"Science Stays True Here"
Advances in Ecological and Environmental Research (ISSN 2517-9454, USA), 168-176 | Science Signpost Publishing

Data Management System of the UNECE ICP Vegetation

Alexander Uzhinskiy[1], Gennady Ososkov[1], Marina Frontasyeva[2]

1. Laboratory of Information Technologies.

2. Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, 6, Joliot-Curie, str.,141980, Dubna, Moscow Region, Russian Federation.

Received: April 02, 2019 / Accepted: May 10, 2019 / Published: June 25, 2019

Abstract: The aim of the UNECE International Cooperative Program (ICP) Vegetation in the framework of the United Nations Convention on Long-Range Transboundary Air Pollution (CLRTAP) is to identify the main polluted areas of Europe, produce regional maps and further develop the understanding of the long-range transboundary pollution. The Data Management System (DMS) of the UNECE ICP Vegetation consists of a set of inter-connected services and tools deployed and hosted at the Joint Institute of Nuclear Research (JINR) cloud infrastructure. DMS is intended to provide its participants with modern unified system of collecting, analyzing and processing of biological monitoring data. Motivation, basic principles and architecture of the DMS are presented.

Keywords: environmental monitoring, data management, cloud platform, intellectual data processing, UNECE ICP Vegetation, air pollution, mosses, heavy metals, neural networks.

1. Introduction

Air pollution has a significant negative impact on the various components of ecosystems, human health, and ultimately, cause significant economic damage. Increased ratification of the Protocols of the Convention on Long-range Transboundary Air Pollution (LRTAP) is identified as a high priority in the new long-term strategy of the Convention. Full implementation of air pollution abatement policies is particularly desirable for countries of Eastern Europe, the Caucasus and Central Asia (EECCA) as well as South-Eastern Europe (SEE). Atmospheric deposition study of heavy metals, nitrogen, persistent organic compounds (POPs) and radionuclides is based on the analysis of naturally growing mosses through moss surveys carried out every 5 years [1]. Due to intense activity of the Joint Institute for Nuclear Research (JINR) as a coordinator of the moss surveys since 2014, Armenia, Azerbaijan, Georgia, Kazakhstan, Moldova, Mongolia and Vietnam joined the moss survey for 2015/2016. Nowadays the UNECE International Cooperative Program (ICP) Vegetation [2] is realized in 39 countries of Europe and Asia. Mosses are collected at thousands of sites across Europe and their heavy metals (since 1990), nitrogen (since 2005), POPs (persistent organic compounds, pilot study in 2010) and radionuclides (since 2015) concentrations are determined. The goal of this program is to identify the main polluted areas, to produce regional maps and to model the long-range transboundary pollution [3].

2. Experiment and data interpretation

Sampling is carried out in compliance with the internationally accepted guidelines [4]. Such analytical techniques as AAS, AFS, CVAAS, CVAFS, ETAAS, FAAS, GFAAS, ICP-ES, ICP-MS, as well as INAA are used for elemental determination. A total of 13 elements are reported to the Atlas (As, Cd, Cr, Cu, Fe, Hg, Ni, Pb, V, Zn, Al, Sb, and N). Nowadays POPs (whichever determined) and radionuclides (namely, ^{210}Pb and ^{137}Cs) are accepted for air pollution characterization. The results are reported as a number of sampling sites, minimum, maximum and median concentrations in mg/kg. The data interpretation is based on Multivariate statistical analysis (factor analysis), description of sampling sites (MossMet information package) and distribution maps for each element. Examples of maps are presented in Fig. 1.

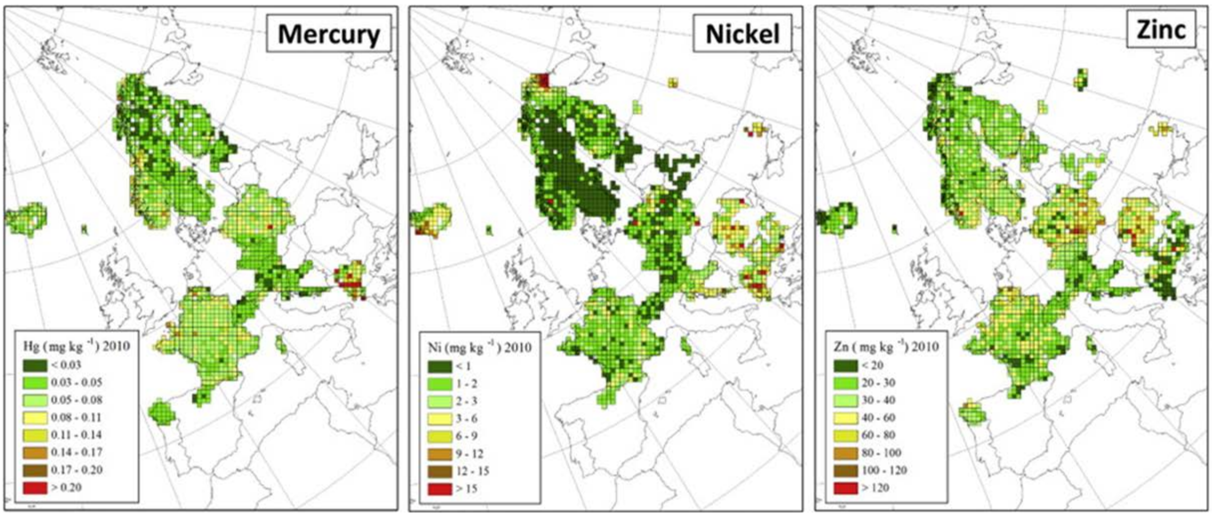


Figure. 1. Examples of distribution maps [3]

3. Motivation

The UNECE ICP Vegetation program is a very important project, but it had a serious drawback related to its weak adoption of modern informational technologies. There are dozens of respondents in the existing monitoring network and their number is increasing. Information on collecting and processing of samples was carried out manually or with minimum automation. Until 2014, data mostly was stored in Excel files and was aggregated manually by the coordinator. Files from respondents were usually passed to the coordinator by email. There was no common standards in data transfer, storing and processing software. Such situation does not meet the modern standards for quality, effectiveness and speed of research. Lack of a single web-platform that provides comprehensive solution of biological monitoring and forecasting tasks was one of the major problem for research.

The aim of the project was to develop DMS using modern analytical, statistical, programmatic and organizational methods to provide the scientific community with unified system of gathering, storing, analyzing, processing, sharing and collective usage of biological monitoring data.

The DMS elements are to facilitate IT-aspects of all biological monitoring stages starting from a choice of collection places and parameters of samples description and finishing with generation of pollution maps of a particular area or state-of-environment forecast in the long term. Mechanisms and tools for association of participants of heterogeneous networks of biological monitoring are provided in the DMS. That enables verifying obtained results and optimizes research. The open part of the DMS can be used for informing public authorities, local governments, legal entities and individuals about state-of-environment changes.

One more important aspect of ecological researches relates to various statistical methods applied to process collected data. Modern approaches to explore air pollutions provided by heavy metals, nitrogen, POPs and radionuclides include as a mandatory part of multivariable statistical and intellectual data processing. Latest tendencies in that field include extension of a set of georeferenced data integrated in data processing of surveyed material. So it is not limited by geographical, topographical or geological information, what is traditional in such cases, but also includes, for example, satellite imagery and their products, topographic high-precision data derived from aerial photography, etc. These new data classes, contrary to the traditional ones, are characterized by a high resolution and dynamic nature, for example, satellite images represent a reflection of solar radiation, which depends on the time of day, season, cloud cover, etc. This in turn greatly increases the amount of data to be processed. The task of integration of different data types is tied to the problem of the development of new models and algorithms – such as neural networks [5], self-organizing maps [6], etc during the study of dynamic properties of environmental processes among other things.

So, one more aim of our project is to develop modern software tools for multivariable statistical and intellectual data processing oriented on the GIS-technology.

4. Architecture and technologies

To optimize the whole procedure of data management, it was proposed to build a DMS consisting of a set of interconnected services and tools to be developed, deployed and hosted in the JINR cloud infrastructure [7].

Such an approach also allows scaling cloud resources up and down according to the service load. When some particular component will require more resources the cloud can provide them without affecting other components. This increases the efficiency of hardware utilization as well as the reliability and availability of the service itself for end-users. Such auto-scaling behavior will be achieved by using the OneFlow component of the OpenNebula platform, which the JINR cloud is based on.

We defined requirements for the DMS and it components. The general architecture of the platform and technologies used are depicted in Fig. 2.

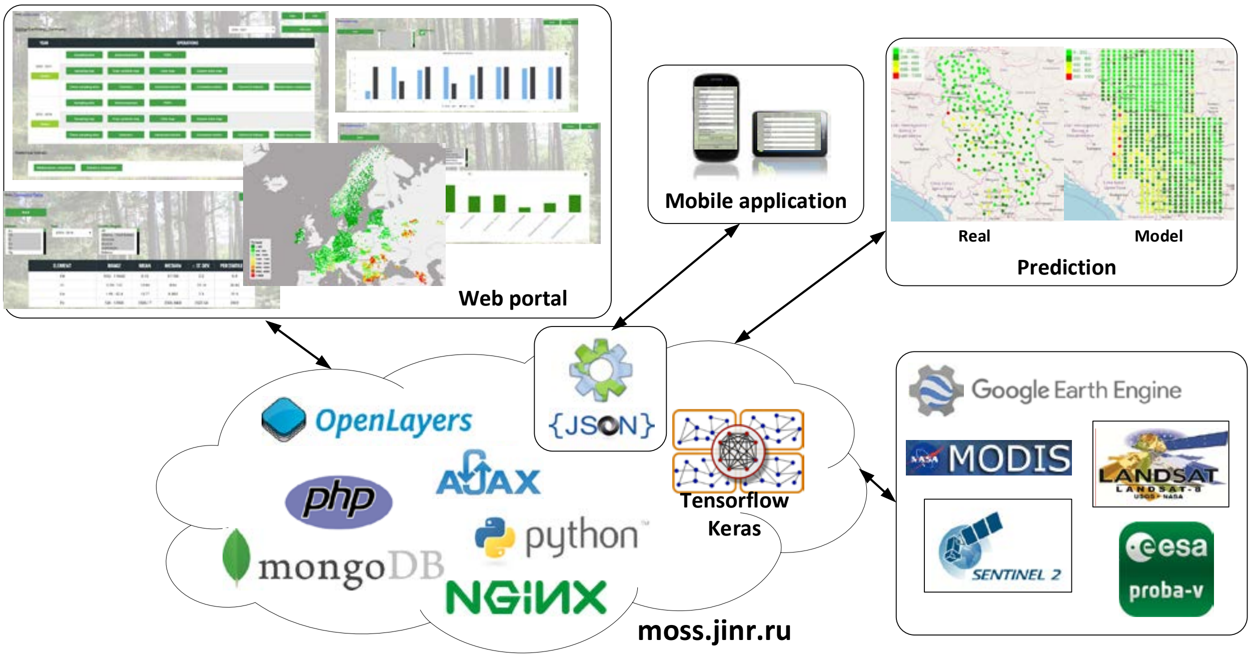
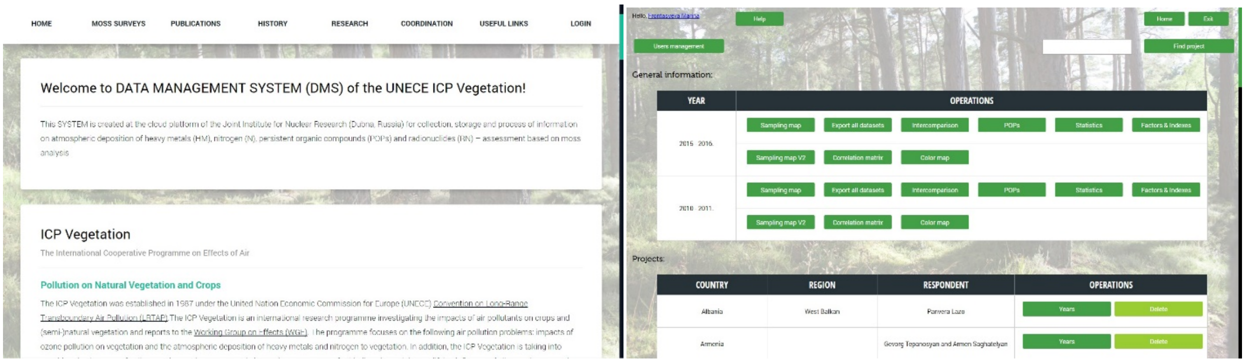


Figure 2. General architecture of the DMS and technologies used

A preliminary data analysis of raw data provided by the contributors was performed. The data samples can have 10 to 100 metrics depending on the collecting area. Interlaboratory comparison and POPs data also have different variegated structure. While project evaluates, it will be necessary to store data of other projects and experiments that will also have different structure. Traditional relation databases are ineffective in current situation and no relational MongoDB was used. The web-server of the DMS is Nginx since it can provide needed performance. The portal back-end developed with PHP that also used for calculation of the factors, indexes and other statistic parameters.

The web-portal has responsive design that adjusts to different screen sizes. The portal allows multilevel access to the data and has advanced data processing and reporting mechanisms. There are two parts of the portal – public and private ones. A general information about the project and the platform is presented in the public part. The private part can be accessed only by authorized contributors and are used for data management and analysis (see Fig. 3).



Welcome to DATA MANAGEMENT SYSTEM (DMS) of the UNECE ICP Vegetation!

This SYSTEM is created at the cloud platform of the Joint Institute for Nuclear Research (Dubna, Russia) for collection, storage and process of information on atmospheric deposition of heavy metals (HM), nitrogen (N), persistent organic compounds (POPs) and radionuclides (RN) – assessment based on moss analysis

ICP Vegetation

The International Cooperative Programme on Effects of Air

Pollution on Natural Vegetation and Crops

The ICP Vegetation was established in 1987 under the United Nation Economic Commission for Europe (UNECE) Convention on Long-Range Transboundary Air Pollution (LRTAP) The ICP Vegetation is an international research programme investigating the impacts of air pollutants on crops and (semi-)natural vegetation and reports to the Working Group on Effects (WGE). The programme focuses on the following air pollution problems: impacts of ozone pollution on vegetation and the atmospheric deposition of heavy metals and nitrogen to vegetation. In addition, the ICP Vegetation is taking into

Hello, Frontasyeva Marina

General information:

YEAR	OPERATIONS
2015 - 2016.	Sampling map, Export all datasets, Intercomparison, POPs, Statistics, Factors & Indexes Sampling map V2, Correlation matrix, Color map
2010 - 2011.	Sampling map, Export all datasets, Intercomparison, POPs, Statistics, Factors & Indexes Sampling map V2, Correlation matrix, Color map

Projects:

COUNTRY	REGION	RESPONDENT	OPERATIONS
Albania	West Balkan	Panvera Lazo	Years Delete
Armenia		Gevorg Tepanosyan and Armen Saghatelyan	Years Delete

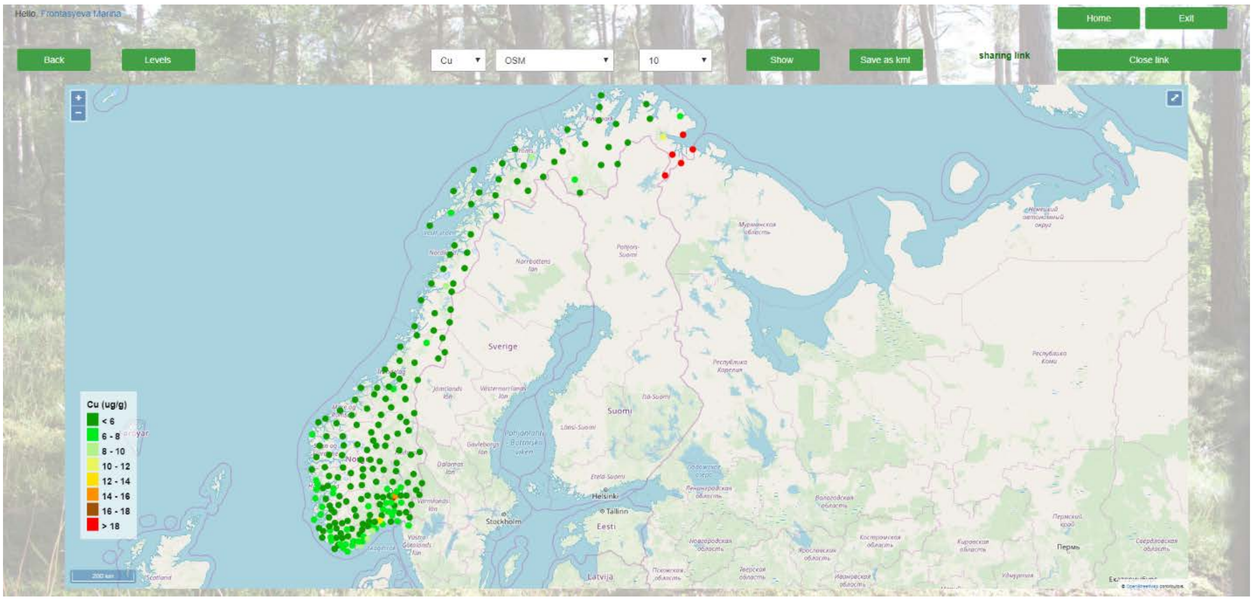
Figure 3. DMS interfaces. Left – public part. Right – private part

Contributors can manage their sampling, inter laboratory comparison and POPs data at the private part of the DMS. Data can be imported from and exported to Excel files. Verification of imported data is done automatically and allows finding most of human made mistakes.

During processing data for Atlas 2015-2016 we have meet misspelling in moss names, wrong coordinates, negative concentration and many other problems with data. Now for Atlas 2020-2021 we have the mobile application that allows filling in required by the UNECE ICP Vegetation manual information about sampling sites. The application automatically sets longitude and latitude of the sampling site, controls correctness of the input data and allows capturing photos of the moss samples and the nearest area. The application integrated with the DMS and all information about sampling sites can be imported to the system.

At the DMS participants can get some statistic parameters of their data, correlation between element concentration, contaminations factors, geo-indexes and so on.

We are using the OpenLayers library and few simple backgrounds to generate maps. Data can be represented as sampling map, color map (where color represent concentration) and graduated symbol map see Fig. 4. Contributors can share their maps or statistic metrics so it can be accessed with no credentials from all over the world.



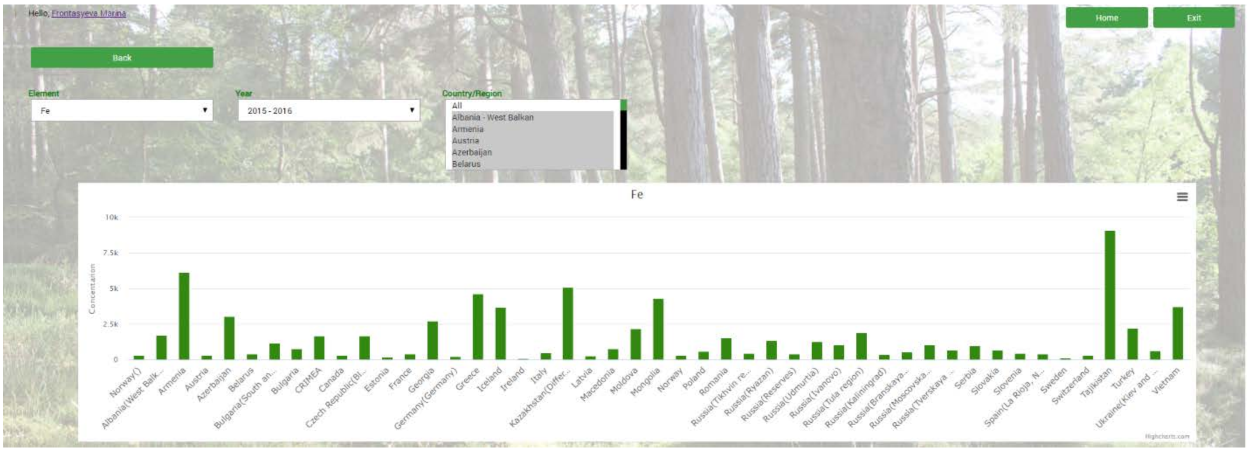
Hello, Frontasyeva Marina

Cu
OSM
10

Figure 4. Color graduated map of Cu at Norway

Mostly participant can access only their own data, but in some case they can use special tools to get general information from others contributors. For example, it is useful to execute mean values comparison of elements concentration see Fig. 5. Contributors can always forbid joint operation with their data at their profile. As soon as we have information for few atlases, it is possible to check some historical trends and build some charts and graphs.

Coordinators of the program can access to any contributor's data and tools and they have ability to execute group operation with data.



Hello, Frontasyeva Marina

Element: Fe

Year: 2015 - 2016

Country/Region:
All
Albania - West Balkan
Armenia
Austria
Azerbaijan
Belarus

Figure 5. Comparison of the Fe concentration mean values

Prediction is an important step of data analysis of any ecological survey. ArcGIS and some interpolation methods was generally used for concentration prediction and distribution map creation [8], but we were looking for some new ways. We are trying to use satellite imagery data and artificial neural network to predict concentration. General idea is to use data that we can get from satellite images together with sampling data from DMS to learn NN and then use only data from satellite images to predict concentration. There are open programs like LandSat [9], MODIS [10], Sentinel 2 [11] that provide free access to their data. One can search their database and find necessary images. Special software such as ENVI or ERDAS can be used to process images.

It is not useful because images are of gigabyte size, and we should have few of them to cover the region; some software exists to search through image archives, but the functionality of these programs is rather poor, and they often work with only one satellite data source; it is almost impossible to automate the process, and even if we could, the consumption of resources would be too huge. Alternatively, Google Earth Engine platform can be used. It combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface. The platform has advanced mechanisms to search, process and analyze satellite data.

We have create a piece of software that takes coordinate of the sampling site from DMS and calculate indexes from different programs satellite images for them and then calculate the correlation between the contamination and indexes. We had analyzed information for seven countries where the number of sampling sites were more than 200: Norway, France, Germany, Sweden, Rumania, Serbia, and Iceland. We used more than 20 satellite programs/products, and different types of indexes for them to find a correlation. We have found several regions were few elements have rather good correlation with satellite indexes [12]. Nature of such correlations can be an objective of the independent study. Data from DMS and satellite indexes were used to train different statistical models. Best models can use only new satellite indexes to predict concentrations. Such approach was used to predict Sb for Norway, U for Romania and Mn for Serbia see Fig. 6.

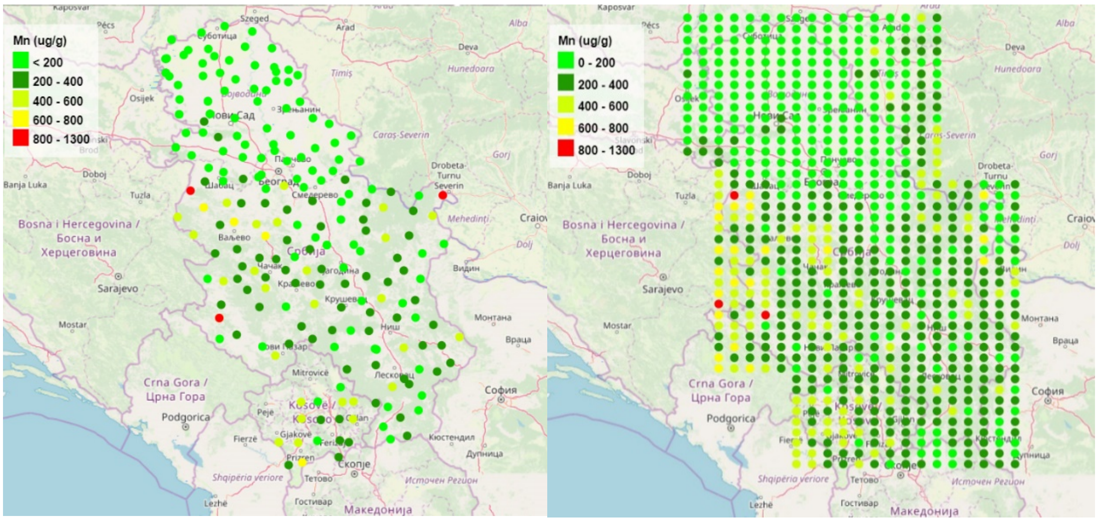


Figure 6. Mn for Serbia. Left – real data. Right – model prediction.

We keep on searching connection of the contamination and satellite indexes and testing new models and approaches.

5. Conclusion

The study of migration and deposition of toxic pollutants, which include heavy metals, persistent organic pollutants and radionuclides, the influence of pollutants on the various components of the natural and urban ecosystems is the key problem of modern biogeochemistry and environmental studies. The DMS is using modern analytical, statistical, programmatic and organizational methods to supply the scientific community with unified system of collecting, analyzing and processing of biological monitoring data. The project is carried out in the framework of the International Cooperative Program ICP Vegetation. Currently all key elements of the DMS have been realized. We embodied the platform requirements and it components in its general architecture. It allows organizing data collection quite flexible in respect to heterogeneity of raw data, then preliminary data analysis and verification. Contributors are able now to generate maps in desirable format and make basic statistic calculations. Our important achievement is that the platform has now advanced mechanisms for intellectual data processing oriented on the GIS-technology. Experiments on predicting element concentrations using neural networks and satellite imagery are in progress.

References

[1] United Nations Economic Commission for Europe International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops [Electronic resource]: http://icpvegetation.ceh.ac.uk/

[2] H. Harmens and G. Mills (Eds.) Air Pollution: Deposition to and impacts on vegetation in (South-East Europe, Caucasus, Central Asia (EECCA/SEE) and South-East Asia. Report prepared by ICP Vegetation, March 2014. ICP Vegetation Programme Coordination Centre, Centre for Ecology and Hydrology, Bangor. UK. 2014. 72p.

[3] H. Harmens, D.A. Norris, K. Sharps, G. Mills, M. Frontasyeva et al. Environmental Pollution. 2015. p. 93-104.

[4] HEAVY METALS, NITROGEN AND POPs IN EUROPEAN MOSSES: 2015 SURVEY: http://icpvegetation.ceh.ac.uk/publications/documents/MossmonitoringMANUAL-2015-17.07.14.pdf

[5] J. Alijagić. Application of multivariate statistical methods and artificial neural network for separation natural background and influence of mining and metallurgy activities on distribution of chemical elements in the Stavnja valley (Bosnia and Herzegovina). PhD thesis. University of Nova Gorica. 2013.

[6] G. Žibret, R. Šajn, Mathematical Geosciences. 2010,42(6): 681–703.

[7] A.V. Baranov, N.A. Balashov, N.A. Kutovskiy, R.N. Semenov. Physics of Particles and Nuclei Letters. 2016 ISSN 1547-4771 eISSN: 1531-8567, vol. 13, No. 5, pp. 672–675, DOI: 10.1134/S1547477116050071.

[8] A. Buse et al. Heavy metals in European mosses: 2000/2001 survey. UNECE ICP Vegetation Coordination Centre, Centre for Ecology and Hydrology, Bangor, UK. 2003.

[9] Landsat program home page - https://landsat.usgs.gov/

[10] MODIS (Moderate Resolution Imaging Spectroradiomete) program home page - https://modis.gsfc.nasa.gov/

[11] Sentinel program home page - http://www.copernicus.eu/main/sentinels

[12] Alexander Uzhinskiy, Gennady Ososkov, Pavel Goncharov, Marina Frontsyeva, Combining satellite imagery and machine learning to predict atmospheric heavy metal contamination // CEUR Workshop Proceedings, Vol-2267, ISSN 1613-0073, p. 351-358, 2018