DEVELOPMENT OF 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, its state of the art, main priorities and research activities are described. One of the high priority tasks for Segment members is development of 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, data processing components for disciplinary and multidisciplinary applications. Thereby the principles of architectural organization of intelligent data processing systems are determined in this paper.

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

1 INTRODUCTION

Modern scientific researches connected with the search of answers on global challenges arousing in the beginning of XXI century are interdisciplinary and focused on solving of bad structured tasks. The example of such researches is the analysis of sustainable development processes in global and regional context (Zgurovsky, Stratukha, Melnichenko, Voitko, .Boldak, Yefremov et al., 2010).

When we talk about interdisciplinary (Somervill, & Rapport, 2000) we mean the usage of data which 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 complete presentation of object of research are developed. As a rule in terms of such researches the formal models are

absent, but at the same time there is a possibility to use the results of objective measurements. Notably the tasks of such researches can be described as bad structured (Newell, & Simon, 1972) for deciding which of the results of objective measurements and subjective expert assessment are used.

To deal with such tasks the methods of scientific calculations (Yang, 2008) that are the essence of intelligent data processing concept (Yang, 2008) are used more and more often. The concept consists of organization of the detection process in "raw" data of unknown nontrivial practically useful knowledge which can be interpreted and may be useful for decision making in different spheres of human activity.

It is clear that concept of intelligent data processing is oriented not only on the usage of special program tools but also on special information-communicative infrastructure, which allows to use huge volumes of data of different origin and its processing for search of solutions of interdisciplinary tasks.

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

The aim of this paper is to describe state-of-the-art of Russian-Ukrainian WDS Segment and to determine the principles of architectural organization of intelligent data processing system on the basis of which the components of open program system can be created and integrated, which give the user the data and tools to solve the challenges of interdisciplinary researches.

2 STATE OF THE ART

Five Russian ICSU World Data Centers (WDC) for Oceanography, Meteorology, Rockets, Satellites and Rotation of the Earth, Solar-Terrestrial Physics and Solid Earth Physics more than 50 years collect, analyze, archive and distribute data for the broad spectrum of observatory types. The Centers provide open and convenient access to great volumes of data, permanently increase the information resources in the Internet. At 2006 Ukrainian WDC for Geoinformatics and Sustainable Development was formed, it is one of 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 establishment of Scientific Council of Russian and Ukrainian World Data Centers and forming of the Russian-Ukrainian segment of WDCs (Segment).

The top priorities for our Segment are following:

- Integration to new WDS and effective cooperation with WDS members
- Providing safety of data from non-digital data carriers and its digitization
- Establishing data quality policy
- Development of data providers' infrastructure
- Formation of common information space for Russian-Ukrainian WDS segment

Creation of 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 approach would make possible to create a single access point to all data and services of Russian-Ukrainian WDS Segment. That would also provide a flexible framework for unified data processing toolkit development. The main features of this common information space:

- Flexible and scalable cross-platform open source-based architecture (e.g. SOA)
- Centralized data & 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 their interests to join the new ICSU World Data System and have successfully completed the necessary stages of certification. Today all Russian and Ukrainian WDCs have obtained status of regular WDS members.

2.1 WDCs in Obninsk

There are three World Data Centers established on the basis of 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.

WDC for Meteorology has commitment to provide long term secure preservation and dissemination of meteorological data and products both for global and regional scales. Data holding contains observed meteorological data for the long period from XVIII century to the present days. Data are collected from the Russian network of meteorological stations and from the other parts of the globe by means of telecommunication system of WMO. Data are validated, checked and updated in continuous way. 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/).

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

WDC for rockets, satellites and rotation of the Earth collects meteorological data of national and international rockets and satellites, data on Earth rotation. Necessary quality assessment and quality control procedures are applied when required. Data holding contains observed rockets and satellites meteorological data for the period from 1966 up to present days. 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 original data description language (DDL) and hierarchical DBMS AISORI developed in RIHMI-WDC. This ensure 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 metadata base. The metadata elements enable to identify type and origin of data, their spatial and temporal coverage, other characteristics necessary to ensure authenticity of data sets. Metadata are used to prepare published and electronic Data Catalogues. These are posted on RIHMI-WDC Web site (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 have been developed and applied to avoid data duplicates within global data sets. Also to ensure metadata integrity the directories of organizations, maritime research projects, research vessels are regularly validated to be consistent with international counterpart directories of IODE, GCMD, ICES, WMO, EDMO, EDMERP.

To achieve a high rate of data processing and provide effective online user access to the data they are loaded into relational Oracle DBMS which is a component of integrated web-technology. The last one enables user to discover the data of user's interest, to retrieve, browse and view them in a table form or as a plot, map or diagram.

For exposing on the Web metadata consistent with ISO 19115 standard there are two international metadata profile being in use within RIHMI-WDC: WMO core profile for meteorological data (http://www.wmo.int/pages/prog/www/WDM/Metadata/documents.html) and CDI for oceanographic data (http://www.seadatanet.org/Standards-Software/Metadata-formats). WDCs also strictly adhere to international standards for data exchange, in particular BUFR/CREX data format developed under WMO umbrella, and

widely used NetCDF and ODV data formats.

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

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

To perform processing functions necessary for data archiving and provision of user access to the data the IBM z9 BC mainframe is employed. IBM System Storage DS8300 offers high performance, higher capacity storage up to 512 Tb. Full Disk Encryption with local key management provides relentless data security. The firewall system also is maintained to secure user access via 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 buildings detached. The direct access of external user to the library is impossible. A set of information service systems provide data stewardship and preservation. In particular 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 WDCs data.

2.2 WDCs in Moscow

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

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 form. WDC for STP converts old analog data into digital form in order to preserve data to provide them to 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 web site 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 supports databases "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; search for specific events in "space weather" in terms of a natural language with fuzzy logic application. The WDC for STP is the participant of the international project in the field of information technologies and geophysics INTERMAGNET.

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, recent movements. Data stored in the Center are 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 results of land and sea expeditions, launches of satellites, special experiments, results of geophysical observations (seismological, geomagnetic etc.) on global networks of observatories. The Center accepts data according to a long term relationship with 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 is accepted only after they have passed examination in these organizations and have received the status of the data intended for the international exchange in the

decision of a commission of experts. The Center provides access to all these data and also serves as information and reference center. Some parts of the data are stored as publications on paper and microfilm, but considerable part are also available in digital electronic form on the WDC for SEP's web site http://www.wdcb.ru/sep/. Data, metadata, thematic databases, inventory catalogs are available on line on the web site. The WDC for SEP is the participant of the international and interdisciplinary project in the field of geology and geophysics InterMARGINS, concerned 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 the modern level of technologies 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 providing permanent online access to data. The user interface is developed for convenient search, browsing, visualization and retrieval of the data on the WDCs web sites.

All data received by WDCs for STP and SEP are analyzed: a 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 is carried out with special QC software. Descriptions of data, data formats and metadata of GCMD DIF standard are prepared.

2.3 WDC in Kyiv

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. WDC-Ukraine was created by the decision of Presidium of the National Academy of Sciences of Ukraine (NAS), Ministry of education and science of Ukraine and Geophysical Center of the Russian Academy of Sciences (GC RAS) from April, 3, 2006 as a subdivision of the Russian World Data Centers "WDC for Solar-Terrestrial Physics" and "WDC for Solid Earth Physics" (Moscow). Afterwards in the 2008 Ukrainian branch of WDC obtained an independent status as World Data Center for Geoinformatics and Sustainable Development.

The World Data Center in Ukraine should afford access to global information resources of the ICSU on Earth sciences, planetary and space physics, and related subjects for the Ukrainian scientific community and to provide acquisition and storage of national scientific data on the above disciplines and their presentation to the world community. WDC-Ukraine collects, processes, analyses and disseminates global and national data necessary for sustainable development research. Center is one of leaders in various fields of sustainable development research in Ukraine. Scientific and technical staff of the WDC-Ukraine performs fundamental and applied research and analysis for solving interdisciplinary problems with system nature, particularly quantitative measurement and modelling of sustainable development processes and evaluation of the impact of the global threats set on sustainability in global and regional contexts, it also develops and implements information technologies for solving of 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 system approach have allowed proposing a unique (for the World Data System) network model of functioning of the WDC-Ukraine as 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 conventionally called "Network of networks" (Starostenko, Yatskiv, Lyalko, Ivanov, Rudenko, & Yefremov, 2008). At the session of WDS Scientific Committee 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, educational organizations, and the high-performance computing cluster of the NTUU "KPI" (cluster is a part of the national GRID-infrastructure so in case of need tasks can be distributed among all partners of this network), and on the other hand, focusing of the efforts of the WDC staff on solving interdisciplinary system 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 nature;
- 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 Segment centers fulfill a number of joint projects aimed at development of 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 set of databases and processing algorithms aimed to system prevision of complex anthropogenic and natural systems' behavior»;
- 2009 – 2010 –«Development World Data Centers network for investigation of basics of complex natural and anthropogenic systems' global modeling»;
- 2010 – 2011 –«Development of fundamentals analytical methods of multidisciplinary data for creation 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 the general approach and methods for system adjustment of data of various nature in the distributed multidisciplinary databases infrastructure of the World Data System Russian-Ukrainian Segment for solving of fundamental interdisciplinary tasks of processes correlation in the geospheres system».

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

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

To present WDS in Internet and providing an access to data and metadata of WDS a new web-site was designed and developed by World Data Center for Geoinformatics and Sustainable Development. It is located on the address www.icsu-wds.org and supported by WDC-Ukraine staff. The Web-portal is constructed on the basis of flexible and easily scaled platform allowing, if necessary easily and quickly revise not only material but also 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

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

Segment members are involved increasingly frequently to solve fundamental and applied interdisciplinary problems that need system adjustment of data of various nature and using intelligent data processing technologies. For this purpose we start to unite tools for intelligent data processing in special framework.

4.1 Organization of the intelligent data processing

As it is seen from Figure 1 which shows a typical process of intelligent data processing (Zgurovsky, 2010) it 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.

You can also notice that the same methods can be used to solve different tasks and the same task can be solved by different methods.

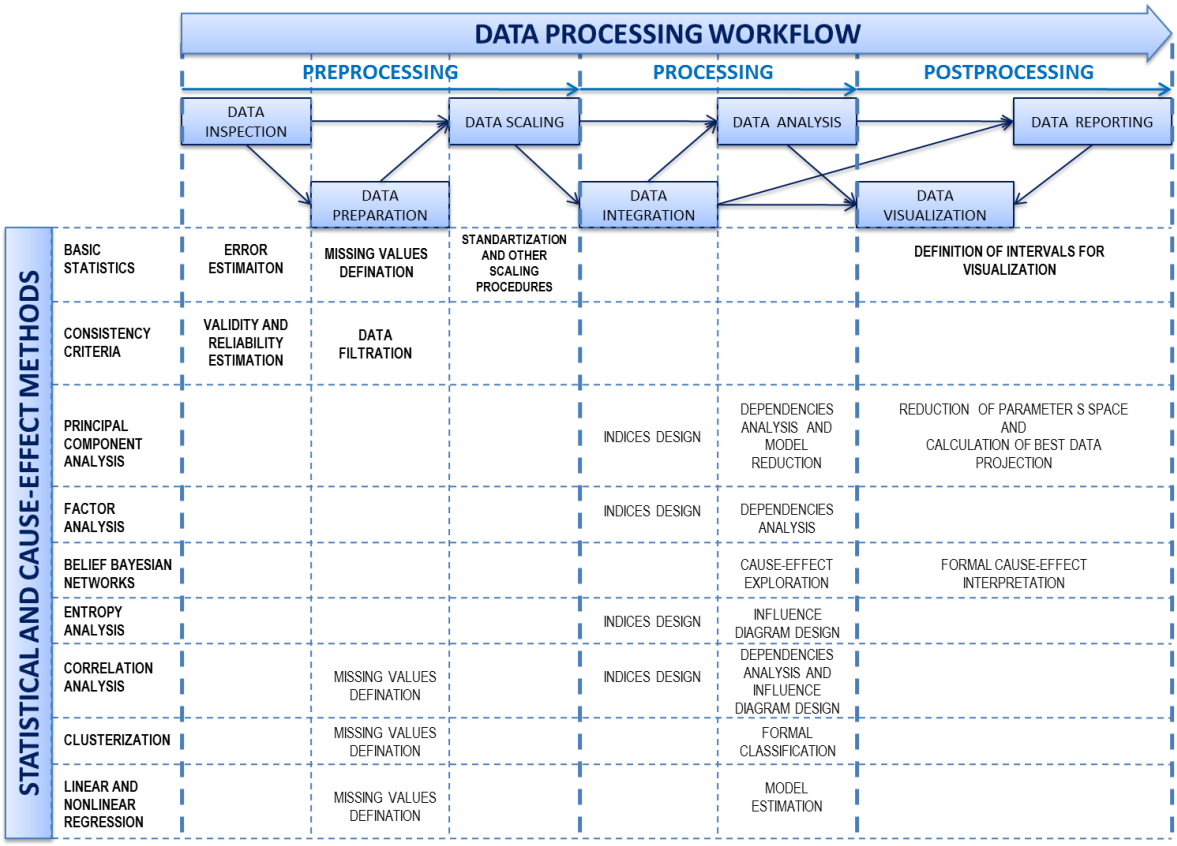

Figure 1. Process on intelligent data processing

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

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

Also it should be pointed out that data which are used can be kept in different storages. That is why it is suitable to develop special unified software-based mechanisms of access to them and provide data with structures of metadata, determined in a result of their diversity on another level from sources of data. Metadata is the combination of some regularity in particular sphere and they are suitable for further combination of different output data. For interdisciplinary researches the problem of development of program methods information transfer to metadata is actual (Shapovalova, Yefremov, & Glukhanik, 2011).

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

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

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

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

4.2 Architectural principles of organization of intelligent data processing system

Choice of architectural principles of building of intelligent data processing system must consider general and specific requirements which are advanced for such systems from side of all interested people. Regardless of list of particular applications, which the intelligent data processing system deals with, it must satisfy the requirements conditionally divided on two categories. One from them is connected with supplying requirements which are advanced by user. The second one is determined by the possibility of quick development of a whole system as well as its components.

From user's view the system must correspond to such requirements:

- system must allow user to process data of such volume which exceeds the possibilities of its equipment;
- system must process raw data format;
- system must be accessible to the wide range of users;
- system must be simple in use even for users who do not know programming.

Systems requirements are as follows:

- system must provide increasing of calculating sources and volumes of data storage due to its workload;
- internal mechanisms of data processing must be hidden from user. They can include algorithms which are author's now-how or use the intermediate information, which is not free in access;
- architecture and software-based mechanisms of system must predict the possibility of adding of new function and (or) system reconfiguration.

These requirements determine the principles of building of intelligent data processing system on architectural and structural levels of their providing.

Possible increasing of sources, needed for data processing, over possibilities of user's equipment forces to use the system division into client and server components in terms of one of architectural principles of system organization. Such client-server organization means that the data processing may be held on some distant servers (Korzhov, 1997).

In this case the client just enquires and gets the results of counting, notably the user equipment is used as the second client. For organization of relation between client and server net protocols of appendixes can be used and the Internet itself can be used as communication environment. In this way the access to system is provided for wide range of users.

Modern net protocols of appendixes level can be divided into two categories: protocols, based on usage of access to information sources and protocols, based on distant call of procedure.

The first category is called RESET (Representational State Transfer) protocols. In case of usage of such architecture the user agents have the possibility to correlate with heterogeneous sources. This correlation is provided with the help of unique interface of standard commands HTTP (GET, POST, PUT and DELETE). In this case the resource contains all information needed for its processing (Flanders, 2009).

Protocols of second categories use mechanisms of distant call of procedures, among which protocol SOAP (Simple Object Access Protocol) is the most widespread. It provides the transfer structural messages on the base of XML. As opposed to REST, SOAP is protected protocol with guaranteed transfer of messages (MacVittie, 2007).

Each of described protocols has its advantages and disadvantages. To provide the possibility of integration of intelligent data processing system with other systems, the target is to realize not one but several protocols with the usage of which the message transfer is held.

The usage of client-server architecture with communication environment on the base of net Internet allows to examine the intelligent data processing system with the point of view of concept of cloud computing, in the terms of which the information is placed and stored on the distant servers, which are accessible with the help of the Internet and is only temporary storage on the client base (Mell, & Grance, 2009). The selection of cloud computing concept hides the mechanisms of operation of data processing tools which is one of the requirements for developing system.

Among such views on the system from the point of view of user, from renting him the hardware devices or virtual calculating environment and up to providing the access to the functionally finished software-based product, for the intelligent data processing system the most appropriate is its presentation as Software as a Service (SaaS) (Kolesov, 2008).

Concept of cloud computing can be realized in systems designed on different architectural principles. In terms of one of them which has the title service-oriented architecture (SOA), the module method is used in the development of software, based on services usage (service) with standard interfaces, when system is examined as combination of autonomic services combined by general communication mechanism.

It should be pointed out that it predicts the necessity of inputting into system architecture of interim chain (strip) which organizes the relation between connections to it services. The methods of services connection to strip are standardized. For user the system strip, its functions provided by dispatcher (he is responsible for calls processing and results delivery (Ferguson, & Stockton, 2005)), realizes the concept of one point of access to the application SaaS.

Thereby, from the user point of view the intelligent data processing system must be a finished application SaaS which allows to store and treat data with the help of server sources. It is reasonable to use principles of service-oriented architecture with dispatcher that realizes the functionality of system strip creating preconditions for possible scaling and system expansion.

5 CONCLUSION

Creation of Russian-Ukrainian WDS Segment and its successful activities became the important step for development of Russian-Ukrainian Data Community and establishment of scientific data infrastructure for Russian and Ukrainian scientific organizations that become 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 stages of interdisciplinary holistic process of work with scientific data. Methods proposed by authors became the basis for organization of information-communication infrastructure for World Data Center of Geoinformatics and Sustainable Development and its partners, part of functional modules implemented as web-services within the scopes of WDC-Ukraine portal (http://wdc.org.ua). Also proposed approaches were successfully used for building intelligent data processing subsystem for Segment common distributed data system – 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. — 245 p.

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