

ICESTAR: DEPLOYMENT OF VIRTUAL OBSERVATORIES IN CYBERSPACE TOWARDS CREATING THE WORLDWIDE FABRIC OF GEOPHYSICAL DATA

V. O. Papitashvili¹, V. G. Petrov², N. E. Papitashvili³

¹SPRL, University of Michigan, Ann Arbor, Michigan, United States, ²IZMIRAN, Troitsk, Moscow Region, Russia, ³SPDF/NASA and QSS Group, Greenbelt, Maryland, United States

A concept of the Worldwide Data Fabric (WDF) emerged from a success of the distributed system of geophysical data providers and World Data Centers (WDC) that served geophysical communities since the International Geophysical Year (1957-58). However, with the advent of Internet and World Wide Web, a need of submitting digital data to some central repositories became less pressing because the data could easily be pulled out directly from provider servers by users or data centers as required. This leads to the concept of “virtual observatories” deployed in cyberspace and interconnected through the worldwide fabric of geophysical data providers and users/centers. The WDC can play here a new, important role – they should now scan the WDF for newly-webbed geophysical data and copy them to the central archives, preserving at least two identical copies (or as many copies as Centers exist) of new datasets within the data fabric. In this way, the WDF would become a self-organized system of the data nodes (providers and users) and data portals (WDCs as “clearinghouses”). However, the WDF concept faces a number of challenges. First, the data providers should agree to make their data available via the Internet through open (but secure) access protocols. Second, because multiple copies of every dataset would now be copied across the WDF, every dataset (original or copied) must be digitally signed by the data provider and then by every data copyist. These signatures should be encrypted within the data file, protecting the file from deliberate hacking and corruption. As the WDF (for all or specific geophysical disciplines) is established, then one can write a specific middleware - either to retrieve needed observational data (i.e., data users), or to collect information from all nodes on the number of demands of a specific dataset for copying (i.e., data providers). The latter would then resolve the long-standing problem of thorough acknowledgements to the data providers. The presentation will address these challenges suggesting some immediate and long-term solutions.

URL: <http://mist.engin.umich.edu/mist/vgmo/vgmo.html>