

ELECTRONIC PUBLISHING FOR ACADEMICS: GEOLOGY AND PHYSICS IN RUSSIAN NETWORK FOR NATURAL SCIENCES

PAVEL PLETCHOV²; ALEXEY KRASHENINNIKOV^{1,*}; ALEKSEI SELIVERSTOV¹;
SERGEI TRUSOV²; VADIM USTIANSKY³

¹ Physics Department, M.V.Lomonosov Moscow State University
119992, Moscow, GSP-2, Leninskie Gory, Russian Federation
avs@phys.web.ru; acrush@phys.web.ru

² Geology Department, M.V.Lomonosov Moscow State University
119992, Moscow, GSP-2, Leninskie Gory, Russian Federation
pavel@cs.ru; serg@students.web.ru

³ Sternberg Astronomical Institute, M.V.Lomonosov Moscow State University
119992, Moscow, GSP-2, Leninskie Gory, Russian Federation
ustiansk@sai.msu.ru

* Corresponding author: acrush@phys.web.ru

Electronic publications have proven to be crucial for supplying the scientific community with up-to-date research results and instruments for both education and research. Present work summarizes the experience of development and maintenance of two natural science oriented servers and their integration with related databases and web-resources. The Russian network for natural sciences pursues the aim of providing the scientific society of the CIS states with high quality tools for peer-reviewed scientific publishing in Russian and other languages with full public availability and an open archive of handbooks and textbooks on geology and physics. Web-based software for scientific calculations, forums, various information services and an ample catalogue of related scientific resources combined with search and classification tools are also presented on the Network. Search forms for geology and physics resources freely distributed among interested third-party sites and means of mutual information exchange based on XML serve the aim of deep integration of scientific resources. Designed and launched with the support of Moscow State University in 1997 and 2000 respectively, the Geo.Web.Ru and Phys.Web.Ru servers provide users with thousands of scientific and popular books, articles and sketches on physics (over 2,200 documents) and geology (over 4,700 documents). The servers are by now among the top-ranked resources for Russian-speaking scientific auditorium according to usage statistics.

Keywords: Portals; Catalogues; Search; Resource Integration

INTRODUCTION

Electronic publications have proven to be crucial for supplying the scientific community with up-to-date research results and instruments for both education and research. The development of the coming generation of scientists requires a deep involvement of students in the scientific information exchange. For beginners it is also important to get the full picture of modern scientific “mise en scene”. This can only be achieved by combining the latest reports with ample reference sources and demonstrations presented in the native language. Providing a wide range of open access resources prepared by experts could foster the intellectual growth of future scientists.

This approach was adopted for development of the natural science oriented servers: “About Geology” (<http://geo.web.ru>) since 1997 and Scientific-educational server on Physics (<http://phys.web.ru>) since 2000. The projects, as conceived from the beginning, aim at providing the

scientific audience with high quality publications submitted by professionals in science and education, with educational content such as textbooks and glossaries.

Historically, the geological server has several names, one of which provides a deeper understanding of its concept: the Russian Virtual Geoscience Network. This implies the effort of the developers to create a portal able to unite the many aspects and realizations of electronic Earth science resources. Later, with the development of the physics branch, the complex has evolved into the Russian Network for Natural Sciences as stated in the header. Moreover, for every scientist it is very important to get the broadest possible view of the investigated problem. Hence, there is a high demand for relevant topical search engines that filter off the background noise level documents resulting from common vocable search requests. This problem may be solved by creating a topical hierarchical resource catalogue combined with means of indexing and search. This has been done for the first time in the area of Earth sciences and physics.

Also, it is well-known that in most cases the pre-publication work of a scientist usually includes wide discussions of preliminary results, which in modern world substantiates the demand for professional communication tools such as forums. Such are provided to users at the geology server and are prepared for launch at the physics server. As for educational component, the servers focus on electronic publishing of popular articles and textbooks including those compiled of lectures read at the faculties and departments of M.V.Lomonosov Moscow State University and other higher education institutes.

The has evolved from There still remain a number of sections that appeared then: a forum for students, a mirror of a bard song collection, etc. But just as it has previously happened to Yahoo!, very soon the collection of hyperlinks grew beyond the level of just a "See also" page and demanded a development of rubricator and search engine.

In 1997 the authors of the Educational network project [1] (then a website devoted to students of the Geology Department of Moscow State University and a collection of links to geological resources) have applied for a Russian Foundation for Basic Research grant. The aim was to create a tool for automated collecting and sorting of references to various geological and geophysical databases and sites, which could be used by virtually any scientist worldwide. In the project called Russian Virtual Network for Earth Sciences a web-enabled object model has been applied to organization of Earth science information resource for the first time.

In 2000 the physics section of www.nature.ru Scientific Network, the latter also developed with active participation of the authors, peeled away to form the Phys.Web.Ru Scientific-Educational server on physics [2]. Based on similar software and ideology, the servers have united to form the Russian Network for Natural Sciences.

USER AND AUTHOR/EDITOR INTERFACES AND EDITORIAL POLICIES

The user interface is developed using a modular structure to allow same script files to serve both readers and chief editors. The capabilities of the interface depend on user role, i.e. the set of rights the user is vested with by server administrator or automatically. When entitled additional rights new components become available to user. At maximum, it's possible to create and edit article text (including LaTeX-based formulae typing) and meta-information, publish articles for common access, change the rubricator section names and their structure, edit keywords, user's personal information, user roles and availability of special administrative applications.

The editorial policies of the servers differ in details, but have common features besides requirements of legislation currently in force. Scientific articles and reviews published have to

convey scientific knowledge, contain original results or references to such. It is advisable that the author be professional in science, teaching or scientific journalism. The copyright belongs to the author and the publisher is granted the permission to reprint.

BASIC DETAILS

Technologically, both of the presented servers represent a three-level architecture. Upper level consists of a necessary number of external interfaces for data representation, import and export. Intermediate level contains the common information and operations bus with a standard unified interface. Lower level consists of an arbitrary number of data storage and processing units (devices). Such architecture has many advantages for large project development, of which one of the most important is the scalability. Additional servers may be plugged in on the run to maintain uniform load distribution and overall performance of the system.

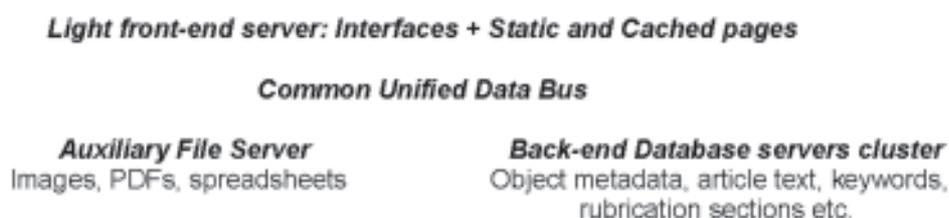


FIGURE 1. DATA STORAGE AND PROCESSING ARCHITECTURE OF THE SERVERS.

The documents in the network are supposed to be viewable to all users virtually independent on the software they dispose. Therefore, the general representation is based on standard HTML with automatically pre-generated GIF insets of LaTeX formulae and other mathematical constructions. In certain cases PDF files of papers are additionally presented as print-versions.

DATA CLASSIFICATION, SEARCH AND REPRESENTATION FOR NATURAL SCIENCE RESOURCES

During the recent years a unified basis for data classification, its search and representation in the Internet are requested. Very frequently geo-information systems (GIS) or cartographic systems are proposed for geological data. However, there is an enormous amount of information unfit for cartographic representation.

Practically all Earth science and physics information flows represented in the Internet may be referred to the following types:

- a) descriptive (articles, books, lectures);
- b) event-related (monitoring, news, conferences);
- c) debatable (considerations, questions/answers);
- d) reference (databases, catalogs, libraries);
- e) interactive resources (simulations, specialized calculations, demonstration programs).

Descriptive, event-related and debatable information flows are served well with conventional content management systems (CMS). Such systems are implemented in many scientific content servers. For example, such Russian-language based CMS maintain in part the information flow processing in <http://Geo.Web.Ru> and <http://Phys.Web.Ru>. The data of types a), b), c) can be rendered in a ‘pseudo-static’ way which allows further integration to the Internet by means of reference, indexing and search. Such integration implies availability to various search engines such as global resource crawlers, local search and internal navigation.

The information on servers is classified using a complex expandable rubrication scheme including a linear branch for publication genre (scientific articles, popular articles, news, annotated references, event announcements and so on), a linear branch for glossary articles (alphabetical) and a hierarchical subject-field tree. The editorial board may introduce more rubrication branches if necessary, for example to sort documents according to target audience. This helps the users to find exactly what they want even if they do not know how to compose a yielding search query. In practice it is convenient to select articles using both branches (certain message types on certain subjects).

Local full-text search within the articles, glossaries, books and other materials stored in the local databases of servers is based on OpenFTS technology implementing the Generalized Search Tree (GiST) concept. These have been developed expressly for Scientific Network member servers and are included in the PostgreSQL since v. 7.3.

The first implementation of a system with additional dynamic links between information objects is represented by the Russian Virtual Geoscience Network (<http://Geo.Web.Ru>). For the first time, to our knowledge, in the natural science area we have developed a complex of search tools much simplifying the use of servers and easing the access to related documents. Among such we should mention the dynamic hyperlink model [1] where special simple markup rules allow the author to refer to the latest and the most relevant data available on the Network. This is achieved by treating some particular types of anchor tag’s href attribute as search queries handling a number of parameters including search keywords, rubricator sections from which the search is requested, distance between keywords in the target document, their position in the document and more. Also, in cooperation with the Scientific Network (www.nature.ru) and Russian Astronomy Network (www.astronet.ru) we have for the first time (at least in Russia) developed a tool for selection-based searching. We consider these as features of a prototype of unified specialized search system at a national scale which can be interfaced with Semantic Web in the future.

External search is based on *X-Ware* technology [3] – a modified search engine of the www.rambler.ru portal – adapted for scientific term stemming, morphology enabled search and topical catalogue based search. Both internal and external search requests are handled with the same forms available in every page of Phys.Web.Ru and Geo.Web.Ru.

RESOURCE INTEGRATION AND TOOLS FOR PROFESSIONALS

The servers of the Russian Network for Natural Sciences try to provide their users with the broadest possible view of physics, geology and geophysics. The means is probably the wide free distribution of search forms allowing for search over all the resources included in the topical catalogues of Geo.Web.Ru and Phys.Web.Ru. There is an active exchange of publications between the two servers and with third-party resources such as www.astronet.ru, www.nature.ru, www.exponenta.ru, etc.

The concept of Earth science and physics resource integration adopted a self-registration approach combined with automated immediate error-checking services and database update robots that help to keep the contents actual. The registration processing included checking of formal requirements imposed by the database structure and expert review of resources submitted to avoid unsolicited non-topical (misleading) links. The submitted information was then placed in queue and automatically expired after two weeks if not accepted for use (inclusion) in the Network by the experts. Expiry of long-down (unavailable for more than two weeks) or for some reason irrelevant database records allowed to maintain reasonable search performance and update rates on noncommercial web-servers provided by the Geology and Physics Departments using the institutional Internet-channel of the University. All errors including expiry as well as any remarks made by the reviewer were automatically reported to submitters by e-mail. The aim was to help authors improve their resources and prepare them for inclusion in the Network, i.e. to create a data representation scheme suitable for indexing with the Network's crawler to make it available through a unified interface. Upon acceptance the submitted resources were indexed and included in the collection of records available for the search engine.

Several software tools have been developed for experts maintaining the topical link-lists referring to geological or geophysical resources, where all the corresponding database records are stored in the same pivot table. Still the links can emerge in an arbitrary number of link-lists and be classified using rubrication of any complexity. The same way the references are periodically cron-checked for availability and removed if unavailable for a long period of time unless such removal is explicitly prohibited by the server administrator.

Integration with various databases and textual information resources is further deepened by inclusion of auxiliary semantic links to relevant documents and search queries in Geo.Web.Ru or Phys.Web.Ru from the database records. For example, such links are provided to the users of the Crystallographic Database for Minerals and their Structural Analogues (<http://database.iem.ac.ru/mincryst/>, includes 2585 items), the "WWW-Mincrust", part of the Russian Virtual Geoscience Network project. The links normally lead to search pages resulting from an automatically generated query including russian and english names of minerals, their spatial symmetry groups, etc.

The geology server also provides its users with online scientific calculation tools presented by numerous geological, petrological, and thermodynamical programs with Web-interface. Among them are:

- Comagmat – a powerful complex of related programs developed to calculate phase equilibria for dry and hydrous natural magmas crystallization. (<http://geo.web.ru/~kbs/>)
- Geothermometers and geobarometers that calculate temperature and pressure of mineral equilibria from the mineral's chemical composition using various petrological models. (<http://geo.web.ru/db/calc/>)
- Regional mineralogy database – collection of 2000 minerals and 7000 corresponding finding sites (<http://geo.web.ru/druza/>)
- Meteorite database provides information about more than 1000 meteorites from Russian Academy of Science collection: meteorite names, finding/fall locations, chemical composition and more (<http://geo.web.ru/db/meteorites/>) The set of instruments for the users of discussed servers also includes such services as event scheduler (information on conferences and schools held on a national or/and a global scale including publication of selected proceedings: http://geo.web.ru/db/full_year.html and http://phys.web.ru/db/full_year.html) and a hierarchical resource catalogue (annotated links to related materials in Russian as well as in other languages).

As for integration, the availability of different databases and interactive resources on the Internet gives rise to both technical and conceptual difficulties. The main problems include different (often incomparable) structures of data, lack of standards for specialized information representation, the 'variety' of database interfaces and differences in tasks of information compilers. In the opinion of the authors, development of auxiliary semantic links of different purposes between data objects is a promising trend in the evolution of common information space.

PEER-REVIEWED PUBLISHING, EDUCATION AND INSTITUTIONAL ARCHIVING

As conceived, the Geo.Web.Ru and Phys.Web.Ru provide the users with high quality peer-reviewed scientific publications and reference data. The news and popular articles are written by professionals in science and education and reviewed by associates of Geology and Physics Departments of Lomonosov Moscow University. In 2001 the Geology Department included the Geo.Web.Ru server in the obligatory mailing list for PhD and professor theses' abstracts. Currently the possibility of providing Phys.Web.Ru with the status of the official e-prints server of the Physics Department is discussed with the administration.

The administration of the Russian Network for Natural Sciences pays much attention to development of the coming generation of scientists that requires a deep involvement of students in the scientific information exchange. The Network focuses on electronic publishing of textbooks including those compiled of lectures read at the Faculties of Lomonosov Moscow State University. The geology server has a specialized interface for this task that classifies the lecture courses present on the server into geological specializations (majors) and years of study (<http://geo.web.ru/db/edu/>). By now there are 12 lecture courses, 10 full textbooks and numerous tests on geological subjects and 13 lecture courses and 10 full textbooks on physics and astronomy. The collections of educational materials and demos are constantly expanded.

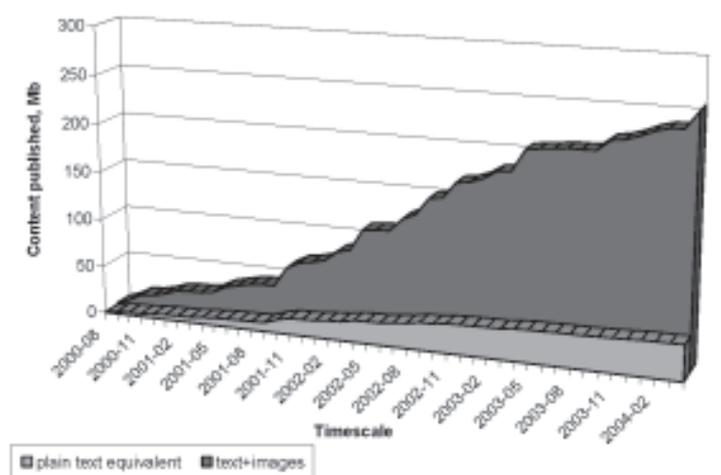


FIGURE 2. GEO.WEB.RU CONTENT DATABASE FILLING DYNAMICS.

USAGE STATISTICS AND CONCLUSIONS

The contents of the Network for Natural Sciences are intensively filled up with original articles submitted by independent authors and permitted reprints of third-party materials. The

dynamics of content growth for 'All about Geology' server (geo.web.ru) is well illustrated with Fig. 2. The content volume may be estimated as several tens of Megabytes per server in plain text equivalent (excluding HTML tags and additional files). Up to 5 Megabytes (1 Megabyte average) of textual information are published every month by the staff and independent authors using the central servers' databases only.

Usage statistics provided by Rambler's Top100 (<http://top100.rambler.ru/>, Figs. 3 and 4) show observable increase in numbers of visitors and page hits per month. The falloffs observed generally correspond to summer vacations period or occasional technical problems with servers or internet-access lines. The trends show a gradual increase of interest to the Russian Network for Natural Sciences.

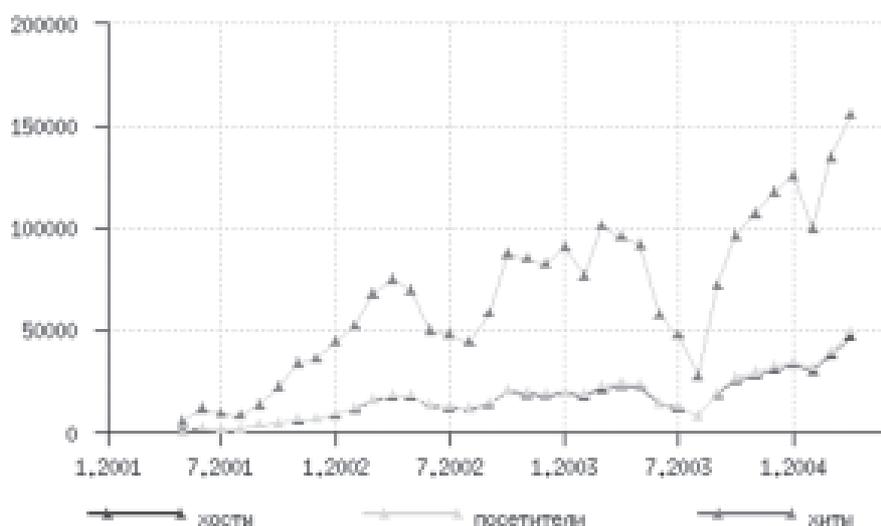


FIGURE 3. GEO.WEB.RU PUBLIC RESOURCE USAGE STATISTICS ACCORDING TO RAMBLER'S TOP100. DARKER CURVE ON THE BOTTOM SHOWING VISITOR HOSTS (UNIQUE IPS), THE LIGHTER CURVE ON THE BOTTOM – VISITING USERS, AND THE UPPER CURVE – NUMBER OF PAGES (URLS) REQUESTED PER MONTH.

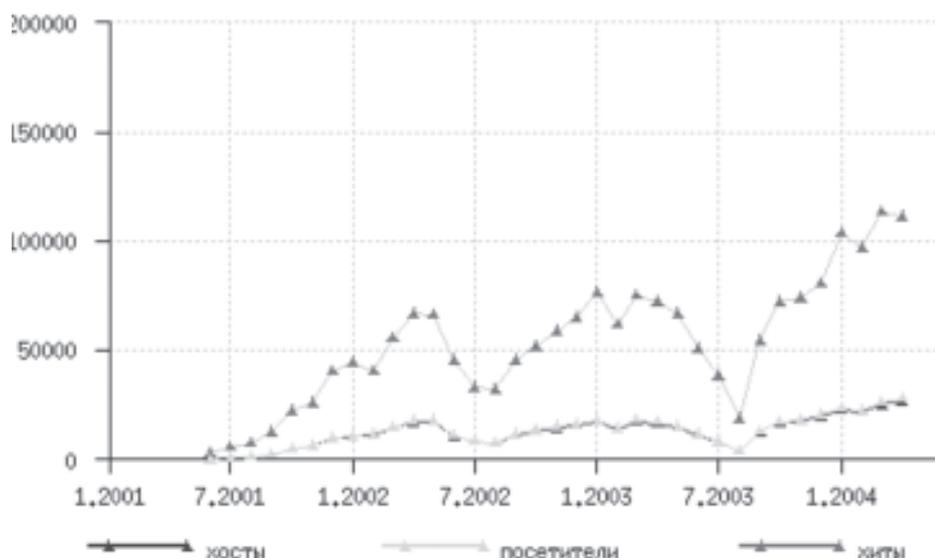


FIGURE 4. PHYS.WEB.RU PUBLIC RESOURCE USAGE STATISTICS ACCORDING TO RAMBLER’S TOP100. DARKER CURVE ON THE BOTTOM SHOWING VISITOR HOSTS (UNIQUE IPS), THE LIGHTER CURVE ON THE BOTTOM – VISITING USERS, AND THE UPPER CURVE – NUMBER OF PAGES (URLS) REQUESTED PER MONTH.

The positive experience described above shows that virtual networks uniting multiple resources around central content oriented servers are in high demand among the scientists of the CIS and of the entire world. Still remaining focused at Russian-speaking audience, given the presence of native English speaking experts in the editorial boards the servers may be expanded with English interfaces and English versions of publications to improve the availability of the Network’s resources therein to foreign scientists. We conclude that the reported approach is well applicable to electronic publishing for academics.

ACKNOWLEDGEMENTS

This work was partially supported by Russian Foundation for Basic Research (project 03-07-90257).

NOTES AND REFERENCES

1. Pletchov P.Yu. and Koshchug D.G. *Educational network of the Geological faculty of the Moscow State University*. In IUGG XXII General Assembly, Birmingham, UK, 1999.
2. Seliverstov A.V., Krashennnikov A.P. *Scientific-educational server on physics Phys.Web.Ru*. In Conference “Scientific Service in the Internet 2002”, Novorossiysk, Russia, September 23-28, 2002, p.194 (in Russian).
3. I.E.Panchenko, A.V.Seliverstov, V.E.Ershov, I.V.Kutyrev, V.O.Ustiansky. *X-Ware platform: building scientific and educational sites, distributed editorial boards and information systems*. In Conference “Scientific Service in the Internet 2003”, Novorossiysk, Russia, September 24-29, 2003, p.309 (in Russian).
4. P.Pletchov, S.Trousov, D.Varlamov, K.Bychkov. *Prototype of the logical basis for geological data representation in Internet*, In XXIII General Assembly of the International Union of Geodesy and Geophysics, Sapporo, Japan, June 30, 2003 - July 11, 2003, vol.B, p.524.