

# Integrating Water Data to Empower Alaska’s Rural Communities

Rachel Lewis<sup>1</sup>, Noah Tsigonis<sup>2</sup>, Arghya Kusum Das<sup>1</sup>

<sup>1</sup>University of Alaska Fairbanks, <sup>2</sup>Arctic Outlook

Fairbanks, AK, USA

{rlewis57, akdas}@alaska.edu, noah.tsigonis@gmail.com

**Abstract**—Alaska’s fragmented water and wastewater sector creates significant challenges for residents and professionals, limiting access to critical resources, support networks, and opportunities for skill development—ultimately affecting safety and efficiency. To address this, we developed a unified hydroinformatics platform designed to serve remote and underserved regions of Alaska by providing essential information, education, and tools that foster safer, healthier communities. This platform integrates water, wastewater, and solid waste data from 11 disparate silos maintained by various government and private entities, each with unique formats and structures. Using Selenium-based web scraping, we automate data collection to ensure real-time updates with minimal manual input. Key features include a user-friendly file upload interface, an intuitive query tool accessible to non-technical users, and a MongoDB backend for efficient data storage and management. Our solution streamlines the organization and dissemination of up-to-date sanitation systems information, supporting informed decision-making. Currently, the platform consolidates data from over 2,000 water operators, 1,500 drinking water systems, 180 wastewater profiles, and 900 solid waste profiles, impacting more than 200 communities statewide.

**Index Terms**—Hydroinformatics, Water information reuse and integration, Social impact of technology.

## I. INTRODUCTION

Hydroinformatics integrates water science, data science, and computer science [1]–[3]. The rise in data volume and diversity enables evidence-based decision-making, improving the management, efficiency, and resilience of water systems [4]. We are expanding hydroinformatics to include wastewater and solid waste.

Hydroinformatics holds great promise for Alaska, but progress is hindered by fragmented, non-standardized water data. Five key agencies manage critical information independently, resulting in 11 disconnected sources with limited integration or interoperability. This fragmentation makes it especially hard for communities—many unaware these datasets exist—to access or use the data for informed decision-making and resource management.

To unlock hydroinformatics’ full potential in Alaska, it’s essential to address data fragmentation. We developed a software tool that automates data collection from various agency websites, standardizes it into CSV files with consistent terms, and includes features for file upload, database management, and query submission—making water data access and analysis more efficient.

Our platform streamlines the flow of unstructured data from various sanitation facility databases. Users provide a webpage

link and create a web scraper using pre-made instructions tailored for domain experts like technical assistance providers and hydroinformatics professionals. These instructions guide webpage navigation, form submission, and data extraction. Users can also add loops to collect related data across multiple pages. Submitted instructions are reformatted into tuples used by the scraper. Successful scraping outputs data as a JSON file, which can be downloaded or uploaded to a MongoDB collection. Users can adjust commands and retry if needed. Additionally, the tool supports CSV/Excel uploads, content editing, and direct database entry. A query manager helps restructure and format data using MongoDB commands, preparing it for final use in the platform.

Our solution streamlines access to vital data, empowering communities to make informed decisions and manage resources more effectively. We’ve integrated information from over 2,000 water and wastewater operators into a user-friendly web app, covering over 1,500 drinking water systems, 180 wastewater profiles, 900 solid waste profiles, and impacting over 200 communities across Alaska. By centralizing this data, the platform supports better monitoring, planning, and sustainable water management statewide—especially in remote regions.

Our code is open sourced on GitHub at <https://github.com/akdasUAF/ArcticOutlook>. Our integrated information can be found on Arctic Outlook’s website: <https://www.arcticoutlook.com/>. To share results, Arctic Outlook has engaged a beta test group—including tribes, tribal health organizations, technical providers, and state agencies—to introduce the web portal and gather feedback.

The rest of the paper is organized as follows. Section II describes the works related to our current effort. In Section III, we described the data sources that we used to integrate Alaska’s sanitation facility data. Section IV describes our software development methodology overview. In Section V, we described the impacts on the Alaskan communities. Finally Section VI concludes our paper.

## II. RELATED WORK

There are multiple efforts to make water data available to users and the public, including the Open Water Data Initiative’s goal of developing a holistic model for building an Open Water Web [5]. Efforts have been undertaken previously to develop data models such as the Observations Data Model

to manage storage and retrieval of observation data while keeping a simple relational format [6]. Additional works include works investigating decision tree mining for model sharing [7] and time-series data mining for water resource management research [8]. New interdisciplinary research fields are emerging such as water quality system informatics that employs cybernetics to collect and digitize water quality data [9]. Data processing models are also beneficial for proposals working to ensure water availability via a storage reservoir in the Xiong'an New Area [10].

Web scraping offers a promising approach to automated data collection given the vast online information available [11], [12]. Python-based tools have scraped Google search results to assess water quality in Germany [13] and logged sensor data from Surabaya reservoirs into a PostgreSQL database [14]. For storing such data, NoSQL databases—especially those without predefined schemas—are well-suited for Big Data applications [15]. MongoDB, built for web applications, supports bulk operations, auto-sharding, and high flexibility [15], [16]. Additionally, schema inference from JSON collections has been explored to aid data integration [17].

In this work, we developed Alaska's first tool to unify fragmented water data, making vital information—especially for rural residents—accessible in one easy-to-use location. Like other prior works, we also used Python-based webscraping and a MongoDB database for our data integration.

### III. DATA SOURCES

To build the unified platform, we identified eleven major databases across multiple websites, detailed in Table I and Table II. These databases offer valuable insights into operator qualifications and sanitation system performance, supporting Alaska's public health and water management efforts. The following sections highlight the key information maintained in these databases, which collect, organize, and share vital data related to sanitation systems, utility operations, and community resources. Each source was carefully selected for its role in supporting Alaska's environmental and community stewardship. By centralizing access to this information, our platform empowers agencies, community leaders, and residents to make informed, data-driven decisions, monitor compliance, and optimize resource allocation for both current and future needs. The following subsections discuss the overview of each database.

#### A. Drinking Water Watch Database

The Drinking Water Program within Alaska's Division of Environmental Health, part of the Department of Environmental Conservation, oversees the Drinking Water Watch (DWW) platform. This online database serves as a centralized resource for detailed, system-specific data on public drinking water systems across the state. Key data fields in DWW include the Public Water System Identification Number (PWSID), system name, point of contact, source type (e.g., groundwater, surface water), operational status, and federal classification (e.g., community, transient non-community). These data elements support effective monitoring, regulatory compliance,

and management of public water systems, providing the public with access to water quality and system status information.

#### B. Water- and Wastewater-Operator and System Database

The Operator Certification Program within Alaska's Division of Water, part of the Department of Environmental Conservation, provides an online platform with two key databases: Operator Search and System Search. These tools offer detailed information about certified operators and the water or wastewater systems they manage, supporting regulatory compliance and transparency.

The Water- and Wastewater Operator database provides details on certified operators, including certification levels and the systems they oversee. The Water- or Wastewater-System database offers comprehensive records for each system, including identification, classification, treatment, and assigned operators with their certification levels. Together, these databases provide valuable insights into operator qualifications and system performance, supporting Alaska's efforts to maintain high standards in public health and water management.

#### C. Remote Maintenance Worker Contact List

The Remote Maintenance Worker (RMW) Program within Alaska's Division of Water, part of the Department of Environmental Conservation, provides a comprehensive Community Water and Sewer Improvement Contact Excel List, essential for managing water and sewer systems in remote Alaskan communities. This resource includes up-to-date contacts for key roles, such as the Project Agency Lead overseeing improvement projects, RMW teams responsible for on-the-ground maintenance, Rural Utility Business Advisors (RUBA) offering utility management guidance, Operator Certification Contacts for credentialing, Drinking Water Program Staff for regulatory compliance, and Solid Waste Program Contacts for waste management support. Regularly updated to reflect personnel changes, this list is a vital resource for community leaders, utility operators, and regulatory bodies.

#### D. Well Log Tracking System Database

The Division of Mining, Land, and Water in Alaska's Department of Natural Resources manages the Well Log Tracking System (WELTS), an online platform that consolidates well logs and geophysical data for water wells across the state. WELTS provides detailed records on well construction, geological formations, and water rights supporting engineers, geologists, and water resource managers, enabling stakeholders to access accurate, up-to-date groundwater data for informed decision-making and resource management.

#### E. Remote Monitoring and Energy Audits Databases

The Alaska Native Tribal Health Consortium (ANTHC) supports rural Alaskan communities with two key offerings: Remote Monitoring and Energy Audits, aimed at improving utility management and energy efficiency.

Remote Monitoring allows operators of water and wastewater systems in remote areas to track critical metrics (such

TABLE I  
DATA SOURCES: DATABASE NAMES AND THE WEBSITE

Database name	Website
Drinking Water Watch Database (DWW)	<a href="https://dec.alaska.gov/DWW/">https://dec.alaska.gov/DWW/</a>
Water- and Wastewater-Operator Database	<a href="https://dec.alaska.gov/Applications/Water/OpCert/Home.aspx?p=OperatorSearch">https://dec.alaska.gov/Applications/Water/OpCert/Home.aspx?p=OperatorSearch</a>
Water and Wastewater System Database	<a href="https://dec.alaska.gov/Applications/Water/OpCert/Home.aspx?p=SystemSearch">https://dec.alaska.gov/Applications/Water/OpCert/Home.aspx?p=SystemSearch</a>
Community Worker and Sewer Improvement Contact List	<a href="https://dec.alaska.gov/Applications/Water/OpCert/community-water-sewer-improvement-contact-list.xlsx/">https://dec.alaska.gov/Applications/Water/OpCert/community-water-sewer-improvement-contact-list.xlsx/</a>
Well Log Tracking System Database (WLTS)	<a href="https://dnr.alaska.gov/welts">https://dnr.alaska.gov/welts</a>
Remote Monitoring Database	<a href="https://anthc.bmon.org/map/">https://anthc.bmon.org/map/</a>
Energy Audit Database	<a href="https://www.anthc.org/what-we-do/rural-energy/audits/">https://www.anthc.org/what-we-do/rural-energy/audits/</a>
Environmental Data Management System (EDMS)	<a href="https://dec.alaska.gov/Applications/Water/EDMS/nsite/map/">https://dec.alaska.gov/Applications/Water/EDMS/nsite/map/</a>
Enforcement and Compliance History Online (ECHO) Database	<a href="https://echo.epa.gov">https://echo.epa.gov</a>
Solid Waste Information Management System (SWIMS)	<a href="https://dec.alaska.gov/Applications/EH/SWIMS/Default.aspx">https://dec.alaska.gov/Applications/EH/SWIMS/Default.aspx</a>
Community Story Maps	<a href="https://dcra-cdo-dcced.opendata.arcgis.com/">https://dcra-cdo-dcced.opendata.arcgis.com/</a>

TABLE II  
DATA SOURCES: DATABASE NAMES AND THE CONTENTS

Database name	Database content
Drinking Water Watch Database	Public Water System Identification Number (PWSID), Water System Name, Point of Contact, Source Type, Operational Status, Federal Classification
Water-and Wastewater-Operator Database	Operator Identification, Certification Details, Assigned Systems
Water and Wastewater System Database	System Identification, System Classification, System Scoring, Assigned Operators
Community Worker and Sewer Improvement Contact List	Project Agency Lead, Remote Maintenance Worker, Rural Utility Business Advisor (RUBA), Operator Certification Contact, Drinking Water Program Staff, Solid Waste Program Contact
Well Log Tracking System Database	Well Log Information, Geophysical Data
Remote Monitoring Database	Enhanced Operational Oversight, Reduced On-Site Presence Requirements, Peace of Mind
Energy Audit Database	Comprehensive Energy Analysis, Cost-Saving and Efficiency Recommendations, Support for Sustainable Operations
Environmental Data Management System (EDMS)	Wastewater System Dashboards, Compliance and Permit Tracking, Data Accessibility and Transparency, Geographic and Site-Specific Information
Enforcement and Compliance History Online (ECHO) Database	Detailed Facility Reports, Compliance and Violation Tracking, Enforcement Actions and Penalties, Public Data Accessibility
Solid Waste Information Management System (SWIMS)	Permit and Compliance Tracking, Inspection and Operational Records, Public Data Accessibility
Community Story Maps	Community Profiles and Demographics, Infrastructure and Resource Mapping, Economic and Environmental Insights

as building temperatures, water levels, fuel levels, etc.) in real-time through an online platform. This program enhances operational oversight, reduces the need for on-site presence of technical assistance providers, and provides alerts to improve system reliability and prevent disruptions.

Energy Audits assess energy use across community facilities, identifying opportunities to reduce consumption and costs. ANTHC offers recommendations for improving energy efficiency, promoting long-term sustainability, and cost savings.

Together, these programs help communities optimize resources, lower operational costs, and adopt sustainable practices in managing essential services.

#### F. Environmental Data Management System (EDMS)

The Environmental Data Management System (EDMS), managed by Alaska’s Division of Water, part of the Department of Environmental Conservation consolidates and streamlines environmental data related to wastewater systems and other assets across the state. This online platform supports compliance, resource management, and data transparency by providing accessible information on system status, regulatory compliance, and wastewater quality assessments. Key features

include wastewater system dashboards with compliance history, permit tracking, and inspection reports, as well as the ability to search site-specific data by location or system type. EDMS serves as a vital resource for government agencies, utilities, environmental professionals, and the public, enhancing data transparency and supporting proactive environmental management across Alaska.

#### G. Enforcement and Compliance History Online (ECHO) Database

The Enforcement and Compliance History Online (ECHO) database, managed by the U.S. Environmental Protection Agency (EPA), provides public access to compliance and enforcement data for regulated facilities nationwide. ECHO tracks compliance histories, enforcement actions, and penalties, promoting transparency and accountability in environmental protection. It covers various facilities, such as wastewater plants, public water systems, and other regulatory facilities, offering detailed reports on permit status, inspections, violations, and enforcement measures. The database also tracks compliance with key laws like the Clean Water and Clean Air Acts, documenting enforcement actions and corrective steps taken. Designed for public access, ECHO allows users to

monitor facility compliance and supports the EPA’s efforts to ensure environmental accountability.

#### H. Solid Waste Information Management System (SWIMS)

The Solid Waste Information Management System (SWIMS), managed by the Solid Waste Program within Alaska’s Division of Environmental Health, which is a part of the Department of Environmental Conservation, is an online database that centralizes data on solid waste facilities statewide. It provides essential information for regulatory compliance, environmental monitoring, and public awareness regarding landfills. Key features include detailed facility profiles with classification, location, and operational details, as well as permit and compliance tracking. SWIMS also offers inspection records and inspection reports, promoting transparency in waste management. Publicly accessible, SWIMS ensures community members, researchers, and stakeholders can easily access up-to-date information on solid waste facility compliance and management practices, supporting Alaska’s environmental protection goals.

#### I. Community Story Maps

The Community Story Maps platform, managed by Alaska’s Division of Community and Regional Affairs (DCRA) within the Department of Commerce, Community, and Economic Development, offers an interactive resource for exploring the state’s communities. It provides geographic and demographic insights, focusing on social, economic, and environmental data. Key features include detailed community profiles with demographic information, infrastructure mapping (e.g., energy and sanitation facilities), and economic and environmental data (e.g., income level and climate data). The platform supports informed decision-making, regional planning, and sustainable development. Community Story Maps are a valuable tool for residents, researchers, and policymakers.

### IV. OUR SOFTWARE SOLUTION

#### A. Software Architecture

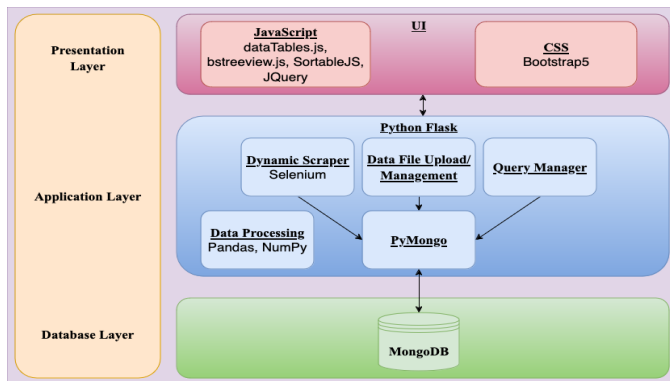


Fig. 1. Architecture Diagram for the Webtool

By utilizing a custom javascript frontend with Bootstrap 5 styling, users can create a unique list of scraper instructions that are designed for each run of the scraper. The server

will format this input into a list of tuples before calling the Selenium scraper. The scraper will then attempt to scrape the provided webpage following the list of instructions created by the user. If this scrape is successful, the results of the scrape will be provided to the user on a redirected webpage. From this page, the user may download the data as a JSON file or upload the file directly to a specified MongoDB database. The user will have access to the debug log to assist with updating the scraper should an error be encountered. A similar debug log will be provided to the user if the scraper fails.

The Flask server serves as the intermediary between the MongoDB backend and the frontend UI, as shown in Fig. 1. This portion of the tool handles converting user input into the different formats required by the Selenium scraper and the query manager. Each instruction command on the frontend corresponds to a function that will process the user data, convert it into the format required, and, in the case of the scraper, perform the action specified by the command. As the data collected by this tool will be updated regularly, it is vital that the various scrapes, queries, and file uploads needed to keep the data archive up-to-date and organized are easy to rerun as needed. For this, the server will also enable the saving of scraper instruction lists, queries, and specific CSV/Excel files for ease of repeating these tasks. This will allow users to set up a scrape or query once and be able to run it in the future with just a few button clicks. The scraped information will then be displayed to the user for manual verification of the data and may be downloaded as a JSON file. The user can launch a localized server for utilization via a Dockerfile.

As the results of the scraper may include records that have more or less information than other records as well as nested data, the usage of MongoDB allows for the storage of these records in a collection immediately after a scrape has been completed. By utilizing the PyMongo library and the three tools outlined above, we are able to insert the scraped data and submit user queries to the MongoDB database if the user has provided the correct connection URI, database name, and collection name. These newly inserted records can then be sorted as needed.

#### B. The Data workflow

1) *Collecting data through web scraping:* In an effort to streamline access to this information and reduce the time needed to keep the sanitation facility data up-to-date, we began work to develop a tool that will enable a user the ability to construct web scraping tasks dynamically based on the information needed at the time of data collection. Typical web scraping scripts are tailored to gather select information from a web page with a known structure that enables the creation of hard coded scrapers, easily tested and updated should the web pages undergo changes. Our tool allows users to create scraping instruction trees for each session, enabling them to build a scraper without needing to develop a separate script for each task with little to no prior experience. To build a scraper instruction tree, a user will use a browser’s debug tool to locate elements and to provide input to each

command to navigate and scrape web pages. As the tool saves these instruction lists, users can easily update them should the webpage change. Tooltips and documentation are provided to explain each command and include usage examples.

2) *Archiving unstructured data in NoSQL database:* As much of the information pulled from these websites will be unstructured, or will be structured differently between websites, using a NoSQL database will allow us to immediately store the collected information and provide a platform to aggregate this data into useful groupings for future data exporting. MongoDB in particular, with its capability of having nested data storage within each document, will allow us to store the scraped data in the same format it was collected in. This, in turn, will allow for an easier time when building this tool, as the scraping program does not also need to enforce restrictions on how or what information can be pulled. It is crucial to design a database structure that will enable a user to quickly set up new collections based on the contents scraped and relate the necessary collections back to the origin point of the data.

3) *Querying the data through intuitive web interface:* As this project developed, new features were required to enable our non-coding users the ability to interact with and modify the existing database in order to format the data into the correct output for each given table on the Arctic Outlook website. Each table on this website is formatted via a CSV file and allows up to one nested level for referencing or building subtables. As our initial data dump may have many nested levels of information from various subpages, our user will need to be able to manipulate the data within the data dump collection to generate different tables for the website. As our user is unfamiliar with MongoDB’s query language, it was decided that we would develop a simple graphical user interface (GUI) to assist with generating queries to restructure the scraped data. The GUI focuses on providing access to the commands utilized in formatting the data for the website. The submitted user input is then converted into a MongoDB query.

4) *The Resulting Website:* A screenshot of the resulting website can be found in Fig. 2. Almost 1,300 people across Alaska are using this website as a one-stop solution for accessing any water- and wastewater-related information. By utilizing the webtool, we will be able to collect, aggregate, and integrate each of the data sources into this website.

## V. COMMUNITY IMPACTS

We unified disparate datasets from over 2,000 water operators, 1,500 drinking water systems, 180 wastewater profiles, and 900 solid waste profiles, 200 communities, into a web-based, user-friendly hydroinformatics platform. This platform is a key step toward overcoming data fragmentation and improving access in Alaska’s sanitation sectors. The Community impacts of our work include:

1) *Enhanced Accessibility and Usability* Communities in remote areas now have streamlined access to essential sanitation data, enabling them to make informed decisions about resource allocation and infrastructure management. The user-

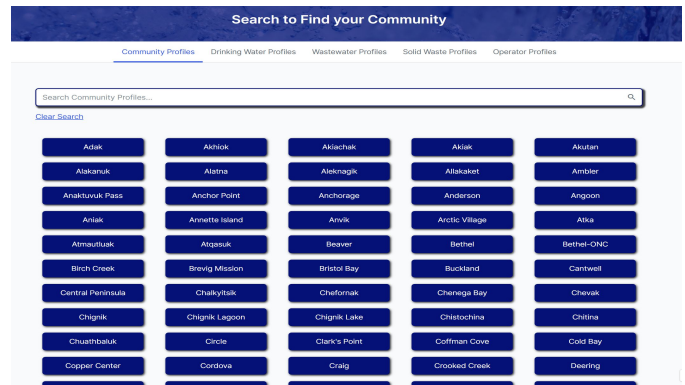


Fig. 2. Screenshot of our website that integrates all the disparate datasets

friendly interface accommodates technical and non-technical users alike, ensuring broad adoption and utilization.

2) *Empowering Underserved Communities* The platform directly supports rural and tribal communities, which often face significant barriers in accessing data and resources. By integrating information from eleven data sources, it bridges the knowledge gap, promoting inclusivity and equity in resource distribution across these underserved areas

3) *Operational Efficiency* Automated data collection and standardized formatting save considerable time and effort for utilities, agencies, and technical assistance providers, allowing them to focus on service improvement rather than data retrieval tasks. By reducing manual processes, the platform minimizes errors, enhances data reliability, and supports utility compliance with both state and federal regulations.

4) *Collaboration and Stakeholder Engagement* Arctic Outlook has engaged a beta group consisting of tribes, tribal health organizations, technical assistance providers, and state agencies, creating a feedback loop to refine the platform. Initial feedback underscores the platform’s value in uncovering previously unknown datasets and enhancing coordination among stakeholders.

5) *Promoting Public Health and Environmental Health* Consolidated data on drinking water, wastewater, and solid waste systems empowers communities to effectively tackle public health and environmental health challenges.

6) *Capacity Building and Education* Our website fosters a culture of data-driven decision-making and long-term sustainability. It helps stakeholders develop technical skills and build confidence in managing data, ultimately enhancing the resilience of sanitation systems in Alaska.

7) *Future Scalability* The platform’s modular architecture enables the integration of additional data sources and expansion into new areas of sanitation infrastructure management. As the platform evolves, it has the potential to serve as a model for other regions facing similar challenges, broadening its impact beyond Alaska.

## VI. CONCLUSION AND FUTURE WORK

In this work, we outlined the development of a webtool capable of aggregating and managing sanitation system data

within the state of Alaska. This tool enables users to scrape, upload, and manage data with the goal of compiling this information in an easy to access website for stakeholders throughout the state. In the future, we hope to continue our work on this tool by expanding some of the core features to improve usability based on feedback from users.

#### ACKNOWLEDGMENT

This research was, in part, funded by the following grant: 1) USGS NIWR G00014344, 2) UAF Arctic Fellow Grant (UAF Center of ICE/Office of Naval Research), 3) AIM AHEAD PAIR (NIH Agreement No. 1OT2OD032581), 4) NSF EP-SCoR RII Track4 #2327456, and 5) NSF MRI #2320196 6) DOE RENEW #DE-SC0025715. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the funding agencies.

#### REFERENCES

- [1] Y. Chen and D. Han, "Big data and hydroinformatics," *Journal of Hydroinformatics*, vol. 18, no. 4, pp. 599–614, Mar. 2016, ISSN: 1464-7141. DOI: 10.2166/hydro.2016.180. eprint: <https://iwaponline.com/jh/article-pdf/18/4/599/1107170/jh0180599.pdf>.
- [2] Z. Vojinovic and M. B. Abbott, "Twenty-five years of hydroinformatics," *Water*, vol. 9, no. 1, 2017, ISSN: 2073-4441. DOI: 10.3390/w9010059.
- [3] C. Makropoulos and D. A. Savić, "Urban hydroinformatics: Past, present and future," *Water*, vol. 11, no. 10, 2019, ISSN: 2073-4441. DOI: 10.3390/w11101959.
- [4] K. Gibert, J. S. Horsburgh, I. N. Athanasiadis, and G. Holmes, "Environmental data science," *Environmental Modelling & Software*, vol. 106, pp. 4–12, 2018, Special Issue on Environmental Data Science. Applications to Air quality and Water cycle, ISSN: 1364-8152. DOI: <https://doi.org/10.1016/j.envsoft.2018.04.005>.
- [5] D. Blodgett, E. Read, J. Lucido, T. Slawecki, and D. Young, "An analysis of water data systems to inform the open water data initiative," *JAWRA Journal of the American Water Resources Association*, vol. 52, no. 4, pp. 845–858, 2016. DOI: <https://doi.org/10.1111/1752-1688.12417>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.12417>.
- [6] J. S. Horsburgh, D. G. Tarboton, D. R. Maidment, and I. Zaslavsky, "A relational model for environmental and water resources data," *Water Resources Research*, vol. 44, no. 5, 2008. DOI: <https://doi.org/10.1029/2007WR006392>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR006392>.
- [7] L. Shanguan and Z. Sun, "Enhancing data quality and real-time sharing performance in water informatics through decision tree mining algorithm," *Measurement: Sensors*, vol. 31, p. 101024, 2024, ISSN: 2665-9174. DOI: <https://doi.org/10.1016/j.measen.2023.101024>.
- [8] W. Deng and G. Wang, "A novel water quality data analysis framework based on time-series data mining," *Journal of Environmental Management*, vol. 196, pp. 365–375, 2017, ISSN: 0301-4797. DOI: <https://doi.org/10.1016/j.jenvman.2017.03.024>.
- [9] H. Liu, Z. Chen, Z. Wang, *et al.*, "Water quality system informatics: An emerging inter-discipline of environmental engineering," *Engineering*, vol. 43, pp. 115–124, 2024, ISSN: 2095-8099. DOI: <https://doi.org/10.1016/j.eng.2024.03.018>.
- [10] X. Li, F. Neng, J. Zhao, and N. Zhang, "Data analysis of water supply demand and study on scale of water transfer project in xiong'an new area," in *2022 8th International Conference on Hydraulic and Civil Engineering: Deep Space Intelligent Development and Utilization Forum (ICHCE)*, 2022, pp. 1–6. DOI: 10.1109/ICHCE57331.2022.10042504.
- [11] C. Lotfi, S. Srinivasan, M. Ertz, and I. Latrous, "Web scraping techniques and applications: A literature review," in Jan. 2021, pp. 381–394, ISBN: 9789391842086. DOI: 10.52458/978-93-91842-08-6-38.
- [12] S. Upadhyay, V. Pant, S. Bhasin, and M. K. Pattanshetti, "Articulating the construction of a web scraper for massive data extraction," in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017, pp. 1–4. DOI: 10.1109/ICECCT.2017.8117827.
- [13] L. Saal and A. S. Ruhl, "Automated scraping and analyses of drinking water quality data," *International Journal of Hygiene and Environmental Health*, vol. 255, p. 114295, 2024, ISSN: 1438-4639. DOI: <https://doi.org/10.1016/j.ijheh.2023.114295>.
- [14] R. P. N. Budiarti, N. Widyatmoko, M. Hariadi, and M. H. Purnomo, "Web scraping for automated water quality monitoring system: A case study of pdam surabaya," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2016, pp. 641–648. DOI: 10.1109/ISITIA.2016.7828735.
- [15] B. Jose and S. Abraham, "Exploring the merits of nosql: A study based on mongodb," in *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, 2017, pp. 266–271. DOI: 10.1109/NETACT.2017.8076778.
- [16] V. N. Gudivada, D. Rao, and V. V. Raghavan, "Nosql systems for big data management," in *2014 IEEE World Congress on Services*, 2014, pp. 190–197. DOI: 10.1109/SERVICES.2014.42.
- [17] A. A. Frozza, R. d. S. Mello, and F. d. S. d. Costa, "An approach for schema extraction of json and extended json document collections," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 2018, pp. 356–363. DOI: 10.1109/IRI.2018.00060.