SciDataCon 2025



Contribution ID: 267

Type: Poster

Title: Enabling Trustworthy and FAIR AI for Transboundary Aquifer Resilience: Challenges and Opportunities for Reproducible, Responsible, and Open Science

Monday 13 October 2025 19:10 (20 minutes)

Artificial intelligence (AI) offers powerful potential to address pressing challenges in transboundary water management, especially in regions with insufficient infrastructure for in-situ water quantity and quality monitoring and modeling. However, the successful application of AI in this context depends on more than algorithmic accuracy and can be challenging to achieve even in a system with robust data that follows best practices to ensure AI readiness. In transboundary regions where groundwater resources are shared by more than one country, data collection and standards can vary dramatically across borders. Challenges associated with ensuring reproducibility and interoperability are exacerbated for transboundary systems due to this reality. These issues intersect directly with broader themes of responsible science in the era of FAIR (Findable, Accessible, Interoperable, and Reusable) data and AI.

The Groundwater Resilience Assessment through iNtegrated Data Exploration for Ukraine (GRANDE-U) project exemplifies both the promise and complexity of such transboundary collaborations. Focused on aquifer resilience in Ukraine and neighboring countries, the GRANDE-U project brings together researchers from the U.S., Ukraine, Poland, Latvia, Lithuania, and Estonia, with support from the U.S. National Science Foundation and parallel research agencies in the European partner countries. It combines physics-based and machine learning models with satellite data to monitor and predict groundwater storage and flows across national borders. This integrated approach hinges on the willingness and ability of each country to share data openly, follow interoperable data standards and database conventions, and conduct joint data analysis and AI modeling, leveraging the complementary expertise of the country teams.

The FAIR data principles provide a crucial foundation for such collaborations. Data collection protocols and hydrogeologic descriptions vary across GRANDE-U's partner countries, as does the density of in-situ observations. Thus, the GRANDE-U project has prioritized the development of a consolidated spatio-temporal database of satellite and in-situ groundwater and surface water observations. The database is structured to support consistent and interoperable data standards across the transboundary regions to create an AI-ready foundation.

The database schema was iteratively refined with feedback from all partner countries, ensuring alignment with both scientific objectives and policy priorities. This process included ensuring clarity across the international research team on data provenance, modeling assumptions, and validation methods. Reproducibility was prioritized across every stage and component of the research. All machine learning workflows were shared as executable Jupyter notebooks, detailing training data exploration, feature engineering, hyperparameter tuning, and spatial transferability. These transparent modeling practices were reinforced through technical webinars.

This collaboratively developed database serves as the foundation for two critical components: downscaling satellite-based groundwater estimates using local hydrogeologic context, and extending AI models to broader transboundary regions. A major milestone was the development and validation of novel algorithms to downscale GRACE/GRACE-FO satellite data, with successful application along the Poland-Ukraine border. These methods leveraged high-resolution geologic, topographic, and land cover datasets, and in-situ groundwater monitoring wells, to achieve accurate modeling using a range of machine learning models, including random forest regressor and boosting techniques (Pearson R > 0.8 in porous aquifers). The results underscore the potential of machine learning to fill observational gaps and enable groundwater modeling in regions with sparse field data —capabilities particularly crucial for conflict-affected areas such as Ukraine.

Analyzing collaboration networks in transboundary groundwater research offers valuable insights for strengthening international partnerships and designing more effective workforce development strategies. By identifying key contributors, interdisciplinary linkages, and institutional gaps, such network analysis can guide the organization of complementary expertise across hydrogeology, remote sensing, and AI, and foster scientific ecosystems that leverage unique expertise to move innovation forward in ways that are not possible otherwise. In GRANDE-U, this approach has been used to study the evolving structure of global groundwater research networks, leveraging bibliographic data and interactive visual analytics. These findings have directly informed the project's training and capacity-building efforts, including webinars and workshops for students and early-career researchers focused on reproducible AI workflows, FAIR data practices, and spatial modeling. Participants from Ukraine, the U.S., and multiple European countries have engaged in these sessions, helping build an internationally connected, interdisciplinary community. By making training materials and data pipelines openly available, GRANDE-U advances a culture of transparency, collaboration, and responsible data stewardship across borders.

While AI offers transformative opportunities for sustainable groundwater management across borders, its benefits can only be realized within a purposeful framework that emphasizes rigorous, responsible, and reproducible science. The GRANDE-U experience demonstrates how such a framework can succeed by prioritizing FAIR data and open research practices, and by actively engaging diverse transdisciplinary teams in collaborative modeling and knowledge-sharing across political and institutional boundaries. GRANDE-U leverages state-of-the-art practices catalogued by the FAIR in ML, AI Readiness, AI Reproducibility (FARR) Research Coordination Network, especially for maintaining AI readiness in its data products.

GRANDE-U support under the NSF IMPRESS-U program (awards 2409395 and 2409396) is gratefully acknowledged.

Primary authors: ZASLAVSKY, Ilya (San Diego Supercomputer Center, UCSD); ATKINS, Ashley (San Diego Supercomputer Center, UCSD); KIRKPATRICK, Christine (San Diego Supercomputer Center / CODATA)

Presenter: ZASLAVSKY, Ilya (San Diego Supercomputer Center, UCSD)

Session Classification: Poster Session

Track Classification: SciDataCon2025 Specific Themes: Rigorous, responsible and reproducible science in the era of FAIR data and AI