DOI: 10.37943/15JUNJ2506

Meiran Zhiyenbayev
Lecturer, Department of Artificial intelligence and Big Data
meiran1991@gmail.com, orcid.org/0009-0005-0497-0965
Al-Farabi Kazakh National University, Kazakhstan

Assel Ospan
Senior Lecturer, Department of Artificial intelligence and Big Data
ospan.assel@kaznu.kz, orcid.org/0000-0002-1860-6997
Al-Farabi Kazakh National University, Kazakhstan

Nadezhda Kunicina
PhD, Professor, Faculty of Electrical Engineering and Ecology
Nadezda.Kunicina@rtu.lv, orcid.org/24438414500
Riga Technical University, Riga, Latvia

Madina Mansurova
Candidate of Physical and Mathematical Sciences
Head of Department of Artificial intelligence and Big Data
mansurova.madina@gmail.com, orcid.org/0000-0002-9680-2758
Al-Farabi Kazakh National University, Kazakhstan

Roman Titkov
Programmer, Department of Artificial intelligence and Big Data
roman.titkov@kaznu.edu.kz, orcid.org/0009-0003-8579-3600
Al-Farabi Kazakh National University, Kazakhstan

SYSTEMATIC DATA PROCUREMENT IN AN OWL-EMBEDDED INFORMATION AND ANALYTICAL FRAMEWORK FOR THE MONITORING OF WATER RESOURCES IN THE ILE-BALKHASH BASIN

Abstract: The world is facing an escalating water shortage crisis, with dire consequences for ecosystems, human health, and socio-economic development. This article explores the multifaceted nature of the water shortage problem of Ile-Balkhash basin that falls under the jurisdiction of the Balkhash-Alakol Republic of Kazakhstan, its underlying causes, and the complex web of challenges it presents. The predicament in the Ile-Balkhash basin is a complex interplay of various factors. Climate change has led to erratic precipitation patterns, exacerbating the problem. The simultaneous rise in population places additional stress on the already limited water resources. Moreover, inefficient water management practices have perpetuated the issue, hindering the equitable distribution of water. The challenge of conducting a comprehensive basin analysis is a formidable task due to the numerous variables and indicators involved. It demands an enormous amount of time and effort. To address this issue, a web application framework integrated with the Web Ontology Language database, allowing for the execution of advanced queries to extract valuable insights from various objects and indicators, has been developed. The database underpinning this system is meticulously compiled, drawing upon data from the Hydrological monitoring of water bodies and the National

Received: 21.08.2023 Accepted: 23.09.2023 Published: 30.09.2023

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

Hydrometeorological Service of the Republic of Kazakhstan. These sources provide critical data that forms the bedrock of the analysis. Recognizing the importance of data storage and management in this endeavor, integrated components have been established. These components play a pivotal role in structuring the diverse data sources and maintaining their currency. The water shortage issue in the Ile-Balkhash basin serves as a stark reminder of the urgency with which such crises must be addressed. The tools and methods offer hope and underscore global water management sustainability.

Keywords: monitoring of water resources; owl ontology; Ile-Balkhash basin; geo-data management; water shortage crisis; data package

Introduction

The Ili-Balkhash basin is one of the most significant lake ecosystems on a global scale and covers a significant area of 413,000 km² - an amazing natural complex. It is noticeable that 85% (353,000 km²) belongs to Kazakhstan, while the remaining 15% is in China. Within Kazakh territory are the regions of Almaty, Zhambyl and Karaganda, including the vibrant metropolis of Almaty, which has a population of 2.5 million.

Five significant rivers flow into Lake Balkhash: Ili (1439 km), Karatau (390 km), Ayaguz (492 km), Lepsy (418 km) and Aksu (316 km) [1].

Potential hydropower reservoirs in the basins of Lake Balkhash amount to 63.5 billion kWh, which is almost 40% of the total volume of Kazakhstan. Specifically, the Ili River brings 7008 million kWh, which is 18.2% of the total potential of all rivers in the Ili-Balkhash basin, equal to 35.5 billion kWh. At the moment, irrigated agriculture is consuming 70% of water resources. Rising demand for water, especially in China, is raising concerns about the lake's safety and environmental situation. This problem is exacerbated by climate change and the degradation of glaciers in the basin. A critical calculation from the two-year EU Balkhash project shows that a minimum annual inflow of 12 km³ from China needs to be maintained, a value that has now been reduced to 8 km³ [2].

Identifying patterns and developing strategies to mitigate human impacts and climate change is becoming important. The principles and approaches of water resources management play a leading role in the formation of a science-based strategy for the protection and management of water resources. The study used data from 63 monitoring stations with various sensors covering the period from January 1, 2001 to the present, which is depicted in Fig. 1. These sensors recorded a variety of parameters including water level, flow, temperature, flooding, and surface water quality [3,4].

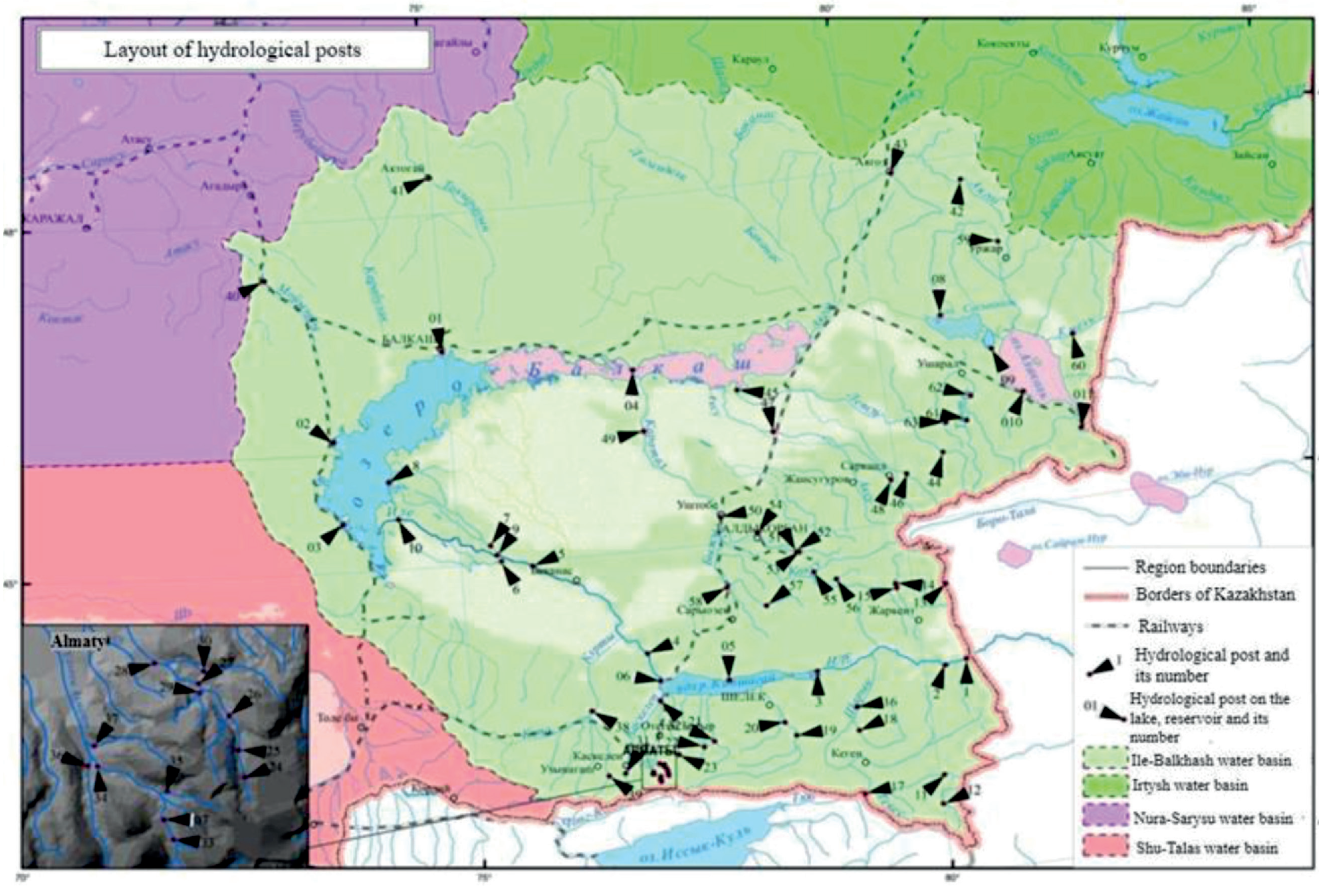

Figure 1. The layout of the hydrological posts of the Ili-Balkhash basin

Literature review and problem statement

The Ili-Balkhash Basin is a vital region not only in Kazakhstan, but also plays a crucial role in Central Asia, and the sustainable management of its water resources is essential for both environmental conservation and human livelihoods in the area. The Ili-Balkhash Basin encounters multiple environmental issues, such as a shortage of water due to excessive utilization and the effects of climate change which is the main issue that has a direct effect on water level in the Ili-Balkhash Basin, particularly in Lake Balkhash [5]. Additionally, there are worries about water quality due to contamination stemming from industrial and agricultural practices [6]. These problems have consequences for the general well-being of Lake Balkhash and the ecosystems in its vicinity [7]. Since there are multiple heterogeneous data from different sources, it challenges researchers and analysts to effectively study this basin. As a solution, a monitoring web framework application with extracting objects from owl-based databases with advanced querying is provided. In the following sections, the approach based on Web Ontology Language with organized heterogeneous data in a way that allows to run advanced queries to get more insights and to do a proper analysis will be described.

OWL-Embedded Information refers to the use of the Web Ontology Language (OWL) within information systems, databases, or applications to represent and manage knowledge in a structured, machine-readable format. Using OWL-Embedded Information offers numerous benefits in various domains due to its capacity for precise knowledge representation and semantic interoperability. There are several compelling reasons for using OWL-Embedded Information, along with numerous benefits associated with its implementation. Precise Knowledge Representation - OWL allows you to represent knowledge with a high degree of precision and

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

expressiveness. It enables you to define complex concepts, relationships, and constraints in a formal manner, making it suitable for representing diverse and intricate information [8]. Semantic Interoperability: by using OWL-Embedded Information systems can achieve semantic interoperability. This means that different systems can understand and exchange data in a way that ensures a common and accurate interpretation of the information, even when dealing with diverse data sources or applications [9]. OWL's formal semantics and logical foundation enable automated reasoning. Systems that embed OWL can perform logical inference and deduction, allowing them to derive new knowledge from existing data, detect inconsistencies, and make intelligent decisions based on this reasoning [10]. When working with data from various sources, OWL can serve as a common ontology that unifies data and enables seamless integration. This reduces the complexity of data integration tasks and ensures consistency in data representation [11]. Systems incorporating OWL-Embedded Information can support semantic search capabilities. Users can search for information based on the meaning and relationships within the data, leading to more relevant and precise search results compared to keyword-based searches [12]. OWL-embedded systems can be designed to evolve over time as knowledge grows or changes. This adaptability is particularly useful in applications like recommendation systems, where the ability to learn and adapt is crucial [13]. Knowledge Sharing and Collaboration: OWL-Embedded Information facilitates knowledge sharing and collaboration among individuals and organizations. It provides a standardized way to represent and exchange data, making it easier for multiple parties to work together effectively [14].

Recent studies have highlighted the urgency of addressing these issues. For instance, de Boer et al. [15] conducted a comprehensive analysis of the impact of climate change on the Ili-Balkhash Basin's water resources, providing critical insights into the challenges we face. In a recent study by Nurtazin et al. [16], the authors investigated the potential use of advanced monitoring technologies to improve water resource management in the Ili-Balkhash Basin, offering innovative solutions for sustainable water utilization. Additionally, Nurtazin et al. [17] explored the extent of water contamination in the basin, shedding light on the sources and implications of this issue. Furthermore, a study by Hein et al. [18] discussed the long-term ecological impacts of water resource management in the region, emphasizing the need for comprehensive environmental conservation measures. Finally, JWeili et al. [19] provided insights into the role of international cooperation in addressing water resource challenges in the Ili-Balkhash Basin, highlighting the importance of collaborative efforts.

Given the heterogeneous data from various sources, it is crucial for researchers and analysts to effectively study this basin. To address this, the development of a monitoring web framework application that extracts data from OWL-based databases and supports advanced querying is proposed. In the following sections, the approach which is based on the Web Ontology Language will be described. By organizing all heterogeneous data in this manner, it is possible to run advanced queries to gain deeper insights and conduct a rigorous analysis.

In summary, using OWL-Embedded Information offers a structured and powerful approach to managing and leveraging knowledge within information systems. It enhances interoperability, reasoning capabilities, data integration, and search accuracy, making it a valuable tool in diverse applications across various industries.

Architecture

The OWL-Embedded Information and Analytical Framework is the platform where each step is automated from extracting source data to building ontology and extracting data from ontology using web interface. First part of this application takes you through the entire ETL process which involves extracting data from one or multiple sources in different formats such as pdf, csv, xlsx and so, reshaping it to align with the desired target structure that will be used to importing it into a designated data warehouse, as shown in Fig. 2, and developing ontology for the Ili-Balkhash basin. When our data is clean and structured in commonly used data format, it becomes possible to use that data to create and develop our ontology.

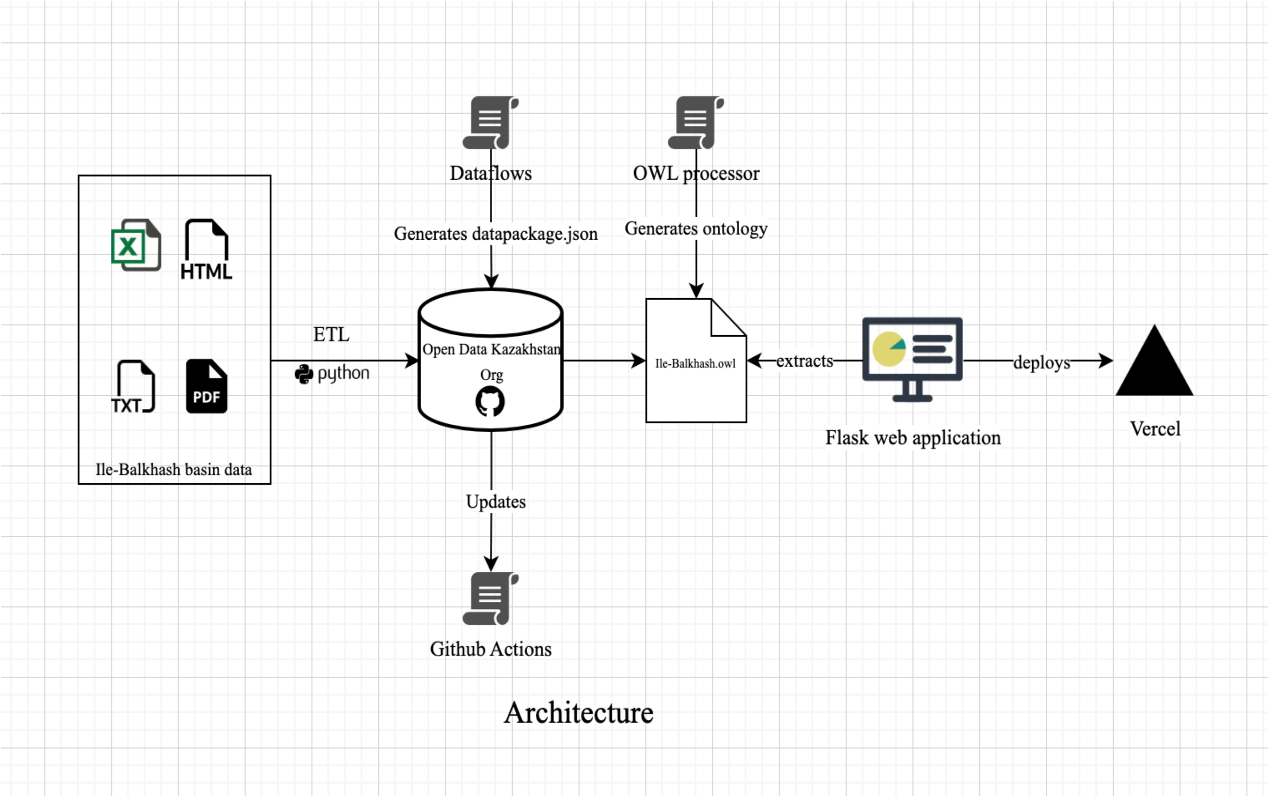

Figure 2. Architecture for OWL-Embedded Information and Analytical Framework

The ETL procedure for producing organized datasets is constructed utilizing the Python programming language, incorporating automation through certain open-source technologies. Within the data cleansing phase, the primary responsibility is assumed by the dataflows library, which encompasses essential functions for tidying, parsing, and reorganizing data [20]. This library helps us to build pipelines for extracting, cleaning, and loading our datasets. This library consists of all necessary classes and functions for the data preparation step. Another advantage is that it has a pre-build data package class that we describe in the following sections.

When examining the documents, it becomes evident that they lack standardization and structure. They predominantly exist in Excel format, featuring varying structures and data formats that are unsuitable for constructing dashboards and conducting analytical work. Therefore, our objective is to establish an automated data pipeline comprising collection, cleansing, formatting, and packaging stages. These processes will be scheduled and executed based on the publication dates of the source materials.

In the construction of our dataset, lightweight yet comprehensive data standards known as the Data Package are employed. This serves as a fundamental container format for describing a collection of data within a single package. It offers a foundation for the convenient delivery, installation, and management of datasets. A Data Package has the flexibility to encompass various types of data. Simultaneously, Data Packages can be tailored and enriched to cater to specific data types; for instance, Tabular Data Packages are designed for tabular data, while Geo Data Packages are optimized for geographical data, among others. Fig. 3 illustrates the

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

metadata JSON file, which contains essential information about the data, such as authorship, licensing, publication date, row count, data profile, hash, and metadata pertaining to each resource.

This metadata encompasses details like format, encoding, name, title, and schema, which describes each data field, as exemplified in Fig. 4. It is worth noting that the Data Package framework is also endorsed by the Open Knowledge Foundation, and its specifications are regularly updated and made accessible through various channels and posts [21]. Once our data is ready to use, it's possible to start creating our owl ontology using python processors.

```
"bytes": 3794078,
"count_of_rows": 122100,
"hash": "49c849a3bc4ff428a7aa4e39cab9f6f8",
"name": "ili-balkhash-basin-ontology-data",
"profile": "data-package",
"resources": [
  {
    "bytes": 1761990,
    "dialect": {
      "caseSensitiveHeader": false,
      "delimiter": ",",
      "doubleQuote": true,
      "header": true,
      "lineTerminator": "\r\n",
      "quoteChar": "\"",
      "skipInitialSpace": false
    },
    "encoding": "utf-8",
    "format": "csv",
    "hash": "5a96136da0534a2e10cb3b72a8b668a4",
    "name": "water-consumption",
    "path": "data/water-consumption.csv",
    "profile": "tabular-data-resource",
    "schema": {
      "fields": [
        {
          "format": "default",
          "name": "device_code",
          "type": "integer"
        },
        {
          "format": "default",
          "name": "date",
          "type": "string"
```

Figure 3. The sample generated metadata about the package
(github.com/open-data-kazakhstan/ili-balkhash-basin-ontology-data [22])

In the second part, the web application framework using Flask, which extracts data from the generated owl ontology and visualizes our data is built. Owl embedded query section helps to run advanced queries directly from the platform and generates table views for the result. In addition, we added some graph views to make our analysis more readable and clear. Detailed description of each functionality will be described in the following sections.

```
    "bytes": 1761990,
    "dialect": {
      "caseSensitiveHeader": false,
      "delimiter": ",",
      "doubleQuote": true,
      "header": true,
      "lineTerminator": "\r\n",
      "quoteChar": "\"",
      "skipInitialSpace": false
    },
    "encoding": "utf-8",
    "format": "csv",
    "hash": "5a96136da0534a2e10cb3b72a8b668a4",
    "name": "water-consumption",
    "path": "data/water-consumption.csv",
    "profile": "tabular-data-resource",
    "schema": {
      "fields": [
        {
          "format": "default",
          "name": "device_code",
          "type": "integer"
        },
        {
          "format": "default",
          "name": "date",
          "type": "string"
        },
        {
          "decimalChar": ".",
          "format": "default",
          "groupChar": "",
          "name": "value",
          "type": "number"
```


Figure 4. This sample generated metadata about each data field
(github.com/open-data-kazakhstan/ili-balkhash-basin-ontology-data[22])

Ontology-based approach

There are various methods for outlining our logical model and constructing the ETL process, including graph-based, UML-based, BPMN-based, and ontology-based approaches. Upon closer examination of their conceptual, logical, and physical models, it becomes apparent that they differ in terms of data formats and modes, which may not align with our desired outcome [23]. In our particular case, the ontology-based approach stands out as the most suitable option. It offers the advantage of working not only with structured data but also with semi-structured data. The main challenge lies in ensuring that an ETL developer invests a bit of effort into the careful design of an enriched ontology database. Therefore, we have opted for a semi-automated approach to strike the right balance.

Once the approach is decided, the next step is to minimize the total cost which consists of time and resources, while maximizing the quality of data from several resources by the following formula:

$$Z = \sum_{i \in I} \sum_{j \in J} (X_{ij} \cdot Cost_{ij}) \quad (1)$$

where

Z is the total cost, which is in minutes, and it will be minimized by optimizing scripts and tasks,

i – an identifier for each data source. In proposed ETL process, there are 4 sources, i from 1 to 4, such as:

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

- IoT Sensor – Water Consumption (i=1)
- IoT Sensor – Water Level (i=2)
- National Hydrometeorological Service (i=3)
- Bureau of National statistics (i=4)
- j – an identifier for each data processing task (extraction, transformation, loading):
- Extraction (j=1)
- Transformation (j=2)
- Loading (j=3)

Let X_{ij} be a binary variable with i and j.

$Cost_{ij}$ – represents the cost (time, resources) of performing task j on data source i.

$X_{ij}=1$ for all i since we load all sources.

There are several constraints which we need to carefully analyze such as how each data source should be processed, task dependency, data quality, resource size, and IoT data availability. Cost (time in minutes) of using each data source for each task described in Table 1:

Table 1. Cost (time in minutes) for each source tasks

Data Source	Extraction (j=1)	Transformation (j=2)	Loading (j=3)
IoT Sensor – Water Consumption	5	10	3
IoT Sensor – Water Level	5	10	3
National Hydrometeorological Service	3	10	3
Bureau of National statistics	3	5	3

By calculating total cost Z using formula (1) would be 63 minutes. This represents the time required for the ETL process according to the selected data source-task assignments, with costs in minutes as requested. It will be minimized by optimizing sql scripts and Python scripts by changing order of load.

Another important constraint is data quality that plays a crucial role in developing ETL systems. In order to calculate data quality coefficients for each section in Table 3, the average for all the sources that will be the constraints while optimizing the formula can be found. In the proposed system, sources are IoT and websites, so there will be some downtime issues (Table 4).

Table 2. Data Quality Measures

Data Source	Data Quality (0 to 1)
IoT Sensor – Water Consumption	0.95
IoT Sensor – Water Level	0.95
National Hydrometeorological Service	0.99
Bureau of National statistics	0.99

Table 3. Resource Usage (CPU %) of using each data source for each task

Data Source	Extraction (j=1)	Transformation (j=2)	Loading (j=3)
IoT Sensor – Water Consumption	7	15	10
IoT Sensor – Water Level	8	15	10
National Hydrometeorological Service	5	10	5
Bureau of National statistics	5	10	5

Table 4. Resource Usage (CPU %) of using each data source for each task

Data Source	Extraction (j=1)	Transformation (j=2)	Loading (j=3)
IoT Sensor – Water Consumption	7	15	10
IoT Sensor – Water Level	8	15	10
National Hydrometeorological Service	5	10	5
Bureau of National statistics	5	10	5

Maximum Resource Capacity ⩽ 35% CPU.

By carefully considering these constraints, there are seen the best ways to make the ETL procedures more efficient, even using multiple tasks if needed. These constraints also help to control costs when new data sources are added, it becomes possible to adjust the tasks accordingly. For our web-based ontology and web application, these constraints ensure smooth operation with our existing data sources. However, as data quality changes, IoT data availability fluctuates, and resource usage evolves, new challenges and constraints might be encountered.

Data storage and management

The role of data storage and administration is pivotal within any data management framework. In our case, we utilize the GitHub organization called «Open Data Kazakhstan» [24]. This organization was initially established as an open-source data center alternative for researchers and analysts. It's important to note that all the data housed within this organization is meticulously maintained, ensuring cleanliness and a well-structured format. The organization serves as a focal point for data engineering professionals seeking to actively participate in the augmentation and upkeep of datasets, with the primary objective of expanding the data-driven community within our nation. Within the research paper, datasets via this organization with the overarching goal of making all pertinent information accessible to the general public are disseminated. This approach is driven by our commitment to facilitating the utilization of our data by fellow researchers and analysts, thus fostering their capacity to contribute to and address complex challenges within the substantial domain of water resource management in the Republic of Kazakhstan. The prevailing water scarcity issue in our region accentuates the urgency and importance of this endeavor.

To ensure data currency, we've implemented GitHub Actions [25], which execute at regular intervals based on data-specific parameters outlined in the configuration file, as depicted in Fig. 5. GitHub Actions stands as a robust utility enabling developers to automate their operational sequences and enhance the efficiency of their development procedures. Employing GitHub Actions for data updates offers several benefits, including the automation of data refreshes, robust version control, support for collaborative efforts, seamless integration with various tools, and active engagement within the open-source community.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

```
name: Ile balkhash basin ontology data pipeline
on:
  push:
    branches:
      - master
  schedule:
    - cron: "0 0 1 * *"

jobs:
  build:
    runs-on: ubuntu-22.04
    steps:
    - uses: actions/checkout@master
    - name: Build the data and create local changes
      uses: actions/setup-python@v1
      with:
        python-version: '3.11.2'
        architecture: x64
    - run: |
        pip install -r requirements.txt
        python scripts/process.py
    - name: Commit files
      run: |
        git config --local user.email "action@github.com"
        git config --local user.name "GitHub Action"
        git commit --allow-empty -m "Auto-update of the data package" -a
    - name: Push changes
      uses: ad-m/github-push-action@master
      with:
        github_token: ${{ secrets.gh }}
```

Figure 5. The generated actions.yml for Github actions (workflows/actions.yml [19])

Web Ontology preparation

Creating an OWL ontology from Excel and CSV data using Python with the Owlready2 library is straightforward. Owlready2 is a Python library for ontology-oriented programming and allows you to work with OWL ontologies in a Pythonic way. There are several steps to create owl ontology from our Ili-Balkhash basin datasets:

- Prepare python environment with installed Owlready2 library
- Organize your data in Excel or CSV format and ensure that it's well-structured. Create clear columns for subjects, predicates (properties), and objects (values).
- Convert Data to RDF and Create Ontology
- Define and refine your ontology structure by adding more classes, properties, and individuals as needed.
- Validate your ontology and infer new facts within your ontology.
- Save it in the OWL format.

Data visualization

Our web application has been developed utilizing the Flask framework, leveraging the Python programming language. The primary objective of this web-based portal is to facilitate direct querying of our OWL-based ontology database, enabling efficient and rapid monitoring and analysis of objects within the Ili-Balkhash basin. This portal is structured into three distinct sections:

The first section, as depicted in Fig. 6, is dedicated to query execution. It offers the advantage of enabling users to formulate and execute queries.

Fig. 7 illustrates the second section, designed for visual data representation. This section provides data in tabular form, complete with columns and filtering capabilities, enhancing data comprehension and exploration.

Additionally, Fig. 8 showcases the visual graphics section, which offers dynamic visualizations to provide deeper insights into the data, aiding in the interpretation of complex trends and patterns.

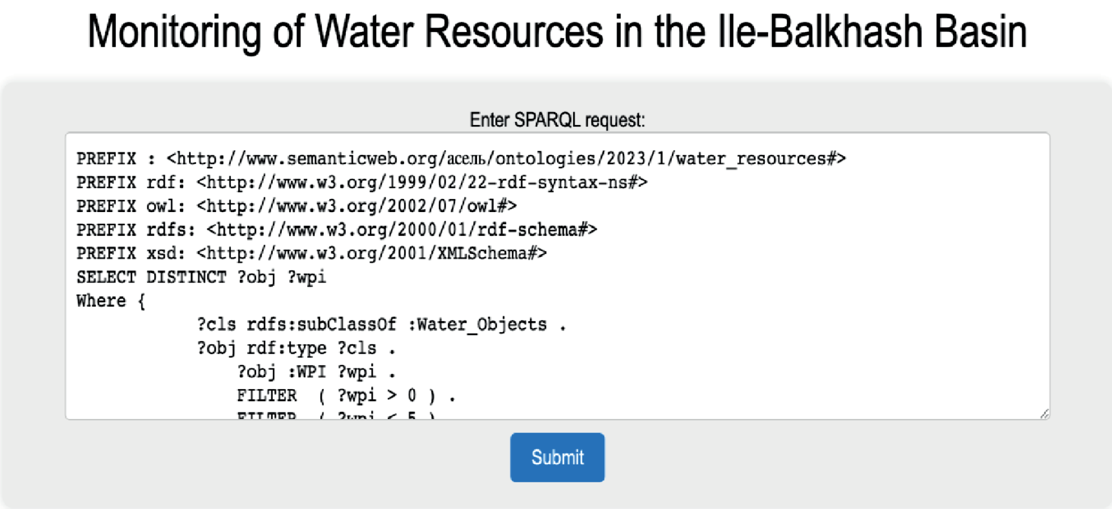

Figure 6. The 1st section, sparql embedded code submission: Network visualization of all keywords (github.com/open-data-kazakhstan/owl-embebded-web-app[26])

The study adopts an all-encompassing methodology to tackle the intricacies of the Ile-Balkhash basin, leveraging OWL-Integrated Information and Analytical Systems. A standout feature is the generation of OWL ontologies from structured data in Excel and CSV formats, facilitated by Python's Owlready2 library. This library simplifies ontology-focused programming in a Python-friendly manner. The procedure for ontology creation encompasses setting up the Python environment, structuring data in Excel or CSV, transforming this data into RDF, elaborating and fine-tuning the ontology, and ultimately storing it in OWL format.

The web-based interface for this research is built on the Flask framework and Python. It functions as an online portal for immediate querying of the OWL-centric ontology database. The interface is segmented into three functional areas: query formulation and execution, visual representation of data, and graphical visualizations. Each segment fulfills a distinct role, from enabling query formulation and execution to offering intricate visualizations for a deeper understanding of the data.

Data governance and administration are meticulously orchestrated via the «Open Data Kazakhstan» GitHub organization, and GitHub Actions are deployed for data currency assurance. This multi-faceted approach not only optimizes the data workflow but also boosts developmental efficiency and encourages teamwork.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 15, SEPTEMBER 2023

All	All
Ulba	4.76
Balkhash	4.52
Uba	3.45
Bukhtarma	3.22
Ilek	3.2
Khorgos	3.08
Shu	2.84
Ile	2.25
Ili	2.25

Figure 7. The 2nd section: Table representation of data (github.com/open-data-kazakhstan/owl-embebded-web-app[26])

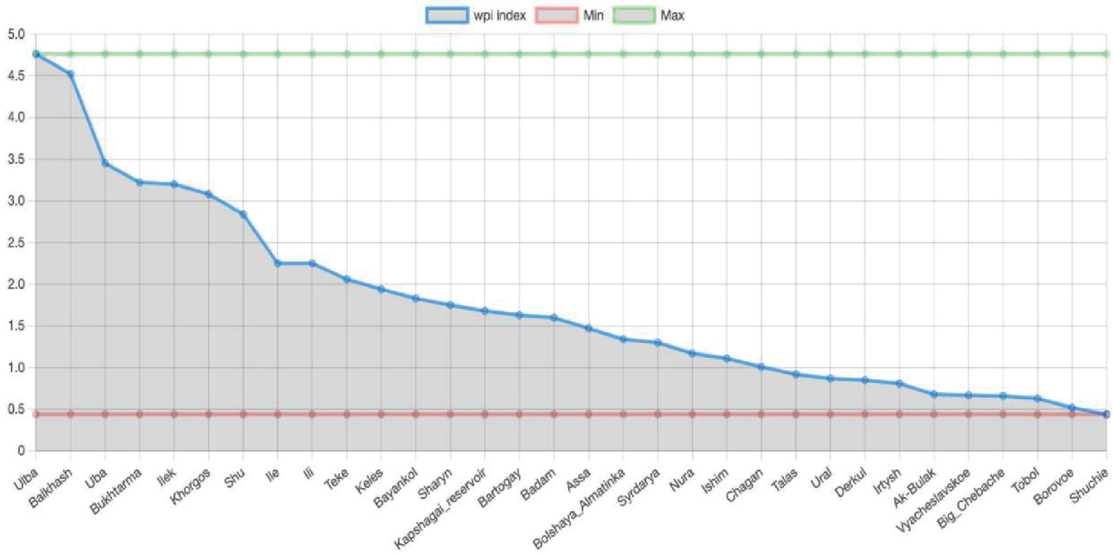

Figure 8. The 3rd section: Dynamic visualization of data (github.com/open-data-kazakhstan/owl-embebded-web-app[26])

Results

The results show that the capacity to scrutinize objects within the basin is greatly enhanced through the comprehensive population of the OWL-based database with diverse and disparate data sources. This augmentation empowers to execute advanced queries, thus unearthing deeper insights and forging connections between various indicators. This monitoring system proves instrumental in broadening our analysis by considering additional elements such as temperature and precipitation, which significantly influence the overall state of the basin, as indicated in reference [27]. Furthermore, the robust data management system ensures the perpetuity of data relevance, affording the flexibility to seamlessly append new objects and sources as needed. This adaptability guarantees that our information remains current and comprehensive, reinforcing the efficacy of our investigative efforts.

Conclusion

The study presents a pioneering methodology for the management of intricate hydrological resources through the use of OWL-Integrated Information. The Owlready2 library's role in generating OWL ontologies from well-organized Excel and CSV data is especially significant. This allows for accurate knowledge depiction and enables automated logical reasoning, thereby amplifying the system's ability to make data-driven intelligent choices.

The developed web interface serves as a potent instrument for scholars and data analysts. It not only allows for immediate querying of the OWL-centric ontology database but also incorporates advanced functionalities like intricate visualizations and data filtration. This is particularly advantageous for deciphering complex data patterns and trends related to the Ile-Balkhash basin.

Data governance and administration have been carefully planned. The deployment of GitHub Actions for automated data updates and the «Open Data Kazakhstan» GitHub organization for data storage are innovative strategies that add resilience to the study. These elements assure data currency and promote collaborative endeavors, which are vital for tackling urgent challenges in the realm of water resource management.

In summary, the study offers a structured, efficient, and potent methodology for the utilization and governance of knowledge within informational ecosystems. It holds substantial implications for the management of water resources, particularly in areas like the Republic of Kazakhstan, where water scarcity is an urgent concern. The study serves as a compelling example of the transformative potential of OWL-Integrated Information in effecting significant scientific and environmental advancements.

Supplementary Materials

ETL pipeline is located under Open Data Kazakshtan Organization in Github https://github.com/data-engineering-kaznu/ili-balkhash-basin-ontology-data

Dashboard is published and available online at https://ili-balkhas-basin-ontology.vercel.app/

OWL ontology processor source code is located in Github https://github.com/Titrom025/PyTableMiner/

OWL-Embedded Information and Analytical Framework source code is located in Github https://github.com/data-engineering-kaznu/owl-embebded-web-app

Acknowledgment

This work was carried out and sponsored within the framework of the scientific project AP09261344 "Development of methods for automatic extraction of spatial objects from heterogeneous sources for information support of geographic information systems".

References

[1] International Lake Environment Committee. (2023). *Lake Balkhash. World Lakes Database*. https://wldb.ilec.or.jp/Display/html/3571

[2] Dolbeshkin, M., Puzikov, E., & Povetkin, R. (2017). Modeling of transboundary Ile-Balkhash river basin in Central Asia. *News of the National Academy of sciences of the Republic of Kazakhstan*, 5, 171-180.

[3] Hydrological monitoring of water bodies of the Republic of Kazakhstan. (2023). Retrieved from http://ecodata.kz:3838/app_hydro/

[4] National Hydrometeorological Service of the Republic of Kazakhstan. (2023). *Monthly State of the Environment Newsletter*. https://www.kazhydromet.kz/ru/ecology/ezhemesyachnyy-informacionnyy-byulleten-o-sostoyanii-okruzhayuschey-sredy/2023

[5] Cherednichenko, A.V., Cherednichenko, Al.V., Cherednichenko, V.S., Storozhenko, N.D., Kozhachmetova, E.P., Kupchishin, A.I., & Abdrahimov, R.G. (2019). Dynamic of Balkhash lake level under climate change conditions. Institute of Biology and Biotechnology at, al Farabi University, Almaty, Kazakhstan.

[6] Shen, B., Wu, J., Zhan, S., & Jin, M. (2021). Residues of organochlorine pesticides (OCPs) and polycyclic aromatic hydrocarbons (PAHs) in waters of the Ili-Balkhash Basin, arid Central Asia: Concentrations and risk assessment. *Chemosphere*, 273.

[7] Isbekov, K.B., Tsoy, V.N., Crétaux, J.-F., Aladin, N.V., Plotnikov, I.S., Clos, G., Berge-Nguyen, M., & Assylbekova, S.Z. (2019). Impacts of water level changes in the fauna, flora, and physical properties over the Balkhash Lake watershed. *Lakes Reserv. Res. Manag.*

[8] Allemang, D., & Hendler, J. (2020). Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL. *ACM Books*, 3.

[9] Patel, A., Debnath, N.C., & Bhushan, B. (2022). Semantic Web Technologies. Research and Applications. *CRC Press.*

[10] Heath, T., Bizer, C., & Berners, L. (2023). Linked Data - The Story So Far. Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web (1st ed.). *ACM Books*, 52.

[11] Po, L., Bikakis, N., & Desimoni, F. (2020). Linked Data Visualization: Techniques, Tools, and Big Data (Synthesis Lectures on the Semantic Web: Theory and Technology). *Morgan & Claypool Publishers.*

[12] Galitsky, B. (2022). Multi-case-based reasoning by syntactic-semantic alignment and discourse analysis. *Artificial Intelligence for Healthcare Applications and Management*, 2, 154-196.

[13] Motik, B., Patel-Schneider, P., & Parsia, B. (Eds.). (2012). OWL 2 Web Ontology Language. *Structural Specification and Functional-Style Syntax*, 2.

[14] Gyrard, A., Datta, S.K., & Bonnet, C. (2018). A Survey and Analysis of Ontology-Based Software Tools for Semantic Interoperability in IoT and WoT Landscapes. *IEEE*.

[15] de Boer, T., Paltan, H., Sternberg, T., & Wheeler, K. (2021). Evaluating Vulnerability of Central Asian Water Resources under Uncertain Climate and Development Conditions: The Case of the Ili-Balkhash Basin. *Water*, 13, 615. https://doi.org/10.3390/w13050615

[16] Nurtazin, S., Thevs, N., Iklasov, M., Graham, N., Salmurzauli, R., & Pueppke, S. (2019). Challenges to the sustainable use of water resources in the Ili River basin of Central Asia. *E3S Web Conf.*, 81, 01009. https://doi.org/10.1051/e3sconf/20198101009

[17] Nurtazin, S., Pueppke, S., Ospan, T., Mukhitdinov, A., & Elebessov, T. (2020). Quality of Drinking Water in the Balkhash District of Kazakhstan's Almaty Region. *Water*, 12, 392. https://doi.org/10.3390/w1202039

[18] Hein, T., Funk, A., Pletterbauer, F., et al. (2019). Management challenges related to long-term ecological impacts, complex stressor interactions, and different assessment approaches in the Danube River Basin. *River Res Applic.*, 35, 500–509. https://doi.org/10.1002/rra.3243

[19] Duan, J., Zou, S., Chen Y., Nover, D., Fang, G., & Wang, Y. (2020). Sustainable water management for cross-border resources: The Balkhash Lake Basin of Central Asia, 1931-2015. *Journal of Cleaner Production*, 263, 121614. https://doi.org/10.1016/j.jclepro.2020.121614

[20] Dataflows library. (2023). Retrieved from https://github.com/datahq/dataflows

[21] Zhiyenbayev, M. (2017). Data Package v1 Specifications. What has Changed and how to Upgrade. *Open Knowledge Lab*. https://okfnlabs.org/blog/2017/10/11/upgrade-to-data-package-specs-v1.html

[22] Ile-Balkhash basin ontology data Republic of Kazakhstan. (2023). Retrieved from https://github.com/data-engineering-kaznu/ile-balkhas-basin-ontology-data/blob/master/datapackage.json

[23] Ali, S.M.F., & Wrembel, R. (2017). From conceptual design to performance optimization of ETL workflows: current state of research and open problems. *The VLDB Journal*, 26, 777–801. https://doi.org/10.1007/s00778-017-0477-2

[24] Open Data Kazakhstan. (2023). Retrieved from https://github.com/open-data-kazakhstan

[25] Github action. (2023). Retrieved from https://github.com/data-engineering-kaznu/ile-balk.

[26] Owl embedded web framework. (2023). https://github.com/data-engineering-kaznu/owl-embedded-web-app

[27] Guo, L., & Xia, Z. (2014). Temperature and precipitation long-term trends and variations in the Ili-Balkhash Basin. *Theoretical and Applied Climatology*, 115(1-2), 219-229.