Article

Not peer-reviewed version

The Development of the Monitoring Water Resources Ontology as a Research Tool for Sustainable Regional Development

Assel Ospan *, Madina Mansurova , Vladimir Barakhnin , Aliya Nugumanova , Roman Titkov

Posted Date: 29 June 2023

doi: 10.20944/preprints202306.2065.v1

Keywords: knowledge graph; ontology; semantic web; water resources monitoring; open data; sustainable development; RDF triples; SPARQL

 doi:10.20944/preprints202306.2065.v1

Article

The Development of The Monitoring Water Resources Ontology As A Research Tool for Sustainable Regional Development

Assel Ospan [1,*], Madina Mansurova [2], Vladimir Barakhnin [3], Aliya Nugumanova [4] and Roman Titkov [5]

1 Al-Farabi Kazakh National University; asselyaospan@gmail.com; https://orcid.org/0000-0002-1860-6997
2 Al-Farabi Kazakh National University; mansurova.madina@gmail.com; https://orcid.org/0000-0002-9680-2758
3 Federal Research Center for Information and Computing Technologies; barakhnin@ngs.ru; https://orcid.org/0000-0003-3299-0507
4 Astana IT University; yalisha@yandex.kz; https://orcid.org/0000-0001-5522-4421
5 Federal Research Center for Information and Computing Technologies; r.titkov@g.nsu.ru; https://orcid.org/0009-0003-8579-3600
* Correspondence: asselyaospan@gmail.com; Tel.:+7-708-151-22-39

Abstract: The development of knowledge graphs for water resources as a tool for studying sustainable regional development is a relevant task, as the increasing deterioration of water bodies affects the ecology, economy of the country, as well as the health of the population living near water bodies. Solving this problem requires the development of water resource monitoring methods based on semantic data analysis is required to detect and track sources of pollution and provide recommendations to support decision-making. This research proposes a method of semantic analysis for modeling knowledge graphs and constructing rules that enable the integration of data from heterogeneous sources, semantic compatibility, support for reasoning and inference, facilitate queries and search, and represent knowledge in the field of water resource monitoring. As a result, an ontology based on the SWRL language is created, which enables the expression of complex relationships and derivation of new knowledge, SSN for interaction between the ontology and sensor data, and Time Ontology, which models and analyzes information related to temporal aspects.After populating the ontology and establishing logical connections, SPARQL queries are implemented to retrieve new knowledge about water resources.

Keywords: knowledge graph; ontology; semantic web; water resources monitoring; open data; sustainable development; RDF triples; SPARQL.

1. Introduction

Water resources play an important role in human life and ecosystems, and their effective monitoring and management are essential for sustainable development. Thus, the resolution adopted by the UN General Assembly in 2016 "International Decade for Action "Water for Sustainable Development", 2018–2028" [1] emphasizes the need for increased attention to sustainable development and integrated water resources management in order to achieve socio-economic and environmental objectives and to the implementation and promotion of relevant programs and projects. The President of Kazakhstan, Kassym-Zhomart Tokayev, said that by 2050 Kazakhstan could enter the category of countries in dire need of water [2].

Monitoring of the Ili-Balkhash basin (hereinafter IBB) is extremely important because of the importance of this basin as the largest lake ecosystem and a source of valuable natural resources, since almost 40% of all hydropower resources of Kazakhstan are concentrated in IBB. The basin of Lake Balkhash occupies a vast territory and covers several regions of Kazakhstan and China [3]. In the Kazakh part of the basin is located the city of Almaty—a large metropolis with a population of 2.5 million. According to environmentalists, Lake Balkhash is threatened by an ecological catastrophe: its main tributary, the Ili River, is getting smaller every year due to a decrease in water volume [4].

In the light of the above, there is an obvious need to create an intelligent system for monitoring the water resources of the Ili-Balkhash basin in order to gain new knowledge about the state of water use facilities and make informed management decisions based on them. When creating such systems, it becomes necessary to develop technologies capable of processing and analyzing large volumes of semantically related data related to water resources. One of the key features of this work is the use of a variety of data sources, including physical sources – sensor sensors, web sources - official reports, documents and other information resources in the field of water consumption and ecology [5–9]. The use of heterogeneous data sources makes it possible to obtain more complete and reliable information about the quality, quantity and condition of water resources.

Since water monitoring consists of many parameters such as level, flow, quality, chemical composition, etc., it was necessary to organize the processing of these parameters in such a way as to ensure the integration of water data for effective search, semantic analysis and reasoning. Thus, it was decided to develop a knowledge graph for this subject area [10]. The knowledge graph for monitoring water resources has the following advantages:

1. provides a framework for integrating data from various sources, including sensors, monitoring stations, databases and external water body datasets;
2. allows you to capture contextual information, which presents spatial and temporal measurements of water level and flow, water quality parameters, hydrological factors, which provide an opportunity to get a comprehensive understanding of the complex interactions and dynamics in the aquatic ecosystem;
3. displays the semantic relationships between water resources objects, which allows for deep data analysis, facilitating queries and predicting the consequences of changes in the state of waters;
4. detects factors influencing water quality or level, allowing you to explore each node and the connections between them to identify hidden patterns and anomalies;
5. provides a decision support system related to the monitoring of water resources, making it possible to generate recommendations with the help of subject matter experts;
6. is easily scalable and flexible, which allows it to be expanded and adapted as new data becomes available or monitoring requirements or parameters change.

The knowledge graph is built using ontology and semantic models. This paper presents an ontology model based on the SWRL [11] language, which provides the ability to express complex relationships and derive new knowledge in ontologies. The use of the technology of the semantic sensor network SSN [12] makes it possible to interact with data from sensors and observations, ensuring the integration and processing of these data. In addition, Time Ontology [13] is used, which models and analyzes information related to temporal indicators, which is especially important in the context of water resources monitoring.

The further structure of the work is as follows. Section 2 contains a list of related works; Section 3 briefly characterizes the study area. Section 4 contains a description of data sources and methods for their processing, including ontology construction methods, as well as its structure (modules, properties, and rules). Section 5 formulates the main results of the work, Section 6 contains a discussion to conduct a general analysis of the purpose of this work, Section 7 Conclusion, which presents plans for further research.

2. Related Works

The application of the ontology for monitoring water resources is successfully practiced all over the world. The authors of [14] proposed a semantic web method for modeling ontology and constructing rules for monitoring river water quality, which also used observations from sensors. As a result of their work, a system was obtained to track the entire process of pollution events and warn of approaching epidemiological values. This work has been very helpful in developing the third module of our proposed water monitoring ontology (see Section 4.3).

Before creating an ontology, it was necessary to reformat the data of their heterogeneous sources into a machine-readable format; in our case, to replenish the knowledge graphs, the triplet format (XML, RDF, JSON) was required. This practice was successfully applied in [15], where the authors presented an architecture for integrating heterogeneous sources into an unified knowledge graph, from which useful knowledge about water quality was successfully extracted by analyzing the physicochemical and biological properties using spatiotemporal values and normative documents about water quality. The methods outlined in this work allowed us to develop our own architecture for the process of creating a water monitoring ontology, shown in Figure 5, which we expanded with three methods of data parsing and the inclusion of additional data on the level and flow of water.

Sensor data play a key role in monitoring water objects, so the SSN ontology provides interaction with sensor and observation data. A successful application of the SSN ontology is presented in [16], where an ontology was developed using data from hydrological sensors, by importing Time Ontology and instantiating hydrological classes and establishing reasoning rules. Based on this work, we developed Module 2.

The SWRL and SSN methods played the basis for creating linked data, since these methods allow you to create complex relationships and derive new knowledge from the ontology. The SWRL method has been successfully applied in solving problems from other areas, such as the diagnosis and treatment of vector-borne diseases [17], solving the problem of knowledge sharing about complex engineering systems for diagnostics [18], to facilitate preventive maintenance in industry using fuzzy clustering. Also in [19,20] the results of successful use of SSN are presented.

For several years, we have been monitoring the dynamics of the width of the Ili River on the border with China using remote sensing images from Landsat and Sentinel, the results of which were published in [21]. Long-term analysis shows that since 1980 the width of the Ili River has decreased from 400 m to 270 m, which once again proves the decrease in the water level of the river. The results of this work prompted us to continue the analysis and create a unified system for monitoring Ili water resources. Module 4 includes socio-economic indicators in the regions fed by the Ili River, including disease statistics, since in [22] the authors present their evidence of the impact of surface water quality on diseases of the population living near water objects. A correlation was also found between the growth of the population of the Almaty region and the decrease in the quantity/deterioration of the quality of waters at the IBB.

3. Study Area

The Ili-Balkhash basin is one of the largest lake ecosystems on the planet and is a unique natural complex that occupies a vast territory of 413 thousand km^2, 85% of which (353 thousand km^2) is located on the territory of Kazakhstan and 15%—on the territory of China. The Kazakhstan part of the Ili-Balkhash basin includes the territory of Almaty, Zhambyl and Karaganda regions. There is a large metropolis in the basin - the city of Almaty, with a population of 2.5 million. 5 large rivers flow into Lake Balkhash: Ili (length 1439 km), Karatal (390 km), Ayaguz (492 km), Lepsy (418 km), Aksu (316 km). The total volume of potential hydropower resources in the basins of Lake Balkhash is 63.5 billion kWh, almost 40% of all hydropower resources of Kazakhstan are concentrated here, and the potential of the Ili River is 7008 million kWh, which is 18.2% of the potential of all IBB rivers, namely 35.5 billion kWh. One of the active water users in the Ili-Balkhash basin is irrigated agriculture. It existed here more than a century ago: for example, in 1915 the area of irrigated land was 290 thousand hectares [23].

According to environmentalists, Lake Balkhash is threatened by an ecological catastrophe: its main tributary, the Ili River, is getting smaller every year due to a decrease in the volume of water. Now we are working with China only on an agreement on water distribution. During the implementation of the two-year EU project on Balkhash, it was calculated that a minimum flow from China of 12 km3 per year should be maintained. But now it has significantly decreased to 8 km^3 [24].

At present, the main consumer of water in this basin, both on the territory of the Republic of Kazakhstan and on the territory of China, is irrigated agriculture, for the needs of which more than 70% of water resources are spent. As recent data show, the use of water for irrigated agriculture tends to increase, especially in China, which raises concerns for the safety of Lake Balkhash and for the ecological situation in the basin. The situation is also aggravated by global and regional climate change and the degradation of glaciers in the basin.

Therefore, studying the state of the basin in terms of the impact of economic activity on the environment, identifying trends and developing recommendations for mitigating anthropogenic impact and climate change play an important role in developing a science-based strategy for managing the use and protection of water resources and applying the principles and approaches of Integrated Water Management (IWM).

The problems of IWM implementation in Kazakhstan, as well as the data obtained in the course of this study, are relevant and in demand [24]. This study used data from 63 monitoring stations (Figure 1), where various sensors were installed, and observations that collected data from January 1, 2001 to the present. Observation parameters obtained by these sensors include water level, flow and temperature, information about floods, as well as the state of surface water quality [5,6].

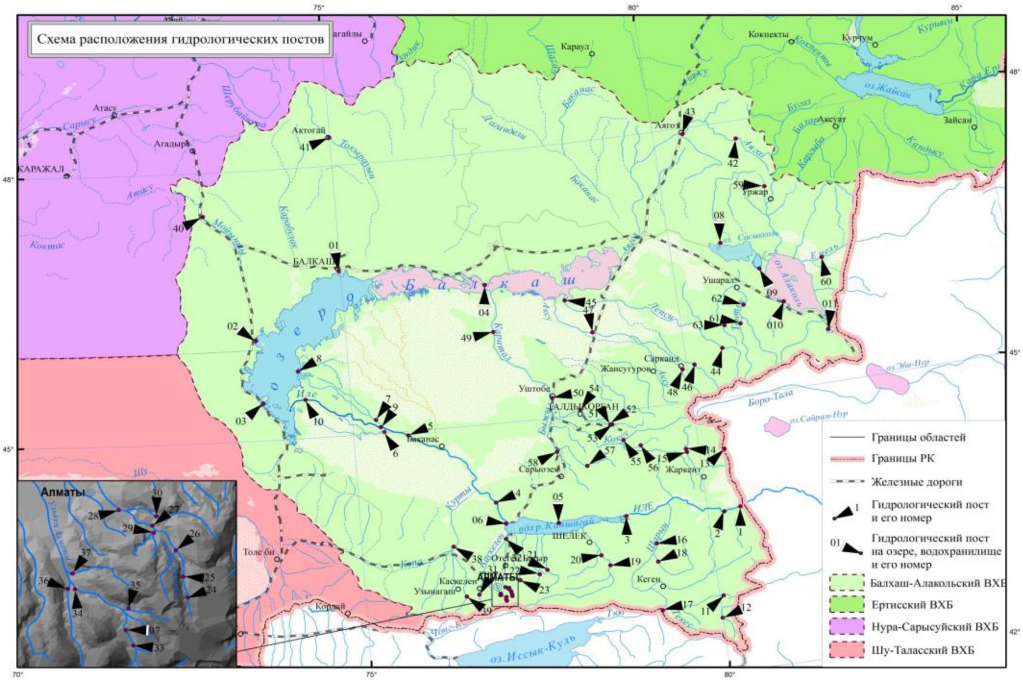


Figure 1. The layout of the hydrological posts of the Ili-Balkhash basin

4. Materials and Methods

4.1. Data Sources and Pre-Processing

To build an ontology for monitoring the water resources of the IBB, it was necessary to collect data. The problem was that there is no unified database for the water resources of Kazakhstan, so the data had been collected from heterogeneous sources and divided into 4 groups: 1. Water Regularions, 2. Sensor Data, 3. Water Objects, 4. Socio-economic Indicators.

For Water Regularions group, we collected water regulation data, which set thresholds for water use and their quality, which is measured by the Water Pollution Index (WPI) parameter. According to the WPI parameter, the classes of surface water pollution and their characteristics are determined, which are presented in Table 1. The data for the table were extracted from the Information and Legal System of Normative Legal Acts of the Republic of Kazakhstan from the paragraph on the Water Code of the Republic of Kazakhstan [8]. The data on this web page is presented in text and tables as shown in Figure 2, so the water regulation data was extracted using the QURMA tool into a structured format [26] (Figure 6).

Table 1. Determination of Water Classes by WPI indicators.

Class	Water Quality Characteristic	Water Pollution Index (WPI)	Domestic and Drinking Water Use	Domestic Water Use
I class	"very clean"	0,0-0,3	quite suitable	suitable
II class	"clean"	0,3-1,0	suitable	suitable
III class	"moderately polluted"	1,0-2,5	suitable for cleaning	suitable
IV class	"polluted"	2,5-4,0	not suitable	not suitable
V class	"dirty"	4,0-6,0	not suitable	not suitable
VI class	"very dirty"	6,0-10	not suitable	not suitable
VII class	"extremely dirty"	>10	not suitable	not suitable

(**a**)

5	Алюминий (Al^{3+})	0,5	с.-т.	2
6	Аммиак (по азоту)	2,0	с.-т.	3
7	Барий (Ba^{2+})	0,7	с.-т.	2
8	Бериллий (Be^{2+})	0,0002	с.-т.	1
9	Бор (B, суммарно)	0,5	с.-т.	2
10	Ванадий	0,1	с.-т.	3
11	Висмут	0,1	с.-т.	2
12	Вольфрам	0,05	с.-т.	2
13	Железо (Fe, суммарно)	0,3	орг.	3
14	Кадмий (Cd, суммарно)	0,001	с.-т.	2
15	Кобальт	0,1	с.-т.	2
16	Кремний	10,0	с.-т.	2
17	Литий	0,03	с.-т.	2
18	Марганец (Mn, суммарно)	0,1	орг.	3
19	Медь (Cu, суммарно)	1,0	орг.	3
20	Молибден (Mo, суммарно)	0,25	с.-т.	2
21	Мышьяк (As, суммарно)	0,05	с.-т.	2

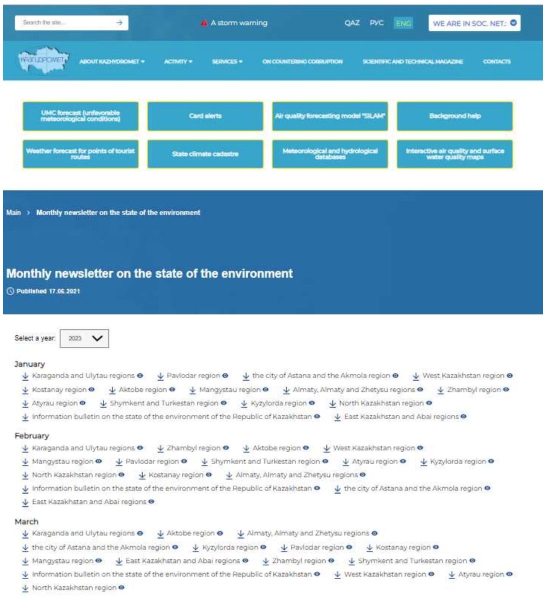
(**b**)

Figure 2. Information and legal system of normative legal acts on the Water Code of the Republic of Kazakhstan: (**a**) Web page; (**b**) type of data storage on this web page.

Sensor Data is the basis of the knowledge graph of water resources monitoring, as it contains daily/monthly/quarterly observations from sensors about the level, flow and quality of water. Records were collected from the installed gauging stations in the basin, and all records from the sensors are uploaded by the National Hydrometeorological Service of the Republic of Kazakhstan [6] to their website in the form of a report in PDF files, examples of these files are presented in Figure 3. A total of 63 hydro stations have been installed at the IBB, which have been collecting data since 1995, dividing data into 2 categories: 1. Observations on the level and flow of water [5]; 2. The state of the quality of surface waters [6]. Category 1 contains data on the following parameters: location of hydrological posts, water level, state of a water object, water temperature, meteorological conditions, water discharges, ice thickness and snow depth on ice, ice phenomena at the site of the post, information about floods and rain floods. Here, all data is already recorded in a structured file format (CSV, Excel, SQL), and stored on the official Eco-data website [5]. Category 2 stores observations of surface water quality in the Almaty region, carried out at 42 gates of 22 water objects (rivers Ili, Tekes, Korgas, Kishi Almaty, Esentai, Ulken Almaty, Chilik, Charyn, Bayankol, Kaskelen, Karkara, Esik, Turgen, Talgar, Temirlik, Karatal, Aksu, Lepsi, lakes Ulken Almaty, Alakol, Balkhash and reservoirs

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

Kapchagay), where 44 physical and chemical indicators of quality are determined: temperature, suspended solids, transparency, pH, dissolved oxygen, BOD5, COD, main ions of salt composition, biogenic elements, organic substances (petroleum products, phenols), heavy metals, pesticides.



(**a**)

Результаты мониторинга качества поверхностных вод на территории Алматинской, Жетысуской областях и г. Алматы

Основным нормативным документом для оценки качества воды водных объектов Республики Казахстан является «Единая система классификации качества воды в водных объектах» (далее – Единая Классификация).

По Единой классификации качество воды оценивается следующим образом:

Таблица 8

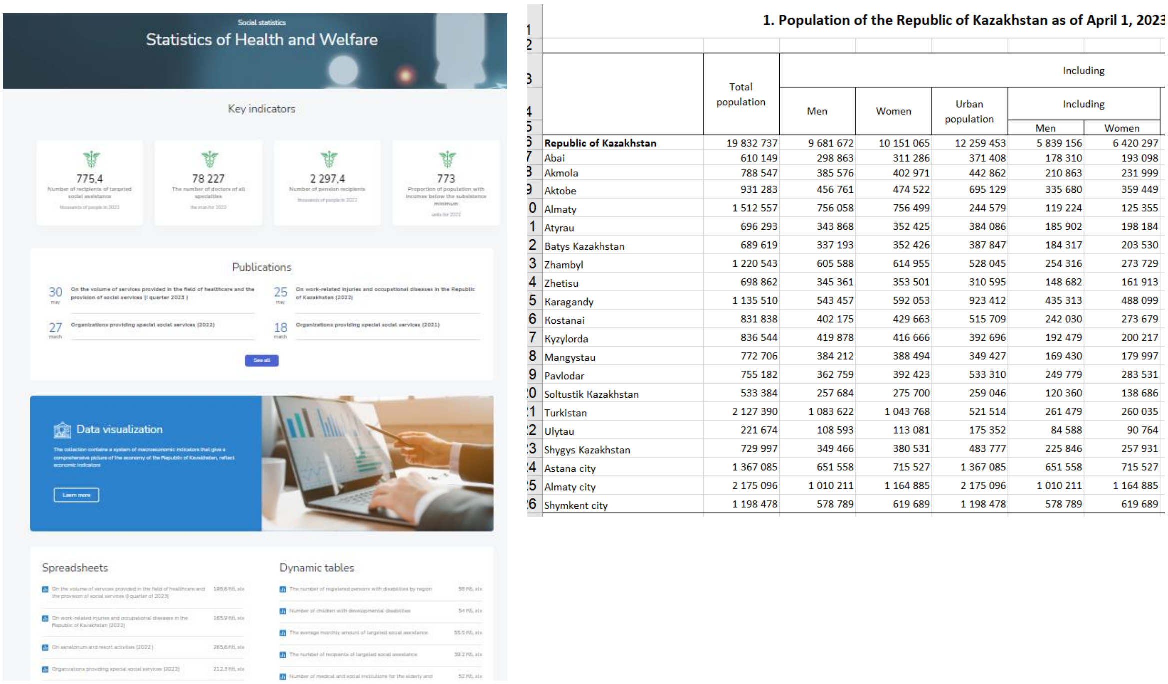
Наименование водного объекта	**Класс качества воды**		**Параметры**	**ед. изм.**	**концентр ация**
	январь 2022 г.	**январь 2023г.**			
река Киши Алматы	3 класс	4 класс	Магний	мг/дм^3	41,2
река Есентай	2 класс	2 класс	ХПК	мг/дм^3	22
река Улькен Алматы	2 класс	2 класс	Нитрит анион	мг/дм^3	0,121
река Иле	3 класс	3 класс	Магний	мг/дм^3	24,1
река Шилик	3 класс	4 класс	Магний	мг/дм^3	37
река Шарын	3 класс	4 класс	Взвешенные вещества	мг/дм^3	13
рекаТекес	3 класс	3 класс	Магний	мг/дм^3	25,3
река Коргас	2 класс	2 класс	Нитрит анион	мг/дм^3	0,177
река Баянкол	2 класс	3 класс	Фосфор общий	мг/дм^3	0,209
река Есик	2 класс	4 класс	Взвешенные вещества	мг/дм^3	15
река Каскелен	3 класс	3 класс	Магний	мг/дм^3	21,6
река Каркара	3 класс	3 класс	Фосфор общий	мг/дм^3	0,227
			Магний	мг/дм^3	28,7
река Тургень	3 класс	4 класс	Взвешенные вещества	мг/дм^3	15
река Талгар	2 класс	2 класс	Фосфор общий	мг/дм^3	0,183
река Темерлик	3 класс	4 класс	Взвешенные вещества	мг/дм^3	16
река Лепси	2 класс	3 класс	Магний	мг/дм^3	25,55
река Аксу	1 класс*	3 класс	Магний	мг/дм^3	25,3
река Каратал	3 класс	3 класс	Магний	мг/дм^3	21,567

(**b**)

Figure 3. Web page of the National Hydrometeorological Service of the Republic of Kazakhstan: (**a**) Monthly newsletter on the state of the environment; (**b**) surface water quality monitoring report for Almaty region for 2022–2023.

The source for Water Objects is the open encyclopedia Wikipedia [9], from which general data on the IBB, its rivers and lakes were taken, with data such as physical and geographical description, soils and vegetation, hydrography, glaciers, hydropower resources, knowledge of river flow and economic activity.

And the last item is Socio-economic indicators about the regions where the water bodies of the basin are located, obtained from the Bureau of National Statistics of the Republic of Kazakhstan [7]. Here, all statistical records are stored in excel and text files, as shown in Figure 4. Since there are many socio-economic indicators, indicators were chosen that are influenced by the resources and quality of nearby water objects. Thus, the following indicators were chosen: population, birth rate/mortality, diseases of the circulatory system associated with iodine deficiency, malignant neoplasms, acute infections of the upper respiratory tract, life expectancy, types and activities of industrial companies. Table 2 presents 4 groups of sources and the content of the indicators.



	Total population	Men	Women	Urban population	Including: Men	Including: Women
Republic of Kazakhstan	19 832 737	9 681 672	10 151 065	12 259 453	5 839 156	6 420 297
Abai	610 149	298 863	311 286	371 408	178 310	193 098
Akmola	788 547	385 576	402 971	442 862	210 863	231 999
Aktobe	931 283	456 761	474 522	695 129	335 680	359 449
Almaty	1 512 557	756 058	756 499	244 579	119 224	125 355
Atyrau	696 293	343 868	352 425	384 086	185 902	198 184
Batys Kazakhstan	689 619	337 193	352 426	387 847	184 317	203 530
Zhambyl	1 220 543	605 588	614 955	528 045	254 316	273 729
Zhetisu	698 862	345 361	353 501	310 595	148 682	161 913
Karagandy	1 135 510	543 457	592 053	923 412	435 313	488 099
Kostanai	831 838	402 175	429 663	515 709	242 030	273 679
Kyzylorda	836 544	419 878	416 666	392 696	192 479	200 217
Mangystau	772 706	384 212	388 494	349 427	169 430	179 997
Pavlodar	755 182	362 759	392 423	533 310	249 779	283 531
Soltustik Kazakhstan	533 384	257 684	275 700	259 046	120 360	138 686
Turkistan	2 127 390	1 083 622	1 043 768	521 514	261 479	260 035
Ulytau	221 674	108 593	113 081	175 352	84 588	90 764
Shygys Kazakhstan	729 997	349 466	380 531	483 777	225 846	257 931
Astana city	1 367 085	651 558	715 527	1 367 085	651 558	715 527
Almaty city	2 175 096	1 010 211	1 164 885	2 175 096	1 010 211	1 164 885
Shymkent city	1 198 478	578 789	619 689	1 198 478	578 789	619 689

(**a**) (**b**)

Figure 4. Web page of the Bureau of National Statistics of the Republic of Kazakhstan; (**a**) Webpage with uploaded reports; (**b**) excel files of statistics data for 2023.

Table 2. Heterogeneous data sources

Heterogeneous Data Sources	Source	Content
I Water Regulations	Order on Approval of the Sanitary Rules "Sanitary and Epidemiological Requirements for Water Sources, Places of Water Intake for Domestic and Drinking Purposes, Domestic and Drinking Water Supply and Places of Cultural and Domestic Water Use and Safety of Water Bodies" [8]	1. General Provisions; 2. Sanitary and epidemiological requirements for water sources; 4. Indicators of drinking water quality; 5. Microbiological and parasitological indicators of drinking water quality; 6. Hygienic standards for the content of harmful substances in drinking water; 7. Quantity, frequency of water sampling; 8. List of indicators.
II Sensor Data	Daily hydrological bulletin of the Republic of Kazakhstan [5]	1. Location of hydrological posts; 2. Water level; 3. The state of the water object; 4. Water temperature; 5. Weather conditions; 6. Water consumption; 7. Thickness of ice and height of snow on ice; 8. Ice phenomena at the site of the post; 9. Information about floods and rain floods.
	Monthly State of the Environment Newsletter [6]	1. The main sources of air pollution; 2. The state of the quality of atmospheric air; 3.The chemical composition of atmospheric precipitation. 4. The state of the quality of surface waters; 5. Radiation environment.
III Water Objects	Information about IBB from Wikipedia [9]	1. Physico-geographical description; 2. Soils and vegetation; 3. Hydrography; 4. Glaciers; 5. Hydropower resources; 6. Knowledge of river flow; 7. Economic activity.

IV Socio-economic Indicators	Bureau of National Statistics of the Republic of Kazakhstan [7]	1. Population; 2. Birth/mortality of the population; 3. Diseases of the circulatory system associated with iodine deficiency; 4. Malignant neoplasms; 5. Acute infections of the upper respiratory tract; 6. Life expectancy; 7. Types and activities of industrial companies.

4.2. Methods for Creating an Ontology

Water resources monitoring is essential to understand the status and trends of water resources development, identify potential problems, implement decision-making processes, and appropriate management and conservation strategies. This helps ensure the sustainable use of water, protect ecosystems, and protect human health and well-being. Based on the studies of water resources monitoring, the most important aspects were identified, which include the systematic observation and analysis of water resources to assess their quality, quantity, and general condition. In general, water resources monitoring provides for the following activities:

1. water quality monitoring, which includes the measurement and analysis of various physical, chemical and biological characteristics of water;
2. monitoring of water quantity, aimed at assessing the availability and volume of water resources;
3. monitoring the ecological aspects of water bodies, which includes the study of organisms, habitats and biodiversity present in the aquatic environment;
4. water use monitoring, which includes tracking the consumption, distribution and use of water resources for various purposes such as agriculture, industry, domestic use and recreation;
5. monitoring of the early warning system to detect and alert potential risks and emergencies, which may include monitoring factors such as water levels, flow rates, weather conditions and water quality parameters to provide timely warnings of floods, droughts, pollution incidents or other water hazards;
6. monitoring the regulatory framework and compliance with water quality standards and environmental regulations to ensure compliance with legal requirements, permit conditions and water quality recommendations established by government authorities.

The proposed ontology of water resources monitoring was created by applying the SWRL and SSN semantic network methods. The SWRL method based on the integration of OWL and RuleML was proposed in 2019 [11], and now this method is used to solve many problems:

- defining rules in ontologies for expressing complex relationships and inferring new knowledge;
- the application of logical reasoning to obtain new facts or conclusions;
- creation of intelligent systems and expert systems using reasoning based on rules;
- development of intelligent agents that can reason and make decisions autonomously.

SWRL rules consist of an antecedent (also called a body) and a consequent. The antecedent indicates the conditions that must be met for the rule to be applicable, and the consequent indicates new information or conclusions that can be drawn when the rule is triggered, i.e., if the antecedent is true, then the consequent must also be true:

rule ::= 'Implies(' [URIreference] { annotation } antecedent consequent ')'
antecedent ::= 'Antecedent(' { atom } ')'
consequent ::= 'Consequent(' { atom } ')'

The semantic sensor network SSN was proposed in 2017 [12], the main goal of this ontology is interaction with sensor and observation data. Nowadays SSN provides integration, detection and interaction of sensors from various sources, and also solves many other problems:

- allows you to integrate sensor data from heterogeneous sources;
- improves detection of sensor data by providing a standardized view;
- allows you to apply semantic analysis methods to sensor data;
- allows enriching sensor data with contextual information;

- makes it particularly effective in IoT environments where numerous sensors and devices generate huge amounts of data.

Also, an important part of the ontology is occupied by Time Ontology [13], which is used to model and analyze time-related information. Time Ontology is very useful for creating our ontology, as it has the following properties: it provides a structured and formal representation of temporal concepts such as intervals, durations, points in time, and temporal relationships; supports the integration and compatibility of time-related data from different sources; allows you to model and track events over time, facilitates the analysis and understanding of historical data, supports the formulation and execution of time-based queries on temporal data; provides a link to the semantic web and related data.

In the process of creating the IBB Water Resources Monitoring Ontology, data had been collected from various sources and a thorough analysis was made to select needed data that would provide an effective ontology. Based on the successful work on creating an ontology for water resources [14,15,16], based on SWRL, SSN and Time Ontology, indicators were identified that can be used to perform semantic analysis.

After defining the methods and rules for creating an ontology, it was necessary to develop its architecture. The right choice of architecture is especially important in our situation, since all the data for monitoring water resources were in heterogeneous sources. The architecture is shown in Figure 5, where the left column shows our heterogeneous data sources, which are divided into 4 groups:

1. Water Regulations - consists of the regulatory rules of the Water Code of the Republic of Kazakhstan, all data are presented on web pages (.html format).
2. Sensor Data - data from sensors that are installed at the gauging stations of the basin and contain information from 1995 on days, months and years. All sensor records are presented in the form of reports (PDF format) on the website of the National Hydrometeorological Service of the Republic of Kazakhstan [6].
3. Water Objects - information about water objects of the IBB, which includes general data, such as the volume of water, the length of rivers, etc. All data is taken from the open encyclopedia Wikipedia (.html format).
4. Socio-economic indicators - here are collected indicators from the Bureau of National Statistics of the Republic of Kazakhstan, namely those that are affected by the quality and quantity of water.

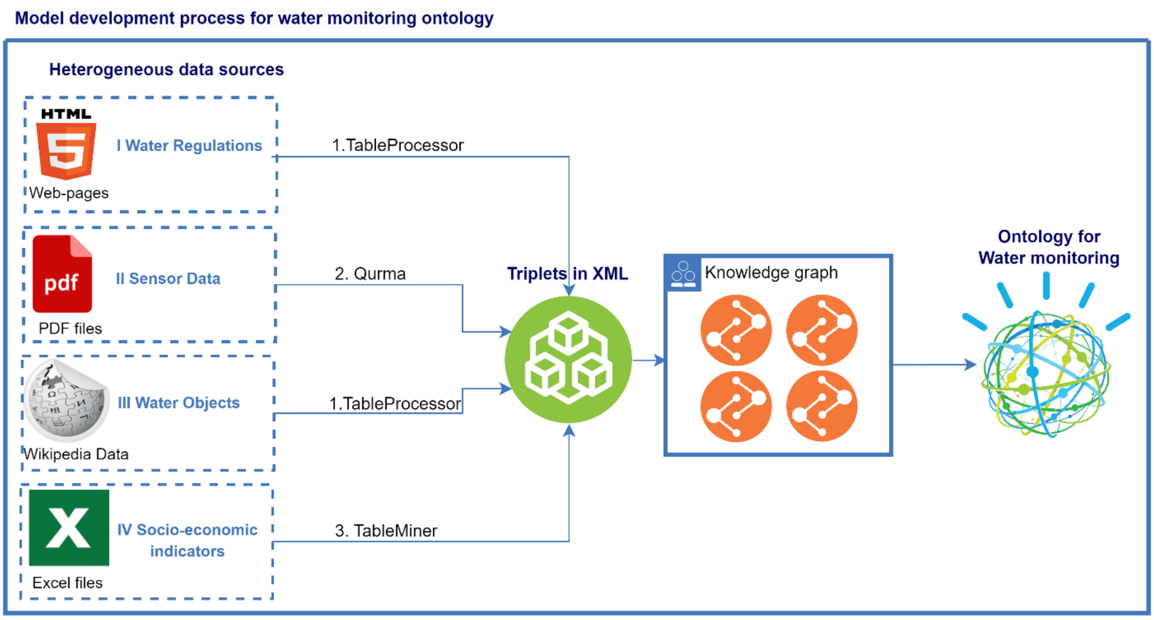


Figure 5. Architecture of the process of creating an ontology of monitoring water resources.

 doi:10.20944/preprints202306.2065.v1

Next, we see arrows that head towards to triples in XML, here we need to pay attention to the records of arrows - these are the methods by which data was derived from non/semi-structured sources in the form of triplets: object-attribute-value. The first method is TableProcessor [25] developed by we, that uses semantic analysis to extract useful information from web pages and save the data in a machine-readable form, in JSON format. Our second method is Qurma [26], this method extracts useful knowledge from pdf-files to replenish the knowledge base, and all data is extracted in triplet format. The third tool TableMiner [27] is designed to extract data from excel files, which performs semantic analysis, it is based on machine learning.

All three of these methods were used to represent all of our data as triplets in XML format, as shown in Figure 6, in order to further build an ontology. After presenting all the data in the form of triplets, a knowledge graph is built based on logical connections and relationships. All logical connections and ways of obtaining new knowledge are described in the next subsection Ontology Modules.


```
<?xml version="1.0" encoding="UTF-8"?>
<water_quality>
  - <row>
      <index>0</index>
      <Rivers_KZ>Irtysh </Rivers_KZ>
      <Type_of_water_objects>river</Type_of_water_objects>
      <Regions>East Kazakhstan</Regions>
      <WPI__April_2005>3.13</WPI__April_2005>
      <WPI__March_2006>2.3</WPI__March_2006>
      <WPI__April_2006>2.35</WPI__April_2006>
      <Ingredients_and_indicators_of_water_quality>Copper</Ingredients_and_indicators_of_water_quality>
      <Average_concentration>0.0019</Average_concentration>
      <Multiplicity_of_exceeding_the_MPC>1.9</Multiplicity_of_exceeding_the_MPC>
      <Classes>III class </Classes>
      <Water_quality_characteristic>"moderately polluted"</Water_quality_characteristic>
    </row>
```


Figure 6. Representation of water quality data as triplets in XML format

4.3. Ontology Modules

Building an ontology consists of seven steps:

1. determine the subject area and scope of the ontology;
2. consider the possibility of reusing existing ontologies;
3. list important terms in the ontology (main classes);
4. define classes and class hierarchy;
5. define class properties;
6. define threshold data values;
7. create entity instances.

Based on the collected data, an ontology had been developed, shown in Figure 7, which links all the above data, forming a unified system that allows monitoring water resources.

The ontology consists of 5 modules, and these modules are interconnected, which allows us to extract useful data by the necessary objects or by time intervals.

***I. Water Regulations Module (WR)*:** the development of this module was built using the regulatory rules of the Water Code [8]. This module contains physicochemical and microbiological parameters, sampling points, hydrological units and associated methods. There are 2 classes defined in this module:

1.1. The *WR:Water* Class is used to describe the quality of water objects based on water regulations. This class has 5 subclasses, each of which describes pollution classes according to the water pollution index (WPI) (Table 1).

1.2. The *WR:Water Hygiene Standards* Class describes the chemical and microbiological composition of water, on the basis of which water pollution classes are determined. This class consists of 4 data properties: oil products, surfactants (SAS), organic substances, inorganic substances (cations).

 doi:10.20944/preprints202306.2065.v1

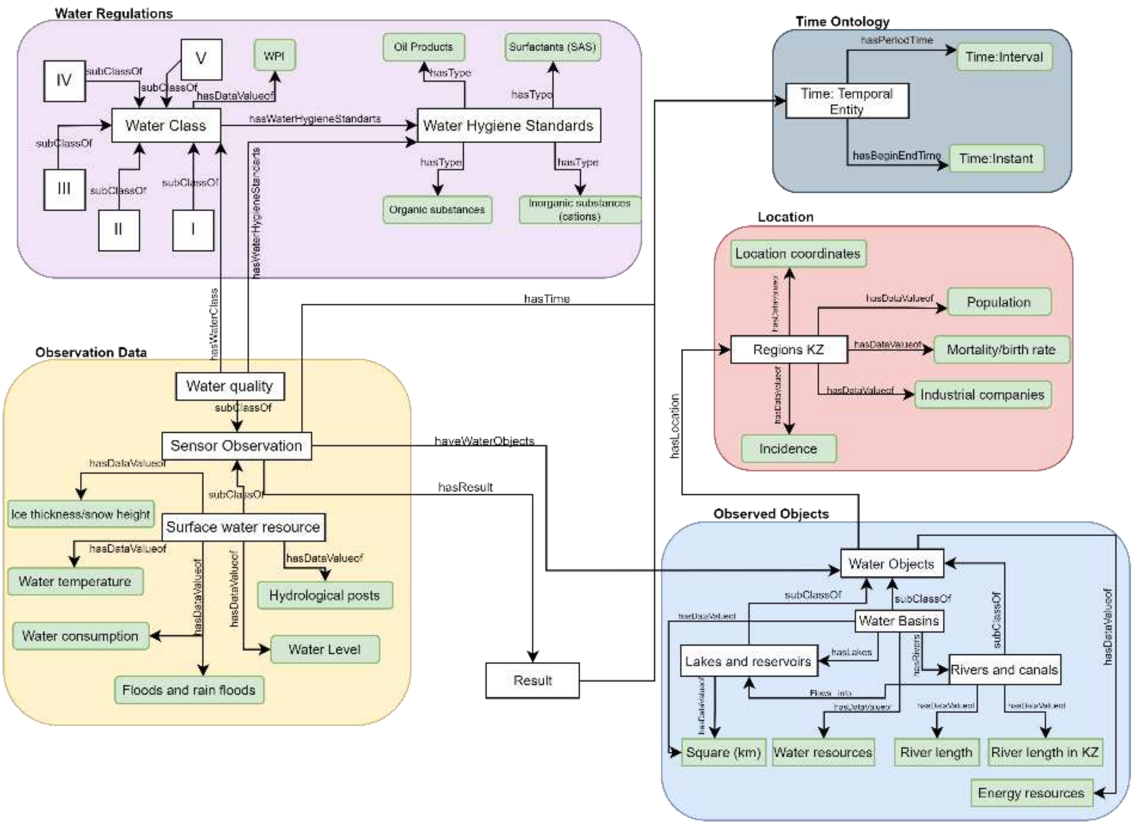


Figure 7. Scheme of ontology of water resources monitoring.

II. Observation Data Module (OD): this module is based on data from sensors that are installed at the hydro stations in the basin and contain information since 1995 by day, month and year. This module contains 1 superclass Sensor Observation and 2 of its subclasses: surface water resource and water quality:

2.1. The *OD:surface water resource* class contains annual data on the regime and resources of land surface waters and describes water parameters in the time interval, such as codes of hydrological posts, water level, water temperature, water discharges, ice thickness and snow depth on ice, information about flood and rain flood.

2.2. The *OD:water quality* class contains monitoring data on surface water quality in the territory of the IBB at 42 gates of 22 water bodies (rivers Ili, Tekes, Korgas, Kishi Almaty, Esentai, Ulken Almaty, Chilik, Charyn, Bayankol, Kaskelen, Karkara, Esik, Turgen, Talgar, Temirlik, Karatal, Aksu, Lepsi, lakes Ulken Almaty, Alakol, Balkhash and reservoir Kapchagai). When studying surface waters, 44 physical and chemical indicators of quality are determined in the taken water samples: temperature, suspended solids, transparency, hydrogen index (pH), dissolved oxygen, BOD5, COD, main ions of the salt composition, biogenic elements, organic substances (petroleum products, phenols), heavy metals, pesticides.

III. Observed Objects Module (OO): this module contains the main water objects in the water-economic basin. There is 1 superclass Water objects, which has 3 subclasses: water basins, lakes and reservoirs, rivers and canals:

3.1. of the rivers in Kazakhstan and where the river flows into.The *OO:water basins* class describes basins that contain subclasses: lakes and reservoirs, rivers and canals; and data properties such as basin area, water resource, and energy resource.

3.2. The *OO: lakes and reservoirs* class also contain area, water and energy data properties.

3.3. The *OO:rivers and canal* class contains the properties of the data and the length of the rivers, the length

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

IV. Time Ontology Module: Based on the OWL-2 DL ontology of time concepts for describing time properties, which has two subclasses: Time:Interval and Time:Instant.

4.1. *Time:Interval* describe the length in a certain interval.
4.2. *Time:Instant* describes describes one set time, where start and end must match..

V. Location Module (L): Location Module (L): describes the location, specifically the regions of Kazakhstan since each water body belongs to a region or city. This module consists of 1 main class Regions, and its properties:

5.1. *L: Regions* class contains data on all regions and cities of Kazakhstan, and all these regions have coordinates. The Regions class has indicators such as population, births/deaths, disease rates, and industrial companies.

4.4. Water Ontology Properties

Properties include object pointer properties and data properties. Our ontology defines 4 types of data properties:

1. *'hasDataValueof'*: contains observation data, fixed and historical data, in different measurement types and scales;
2. *'hasType'*: expresses that the object has data types - here this data property expresses the water quality composition data type, such as oil products, surfactants (SAS), organic substances, inorganic substances (cations));
3. *'hasPeriodTime'*: describes the length in a certain period of time (from 2000 to 2023);
4. *'hasBeginEndTime'*: describes one set time where start and end must match.

Properties-pointers to ontology objects create a unified system that connects both classes within modules and classes between modules. Object properties can be divided into seven categories:

1. The *subClassOf* property: This property describes the relationship between a top-level class and its subclass. The subclass inherits all attributes and operations from the superclass, so for example the WR:Water Class class has a WPI data property, and is associated with the WR:Water Hygiene Standards class, which means that all 5 of its subclasses have the same properties, they differ only in different WPI and quality indicators water.
2. The *hasWaterHygieneStandards* property: creates a relationship between the classes within module 1 - WR:Water Class and WR:Water Hygiene Standards, and between modules 1 and 2 with the Sensor Observation class. In the first case, the connection provides data integrity for determining the level of water pollution class based on data from the normative documents for the water code, which defines the limit values of chemicals in the composition of water, according to which water pollution classes are established. In the second case, the pointer property binds the OD:Sensor Observation class, which contains annual observations from the hydrological posts of each basin on such indicators as the regime and resource of surface waters, water level, water temperature, water discharges, ice thickness and snow depth on ice, flood and rainfall information and water quality.
3. The *hasWaterClass* property: associates the *OD:Water Quality* subclass from Module 2, which consists of microbiological water quality sensor sampling points observations, to the *WR:Water Class* class from Module 1.
4. The *haveWaterObjects* property: associates observation data from sensors with water objects such as basins, rivers and canals, lakes and reservoirs. In this ontology, the *OD:Sensor Observation* class from Module 2 creates a relationship with the *OO:Water Objects* class from Module 3. All sampling points and water state measurements are carried out at hydro posts, each hydro station is associated with a specific water object, and each water object has its own fixed properties, such as the square of a water object, the length of rivers and canals, water and energy resources. By creating a link between these classes, you can get information about each water object, as well as track the dynamics over a certain period of time.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

5. The *hasTime* property: links the data from the *OD:Sensor Observation* class to the *Time: Temporal Entity* class from the Time Ontology Module for temporal concepts, since all sensory data is recorded daily.
6. The *hasLocation* property: is used primarily for coordinate binding of water bodies to regions. In our ontology, the *OO:Water Objects* class is associated with the *Locations:Regionz KZ* class.
7. The *hasResult* property: the last link based on which the decision is made. Here, this pointer property binds the *OD:Sensor Observation* to the *Result* class, and this class is bound to the *Time: Temporal Entity* to track the result for a certain period of time.

4.5. Ontology Rules

Thus, we have developed our ontology by defining relationships between different modules to provide a spatiotemporal and legal context for assessing water quality in the study area. Figure 2 shows an overview of the developed ontology network maintained by Protégé [28] and expressed in OWL2. The ontology rules are built based on the SWRL and SSN.

Rule 1: the hasWaterHygieneStandards rule establishes the pollution classes of water objects based on the normative rules of the Water Code of the Republic of Kazakhstan. Classes are determined based on the WPI - this indicator is set by national regulatory authorities depending on the water object and its location. For water bodies of the Republic of Kazakhstan, the following parameters are set for determining the WPI as shown in Table 3.

Table 3. Hygienic standards for the content of chemicals in water to determine WPI.

№	Substance_Name	Standards (MPC), not More than in mg/L
1	Total mineralization (dry residue)	>1000
2	general hardness	>7.0 (mg-eq./L)
3	Oil products, total	>0,1
4	Surfactants (SAS), anionic	>0,5
5	Inorganic substances (cations)	Depends on the type of chemical substance
6	organic substances	Depends on the type of chemical substance

As noted above in Section 4.1, observations of the quality of surface waters in the Almaty region were carried out at 42 gates of 22 water bodies, and when studying surface waters, 44 physical and chemical quality indicators are determined in the selected water samples. Monitoring of the quality of bottom sediments and coastal soil was carried out at 14 control points of the Ili River, Lakes Balkhash and Alakol [6].

Based on these indicators, the WPI is calculated, based on which Rule 2 is determined.

Rule 2: the *hasWaterClass* rule is used to determine the suitability / unsuitability of water for domestic and drinking water use (Table 2). Sensory observations record the results of samples in the form of chemical substances and their concentrations. With the help of Rule 2, the *Water_Class* class is automatically determined, where the WPI limits are specified in the ontology. Figure 8 shows the setting of WPI limits for the *Water_Class*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

Figure 8. Populating the Water_Class with WPI Limits.

Rule 3: the *OD:Surface Water Resource* rule describes surface water resources with parameters such as code of hydrological posts, water level, water temperature, water discharge, ice thickness and snow depth on ice, high water and rain flood information. Based on this rule, statistical indicators are determined, for example, based on observation data, it is possible to calculate the maximum / minimum values of water discharges, information about recurring floods at a certain time, and at which hydro stations critical or abnormal values are observed.

Rule 4: the *hasWaterObjects* rule associates sensor observations with water objects from the Observed Objects Module. Here, water bodies are divided into 4 types: lakes, reservoirs, rivers and canals. Each water body, depending on the type, has parameters such as the squre of a lake or reservoir, the length of rivers or canals, a hydrological resource, and an energy resource. With this rule, the data from the sensors are divided into groups, and you can compare the flow and water level readings with the total volume or length of the water body.

Rule 5: the *hasTime* rule establishes temporal information that can be used to analyze temporal relationships, perform temporal queries, and support various temporal analyzes and applications: *tl:Instant*: represents a specific moment in time; *tl:Interval*: represents a time interval or duration; *tl:TemporalEntity*: represents a generic temporary entity. This rule associates an observation with the time it was made, or an event with a specific interval of time. You can also define properties for temporal comparisons.

Rule 6: the *hasLocation* rule performs coordinate binding of water objects to regions. The regions of Kazakhstan have socio-economic indicators: population, birth/death rate of the population, disease rates and industrial companies. This is one of the most important rules, since it can be used to identify the correlation between the pollution of water bodies and the incidence of the population, as well as to determine other similar correlations.

5. Results

The schematic structure of the ontology is implemented in Protégé, as shown in Figure 9. So at first we created a skeleton, that is, we defined the main classes and their subclasses: Regions KZ, Sensor Observation, Timestamp, Water Class, Water Hygiene standards, Water Objects.

 doi:10.20944/preprints202306.2065.v1

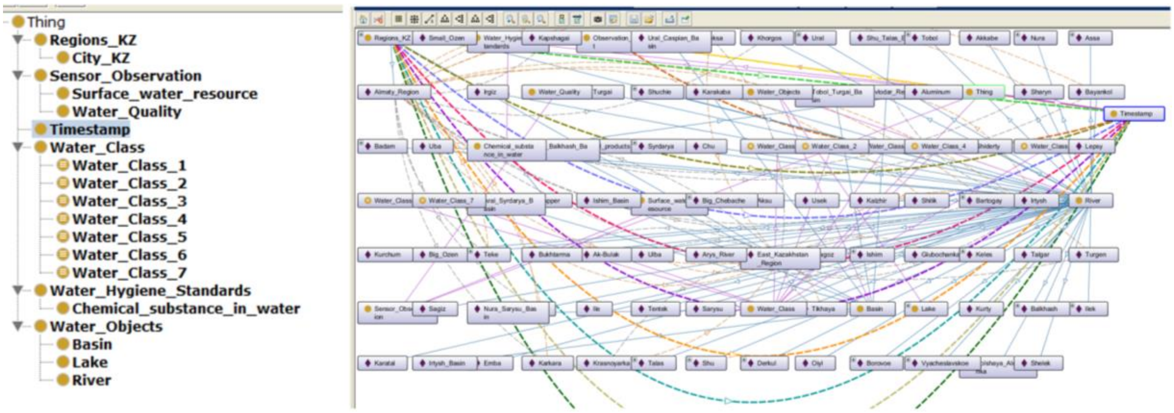


Figure 9. Schematic structure of the Ontology implemented by Protégé.

Further, using the owlready library [29], the ontology was imported into Python. This library provides the ability to create, modify, load and save ontologies, as well as perform various operations with them, such as searching for class instances, checking instance properties, etc. Then our data is imported, which is already reformatted in XML format. Next, the ontology is analyzed, which consists of the following steps:

1. analysis of objects;
2. property analysis;
3. relationship analysis;
4. analysis of classes of objects.

During these stages, the listpoooooooos and dictionaries listed below are completed:

```
object_names = []
data_properties = []
object_properties = []
class_names = []
name2object = {}
name2data_property = {}
name2object_property = {}
```

The first stage is the analysis of ontology objects, during which all objects from the ontology are extracted, their names are normalized (reduced to lower case) and entered into the “object_names” list. Also, a pair is entered into the “name2object” dictionary: the name of the object is a link to the object. These auxiliary data structures are needed to facilitate the task of parsing tables. Subsequent modules use the "object_names" list to find objects, and then use "name2object" to get a reference to the object in the ontology, for example, to add a new property to the object. Using a link to an object allows you to change the object directly in the ontology. Similar actions are carried out for properties (data properties) and relations (object properties). For properties and relationships, accessing the object by reference is necessary to validate the property. As a result, we get a populated ontology with Object Properties, Data Properties, Individuals and logical relationships between them, the results are shown in Figure 10.

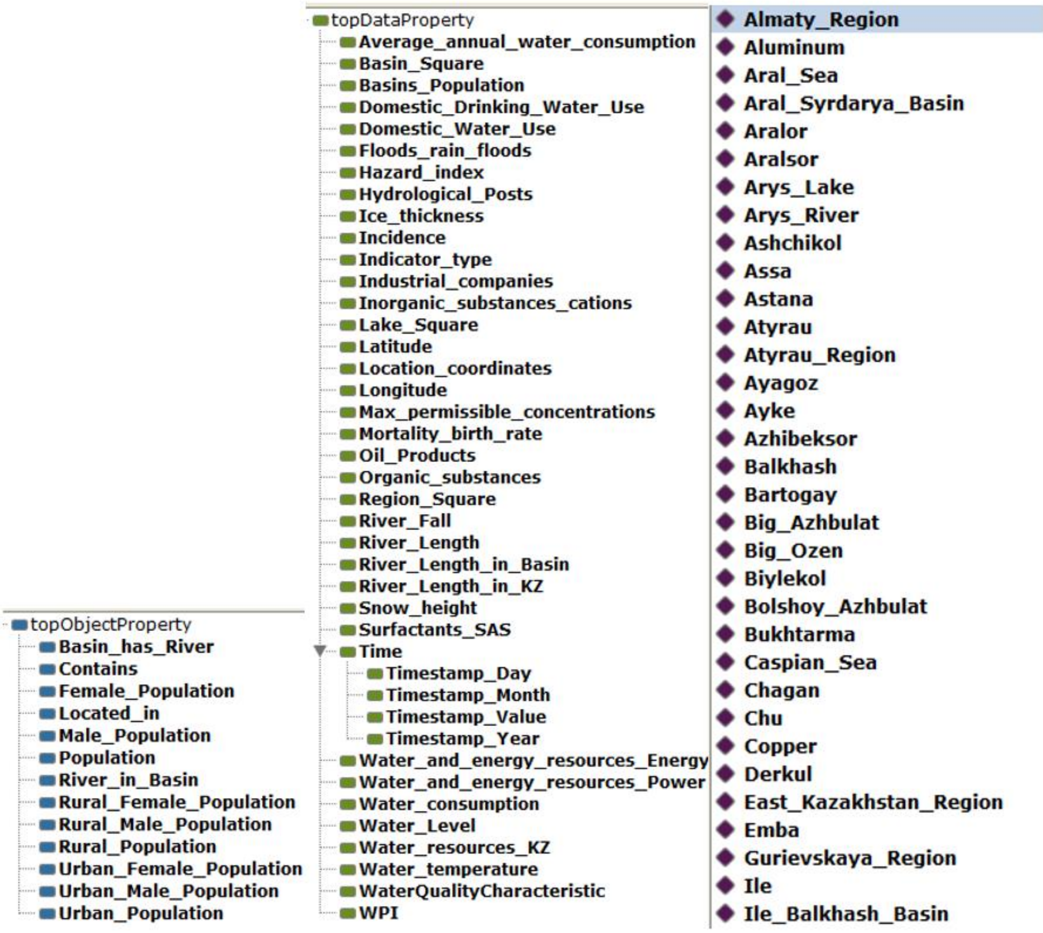


Figure 10. Object Properties, Data Properties, Individuals automatically populated using owlready.

After populating the ontology, it is necessary to check the correctness of the related data. To do this, we will choose our research area - IBB. Let's check for the Balkhash basin and the Ili river - only 7 Object Properties are defined for Balkhash and Ili, which show the following relationships: Balkhash is located in the Almaty and Zhambyl regions - this is the implementation of the links of the Regions and Water Objects classes, in the Balkhash basin there are the Charyn, Kurty, Ile, Chelek is an implementation of Basin River subclass links; in addition, 12 properties are defined that point to numeric attributes, as can be seen in Figure 11.

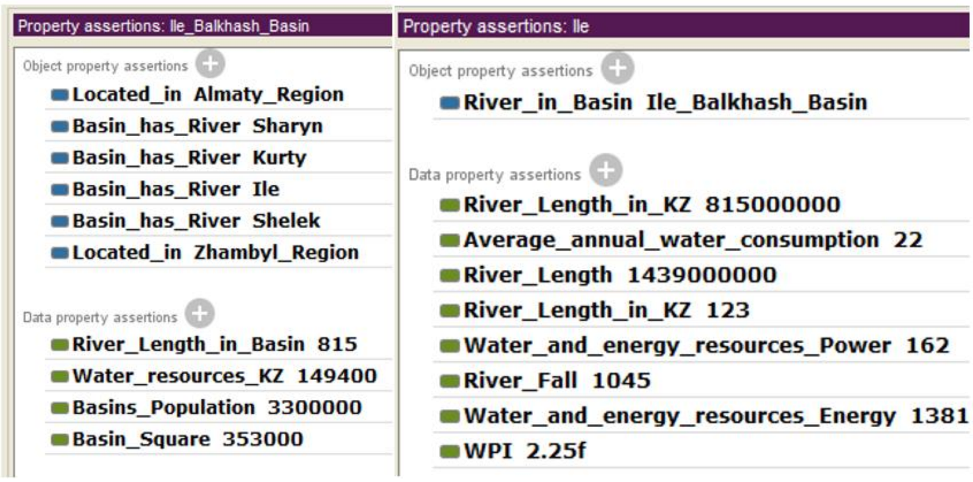


Figure 11. Defined Object Properties and Data properties for IBB.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

Since the main goal of creating this ontology is to establish the impact of the quality and quantity of IBB water resources on the socio-economic situation of the region to which they belong, Figure 12 shows all the associated Object Properties for the Almaty region, since the IBB is mainly located here. For the Almaty region, first of all, social indicators were determined, such as the urban population, the rural population, of which how many women/men, life expectancy, as well as morbidity rates for this region: diseases of the circulatory system associated with iodine deficiency, malignant neoplasms, acute infections of the upper respiratory tract. The impact of water quality on the health of the population living near large water objects was proved in [22]. Based on it, we can draw a conclusion from our data about the relationship between the incidence of the population and the quality of water resources.

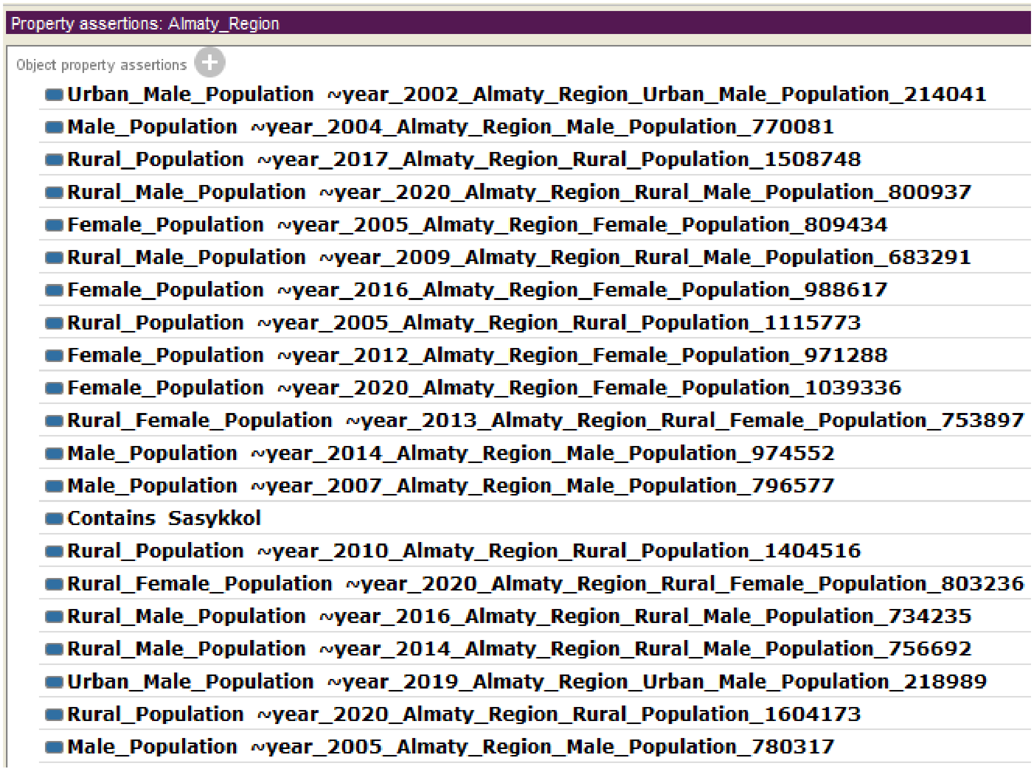


Figure 12. Related Object Properties for Almaty region

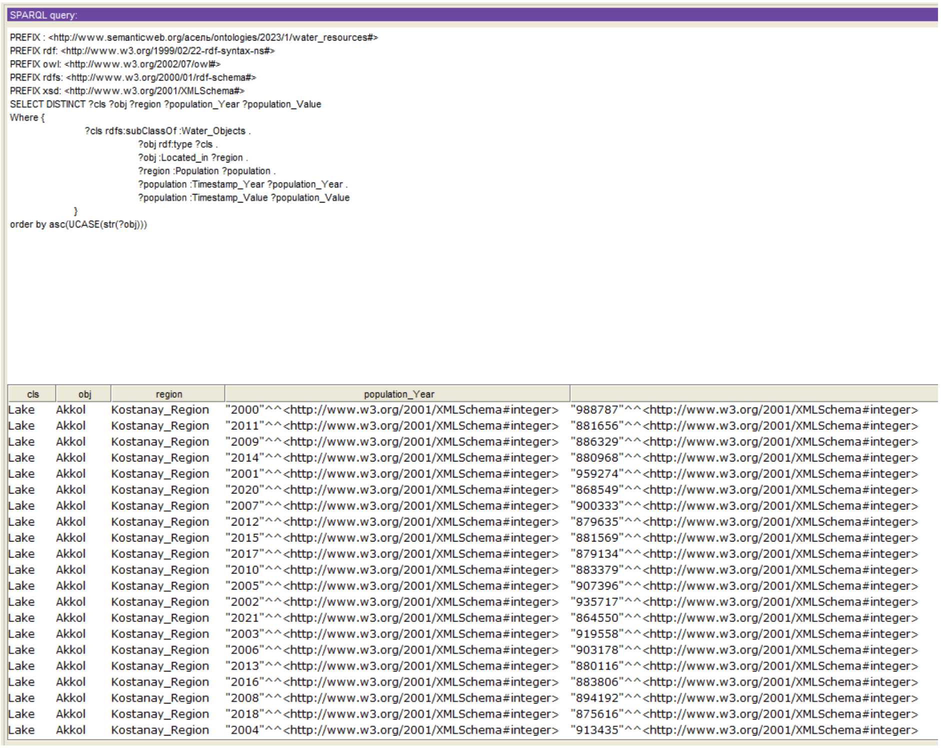
Now that the ontology has been populated and logical relationships have been established, we need to implement queries in SPARQL that will help us derive new knowledge. For our ontology, the following queries are implemented:

1. Query 1 implements the derivation of water objects, the regions to which these objects belong and the population of this region in the time interval, as shown in Figure 13.
2. Query 2 displays water objects and their WPI indicators, their pollution class based on WPI indicators. The query result is shown in Figure 14.
3. Query 3 displays water objects with given WPI indicators provided that 0<WPI<5, as shown in Figure 15.
4. Query 4 displays water bodies with high WPI values and the number of people suffering from diseases of the circulatory system associated with iodine deficiency. The query result is shown in Figure 16.

Thus, based on the created ontology, it is possible to implement many queries to obtain useful knowledge that we could not get if the data were stored in tables or other files.

Since we are using a Temporal Ontology, we can query and retrieve information based on temporal criteria, such as retrieving data within a specific time range or finding events that happened before or after a given timestamp.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2023 doi:10.20944/preprints202306.2065.v1

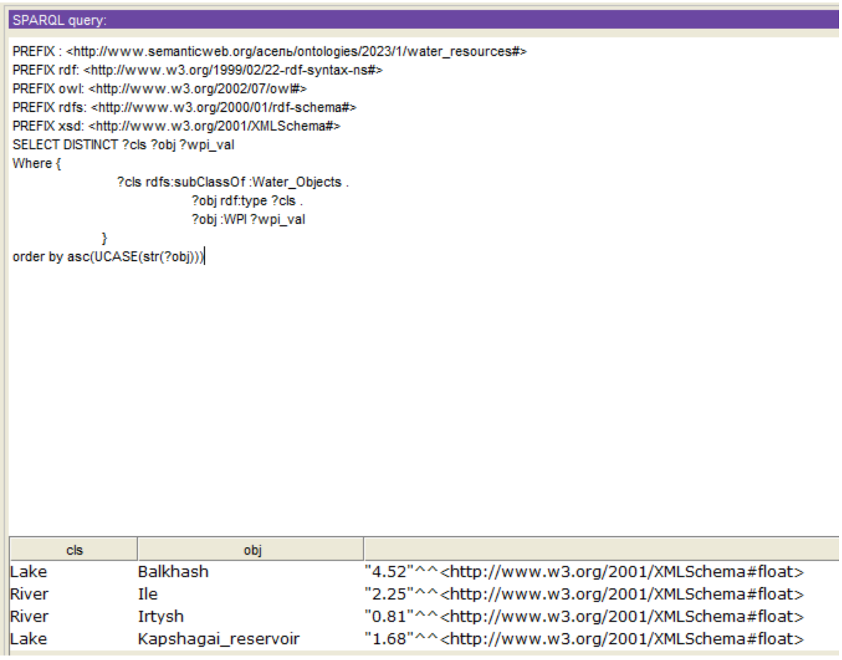


SPARQL query:

```
PREFIX : <http://www.semanticweb.org/асель/ontologies/2023/1/water_resources#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?cls ?obj ?region ?population_Year ?population_Value
Where {
        ?cls rdfs:subClassOf :Water_Objects .
                ?obj rdf:type ?cls .
                ?obj :Located_in ?region .
                ?region :Population ?population .
                ?population :Timestamp_Year ?population_Year .
                ?population :Timestamp_Value ?population_Value
        }
order by asc(UCASE(str(?obj)))
```

cls	obj	region	population_Year	
Lake	Akkol	Kostanay_Region	"2000"^^<http://www.w3.org/2001/XMLSchema#integer>	"988787"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2011"^^<http://www.w3.org/2001/XMLSchema#integer>	"881656"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2009"^^<http://www.w3.org/2001/XMLSchema#integer>	"886329"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2014"^^<http://www.w3.org/2001/XMLSchema#integer>	"880968"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2001"^^<http://www.w3.org/2001/XMLSchema#integer>	"959274"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2020"^^<http://www.w3.org/2001/XMLSchema#integer>	"868549"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2007"^^<http://www.w3.org/2001/XMLSchema#integer>	"900333"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2012"^^<http://www.w3.org/2001/XMLSchema#integer>	"879635"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2015"^^<http://www.w3.org/2001/XMLSchema#integer>	"881569"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2017"^^<http://www.w3.org/2001/XMLSchema#integer>	"879134"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2010"^^<http://www.w3.org/2001/XMLSchema#integer>	"883379"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2005"^^<http://www.w3.org/2001/XMLSchema#integer>	"907396"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2002"^^<http://www.w3.org/2001/XMLSchema#integer>	"935717"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2021"^^<http://www.w3.org/2001/XMLSchema#integer>	"864550"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2003"^^<http://www.w3.org/2001/XMLSchema#integer>	"919558"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2006"^^<http://www.w3.org/2001/XMLSchema#integer>	"903178"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2013"^^<http://www.w3.org/2001/XMLSchema#integer>	"880116"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2016"^^<http://www.w3.org/2001/XMLSchema#integer>	"883806"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2008"^^<http://www.w3.org/2001/XMLSchema#integer>	"894192"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2018"^^<http://www.w3.org/2001/XMLSchema#integer>	"875616"^^<http://www.w3.org/2001/XMLSchema#integer>
Lake	Akkol	Kostanay_Region	"2004"^^<http://www.w3.org/2001/XMLSchema#integer>	"913435"^^<http://www.w3.org/2001/XMLSchema#integer>

Figure 13. Query for displaying water objects, regions to which these objects belong and the population of this region in the time interval.

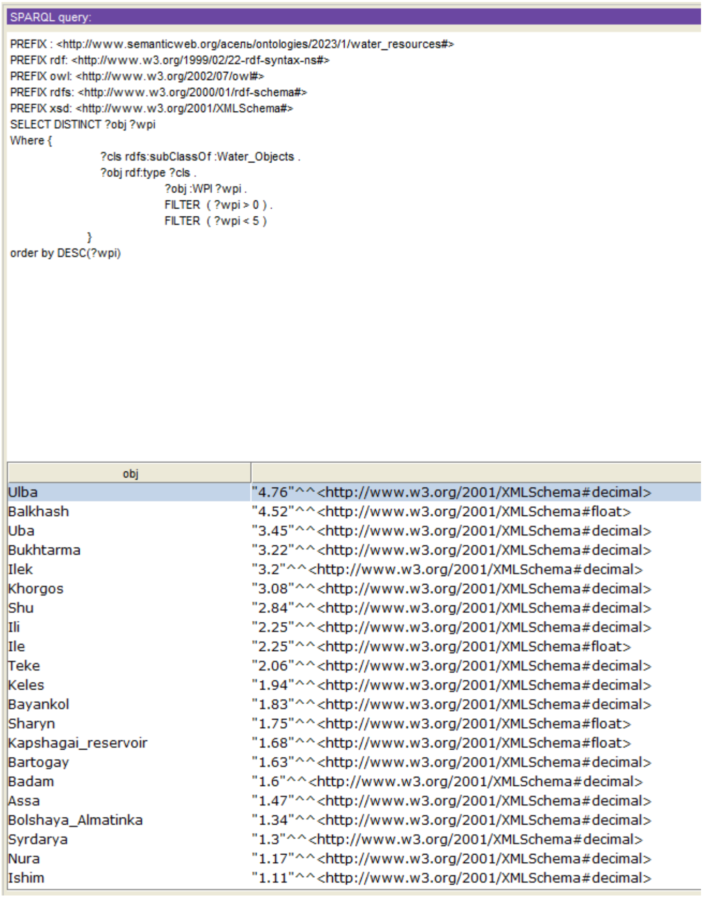


SPARQL query:

```
PREFIX : <http://www.semanticweb.org/асель/ontologies/2023/1/water_resources#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?cls ?obj ?wpi_val
Where {
        ?cls rdfs:subClassOf :Water_Objects .
                ?obj rdf:type ?cls .
                ?obj :WPI ?wpi_val
        }
order by asc(UCASE(str(?obj)))
```

cls	obj	
Lake	Balkhash	"4.52"^^<http://www.w3.org/2001/XMLSchema#float>
River	Ile	"2.25"^^<http://www.w3.org/2001/XMLSchema#float>
River	Irtysh	"0.81"^^<http://www.w3.org/2001/XMLSchema#float>
Lake	Kapshagai_reservoir	"1.68"^^<http://www.w3.org/2001/XMLSchema#float>

Figure 14. Query for displaying water bodies and their WPI indicators

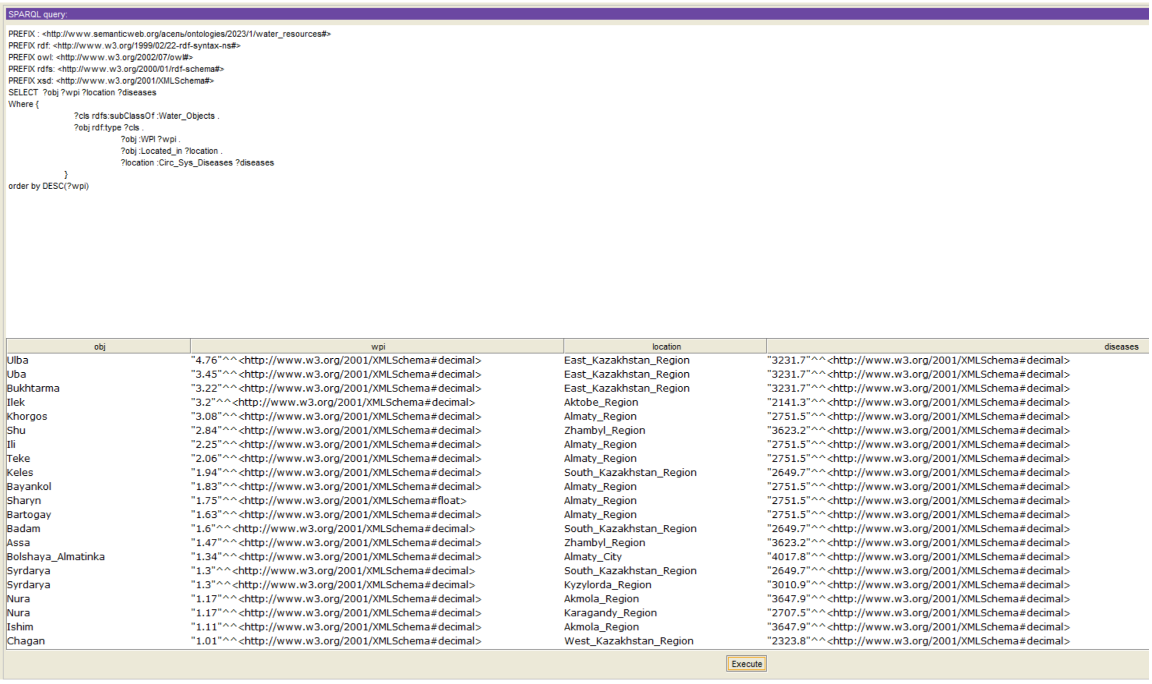


SPARQL query:

```
PREFIX : <http://www.semanticweb.org/асель/ontologies/2023/1/water_resources#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?obj ?wpi
Where {
        ?cls rdfs:subClassOf :Water_Objects .
        ?obj rdf:type ?cls .
            ?obj :WPI ?wpi .
            FILTER ( ?wpi > 0 ) .
            FILTER ( ?wpi < 5 )
      }
order by DESC(?wpi)
```

obj	
Ulba	"4.76"^^<http://www.w3.org/2001/XMLSchema#decimal>
Balkhash	"4.52"^^<http://www.w3.org/2001/XMLSchema#float>
Uba	"3.45"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bukhtarma	"3.22"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ilek	"3.2"^^<http://www.w3.org/2001/XMLSchema#decimal>
Khorgos	"3.08"^^<http://www.w3.org/2001/XMLSchema#decimal>
Shu	"2.84"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ili	"2.25"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ile	"2.25"^^<http://www.w3.org/2001/XMLSchema#float>
Teke	"2.06"^^<http://www.w3.org/2001/XMLSchema#decimal>
Keles	"1.94"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bayankol	"1.83"^^<http://www.w3.org/2001/XMLSchema#decimal>
Sharyn	"1.75"^^<http://www.w3.org/2001/XMLSchema#float>
Kapshagai_reservoir	"1.68"^^<http://www.w3.org/2001/XMLSchema#float>
Bartogay	"1.63"^^<http://www.w3.org/2001/XMLSchema#decimal>
Badam	"1.6"^^<http://www.w3.org/2001/XMLSchema#decimal>
Assa	"1.47"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bolshaya_Almatinka	"1.34"^^<http://www.w3.org/2001/XMLSchema#decimal>
Syrdarya	"1.3"^^<http://www.w3.org/2001/XMLSchema#decimal>
Nura	"1.17"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ishim	"1.11"^^<http://www.w3.org/2001/XMLSchema#decimal>

Figure 15. Query for inferring water bodies with given WPI indicators



SPARQL query:

```
PREFIX : <http://www.semanticweb.org/асель/ontologies/2023/1/water_resources#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT  ?obj ?wpi ?location ?diseases
Where {
        ?cls rdfs:subClassOf :Water_Objects .
        ?obj rdf:type ?cls .
            ?obj :WPI ?wpi .
            ?obj :Located_in ?location .
            ?location :Circ_Sys_Diseases ?diseases
      }
order by DESC(?wpi)
```

obj	wpi	location	diseases
Ulba	"4.76"^^<http://www.w3.org/2001/XMLSchema#decimal>	East_Kazakhstan_Region	"3231.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Uba	"3.45"^^<http://www.w3.org/2001/XMLSchema#decimal>	East_Kazakhstan_Region	"3231.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bukhtarma	"3.22"^^<http://www.w3.org/2001/XMLSchema#decimal>	East_Kazakhstan_Region	"3231.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ilek	"3.2"^^<http://www.w3.org/2001/XMLSchema#decimal>	Aktobe_Region	"2141.3"^^<http://www.w3.org/2001/XMLSchema#decimal>
Khorgos	"3.08"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Shu	"2.84"^^<http://www.w3.org/2001/XMLSchema#decimal>	Zhambyl_Region	"3623.2"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ili	"2.25"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Teke	"2.06"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Keles	"1.94"^^<http://www.w3.org/2001/XMLSchema#decimal>	South_Kazakhstan_Region	"2649.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bayankol	"1.83"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Sharyn	"1.75"^^<http://www.w3.org/2001/XMLSchema#float>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bartogay	"1.63"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_Region	"2751.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Badam	"1.6"^^<http://www.w3.org/2001/XMLSchema#decimal>	South_Kazakhstan_Region	"2649.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Assa	"1.47"^^<http://www.w3.org/2001/XMLSchema#decimal>	Zhambyl_Region	"3623.2"^^<http://www.w3.org/2001/XMLSchema#decimal>
Bolshaya_Almatinka	"1.34"^^<http://www.w3.org/2001/XMLSchema#decimal>	Almaty_City	"4017.8"^^<http://www.w3.org/2001/XMLSchema#decimal>
Syrdarya	"1.3"^^<http://www.w3.org/2001/XMLSchema#decimal>	South_Kazakhstan_Region	"2649.7"^^<http://www.w3.org/2001/XMLSchema#decimal>
Syrdarya	"1.3"^^<http://www.w3.org/2001/XMLSchema#decimal>	Kyzylorda_Region	"3010.9"^^<http://www.w3.org/2001/XMLSchema#decimal>
Nura	"1.17"^^<http://www.w3.org/2001/XMLSchema#decimal>	Akmola_Region	"3647.9"^^<http://www.w3.org/2001/XMLSchema#decimal>
Nura	"1.17"^^<http://www.w3.org/2001/XMLSchema#decimal>	Karagandy_Region	"2707.5"^^<http://www.w3.org/2001/XMLSchema#decimal>
Ishim	"1.11"^^<http://www.w3.org/2001/XMLSchema#decimal>	Akmola_Region	"3647.9"^^<http://www.w3.org/2001/XMLSchema#decimal>
Chagan	"1.01"^^<http://www.w3.org/2001/XMLSchema#decimal>	West_Kazakhstan_Region	"2323.8"^^<http://www.w3.org/2001/XMLSchema#decimal>

Execute

Figure 16. Result of the Query for the deriving of water objects with high WPI values and the quantity of people suffering from diseases of the circulatory system associated with iodine deficiency.

6. Discussion

The key question addressed in this study is the development of an ontology for water resources as a tool for studying sustainable regional development. By developing an ontology specific to environmental tasks, particularly the assessment of sustainable water use, the aim was to enhance understanding of the complex interrelationships between environmental factors, human activities, and the Sustainable Development Goals (SDGs) adopted by the United Nations General Assembly in 2015 [30]. For instance, Goal 6, "Clean Water and Sanitation," from this list entails:

- Ensuring access to clean water and sanitation services for all people.
- Ensuring sustainable use and management of water resources, including the preservation of related ecosystems, the protection and restoration of aquifers, reservoirs, and water ecosystems.
- Reducing water pollution and improving water quality, including the reduction of harmful chemical discharges and improving wastewater treatment.
- Improving water resource efficiency, reducing water losses in various sectors, including agriculture, industry, and urban infrastructure.
- Protecting and restoring ecosystems associated with water resources, such as rivers, lakes, aquifers, and wetlands, in order to maintain their ecological integrity and diversity.

In this section, we discuss the significance of the obtained results for monitoring the achievement of the mentioned goals. First and foremost, the ontology adapted to the context of water resources proved to be a valuable tool for organizing and structuring knowledge. The ontology serves as a conceptual framework that captures the interconnections and interdependencies among various entities, such as water bodies, economic entities, sources of pollution, humans, flora, and fauna. With the help of the ontology, we were able to create a knowledge base that facilitated the integration of diverse datasets and enabled efficient extraction of the required information. We developed sample queries to the knowledge base that allowed us to assess the impact of human activities on the qualitative and quantitative indicators of water resources.

One of the significant advantages of ontology-based approaches is the ability to perform advanced data analysis and logical reasoning. By applying semantic analysis methods, we were able to uncover hidden relationships and derive new knowledge from existing data. For example, using the ontology, we discovered previously unnoticed connections between classes of water pollution and the increase in the population's incidence of circulatory system diseases, as well as between the loss of biodiversity in river fauna and classes of river pollution. This finding highlights the potential of ontological analysis in uncovering complex ecological patterns and making data-driven decisions.

Moreover, the ontology-based approach facilitates knowledge sharing and collaboration among researchers, policymakers, and stakeholders. By providing a shared vocabulary and a common understanding of ecological concepts, the ontology serves as a bridge across different disciplines and enables effective communication. This aspect of collaboration is particularly relevant in the context of regional sustainable development planning, where interdisciplinary efforts and stakeholder engagement are crucial. The ontology acts as a central knowledge repository, promoting interdisciplinary collaboration and supporting a holistic approach to sustainable development goals.

In [31], a knowledge organization system based on an ontology is proposed, which models the key elements of the United Nations' global system of sustainable development goal indicators. This system currently includes 17 goals, 169 targets, and 231 unique indicators, along with over 450 related sets of statistical data supported by the global statistical community for monitoring progress towards the SDGs and the dataset containing these elements. In addition to formalizing and establishing unique identifiers for the components of the SDGs and their indicator system, the ontology includes mapping each goal, target, indicator, and data set to their corresponding terms and subjects in the United Nations Bibliographic Information System (UNBIS) and the EuroVoc Thesaurus, facilitating multilingual semantic search and content linking.

As noted by the authors of this work, in order to promote a holistic approach through coordinated policies and actions involving governments of different countries and stakeholders from all levels of society, it is crucial to develop tools that facilitate the discovery and analysis of interrelationships among various global SDG indicators derived from different data sources,

 doi:10.20944/preprints202306.2065.v1

information, and knowledge supported by different stakeholder groups. Essentially, this is an attempt to provide stakeholders with a means to publish data using shared terminology and URIs centered around SDG concepts, which helps enhance semantic compatibility of diverse data and information related to SDG information assets provided by various societal layers. In this work, we contribute to the formation of such an information asset, providing indicators and data for SDG 6 "Clean Water and Sanitation" at the regional level. Although we have not yet aligned our ontology with the system proposed by the authors, this task represents a technical step that is significantly facilitated by employing an ontological approach.

However, it is important to acknowledge the limitations and challenges associated with ontology development in the domain of water resources. Developing a comprehensive and accurate ontology requires significant domain knowledge and substantial effort. The process of ontology development involves iterative refinement and validation, which can be time-consuming and resource intensive. Additionally, supporting and updating the ontology is necessary to ensure its relevance and alignment with new ecological knowledge. Therefore, future research should focus on automatically expanding the ontology to cover additional aspects and further validating its usefulness in real-world applications.

7. Conclusions

In conclusion, we have successfully developed a knowledge graph based on ontologies using SWRL, SSN, and Time Ontology methods for monitoring the water resources of the Ili-Balkhash basin. By applying these methods, we have obtained an effective tool that allowed us to formalize new knowledge and rules, integrate data from various sensors and devices, providing more comprehensive and accurate information about the state of water resources, and account for temporal aspects, which is important for analyzing changes over time.

Furthermore, an important step was the data collection from heterogeneous sources using tools like Qurma, TableProcessor, and Table Miner. These tools enabled us to extract information from different data formats and transform it into a structured form, facilitating the integration of data into the knowledge graph.

The result of this research is the acquisition of new knowledge about the level, discharge, and quality of water resources, which significantly impact the socio-economic indicators of each region where these water bodies are located. This knowledge can be used for informed decision-making and the development of measures for sustainable water resource management.

Further work in this area aims to apply machine learning methods to work with our ontology, which will allow predicting future epidemiological zones, floods, or droughts, as well as supporting decision-making in the field of water resources. The application of machine learning will help uncover hidden patterns and trends, enabling more accurate forecasting and determination of effective measures for sustainable water resource management.

Author Contributions: Ospan A. developed an ontology model and rules, collected data, participated in the development of parsers for extracting useful knowledge, performed an experiment, wrote this paper. Mansurova M. suggested ideas and research methods, participated in the development of Qurma and TableProcessor tools for data extraction, advised in writing this article. Barakhnin V.—suggested ideas and methods for creating an ontology, participated in the creation of the TableProcessor tool for extracting data from excel files, suggested the idea of using the integration of socio-economic indicators with data on water resources, advised on writing this article. Nugumanova A.—the main author of the creation of the Qurma parser for extracting useful knowledge from PDF files, suggested ideas and research methods, wrote a discussion in this paper, advised on writing this article. Titkov R.—enriched the XML ontology with data, performed the tasks of implementing queries in SPARQL. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: 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”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Resolution adopted by the General Assembly on 21 December 2016. 71/222. International Decade for Action, "Water for Sustainable Development", 2018–2028. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N16/459/99/PDF/N1645999.pdf (accessed on December 2018)
2. The President of the Republic of Kazakhstan K-Zh.K.Tokayev said when Kazakhstan can become a country in dire need of water. Токаев сказал, когда Казахстан может стать "остро нуждающейся в воде" страной: 19 апреля 2023, 11:57 - новости на Tengrinews.kz (accessed on April 2023)
3. International Lake Environment Committee. "Lake Balkhash". World Lakes Database. https://web.archive.org/web/20110716085418/http://www.ilec.or.jp/database/asi/asi-54.html (accessed on October 2022)
4. Azattyq Rýhy - information and analytical agency. Why Balkhash is on the verge of disaste? https://rus.azattyq-ruhy.kz/interview/42487-pochemu-balkhash-na-grani-katastrofy (in russian) (accessed on October 2022)
5. Hydrological monitoring of water bodies of the Republic of Kazakhstan. http://ecodata.kz:3838/app_hydro/ (June 2023)
6. National Hydrometeorological Service of the Republic of Kazakhstan. "Monthly State of the Environment Newsletter". https://www.kazhydromet.kz/ru/ecology/ezhemesyachnyy-informacionnyy-byulleten-o-sostoyanii-okruzhayuschey-sredy/2023 (accessed on June 2023)
7. Bureau of National statistics. Agency for strategic planning and reforms of the Republic of Kazakhstan. https://stat.gov.kz/en/ (accessed on June 2023)
8. Information and legal system of normative legal acts of the Republic of Kazakhstan. water code. Water Code. https://adilet.zan.kz/rus/docs/K030000481_/k030481.htm (accessed on January 2023)
9. Ili (river). https://en.wikipedia.org/wiki/Ili_(river) (accessed on May 2023)
10. IBM. What is a knowledge graph? https://www.ibm.com/topics/knowledge-graph#:~:text=A%20knowledge%20graph%2C%20also%20known,the%20term%20knowledge%20%E2%80%9Cgraph.%E2%80%9D (accessed on June 2023)
11. Horrocks, I.; Patel-Schneider, P.F.; Boley, H.; Tabet, S.; Grosof, B.; Dean, M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Available online: https://www.w3.org/Submission/2004/SUBM-SWRL-20040521/ (accessed on 30 December 2019).
12. Haller, A.; Janowicz, K.; Simon Cox, C.; Le Phuoc, D.; Taylor, K.; Lefrançois, M. Semantic Sensor Network Ontology. Available online: https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/ (accessed on 30 December 2019).
13. Cox, S.; Little, C. Time Ontology in OWL, W3C Proposed Recommendation 7 September 2017. Available online: http://www.w3.org/TR/owl-time (accessed on 12 November 2017).
14. Wang, X.; Wei, H.; Chen, N.; He, X.; Tian, Z. An Observational Process Ontology-Based Modeling Approach for Water Quality Monitoring. Water 2020, 12, 715. https://doi.org/10.3390/w12030715.
15. Juan D.; Rondón D.; Luis M. Vilches-Blázquez, Characterizing water quality datasets through multi-dimensional knowledge graphs: a case study of the Bogota river basin. *Journal of Hydroinformatics* 1 March 2022; 24 (2): 295–314. doi: https://doi.org/10.2166/hydro.2022.070
16. Wang, C.; Chen, N.; Wang, W.; Chen, Z. A Hydrological Sensor Web Ontology Based on the SSN Ontology: A Case Study for a Flood. ISPRS Int. J. Geo-Inf. 2018, 7, 2. https://doi.org/10.3390/ijgi7010002
17. Ritesh, C.; Sadhana, T.; Sonali, A.; Navjot, S. Semantic web-based diagnosis and treatment of vector-borne diseases using SWRL rules. *Knowledge-Based Systems*, 2023, 110645, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2023.110645.
18. Xilang, T.; Mingqing, X.; Bin, H.; Dongqing, P. Exchanging knowledge for test-based diagnosis using OWL Ontologies and SWRL Rules. *Procedia Computer Science*, Volume 131, 2018, Pages 847-854, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.04.279.
19. Bi-Hui, Y.; Wang, H.; Xin-Peng, D.; Xue-Zhen, Zh. Design and implementation of a semantic gateway based on SSN ontology. *Procedia Computer Science*, Volume 183, 2021, Pages 432-439, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.02.081.
20. Ritesh, C.; Sonali, A.; Navjot, S. Semantic sensor network ontology based decision support system for forest fire management. *Ecological Informatics*. Volume 72, 2022, 101821, ISSN 1574-9541. https://doi.org/10.1016/j.ecoinf.2022.101821.

21. Ospan, A.; Mansurova, M.; Kakimzhanov, E.; Aldakulov, B. KazRivDyn: Toolkit for Measuring the Dynamics of Kazakhstan Rivers with a Graphics Based on Google Earth Engine. *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Nur-Sultan, Kazakhstan, 2021, pp. 1-4, doi: 10.1109/SIST50301.2021.9465902.
22. Moiseyenko, T.; Megorskiy, V.; Gashkina, N.; Kudryavceva, L. Impact of water pollution on the health of the population of the industrial region of the north. *Water resources.* Pp. 199-208. https://www.researchgate.net/publication/346787929_VLIANIE_ZAGRAZNENIA_VOD_NA_ZDOROVE_NASELENIA_INDUSTRIALNOGO_REGIONA_SEVERA/link/5fd0bb39a6fdcc697befbe53/download (accessed on January 2010).
23. Ili-Balkhash basin. https://ru.wikipedia.org/wiki/Или-Балхашский_бассейн (in russian) (accessed on June 2022).
24. Espolov, T.; Tleulesova, A.; Zheksembayeva, G. Ile-Balkhash transboundary basin: problematic situation and ways to solve it. Research, results. 2012, Almaty. https://articlekz.com/article/12802#gsc.tab=0 (accessed on November 2012).
25. Barakhnin, V.; Mansurova, M.; Grigorieva, I.; Kozhemyakina, O.; Ospan, A. TableProcessor: The Tool for the Analysis and the Interpretation of Web Tables to Create the Geo Knowledge Base of Kazakhstan. In: et al. *Artificial Intelligence in Models, Methods and Applications.* AIES 2022. Studies in Systems, Decision and Control, vol 457. Springer, Cham. https://doi.org/10.1007/978-3-031-22938-1_15
26. Nugumanova, A.; Apayev, K.; Baiburin, Y.; Mansurova, M.; Ospan, A. Qurma: a table extraction pipeline for knowledge base population. ***Journal of Mathematics, Mechanics and Computer Science***, [S.l.], Volume 114, n. 2, june 2022. ISSN 2617-4871. doi: https://doi.org/10.26577/JMMCS.2022.v114.i2.08. (accessed 25 May 2023)
27. Zhang, Z. Effective and Efficient Semantic Table Interpretation Using TableMiner $^{+}$. *Semantic Web*, Volume. 8, no. 6, pp. 921-957, 2017. 7 August 2017. http://dx.doi.org/10.3233/SW-160242
28. Musen, M.A. The Protégé project: A look back and a look forward. AI Matters. *Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, 1(4), June 2015. DOI: 10.1145/2557001.25757003.
29. **Lamy JB.** Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies**.** ***Artificial Intelligence In Medicine*** **2017**; 80:11-28 DOI: 10.1016/j.artmed.2017.07.002
30. UN 2023. The Sustainable Development Goals. https://www.un.org/sustainabledevelopment/ (accessed on June 2023)
31. Joshi A. et al. A knowledge organization system for the united nations sustainable development goals //The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18. – Springer International Publishing, 2021. – C. 548-564. DOI: 10.1007/978-3-030-77385-4_33.