

TITLE

GORMan: A framework for **G**overnance, **O**ntologies, and **R**isk **M**anagement of scalable, cross-sector integration of secure energy-water infrastructure information

AUTHORS & AFFILIATIONS

Hillary K. Fishler¹, Alexander B. May¹, Katie Knight¹, Narcisa Pricope², and Luc Rébillout³
¹Oak Ridge National Laboratory, ²Mississippi State University, ³University of Mississippi,

FOCAL AREA

This white paper addresses the inherent challenges in applying data-driven approaches to integrated energy-water solutions across the US – due to the lack of standardization in collection, storage, governance, management, curation and risk mitigation among data owners across scales and agencies. To ensure efficient research and policy development related to the water-energy nexus, we propose the review, refinement, and necessary additions to existing ontologies for the aggregation, use, sharing, and protection of platforms connecting critical energy and water infrastructure data, reducing current challenges to its accessibility and discoverability, while mitigating the compounding risks of national and homeland security.

EXISTING CHALLENGE:

Across the US, growing concerns surrounding stresses on water-power systems, such as the proliferation of large load additions like data centers, water quantity-quality risks, agricultural and residential use, and protection of these complex networks necessitate data-driven approaches in the development of technologies and policies across scales to address current and future system strains.¹ However, data-driven policies leveraging tools and analyses across scales are currently hampered by the fragmented nature of both energy and water data types, collection, storage, and security across purposes, agencies, and jurisdictions.²

Both water and energy information are collected, stored, and governed by a range of organizations across sectors and scales, for a wide variation of purposes and applications³. For example, among federal, state, and local agencies, data collection for water systems (i.e. riparian flows, water quality, hydrology, reserves, etc.) varies greatly from that of utilities (i.e. usage trends, pricing, etc.) – mirroring known challenges in the energy and electricity sector. Even within sectors, the size and capacity of such organizations to collect, store, and manage this information can vary widely.⁴ Particularly for smaller entities, patchwork modeling, tracking, monitoring, and updating of information may hamper informed decision-making about system health and longevity.⁵ Across the country organizational information sharing may be deliberately non-transparent to safeguard against public or political ramifications of the maintenance, use, and exposure of highly contentious water and energy resources.^{6,7,8}

Fragmentation and incompleteness of this data for research and decision-making is further complicated by the sensitive nature of these datasets related to domestic water and energy. Leveraging both open and protected data and promoting interoperability of cross-sector datasets to better inform research, policymaking, and investments into coupled water-power systems resilience, presents elevated risk to domestic safety and national and homeland security⁹. Integration of open datasets with those representing critical energy infrastructure information exponentially amplify vulnerabilities to the nation's cyber-physical systems, particularly when operationalized by research institutions, government agencies, utilities, and industry.¹⁰

For architectures intended to capture, link, and amalgamate this sensitive data, existing challenges related to non-standard methods of data collection, reporting, and adherence to legal frameworks and policies across federal, state, and local jurisdictions, agencies, and scale – demonstrate heightened risks to US national security

and homeland security if misused.¹¹ These risks are further intensified with the fragmented collection, classification, storage, naming conventions, and use cases across both open and controlled architectures. And, alongside backlogs of fragmented information, water and energy dataset security outside of the utility industry, while regulated, is not consistently compliant with national security standards. Thus, without standardized processes for water-energy data governance, management, and curation – collaboration efforts among government agencies and researchers for water-energy resilience policies and investments may inadvertently lead to inefficient or underdeveloped solutions, with further fragmentation in cross-sector, multi-scale decision-making processes and long-range strategic planning.¹²

NEAR-TERM OPPORTUNITIES

Our team proposes the development of a rigorous, and unified framework for data governance, anomaly-detection, tagging, and curation of datasets related to critical energy and water infrastructure information data. Modeled after our dynamic framework for sensitive critical energy infrastructure data (for DOE’s Office of Electricity and Grid Deployment Office)¹³ we will employ National Institute of Standards and Technology’s (NIST) Research Data Framework (RDaF) conventions, enforce DOE and federal policies, and adopt nested standards for Critical Energy Infrastructure Information, Private Energy Infrastructure Information, Personally Identifiable Information, Controlled Unclassified Information, and open-and-shared data to ecosystem aggregation of utility, agency, and laboratory databases, monitoring platforms, datasets, and data lakes related to domestic water resources, energy data, and their connected use cases and purposes.¹⁴¹⁵¹⁶

While sharing platforms increase data accessibility and enable data-driven decision-making, these infrastructures have inherent barriers related to data governance and management¹⁷, limited curation capabilities, and risk mitigation, particularly for sensitive, private, and proprietary data in existing ontologies.¹⁸ Because open accessibility to water, power, and intersecting critical infrastructure data may inflate opportunities for exploitation of this sensitive data, anticipatory lifecycle data governance and management workflows are imperative from a project’s launch to reduce vulnerabilities.¹⁹²⁰ The following activities will require coordinated efforts among industry, academia, government, and other stakeholders throughout the lifecycle of the project to ensure the most risk-averse, useful, and applicable conventions and standards across scales, sectors, domains, and purposes.

Objective 1: Advance theory and practice of governance, multimodal curation, and management across data types and architecture. Proper curation of sensitive data requires file-level review for quality assurance and risk minimization and must capture information about overarching data governance and management policies. Consequently, metadata created during the curation process is moving beyond discovery and findability (“the ‘F’ in FAIR”)²¹ to providing information about access protocols, while systematically applying governance policies through controlled vocabularies and automatic tagging. This added layer of metadata provides the capability for auditable events, further ensuring risk mitigation.²²

Objective 2: Create a scalable application and risk mitigation across the data lifecycle. Water data across agencies, jurisdictions, scales, and use cases, is inherently sensitive, particularly when collated into like-datasets, increasing exposure to potential vulnerabilities, simply by the nature of their connectivity. This becomes particularly challenging when multiple streams of critical infrastructure data, (i.e. transmission electric grid data, food supply information, and emergency management, etc.) are aggregated for multi-systems decision-making. Here, the need for risk assessment and mitigation for national security purposes is heightened and magnified.

Objective 3: Develop standards and workflows for integration across jurisdictions. Today, there is no standard for scalable cross-agency data sharing, workflow connections, nor standard for tagging, denoting dataset anomalies, naming conventions, nor use cases. Particularly where human interpretation of regulatory

practices and privacy-preserving standards induce challenges in the usability and accessibility of critical infrastructure data and given a lack of existing tools or workflows for governance, management, curation, and maintenance in current architectures, the proposed methodology has applications to other agencies and sectors requiring similar controls for sensitive data.

Objective 4: Employ classification and naming conventions, shared definitions, and standards to enable autoclassification, seamless platform development across use cases and users, assign role-based access controls, and promote data-driven analysis and a common language among researchers and decision-makers across scales, organizations, and jurisdictions.²³

SUCCESS MEASURES

Rigorous review of existing ontologies and repeatable standards across scopes and scale are critical for pivots or expansions of water-power/water-energy nexus mission and objectives – particularly given the rapid nature of demands to these coupled systems from large load additions, population increase, and industrial use, among others. Success of water-energy system interventions and resilience efforts cannot be measured without cross-disciplinary understanding of the need, purpose, and scope of available information. To enable greater return-on-investment and magnify societal impact of future water-power projects and programs, we propose scalable conventions, language, and controls across the data landscape, leading to:

- Greater data interoperability through established use cases, user communities, documentation, and applied data governance frameworks, and reproducible curation models throughout the entire data lifecycle – leading to the publication of well-stewarded datasets;
- Enhanced discoverability and uniformity through rigorous review and refinement against existing ontologies, data, metadata, and validation methodologies to enable classification of key metadata terms through machine learning, allowing users to sift through approximately a terabyte of data;
- Holistic risk assessment and mitigation planning, with quality assurance reviews for stored and backlogged data, and established quality standards for future dataset integration;
- The development of a machine-readable controlled vocabulary providing clear terms of access for users;
- Creation of a robust data-discovery layer, employing guidelines and controls for role-based access, propriety, private, and critical infrastructure information to expose the right dataset to the right user, and
- Secure workflows for merging and linking tools, policies, and cyber-physical data architecture at scale.

Shared languages and enhanced understanding of available and needed information by water-power system stakeholders across the country can accelerate needed interventions and strategies for resilient water, wastewater, and energy systems, which are critical to societal wellbeing, cohesive emergency management, and national security. Through the development of proactive standards for a more universal approach to data governance, management, and protection of this currently highly fragmented and varied information, our development of conventions across disciplines, jurisdictions, agencies, and scales is poised to provide an essential foundation for cross-sector collaboration to tackle urgent, emerging, and compounded challenges at the nexus of complex water-power systems.

REFERENCES

- ¹ Iyer, G., Fishler, H. K., and Chinthavali, S. 2025. "Data Centers as Positive Disruptions to the Electric Grid." ORNL White Paper. Oak Ridge National Laboratory.
- ² Sprague, Lori A., Gretchen P. Oelsner, and Denise M. Argue. "Challenges with Secondary Use of Multi-Source Water-Quality Data in the United States." *Water Research* 110 (2017): 252–61. <https://doi.org/10.1016/j.watres.2016.12.024>.
- ³ Internet of Water. National Assessment of Water Data Legislation. Durham, NC: Nicholas School of the Environment, Duke University, 2020.
- ⁴ U.S. Geological Survey. "National Water Dashboard." Reston, VA: USGS, n.d. <https://dashboard.waterdata.usgs.gov/app/nwd/en/>.
- ⁵ U.S. Environmental Protection Agency. "Water Quality Data." EPA Water Data and Tools. Washington, DC: EPA, n.d. <https://www.epa.gov/waterdata/water-quality-data>.
- ⁶ Josset, Louise, Maura Allaire, Charlene Hayek, James Rising, Christopher Thomas, and Upmanu Lall. "The U.S. Water Data Gap—A Survey of State-Level Water Data Platforms to Inform the Development of a National Water Portal." *Earth's Future* 7, no. 4 (2019): 433–49. <https://doi.org/10.1029/2018EF001099>.
- ⁷ Aspen Institute. *Internet of Water: Sharing and Integrating Water Data for Sustainability*. Washington, DC: Aspen Institute, 2017.
- ⁸ Blodgett, David, Jessica Lucido, and James Kreft. "Progress on Water Data Integration and Distribution: A Summary of Select US Geological Survey Data Systems." *Journal of Hydroinformatics* 18, no. 2 (2016): 226–37. <https://doi.org/10.2166/hydro.2015.067>.
- ⁹ Iyer, G., Chinthavali, S., and Fishler, H.K. 2025. "Modeling Energy Growth Associated with Data Centers - a holistic platform for accelerated, informed, & energy-resilient data center deployment." ORNL White Paper. Oak Ridge National Laboratory.
- ¹⁰ May, A. B., and H. K. Fishler. "Data Portal Standards for Discovery, Access, and Usability across the DOE National Laboratory Complex." Forthcoming, 2025.
- ¹¹ Open Data Handbook. "What Is Open Data?" Open Data Handbook. Open Knowledge Foundation, n.d. <https://opendatahandbook.org/guide/en/what-is-open-data/>.
- ¹² Smith, Jacob, and Jonas B. Sandbrink. "Governing Dual-Use Data: Policy Frameworks for Secure and Open Science." *Nature Machine Intelligence* 4 (2022): 987–95. <https://doi.org/10.1038/s42256-022-00567-8>.
- ¹³ Oak Ridge National Laboratory. "Open Energy Data Portal." Oak Ridge, TN: ORNL, n.d. <https://openenergyhub.ornl.gov/pages/home/>.
- ¹⁴ U.S. Department of Energy. *Data Stewardship Playbook*. Washington, DC: U.S. Department of Energy, 2025.
- ¹⁵ National Institute of Standards and Technology. "Sensitive Information." Computer Security Resource Center Glossary. Gaithersburg, MD: NIST, n.d. https://csrc.nist.gov/glossary/term/sensitive_information.
- ¹⁶ National Institute of Standards and Technology. "Role-Based Access Control." Computer Security Resource Center Glossary. Gaithersburg, MD: NIST, n.d. https://csrc.nist.gov/glossary/term/role_based_access_control.
- ¹⁷ Research Data Governance Framework. Nashville, TN: Vanderbilt University, n.d. <https://researchdatagovernance.vanderbilt.edu/>.
- ¹⁸ Office of Gas and Electricity Markets (Ofgem). *Data Best Practice Guidance*. London: Ofgem, n.d.
- ¹⁹ National Archives and Records Administration. "Critical Energy Infrastructure Information (CEII)." Controlled Unclassified Information (CUI) Registry. Washington, DC: NARA, n.d. <https://www.archives.gov/cui/registry/category-detail/critical-energy-infrastructure-information>.
- ²⁰ Sheridan, Helen, et al. "Data Curation through Catalogs: A Repository-Independent Model for Data Discovery." *Journal of eScience Librarianship* 10, no. 3 (2021): e1203. <https://doi.org/10.7191/jeslib.2021.1203>.
- ²¹ Hanisch, Robert J., et al. "FAIR Data Principles in Practice: Implementation Challenges and Opportunities." *Data Science Journal* 23, no. 1 (2024): 12. <https://doi.org/10.5334/dsj-2024-012>.
- ²² Oxford Common File Layout (OCFL) Specification. Version 1.1. OCFL Community, 2023. <https://ocfl.io/>.
- ²³ New York University Health Sciences Library. "NYU Health Sciences Data Catalog." New York: NYU, n.d. <https://datacatalog.med.nyu.edu/>.