercialization of OpenFIDO.

Key Results

The OpenFIDO project achieved four important results that addressed the challenge of open data analytics for electric utilities:

1. OpenFIDO identified the key data ingestion, curation, and delivery infrastructure requirements to support critical electricity use-cases, specifically integration capacity analysis, electrification, tariff design, and resilience analysis. The capabilities provided by OpenFIDO reduced the time and effort required to collect data needed for various analyses, improved the reproducibility of analyses conducted, and facilitated sharing of results produced by analysis runs in OpenFIDO.
2. OpenFIDO delivered a new open-source utility data analytics workflow management platform specifically designed for interoperability across multiple organizations using workstation-only, on-premises servers, and cloud infrastructure services. The platform is delivered in identical form on a wide variety of computing systems ranging from personal laptops and desktop workstations to on-premises servers and large-scale cloud-deployed infrastructure.

3. OpenFIDO supports an open-source analytic tool delivery environment, including the methods, pipelines, and workflows that support key utility use-cases identified in the project, as well as future products based on as-yet-unidentified use-cases. Testing and validation of OpenFIDO was conducted at Southern California Edison and National Grid.
4. OpenFIDO was commercialized through Linux Foundation Energy as a scalable utility data analytics platform by one or more vendors and cloud services hosting platforms.

The OpenFIDO project met its main objectives, that is, (1) addressing the data exchange and analysis challenges of an investor-owned utility, (2) developing and testing a platform to manage data and models needed by emerging data-intensive analysis and agent-based simulation tools, and (3) demonstrating its application for data exchange and analysis capabilities associated with use-cases identified by investor-owned utilities.

Benefits to California

OpenFIDO enables regulators, researchers, and other interested stakeholders to access and analyze electric grid data that has historically been inaccessible due to the use of proprietary software tools. The data exchange platform increases the transparency of the electric grid and empowers stakeholders to work with utilities and regulators to integrate distributed energy resources more quickly, reliably, and cost-effectively.

In addition, OpenFIDO provides a framework with which all tools used in distributed energy resource financing, planning, and permitting processes can interoperate. The capabilities of OpenFIDO help utilities and other stakeholders more reliably and efficiently exchange system model data with analysts, regulators, and vendors. The reduction in labor intensity and cost of staff training will improve utility staff productivity, help expedite utility resource integration reviews, and simplify utility regulator compliance activities. Collectively, these work reductions will ultimately result in savings to ratepayers.

CHAPTER 1: Introduction

The electric utility industry is highly reliant on data, analytics, models, and simulation to plan and operate distribution systems. Data comes from many sources including customers, system operations, electric markets, and partner organizations. Analytics are used to develop an understanding of the system and assist in producing action plans based on past behavior.

Utilities develop and maintain detailed models of their systems to study their future behavior, plan for growth and adoption of new technologies, and sustain reliable and cost-effective operations. Simulations ensure that devices and equipment continue to function properly and expansion plans meet existing and emerging requirements. Simulators are increasingly used for essential studies such as reliability and resilience analysis, distributed energy resource (DER) interconnection permitting, tariff design, and load electrification impact studies, among other things.

Climate change and policy responses are also changing analysis demands for these simulation tools. There is an increasing amount of environmental, advanced technology, and consumer behavior data that needs to be incorporated in simulations for results to remain valid. For example, satellite and light detection and ranging (LiDAR) data of vegetation changes in proximity to power lines is now available for all of California from California Forest Observatory (www.forestobservatory.com). This data can be fused with network models to determine which power lines are at risk from vegetation falls during high-wind events. Similarly, the availability of automated meter infrastructure (AMI) data enables nearly real-time generation of customer-class load shape data for use in tariff, electrification, and customer load clustering analysis.

Existing methods of various power system analyses are less valid as key assumptions hold less frequently. The system analyses made by the utility are obsolete or evolving quickly in the face of changing and emerging requirements. For example, reliability studies do not consider the role of battery storage in performance metrics like System Average Interruption Duration Index System Average Interruption Frequency Index, and Customer Average Interruption Frequency Index where customers can provide backup power in ways that are not considered in the metrics' original purpose, which is to provide an aggregate measure of how long customers are without an energy source.

Emerging conditions as higher penetration of DERs are adopted have high uncertainty and are often more data intensive. This is particularly true for studies involving the environment surrounding utility equipment, such as public safety power shutoffs (PSPS). Access to vegetation and weather data is vital to access the long-term performance of distribution systems, as well as forecasting near- and long-term interactions between vegetation and power systems. This data is many orders of magnitude larger and more difficult to manage than traditional utility data, including Supervisory Control and Data Acquisition (SCADA) and AMI data.

Data analysis is also an increasingly simulation-centric activity. Historically, many utility analytics used reduced-order models that do not require highly detailed asset data. This had a significant advantage because utilities did not need to use simulations to run these detailed models and typically could avoid the need for Monte-Carlo simulations to determine the possible ranges of outcomes. With the advent of agent-based simulations that can perform quasi-steady time-series simulations on high-performance computing infrastructure, running large-scale simulations many times is now possible, and new kinds of analysis methods are emerging that cannot be performed using reduced-order models. For example, comprehensive integration capacity analysis (ICA) requires running multiple time-series simulations with different levels of resources to determine when voltage regulator controls become unable to compensate for local generation resources.

OpenFIDO is designed to support a growing reliance on simulation at scale to provide important analysis results for climate change response and technology integration challenges. By allowing researchers, vendors, and users to access an open-source library of analysis methods, tools, and workflows, OpenFIDO provides an environment in which methods can be developed, evaluated, and deployed easily and flexibly while reducing concerns about ongoing support, licensing fees, or compatibility.

CHAPTER 2: Project Approach

OpenFIDO provides a comprehensive open-source data integration and exchange approach pioneered on the open-source VADER platform developed at Stanford University (Chassin 2014). OpenFIDO provides a flexible data analytics and visualization environment for multiple data sources, including AMI data, distribution system models, and environmental data. OpenFIDO can be installed in secure cloud infrastructure, on-premise data centers behind enterprise firewalls, and on a standalone system on a workstation or laptop computer. OpenFIDO is used to facilitate the use of a common collection of DER planning and analysis tools. OpenFIDO already supports a number of input data formats, general comma-separated value (CSV) and JavaScript Object Notation (JSON) data that enable large data ingests such as AMI, SCADA, and climate data, as well as network models from distribution modeling tools such as GridLAB-D and CYME.

Concept of Operations

OpenFIDO was developed with a multi-organization concept of operations designed to allow users across multiple organizations to import, analyze, and share results with a high degree of access control granularity independent of the type and structure of the data. The complexity of integrating these concepts into one area of operation was challenging to deliver in a general purpose, open-source, and highly secure platform. The focus of the development and validation of OpenFIDO up to this point has been to ensure and verify high reliability, security, and flexibility for the users.

OpenFIDO uses a concept of operations founded on the following principles, which are detailed in the subsections below.

1. **Platform portability:** OpenFIDO runs on Docker, an open platform for running applications that enables separation of the application from the host system on which it is running.
2. **Open-source distribution:** OpenFIDO is completely open-source and may be used by anyone for any purpose without licensing fees to the creators of the platform.
3. **Scalable computing:** OpenFIDO can run on most platforms that support Docker, and thus can be run on anything from a laptop computer to ultra-large scale cloud infrastructure.
4. **Open analytics:** OpenFIDO methods, pipelines, and workflows are available on GitHub and can be published and used by anyone at any time on all deployments of OpenFIDO.
5. **Application-level security:** OpenFIDO requires full authenticated access to all endpoints at all times. Users must login to access data artifacts and analysis results.

6. **Endpoint interfaces:** All application endpoints may be accessed using standard URLs, provided access is granted to the requestor.
7. **Separate formats and semantics:** All data can be delivered in multiple formats without consideration of data semantics, that is, data transmitted can be transmitted as CSV or JSON regardless of the contents, for example, AMI, SCADA, or weather, and vice-versa. Another example is AMI data can be transmitted as either CSV or JSON, as determined by the recipient.
8. **Organization-oriented access control:** All users and data may be partitioned and access controlled based on membership to organizations within the platform, for example, different jurisdictions, different utilities, different groups within a utility, and different teams within a group.
9. **Navigation tools:** Users and administrators can access and manage previous analysis results, view graphs and download data for local use in writing reports, and manage settings such as names, descriptions, passwords, and access control lists.

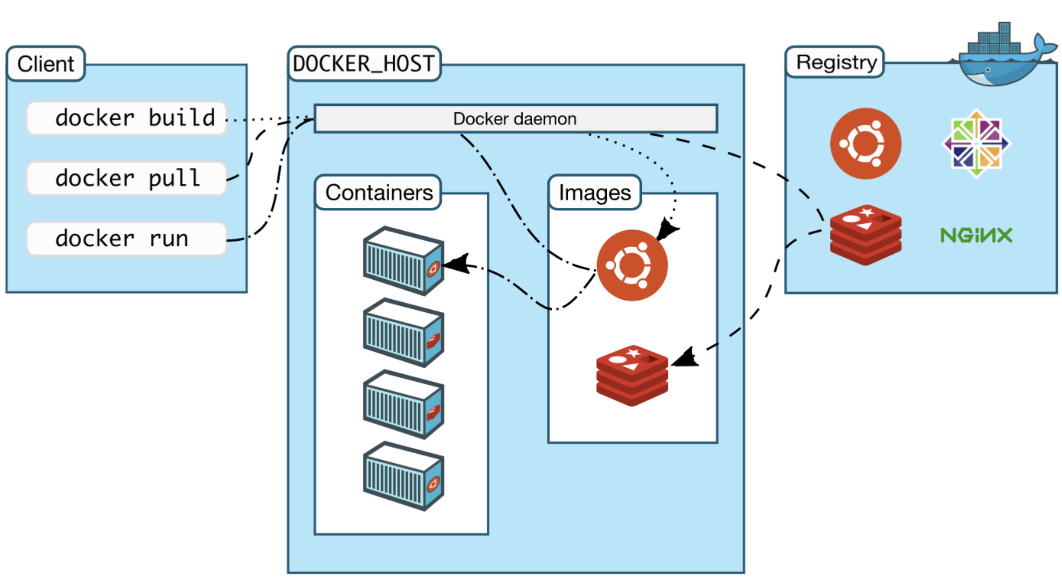
In addition, a number of synergistic activities were undertaken to encourage adoption of the OpenFIDO support for advanced simulation tools by the user community.

Platform Portability

OpenFIDO runs on Docker, an open-source platform for running applications that enables separation of the application from the infrastructure on which it is running. Docker is supported on most host systems, including ApplemacOS, Microsoft Windows, and Linux systems.

Docker uses a client-server architecture shown in Figure 1, where the Docker **client** talks to a Docker **daemon** that does all the work of building, running, and distributing Docker containers. The Docker client and daemon can run on the same system, or users can connect a Docker client to a remote Docker daemon. The client and daemon communicate via a Representational State Transfer Application Programming Interface (REST API) using Unix sockets or a network interface. Another widely used Docker client is **Docker compose**, which is used to build applications that use sets of containers that are coordinates. This is how OpenFIDO is built and delivered so that it provides consistent support for the widest possible range of host systems.

Figure 1: Docker Client-Server Architecture



Source: Docker

Open-Source Distribution

OpenFIDO is completely open-source under a BSD-3 License and may be downloaded and used without licensing fees to the creators of the platform. Content creators for OpenFIDO may choose to use OpenFIDO to provide additional methods, tools, and/or data for a fee, but the underlying system will always be available at no cost to the user.

The following open-source resources are publicly available on GitHub.

- http://github.com/openfido/ (GitHub Organization)
    - openfido-client: main client image
    - openfido-utils: service utilities image
    - openfido-app-service: application services image
    - openfido-auth-service: authentication services image
    - openfido-workflow-service: workflow manager images
    - cli: command line interface
    - cyme-extract: CYME to GridLAB-D model conversion pipeline
    - resilience: Grid Resilience Intelligence Platform anticipation analysis pipeline
    - gridlabd: general HiPAS gridlabd simulation, data, and model converters pipeline
    - ica-analysis: integration capacity analysis pipeline
    - loadshape: load shape analysis pipeline
    - weather: weather forecasting and historical data access pipeline
    - census: TIGER shape files for census regions and tract mapping pipeline
    - address: general address resolution for geographic data pipeline
    - tariff_design: tariff design analysis pipeline
    - electrification: electrification analysis pipeline

Scalable Computing

The long-term goal of OpenFIDO is to support all major cloud platforms, enable on-premises deployment, and workstation/laptop operations. Scalability and portability are therefore primary design considerations throughout the development of the OpenFIDO platform. All aspects of OpenFIDO were integrated within a containerization environment using the Docker toolset. This allows the full functionality of the platform to run across a wide array of deployment contexts, including on-premise servers, workstation, laptop, and cloud systems.

Additionally, while the initial implementation was deployed to the Amazon Web Services (AWS) cloud environment, the development of the platform has avoided any vendor lock-in to AWS-specific technologies. As such, OpenFIDO can run within any environment that supports containerization technologies including Azure, Google Cloud Platform Kubernetes, a single user's desktop, etc. Proof of this portability was established and successfully tested via the development of a single integrated container that can be easily pulled and run with minimal configuration.

Open Analytics

OpenFIDO methods, pipelines, and workflows are available on GitHub and can be published and used by anyone at any time on all deployments of OpenFIDO. OpenFIDO users may specify the environment (DockerHub Repository), pipeline code (Git Clone URL), and version (Repository Branch) of any pipeline they use, as shown in Figure 2.

Figure 2: Pipeline specification

Source: https://github.com/openfido

Application-Level Security

OpenFIDO requires full authenticated access to all endpoints at all times. Users must login to obtain access to results. This security is provided above and beyond the security provided by the hosting platform, for example, Amazon AWS, Google Cloud, Microsoft Azure, on-premises services, or the local workstation firewalls and user credentials.

Endpoint Interfaces

All application endpoints may be accessed using standard URLs, provided access is granted to the requestor. Endpoints are generated for each input and output artifact using a globally unique identifier. These endpoints may have limited or unlimited lifetime. The current release of OpenFIDO provides only limited duration endpoints to ensure long-term security of the data. It has not been determined which future release of OpenFIDO will support unlimited endpoint lifetime or how data managers will manage the duration.

Separate Formats and Semantics

To achieve open analytics, OpenFIDO uses common files formats, for example, GLM, CSV, and JSON, and semantics, for example, labeled columns CSV, GridLAB-D models and templates, to ensure that data and models can be moved from tool to tool with minimal manual intervention.

Data is generally transferred using CSV and JSON file formats, depending on whether the data can be represented in 2D tabular form. File semantics include data with and without labels. When labels are not present, method code may use Python libraries such as NumPy or Pandas modules. Only Pandas modules may be used on labeled data.

Models must be tagged with semantic labels, including "application," "version," and all tags required by the application version indicated. Only JSON file formats may be used for models. Some OpenFIDO pipelines may support other file extension formats, for example, GLM, MDB, and additional semantics, for example, CYME, OMF. But these pipelines generally serve as file format converters because they either only output JSON or only accept JSON input.

All methods, pipelines, and workflows are published with support files to assist in use and deployment. These include the following:

- openfido.sh: the default script to run a pipeline
- openfido.json: the default configuration for a method
- manifest.json: the pipeline run data requirements script
- requirements.txt: the list of required python modules

Note that at the current time, workflows are not directly supported by GitHub repositories in the OpenFIDO user interface and can be managed only using the OpenFIDO API.

Organization-Oriented Access Control

All users and data may be partitioned and access controlled based on membership in organizations specified within the instance of OpenFIDO, for example, different jurisdictions, different utilities, different groups within a utility, or different teams within a group.

Navigation Tools

Users and administrators can access and manage previous analysis results, view graphs and download data for local use in writing reports, and manage settings such as names, descriptions, passwords, and access control lists.

Administrators may invite users with an email address. When the user accepts the invitation, they are added to the authorized user list and granted password-controlled access. Administrators can also delete users. Although the user's access is revoked, data associated with the user is retained and remains available to other users if access is granted.

Administrators may also create organizations and appoint users to administer these organizations, with the same invite/delete/create administrative rights limited to that organization.

CHAPTER 3: Results

The OpenFIDO project achieved four important results that addressed the challenge of open data analytics for electric utilities:

1. OpenFIDO identified the key data ingest, curation, and delivery infrastructure requirements to support critical electricity use-cases, specifically ICA, electrification, tariff design, and resilience analysis. The capabilities provided by OpenFIDO reduce the time and effort required to collect data needed for various analyses, improved the reproducibility of analyses conducted, and facilitated sharing of results produced by analysis run in OpenFIDO.
2. OpenFIDO delivered a new open-source utility data analytics workflow management platform specifically designed for interoperability across multiple organizations using workstation-only, on-premises servers, and cloud infrastructure services. The platform is delivered in identical form on a wide variety of computing systems ranging from personal laptops and desktop workstations to on-premises servers and large-scale cloud-deployed infrastructure.
3. OpenFIDO supports an open-source analytic tool delivery environment, including the methods, pipelines, and workflows that support key utility use-cases identified in the project, as well as future products based on as-yet-unidentified use-cases. Testing and validation of OpenFIDO was conducted at a major California IOU and with a New York system operator.
4. OpenFIDO was commercialized through LF Energy as a scalable utility data analytics platform by one or more vendors and cloud services hosting platforms.

The OpenFIDO project met its main objectives, that is, (1) addressing the data exchange and analysis challenges for IOUs, (2) developing and testing a platform to manage data and models needed by emerging data-intensive analysis and agent-based simulation tools, and (3) demonstrating its application for data exchange and analysis capabilities associated with use-cases identified by IOUs.

The application services were deployed and validated on AWS under the openfido.org domain. In addition, ten pipelines were made public for OpenFIDO users, including the four main use-cases, hosting capacity, tariff design, electrification, and resilience analysis, which use HiPAS GridLAB-D templates. In addition, the following pipelines were validated and deployed to OpenFIDO: distribution system simulation, load shape clustering, weather data, census geography, address resolution, and CYME model extraction.

The real-time validation status of both the application services and pipelines can be viewed at the main OpenFIDO repository page located at https://source.openfido.org/. The application services build, validation, and deployment status are shown in Figure 3.

Figure 3: Application Status

App Service	Auth Service	Workflow Service	Client Service
Test-build passing	Test-build passing	Test-build passing	Test-build passing
Deploy Staging passing	Deploy Staging passing	Deploy Staging passing	Deploy Staging passing
Deploy Production passing	Deploy Production passing	Deploy Production passing	Deploy Production passing

Source: https://github.com/openfido

The pipeline validation status is shown in Figure 4. (Note: failures are shown for illustrative purposes. In the final release of OpenFIDO, these pipelines are passing.)

Figure 4: Pipeline Status

Pipeline	Status
Tariff Design	validation failing
Loadshape	validation passing
Weather	validation passing
HiPAS GridLAB-D	validation passing
Census	validation passing
Resilience	validation failing
Hosting Capacity	validation passing
Electrification	validation failing
Address	validation passing
Cyme Converter	validation passing

Source: https://github.com/openfido

Use-case pipelines (hosting capacity, tariff design, electrification, and resilience) are additionally validated by the HiPAS GridLAB-D template (see https://github.com/slacgismo/gridlabd-template). In addition, load shape clustering is implemented as a HiPAS GridLAB-D template. The remaining pipelines are implemented as HiPAS GridLAB-D geodata packages (for example, Census, Address), converters (for example, Cyme extract), and subcommands (Weather). HiPAS GridLAB-D itself is validated as a complete tool (see https://github.com/slacgismo/gridlabd for details).

Synergistic Activities and Use-Cases

OpenFIDO was developed in coordination with and in support of two other California Energy Commission (CEC) projects that sought to advance the state of the art for power system simulation and analysis within the utility industry. These projects in consultation with their TACs, as well as two U.S. Department of Energy (DOE) projects, and one industry association recommended the use-cases for the development of OpenFIDO. In this section, the use-cases originating in these projects are described.

CEC High-Performance Agent-Based Simulation

CEC funded the High-Performance Agent-based Simulation (HiPAS – EPC-17-046) effort to modernize GridLAB-D, an open-source agent-based simulator developed by the DOE at Pacific Northwest National Laboratory (Schneider 2011). GridLAB-D has been used to study utility problems such as conservation voltage reduction under changing load compositions (Sevlian 2017). GridLAB-D has emerged in recent years as a powerful tool for studying emerging utility problems such as renewable resource integration, rate design in the presence of high penetrations of renewable resources, and high end-use load electrification impacts on distribution systems.

Tariff Design

HiPAS GridLAB-D's effectiveness and capabilities are demonstrated using a Tariff Design use-case, which explores the impact of new rate designs under California IOUs and Community Choice Aggregation Programs. This template allows for the user to execute a full consumer cost analysis on commercial, residential, or industrial loads. This use-case requires a significant amount of data ingestion and user input, which is automated by OpenFIDO. The pipeline allows the modeler to input the rate design structure, as well as regional preferences, weather, load characteristics, and distribution system characteristics. OpenFIDO, then leverages the analytics built into GridLAB-D to generate customer billing profiles based on the given specifications. The advantage of this process is that it can be automated to study a wide range of electricity rates for a variety of load profiles within the same distribution network.

Load Electrification

California is at the forefront of deep electrification. To prepare the grid for an ambitious decarbonization goal, the electrical utility stakeholders are seeking tools and strategies that help mitigate the effects of building and transportation electrification on the distribution grid. In combination, HiPAS and OpenFIDO allow the user to explore different scenarios of electrification and the impacts on the grid. These tools can give insight into management of coincident and non-coincident load peaks, as well as reveal the magnitude of the afternoon ramp in the high-electrification scenario. OpenFIDO allows for ingestion and manipulation of data to solve challenging electrification forecasting issues. The typical data sets that enable the use-cases described below involve appliance loading data, census data, and location services data to name a few.

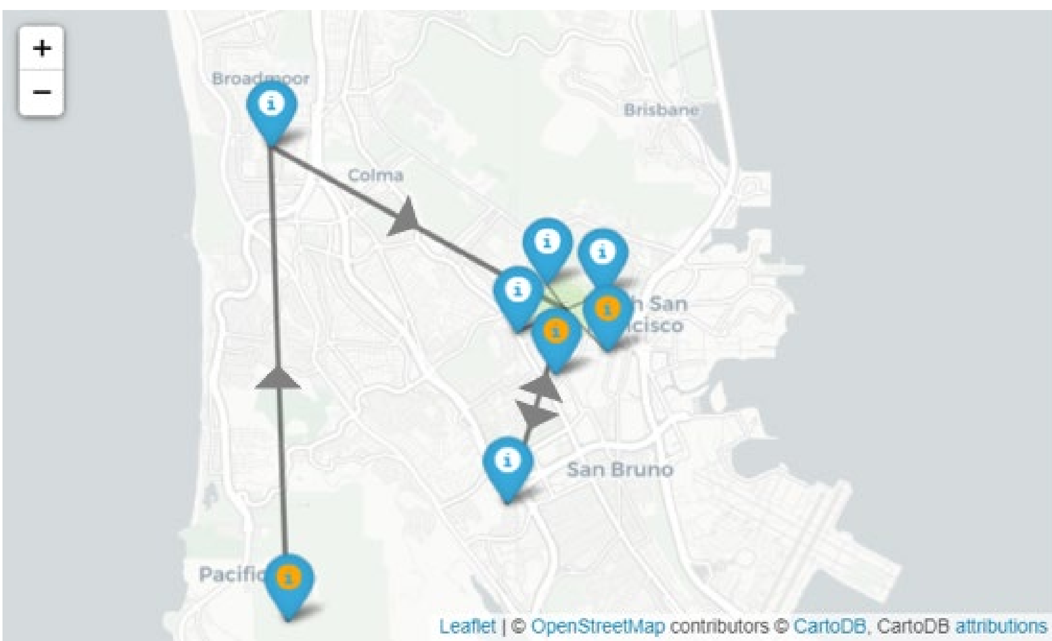
Modeling Impacts of Increasing Residential Building Electrification:

Incremental impacts of building electrification on the distribution system were considered by exploring the replacement of natural gas appliances in the following four categories: heating/cooling/ventilation, water heating, cooking, and clothes drying. The use-case does not consider a one-to-one replacement of the natural gas heating unit due to the likelihood of the unit being replaced with heat-pump technology, which can increase cooling load in the summer temperature highs that was not present prior to electrification. In addition, the scenario considers the thermal integrity of the buildings, which presumes that the building being modelled is upgraded during electrification to account for items such as additional insulation, which changes the behavior of the load. Modeling the explicit parameters and accounting for electrical and thermal effects using GridLAB-D infer the propagating effects of the load on the distribution grid. The tool leverages a number of additional house definition parameters, including thermostat settings, house size, thermal integrity, appliance contents, and typical usage specifications. The use-case deploys the template configuration within HiPAS GridLAB-D and uses OpenFIDO framework to input the necessary parameters modeled in the use-case.

Modeling Impacts of Transportation Electrification:

Capturing the grid impacts due to deep electrification of the transportation sector can be achieved using the OpenFIDO platform by analyzing the human behavior through location services data to model the expected driving profiles in a fully electric transportation sector as illustrated in Figure 5.

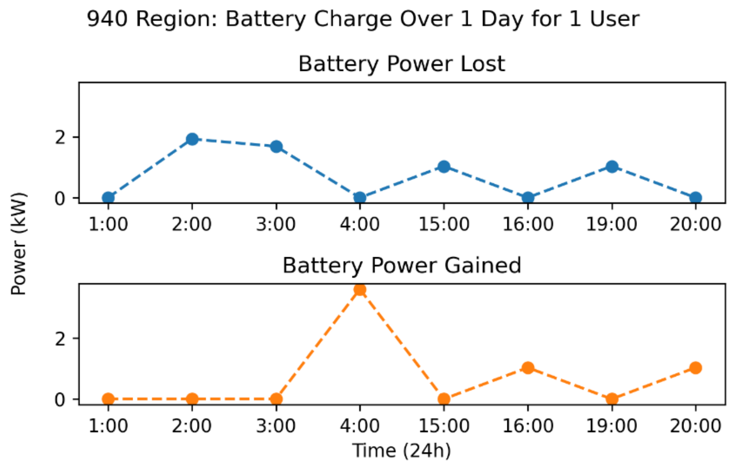
Figure 5: Sample Travel Route Between Two Zip Codes as Identified by the Location Data



The results include time spent at each location and estimated electric energy required by the vehicle for the trip (Newman 2020).

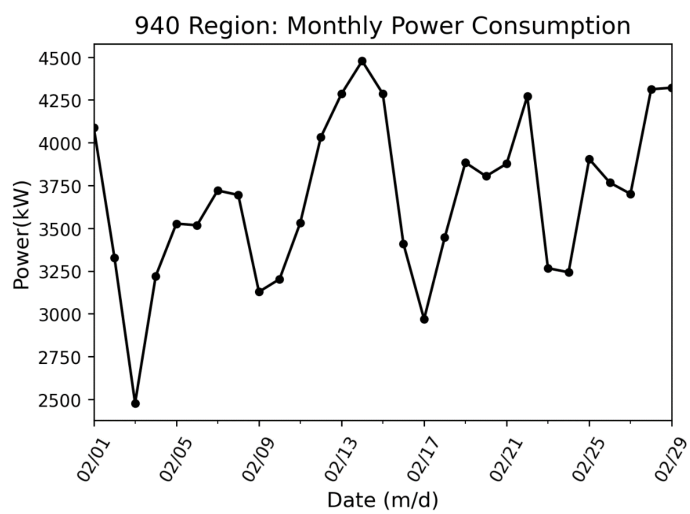
Source: https://github.com/openfido

Figure 6: Estimated Charging Profile Given Duration and Driving Profile



Source: https://github.com/openfido

Figure 7: Estimated Power Consumption by a Specific Zip Code Given Driving Profiles Defined in Location Services Database



Source: https://github.com/openfido

This information, shown in Figures 6 and 7, is pre-processed using OpenFIDO and was combined with grid data to determine the baseline charging profiles and give rise to the constrained optimization problem to advise the grid-optimal charging locations. This in turn affected the driving / charging behavior as part of HiPAS GridLAB-D analytics.

CEC GridLAB-D Open Workspace (GLOW)

CEC funded the GLOW project to deliver a web-based distribution planning platform for GridLAB-D. The open-source graphical user interface of GLOW aims to augment the command-line interface for GridLAB-D in a more intuitive, user-friendly manner, contributing to wider use of the simulation technology. The purpose of the graphical interface is to make electricity modeling software accessible for all parties that might be interested, not just those proficient in data manipulation and command line programs. GLOW brings GridLAB-D to a new base of users and allow for easier testing, application, and expansion of the tool. GLOW shares the same ICA methodology as HiPAS GridLAB-D and OpenFIDO.

Integration Capacity Analysis

In California, IOUs are required to complete system-wide ICA to determine the maximum node level hosting capacity for a circuit to remain within key power system criteria. System ICA results are used to expedite interconnection permitting for DER additions. ICA with DER growth scenarios are also required for all annual IOU distribution system planning processes. In addition, a CPUC ruling in January 2021 added new requirements, including support for customers adding electric vehicle chargers and reducing natural gas usage, reducing the amount of data redacted from ICA maps, improving ICA data practices to avoid undetected errors, and more broadly improving the ICA process.

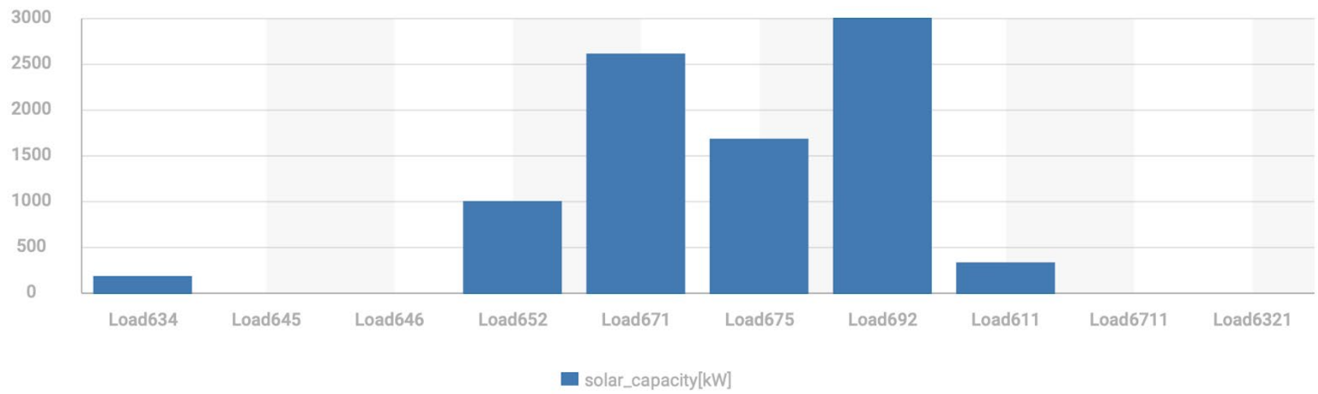
Table 1: Properties That May be Checked for Violations During ICA

Class	Property
Underground & Overhead Lines	Current Rating (summer)
	Current Rating (winter)
Transformers	Power Rating
	Top Oil Rise
	Winding Hot Spot Rise
Regulator	Raise Taps
	Lower Taps
	Continuous Rating
Triplex Meter	Nominal Voltage
Meter	Nominal Voltage

Source: SLAC National Accelerator Laboratory

The ICA program on the OpenFIDO platform quantifies the potential DER generation that can be connected without violating distribution system constraints. The process uses a system-wide iterative power flow simulation. The DER generation level is varied at each node, independently. At each iteration, the simulation checks all lines, transformers, regulators, and meters for constraint violations. For details on the properties that are checked against constraints, see Table 1. If a violation occurs, the details of the violation are recorded. The ICA value for that load bus is the maximum power injection associated with any violation criteria, as shown in Figure 8.

Figure 8: Maximum Power Injection at Each Load Bus of Institute of Electrical and Electronics Engineers (IEEE)-13 Model



Source: SLAC National Accelerator Laboratory

DOE Advanced Load Modeling

The DOE Office of Electricity Advanced Grid Modeling Program funds an ongoing effort to provide utilities and reliability organizations advanced load model data in support of annual interconnection reliability planning studies performed by the North American Electric Corporation (NERC). This program includes efforts to (1) study the heating and cooling sensitivity nationwide to understand the effect of the shift away from resistive heating toward heat-pumps, (2) redevelop and deploy the commercial and residential building model originally created for the Western Electricity Coordinating Council, and (3) create models to support electrification impact studies, including decarbonization of residential end-uses and vehicle charging infrastructure growth scenarios.

LoadInsight

The US DOE's Office of Technology Transitions funded the commercialization of LoadInsight to leverage the benefits of commercial data analytics platforms like VADER and Open Modeling Framework by integrating advanced load data analytics developed for DOE under previous projects conducted at SLAC, Stanford, Pacific Northwest National Lab, and Lawrence Berkeley National Lab. For example, Visualization and Insight System for Demand Operations and Management uses AMI data to address a vital utility need to perform more accurate load forecasts and design efficient and fair tariffs. This is achieved by identifying the most common load shapes found among their customers and determining how many customers correspond to each load shape. Utilities can then maximize their customer program impacts while minimizing their operational impacts. Utilities can quickly identify how many and which customers would benefit from an evening time-of-use (TOU) rate rather than midday TOU rate. Similarly, the Advanced Load Modeler (ALM) from the DOE Office of Electricity's Advanced Grid Modeling Program uses SCADA data to identify the load compositions NERC requires for the composite load models used in transmission reliability and resource adequacy studies. The merger of VISDOM and ALM tools allow each to validate the other to improve overall accuracy of analysis results.

The OpenFIDO platform is based on the same architecture as LoadInsight and has many common implementation elements including the same overall user experience, similar data ingestion, job control, and data delivery infrastructure, and the same user credentials and access control mechanisms. While LoadInsight's target use-case is primarily centered on end-use load analysis, OpenFIDO is much more general as it supports many other types of analyses needed by utilities beyond just the end-use load shape, load composition, and load weather and/or price sensitivity supported by LoadInsight. Many of LoadInsight's capabilities were adopted by OpenFIDO, and vice-versa. As these products both mature and progress they will likely converge while maintaining their distinct target audience and branding.

DOE Grid Modernization Lab Consortium (GMLC) Grid Resilience Intelligence Platform

Extreme weather events such as Atlantic hurricanes and regional windstorm events pose an increasing threat to the nation's electric power systems and the associated socio-economic systems that depend on reliable delivery of electric power. While utilities have software tools available to help plan their near- and long-term operations, these tools do not include capabilities to help them plan for and recover from extreme events. Software for resilience-by-design and fast recovery is not widely available commercially, and research efforts in this area are preliminary. With this in mind, DOE GMLC funded a multi-lab initiative to develop the Grid Resilience and Intelligence Platform (GRIP) to develop and deploy a suite of novel software tools to anticipate, absorb, and recover from extreme events. The tools are integrated into an extensible and open platform, aiding future efforts in this area. This research aids the DOE Office of Electricity and Energy Efficiency and Renewable Energy Resilient Distribution Grid research and development mission of developing cutting edge resiliency technologies that are deployed to utilities and reducing outage costs to meet the Multi-Year Program Plan goal and DOE major outcome of a "10 percent reduction in the economic costs of power outages by 2025."

The project uses HiPAS GridLAB-D as the baseline power flow solver for the platform. The GRIP project enhanced the existing powerflow solver to include a pole failure and vegetation vulnerability solver that allow the user to evaluate vulnerability metrics for grid assets during extreme weather conditions.

Pole Failure:

Using the Southern California Edison pole data format and the capabilities from the CYME to GridLAB-D converter OpenFIDO ingests both data sources and links the pole data to the electrical grid distribution models from CYME and SPIDAcalc to compile a model populated with pole parameters. The output of this use-case illuminates the vulnerable components of the grid and their failure timelines, which can be used to plan and dispatch preventive action.

Vegetation Analysis:

Using Salo Sciences database generated from the California Forest Observatory, which contains information regarding grid-asset surrounding vegetation, OpenFIDO ingests LIDAR data such as tree height, vegetation density, and surface fuels and integrates it with HiPAS

GridLAB-D analytics to generate geographic data, which identifies asset vulnerabilities due to vegetation overgrowth and decline.

National Rural Electric Cooperative Association (NRECA) Open Modeling Framework

The NRECA deployed GridLAB-D as part of the Open Modeling Framework (OMF). The OMF data format is called OMD, and it is a JSON representation of a GLM file similar to the standard GridLAB-D JSON file with additional data related to the OMF interface. OpenFIDO can convert GLM and JSON models to and from OMF using the NRECA OMF file input/export functionality. Although this is currently supported through HiPAS GridLAB-D, the method for this conversion has not yet been deployed in an OpenFIDO GitHub repository. This will likely be included in a future release of OpenFIDO.

Technology/Knowledge Transfer/Market Adoption (Advancing the Research to Market)

The project team developed a combined Knowledge Transfer Plan for OpenFiDO, GLOW, and HiPAS. This Knowledge Transfer Plan aims to provide a platform for lessons learned and knowledge gained over the course of the three interrelated initiatives.

The target audiences include the staff at the CEC, the CPUC, load-serving entities, DER vendors, public interest organizations, research institutions, and other interested parties.

The Knowledge Transfer Plan includes project fact sheets and presentation materials for public distribution. Project fact sheets and presentation materials shared in TAC meetings or ad hoc workshops and webinars were all published on the project website hosted by Gridworks.

Over the course of the project, the team held biannual public meetings with the TAC in the spring and fall. The meetings were an opportunity to update the TAC and other interested stakeholders on the project status, including sharing methods, validation models, and model results. Over five TAC meetings between 2018 and 2020, 172 people participated, heard updates, and provided input on the development of OpenFIDO.

Training materials were developed and deployed when the OpenFIDO project was completed. To maximize efficiency, training materials were developed and deployed in concert with materials to support GLOW and HiPAS.

OpenFIDO was submitted for adoption by LF Energy, which is an open-source foundation focused on the power systems sector and hosted within the LF. LF Energy provides a neutral, collaborative community to build the shared digital investments needed to transform the world’s relationship to energy.

LF Energy brings together stakeholders to solve the complex, interconnected problems associated with the decarbonization of energy by using resilient, secure, and flexible open-source software. The digitalization of power systems enables the abstraction of the world’s largest machine into composable software defined infrastructure. Digitalization also means that operators can “network electrons” by orchestrating the metadata about an electron in ways never before possible. Digitalization facilitates a radically energy-efficient future. When every electron counts, renewable and distributed energy provides humanity with the tools to address climate change by decarbonizing the grid, powering the transition to e-mobility, and supporting the urbanization of world populations.

LF Energy leverages transparent, open-source development best practices, along with existing and emerging standards, to efficiently scale, modernize and digitally transform the power systems sector. By providing frameworks and reference architectures, LF Energy minimizes toil and alleviates pain points such as cybersecurity, interoperability, control, automation, virtualization, flexibility, and the digital orchestration and balancing of supply and demand.

Key Findings

National Grid Load Forecasting Study

OpenFIDO's performance using the distribution system simulation pipeline is based on the results of running the National Grid 2022 load forecasting study using HiPAS GridLAB-D. Three metrics of performance were evaluated during this study and compared to the same study conducted by National Grid the previous year using an older version of GridLAB-D without OpenFIDO. The metrics and performance results are as follows.

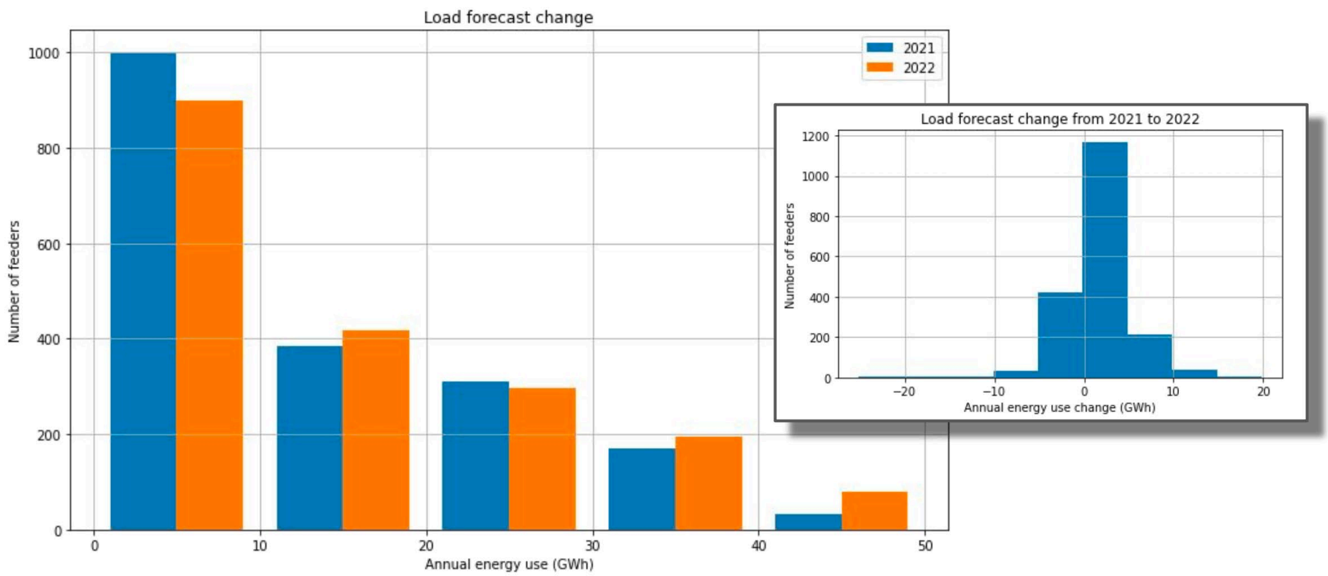
Accuracy of Results

A total of 1,920 feeder models in CYME were converted using the CYME to GLM converter, and 1,871 of these models converted successfully without manual intervention (97.5% success). In the model requiring manual intervention, the follow problems were observed with the CYME models, all of which required manual corrections based on engineering judgement and consultation with the modelers:

- Load objects had nominal voltages that did not match the load bus on the network.
- Capacitor switches settings, phases, and control did not match or were incorrectly set.
- Regulator bandwidth units and/or values were incorrect.
- Load magnitudes exceeded reasonable/feasible values for the given phases.
- Photovoltaic generation exceeded hosting capacity limit for the feeder.
- Triplex lines were not modeled correctly resulting in loss discrepancies.
- Loads were not connected resulting in kilovolt amps discrepancies.
- Assets were not connected to the correct phases.
- Transformer modeling errors resulting in incorrect secondary voltages.

The load forecasts of individual feeders' total annual energy usage were compared with the previous year. The mode of the forecast change was found to be between 0 and 5 GWh/y (<1 percent) with more than 99 percent of the changes between -10 and +15 GWh/y, as shown in Figure 9.

Figure 9: Load Forecast Energy Use Change for New York from 2021 to 2022.



Source: SLAC National Accelerator Laboratory

Simulation Speed

The annual load forecast requires an 8,760-hour simulation of each feeder. The 2021 study was run using the original version of the GridLAB-D software without OpenFIDO and required 114 hours to complete on the AWS Windows servers for which it was designed. The same study was run on an AWS Linux server with the new version of GridLAB-D that OpenFIDO uses and required less than 3.5 hours to complete.

A speed test of hosting capacity analysis was complete for both the 25 DOE Taxonomy Feeders and 476 of the National Grid feeder. The following performance metrics were identified.

Table 2: Hosting Capacity Analysis Performance

Study	Runtime seconds as a function of nodes in model (kilonodes)	Runtime seconds as a function of line in model (kilobranches)	Runtime seconds as a function of number of DERs
Taxonomy Feeders	$t = 52n^2 - 4.9n$	(na)	$t = 0.57n + 2.95$
National Grid Feeders	$t = 1.35n$	$t = 1.08n^2 - 0.18n$	(na)

Source: SLAC National Accelerator Laboratory

Cost of Operations

The National Grid study in 2021 required 17 terabytes (TB) of storage to complete. The 2022 required 1.1 TB. The reduction was mainly attributed to a significant reduction in the number of warnings generated by conditions resulting from CYME to GLM converter problems.

The average runtime of the 2021 study was roughly 25,600 hours while the average runtime of the 2022 study using the load forecasting pipeline was less than 4.5 hours. The AWS host was changed from Windows to Linux (version: c5a.24xlarge). This resulted in a cost reduction from roughly $113,000 in 2021 to only $20.25 in 2022. The workflow was also changed from a sequential run to an optimized parallel run.

Recommendations

AWS Automated Deployment

One potential avenue of improvement for the OpenFIDO platform would be to generate an interactive front end for organizations to use to automatically deploy the application on their own AWS infrastructure.

This would require the development and implementation of SLAC's own Terraform script with an external configuration reference system, as well as a web page to allow prospective users and organizations to familiarize themselves with the platform and the deployment requirements.

The platform will provide users a configuration form that, when filled out and submitted online, will automatically deploy the entire AWS infrastructure for them.

The platform will also support a database table to preserve basic use-case information (for example, organization name and basic contact information, as well as general logging). It will not save sensitive information, such as AWS account credentials, nor will it process any financial transactions. Such information is only needed for one-time use to deploy the application. All users must have their own AWS accounts to use this service.

Additional features will include the option to download a copy of their submitted configuration (default will be to save a copy, however it can be toggled off). Users will also be able to upload a saved configuration file.

UI System Upgrade

The OpenFIDO user interface (UI) can also see additional improvements in multiple areas.

Workflow UI

One primary area of interest would be adding in a workflow UI to support chaining together data from multiple workflows. Currently, once a user submits a workflow, they will need to download the output artifacts and then re-upload them into any follow-on workflows.

This UI would eliminate that requirement. A simplified implementation would allow a user to select a saved workflow run while setting up a new pipeline run, and it would automatically upload all the saved workflow artifacts as inputs into the new run.

A more advanced implementation could see a user creating a pipeline chain. This would include adding a dedicated interface, where users can add multiple pipelines in a linear processing order.

This system would add a pipeline-outputs display to include all expected output files that will be passed to the next pipeline in the chain, along with a pipeline-inputs display to show all necessary input files.

The pipeline inputs display would show which inputs will be met from a previous pipeline step in the chain. Then, all the users would need to do is upload any additional dependencies for each step in the chain and click a "run chain" button.

Environment Variable Support

Another area of improvement for the workflow process and UI would be to include an environmental variables section in either the user's profile, or as a submittable file to a pipeline run.

Many potential pipelines may depend on data from a resource requiring credentials. This system would allow users to provide credentials or any other necessary environmental data to process their pipelines.

A profile-based implementation would create the easiest user experience, as they would only need to set their environment variables once in their user account for them to be applied to all workflows. However, a drawback to this approach would mean the need to store potentially sensitive credentials.

The alternative approach of allowing users to submit a standardized ENV file, an extension handling environment variables, would require more work from the user end but would alleviate the burden of protecting sensitive information from the system. This approach would create a specialized set of logic to handle submitted ENV files, allowing their use in a workflow run while removing the file as a saved artifact from both input and output files.

Change pipeline workflows to use Apache Beam

Apache Beam is an open-source, unified programming model that enables developers to implement batch and streaming data processing jobs with ease. Here are some advantages of using Apache Beam:

- **Platform Independence:** Apache Beam supports multiple execution engines such as Apache Flink, Apache Spark, and Google Cloud Dataflow, allowing users to run data processing jobs across multiple platforms without needing to rewrite code.
- **Flexibility:** With Apache Beam, users can write data processing code in a language of choice, such as Python, Java, or Go. This flexibility allows users to employ their preferred language and still benefit from the features of Apache Beam.
- **Scalability:** Apache Beam provides horizontal scalability, allowing scaling up or down of processing jobs as per data volume and processing requirements. This scalability ensures handling of large datasets and efficiently processing them without any performance issues.

However, using Apache Beam also has some drawbacks, which are worth noting:

- Steep Learning Curve: Apache Beam has a steep learning curve, and developers need to have a good understanding of the programming model, execution engines, and related concepts. This could lead to a significant learning investment for developers who are new to the tool:

- Overhead: Apache Beam provides several abstractions and APIs that make it easier to write scalable and efficient data processing jobs. However, this abstraction comes with some overhead, which could result in some performance degradation when compared to using a lower-level programming model.
- Transitioning to Apache Beam would require a heavy analysis to rewrite/rebuild existing OpenFIDO infrastructure and to reorient the application's data flow and structure.

In general, Apache Beam offers a flexible and scalable platform-independent data processing model that can help developers handle large datasets efficiently. However, it requires some investment in learning and may not natively support some advanced features.

Recommendation System

OpenFIDO users might not know which pipelines are good to use for follow-on data processing or have not considered next steps in their data analysis. Other users also might consistently re-use the same pipeline series in their analyses.

For these cases, adding a recommendation system can help users discover additional pipelines that can relate to the data they are processing. This can be presented to users as two separate recommendation lists following a pipeline run:

1. The "user" recommendation list, which will be based on which pipelines the user usually runs next after completing their current pipeline run.
2. The "peers" recommendation list, which will be based on what pipelines other users have run after completing this current pipeline run.

Additionally, this system can use the data-forwarding from the workflow UI upgrades to enable a user to easily pass their result artifacts to the new pipeline, should they select one of the recommendations.

Page Load Speed

By implementing pagination in the OpenFIDO application, users can reap a variety of benefits, especially when it comes to displaying large amounts of data. For instance, the OpenFIDO pipeline display currently loads all pipelines for a given organization, which can lead to slower page load speeds and decreased page reactivity.

By adding pagination, one can significantly improve the website's performance and reactivity. This means that users will be able to navigate through the pipelines more quickly and easily. Additionally, implementing pagination for workflow runs ensures that the interface remains responsive even when displaying a large number of runs for a given pipeline.

Overall, implementing pagination on a React front-end website can help ensure that users have a smooth and seamless experience, even when dealing with a significant amount of data. By making it easier to navigate through large data sets, you can help users find what they are looking for more quickly and efficiently.

CHAPTER 4: Conclusion

Ongoing challenges still remain for data sharing and data exchange among stakeholders in electric power system planning, like operations and regulators. The development and release of OpenFIDO addressed some key elements of this challenge. Among the most important of OpenFIDO achievements are the following.

Scalable computing: OpenFIDO runs on an open-source interoperability platform called Docker that facilitates running applications on a wide-variety of host systems, including laptops, workstations, on-premises servers, and cloud infrastructure.

Open-source distribution: OpenFIDO is completely open-source and may be used by anyone for any purpose without licensing fees to the creators of the platform.

Open analytics: OpenFIDO methods, pipelines, and workflows are available on GitHub and can be published and used by anyone at any time on all deployments of OpenFIDO. In addition, anyone can submit new methods, pipelines, and workflows to OpenFIDO for private or public use.

Future Research and Development

Several new pipelines are currently proposed or in development by other projects, including the following:

- **Load decomposition:** analysis to permit extraction of end-use load shapes from AMI data and SCADA data
- **Census data:** access consumer demographic data to enhance load models using economic and population data
- **Advanced building loads:** utilize satellite data to identify the location, size, and type of buildings in communities, and automatically link the building models to power system network models
- **Grid resilience:** PSPS alternative/standardized outage optimization methodology, and long-term climate change impacts on electric load, distribution infrastructure, and data-driven asset planning and hardening

Advanced Workflow Research and Development

The more advanced workflow implementation is available in the application services, but it is not available to users of the interface. In addition, workflow management tools such as Apache Beam are now widely accepted as efficient and scalable solutions that OpenFIDO could use in place of the existing workflow service. The project team recommends follow-up development to incorporate modern batch and streaming analysis technology in future OpenFIDO deployments.

Shared/public data/models Security and Integrity

The current data sharing model uses widely accepted digital security methods. However, this may not be fully satisfactory for utilities to satisfy critical infrastructure cyber-security standards. The project team recommends follow-up research and development work to identify and deploy digital artifact sharing mechanisms that allow utilities to share selected data from software, databases, and repositories with authorized users in a reliable and secure manner.

Legacy Data Handling and Data Aging Strategies

Historical data is a critical component in developing long-term system performance models such as building loads, asset degradation, and human behavior. Consistent long-term access to weather, demographic, building stock, energy consumption, and power demand data is an ongoing challenge and a barrier to adoption of new analytics tools and methods, particularly as modern and more advanced data management and data sharing capabilities become widely accepted. The project team recommends further research and development into data handling strategies to future proof critical existing data sets and ensure that new data analytics have access to much of the valuable existing historical data that makes these analyses worth deploying at scale.

GLOSSARY AND LIST OF ACRONYMS

Term	Definition
AMI	Automated Meter Infrastructure
AWS	Amazon Web Services
CAIFI	Customer Average Interruption Frequency Index
CEC	California Energy Commission
CSV	Comma-Separated Value
DER	Distributed Energy Resource
DOE	Department of Energy
GLOW	GridLAB-D Open Workspace
HiPAS	High Performance Agent-Based Simulation
ICA	Integrated/Hosting Capacity Analysis
IEEE	Institute of Electrical and Electronics Engineers
IOU	Investor-Owned Utilities
JSON	JavaScript Object Notation
LF	Linux Foundation
LiDAR	Light Detection and Ranging
NERC	North American Electric Corporation
OpenFIDO	Open Framework for Integrated Data Operations
PSPS	Public Safety Power Shutoffs
REST API	Representational State Transfer Application Programming Interface
SAIDI	System Average Interruption Duration Index
SAIFI	System Average Interruption Frequency Index
SCADA	Supervisory Control and Data Acquisition
SLAC	Stanford Linear Accelerator Center
TAC	Technical Advisory Committee
VADER	Visualization and Analytics of Distribution Systems with High Penetrations of Distributed Energy Resources
WECC	Western Electricity Coordinating Council

References

Chassin, David P, Jason C Fuller, Ned Djilali. 2014. GridLAB-D: An agent-based simulation framework for smart grids. https://www.hindawi.com/journals/jam/2014/492320/. Journal of Applied Mathematics.

Newman J, A Teyber, D Chassin, Development of Methodology to Forecast Charging Demand for Electric Vehicles Using Location Services Data, SULI Program Report, August 2020. URL: https://tinyurl.com/y5mfvr74.

Schneider, KP, JC Fuller, D Chassin. 2011. Evaluating conservation voltage reduction: An application of GridLAB-D: An open source software package. https://ieeexplore.ieee.org/abstract/document/6039467. IEEE Power and Energy Society General Meeting.

Sevlian, Raffi Avo, Jiafan Yu, Yizheng Liao, Xiao Chen, Yang Weng, Emre Can Kara, Michelangelo Tabone, Srini Badri, Chin-Woo Tan, David Chassin, Sila Kiliccote, Ram Rajagopal. 2017. VADER: Visualization and Analytics for Distributed Energy Resources. https://arxiv.org/pdf/1708.09473.pdf. arXiv.

Project Deliverables

The following project deliverables are available upon request by submitting an email to ERDDpubs@energy.ca.gov. These documents may also be obtained from the OpenFIDO GitHub repository https://github.com/slacgismo/openfido.

- Final report
- Data exchange requirements and assessment presentation
- Data exchange implementation and validation plan
- Data exchange implementation and validation plan presentation
- Implementation report (Critical Project Review Report #1)
- Testing and validation presentation
- Testing and validation report
- Development and user training documentation
- Final product release report
- Production report (Critical Project Review #2)
- Kickoff meeting benefits questionnaire
- Final meeting benefits questionnaire
- Initial fact sheet
- Final fact sheet
- Technology knowledge transfer plan
- Technology knowledge transfer report

CALIFORNIA ENERGY COMMISSION

CALIFORNIA NATURAL RESOURCES AGENCY

ENERGY RESEARCH AND DEVELOPMENT DIVISION

Appendix A: OpenFIDO GitHub Product Repositories

March 2024 | CEC-500-2024-026

APPENDIX A:
OpenFIDO GitHub Product Repositories

The following repositories are available to the public as part of the production release of OpenFIDO.

Application Services

- Main application: https://github.com/openfido/openfido-app-service
- Authentication: https://github.com/openfido/openfido-auth-service
- Workflow: https://github.com/openfido/openfido-workflow-service
- User experience: https://github.com/openfido/openfido-client-service

Pipelines

- Tariff design: https://github.com/openfido/tariff_design
- Load shape analysis: https://github.com/openfido/loadshape
- Weather data: https://github.com/openfido/weather
- HiPAS GridLAB-D simulation: https://github.com/openfido/gridlabd
- Census geographic data: https://github.com/openfido/census
- Resilience analysis: https://github.com/openfido/resilience
- Hosting capacity analysis: https://github.com/openfido/hosting_capacity
- Electrification: https://github.com/openfido/electrification
- Address resolution: https://github.com/openfido/address
- Cyme converter: https://github.com/openfido/cyme_extract