DEVELOPMENT OF THE WDS RUSSIAN-UKRAINIAN SEGMENT

Marsel Shaimardanov[1*], Alexei Gvishiani[2], Michael Zgurovsky[3], Alexander Sterin[4], Alexander Kuznetsov[5], Natalia Sergeyeva[6], Evgeny Kharin[7], and Kostiantyn Yefremov[8]*

*[*1]All-Russian Scientific and Research Institute of Hydrometeorological Information – World Data Center, WDC for Meteorology, 6, Koroleva St., Obninsk, Kaluga Region, 249035 Russian Federation*
Email: marsel@meteo.ru
[2]Geophysical Centre of Russian Academy of Science, 3, Molodezhnaya St., 119296 Moscow, Russian Federation
Email: a.gvishiani@gcras.ru
[3]National Technical University of Ukraine "Kyiv Polytechnic Institute", 37, Peremohy ave., 03056 Kyiv, Ukraine
Email: zgur@zgurov@kiev.ua
[4]All-Russian Scientific and Research Institute of Hydrometeorological Information – World Data Center, WDC for Rockets and Satellites and Rotation of the Earth, 6, Koroleva St., Obninsk, Kaluga Region, 249035 Russian Federation
Email: sterin@meteo.ru
[5]All-Russian Scientific and Research Institute of Hydrometeorological Information – World Data Center, WDC for Oceanography, 6, Koroleva St., Obninsk, Kaluga Region, 249035 Russian Federation
Email: kuznet@meteo.ru
[6]Geophysical Centre of Russian Academy of Science, WDC for Solid Earth Physics, 3, Molodezhnaya St., 119296 Moscow, Russian Federation
Email: nata@wdcb.ru
[7]Geophysical Centre of Russian Academy of Science, WDC for Solar-Terrestrial Physics, 3, Molodezhnaya St., 119296 Moscow, Russian Federation
Email: kharin@wdcb.ru
**[8]National Technical University of Ukraine "Kyiv Polytechnic Institute", WDC for Geoinformatics and Sustainable Development, 37, Peremohy ave., 03056 Kyiv, Ukraine*
Email: k.yefremov@wdc.org.ua

ABSTRACT

Establishment of the Russian-Ukrainian WDS Segment and its state of the art, main priorities and research activities are described. One of the high priority tasks for Segment members is development of a common information space - transition from Legacy Systems and individual services to a common, globally interoperable, distributed data system that incorporates emerging technologies and new scientific data activities. The new system will build on the potential and added value offered by advanced interconnections between data management and data processing components for disciplinary and multidisciplinary applications. Thus, the principles of the architectural organization of intelligent data processing systems are discussed in this paper.

Keywords: World Data System, ICSU, Intelligent data processing, Interdisciplinary research, Heterogeneous data sources

1 INTRODUCTION

Modern scientific research in concordance with the search for answers to global challenges arising in the beginning of the twenty-first century is interdisciplinary and focused on solving badly structured tasks. An example of such research is the analysis of sustainable development processes in a global and regional context (Zgurovsky, Stratukha, Melnichenko, Voitko, .Boldak, Yefremov, et al., 2010).

When we talk about interdisciplinary research (Somervill & Rapport, 2000), we mean the usage of data that is quantitative or qualitative assessments that characterize different phenomena or objects. On the basis of this data and models developed in different scientific spheres, the generalized (interdisciplinary) models of a complete presentation of an object of research are developed. As a rule in terms of such research, formal models are

absent, but at the same time, it is possible to use the results of objective measurements. Notably the tasks of such research can be described as badly structured (Newell & Simon, 1972) when deciding which of the results of the objective measurements and subjective expert assessment are to be used.

To deal with such tasks, the methods of scientific calculations (Yang, 2008) that are the essence of intelligent data processing concepts (Yang, 2008) are used more and more often. These concepts consist of organizing the detection process into "raw" data of unknown, nontrivial, practical and useful knowledge that can be interpreted and may be useful for decision making in different spheres of human activity.

It is clear that the concept of intelligent data processing depends not only on the usage of special program tools but also on a special information-communicative infrastructure. This allows the use of huge volumes of data of different origins and processing in the search for solutions of interdisciplinary questions.

To harmonize the actions for the establishment of such infrastructure and to organize a common information space to maintain the acquisition, handling, and exchange of data and solve the fundamental and applied interdisciplinary problems, in 2008 the Russian and Ukrainian World Data Centers united into a Segment (Zgurovsky, Gvishiani, Yefremov, & Pasichny, 2010).

The aim of this paper is to describe the state-of-the-art Russian-Ukrainian WDS Segment and to examine the principles of the architectural organization of an intelligent data processing system on the basis of which the components of an open program system can be created and integrated. This is a system that gives the user the data and tools to solve the challenges of interdisciplinary research.

2 STATE OF THE ART

For more than 50 years, the five Russian ICSU World Data Centers (WDC) for Oceanography, Meteorology, Rockets, Satellites and Rotation of the Earth, Solar-Terrestrial Physics and Solid Earth Physics have collected, analyzed, archived, and distributed data from a broad spectrum of observatory types. The centers provide open and convenient access to large volumes of data, permanently increasing the information resources on the Internet. In 2006, the Ukrainian WDC for Geoinformatics and Sustainable Development was formed. It is one of the leaders in various fields of sustainable development research in Ukraine, and it collects, processes, analyses, and disseminates global and national data necessary for sustainable development research.

Several meetings of the ICSU World Data Centers in Russia and Ukraine were held in Obninsk, Moscow, and Kyiv in 2008-2010. This activity has resulted in the establishment of the Scientific Council of Russian and Ukrainian World Data Centers and the formation of the Russian-Ukrainian segment of WDCs (Segment).

The top priorities for our Segment are the following:

- Integrating with new WDSs and effective cooperation with WDS members
- Providing safety of data from non-digital data carriers and its digitization
- Establishing a data quality policy
- Developing a data providers' infrastructure
- Forming a common information space for the Russian-Ukrainian WDS segment

The creation of a common information space for our segment is a high priority task. It is important to provide unified data formats and develop unified tools for efficient data exchange. Such an approach would make it possible to create a single access point to all data and services of the Russian-Ukrainian WDS Segment. That would also provide a flexible framework for development of a unified data processing toolkit. The main features of this common information space are:

- Flexible and scalable cross-platform, open source-based architecture (e.g., SOA)
- Centralized data and services registry
- Easy integration with existing systems (using SOAP, WSDL, UDDI, etc.)
- Single access point
- Easily created and customized UI, based on existing services
- Common approach for acquiring data from various data sources

All segment members expressed interest in joining the new ICSU World Data System and have successfully completed the necessary stages of certification. Today all Russian and Ukrainian WDCs have obtained the status of regular WDS members.

2.1 WDCs in Obninsk

There are three World Data Centers established on the basis of the All-Russian Scientific and Research Institute of Hydrometeorological Information – World Data Center (RIHMI-WDC):

- World Data Center for Meteorology;
- World Data Center for Oceanography;
- World Data Center for Rockets, Satellites, and Rotation of the Earth.

The WDC for Meteorology has a commitment to provide long term secure preservation and dissemination of meteorological data and products on both global and regional scales. The data holding contains observed meteorological data for the period from the eighteenth century to the present. Data are collected from the Russian network of meteorological stations and from other parts of the globe by means of the WMO's telecommunication system. Data are validated, checked, and updated continuously. Specialized high quality data sets for climate study are of particular importance and are provided for online access (http://meteo.ru/english/climate/cl_data). Climate surveys are published on a regular basis (http://meteo.ru/english/climate/bulletin/).

The WDC for oceanography collects data and products of national and international projects in the field of physical and chemical oceanography. The data holding contains observed oceanographic data collected from single observational platforms (research vessels, buoys, and other devices) and coastal stations for the period from the nineteenth century to the present. The WDC also provides additional products, including data analyses, maps of data distributions, and data summaries. Online access to the data and products is made available through http://www.meteo.ru/mcd/ewdcoce.html.

The WDC for rockets, satellites, and rotation of the Earth collects meteorological data of national and international rockets and satellites and data on Earth rotation. Necessary quality assessment and quality control procedures are applied when required. The data holding contains observed rocket and satellite meteorological data for the period from 1966 to the present. Online access to the data and products is provided through the entry point http://www.meteo.ru/mcd/ewdcroc.html.

Internal data management is based on the original data description language (DDL) and hierarchical DBMS AISORI developed at the RIHMI-WDC. This ensures data and metadata identity, integrity, and efficient archiving and usage.

All data and information submitted to the WDCs are classified and registered in catalogues and directories of the metadata base. The metadata elements enable identification of the type and origin of the data, their spatial and temporal coverage, and other characteristics necessary to ensure the authenticity of the data sets. The metadata are used to prepare published and electronic data catalogues. These are posted on the RIHMI-WDC website (http://meteo.ru/english/data_b/).

All data are checked and validated by means of a set of visual and automatic QC procedures in accordance with WMO and IOC Manuals and Guides. The numerical criteria used in QC procedures are being updated on the base of continuous climatological research. The enhanced procedure for duplicates check has been developed and applied to avoid data duplicates within global data sets. Also, to ensure metadata integrity, the directories of organizations, maritime research projects, and research vessels are regularly validated to be consistent with international counterpart directories of IODE, GCMD, ICES, WMO, EDMO, and EDMERP.

To achieve a high rate of data processing and provide effective online user access to the data, the data are loaded into a relational Oracle DBMS, which is a component of integrated web-technology. This final step enables the user to discover the data of his/her particular interest and to retrieve, browse, and view the data in table form or as a plot, map, or diagram.

To place on the web metadata consistent with ISO 19115 standards, there are two international metadata profiles used by the RIHMI-WDC: the WMO core profile for meteorological data (http://www.wmo.int/pages/prog/www/WDM/Metadata/documents.html) and the CDI for oceanographic data (http://www.seadatanet.org/Standards-Software/Metadata-formats). The WDCs also strictly adhere to international standards for data exchange, in particular the BUFR/CREX data format developed under the WMO umbrella and the widely used NetCDF and ODV data formats.

The collection and dissemination of international publications (atlases, gazetteers, reference books, manuals, and guidelines) are also the responsibility of the world data centers hosted by RIHMI-WDC. Catalogs of international publications in the fields of meteorology, oceanography, rockets, and satellites collected by WDC are available online (http://meteo.ru/english/publish/).

Each world data center uses the RIHMI-WDC technical infrastructure and software utilities for long-term data preservation and dissemination. All procedures used in the RIHMI-WDC for long-term data storage (holdings with appropriate conditions, periodic checks, and recovery as necessary) are applied to the WDCs data and information along with national data and information.

To perform the processing functions necessary for data archiving and provision of user access to the data, the IBM z9 BC mainframe is employed. The IBM System Storage DS8300 offers high performance, higher capacity storage up to 512 Tb. Full Disk Encryption with local key management providing exhaustive data security. Also the firewall system is maintained to secure user access via the Internet to the operational data base and other information resources.

To ensure long-term and safe storage, all data are archived within two robotic IBM System Storage 3500 Tape Libraries. One of these is the main library, and another one is a mirror library used for data backup and recovery. The libraries are located in two detached buildings. Direct access by an external user to the library is impossible.

A set of information service systems provide data stewardship and preservation. In particular, a technological schema of backup and long-term storage is based on IBM Tivoli Storage Manager software and Content Manager on Demand. This solution provides cost-effective functionality, scalability, and ease of use for the entry-level storage user.

All of these developments serve to ensure long-term data preservation and free timely access to the WDC's data.

2.2 WDCs in Moscow

The WDC for Solar-Terrestrial Physics and the WDC for Solid Earth Physics, Moscow make up the parts of the Geophysical Center of the Russian Academy of Sciences (GC RAS).

The WDC for Solar-Terrestrial Physics (WDC for STP) was one of the original data centers established in the USSR by the Academy of Sciences of the USSR in 1957 to support the IGY. The WDC for STP holds data sets relating to solar activity and interplanetary phenomena, ionosphere, geomagnetic variations, and cosmic rays. The center maintains and provides services for the archive of historical and modern results of geophysical observations on global networks of observatories. The data are available in different traditional forms, e.g., paper, microfilms and microfiches, and electronic formats. The WDC for STP converts old analog data into digital form in order to preserve the data for scientists in convenient ready-to-use form. All data are registered in the computer database and listed in the data availability catalog. Digital data in non-standard formats, metadata, and data availability catalogs are available at free access on the WDC for STP's website: http://www.wdcb.ru/stp/index.en.html. The World Data Center for Solar-Terrestrial Physics together with Laboratory of Network Information Technologies of GC RAS and National Geophysical Data Center of USA support the databases in the "Space Physics Interactive Data Resource" (SPIDR). SPIDR is a distributed network of databases and service programs that are synchronized in real time and allow the user concurrent access to a network of thematic databases: interactive visualization of time series, maps, and images and sampling of the multidisciplinary data and the search for specific events in "space weather" in terms of a natural language with fuzzy logic application. The WDC for STP is a participant in INTERMAGNET, an international project in the field of information technologies and geophysics.

The WDC for Solid Earth Physics (WDC for SEP) maintains extensive archives of data on seismology, geomagnetism (the main magnetic field), archeo- and paleomagnetism, gravimetry, geothermy, and recent movements. Data stored in the center were obtained during the International Geophysical Year (1957-1958) and subsequent international projects, such as the "Upper Mantle", "Geodynamics Project", "International Polar Year 2007-2008", et al. They are the results of land and sea expeditions, launches of satellites, special experiments, and results of geophysical observations (seismological, geomagnetic, etc.) on global networks of observatories. The center accepts data according to a long term relationship with the producers of the data: separate stationary observatories, regional, national, and international observatory networks' data-processing and analytical centers. The data provided by scientific institutes and other organizations are accepted only after they have passed examination in these organizations and have received the status of data intended for international exchange by decision of a commission of experts. The center provides access to all these data and also serves as an information and reference center. Some parts of the data are stored as publications on paper

and microfilm, but a considerable part is also available in digital electronic form on the WDC for SEP's website http://www.wdcb.ru/sep/. Data, metadata, thematic databases, and inventory catalogs are available on line on the website. The WDC for SEP is a participant in InterMARGINS, an international and interdisciplinary project in the field of geology and geophysics, concerned with all aspects of continental margin research.

The WDC for STP and WDC for SEP give data, metadata, and other products to scientists all over the world without restrictions and are free-of-charge. The centers provide all conditions for long-term secure preservation and dissemination of these data. The centers continuously improve implementation of new technologies of data maintenance, software, and hardware. They aspire to have a modern level of technology for collection, handling, transmission, and storage of data and information and to consider new scientific requirements. The centers have modern technical facilities, use modern technologies for data processing, and provide permanent online access to data. The user interface is developed for convenient searching, browsing, visualization, and retrieval of the data on the WDCs websites.

All data received by the WDCs for STP and SEP are analyzed: the discipline, type of observations, period of observations, geographical territory to which the observations are related, etc. are defined. Data are registered and are placed in an appropriate section of the archive. The archive is structured. Reserve copies are made for electronic data. Data analysis and data quality control are carried out with special QC software. Descriptions of data, data formats, and metadata of GCMD DIF standards are prepared.

2.3 WDC in Kyiv

The World Data Center for Geoinformatics and Sustainable Development (WDC-Ukraine) is situated in the Institute for Applied System Analysis (IASA) of the National Academy of Sciences (NAS) of Ukraine and Ministry of Education and Science, Youth and Sport (MESYS) of Ukraine in the structure of the National Technical University of Ukraine the "Kyiv Polytechnic Institute" (NTUU "KPI"), Kyiv, Ukraine. The WDC-Ukraine was created on April 3, 2006, by the decision of the Presidium of the National Academy of Sciences of Ukraine (NAS), the Ministry of Education and Science of Ukraine, and the Geophysical Center of the Russian Academy of Sciences (GC RAS) as a subdivision of the Russian World Data Centers "WDC for Solar-Terrestrial Physics" and "WDC for Solid Earth Physics" (Moscow). In 2008, the Ukrainian branch of the WDC obtained independent status as the World Data Center for Geoinformatics and Sustainable Development.

The World Data Center in Ukraine affords access to the global information resources of ICSU on Earth sciences, planetary and space physics, and related subjects for the Ukrainian scientific community, provides acquisition and storage of national scientific data on the above disciplines, and presents the data to the world community. The WDC-Ukraine collects, processes, analyses, and disseminates global and national data necessary for sustainable development research. The center is one of the leaders in various fields of sustainable development research in Ukraine. The scientific and technical staff of the WDC-Ukraine performs fundamental and applied research and analysis for solving interdisciplinary problems with a systemic nature, particularly quantitative measurement and modeling of sustainable development processes and evaluation of the impact of global threats to sustainability in global and regional contexts. It also develops and implements information technologies for solving a wide range of tasks connected with the collection, exchange, processing, and analysis of interdisciplinary data and solving different tasks of the applied system analysis.

A deep study of World Data Centers in other countries, the interdisciplinary orientation of the IASA, and a systematic approach have allowed the establishment of the WDC-Ukraine as a unique (for the World Data System) network model of a unified interdisciplinary national data center. According to this model, each research area is supervised by one or several scientific organizations of the National Academy of Sciences of Ukraine. Here are some of them:

- Institute for Applied Systems Analysis NAS of Ukraine and MESYS of Ukraine (system coordination of interdisciplinary data, sustainable development);
- S. I. Subbotin Institute of Geophysics NAS of Ukraine (data on seismology, gravimetry, heat flow, archeo- and paleomagnetism, and magnetic measurements);
- Scientific Center for Aerospace Research of the Earth, Institute of Geosciences NAS of Ukraine (aerospace pictures to be used in geology, ecology, agriculture, forestry, and water industry to predict risks of natural and technogenic processes, global environmental changes, and catastrophic processes);
- Main Astronomical Observatory NAS of Ukraine (space geodesy and geodynamics; cosmic rays);
- Marine Hydrophysical Institute NAS of Ukraine (oceanology and hydrometeorology);
- Institute of Geography NAS of Ukraine (cartography);

- Chornobyl Center for Nuclear Safety, Radioactive Waste, and Radioecology.

The network model was first presented on October 7, 2009, at the special WDC session "Emerging Technologies and Opportunities for Global Data Management and Exchange" within the framework of the CODATA-2008 conference (October 5–8, 2008, Kyiv, Ukraine), where it was approved and named the "Network of Networks" (Starostenko, Yatskiv, Lyalko, Ivanov, Rudenko, & Yefremov, 2008). At a session of the WDS Scientific Committee held on October 13–14, 2009, in Paris, this model was taken as a sample for other WDCs.

In Ukraine, such an approach, on the one hand, allows efficient use of the technological capabilities of the Ukrainian Research and Academic Network (URAN), which unites Ukrainian scientific and educational organizations and the high-performance computing cluster of the NTUU "KPI" (this cluster is a part of the national GRID-infrastructure so that in case of need, tasks can be distributed among all partners of this network) and, on the other hand, focuses the efforts of the WDC staff on solving interdisciplinary systemic problems important for all the WDC partners:

- Provision of all main phases of data management (collection, quality assurance, storage, processing, sharing, reporting, and long-term stewardship) for scientific data of various natures;
- Development of mathematical models, methods and tools for assessment, and decision-making in complex systems;
- Development and support of information systems and services focused on data analysis and processing.

3 JOINT PROJECTS

To realize the aims of the segment, centers fulfill a number of joint projects aimed at the development of the World Data System and its Russian-Ukrainian Segment. These projects are supported by Basic Research Foundations and Academies of Sciences from both countries:

- 2008 – 2009: Development of a set of databases and processing algorithms aimed at systematically previewing complex anthropogenic and natural systems' behavior;
- 2009 – 2010: Development of a World Data Centers' network for investigation of the basics of complex natural and anthropogenic systems' global modeling;
- 2010 – 2011: Development of fundamental analytical methods of multidisciplinary data for creation of an integrated access system to information resources of the World Data Centers in Russia and Ukraine;
- 2011 – 2012: Development of indices and indicators of Ukrainian and Russian regions sustainable development based on combined usage of causality and stochastic semantics;
- 2012 – 2013: Development of a general approach and methods for systematic adjustment of data of various natures in the distributed multidisciplinary databases infrastructure of the World Data System Russian-Ukrainian Segment for solving fundamental interdisciplinary tasks of process correlation in the geospheres system.

According to the basic principles of WDS development, one of which is a transition from existing stand-alone WDCs and individual services to a common globally interoperable distributed data system, some projects were targeted to the creation of the Common Data Catalogue (single access point) that gave access to heterogeneous data sources using agent-oriented and ontology-based approaches. A pilot version of the catalogue was developed and given access for testing purposes via the Internet. Such an approach will allow uniting a practically unlimited quantity of diverse data sources into a single heterogeneous environment that will be transparent to the user and organize an intelligent selection of the data sets according to user queries. Now we continue working on the implementation of the services in this system to organize intelligent data processing with the use of an adaptive approach. The user can formulate his or her query in the subset of the natural language.

Besides the development of the fundamental basis and methods for interdisciplinary data analysis and integration of access to information resources of the Russian and Ukrainian WDCs, there are also projects aimed at the development of an intelligence GIS "Russia-Ukraine" for the support of fundamental and applied interdisciplinary studies of complex systems of different natures and on the creation of an interregional information node for collecting and processing the data from the Russian-Ukrainian segment of INTERMAGNET.

To present the WDS on the Internet and provide an access to the data and metadata of the WDS, a new website was designed and developed by the World Data Center for Geoinformatics and Sustainable Development. It is located at www.icsu-wds.org and supported by the WDC-Ukraine staff. This web portal is constructed on the basis of a flexible and easily scaled platform allowing, if necessary, easy and fast revision not only of the material but also of the portal structure. The site has its proper member zone, organized by integrating MediaWiki and GoogleDocs, in the framework of which members of WDS-SC actively exchange materials.

4 COMMON INTELLIGENT DATA PROCESSING SYSTEM

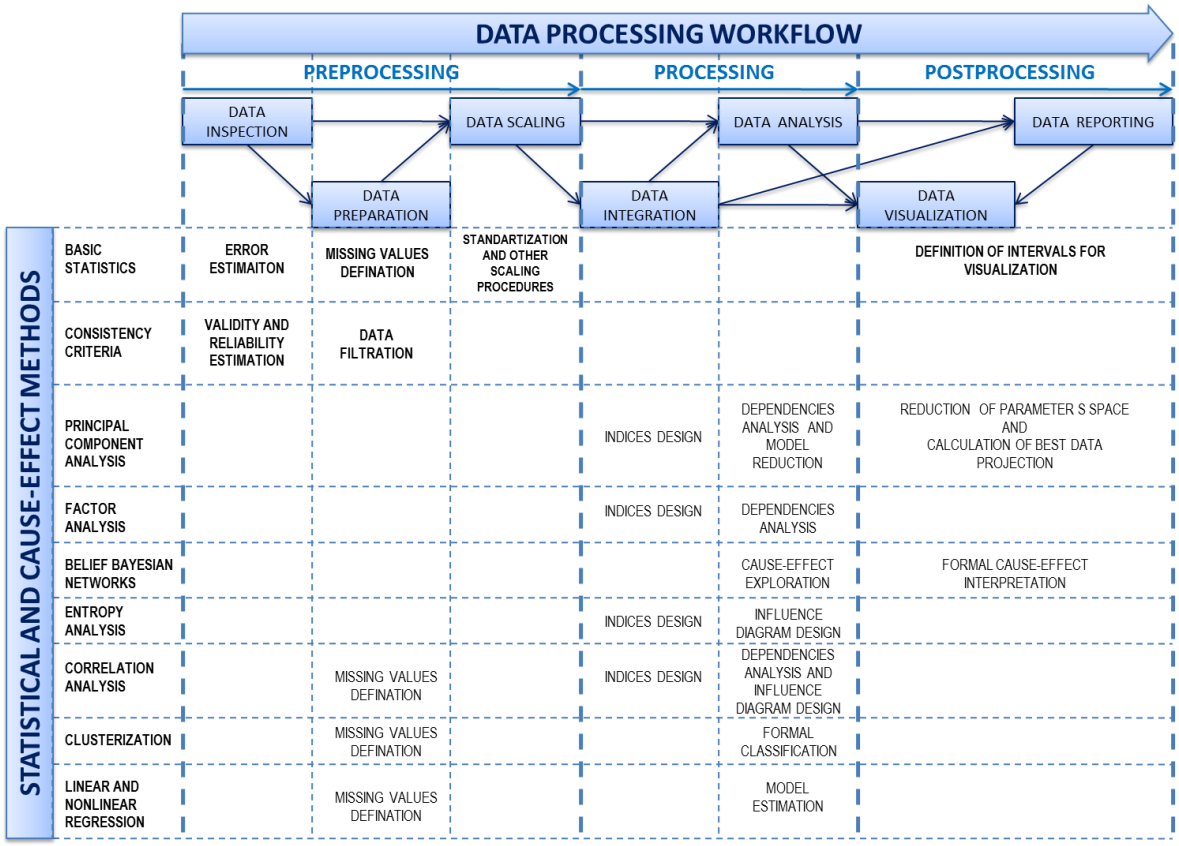
The Russian-Ukrainian WDS Segment provides a wide range of data for a wide diversity of disciplines: Seismology, Gravimetry, Heat Flow, Magnetic Measurements, Archeo- & Paleomagnetism, Solar Activity and Interplanetary Mediums, Cosmic Rays, Ionospheric Phenomena, Geomagnetic Variations, Space Geodesy and Geodynamics, Oceanography, Meteorology, Cartography, Remote Sensing, Sustainable Development, etc.

Segment members are involved with increasing frequency in solving fundamental and applied interdisciplinary problems that need systematic adjustment of data of various natures and the use of intelligent data processing technologies. For this purpose we have started to unite tools for intelligent data processing into a special framework.

4.1 Organization of intelligent data processing

Figure 1 shows a typical process of intelligent data processing (Zgurovsky, 2010) that can be divided into stages of preliminary processing, analysis, and post-processing of data. Each of these stages solves special tasks with the usage of methods of multidimensional statistic analysis, statistic-expert methods, and other methods. The same methods can be used to solve different tasks and the same task can be solved by different methods.

Figure 1. Process on intelligent data processing



That is why we can point out software-based means that use statistical and expert methods, regardless of context of their usage, to solve particular tasks in data processing. These software-based means can be organized as universal software-based libraries in the programs responsible for dealing with the task of data processing in terms of general process, shown in Figure 1.

It should be noted that the possibility of organizing separate programs and libraries in terms of a general calculating process depends on the presence of data compatibility and results. Also there should exist general models of data that will be used for the specification of parameters and results of the work of particular programs.

Further, it should be pointed out that the data that are used can be kept in different storage areas. That is why it is important to develop special unified software-based mechanisms to access the data and provide the data with the structure of metadata, determined from the data sources. Metadata are combinations of some regularity in a particular sphere, and they are suitable for further combinations of different output data. For interdisciplinary research, there is a real problem in the development of program methods information transfer to metadata (Shapovalova, Yefremov, & Glukhanik, 2011).

Finally, every process of intelligent data processing should be performed as a tool for the solution of a particular application accessible to the user.

Paying attention to the above, the module organization of software-based methods of a system of intelligent data processing, divided into four levels, is proposed:

- The level of data model presentation used by higher level software methods for data and results specification;
- The level of software implementation of data processing methods that can be used for solving different tasks;
- The level of software implementation of stages and the overall process of data processing for a particular application task;
- The level of software implementation of joint mechanisms of access to data, data visualization, reports generation, etc.

The described organization under conditions of a sufficient spectrum of software methods on each level actually allows combining accessible software-based components for solving different tasks of intelligent data processing, which supports the concept of "quick" development.

4.2 Architectural principles of organization of an intelligent data processing system

The choice of architectural principles for building an intelligent data processing system must consider both the general and specific requirements that are advanced for such systems by all interested people. Regardless of the list of particular applications that the intelligent data processing system deals with, the system must satisfy requirements conditionally divided into two categories. One category is connected with supplying requirements advanced by the user. The second is determined by the possibility of quick development of the whole system as well as its components.

From the user's viewpoint, the system must correspond to these requirements:

- The system must allow the user to process data of a volume that exceeds the possibilities of its equipment;
- The system must process using a raw data format;
- The system must be accessible to a wide range of users;
- The system must be easy to use even for users unfamiliar with programming.

The system's requirements are as follows:

- The system must provide an increase of calculating sources and volume of data storage due to its workload;
- The internal mechanisms of data processing must be hidden from the user. These can include algorithms that are the author's know-how or use intermediate information for which there is not free access;
- The architecture and software-based mechanisms of the system must predict the possibility of adding a new function and (or) system reconfiguration.

These requirements determine the principles of building an intelligent data processing system on architectural and structural levels provided by the user and the system.

The possibility of a need for increased sources for data processing, which may exceed the user's equipment, forces the division of the system into client and server components. Such a client-server organization may require that the data processing be accomplished on distant servers (Korzhov, 1997).

In this case, the client simply enquires and receives the results while the user's equipment is used as the second client. To organize the relationship between client and server, net protocols of appendices can be used, and the Internet itself can be used as a communication environment. In this way, access to the system is provided for a wide range of users.

The level of modern net protocols of appendices can be divided into two categories: protocols based on the usage of access to information sources and protocols based on an external call for procedures.

The first category is called RESET (Representational State Transfer) protocols. In RESET, the user agents can communicate with heterogeneous sources. This communication is provided with the help of a unique interface of standard HTTP commands (GET, POST, PUT, and DELETE). In this case the resource contains all the information needed for its processing (Flanders, 2009).

Protocols of the second category use mechanisms of external call for procedures, among which SOAP (Simple Object Access Protocol) is the most widespread. It provides the transfer of structural messages through XML. As opposed to RESET, SOAP is a protected protocol with guaranteed transfer of messages (MacVittie, 2007).

Each of the described protocols has its advantages and disadvantages. To provide the possibility of integration of an intelligent data processing system with other systems, the goal is to use not one but several protocols to transfer the message.

The use of client-server architecture with a communication environment based on the Internet allows the examination of the intelligent data processing system from the point of view of cloud computing, where the information is placed and stored on distant servers, accessible with the help of the Internet, and is only temporarily stored on the client base (Mell, & Grance, 2009). Selecting cloud computing hides the operation mechanisms of the data processing tools, which is one of the requirements for system development.

From the point of view of the user, from renting the hardware devices or virtual calculating environment to accessing the functionally finished software-based product, the most appropriate intelligent data processing system is Software as a Service (SaaS) (Kolesov, 2008).

The concept of cloud computing can be put to use in systems designed on different architectural principles. For service-oriented architecture (SOA), the module method is used in the development of software, based on the use of standard interfaces when the system is a combination of autonomic services and a general communication mechanism.

It should be pointed out that it is necessary to input an interim chain (strip) into the system architecture, which organizes the relationship between connections and services. The methods of connecting the services to the strip are standardized. For users, the strip system strip, with its functions provided by a dispatcher (the person responsible for call processing and delivery of results (Ferguson, & Stockton, 2005)), is their point of access to the application SaaS.

Therefore, from the user's point of view, the intelligent data processing system must be the completed application SaaS, which allows the storage and treatment of data with the help of server sources. It is reasonable to use the principles of a service-oriented architecture with a dispatcher to create preconditions for possible scaling and system expansion.

5 CONCLUSION

The creation of the Russian-Ukrainian WDS Segment and its successful activities was an important step in the development of the Russian-Ukrainian Data Community and the establishment of a scientific data infrastructure for Russian and Ukrainian scientific organizations, which were made possible thanks largely to the implementation of joint bilateral projects with financial support from the Russian and Ukrainian Academies of Sciences and Basic Research Foundations of both countries.

Intelligent data processing is one of the stages of the interdisciplinary holistic process of working with scientific

data. The methods proposed by the authors have become the basis for the organization of the information-communication infrastructure for the World Data Center of Geoinformatics and Sustainable Development and its partners, part of the functional modules implemented as web-services within the scope of the WDC-Ukraine portal (http://wdc.org.ua). Also the proposed approaches were successfully used for building an intelligent data processing subsystem for the common segment data distribution system – the Common Data Catalogue.

6 REFERENCES

Ferguson, D. & Stockton, M. (2005) *SOA programming model for implementing Web services, Part 1: Introduction to the IBM SOA programming model*. Retrieved October 20, 2012 from the World Wide Web: http://www.ibm.com/developerworks/library/ws-soa-progmodel/index.html

Flanders, J. (2009) An Introduction To RESTful Services With WCF. *MSDN Magazine*, January 2009. Retrieved October 23, 2012 from the World Wide Web: http://msdn.microsoft.com/en-us/magazine/dd315413.aspx

Kolesov, A. (2008) SaaS Model — in World and in Russia. *Byte*, 10(119). Retrieved November 12, 2012 from the World Wide Web: http://www.bytemag.ru/articles/detail.php?ID=12825

Korzhov, V. (1997) Multilevel client-server system. *Networks*, 06. Retrieved October 17, 2012 from the World Wide Web: http://www.osp.ru/nets/1997/06/142618/#part_1 (in Russian)

MacVittie, L. (2007) REST as alternative for SOAP. *Networks and Communication Systems*, 1. Retrieved October 22, 2012 from the World Wide Web: http://www.ccc.ru/magazine/depot/07_01/read.html?0502.htm.

Mell, P. & Grance, T. (2011) The NIST Definition of Cloud Computing. *National Institute of Standards and Technology, Information Technology Laboratory,* SP 800-145.

Newell, A. & Simon, H. (1972) *Human problem solving*, Englewood Cliffs, NJ: Prentice-Hall

Shapovalova, S.I., Yefremov, K.V. & Glukhanik, A.I. (2011) Organization of integrated access to information resources. *Proceedings of the XI International Conference "Intelligent Analysis of Information"*. Kyiv, Ukraine (in Russian)

Somervill, M. & Rapport, D. (2000) *Transdisciplinarity: Recreating Integrated Knowledge*. Oxford, UK: EOLSS Publishers Co. Ltd

Starostenko, V., Yatskiv, Ya., Lyalko, V., Ivanov, V., Rudenko, L., & Yefremov, K. (2008) Ukrainian science data: mutual goals and approaches, *Proc. 21st Int. CODATA Conf.*, Kyiv, Ukraine

Yang, X. S. (2008) *Introduction to Computational Mathematics*, World Scientific Publishing, 2008. p245.

Zgurovsky, M.Z. (2010) System Adjustment of Various nature DATA for Global Modelling of Sustainable development. *Proc. 22nd Int. CODATA Conf.,* Cape Town, South Africa

Zgurovsky, M.Z, Stratukha, G.A., Melnichenko, A.A., Voitko, S.V., .Boldak, A.A., Yefremov, K.V. et al. (2010) *Sustainable development analysis – global and regional contexts. P.1. Global analysis of quality and security of life*, Kyiv, Ukraine: NTUU "KPI"

Zgurovsky, M.Z, Stratukha, G.A., Melnichenko, A.A., Voitko, S.V., .Boldak, A.A., Yefremov, K.V. et al. (2010) *Sustainable development analysis – global and regional contexts. P.2. Ukraine in the sustainable development indicator analysis*, Kyiv, Ukraine: NTUU "KPI"

Zgurovsky, M.Z., Gvishiani, A. D., Yefremov, K.V. & Pasichny, A.M. (2010) Integration of the Ukrainian science into the world data system. *Cybernetics and Systems Analysis*, Vol. 46, No 2, pp 211-219

(Article history: Available online 19 January 2013)