Russian Journal of Earth Sciences

Special Issue: "Data Science, Geoinformatics and Systems Analysis in Geosciences"

Key Aspects of Information Technology for Maintaining of the Unified State Fund of Data on the State of the Environment, Its Pollution

E. V. Kolesnikov[1], **V. M. Shaimardanov**[1]

[1] All-Russia Research Institute of Hydrometeorological Information, World Data Center (RIHMI-WDC), Obninsk, Russia

* **Correspondence to:** Kolesnikov Evgeny Vladimirovich, ekolesnikov@meteo.ru

Abstract: Information from the Unified State Fund of Data on the State of the Environment and its Pollution (USFD), which consolidates large amounts of data from hydrometeorological observations and pollution monitoring, is one of the key factors affecting the life of modern world society and the stable economic development of every state on the planet Earth. This effect is due to the presence of particularly dangerous hydrometeorological phenomena and adverse weather conditions that can cause social and economic damage. In this regard, the tasks of collecting, accounting, long-term storage and timely provision of such information to consumers of various sectors of economic activity for use in decision-making remain relevant. This article discusses the key aspects of the information technology of the USFD, which is formed as a result of the activities of the Federal Service of Russia for Hydrometeorology and Pollution Monitoring (Roshydromet). As a result of the work, a description of the technological process of maintaining USFD, statistical data on the demand and use of USFD information and a new documentary information search system solving the problem of accounting and long-term storage of USFD information resources are presented. The system was developed using modern information technologies, international standards for describing information resources and taking into account the experience of foreign organizations engaged in activities similar to Roshydromet. The direction of further development of digital technologies for operational public service with information products created on the basis of verified USFD information resources is also outlined.

Keywords: Hydrometeorology, Environmental monitoring, Big data, Long-term storage, Information systems and technologies, Retrieval system, Databases, Metadata, Information resources.

Citation: Kolesnikov, E. V., and V. M. Shaimardanov (2025), Key Aspects of Information Technology for Maintaining of the Unified State Fund of Data on the State of the Environment, Its Pollution, *Russian Journal of Earth Sciences*, 25, ES2007, EDN: RCERJT, https://doi.org/10.2205/2025ES000966

Research Article

Received: 15 November 2024
Accepted: 15 April 2025
Published: 23 May 2025

1. Introduction

Regular monitoring of the state of the natural environment in Russia and the storage of collected information has been carried out for two centuries. At the beginning of the XIX century, many proposals were submitted for the creation of regular hydrometeorological observations in Russia, which formed the basis for the project of organizing the service, curated by a talented scientist and organizer, academician Adolf Yakovlevich Kupfer. In 1848, the activities of the organization headed by Kupfer led to the creation on its basis of the Main Physical Observatory (MPO), known today as the Main Geophysical Observatory named after A.I. Voeikov. It was from this moment that the creation of a network of meteorological observations, the replication of observation materials, the creation of meteorological instrumentation and an instrument verification system began. Today, this function is performed by Roshydromet, as a result of which the USFD is formed. This is

a unique data collection that contains an orderly, constantly updated set of documented information on environmental pollution monitoring, hydrometeorology and related fields (aerology, hydrology, meteorology, oceanology, heliogeophysics, etc.).

Information products created on the basis of USFD data are actively used in almost all major sectors of the economy, for example, in construction, architecture and design, energy, transport and communications, aviation, agriculture, as well as in scientific research and the activities of executive authorities. This once again confirms the relevance of the tasks of collecting, long-term storage and timely operational access to this kind of information. Figure 1 shows statistics on the number of requests and users over the past 13 years.

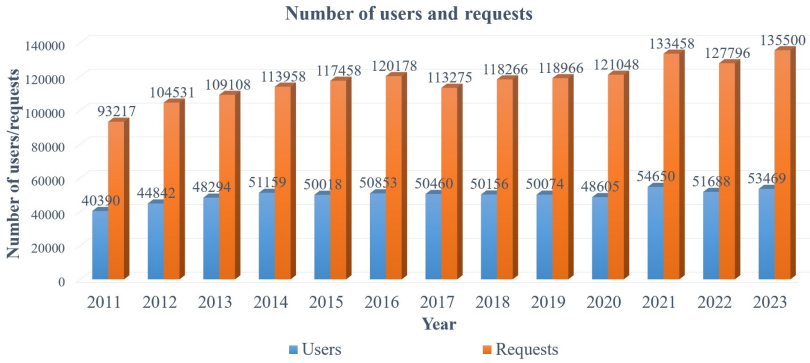

Figure 1. Number of requests and users.

A large proportion of requests, not counting individuals, falls on architecture, design and construction, which is reasonable. Today, no construction is possible without an important component - a set of engineering and survey works, which include engineering and hydrometeorological surveys.

Figure 2 shows the distribution of users by types of primary economic activity for 2023. Specialists engaged in survey work must study the components of the surrounding natural environment at the construction site: water bodies, natural and climatic conditions, land resources, atmospheric air, soil cover, etc. It is especially important to have information about particularly dangerous hydrometeorological processes and phenomena. Nature itself, due to constant changes in environmental conditions, regularly confirms the relevance of survey work and forces a person to conduct such research for the construction of objects of any level of complexity, be it a house, a bridge, a pipeline or a power transmission line pole.

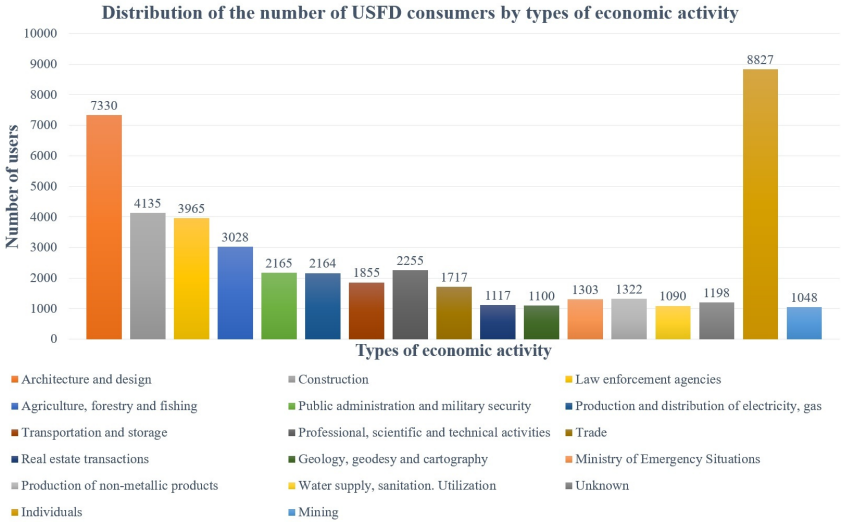

Figure 2. Distribution of users by types of primary economic activity for 2023.

The tasks of accounting and long-term storage of information resources are not new in the field of information systems and technologies. However, the application of a particular

technological solution largely depends on the subject area, namely, on the features of its information resources – storage objects, the management of which must be implemented. In the context of the USFD, not many works have been published on these tasks at the moment. This is due to the specificity of the subject area. Attempts to solve the problems of long-term storage and accounting of information resources of the USFD are presented in works [*Dolgih et al.*, 2014a,b]. However, the project was not fully implemented.

Thus, the aim of the work is to describe the information process of the technology of maintaining the USFD, to note the relevance of the tasks of collection and long-term storage, to present a new documentary information retrieval system of accounting and long-term storage, taking into account the experience of domestic and foreign colleagues in solving these problems.

2. Technology for maintaining USFD

Figure 3 shows the general structural diagram of the technological process of maintaining the Unified State Hydrometeorological Data. The structure of Roshydromet, which participates in the process of maintaining the Unified State Hydrometeorological Data, can be divided into several logical levels.

At the first level (local), there is the State Observation Network of Observation Units, which collects and processes primary observation data for the following types: meteorological, aerological, agrometeorological, synoptic, hydrological, marine hydrometeorological, geophysical, heliogeophysical, pollution monitoring, satellite.

At the second level (regional) are the Territorial Administration for Hydrometeorological and Environmental Monitoring (ROHEM), which perform tasks on processing and accumulating primary observation data and ensuring the functioning of their observation units. Scientific research institutions are also located at this level. The subject of the activities of the research institutions of Roshydromet is scientific research in the field of hydrometeorology, related areas and pollution monitoring. Each research institute is assigned specific types of observations in the context of which measurement methods and tools are being developed and improved, as well as methodological and software for processing and accumulating primary observational data.

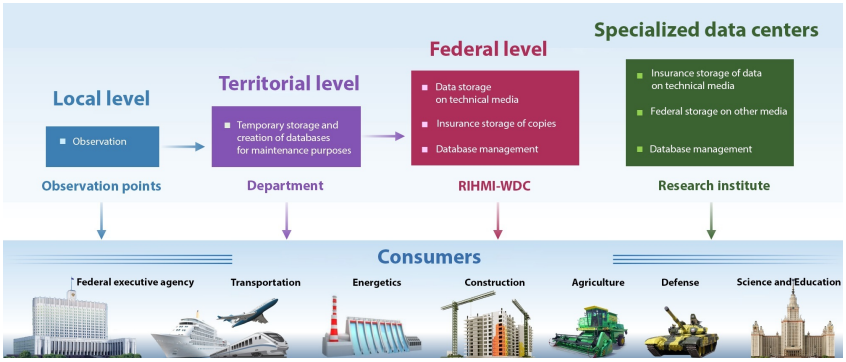

Figure 3. Structural diagram of maintaining the USFD.

To solve the urgent problems of centralized accounting and long-term storage of the USFD, the All-Russian Research Institute of Hydrometeorological Information – World Data Center (FSBI "RIHMI-WDC") was created in the city of Obninsk [*History…*, 2024].

3. Composition and structure of the USFD

Today, the USFD includes information resources included in the list of governance document (GD) 52.19.143 "List of documents of the archival data fund on the state of the environment, its pollution", which determines the terms and place of storage of each information resource. Figure 4 shows the distribution of the volume of information in the USFD by discipline on paper and electronic media. As of January 1, 2024, the USFD is located on the following types of media:

- Photo media (microfilms) – 857,653 storage units.
- Paper media – 2,840,289 storage units (of which 2,710,579 units are of permanent storage period, 129,710 units are of temporary storage period).
- Electronic media (optical disks, cartridges, magnetic tapes, tape libraries – with a total volume of information of 2.8 TB of verified series of primary observation data and 30 TB of primary observation data.

It is also worth noting that paper documents undergo a digitization stage using scanning. At the moment, about 500 thousand documents of the USFD have been scanned, with a total volume of about 150 TB. This solution is described in the work [*Kolesnikov et al.*, 2021; 2022].

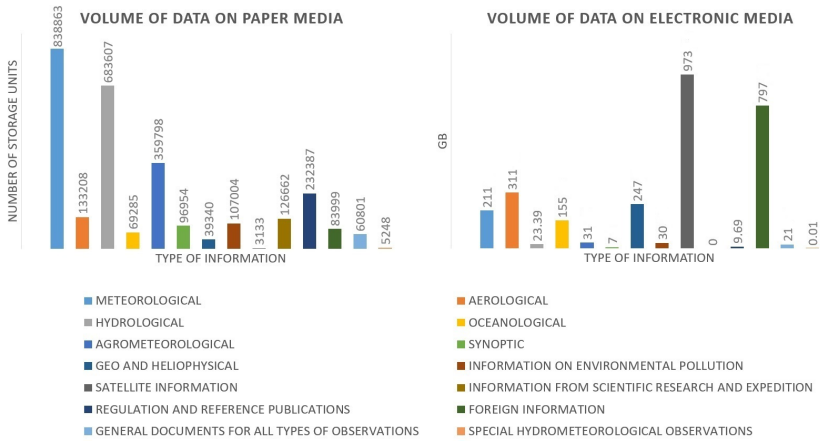

Figure 4. Distribution of the volume of information in the USFD by discipline on paper and electronic media.

The collection of USFD data consists of about 50 million files with observational data and is unique. Its information component meets the criteria of Big Data. We are talking about both the 3V and 5V criteria. The technological component of maintaining the USFD began to form in the early 1970s of the last century, long before the concept of Big Data appeared. The foundations of the "computer infrastructure" for meteorological and climate research using observation data were laid – from a modern perspective, it can be called "Seamless Technology". This largely anticipated those areas of development of RIHMI-WDC that we today attribute to Big Data. Figure 5 shows the corresponding IT infrastructure, which ensures a stable ability to perform the task of maintaining the USFD.

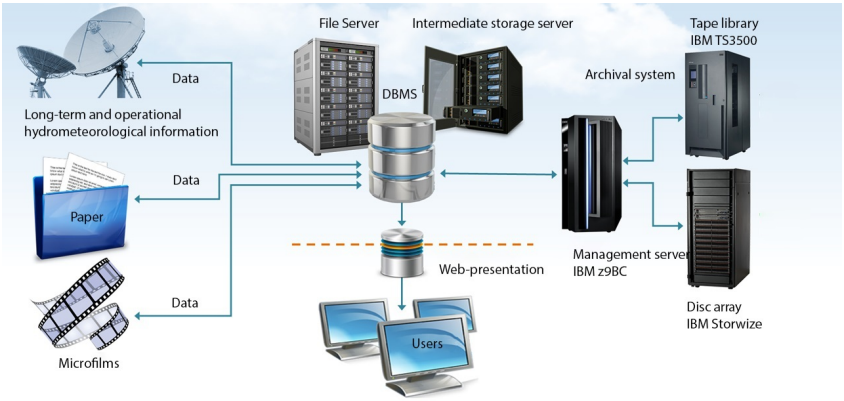

Figure 5. IT infrastructure for maintaining the USFD.

To store a large volume of information resources, the USFD uses a hardware and software complex based on tape libraries [*Shaimardanov et al.*, 2014]. Storage systems based on magnetic tape technology are currently the most reliable means for ensuring long-term storage of digital information.

4. The problem of accounting and long-term storage of the USFD

The main problem of accounting and searching for information resources of the USFD is that the USFD information resources are documents that contain factual information about the state of the environment. In this case, the search for the necessary information resource is reduced to a visual review of the catalog structure of documents. Having analyzed the works [*Dolgih et al.*, 2014a,b], in which attempts were made to solve the problems of accounting and searching, as well as the experience of foreign colleagues (public portals of foreign electronic archives [*Databases…*, 2024; *National Meteorological Library*, 2024; *Weather…*, 2024]), we can conclude that the solution to the problem is reduced to the creation of an information system based on an information model of metadata that allows describing each information resource as a unique storage object, similar to the accounting of books in a library.

The USFD information refers to geographic (spatial) information. However, the object of the study is not the data on the state of the environment itself, but the documents that contain them. Therefore, in order to create a metadata model, it is necessary to consider the standards of librarianship and publishing. Such materials include the Dublin Core specification (DC), which is widely used both in domestic and international practice. The corresponding international standards have been developed on its basis: ISO 15836-1:2017, ISO 15836-2:2019, GOST R 7.0.10-2019, GOST R ISO 15836-2-2022.

5. Results

The analysis of information systems of foreign organizations, similar to Roshydromet, showed that all systems belong to the class of documentary ones and in all systems for searching information resources there is a trend of using attribute search.

The key result of this work is the presentation of the metadata information model, which allows identifying each USFD information resource as a unique storage unit. This model is the basis for the information retrieval system for accounting of information resources of the USFD. Table 1 shows the attribute composition of the metadata model (47 elements), formed on the basis of the previously presented standards of the Dublin Core specification.

Table 1. Attributive composition of the model of metadata of information resources of the USFD.

GROUP OF IDENTIFICATION ELEMENTS		
No.	Field	Discription
1	Title	Document name
2	Alternative Title	Alternative document name
3	Subject	Subject (aerology, hydrology, meteorology, etc.)
4	View of observation	Type of observations
5	Collection	Article number in RD 52.19.143
6	Array	The array that contains the document
7	Inventory book	The inventory book in which the document is entered
8	Inventory number	Inventory number of the document
9	Identifier	Unique Composite Index: Subject_View of observation_Collection_ Array_ Inventory number
GROUP OF ELEMENTS OF MAIN PROPERTIES		
10	Volume	Document size
11	Resource type	The class or type to which the document belongs

Continued on next page

Table 1. Attributive composition of the model of metadata of information resources of the USFD. (Continued)

12	File format	Document file format
13	Access rights	Information about who can access the document
14	Frequency	Frequency with which the array is replenished with a document
15	Link	Link to the document in the file system
16	Media type	Physical medium of the document
17	Creator	Entity responsible for creating the document
18	Publisher	Entity responsible for providing the document
19	Co-executor	Entity responsible for contributing to the creation of the document
	PHYSICAL PLACEMENT OF THE DOCUMENT	
20	Location	Physical location of original document
21	Archival Repository Number	Archive storage number
22	Stack Number	Rack/cabinet number
23	Section Number	Section number
24	Shelf Number	Shelf number
	GROUP OF TIME CHARACTERISTICS ELEMENTS	
25	Date created	Document creation date
26	Date of publication	Document publication date
27	Date accepted	Document adoption date
28	Date copyrighted	Copyright date
29	Date submitted	Document submission date
30	Date modified	Modification date
31	Date available	Date the document was or will be available
32	Date valid	Date (often a range of dates) of the document's validity
	GROUP OF ELEMENTS OF CONNECTION WITH OTHER OBJECTS	
33	Has part	A related document that is physically or logically included in the document being described
34	Has version	A related document that is a version, revision, or adaptation of the document being described
35	Has format	A related document that is substantially the same as an existing document being described but is in a different format
36	Requires	A related document that is required by the document being described to support its operation, presentation, or coordination
37	Replaces	A related document that is superseded, replaced, or continued by the document being described
38	Is part of	A related document that is physically or logically included in the document being described
39	Is version of	A related document of which the document being described is a version, revision, or adaptation
40	Is format of	An existing related document that is substantially the same as the document being described but is in a different format

Continued on next page

Table 1. Attributive composition of the model of metadata of information resources of the USFD. (Continued)

41	Is required by	A related document that requires the document being described to support its operation, presentation, or coordination
42	Is replaced by	A related document that supersedes, replaces, or is a continuation of the document being described
	GROUP OF ELEMENTS OF THE CONTENT OF OBJECTS	
43	Start date of observations	Environmental monitoring start date
44	End date of observations	Environmental monitoring end date
45	Observation area	Observation area
46	Language	Document language
47	Annotation	Document description

6. Discussion

The information retrieval system, which has been implemented into trial operation, provides accounting, operational remote search and provision of information resources of the Unified State Register of Documents. However, for the class of users who are not familiar with the composition and structure of the Unified State Register of Documents, problems may arise in working with documents. In this case, the user needs to visually search for actual observation data in the provided information resource. This significantly slows down the process of obtaining data. Therefore, the further development of information technologies for operational customer service consists in the development of new digital services that will contain the actual data of the EGFD. In other words, each information resource of the Unified State Register of Documents will be transformed into a relational database structure that will allow creating search queries to the data. Work in this direction is already underway and at the first stage a prototype of specialized software has been developed [*Peretyatko*, 2022; 2023], which will form the basis of the new factographic information-retrieval system.

7. Conclusions

This paper presented the information process of the technology for maintaining the Unified State Register of Documents, its composition and structure, and noted the relevance of the tasks of collecting and long-term storage of Unified State Register of Documents information in view of its demand for decision-making in various sectors of economic activity. As a solution to these problems, a new documentary information retrieval system for accounting and long-term storage was presented, taking into account the features of the information resources of the Unified State Register of Documents. It was also announced that the technology of operational maintenance with actual observation data would be further developed.

References

Databases. Publication of various survey and observation data and so forth. — 2024. — URL: https://www.jamstec.go.jp/e/database (visited on 10/14/2024).

Dolgih S. G., Shaimardanov V. M. Creating a data management technology in automated archiving system // Proceedings of the Voeikov Main Geophysical Observatory. — 2014a. — No. 572. — P. 162–174. — EDN: SZXXSH. — (In Russian).

Dolgih S. G., Shaimardanov V. M. The development of technology of data management in an automated archival system // Proceedings of the Voeikov Main Geophysical Observatory. — 2014b. — No. 575. — P. 204–214. — EDN: TIQKSX. — (In Russian).

History of the Federal State Budgetary Institution "All-Russian Research Institute of Hydrometeorological Information - World Data Center". — 2024. — URL: http://meteo.ru/about/history (visited on 10/14/2024). — (In Russian).

Kolesnikov E. V., Shaimardanov V. M. Electronic fund of scanned hydrometeorological documents. Database RU 2021621423. — 2021. — EDN: CGZKFV. — (In Russian).

Kolesnikov E. V., Shaimardanov V. M. Development of an Automated Paper Documents Scanning Technology for the Russian Meteorological Service // Russian Journal of Cybernetics. — 2022. — 1(9). — P. 28–38. — DOI: 10.51790/2712-9942-2022-3-1-4. — (In Russian).

National Meteorological Library. — 2024. — URL: http://www.bom.gov.au/library (visited on 10/14/2024).

Peretyatko L. O. The National Hydrometeorological Archive as Big Data. Appropriate Technologies and Tools // Russian Journal of Cybernetics. — 2022. — Vol. 3, no. 4. — P. 98–101. — DOI: 10.51790/2712-9942-2022-3-4-11. — (In Russian).

Peretyatko L. O. The National Hydrometeorological Archive as Big Data. A Universal Parser of the Hydrometeorological Data Description Language // Russian Journal of Cybernetics. — 2023. — Vol. 4, no. 2. — P. 47–52. — DOI: 10.51790/2712-9942-2023-4-2-07. — (In Russian).

Shaimardanov V. M., Shaimardanov M. Z. Development of Roshydromet automated archived data system // Uchenye zapiski RGGMU. — 2014. — No. 36. — P. 60–66. — EDN: TQUBAD. — (In Russian).

Weather and climate change. — 2024. — URL: https://www.metoffice.gov.uk (visited on 10/14/2024).