

# A Virtual Global Magnetic Observatory Network: VGMO.NET

Vladimir O. Papitashvili<sup>1\*</sup>, Valeriy G. Petrov<sup>1†</sup>, Anshuman B. Saxena<sup>1‡</sup>, C. Robert Clauer<sup>1</sup>, and Natalia E. Papitashvili<sup>2</sup>

<sup>1</sup>Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

<sup>2</sup>Space Physics Data Facility and QSS Group, Inc., NASA/GSFC, Greenbelt, Maryland 20771, U.S.A.

(Received April 30, 2005; Revised September 12, 2005; Accepted September 12, 2005; Online published June 2, 2006)

Since the mid-1990s, many geomagnetic datasets began appearing on the World Wide Web. Often these data were not submitted to the World Data Centers as recommended since the International Geophysical Year (1957–58). As a result, existing data become naturally distributed creating an urgent need for more sophisticated search engines capable of identifying and retrieving the data from Web for scientific analyses. We introduce a Virtual Global Magnetic Observatory concept for “pulling” geomagnetic data distributed worldwide. The VGMO moves information exchange from simple file transfers to a higher level of abstraction, forming the distributed databases through establishment of self-organizing data nodes. We present the first working VGMO prototype: an “on-demand” objects-building Internet application that is transparent in its internal data management to the external users/clients. The VGMO server continuously builds data-objects only from client requests by going through a pre-set list of Web-based data nodes (including WDCs). As the retrieved data are added to the server (or node) database, future requests of the same interval would not force a new Web search. Furthermore, new nodes can be made available to others through the VGMO network, building the worldwide geomagnetic data “fabric” in a platform-independent and location-neutral environment of newly “webbed” digital data. Application is on a server at <http://mist.engin.umich.edu>.

**Key words:** Geomagnetic field, data management, virtual observatory.

## 1. Introduction

With the unprecedented increase in the supply and demand for scientific data, today scientists must put a large amount of their research effort into identifying and acquiring relevant datasets. Centralized data repositories often serve as a single point of acquiring and disseminating scientific data. One such central data repository is the World Data Center System (Guide to WDC System, 1996), which supports international exchange of geophysical observations in accordance with principles set forth by the International Council for Science at the time of the International Geophysical Year (IGY, 1957–58). These Centers collect data and publications for various geoscience disciplines: glaciology, meteorology, oceanography, space sciences (from rockets and satellites), solar-terrestrial physics (solar and interplanetary phenomena, solar flare associated events and corresponding ionospheric phenomena, geomagnetic variations, aurora, cosmic rays, airglow), and solid-earth geophysics (seismology, gravimetry, tsunami, marine geology and geophysics, recent tectonic movements, Earth’s rotation and tides, paleomagnetic and archeomagnetic measurements, volcanology, geothermic observations).

One of the important issues in maintaining data collec-

tions in the WDCs is the need to acquire newly obtained observations. By the above-mentioned principles, the data providers (i.e., observers at the discipline observatories) must submit their data to the appropriate data center. This type of the data collection can be called as a “push data” paradigm: the data will not appear in the public domain (i.e., in the data centers) unless being “pushed out” by the data providers (Fig. 1, two left panels). The nature and formats of the recorded data and their exchange are specific for each discipline; thus, the data resulting from experimental setups are often found unsuitable for immediate submission to the WDC.

For example, the World Data Centers accept both variation/research data and absolute observatory data, but the data must be clearly identified as to which type they are to be archived and distributed. However, the quality of variation data does not often satisfy the WDC criteria (set mainly from and for the standard magnetic observatory data); therefore, the geomagnetic variations data submitted to the WDCs may not be necessarily archived there.

Furthermore, since IGY the data submission to WDCs remains voluntary; as a result, data providers often try to avoid additional steps in data processing needed to satisfy the WDC standards, and instead choose to publish and exchange the data on their own, especially with the advent of the Internet and World Wide Web in 1990s. Therefore, nowadays to get scientific data from various, mostly distributed sources, a scientist may have to search through a number of World Data Centers and various research or government institutions running geomagnetic observatories in many countries. The latter requires establishing personal

\*Also at National Science Foundation, Arlington, Virginia, U.S.A.

†Also at IZMIRAN, Troitsk, Moscow Region, Russia.

‡Now at TATA Consultancy Services, Bangalore, India.

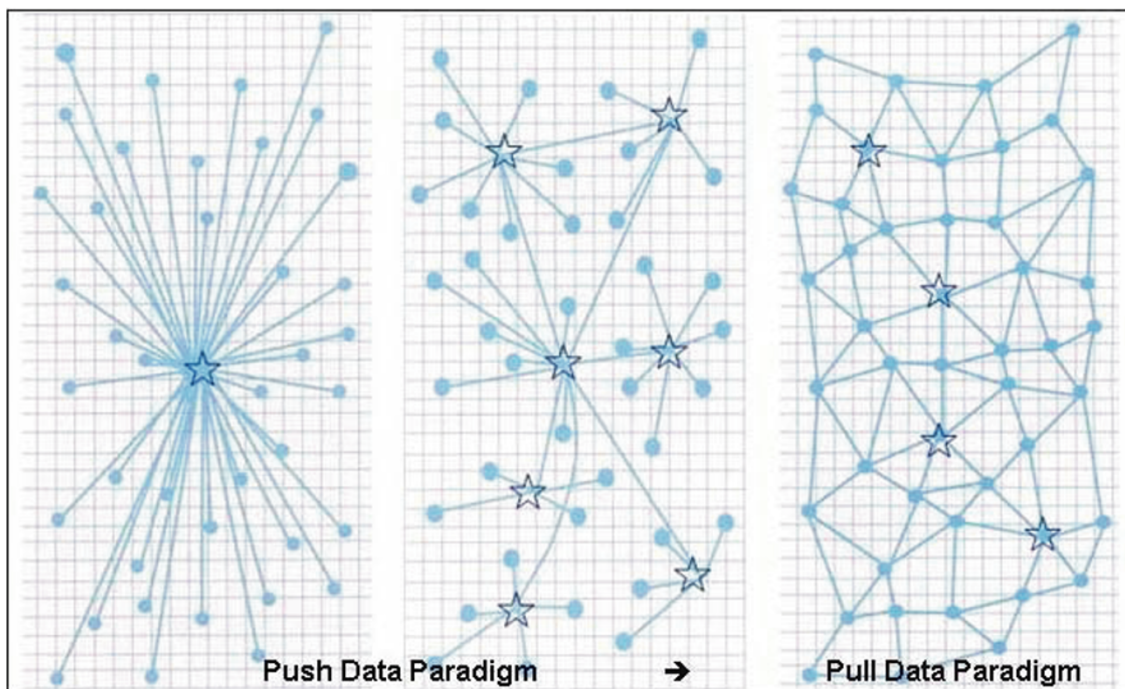


Fig. 1. Graphical representation of the old “Push Data” concept (two left panels) and its development into a new data collection and dissemination paradigm of “Pulling Data” from the Internet. Diagrams shown in the figure background have been initially designed by the RAND Corporation to illustrate propagation of the messages through communication networks (Campbell, 2004).

contacts promising some sort of acknowledgements for the use of requested data. As the required data are identified and received (via snail-mail, airmail, e-mail, or Internet’s “file transfer protocol” FTP), only then the data can be ingested into a local database and processed using mostly user-built codes. Searching and then converting data from various formats into something useful for local processing and analysis is often very time consuming, labor intensive, and sometime frustrating if the collected data do not deliver what the scientist expected in the beginning of the data search.

Thus, until recently digital geomagnetic data were exchanged mostly through personal communication—on floppy or optical discs, via customized FTP access, etc. This imposed a significant communication overhead—delays between responses, need to develop more contacts, and any changes in the structure of remote databases would require renegotiation for the access. Since mid-1990s, with the rapid growth of Internet and increased power of desktop personal computers, many research groups have started publishing their data through the World Wide Web; in fact, more datasets are now available on the Web as compared to the initial WDC holdings. By publishing data through the Web, one not only avoids additional work for submissions to WDCs, but also achieves greater visibility amongst scientific user communities. Although it seems that the described trend puts more and more geomagnetic data into the public domain, it poses a significant challenge for the efficient data sharing, creating a need for more sophisticated search engines capable of identifying, cataloguing, and accessing the “webbed” databases for scientific analyses.

In 2000, we built a system that provided access to some

Greenland and Antarctic magnetometer data via the Polar Conjugate Facility deployed at the Danish Meteorological Institute, DMI (Papitashvili and Papitashvili, 2001; Papitashvili *et al.*, 2002a; [http://web.dmi.dk/fsweb/projects/pcf/pcf\\_form.html](http://web.dmi.dk/fsweb/projects/pcf/pcf_form.html)). Although this system was “hardwired” to specific, proprietary formats of geomagnetic data stored at DMI and other participating institutions, the PCF allowed us to test successfully a “local/remote access” approach that led us to the VGMO.NET concept.

In this paper, we present the design philosophy and concept of an Internet-based middleware application, a Virtual Global Magnetic Observatory, identified as a distributed network of the data nodes (VGMO.NET) randomly interconnected via Internet from which the data are simply “pulled out” only when requested (Fig. 1, right panel). This approach constitutes a new “pull data” paradigm where data reside on their “provider” servers, but the VGMO.NET (as a whole; Fig. 2) identifies all available geomagnetic data repositories, manages access to the remote databases, and performs online processing of the requested or newly acquired data (Papitashvili *et al.*, 2002b, 2003). The VGMO.NET concept moves information exchange from a simple file transfer to a higher level of abstraction, forming the “worldwide fabric” of distributed geomagnetic data through establishment of the self-populating and self-organizing data nodes. The middleware architecture of VGMO.NET provides to the worldwide geomagnetic community a new way for sharing data and functionality in a platform-independent and location-neutral environment (Papitashvili *et al.*, 2004).

The VGMO.NET model can be easily applied to the standard holdings at the World Data Centers, as well as to

### VGMO.NET Data Collection/Distribution System

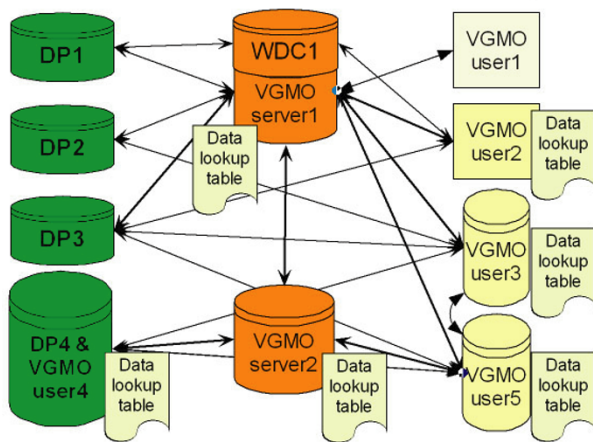


Fig. 2. The VGMO.NET architecture unleashed. The VGMO server can be deployed at any node of the data network (“data fabric”) shown in Fig. 1 (right panel): these could be the WDC data servers (shown above as WDC/VGMO server1) or any of the users’ data servers (VGMO server2), etc. These servers interact with the local and remote databases (shown as DP<sub>x</sub>) which could be independent or collocate with the VGMO’s nodes. The VGMO users can be “passive” (i.e., using the VGMO.NET but not opening their own data to the network; see “VGMO user1” above), or “active” users contributing to the data network (e.g., “VGMO user2, 3 and 5” above).

other geomagnetic data collection systems like the INTERMAGNET (<http://www.intermagnet.org>), Circum-pacific 210° Magnetic Meridian Network (<http://stdb2.stelab.nagoya-u.ac.jp/mm210/>), magnetometer arrays in the Arctic (<http://space.augsburg.edu/space/>; <http://bluebird.phys.ualberta.ca/carisma/>; <http://www.ava.fmi.fi/image/>) and Antarctic (<http://mist.engin.umich.edu>; <http://dabs.nerc-bas.ac.uk/dabs/>)—to name just a few.

The proposed “pull data” concept forms a basis of the Electronic Geophysical Year (*eGY*) initiative (<http://www.egy.org>), which sets out the principles of geoscience data stewardship for accomplishing a 21st Century vision of an “Information Society” where human beings share the benefits of free and open access to data and information. With *eGY*, these principles are enunciated in a geoscience context, but they are applicable across all of the sciences and beyond (Baker *et al.*, 2004; Papitashvili *et al.*, 2005a, b, c).

## 2. Related Efforts

Several methods have recently been proposed for efficient sharing of scientific data. Four open-source data management systems, namely the International Virtual Observatory Alliance (IVOA; (<http://www.ivoa.net>), Virtual Solar Observatory (VSO; <http://vso.nso.edu>), Live Access Server (LAS; <http://ferret.wrc.noaa.gov/Ferret/LAS/>), and Distributed Oceanographic Data System (DODS; <http://www.unidata.ucar.edu>) are the most prominent efforts in networking the astronomy, space, and geosciences data. The IVOA mission is to facilitate international coordination and collaboration necessary for the development and deployment of the tools, systems, and organizational structures necessary to enable international utilization of astro-

nomical archives as an integrated and interoperating virtual observatory (VxO).

The VSO is a new tool for investigating the physics of the Sun and its impact on the Earth environment. This tool allows addressing one of the central challenges of solar research: the need to locate, correlate, absorb, and analyze data from a wide array of scientific instruments that measure the Sun on spatial and temporal scales that range over seven orders of magnitude. The VSO is a scalable environment for searching, integrating, and analyzing databases distributed over the Internet. A key element of the VSO is an integrated data-mining and analysis capability that can be applied both across and within databases.

LAS is a highly configurable Web server designed to provide flexible access to climate scientific data by enabling the data provider to: (a) unify access to multiple types of data in a single interface, (b) create thematic data servers from distributed data sources, (c) offer derived products “on-fly”, (d) remedy metadata inadequacies if data are poorly self-described, and (e) offer unique products such as visualization styles specialized for the specific data.

DODS is a middleware that provides network access from the OPeNDAP-enabled clients machines to the data residing on the OPeNDAP-enabled servers (<http://opendap.org>). Applications can be transformed into the DODS clients, and thus achieve access to the remote datasets by linking them with the Application Program Interfaces (API; a set of routines, protocols, and tools for building software applications) provided in the DODS API toolkit. The DODS uses a “client-server” model where a client sends a data request across the Internet to a server; the latter answers with the requested data. The DODS server manages a set of scripts that are specific to the format of the dataset it serves. A typical DODS script fetches a selection of data from the available databases, converts these data into a binary format, packages it with some descriptive information, and then sends to the client.

However, the data management system most close to the VGMO described in this paper is the Space Physics Interactive Data Resource (SPIDR), designed “to allow a solar-terrestrial physics customer to intelligently access and manage historical space physics data for integration with environment models and space weather forecasts” (<http://spidr.ngdc.noaa.gov/spidr/>). The SPIDR consist of a distributed network of synchronous (but historical, already submitted to WDCs) databases and Java middleware servers accessed via the World Wide Web. The SPIDR main goal is to enable easy data mirroring between the World Data Centers (currently located in the U.S.A., Australia, China, Japan, Russia, and South Africa) to eliminate “the network bottlenecks associated with transcontinental links”.

We note that the above-mentioned approaches present distributed datasets as a unified virtual database by using specialized networking tools on a server’s side. Our approach differs from the systems described above in the underlying assumption about location of data. We assume that a large amount of data is published through the World Wide Web, and our aim is to provide a unified interface to this geographically widespread data repository. For example, in addition to accessing free http/ftp servers, we can in-

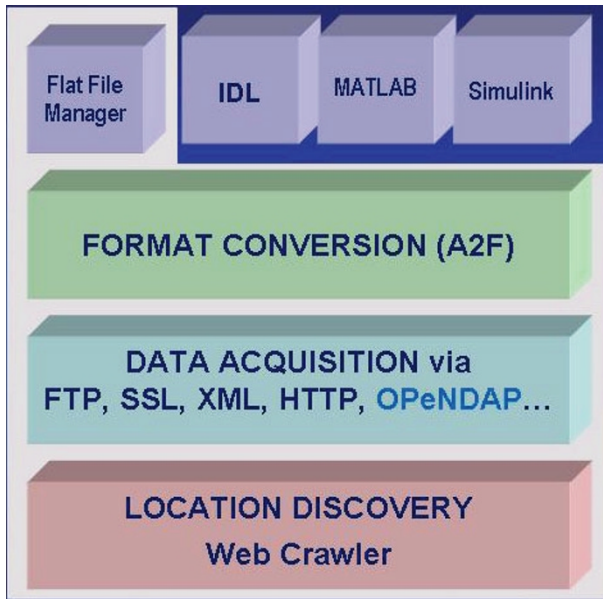


Fig. 3. The layered architecture of VGMO.NET concept. As shown, three lower layers constitute the internal part of the VGMO node, managed by the Flat File Manager atop. The current version of the FFMN is written in Fortran 90, but it can be replaced or amended by any of the high-level programming module(s) as shown above. The “manager” module organizes the data requests and puts the retrieved data into the temporary flat file, which will later be provided to the user who requested the data. In this, the “manager” provides various graphic and browsing functionalities (even some specific data manipulation functions), helping the user to make an intelligent decision on the usefulness of the retrieved data.

corporate into VGMO.NET the access to proprietary data repositories located at the OPeNDAP-enabled servers. The SPIDR system has all potential capabilities and functionalities to become one of the VGMO.NET portals if the SPIDR operators adopt the “pull data” collection paradigm.

### 3. VGMO.NET System Design

Our design objectives for the VGMO.NET concept were the following: (a) to identify prospective geomagnetic data repositories; (b) to provide transparent access to the remote geomagnetic databases through a common Web-based, Java-enabled interface; (c) to perform online processing of the acquired data sets; (d) to provide the ability to construct self-organized geomagnetic databases on individual, user-owned machines; and (e) to provide the ability to feed the collected data in various geospace models. These self-organized data sub-centers can then be made available to other users through future request chains, helping to build a “fabric” of data archives and crosscut access points.

As shown in Fig. 3, VGMO.NET has a four-tiered architecture. The lowest layer of the system forages the Web for the known and new geomagnetic data repositories. The heart of this “location discovery” module is an http-based “GeoMagnetic Crawler” (GeoMaC) that provides an access to the known geomagnetic data repositories for the further data downloading upon the VGMO.NET user requests. The Data Acquisition Layer (DatAc) relies on GeoMaC for obtaining the information required to interact with the remote data repositories. The DatAc first tries to service the received request for data locally; however, if the requested

```

Name of header and data files:      VOS01
Date files created:                13-May-2002
Record length of data file, in bytes: 20
Number of columns:                 4
Number of rows:                    3137310
Flag for missing data:             -0.10E+33

```

#	name	units	source	type	loc
1	Time	seconds		T	1
2	VOSE	nT	Antarctic magnetometer	R	9
3	VOSH	nT	Antarctic magnetometer	R	13
4	VOSZ	nT	Antarctic magnetometer	R	17

```

NOTES:
Start time = 01-JAN-01 00:02:00.000
End   time = 31-DEC-01 23:58:00.000

Antarctic magnetometer high-resolution data
END

```

Fig. 4. An example of the Flat File Manager header file. This file provides the formatting information and metadata for the specific binary data file. This pair of files (the header and binary) constitutes a basic data unit in the VGMO.NET architecture.

dataset is not found in the VGMO.NET server database, a number of known remote sites are contacted and data are downloaded as requested using the specific site information from the shared data structures. This requires knowing the supported access protocols, available tree structures, and file-naming conventions at these remote locations. These details are maintained by GeoMaC in shared, location-specific data structures, populated with information obtained through various means, like personal communication, prior knowledge, or accrued info from the newly “webbed” site.

The third layer converts downloaded data into a flat-file format using various built-in conversion routines. We separate this layer because it is structured as a system of various plug-ins, which can be modified independently if a new format is found or an old one changed.

Finally, any of the tools built into the top “visualization” layer can be used to perform various analyses of the downloaded and converted datasets. One of the highlights in our current implementation is the development of an integrated visualization tool, the Flat File Manager (FFMN; see Fig. 4 and Section 4). Apart from the data visualization analysis, the user can choose to store the data locally building or expanding the local database. If the user decides to open his/her local database to the Internet (e.g., either as a simple ftp directory or as a node of VGMO.NET), then this local self-organized data sub-center becomes available through the future VGMO.NET requests chain. In summary, all of the VGMO.NET portals and nodes form the distributed “data fabric”.

#### 3.1 Local database

In spite of tremendous efforts to introduce a universal geomagnetic database format, many “webbed” geomagnetic sites still offer a variety of different, sometime proprietary file formats; although we hope that the recently approved IAGA-2002 Data Exchange Format will become dominant in years to come (<http://www.iugg.org/IAGA/>; <http://www.ngdc.noaa.gov/IAGA/vdat/iagaformat.html>). In its nature, the new IAGA-2002 format is an ASCII “replica” of the flat-file format described below.

To combine data from different sources into the VGMO.NET database, we use a FLATDBMS format intro-

duced by Smith and Clauer (1984), where a dataset consists of two separate files sharing the same file name, but carrying different file extensions: a header file (\*.hed) which provides description (in ASCII) of the corresponding data file, and the binary data file itself (\*.dat). This “flat-file” format has the advantage of the ASCII formatting in the readable/editable data file description (i.e., providing meta-data on the dataset), as well as its binary data file presentation offers the compact data storage suitable for the fast random access.

Figure 4 shows an example of the geomagnetic header file. Here all lines before NOTES are mandatory specifying the data elements offsets (last column “loc”) in every data record; this allows the user to construct a correct format statement for reading the binary data by a software code. The content between the NOTES and END lines is free-formatted; the user can write here any description (meta-data) of the particular dataset.

As described in the header file, the VGMO.NET data file is a binary, fixed-length record file. First data item in the record is a real\*8 time record (date and time), followed by real\*4 data records. We choose to record time in this file as a Modified Julian day and its fraction. The Julian Day (JD) is a dating method that is used by astronomers, geophysicists, chronologists, and others who need to have an unambiguous dating system based on continuing day counts. Start of the JD count is from zero settled at noon (12 hours) of January 1 of 4713 BC. The Modified Julian Day was introduced by space scientists in the late 1950’s and it is defined as  $MJD = JD - 2400000.5$ ; the count starts at midnight between November 16 and 17 of 1858. Sanctioning MJD, the International Astronomical Union recommended it as a decimal day count, which is independent of the civil calendars in use. It is convenient giving the dates in this system for all cases where data are collected over long periods. Therefore, in our flat-file database system the real\*8 time record contains the real number where the integer part represents the number of whole days since beginning of 17 November 1858, and a decimal part is the UT time as a fraction of the day. For example, 2003-08-19 18:00 UT will be written as 52870.75 into the real\*8 time record.

### 3.2 Directory structure and naming convention

Figure 5 shows a snapshot of the VGMO.NET server-side database directory tree. The root directory is named as DB. All the data are maintained under this root directory and categorized into annual and monthly subdirectories. At the lowest level, we segregate data based on their source or discipline. For example, all magnetic measurements are placed in the MAG subdirectory, while all satellite measurements can be placed under the SAT subdirectory. Such categorization allows to group various datasets (e.g., ionospheric or plasma data) into our DB structure and manage these data with our FFMN software (if the data in question are time series or series of images organized similarly to geomagnetic data).

We choose to assign intuitive file names that carry relevant information about the content of a given file. Such file names facilitate efficient search through directories. Each file name consists of four parts: a three-letter IAGA station code ([http://www.meteo.be/IAGA\\_WG\\_V.1/](http://www.meteo.be/IAGA_WG_V.1/)), a “YYYYM-

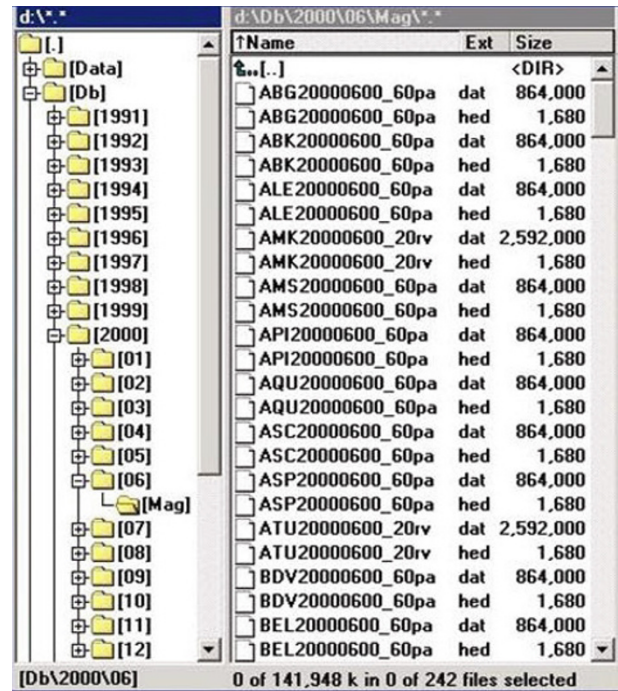


Fig. 5. An example of the directory tree for the database located at the VGMO server or at any of the VGMO nodes if the users would follow the VGMO file-naming and formatting agreements.

MDD” time stamp, double-digit number representing data sampling rate in seconds (60, 20, 10, or 01), and finally special tags that are attached for housekeeping purposes. These tags are: **p** – public access data; **r** – restricted access data; **a** – absolute measurements; **v** – variation measurements. For example, a publicly accessible dataset consisting of 60-sec samples of the absolute geomagnetic measurements made at the Antarctic base Vostok in December 2002 will be stored in a file named `\DB\2002\12\MAG\VOS20021200_60pa`. Note that a double-zero (00) before the underscore in the file name means that this is a monthly file; daily files will be marked here by the numbers from 01 through 31.

### 4. VGMO.NET Implementation

In developing the VGMO concept, we guessed that the scientists might be interested either in searching and downloading geomagnetic data via a Web-based portal or in creating their own self-organized geomagnetic databases similar to our flat-file database managed via FFMN. To meet this, we developed the VGMO.NET middleware in two flavors: (1) as a **standalone, self-supported application** and (2) as a **Web-based portal** to our server-side (<http://mist.engin.umich.edu/mist/vgmo/vgmo.html>) and remote geomagnetic databases. The Web-based implementation demonstrates the feasibility of our notion of building the self-organized databases; it provides a centralized point for accessing geomagnetic data distributed around the World. Any user can request geomagnetic data via the VGMO.NET portal without even knowing where the requested data are located; they either could be stored in one of the VGMO.NET servers or distributed over various remote repositories. On the other hand, users

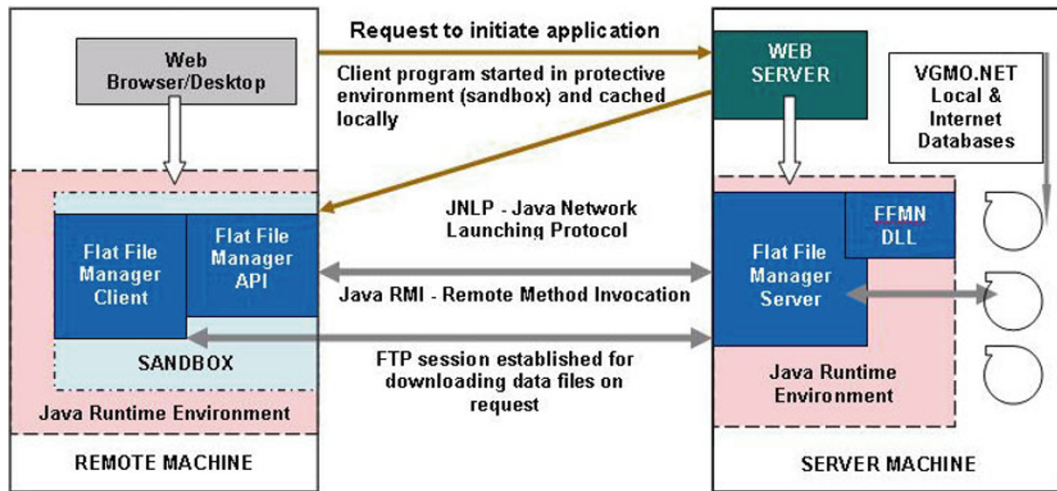


Fig. 6. The Web-based VGMO.NET portal architecture unleashed. Note that the users (left side) must download Java-based client software from the VGMO.NET portal, which will be installed in a “sandbox” environment of the user’s machine. This client software provides the user full graphic and browsing capabilities of the VGMO server.

interested in building their own data sub-centers can install the “standalone” version of our middleware, which utilizes all functions of the VGMO.NET portal and “data fabric” connecting to the VGMO nodes directly, without going through the Web-based portal. These individual, self-organized sub-centers can be then made available to other similar installations if opened to the “data fabric”.

Figure 6 shows the server-client relationships in the VGMO.NET portal. This secure, scalable, platform-independent, and user-friendly software framework provides a remote access to the VGMO.NET portal via the Flat File Manager Client (written in Java 2) that requires a Java Web Start (JWS, <http://java.sun.com/products/javawebstart/>) through the Java Network Launching Protocol (JNLP). The standalone version (as installed at a client machine) can create, manage, and populate the local geomagnetic database via an easy access to the VGMO.NET portal(s) and other “webbed” geomagnetic data repositories, aiming on building the “data fabric” of geomagnetic archives and access points.

#### 4.1 A closer look at the VGMO.NET portal

The Web-based version of the VGMO.NET portal is written in FORTRAN-90; its Java-based client code can be deployed on client machines using the Java Web Start technology (JWS and JNLP). With Java Web Start, simply clicking on a Web link at the client computer can launch the application. If the Java application is not present on the client machine, then Java Web Start would automatically download all necessary files, caching these files on the client computer so that the application is always ready to be re-launched anytime the user wishes.

The very first access (initiation) to the Web-based VGMO.NET portal should be performed via a standard Web browser (e.g., Internet Explorer or Netscape). As connected, the remote, client machine is checked for necessary software. If the software check fails, the client is provided with an option of “one click” installation (currently available only for Windows, but under development for UNIX/Linux). The user can also choose to download

various Java components manually and launch the VGMO client later. As the VGMO client software is downloaded from our Web server and launched locally via JNLP, it runs in a protected environment (sandbox) and does not interact with the client’s machine environment. Subsequent initiations of VGMO.NET from the client machines can be independent of a Web browser; the application can be launched through desktop shortcuts, making launching the Web-deployed application similar to launching a native application.

The VGMO.NET server is implemented as an operating system service. The server listens for session initiation requests and initiates a new process for every session created. For scalability reasons, the server maintains no state about the ongoing sessions. As a result, every request (for downloading data or starting worldwide search) should be complete in itself. Our current implementation identifies a client with the current IP address; thus, multiple VGMO sessions cannot be started from a machine having a single (unique) connection to the Internet.

It is necessary to ensure that a downloaded, potentially untrusted application cannot compromise the security of client machines or local networks. Ideally, one would like to run such application in a restricted environment so that it has limited access to local computing resources such as storage devices and the local network. However, the nature of services expected from VGMO.NET should warrant access to local resources like hard disks. Our framework provides here some flexibility without compromising security. This is accomplished by using the Java Web Start support for digital code signing.

The VGMO.NET client is delivered in a signed Java archive file (a file format used to bundle all components required by a Java applet) and its contents are verified before execution. If verification of a digital signature fails, the application is not launched. This makes it possible for users to verify that a JAR file comes from a trusted source. By notifying the JAR file origin, the client is allowed to make an informed decision whether to grant additional priv-

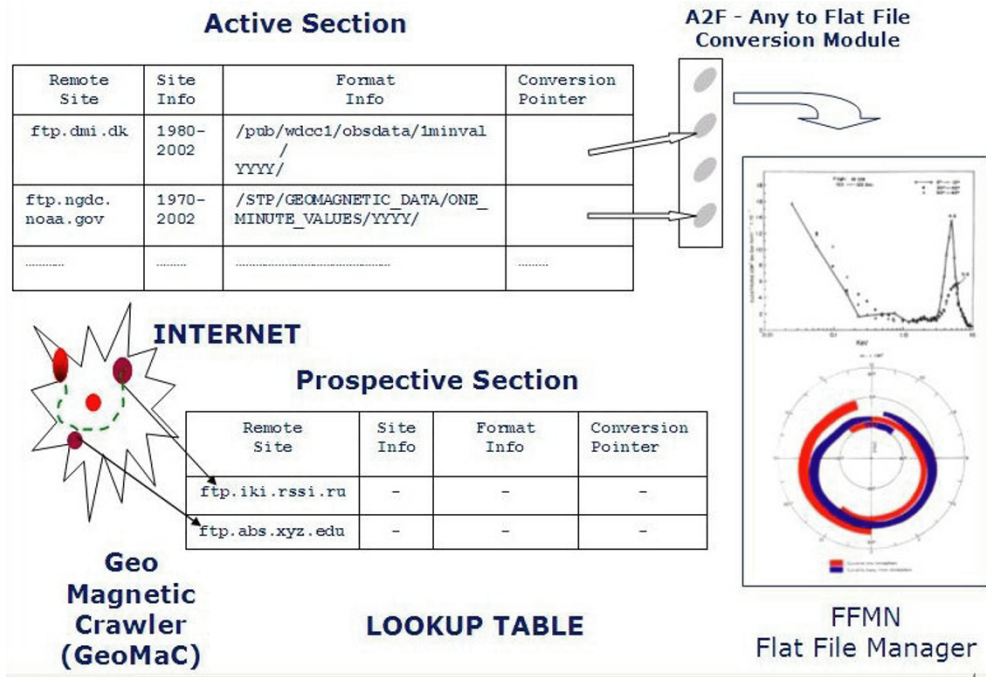


Fig. 7. An example of the lookup table through which the VGMO’s FFMN finds data locally and over Internet.

ileges to the downloaded code, or not. If the user trusts the VGMO.NET service provider, then he or she may choose to grant additional system privileges, such as access to local disks to store newly downloaded data, etc.

**4.2 VGMO.NET lookup tables**

Our current implementation of VGMO.NET maintains a lookup table that contains information about various geomagnetic observatories around the World. This table is maintained at the VGMO.NET portal, but it also provided (and periodically updated) to the VGMO nodes running our “standalone” application. The format of this lookup table is sketched in Fig. 7. The lookup table consists of an **active section** and a **prospective section**. At any time only the active section of the lookup table is used for remote search. We add information about possible data repositories to the prospective section of the lookup table. Once the access mechanism (the format and conversion routine) for these sites is obtained, these entries are moved to the active section of the table.

On receiving the request for a particular dataset, the system first looks in the local (server-side) database. As Fig. 8 shows, if no “matching data” are found locally, the remote sites are contacted for the availability of requested data. Upon initiation of the worldwide search, the client sees a list of sites that will be contacted. The client then can choose to avoid contacting some sites based on his/her personal knowledge of the remote site content. If the requested data are found, they are downloaded and passed to the higher layer, the A2F conversion module. A2F converts these datasets into the flat-files system. In our current implementation, we use the FFMN interface to submit requests for the datasets under interest. FFMN can manage access to multiple datasets (local and remote) simultaneously. Each dataset is organized as a separate flat file (with

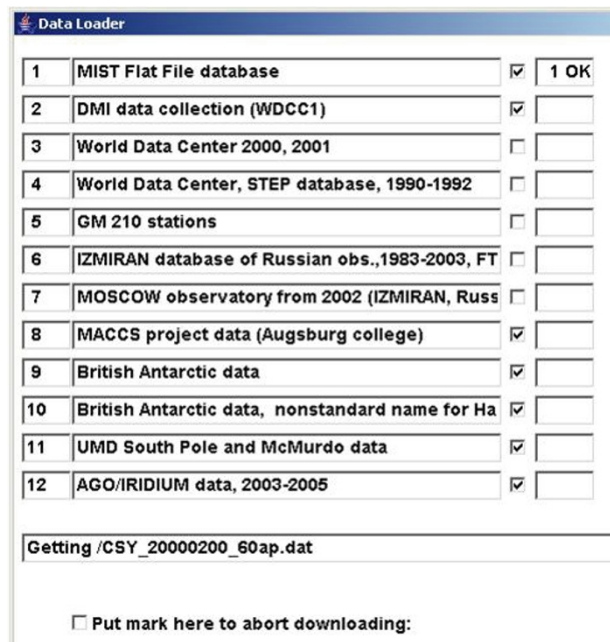


Fig. 8. An example of the VGMO user screen displaying the worldwide search initiated at the VGMO server (portal).

up to 200 columns each) of the ground-based magnetometer observations and can be uniquely identified with the time interval and station code.

**4.3 FFMN client interface**

As an example, Fig. 9 shows the FFMN Main Menu, which allows the user to select up to three datasets (**File**), and then to do certain operations with the selected data (**Action**) according to the pre-set **Options**. The **File** item allows the user to open the server database files or to create a temporary dataset for the selected geomagnetic stations

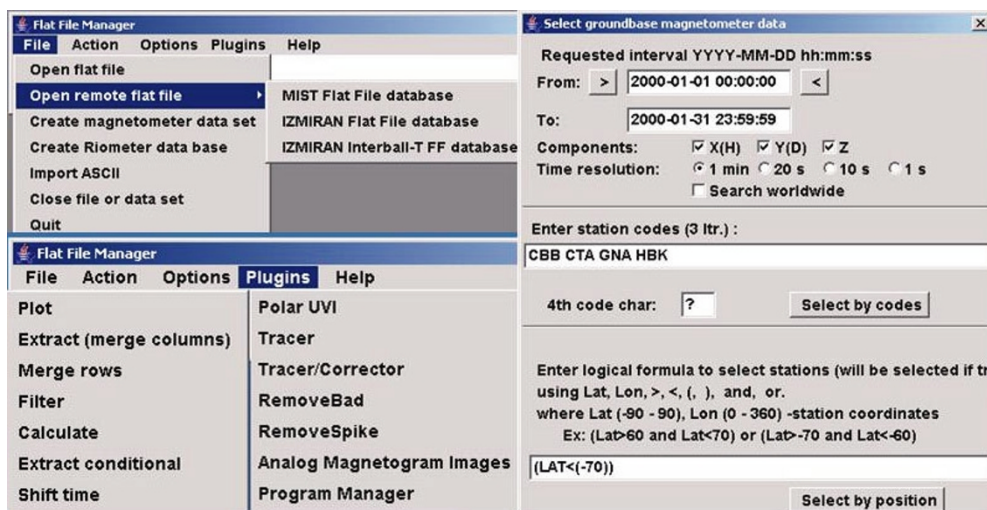


Fig. 9. The VGMO's FFMN at work: examples of different functionalities provided by the server in selecting, retrieving, and manipulating the data.

(selected either by names or by geographic location). If the selected data are found in the server database, then the FFMN server retrieves the requested data for plotting on (and possible uploading to) the remote client machine. In addition, if the “Search worldwide” box is checked, the FFMN server will look for the selected data on a number of remote FTP sites (listed in the FFMN Lookup File); these data are then downloaded, converted to flat files, and added to the FFMN server database. When the new FTP or WWW sites with geomagnetic data are found, they can be easily linked through additions to the FFMN Lookup File.

As seen on the left panel of Fig. 10, the FFMN client shows all data that became available at the server after the local and remote search (that is, included in the temporary dataset for the selected time interval). The user then can **Select All** or mark (+) the data subset for **Action**. The latter includes: **Plot** a stack of magnetograms in accordance with the changeable **Options** menu; **Download** the selected dataset to a file in various formats, including IAGA-2002, and then send this file to the client machine via FTP. If there were a few opened files on the server, then all selected data will be combined into a single file. The **Merge rows** option allows to combine two similar datasets with interlaced time intervals; **Filter** applies the low, high, or band-pass filters to the selected data; **Calculate** allows to enter a formula to compute a new physical quantity from the given columns in the opened data files (e.g., compute the total geomagnetic field intensity from three orthogonal components); and **Shift time** allows to enter an offset timing in one of the datasets for the plotting purposes (e.g., showing delays between time series). The right panel of Figure 10 shows an example where the data from ground-based magnetic observatories are plotted via the FFMN client interface together with data from the ACE spacecraft located at the Lagrangian point 1.5 million kilometers Sunward from the Earth, upstream in the solar wind flow.

The FFMN client is able to create a new dataset from a number of separate magnetometer observations. Note that stations can be either explicitly specified or “interpreted” from a regular expression composed of the combination of

latitude and longitude ranges. FFMN can also be used to combine different data sets. For example, Fig. 11 shows, starting with the bottom trace, one-minute geomagnetic time series from low-latitude observatories San Juan (SJG) and Kakioka (KAK), combined with the 4-minute resolution plasma density from ACE ( $N_p$ ), and hourly mean Dst index (top). Such data subsets can be saved as a new single flat (or ASCII) file and used for further analysis.

In addition to providing various graphical visualizations, FFMN can also be used to perform simple scientific analysis of geomagnetic data like cross-correlations via a stack plot, time shifts, filtering, etc. For tasks that are more specific, the user-written modules (in FORTRAN, C, C++, or IDL languages) can be added as the plug-ins to handle the downloaded data files. While the back-end scientific processing and graphics are accomplished in our application in FORTRAN, the Graphical User Interface (GUI) delivered by FFMN to the client machine is written in Java. We use the native programming interface for Java, the Java Native Interface (JNI, <http://java.sun.com/docs/books/tutorial/>) to glue the two ends together. Programming through the JNI framework ensures platform independence and still allows us to use native codes written in FORTRAN that can best handle scientific tasks.

## 5. VGMO.NET Evaluation

Our experience with VGMO.NET has been exciting but mixed. We were able to demonstrate successfully the effectiveness of our data-sharing tool, but we also realized that many of our design principles are difficult to implement. Our GeoMaC is not yet automatic; we mainly search new geomagnetic data repositories and update our lookup table semi-manually. Furthermore, it is difficult to focus our crawler on the data within the geomagnetic domain, which we attribute to the fact that the Web-posted data generally lack metadata. This lack of metadata makes it hard to search the Web and limits its use. We, therefore, conclude that in lieu of the absence of semantic support for Web documents, it is extremely difficult to publish scientific data through the World Wide Web (Saxena and Papitashvili, 2003).



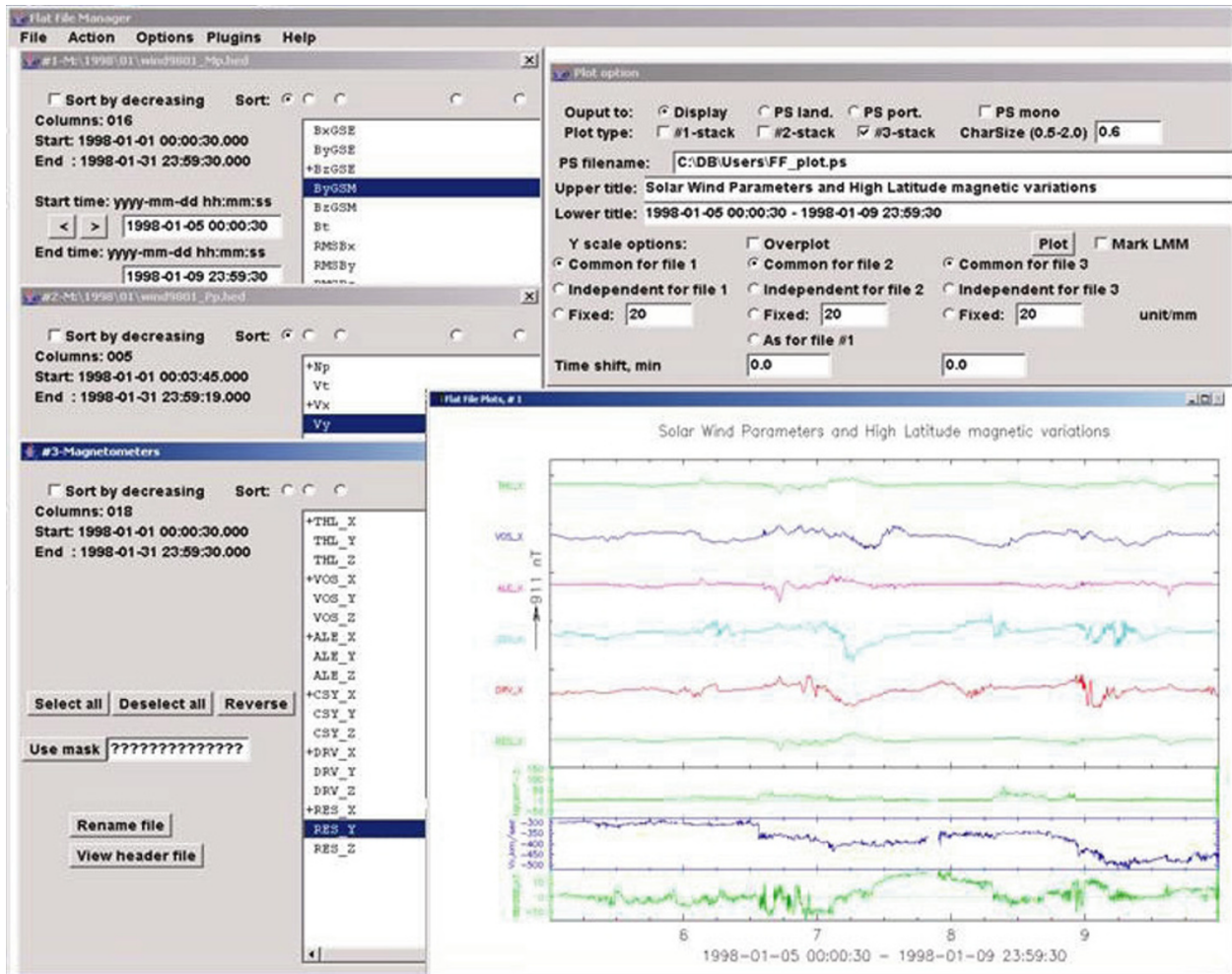


Fig. 10. Same as in Fig. 9: the FFMN search, retrieval, and plot examples.

Furthermore, moving the entries from **Prospective section** into **Active section** of the lookup table was also a major challenge. We often had to use our personal contacts to get the details of a new site, structure of the data repository, and the format of the stored data. In addition, we also observed an interesting phenomenon regarding usage patterns. Users were more inclined to download the entire database as, and when, they learned about it. This is primarily because the end-user machines are becoming more and more resourceful—have increased computing power, huge storage availability, and better network connectivity.

### 6. Conclusions and Future Directions

In this paper, we described a framework for identifying various sources of geomagnetic data that extend from known centralized repositories to widespread unaccounted data sources on the World Wide Web. Our proposed system, VGMO.NET, allows a single point of access to, otherwise, widespread data. It also allows creating individual databases on the client machines in a transparent manner. The success of our system depends on the ability to identify new data sites. Our current implementation of the Web search for geomagnetic data is preliminary; we plan to develop it further by considering some aspects of the intelligent Web crawling for identifying the newly “webbed” dig-

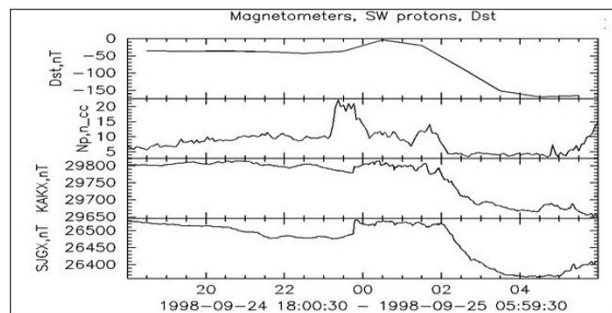


Fig. 11. A visualization sample created by the FFMN client interface for the ground-based magnetometers and Dst index time series combined in a single plot.

ital geomagnetic data.

The proposed VGMO.NET is not specific only to geomagnetic data. Our middleware is written as a framework where the discipline-specific terms and units are stored in a separate HTML file. Replacing these terms and units, we can convert at once our middleware to another type of the Virtual Observatory. The only condition for creating the new VxO is that the stored/handled data must be the time series or series of images; then one can easily build a global

“fabric” of discipline-specific data and crosscutting access points. This universality makes possible easy cross-talk between the Virtual Observatories of various disciplines.

If the World Data Centers adopt the “pull data” paradigm described in this paper, then the WDCs would become even more important. In addition to their role of “permanent archives” for geoscientific data, these Centers would become “clearinghouses”, watching as new digital data appear at the provider sites and copying them to the Centers archives; thus preserving the newly “webbed” data from irreversible loss.

At last, the VGMO.NET concept allows to record all data requests (submitted over the entire “data fabric” within the limited period of time) in the node’s “log files”. These records give the data providers some real means for collecting periodically (via simple reviews of the network logs) statistics of practicality and usefulness of their data, finally “acknowledging” the usefulness of their effort in producing geoscientific data.

**Acknowledgments.** In Michigan, this work was supported by the NSF award ATM-0112720 and NASA grant NAG5-12146.

## References

- Baker, D. N., C. Barton, A. S. Rodger, B. Fraser, B. Thompson, and V. Papitashvili, Moving beyond the IGY: The Electronic Geophysical Year (eGY) concept, *EOS, Trans., American Geophysical Union*, **85**(11), 105–109, 2004.
- Campbell, V., How RAND Invented the Postwar World, *Invention & Technology*, **20**(1), 50–59, 2004.
- Guide to the World Data Center System, ICSU Panel on World Data Centers, NOAA NGDC/WDC-A, Boulder, Colo., 109 pp., 1996 (<http://www.ngdc.noaa.gov/wdc/wdcmain.html>).
- Papitashvili, V. O. and N. E. Papitashvili, A concept of self-populating sites sharing data via World Wide Web, Symposium G5.03, IAGA-IASPEI Joint Scientific Assembly, Hanoi, Vietnam, August 19–31, 2001.
- Papitashvili, V. O., M. Candidi, C. R. Clauer, F. Christiansen, M. J. Engebretson, T. Neubert, O. Rasmussen, A. S. Rodger, T. J. Rosenberg, and J. F. Watermann, Polar Conjugate Facility: New opportunities for old-fashioned analyses of high-latitude geomagnetic data (invited), XXVII SCAR Meeting, Shanghai, PRC, July 15–19, 2002a.
- Papitashvili, V. O., C. R. Clauer, V. G. Petrov, and A. B. Saxena, Virtual Global Magnetic Observatory: Concept and Implementation, Session SH03 “Towards an Integrated Solar-Terrestrial Data Environment”, 2002 Fall AGU Meeting, San Francisco, December 6–10, 2002b.
- Papitashvili, V. O., A. B. Saxena, V. G. Petrov, and C. R. Clauer, VGMO.NET—Realization and testing of a Virtual Global Magnetic Observatory (invited), Session GAV.03 “The Geospace Environment in Near-Real Time: Science and Technology”, IUGG/IAGA General Assembly, Sapporo, Japan, June 30–July 11, 2003.
- Papitashvili, V. O., A. B. Saxena, V. G. Petrov, and C. R. Clauer, A Virtual Antarctic International Magnetometer Network, Session X: eAntarctica: From Data Management to Information Portals, XXVIII SCAR, Bremen, Germany, July 25–30, 2004.
- Papitashvili, V., D. Baker, C. Barton, W. Peterson, and the eGY Team, E-Science for Geoscience: Preparing for the International Polar Year (2007–2008) with the “Electronic Geophysical Year” Initiative (invited), Session US5: The International Polar Year 2007–2008, European Geoscience Union General Assembly 2005, Vienna, Austria, April 24–29, 2005a.
- Papitashvili, V. O., V. G. Petrov, C. R. Clauer, A. B. Saxena, and N. E. Papitashvili, Virtual Global Magnetic Observatory VGMO.NET: A Component of the Electronic Geophysical Year Initiative, Session U08: eGY: e-Science for Geoscience, The 2005 AGU/SEG/NABS/SPD/AAS Joint Assembly, New Orleans, Louisiana, May 23–27, 2005b.
- Papitashvili, V. O. and V. G. Petrov, A Virtual Global Magnetic Observatory: VGMO.NET as a Component of the Worldwide “Data Fabric”, Session GAV01: Magnetic observatories: measurements, quality analysis, and data dissemination, IAGA 2005 Scientific Assembly, Toulouse, France, July 18–29, 2005c.
- Saxena, A. B. and V. O. Papitashvili, Virtual Observatories: A peer-to-peer semantic approach for efficient sharing of scientific data, Special Union Session U22A “Virtual Observatories in Space and Earth Sciences”, 2003 Fall AGU Meeting, San Francisco, December 9, 2003.
- Smith, A. Q. and C. R. Clauer, FLATDBMS: A flexible, source-independent data management system for scientific data, STAR Laboratory Report D106-1984-1, Stanford University, 25 pp., 1984.

---

V. O. Papitashvili (e-mail: [papita@umich.edu](mailto:papita@umich.edu)), V. G. Petrov (e-mail: [vpetrov@engin.umich.edu](mailto:vpetrov@engin.umich.edu)), A. B. Saxena (e-mail: [anshuman.saxena@tcs.com](mailto:anshuman.saxena@tcs.com)), C. R. Clauer (e-mail: [bob.clauer@umich.edu](mailto:bob.clauer@umich.edu)), and N. E. Papitashvili (e-mail: [natasha@mail630.gsfc.nasa.gov](mailto:natasha@mail630.gsfc.nasa.gov))