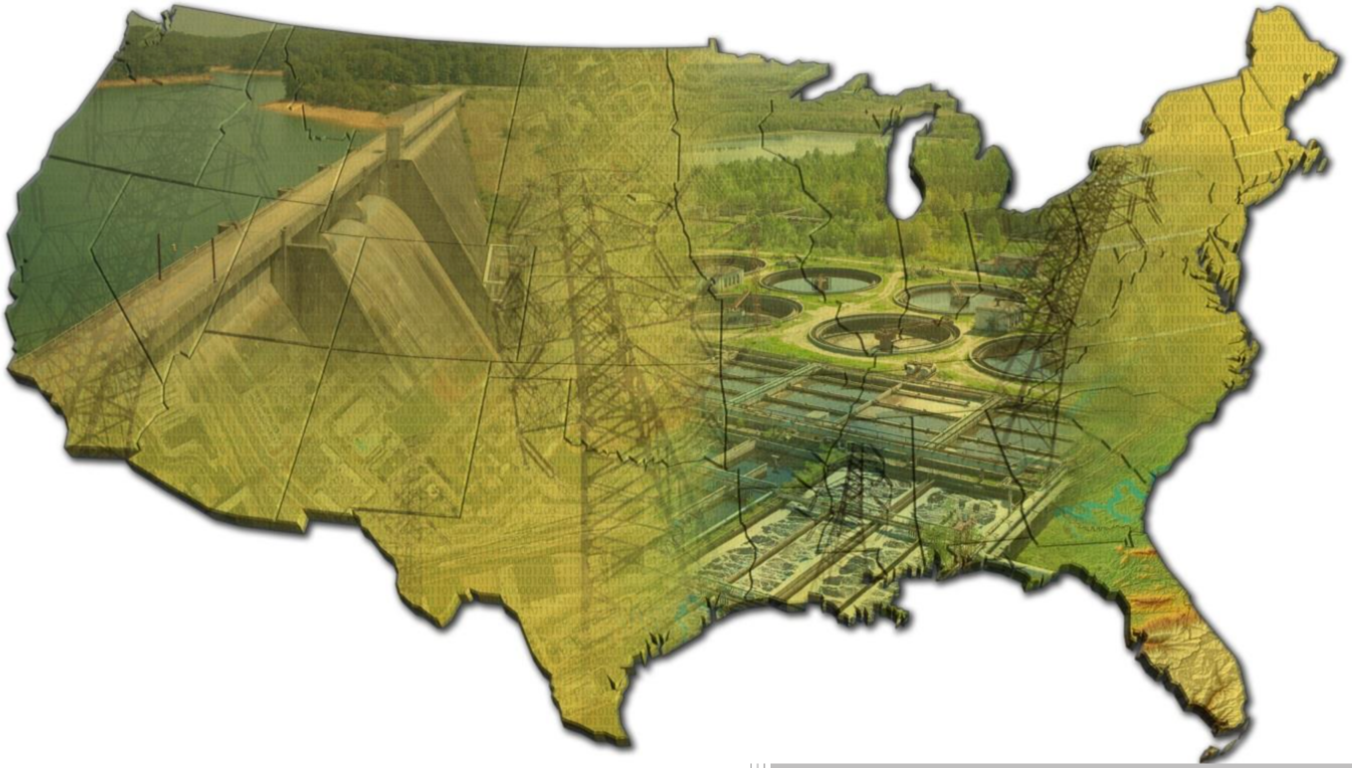
Energy-Water Nexus Knowledge Discovery Framework, Experts' Meeting



Budhu Bhaduri
AJ Simon
Melissa Allen
Jibo Sanyal
Robert Stewart
Ryan McManamay

ORNL/TM-2017/753

**Approved for public release.
Distribution is unlimited.**

DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via US Department of Energy (DOE) SciTech Connect.
Website: `http://www.osti.gov/scitech/`

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-605-6000 (1-800-553-6847)
TDD: 703-487-4639
Fax: 703-605-6900
E-mail: info@ntis.fedworld.gov
Website: `http://www.ntis.gov/help/ordermethods.aspx`

Reports are available to DOE employees, DOE contractors, Energy Technology Data Exchange representatives, and International Nuclear Information System representatives from the following source:

Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831
Telephone: 865-576-8401
Fax: 865-576-5728
E-mail: report@osti.gov
Website: `http://www.osti.gov/contact.html`

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/TM-2017/753

Energy-Water Nexus Knowledge Discovery Framework, Experts' Meeting

Date Published: October 25, 2018

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831-6283
managed by
UT-Battelle, LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725
and
LAWRENCE LIVERMORE NATIONAL LABORATORY
Livermore, CA 94550
operated by
Lawrence Livermore National Security, LLC
for the Department of Energy's National Nuclear Security Administration

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Notation	**Description**	**Page List**
ANL	Argonne National Laboratory	8, 28
BNL	Brookhaven National Laboratory	9
BU	Boston University	9
CADES	Compute and Data Environment for Science	28
CIESIN	Center for International Earth Science Information Network	8
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science, Inc.	9
DOE-BER	Office of Science, Office of Biological and Environmental Research	8, 13
EERE	Office of Energy Efficiency and Renewable Energy	8
EIA	Energy Information Administration	9
EPSA	Office of Energy Policy and Systems Analysis	8, 13
ESGF	Earth System Grid Federation	28
EWN-KDF	Energy-Water Nexus Knowledge Discovery Framework	11, 22
FE	Office of Fossil Energy	8
GCAM	Global Change Assessment Model	18
JHU	Johns Hopkins University	9
LANL	Los Alamos National Laboratory	9
LBNL	Lawrence Berkeley National Laboratory	9
LLNL	Lawrence Livermore National Laboratory	8
NASEO	National Association of State Energy Officials	9
NAU	Northern Arizona University	9
NREL	National Renewable Energy Laboratory	9
ORNL	Oak Ridge National Laboratory	8, 28
PNNL	Pacific Northwest National Laboratory	8
RENCI	Renaissance Computing Institute	9

Notation	Description	Page List
SNL	Sandia National Laboratory	9
UTA	University of Texas at Austin	9

ACKNOWLEDGMENTS

The authors would like to acknowledge the enlightened leadership of the Department of Energy Offices of Energy Policy and Systems Analysis and Biological and Environmental Research and the valuable guidance and feedback of the community of practice around the Energy-Water Nexus.

EXPERT COMMUNITY

Department of Energy

Office of Energy Policy and Systems Analysis (EPSA)
Diana Bauer, Director, Energy Systems Integration Analysis
Zachary Clement, Engineer
Bonita Singal, AAAS Science & Technology Policy Fellow

Office of Science, Office of Biological and Environmental Research (DOE-BER)
Gary Geernaert, Director, Climate and Environmental Sciences
Robert Vallario, Program Manager, Integrated Assessment Research
Justin Jay Hnilo, Program Manager, Climate and Environmental Data Management
David Lesmes, Program Manager, Subsurface Biogeochemical Research

Office of Energy Efficiency and Renewable Energy (EERE)
Hoyt Battey, Program Manager, Market Acceleration and Deployment, Wind/Water Technologies
Simon Gore, NOAA Sea Grant Kanuss Fellow, Wind/Water Technologies

Office of Fossil Energy (FE)
Robie Lewis, Program Manager, Crosscutting Research

Oak Ridge National Laboratory (ORNL)

Budhendra Bhaduri, Corporate Fellow; Director, Urban Dynamics Institute
Jibonananda Sanyal, Team Lead, Scalable and High Performance Geocomputation
Melissa Allen, Research Scientist, Critical Infrastructure and Climate Change
Ryan McManamay, Team Lead, Integrated Water-Energy Systems
Robert Stewart, Team Lead, Geographic Data Sciences

Argonne National Laboratory (ANL)

Ian Foster, Distinguished Fellow; Director, Data Scienc and Lerning Division
May Wu, Lead, Water Analysis

Columbia University

Robert Chen, Director, Center for International Earth Science Information Network (CIESIN)

State University of New York, Buffalo

Varun Chandola, Lead, Data Sciences Group

Pacific Northwest National Laboratory (PNNL)

Katherine Calvin, Research Economist, joint Global Change Research Institute
Robert Link, Senior Computational Scientist, Joint Global Change Research Institute

Lawrence Livermore National Laboratory (LLNL)
AJ Simon, Associate Program Leader, Water Security and Technologies
Dean Williams, Collaborative Institutional Lead, Energy Exascale Earth System Model (E3SM)

University of Texas at Austin (UTA)
Svetlana Ikonnikova, Energy Economist

Lawrence Berkeley National Laboratory (LBNL)
Deb Agrawal, Senior Scientist; Department Head, Data Science and Technology

Brookhaven National Laboratory (BNL)
Kersten Kleese-Van Dam, Director, Computational Science Initiative

National Renewable Energy Laboratory (NREL)
Jordan Macknick, Lead Analyst, Energy-Water-Land
Jon Weers, Senior Web Strategist and Applications Engineer

Sandia National Laboratory (SNL)
Vince Tidwell, Distinguished Member of Technical Staff

Los Alamos National Laboratory (LANL)
Richard Middleton, Deputy Group Leader, Computational Earth Science

Boston University (BU)
Ian Sue Wing, Professor, Department of Earth and Environment

Johns Hopkins University (JHU)
Benjamin Zaitchik, Associate Professor, Earth and Planetary Sciences
Sauleh Siddiqui, Assistant Professor, Department of Civil Engineering

Northern Arizona University (NAU)
Benjamin Ruddell, Associate Professor, School of Informatics, Computing, and Cyber Systems

Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
Jared Bales, Executive Director

Energy Information Administration (EIA)
Shirley Neff, Senior Advisor

National Association of State Energy Officials (NASEO)
Rodney Sobin, Senior Program Director

DC Water
Biju George, Chief Operating Officer

Renaissance Computing Institute (RENCI)
Lea Shanley, Co-Executive Director, South Big Data Innovation Hub

EXECUTIVE SUMMARY

Energy and water generation and delivery systems are inherently interconnected. With worldwide demand for energy growing, the energy sector is experiencing increasing competition for water. With increasing population and changing environmental, socioeconomic, and demographic scenarios, new technology and investment decisions must be made for optimized and sustainable energy-water resource management. These decisions require novel scientific insights into the complex interdependencies of energy-water infrastructures across multiple space and time scales.

On June 13, 2017 an Experts' Meeting was convened for the purpose of informing the design of an integrated data driven modeling, analysis, and visualization capability that can enable efficient local and regional decision-making regarding present and future energy-water infrastructure. This first multi-institute collaborative Energy-Water Nexus Knowledge Discovery Framework (EWN-KDF) will be informed by fundamental science and guided with strategic (federal) policy decisions in the service of ensuring national energy resilience.

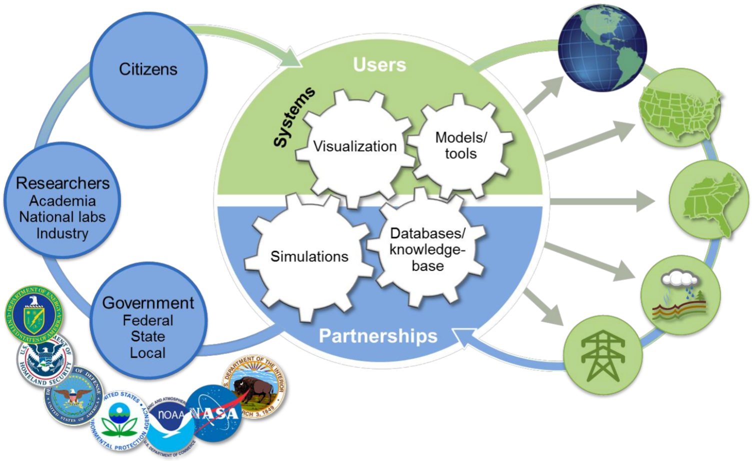


Figure 1. Energy-Water Nexus Knowledge Discovery Framework

The Expert community was tasked with identifying the most critical components of the proposed EWN-KDF, for which the following three objectives had been established:

- Develop a robust data management and geovisual analytics platform, which will provide access to disparate and distributed geospatial Earth science and critical infrastructure data, socioeconomic data, and emergent ad-hoc sensor data;
- Share creation of a powerful toolkit of analysis algorithms and compute resources to empower user-guided data analysis and inquiries; and
- Demonstrate knowledge generation with selected illustrative use cases for the implications of climate change for coupled land-water-energy systems through the application of state-of-the art data integration, analysis, and synthesis.

The experts determined that the first priorities for the EWN-KDF are:

- Provide access to quality data from various sources along with access to meta-data, and facilitate the integration of those data over temporal and spatial dimensions.

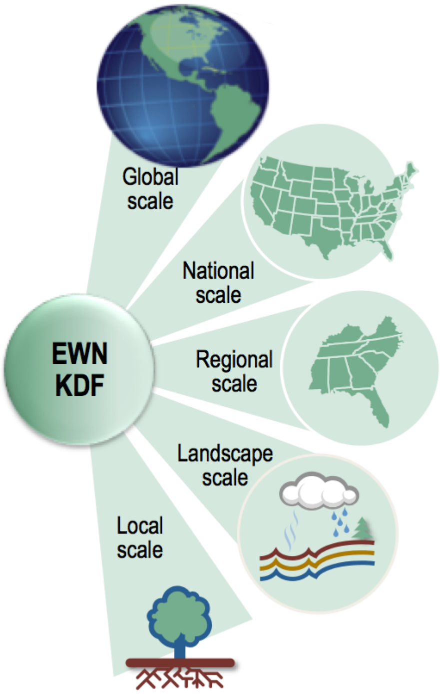
- Facilitate the integration of individual models developed to address a specific geomorphology, infrastructure or socioeconomic systems.
- Provide data and analytical capabilities to manage, visualize and compare model outputs of future regional trajectories for climate, population growth, land use, and economic activity, to test options for adaptation, and resilience, including new technologies.

The following sections summarize the proceedings of that Experts' Meeting and highlights the presentations and discussions leading to the community's recommendations.

1. BACKGROUND

An Experts' Meeting was held on June 13th in Washington, DC to inform the Department of Energy's (DOE) Energy-Water Nexus Knowledge Discovery Framework. This data-driven framework is a critical element within the overall Data, Modeling, and Analysis component of the Department of Energy's Energy-Water Nexus (EWN) crosscut.

Members of the expert panel included those identified as having essential background and expertise to engage in data analysis at the energy-water nexus; integrated, cross-sector modeling; data management, advanced analytic methods, and visualization; and/or development of data layers; and who are candidates for the framework's user community. The input collected from the assembled panel and presented here will inform the design of the framework, including its structure, function, and linkages with other energy-water research.



The Experts' Meeting, led by Diana Bauer (Director, EPSA), Gary Geernaert (Director, of DOE-BER Climate and Environmental Sciences), and Bob Vallario (Program Manager, Integrated Assessment Research), is part of an ongoing investment by the Integrated Assessment Research program in the Office of Science to produce foundational capabilities in environmental and energy research and by the Office of Energy Policy and Systems Analysis to promote complementary work related to analyses, use cases, and data quality. The meeting concentrated on use-inspired perspectives on the scope, design, and phased development of the Framework. Particular attention in this meeting was paid to identifying early use cases that could motivate progress on the framework and strategically align plans among shared energy-water nexus communities. Broader departmental capabilities, needs, and interests were also discussed, and were reflected in the collective expertise of the attendees.

The meeting began with a discussion led by Energy-Water Nexus Knowledge Discovery Framework Principal Investigator, Budhu Bhaduri (Oak Ridge National Laboratory), on the status of and challenges facing the development of the framework. Next, Jay Hnilo (DOE-BER Program Manager for Climate and Environmental Data Management) addressed the assembled community regarding broader integrative data environments. To elucidate potential required capabilites for the system, three illustrative use cases were presented by experts from three different institutions. First, Kate Calvin (Pacific Northwest National Laboratory) discussed analytical capabilities needed for Energy-Water-Land Interactions. Next, AJ Simon (Lawrence Livermore National Laboratory) proposed an automated approach to creating Sankey diagrams associated with issues around energy production, conversion, use, and disposition, and interactions among energy and water in these processes. Finally, Svetlana Ikonnikova (University of Texas at Austin) presented methods and results from a multi-year interdisciplinary study of shale gas oil resources, their future development and associated water and land footprints. The afternoon consisted of three breakout sessions during which participants at the meeting worked through various use cases and prioritized data and analytical needs for obtaining actionable results for each.

2. US DEPARTMENT OF ENERGY'S ENERGY-WATER NEXUS KNOWLEDGE DISCOVERY FRAMEWORK

To prepare the gathered community for discussion around the most actionable components of the emerging Energy-Water Nexus Knowledge Discovery Framework, two speakers led discussions on the status and challenges of the framework itself and of broader integrated data environments in general. The first of the discussions was led by Budhu Bhaduri of Oak Ridge National Laboratory, and the second by Jay Hnilo of the Department of Energy.

2.1 Status and Challenges

The Energy-Water Nexus Knowledge Discovery Framework is conceived as a community of practice around data focusing on integrated data at highest reliable resolutions along physical, infrastructural, and human dimensions. The framework will include accelerated analytics on a consistent platform enabling data, analytics, and visualization as services. It will provide a platform for both the advancement of scientific knowledge and a basis for operational decision making. The user interface will highlight interactive and interoperable data visualization and a capability for high performance scalable analytics. Work in progress for the framework includes emphasis on dynamic collection, integration, management and dissemination of disparate data resources leading to knowledge base creation and analytical tool development.

Recognizing that solutions to global and national energy challenges are often local to regional, the framework will provide for multiscale integration accommodating characterization of interactions among the energy, water, land, agriculture, environment, climate, transportation, cyber, demographics and economics networks. Because making the best decisions depends on understanding the complex potential consequences that the competing approaches are likely to have on energy resilience, economic growth, the environment, etc., the framework will allow for optimization over several model layers and outcomes simultaneously.

The data and information system must be capable of robust representation of past, present, and future dynamics of energy-water systems and must include generation, collection, and analysis of cross-sector data at highest reliable spatial and temporal resolutions to enable interaction with domain models such as integrated assessment models (e.g., Pacific Northwest National Laboratory's Global Change Assessment Model). It must also promote understanding of physical, engineered and human system responses to climatic extremes by integrating observational and model simulated data along biophysical, infrastructural and human dimensions.

Envisioned Capabilities of the EWN-KDF

- **Facilitate** review of data, publications, documents and models
- **Associate** data, knowledge, and people (publications with data; documents with documents)
- **Analyze** spatial analysis with geographic data, scenarios with domain specific models
- **Share** data or analysis results with everyone, selected users (groups), or individuals based on contributor's preference
- **Visualize** spatial overlays and geographic information along with conventional visualization (Tables, graphs, and charts
- **Collaborate** by organizing special interest groups
- **Communicate** on a forum

The system should enable understanding energy-water system dynamics across space and time scales supported by the capability to develop rigorous potential future scenarios with uncertainty analyses that capture likely interactions with climate variability, severe weather events, changes in population, economics, fuels, energy technologies, and energy and transportation infrastructure, creating credible data and knowledge that can provide the foundation for integrated perspectives. Thus, such a federated data and energy-water information system removes the barrier of isolated data silos that lead to “information fragmentation” allowing access to large data volumes across distributed sites, extensible access to data, information, analytical tools from local to global scales, and efficient development of a holistic view across energy sectors facilitating interdependency analysis among critical infrastructures.

Thus, to produce necessary novel scientific insights into the complex interdependencies of energy-water infrastructures across multiple space and time scales, an integrated data driven modeling, analysis, and visualization capability is needed to understand, design, and develop efficient local and regional practices for the energy-water infrastructure. To meet these needs, an Energy-Water Nexus Knowledge Discovery Framework must accomplish two objectives: i) development of a robust data management and geovisual analytics platform that provides access to disparate and distributed geospatial Earth science and critical infrastructure data, socioeconomic data, and emergent ad-hoc sensor data (including Volunteered Geographical Information from sources such as Twitter feeds) to provide a powerful toolkit of analysis algorithms and compute resources to empower user-guided data analysis and inquiries; and ii) demonstration of knowledge generation with selected illustrative use cases for the implications of climate change for coupled land-water-energy systems through the application of state-of-the art data integration, analysis, and synthesis. The proposed architecture to support this workflow is shown in Figure 2.

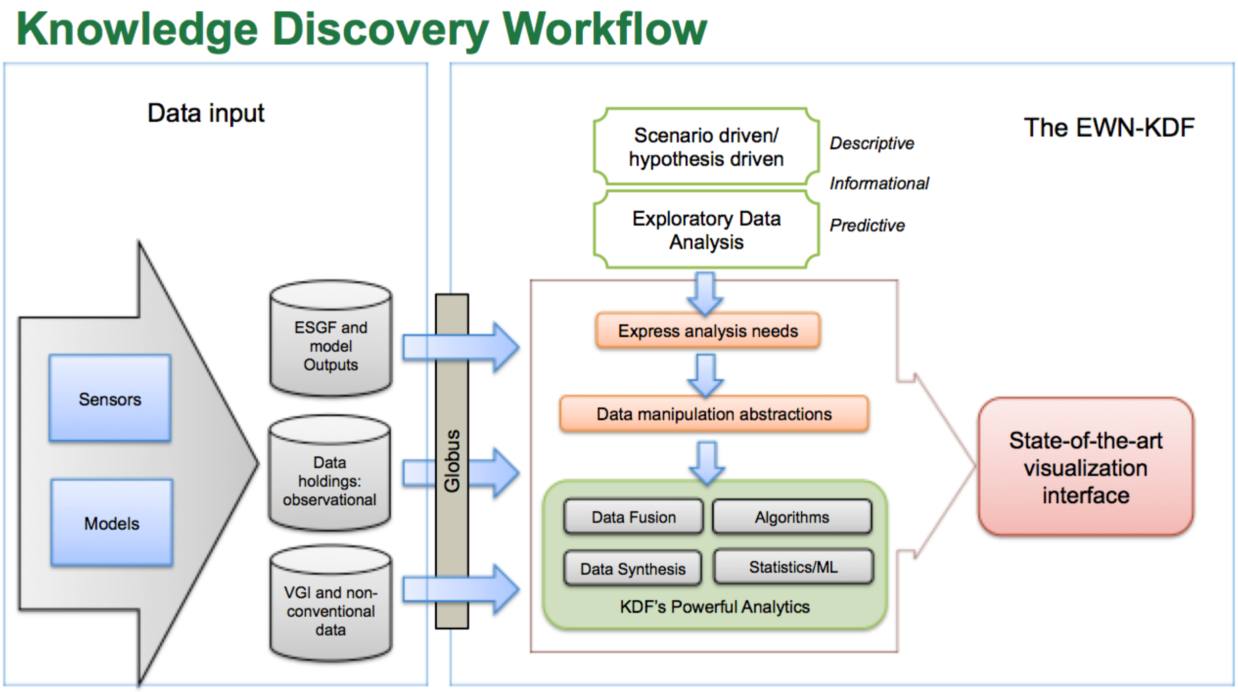


Figure 2. Knowledge Discovery Workflow for Energy-Water Nexus Analytics

Some of the main challenges for the EWN-KDF lie in assuring quality, credibility, and currency for data. Data management requires consensus standards and best practices among partners in the community of practice. Shifting nomenclature must be managed and disparities resolved so that researchers can work from common sets. Thus, a set of standards for formats must be established using quality indices and flags.

Metadata, including geographic and geodetic parameters must be characterized and indexed. With these capabilities in place, potential new products may be defined and criteria for potential new products, data formats, geodetic identification, grids, projections and other factors that cross the boundaries between gridded and non-gridded data can be established. The framework must also be capable of implementing a full categorization of current data needs for the suite of operational tools. To address these challenges, we can leverage experience and expertise from current systems such as the Department of Energy Bioenergy KDF, Systems Biology Knowledgebase (Kbase), Atmospheric Radiaation Measurement data archive (ARM), Carbon Dioxide Information Analysis Center (CDIAC), and Program for Climate Model Diagnosis Intercomparison (PCMDI); the Department of Homeland Security National Infrastructure Simulation and Analysis Center (NISAC) and Homeland Infrastructure Foundation-Level Data (HIFLD) service; and the Oak Ridge National Laboratory Hydro Geographical Information System (HydroGIS) service. Search and analysis services will be harnessed for access-controlled, faceted search across many data sources, and for metadata synthesis and extraction capabilities.

With these data services functioning, we can serve the data to federated analytical capabilities including those of WebWorldWind, the World Spatio-Temporal Analyitical Mapping Project and many others. Together, these resources will help us build an Energy-Water Nexus Knowledge Discovery Framework that uses a data-driven approach for explaining the past, observing the present and predicting future energy system dynamics.

Box 1: Community Engagement Informs the Design of the Knowledge Discovery Framework

- Community engagement achieves two complementary objectives. First, it facilitates the elicitation of relevant research questions that can be addressed through data science, the identification of high value data assets for integration into the EWN-KDF, as well as the identification of critical data gaps that impede research and development. Second, engagement enables the tracking of the use of the EWN-KDF over time as well as reflexive research regarding how such knowledge platforms can be designed to effectively meet the needs of diverse users with diverse questions.
- The Experts' Meeting engaged the community in the identification of Energy-Water Science Drivers, which represent the priority research questions that address critical uncertainties regarding the interactions among climatic change and coupled land/water/energy systems. These priority research questions are the science drivers for the EWN-KDF and represent the most relevant use cases that inform its design.
- Experts were also asked to identify the most critical discrete federated data assets needed for EWN knowledge discovery. They requested specific and disparate multi-dimensional data in diverse formats including observation, modeling, and simulation data assets currently in use by federal agencies and researchers as well as new synthetic data products emerging from, or needed to support, EWN research.

2.2 Broader Integrative Data Environments

An Energy-Water Nexus Knowledge Discovery Framework for the future must accommodate the exponential growth in the volume, acquisition rate, variety, and complexity of scientific data. It must anticipate the magnitude and the character of this growth and maximize its utility through robust cataloging and curation, validation and verification, uncertainty quantification and efficient storage and service. The framework and its management must also foresee and enable the growth of different disciplines working across techniques by integrating simulation, experimental or observational results for complicated data management, analysis, and visualization solutions.

These goals are challenging because metadata standards are varied and in some cases do not exist at all. These issues act to complicate the accessibility, availability and usefulness of high quality research data. Additionally, multiple and complex uncertainties interfere with our ability to manage and mitigate energy and environmental challenges sustainably.

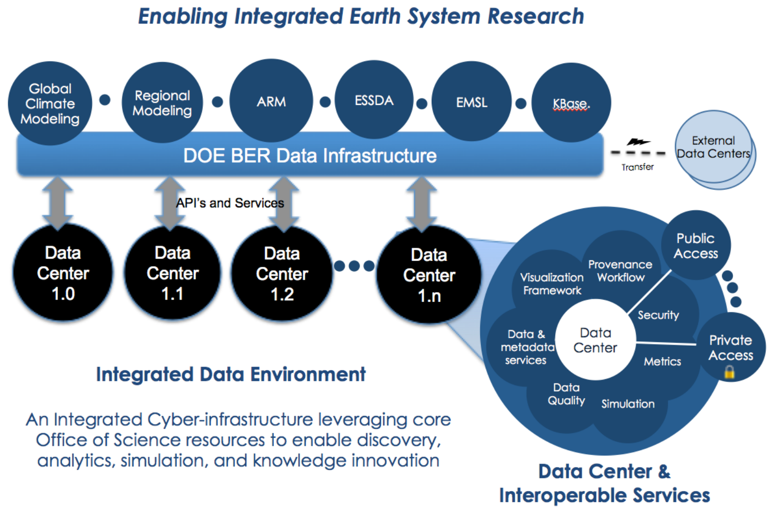


Figure 3. Integrated Earth System Research

However, we are developing this framework at a time when we have a variety of computational and communications tools available to us, and options for integrating these tools in ways that unite the currently fragmented community of practice participating in today's energy-water nexus research. Our challenge is to develop efficient capabilities that enable a predictive understanding of complex, multi-scale, coupled and biologically based environmental systems behavior. These capabilities can and must come together in a new type of integration of in-situ hypothesis-driven experimentation and observations; and technological advances across multiple scales of space, time and system organizations.

3. ILLUSTRATIVE USE CASES

Prominent members of the Energy-Water Nexus community lead discussions on three potential use cases. The use cases were designed to illustrate the diversity of challenges to data management and scientific inquiry for the Energy-Water Nexus Knowledge Discovery Framework. The first use case was presented by Kate Calvin of Pacific Northwest National Laboratory, the second by AJ Simon of Lawrence Livermore National Laboratory, and the third by Svetlana Ilkonnikova of the University of Texas at Austin.

3.1 Use Case 1: Energy-Water-Land Interactions

The motivation and framing for Use Case 1: Energy-Water-Land Interactions was the science question,

> How do constraints on water availability influence the energy and land systems?

and its accompanying hypothesis,

> Limiting water resources will result in shifts in agricultural production away from arid regions and decreased dependence on once-through cooling in rapidly developing economies.

The experimental design proposed for the use case began with the recognition that a model was needed that couples representations of energy, water and land along with interactions among the systems. For this use case, the Global Change Assessment Model (Global Change Assessment Model (GCAM)) answered that need (www.github.com/jgcri/GCAM-core).

GCAM is a global economic model which links Economic, Energy, Land-use, Water, and Climate systems to compute supply and demand for a variety of goods and services (e.g., agricultural commodities, energy carriers, water withdrawals). Using 32 energy/economic regions and approximately 300 land and water regions for "what if" analysis, implications of various technology-rich changes on the energy land and water systems can be explored at 5-year timesteps.

Figure 4. GCAM Framework

Two simulations are run for the use case, 1) a control case and 2) a test case. The control case would evaluate continued evolution of the energy and land use systems into the future unlimited water supply. The test case would assess continued evolution of the energy and land use systems into the future constraining the surface water availability in the future at 2010 levels. Without constraints, water withdrawals increase in the future to meet needs of a growing, increasingly wealthy population. Preliminary results from the simulations show:

- Limiting water availability has strong effects in India, China, and the Middle East.
- Limiting water shifts agricultural production from water constrained regions to other parts of the world.
- Limiting water requires a shift to types of electricity generation that demand lower water withdrawals.

To analyze and visualize both input and output data for the use case, global, spatially-explicit, internally-consistent data sets for all energy, water and land-related information for the historical period is required. Some of the main challenges associated with these analyses are that existing datasets are often collected and served at different spatial, temporal and process resolutions. Thus, consistency and agreement among overlapping variables within these data sets is low. Additionally, some data sets are not available for every location on earth, and some are not available at all. As the model runs, large volumes of data are output as a result. In order to examine these data, which include many variables evaluated across space and time, a robust and visually intuitive presentation must be part of the framework's capability, and must allow the analyst to find unanticipated and interesting dynamics at play. R and Python packages and libraries have been used successfully in the recent past to build information rich, visually compelling presentations.

3.2 Use Case 2: State-Level Energy-Water Sankey Diagrams

The key goal for Use Case 2: State-Level Energy-Water Sankey Diagrams was to visualize the interconnections between energy and water in the broader context of both systems, and to do so at a regional level to inform state-level energy and water policymakers. Eight components of the Energy-Water nexus are examined for this use case:

Energy and Water Comoponents Examined with Sankey Diagrams in the EWN-KDF

- **Energy Production:** coal, oil, gas, wind, biomass, solar, geothermal, hydro, nuclear
- **Energy Conversion:** electricity, fuel
- **Energy Use:** residential, commercial, industrial, transportation
- **Energy Disposition:** sector-wise efficiencies
- **Water Withdrawals:** fresh, saline, surface, ground
- **Water Treatment:** municipal, wastewater
- **Water Use:** residential, commercial, industrial, agricultural
- **Water Disposition:** consumption, surface discharge, ocean discharge, injection

Additionally, interactions between water and energy are investigated:

1. Water Use for Energy: thermoelectric cooling, water impacts of oil and gas production (water/stream injection for production/stimulation and produced water management), irrigation of biomass crops, oil refining, and water use in biofuel refining.
2. Energy Use for Water: municipal water supply and treatment, municipal wastewater treatment, water conveyance, pumping energy for irrigation.

The following data sources and additional data are integrated to meet specific analysis needs for the visualization.

- Energy Information Administration (EIA) State Energy Data System
- US Geological Survey (USGS) Water Use Survey: 2010 and 1995
- EIA Cooling System Water Use Data
- US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Census of Agriculture for 2012 and 2007 (irrigation, groundwater depth)
- USDA NASS Quick Stat data (total grain harvest)
- American Water Works Association (AWWA) 2011 Report Vail report on water management in unconventional oil and gas operations
- Tidwell report on water conveyance energy (Western states)
- Wu report on water-intensity of conventional oil and gas
- Electric Power Research Institute (EPRI) report on water treatment energy intensity

An iterative process of data collection, fusion, analysis and stakeholder review reveals that there are data gaps at the scale, resolution and extent necessary for a highly accurate and detailed examination. These gaps are discovered by working towards specific project goals and the constraints of a Sankey Diagram, which enforces resource conservation and does not handle "looping" gracefully. User questions point to future data collection and analysis needs. For instance, a user may wonder if “annual and state” is really the right temporal and spatial resolution for an analysis of municipal supply and demand when infrastructure constraints may only be relevant during seasonal extremes.

The proposed Energy-Water Nexus Knowledge Discovery Framework could provide a much-improved workflow for creating state-level Energy-Water Sankey Diagrams. The current framework for diagram generation is driven by the diversity of data sources. For example, EIA data can be accessed via API, USGS data is structured Excel (non-normal-form), and many other data are in printed (or unstructured .pdf) tables. Recently, these datasets have been integrated via "PowerPivot," an extended capability built into Microsoft Excel, which enables database-like access to users. This framework encourages (does not force) pre-conditioning data into normal-form. Visualization is accomplished with the commercial eSankey software, reads Excel outputs via an industry-standard connector.

However, this process is labor-intensive in both data import and diagram cleanup, causing Sankey Diagrams to continue to be labor-intensive visualizations. Thus, the desired capabilities for creating Sankey Diagrams in an Energy-Water Nexus Knowledge Discovery Framework include:

- Data Browser with Provenance Management
- Access to online data sets
- Auto/manual updates, history, rollback capabilities
- Manipulation of file-based data sets
- Creation from publications and tabular data
- Local, cloud and "original source" storage.

These capabilities must be combined with a flexible framework offering the ability to write code, perform analyses and accomplish integration in multiple environments (R, Python, SaaS, Excel, etc.). The system should include capability for full version control, documentation of data, code/analysis, and outputs with automatic/suggested unit conversions and re-scaling, geographical aggregation/disaggregation and automatic temporal interpolation (Figure 5).

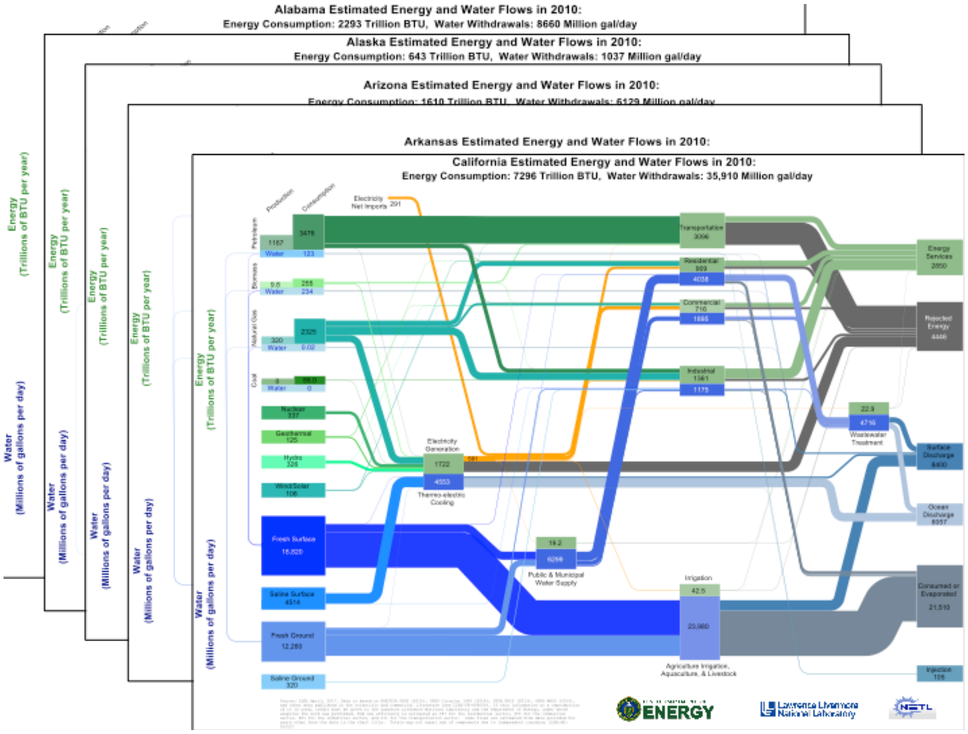


Figure 5. Automated Generation of Sankey Diagrams

3.3 Use Case 3: Water Demand and Supply in Shale Oil and Gas Production

The third use case comes from a multi-year interdisciplinary study of shale gas and oil resources, their future development and their associated water and land footprints. The study was funded by Sloan and Mitchell Foundation to develop the foundational model needed to incorporate new energy-water dynamics. The project was informed and reviewed by the US EIA, USGS, the United Nations Economic Commission for Europe (UNECE) Resource Classification Expert Group, Industry operators, academics, and key non-governmental organizations (NGO).

The study focused on the water-energy nexus of energy resource production, including the following procedures:

1. Water Withdrawal
2. Hydraulic Fracturing (HF) Water Mix
3. Flowback and Produced (FP) Water
4. FP water Treatment and Disposal

Data collection for the study comprised data per well on HF water use and FP water production on an annual basis associated with well location and geologic formation.

Monthly data were used when (rarely) available, and water data were often derived based on production tests. Oil (liquid, condensate) and gas production data were also collected (or derived in the absence of data availability). Data that were lacking included HF water sources, accurate monthly FP water, disposal methods (injections, ponds, processing) and HF and FP water costs.

Results from the study showed both economic and environmental implications. Given the increasing demand for water (oil and gas development, irrigation, power generation, household consumption) and volatility in climate (droughts), water resource rationing (and associated pricing) might be needed in the future (Figure 6). Currently, the producers and the regulators lack information on water processing and disposal (volumes, characteristics, local infrastructure challenges). Consequently, generation of applicable statistics is difficult and comprehensive evaluation is not feasible. Additionally, with increase in operational scale, producers are vulnerable to regulatory risks and are prone to improve technology. Economic signals would help all the sides to value the resource and handle it well.

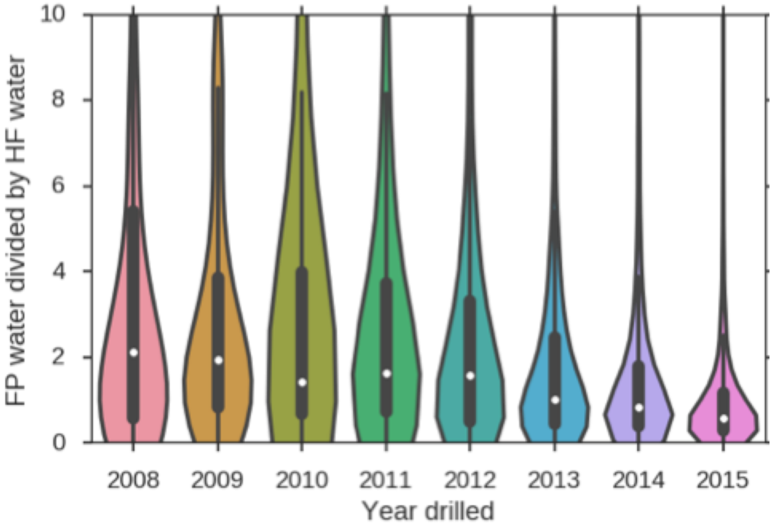


Figure 6. Eagle Ford Shale Play: Implications of New Trends

4. SUMMARY OF BREAKOUT GROUP RESULTS

Participants in the workshop represented a range of stakeholders in the Energy-Water Nexus (EWN) research community. They were assigned to three breakout groups, ensuring that each group had experts in EWN research, data science, and analysis and visualization. Each group brainstormed on additional use cases and was assigned one illustrative use case from the morning session for further analysis. Designated facilitators led these sessions to generate lists of potential use cases focusing on specific questions that the EWN-KDF could answer, lines of scientific inquiry that the KDF could support, or stakeholder objectives that can be met with analysis enabled by the KDF.

4.1 Group 1: Energy and Water Infrastructure Planning and Risk Assessment

Group 1, led by Svetlana Ikonnikova, discussed a use case centered around energy and water infrastructure planning. For this use case, it was determined that analysis on variable time and geographical scales was needed to support investment and regulation decisions. Types of potential users identified for this use case included regulators (“Is supply interrupted in a city/community/state? For how long? Do we need to build additional infrastructure?”), researchers (“Is uncertainty in economic development associated with possible energy-water infrastructure bottlenecks and interruption risks?”), utilities (“How do I prepare against extreme events given different scenarios and their probabilities; what are the risk distributions associated with different pieces of infrastructure?”), energy and water managers (“What is the potential need for water resource to supply the energy sector and water infrastructure rationing?”) and industry (“Facility siting suitability mapping given the map of possible circumstances?”). Key elements of the Energy-Water Nexus are identified as described in the Sankey Diagrams. For the energy sector, production, conversion, use and disposition must be considered. For the water sector, withdrawal, treatment, use and disposition must be considered. Then interaction of these two infrastructures in both directions can be analyzed. Data, data services, analytics and gaps in these needs are summarized in Table 1.

Ideally, the community would like access to a system that has the ability to absorb and handle economic, weather, and climate shocks and shifts affecting energy-water production and consumption, infrastructure functioning, and population dynamics. Such a system could inform operational risk analysis regarding average, extreme and frequent extreme events and could help planners strategize for optimum resiliency of energy and information systems. The system should be able to help balance public vs. industry interests and respond to correction by regulators.

Components of the system should include capabilities for answering questions regarding vulnerabilities associated with power plant siting, population/demand shifts, climate/supply shifts, geographical considerations and extreme weather risks with an eye toward long-term policy contractual and infrastructure plans. These evaluations require integration of mapping layers, aggregation, disaggregation, interpolation, extrapolation, subsetting, formatting and exporting capabilities, along with associated key statistics. Additionally, data from both modeling and measurement should adhere to accepted standards. The platform should offer query capability of data robustness based on these standards.

Existing platforms are available, such as GeoCollaborate Google Earth Engine, The Atmospheric Radiation Measurement (ARM) facility and the Earth System Grid Federation (ESGF) among others. These offer many of the capabilites required by the Energy-Water Nexus including algorithmic processing, statistics and programming interfaces (R, Python). The most important new capability that the EWN-KDF can provide is reasonable integration of all of the types of data that are served on the existing platforms to enable deeper investigation of infrastructure and environmental interdependencies and provide decision support.

Table 1. Energy and Water Infrastructure Planning and Risk Assessment

Key Element	Data	Data Services	Analytics	Gaps
Energy Production	Infrastructure	Download and query access to geographic location and capacity	GIS mapping-models	Data standardization (e.g. P5 and P95), Measurement uncertainty characterization
	Regulatory, Economic	Subsetting, exporting, formatting data from various sources		
Energy Conversion	Infrastructure	Download and query access to geographic location and capacity	Siting, suitability and mapping	
	Regulatory, Economic			
Energy Use	Population/Demographics	Aggregation/Disaggregation		
	Demand/Consumption	Download and query access to historical and projected climate and weather data including probability of extreme events (primarily temperature)	World Spatio-Temporal Analytics and Mapping Project, Earth System Grid Federation	
Water Withdrawal	Availability	Download and query access to historical and projected climate and weather data including probability of extreme events (primarily precipitation),	GIS mapping-models, Earth System Grid Federation	
	Allocation, Regulatory, Economic	Extrapolation/Interpolation		
Water Use	Population	Aggregation/Disaggregation	World Spatio-Temporal Analytics and Mapping Project	
	Demand/Consumption	Gridding/regridding consistency in space and time		
Water for Energy	Systems Characteristics	Uncertainty Quantification/Propagation	GeoCollaborate	Data fusion
Energy for Water	Systems Characteristics	Uncertainty Quantification/Propagation	Google Earth Engine	Data fusion

4.2 Group 2: Comparing Impacts of Drought on the Operation of Electric Generation Units in WECC and SERC Balancing Areas

Led by AJ Simon, this group discussed a use case comparing impacts of drought on the operation of electric generation units. Two specific balancing areas, the Western Electricity Coordinating Council (WECC) and the Southeastern Electric Reliability Council (SERC), were considered so that data and analytical tools appropriate for this scale could be tested in widely varying conditions (climate, weather, terrain, energy mix, demographics, etc.). The group determined that a state-level economic spatial price equilibrium simulation model based on the input-output dataset underlying the Energy-Water Sankey diagrams could inform policymakers and is within the capabilities of the research community. However, it was not clear what specific user needs the resulting model would satisfy. An alternate approach proposed was to begin by analyzing historical interdependence of water withdrawals and electricity generation under different hydrologic regimes and with assets that have different cooling technologies. This type of analysis would need to be performed at the unit level–the primary locus of resilience.

To make the analysis still more complete and sophisticated, the group suggested that the demand side should be included, along with electric power dispatch, in an attempt to capture resilience from re-allocating generation across assets when total capacity is constrained. However, these additions would require simulation, and an integrated model too large and complex to run on what should be a data integration/exploration/analysis platform.

The EWN-KDF will be important to increasing the resilience of energy infrastructure to natural hazards. In particular, exploration of this use case would lead to understanding, under today's electric power system, what effects projected drought could have on electricity generation and capacity. The EWN-KDF could provide a first step toward:

1. Predicting and preparing for supply constraints/curtailments in the current climate
2. Optimizing investment in future assets under climate change

For this and many use cases, the key tradeoffs are among fidelity in the representation of complex interactions, the resolution at which these interactions are captured, and the spatial and temporal scale of the overall analysis. Thus, for robust analysis, users need well-bounded and tractable use case. The main goal should be to catalyze the virtuous cycle of:

Analytical Needs -> Data Gathering -> Further Questions/Analysis...

It is important that analysts have the flexibility to perform analyses of their own design. The group recommended that R and/or Python (and iPython/JuPyter notebooks) interfaces be offered as these frameworks provide analytical flexibility and provenance tracking for both data and analysis. Further provenance management tools for these user-created analyses should also be developed along with available verified open-source code options. Key to the framework is reduction in labor intensity of data management while maintaining transparency, replicability and high quality.

Table 2. Impacts of Drought on the Operation of Electricity Generation Units

Key Element	Data	Data Services	Analytics
Energy Conversion	High Temporal Frequency Generation at the Asset Level: EPA CEMS hourly generation by thermal unit, stratified by cooling type, Monthly hydropower generation	Understanding hourly hydro dispatch, relationship to monthly data, data coverage	How to gap fill missing data on groundwater supplies to thermal units
Energy Use	Population/Demographics	Data cleaning/gap filling, Not having to download raw data for own use	Modeling network effects on the demand side: electricity grid responses to perturbations
	Demand/Consumption	Linking Meteorology, Climate Projections	
Water Withdrawal	Availability	Linking Meteorology, Climate Projections: NWIS stream flow data for gauges proximate to units, UNL Drought Monitor historical data, Reanalysis data to define drought events, Earth system model projections (ESGF/NASA NEX/NARCCAP)	Transform and merge ESM data with emulators to generate projections of supply constraints including thermal conditions
	Allocation	EIA 923 2014-15 water withdrawals	
	Regulatory		Modeling relationship between discharge/cooling water availability/reservoir management under future drought conditions (e.g., climate change)
Water for Energy	Fusion tables among assets (units by cooling type, reservoirs), geographic grids and networks (streams, transmission)	Data QA/QC and provenance management for each source, Algorithms for spatial/temporal merging data sources for each observation, scalability to incorporate new datasets as they become available, Spatial/temporal joins of databases	Ability to run large-scale (1 million obs.) cross-section/time-series regressions to construct statistical emulators

4.3 Group 3: Contrasting two regions in transition, West and Northeast

Discussion in this group was led by Kate Calvin and it dealt with the implications of alternative energy pathways under shocks, stressors and a changing resource base, first as applied to water shortages (drought) and then as it pertained to other climate extremes. The necessity of this use case and its value to society was clear in that extreme events cause failures and demonstrate vulnerabilities. Types of potential users identified for this use case included energy utilities, utility commissions, grid operators, governors, pipeline operators and distributors. The main ideas behind this use case have to do with regions in transition over climate, population and infrastructure.

Data service needs emphasized were those that harmonized cross-sector data and facilitated cross-field analysis. Questions asked for this use case built on the GCAM case study, which considered water as a constraint in the context of resource competition. Questions include:

- How do energy efficiency and renewable energy affect demand for thermoelectric generation and its water footprint?
- How can energy and water be decoupled?
- From utility perspective: how do we reduce demand for energy and how do we produce energy from the resources we have? (DC Water produces a lot of its own energy. Aeration is a specific point of interest because it is the largest energy user and doesn't necessarily have to be driven by electric pumps.)
- What are the constraints and enablers for co-evolution of existing heterogeneous patterns of infrastructure?
- What are the, physical, institutional, governance and cost constraints for capital and operating?
- Can the problem be approached with institution-based user topology?

Exploratory capabilities should allow the user to investigate:

- Characteristic changing regional stressors, including extreme events
- Shocks and gradually evolving stressors
- System benefits vs. individual benefits
- Risk of failure: evaluation of robust decision making approaches
- State 100% renewables targets: how water availability, variability, and policy influence the potential of these targets
- Can we (as a society) afford to meet our coupled water and energy demand today, and will we be able to in the future?
- If state X goes to Y% renewables by year Z, what are the implications for coupled energy-water dynamics within and beyond that state?
- Implications of alternative energy pathways under shocks, stressors, and changing resource base: applied to weather shortages, but with relevance to other extremes

Table 3. Contrasting two Regions in Transition: West and Northeast

Key Element	Data	Data Services
Energy	Current and Past Infrastructure, supply, demand, distribution, Past, Present and Future Climate (trends and extremes): NLDAS + NARCCAP or NEX30, Land Use, Population, Demographics, Regional and Subregional Economics, Technology Profiles	Data Ontology, Provenance and Documentation of Processing; Spatial, Temporal and Process Alignment across Fields and Sectors
Water	Current and Past Infrastructure, Water Use, Water Rights, Water Allocation, Water Supply, Past, Present and Future Climate (trends and extremes): NLDAS + NARCCAP or NEX30, Land Use/Permeability, Regulations/Prohibitions, Water Quality,	
Water for Energy	Stream Temperature, Environmental Flow Requirements (including variability), Water Requirements for Specific Energy Plants, Fine Scale Energy Demands for Water	
Energy for Water	Water Conveyance	

5. HIGH PRIORITY NEEDS AND APPROACHES

Common to the discussions in all of the breakout groups were several identified data and analytical needs. These include:

- Access to observational and derived data sets to parameterize, test, apply, and evaluate models.
- Application of data and analytics at appropriate scales to relevant problems.
- Facilitation of work outside the boundaries of individual models that integrates across disparate physiographic, infrastructure, and socioeconomic systems.
- Data and analytical capabilities for the determination of future regional trajectories for climate, population growth, land use, economic activity, and energy technologies and how they scale over space and time, as well as potential innovations in technology, adaptation, and resilience options

Thus, several developmental priorities in three categories emerged: 1) Data Content and Knowledge Base, 2) Analytical Toolkit, and 3) System Architecture for Data Management.

Three immediate tasks in the Data Content category include:

1. Building an inventory of important data sets, models, visualization tools, and decision support needs at various multimodal and multiresolution scales. The team is surveying the literature to discover these data and tools and to identify critical data gaps.
2. Developing common formats for reconciling gridded and non-gridded data, including metadata standards for geodetic and temporal indexing with cross-sectoral reconciliation. This development is underway, and a demonstration is under preparation.

3. Defining and maintaining standards for quality control, evaluation, calibration, validation and uncertainty analyses for data and modeling.

For the Analytical Toolkit, a first set of in-house analytical tools is undergoing conversion for compatibility with the new system, ultimately to be exposed through a collection of web (RESTFul) and non-web APIs. Among the first tools to be included will be relevant components currently existing in the ORNL World Spatio Temporal Analytics and Mapping Project and in the SUNY/Buffalo WebWorldWind application collection.

Finally, priorities identified for the System Architecture include:

1. Publishing and archiving capabilities for data
2. Support for analysis over heterogeneous data sets
3. Effective access mechanisms across facilities (universities, laboratories, agencies) enabling a virtual laboratory and collaborative ecosystem.

To these ends, initial steps have been taken. Using the example of the Earth System Grid Federation (ESGF) as a starting point, a system that leverages primary data holders, modeling groups, and decision support activities is under construction. The ORNL Compute and Data Environment for Science (CADES) computer, which includes HPC resources, scalable storage, data analysis, and visualization tools, is hosting several virtual machines for users. The CADES infrastructure supports multiple data security levels with separate open-research as well as secure network enclaves, and is combined with the ANL Globus services to provide immediately useful data transfer, sharing, publication, and discovery capabilities, security and scalability.

6. PATH FORWARD

The development of an integrated, multi-layered, federated EWN-KDF provides an unprecedented opportunity for users to coalesce and interrogate a single knowledge system for robust, data driven analysis and visualization.

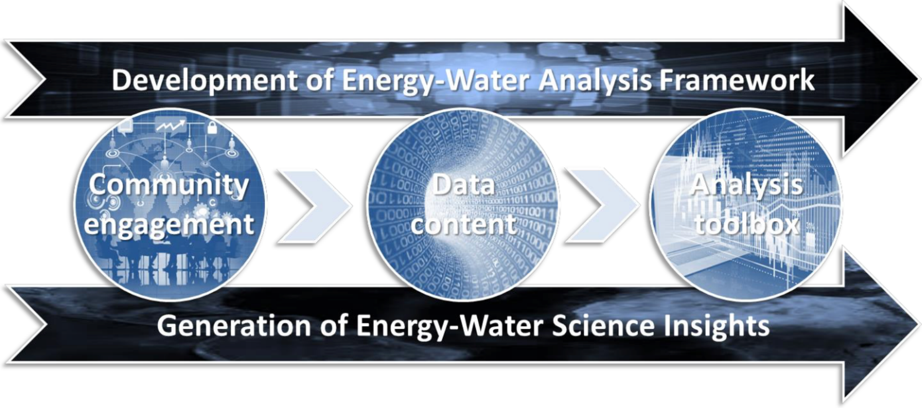


Figure 7. Energy-Water Nexus Knowledge Discovery Framework Path Forward

The most important component in the path forward for the development of the Energy-Water Nexus Knowledge Discovery Framework is continued guidance from the expert user community. Collaborative development ensures the framework's immediate and ongoing relevance and usefulness to the community. This Experts' Meeting has provided key insight into the first priority needs from the system, and will continue to drive its overall development.

Meeting participants also expressed the need for the network to demonstrate its emerging capabilities. Well-designed, rigorously posed and tractable "challenge problems" can drive the development of features and capabilities that can later be expanded to serve the needs of broad scientific inquiry. Such small-scale demonstrations will drive stakeholder engagement, elicit feedback and spark further innovation in energy-water data management. It may be instructive to replicate one of the use cases explored in the meeting (GCAM data, Sankey Diagrams, or unconventional oil and gas data) within the EWN-KDF both to test the new framework against a known challenge and to enhance an analysis result that has broad awareness. The needs expressed by the user community, summarized here, include several key design principles, which will serve as beacons for each incremental accomplishment.

Accessibility: Users of the EWN-KDF should be able to access, manipulate, integrate, analyze, and visualize a broad range of federated data assets relevant to the EWN. This includes not only data on energy and water, but also data regarding other assets, outcomes, or processes that influence, or are influenced by, energy-water dynamics. The framework should be able to accommodate differential access by different classes of users to create an analysis environment for sensitive or proprietary information.

Interoperability: Users should be able to work across disparate data assets and formats including the integration of raster and vector data over varying spatial and temporal scales. In addition to using diverse data, users should also be able to work across different data systems. This includes seamless connections to data hosted by a variety of agencies and organizations.

Flexibility: An EWN-KDF should be able to address diverse analysis challenges by applying statistical, machine learning, as well as other methods to federated as well as user-defined data resources.

Integrity: EWN-KDF users should also be able to access metadata that describes data sources, uncertainties, and quality as well as the provenance of those data and to track changes over time.

Reliability: An EWN-KDF should be a stable analysis platform with minimal service outages and one that generates robust, reproducible results.

Security: The physical and cyberinfrastructure of an EWN-KDF should be secure in order to maintain user profile and scientific data integrity, as well as positive control over system accessibility.

Sustainability: An EWN-KDF should be durable and persistent with the capacity to be readily expanded over time as data storage and compute needs evolve, new data assets and analysis tools become available, and research questions shift to address emerging priorities.

6.1 Data Search, Curation and Provenance

Finally, the Energy-Water Nexus Knowledge Discovery Framework should be able to help the community understand and track data in terms quality level, disagreement across data sets, and gaps across the span of data and within specific data sets.

6.1.1 Data Search and Quality

The architecture should support elastic and faceted search to help users find the data needed to achieve their goals. It should also provide completeness measures, access to metadata, provenance, and other dataset documentation so that users become familiar with the semantics of the data. Data tools within the architecture should provide capabilities for transforming and preparing the data for the use with analytical services provides and/or researchers' own models and tools.

6.1.2 Provenance and Reproducibility

Capturing the provenance of data as scientific discoveries are made, and the successful reproduction of experiments are important. Tracking of provenance should be accomplished using data standards, metadata processing, standardization, and provenance trees. Emerging technologies such as Docker, in combination with virtual notebooks enabled with provenance support, should be employed to allow reproducibility of scientific experiments to facilitate true cross-domain analysis and retain its availability.

6.2 First Steps in the Development of a Usable and Testable Framework

Version 1 of a multi-layered federated platform, the Energy-Water Nexus Knowledge Discovery Framework (EWN-KDF), has been developed and is undergoing testing. This platform utilizes several enterprise-grade software design concepts and standards such as extensible service-oriented architecture, open standard protocols, an event-driven programming model, an enterprise service bus, and adaptive user interfaces to provide a strategic value to the integrative computational and data infrastructure. EWN-KDF is built on the Compute and Data Environment for Science (CADES) environment at Oak Ridge National Laboratory (ORNL). The system contains a Geoserver component which facilitates exploration of the datasets, and is built on Globus platform services, which allow swift and seamless data transfer from vetted sources and enable protected sharing of results with other researchers. Three use cases have been defined and will be tested within the framework using the analytics it provides. Publication of the results obtained by these use cases, along with the framework itself will be made available to the expert community for further evaluation and improvement.

7. APPENDIX A: EXAMPLE DATA SETS

Table 4. Some Important Data Sets for Energy/Water Analysis

Category	Type	Description	Source
Physiographic	Watershed Boundaries	Polygon shapefile: Watershed Boundaries (WBD); time invariant	Natural Resources Conservation Service (NRCS)
		Polygon shapefiles: 4,6,8,10, and 12 digit HUCs	NRCS
		Polygon shapefiles: National hydrography dataset (NHD) catchments	Horizon Systems
	River Network and Spatial Features	Line shapefiles: NHD medium resolution stream reaches (1:100K), includes topology for routing	Horizon Systems
		Line shapefiles: NHD high resolution stream reaches (1:24K), topology exists but no toolsets	US Geological Survey (USGS)
		Polygon shape files: Reservoirs, NHD high resolution	Horizon Systems
		Derived point or line shapefiles of odd water conveyance, NHD high resolution	Horizon Systems
		National Inventory of Dams: characteristics of 87,000 dams in the US	US Army Corps of Engineers (USACE)
	Hydrology	NetCDF files of climate reanalysis at various spatial and temporal resolution	National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL)
		NetCDF files of climate reanalysis at 4km spatial and monthly temporal resolution	PRISM, Northwest Alliance for Computational Science and Engineering
		NetCDF files of climate reanalysis at 1km spatial and daily temporal resolution	Daymet, Oak Ridge National Laboratory (ORNL)
		Raster files: digital elevation maps (DEM) of topography, 30 meter spatial resolution	National Elevation Dataset
		Raster files: digital elevation maps (DEM) of topography, at < 10 meter spatial resolution	USGS
		Geodatabase files: National Land Cover Database (NLCD) 16-class land cover data for 1992, 2001, 2006 and 2011 at 30meter resolution	Multi-resolution Land Characteristics Consortium

		Point location streamflow discharge readings	USGS
		Geospatial layer of Federal Emergency Management Agency (FEMA) 100-year flood zones	FEMA
		Archive of point location runoff, flood, drought readings	USGS
	Climate	Text files of weather and climate measurements, 1800s to current, hourly, daily, monthly, ground station, radar, satellite measurements	National Centers for Environmental Information (NCEI)
		NetCDF files of historical and projected climate conditions	Earth System Grid Federation (ESGF)
Infrastructure	Water Use and Consumption	Excel files of water use by category and sector in the US at the county level, 5-yr temporal resolution	USGS
		Line shapefiles of water returns (volume and location)	Facility Registration System, National Pollution and Discharge Elimination System (NPDES)
		Excel files of powerplant water use, consumption and locations, monthly generator data, 2013 - 2015	EIA (Form 860)
		Excel files of water quality sampling, observation and measurement results (state, county, HUC), 1900 - 2017	Environmental Protection Agency (EPA) STORET
	Energy Production and Consumption	Excel files of powerplant production, monthly generator data 2013 - 2017	EIA (Forms 860, 906, 920, 923)
		Excel files of state electricity generation and consumption by source and sector, 1960-2015	EIA State Energy Data Systems (SEDS)
		Polygon shape files of areas managed by utilities for customers	PLATTS
		Excel files of electric utility sales by census block	EIA Annual Energy Outlook (AEO)
		Excel files of electric utility customers per census block	EIA AEO
		Excel files of US commercial buildings energy consumption and expenditures, electricity, fuel and heat consumption, 1992 - 2012	EIA Commercial Buildings Energy Consumption Survey (CBECS)
		Excel files of US residential energy consumption and expenditures, electricity, fuel and heat consumption, 1993 - 2015	EIA Residential Energy Consumption Survey (RECS)
		Point shapefiles of location and capacity of US substations	Homeland Infrastructure Foundation-Level Data (HIFLD)

		Point shapefiles of location and capacity of US powerplants	HIFLD
Socioeconomic	Population and Demographics	Polygon shapefile of Census blocks	Tigerline/US Census
		LandScan raster files of world population count at 1km resolution (2005 - 2015) and US population at 90m resolution (2007 - 2015)	Oak Ridge National Laboratory (ORNL)
		LandCast raster files of projected US population count at 1km resolution (2030, 2050)	ORNL
		Global Population of the World WMS files of population count and population density at 1km resolution (2000 - 2020), 5-yr temporal resolution	Socioeconomic Data and Applications Center (SEDAC)
		Excel files of US population estimates, decadal counts, annual interpolation	US Census
		Comma delimited text files of US population, demographic and housing summary data, block group spatial resolution, annual temporal resolution	Ammerican Community Survey (ACS), US Census

8. APPENDIX B: INVITED SPEAKERS

Budhendra Bhaduri is the Director of the Urban Dynamics Institute, a Corporate Research Fellow at Oak Ridge National Laboratory (ORNL), and leads the Geographic Information Science and Technology (GIST) group within the Computing and Computational Sciences directorate. He is a founding member of the U.S. Department of Energy's Geospatial Sciences Steering Committee. His primary responsibilities include conceiving, designing, and implementing innovative geocomputational methods and algorithms to solve a wide variety of national and global problems involving population dynamics modeling, natural resource studies, transportation modeling, critical infrastructure protection, and disaster management.

Justin Jay Hnilo is the Program Manager of the Department of Energy Biological and Environmental Research Climate and Environmental Data Management office. Formerly of Lawrence Livermore National Laboratory, Jay developed original climate model diagnostic applications and enhancements using traditional as well as new statistical approaches. He has written advanced diagnostics for model data used domestically and internationally. He has implemented and maintained observational databases, established metadata standards and data file structures, and performed quality control. He has designed techniques to merge and co-locate/geonavigate satellite, insitu and model data, and has written code to quantify and visualize both observational and model-based data to facilitate direct comparison.

Kate Calvin is a research economist working at the Pacific Northwest National Laboratory's Joint Global Change Research Institute (JGCRI) in College Park, Maryland. She has broad expertise in developing and applying computational modeling tools to inform decision-making about complex questions related to food security, water scarcity, energy use, and the impacts of climate change on vital resources. At the Joint Global Change Research Institute, she works with the Global Change Assessment Model (GCAM), an integrated assessment tool developed at PNNL with the support of the U.S. Department of Energy's Office of Science. GCAM enables researchers to explore the drivers, consequences, and responses to global change, taking into account all sectors of the economy and all regions of the world.

AJ Simon leads the Energy Group at Lawrence Livermore National Laboratory. His research focuses on systems analysis and technology assessment of secure and clean energy and water solutions. Recent energy systems research has included wind power forecasting, the water/energy nexus, and nuclear risk analysis. Further research in the group's portfolio spans carbon capture and sequestration, geothermal reservoir management, advanced nuclear fuel cycle analysis, and resource utilization analyses.

Svetlana Ikonnikova is a Research Associate / Energy Economist in the Bureau of Economic Geology at the University of Texas at Austin. Her key interest is in natural gas and natural gas market developments. In the past she studied supply networks delivering natural gas from the Former Soviet Republics to Europe and LNG world market. At present her research is focuses on shale gas, its production outlook and market implications.