Data Management and Sharing Plan
CPS-FR: Physics-Informed Machine Learning for Resilient Microgrid Control

Principal Investigator: Ehsan Reihani

1 Types of Data and Materials

Experimental and simulation data will include physics-informed neural network (PINN) training datasets with microgrid state trajectories, control actions, and system responses, plus hardware-in-the-loop (HIL) real-time control performance data from NVIDIA Jetson AGX Orin platforms. Communication network simulation data will encompass latency patterns, packet loss statistics, IEEE 2030.5 protocol performance metrics, and multi-agent distributed optimization convergence traces. Safety and optimization outputs will feature Control Barrier Function (CBF) traces with constraint satisfaction metrics and economic dispatch results with ADMM convergence logs.

System models and profiles will comprise synthetic microgrid topologies aligned with IEEE test systems, grid-forming inverter models with control parameters, renewable energy generation profiles (solar PV, wind) with realistic variability, critical infrastructure load demand profiles, and communication topologies with latency characteristics. Software repositories will include PyTorch PINN ODE implementations with embedded power system dynamics, multi-agent reinforcement learning algorithms with formal consensus guarantees, GNN-enhanced distributed optimization solvers, CBF safety layers, vendor-agnostic BITW controller firmware for Jetson AGX Orin, and HIL interfaces with real-time control software.

Machine learning artifacts will encompass trained PINNs for state prediction/control, GNN models for optimization acceleration, deep RL policies with stability guarantees, model checkpoints, hyperparameters, and training convergence logs. Documentation will include technical documentation, BITW deployment guides, physics-informed ML tutorials for power systems, curriculum modules for cyber-physical energy systems courses, and industry workshop materials.

2 Standards, Formats, and Metadata

Power system data formats will use IEEE CDF, PSS/E RAW, IEC 61970 CIM for system components, IEEE C37.118 for synchrophasors, OpenDSS for distribution modeling, and PSCAD for electromagnetic transient studies. Machine learning formats will incorporate HDF5 for large numerical datasets, ONNX for model exchange, MLflow for experiment tracking, NumPy/Pandas for structured analysis, and JSON for hyperparameters/metadata. Communication protocols will implement IEC 61850, IEEE 2030.5, Modbus, DNP3, and MQTT/CoAP for IoT microgrid telemetry. Metadata standards will use DataCite schema for dataset citation, Dublin Core for digital resources, Schema.org for web discoverability, reproducible READMEs, and Jupyter notebooks with embedded documentation/visualization.

3 Access, Sharing, and Privacy

Open access policies make all synthetic datasets, source code, and software publicly available through GitHub repositories with complete documentation. PINN models and training

datasets will be released under permissive licenses for research and commercial use, while educational materials use Creative Commons licenses. Publications will be available through institutional repositories and preprint servers.

Controlled access protects sensitive components through anonymization of hardware-specific performance data from NVIDIA Jetson AGX Orin platforms, data-sharing agreements for industrial collaborations with appropriate access controls, synthesized/de-identified critical infrastructure load profiles, and secure portals for communication vulnerability data limited to authorized researchers. Licensing uses MIT for source code, CC BY 4.0 for datasets, Apache 2.0 for ML models, and CC BY-SA 4.0 for documentation/educational materials. Privacy and security implement differential privacy for sensitive data, systematic anonymization to remove PII/proprietary details, secure enclaves for industry collaborations, and regular security audits.

4 Policies for Re-use, Re-distribution, and Derivatives

Attribution requires users to cite original datasets, software, and DOIs with explicit citation guidance for each release. Derivative works are encouraged under chosen licenses (CC BY 4.0, MIT, Apache 2.0) permitting modification/redistribution while requiring documented changes and version control preserving provenance chains. Commercial use and technology transfer are supported through open source terms permitting deployment, partnership agreements for BITW controller adoption, and potential revenue-sharing for substantial implementations. Quality control maintains standards through peer review for contributed data/code, continuous integration testing, enforced documentation standards, and community governance for major software projects.

5 Archiving, Preservation, and Persistent Access

Institutional preservation uses primary archiving at California State University, Bakersfield with mirrored repositories at partner institutions for redundancy and regular integrity checks. Public archives include final datasets in IEEE DataPort, software releases in Zenodo for DOI assignment, publications in arXiv/institutional repositories, and educational materials in disciplinary repositories. Version control maintains histories through Git for source code, Data Version Control (DVC) for datasets, semantic versioning for software, and migration plans for evolving formats. Backup procedures include daily automated backups to geographically distributed servers, cloud storage redundancy, annual disaster recovery testing, and multiple format storage to mitigate obsolescence.

6 Roles and Responsibilities

The Principal Investigator (Ehsan Reihani) provides overall oversight, ensures alignment with project goals and NSF/institutional policy compliance, coordinates across sites/industry partners, and conducts annual plan reviews/updates. The Data Management Coordinator (graduate student or postdoc) provides daily oversight of data collection, processing, and documentation; trains team members on tools/standards; and conducts quality assurance for public releases. Technical team members follow established protocols, maintain thorough documentation, participate in regular data management updates during team meetings, and conduct peer review before public sharing. Industry collaboration management establishes data-sharing agreements with utilities/technology partners, coordinates intellectual property considerations, and facilitates technology transfer balancing open science with commercial interests. This plan will be reviewed annually and updated to reflect project scope changes, technological advances, and evolving best practices ensuring continued NSF compliance and maximum scientific impact.