The Open Access Portal to Vernadsky State Geological Museum of Russian Academy of Sciences Data

Kirill A. Platonov
Vernadsky State Geological
Museum RAS
Moscow, Russia
k.platonov@sgm.ru

Vera V. Naumova
Vernadsky State Geological
Museum RAS
Moscow, Russia
v.naumova@sgm.ru

Vitaliy S. Eremenko
Vernadsky State Geological
Museum RAS
Moscow, Russia
vitaer@gmail.com

Sergey E. Dyakov
Institute of Automation and
Control Processes FEB RAS
Vladivostok, Russia
sergdkv@gmail.com

Iraida A. Starodubtseva
Vernadsky State Geological
Museum RAS
Moscow, Russia
i.starodubtseva@sgm.ru

Abstract

In article share data of the Vernadsky State Geological Museum of RAS, one of the oldest natural-science museum of Russia, which began to be created in 1759, are described and more than 250 years are replenished thanks to scientists, patrons and collectors. Development of the Portal of open access to data of the Museum is discussed. The Portal provides data sets to users on the Internet. It allows any Internet user to research, load and use data for researches and the educational purposes. The standards of metadata and protocols can be used further for integration of data SGM RAS into the Russian and world networks of natural science museum data.

1 Introduction

The application of information and communication technologies in the second decade of XXI century is the necessity not only for large museums, but also for quite small museums, even municipal. Information technologies allow effectively solving many traditional museum problems, including the main museum missions such as storage, research and exposition of museum collections. It is necessary to mention Internet as a powerful tool of museum communication with the world, visitors, including virtual. It is an active implementation of new communication forms, not only through web site, but also through social media. The purpose of such communication is always museum audience expansion and guarantee to the maximum number of people, wherever they are, access to the collections [Ana14].

The basic tendencies in museum information activity are, from one side, intensification of "internal" museum service functions execution (development and implementation of automated record, storage and analysis information systems about museum objects), from other side, the increase of museum expositions interactivity by interactive elements, augmented and virtual systems application and, from the third side, an attraction of public attention to the museum activity through development of "external" information space, providing museum information communication as with public as also with scientific and industrial organizations.

The information technologies are even more important for natural history museums comparing with other museums. The natural history museum collections have enormous information about natural objects, both, exiting and lost (extinct forms of natural life, mined out deposits etc.) This requires the development of information systems additional functions, as existing ones (for example KAMIS) don't meet the natural history museums requirements.

This article describes the application and development issues of information technologies in natural history museums on the example of Vernadsky State Geological Museum RAS. The main directions of Vernadsky State Geological Museum RAS activity are [Mal14]:

1. preservation and augmentation of collections, numbering more than 250 thousands minerals, rocks and fossils;
2. modern scientific research; assistance in expansion of scientific and industrial organizations cooperation;
3. educational and cultural activity for a wide public.

The first step of Internet usage as a tool of museum communication with the world is Natural History museum data open access portal development.

The term "open access" was for the first time mentioned at the Budapest conference on open access in February 2002. The meaning is not practically changed: Open Access is a free, immediate, permanent, fulltext, online access to scientific publications.

The seminar "Information Technologies and Dublin Core Metadata Group operation" within the frames of 69th IFLA General Conference defines the main principals of "Open Archive" ideology: world-wide consolidation of scientific materials archives; open access to archives (metadata); consistent archives and information providers interfaces; usage simplicity; application of existing standards - HTTP, XML, Dublin Core, MARC, MARCXML.

The open access systems to scientific publications, to scientific information archives and to natural history museums data are developed in the present moment for intensification of scientific research and scientific communication development.

Data Portal of Natural History Museum, London; Yale Peabody Museum of Natural History, New Haven; Australian Museum, Sydney and others are the examples of Natural history museums data open access systems.

The funds of the one of the oldest Russian natural history museums Vernadsky State Geological Museum (SGM RAS) started to form in 1759. and for more than 250 years have been supplemented by scientists, patrons and collectors. Minerals, rocks and ores, meteorite, invertebrate fossils, vertebrates and fossil flora are store here. The collections are distributed according to the corresponding inventory: "Minerals", "Rocks and Ores" "Fossil invertebrate", "Fossil vertebrate", "Paleoflora", "Stone products" and "Written sources to museum objects and collections".

Vernadsky State Geological Museum RAS (SGM RAS) (Moscow, Russia) has started work on development and implementation of The Museum Open Data Portal in 2017 (Figure 1).
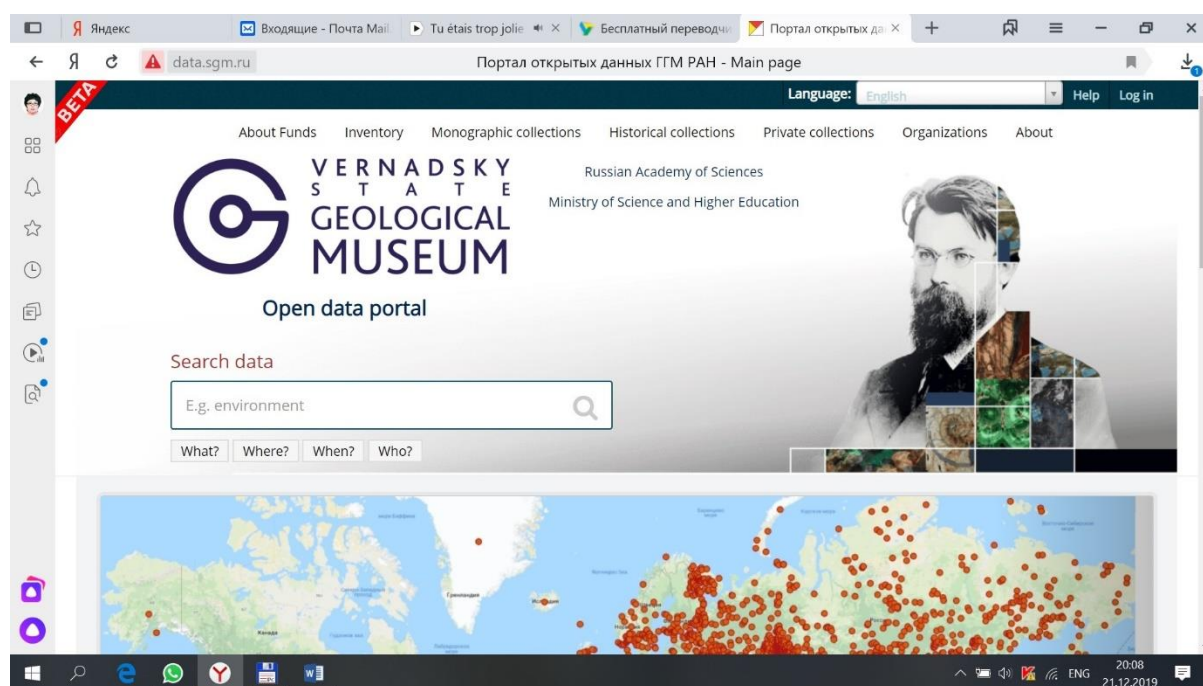

Figure 1: Home page of SGM RAS Open Data Portal

2 Open Data Portal

The Portal provides the user with the set of data online. It allows each scientist or professor/student to study, download or use SGM RAS data for research or education purposes.

The first step in this direction is the reorganization of the museum 's internal database. The base was created on Visual FoxFro in 1995. It is available only to museum employees with different levels of access to data. Its main task is to store and record data on the museum 's collections. The reorganization of the DB made it possible to correct the old data. When new data is entered, the samples are checked for uniqueness and possible errors are monitored. The search in the database is possible both by any of the available values and by any combination thereof. A variety of reports allow you to fully perform accounting functions. Verified data from the internal database, through file import, is sent to the Internet portal.

Within the framework of these works, photographs of stock data and scans of printed materials are organized. The Python programming language has implemented a program that allows to find objects in photos of exhibits and calculate the dimensions of these objects in pixels according to the data from the SGM database and then place the "ruler" on the image, which indicates to the user the real dimensions of the object.

Standards are one of the key components of museum data infrastructure. The introduction of standards ensures compatibility at the data and software level, avoids information loss and opens up new possibilities for data integration and joint processing (Antopolsky et al., 2012). The task being solved uses a Dublin Core-based format.

3 Software platform

As a software platform to implement the Portal, the CKAN platform is a powerful open-source data management system that makes data available by providing tools to optimize data, share, locate, present, and store it.

The CKAN platform is a new type of information systems - data management systems (DMC) based on the principles of "open access" and CODATA works. The storage object of such a system is datasets, i.e. tables. New data enters the system via the user interface or via metadata and OAI data exchange protocols. DOI naming technology is used to ensure the uniqueness of each dataset. Thus, data sets with a DOI name are uniquely identified and have a permanent reference for mention on the Internet and scientific publications.

The CKAN platform is a new type of information systems - data management systems (DMC) based on the principles of "open access" and CODATA works. The storage object of such a system is datasets, i.e. tables. New data enters the system via the user interface or via metadata and OAI data exchange protocols. DOI naming technology is used to ensure the uniqueness of each dataset. Thus, data sets with a DOI name are uniquely identified and have a permanent reference for mention on the Internet and scientific publications.

A module of geographical location name transformation into absolute coordinates was created for samples geographical location representation. OneStreetMap was used for this purpose.

The CKAN system, in addition to tables, is able to store and provide sets of images, web pages, text data, and PDF files.

All museum collections are distributed according to the corresponding inventory: "Minerals", "Rocks and Ores" "Fossil invertebrate", "Fossil vertebrate", "Paleoflora", "Stone products" and "Written sources to museum objects and collections".

Private collections contain samples, collected and presented by private citizens. The most he most significant private collections, stored in Vernadsky State Geological Museum are collections, collected in XIX - beginning of XX centuries and associated with the names of famous Russian state figures – count A.K. Razumovsky, count N.P. Rumyantsev, princes Gagarin, count A.F. Keller, P.G. Demidov. The special place in museum funds have Lidiya Prokhorova minerals collection and N.P. Vishnyakov fossil collection and also the modern Russian political figure Sergey Mironov mineral collection.

The most valuable in museums funds are monographic collections to published scientific articles and monographies, containing typical specimens of new fossil fauna and flora species and types. The collections published in XIX century by national paleontology and stratigraphy founders: Gotthelf Fischer von Waldheim, Charles Roulier, Hermann Trautschhold, A.P. Pavlov, S.N. Nikitin, A.A. Stuckenberg, M.V. Pavlova, V.O. Kovalevsky are among them. Vernadsky State Geological Museum stores more than 300 monographic collections at the present moment. Russian and foreign specialists have been transferring the collections to published works into museum funds, increasing the quantity of collections.

4 Data collections

The Portal provides an access to museum fund information online. Each collected is presented as a set of data. The data sets are a new digital data representation, containing the whole collection information.

The form of collection representation can be different: Internet references, pictures, exhibits description, written sources to museum objects and collections.

The exhibits descriptions are presented in table form. Internet users at the present moment have access to 6 basic data sets, the exhibits in which are described by 150 parameters. The collection is available for users in two modes "Initial view" and "Observer"

The "Observer" mode presents information in three ways: tables, diagram, map. "Table" mode reflects the whole list of exhibits. The search block and filters are fixed in the top right part. The search allows looking for exhibits in each field. The filters allow choosing one or several fixed values typical for exhibits. The search and filters complement each other.

The navigation block is in the center of the Control Panel allowing to switch pages and to change the quantity of records on one page. The switch into "Map" mode represents chosen exhibits by marks on interactive map. The switch on mark shows the exhibit information.

5 Search system

The built-in CKAN Storage and Organization Search System offers good array search capabilities. A study of the system found that the set of CKAN data search and processing operations actually only includes direct comparison operations.

We offer a solution for organizing a thematic search for all collections at the same time.

Search queries are built sequentially. At the first stage there is a search for information in the categories: "Who," What, "When," Where. " Search is carried out on:

- the author of the collection;
- the name;
- the age;
- the geographical location.

It is based on the use of the CKAN API system to obtain a list of data arrays and their attributes, to obtain the csv arrays of interest, to define the csv variety, and to search directly for the resulting data array. This solution is extremely flexible, reliable, and extensible, while search system performance remains satisfactory and memory consumption is minimal through the use of Python generators.

Having received a large number of results in the initial search, we have to clarify it. Search results are currently refined, usually by facet search methods. The results found are immediately distributed to groups (facets) according to predefined characteristics. Often their number in each facet is counted, and the user immediately understands in which direction the search should be narrowed.

The search in the System is available in two ways for users. In the first case, in reply to a user text inquiry, the system gives a list of collections containing the search phrase in its description. The general list has a collection name, brief description and resources types, attached to this collection: table information, exhibits images, publications and monographies describing this collection. The second case suggests thematic search on each collection simultaneously. The search inquiries are built consistently. "Who", "What", "When", "Where" categories are available on the first stage.

The category "What' provides search on the name of the deposits, minerals, rocks and ores. The selection of the value is available out of the specified list. In reply to the request the Portal displays a list of all collections containing the search phrase.

The category "Where" is a common geographical search. The coordinates are to be set by hand or to set a region on an interactive map. The search on countries, geographical regions, districts etc. is also available.

The category "When" allows to search on stratigraphic age. The category "Who" searches on registration authors, donators or collection authors. The system allows a complicated serial information search request.

6 Conclusion

Vernadsky State Geological Museum RAS is planning to assign DOI names to its Open Data Portal in nearest future. It will give possibility to achieve uniqueness of each data set on Open Data Portal. The data sets with DOI names will be identified and will have constant reference for Internet and scientific publications.

The authors of the project assume that the standards of metadata and protocols applied in the construction of the Open Data Portal of the Russian Academy of Sciences will be used in the future to integrate the stock data of the Russian Academy of Sciences into the world network of natural science data.

The basic principles of development of Vernadsky State Geological Museum Open Data Portal are described in Sergey Cherkasov, Vera Naumova and others article [Che18].

The Portal test version is available at the http://data.sgm.ru

The Help with Portal operation is available at the http://data.sgm.ru/help/.

References

[Ana14]   Analytical brief "New information and communication technologies in museums developments" within the frames of UNESCO/MFGS project "Museum management – XXI century: thematic regional trainings UNESCO/ICOM and meetings of SNG countries experts on development of potential in the area of museum management" (2011-2014), Moscow, 2014, 21 p.

[Mal14]   Malyshev Yu.N. A New Stage in Museum Life // Gornaya Promyshlenost, 2014, №4 (116), p. 124

[Che18]   Cherkasov S.V., Naumova V.V., Platonov K.A., Dyakov S.E., Eremenko V.S., Patuk M.I., Starodubtseva I.A. The Basic Principals Of Development Of Open Access To Data Of Vernadsky State Geological Museum Of RAS // Information resources of Russia, 2018, №4 – p.9-14