해외기고

A Model for Distributed Community (e)Science

Peter Arzberger
(PRAGMA)

목 차

1. Introduction

Because quarantine and isolation are the primary means of slowing the spread of SARS, Taiwan's hospitals faced a communication logjam during the 2003 SARS epidemic. Physicians in quarantined hospitals were unable to consult with specialists at other institutions, and on a more personal level, hospital staff and patients had limited contact with their families. As of May 13, 2003, the World Health Organization reported that the respiratory illness had infected 7548 people worldwide, killing 573, and that the outbreak was still on the rise in Asian countries.

On May 15, 2003, in search of expertise for setting up Access Grid sites, Taiwan's National Center for High-performance Computing (NCHC) sent a request for help to Pacific Rim Application and Grid Middleware Assembly (PRAGMA) members. Within hours, offers to assist poured in from around the world, with volunteers ready to provide gear, remote expertise, and Chinese-speaking support staff.

The request was based on a simple question: With all of the technology we have, how do we help our fellow human beings?

This paper also explores how distributed, grassroots communities can take advantage of new technologies to respond to the many challenges we collectively face in the future.

2. Background

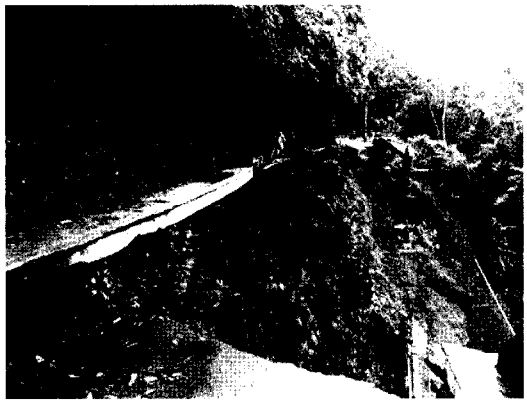
We are living during an information tech nology revolution that is impacting major aspects of society. This revolution is driven by exponential changes in computer, data storage, and optical networking capacity: computing power doubles every 18 months for the same unit price (Moore's law); data storage capacity doubles more rapidly, and optical networking capacity, doubling roughly every 9 months, is faster than both computing

and storage[1]. National investments in the internet alone have had clear and lasting impact on society in only a few years, creating a global post office, a global shopping mall, a global library, and a global university[2].

These trends were recognized throughout the scientific community, and governments began to build a new information infrastructure on top of the networks. In 2000, John Taylor, Director General of (UK) Research Councils (UK), defined e-science as "the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high perfor- mance visualisation back to the individual user scientist.[3]" As defined in a 2003 report to United States National Science Foundation (NSF) by Dan Atkins, "cyberinfrastructure (CI) refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy.[4]"

Since then CI has been extended beyond the physical wired network ends to wireless connected sensors and sensor networks through out our physical and biological world. These extensions allow us to "see" events transpiring at temporal and spatial scales not achieved before.

As noted in a recent report by NSF[5],"at the heart of the cyberinfrastructure vision is the development of a cultural community that supports peer-to-peer collaboration and new modes of education…. Cyberinfrastructure enables distributed knowledge communities that collaborate and communicate across disciplines, distances and cultures. These research and education communities extend beyond traditional brick-and-mortar facilities, becoming virtual organizations that transcend geographic and institutional boundaries."

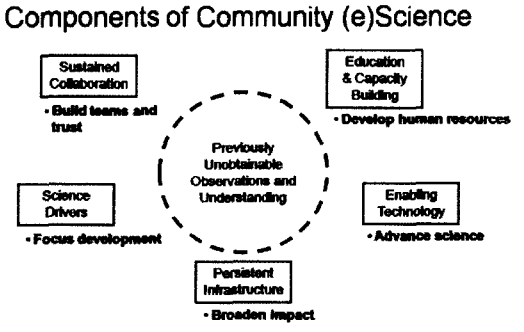


(Image 1) Disruption of transportation system during typhoon, yet instruments keep working.

3. A model of e-science: Five Components that Enable Community Science

1) Stix, Gary, Triumph of Light, Scientific American, January 2001. Has graphic on the three exponential of compute, data storage, and optical network
2) Friedman, Thomas, The Lexus and the Olive Tree, Farrar, Straus & Giroux, 1999.
3) http://www.rcuk.ac.uk/escience/news/firstphase.htm
4) Atkins, D.E., K.K Droegemeier, S.I. Feldman, H. Garcia-Molina, M.L. Klein, D.G. Messerschmitt, P. Messina, J.P. Ostriker, M.H. Wright, 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure (National Science Foundation, Arlington, VA, January 2003): http://www.communitytechnology.org/nsf_ci_report/
5) Cyberinfrastructure for 21st Century Discovery, NSF Report, Overview by A Bement, http://www.nsf.gov/pubs/2007/nsf0728/index.jsp

What are the CI and other components to create successful community science teams and community organizations? We propose that there are at least five interlocking pieces[6] (see adjacent Figure 1) to sustain these new distributed communities.



3.1 Science Drivers

These are the applications that require development of new technologies to make progress and often motivate teams to work together. Example CI project in the US include the sharing and analyzing of brain images through the Biomedical Informatics Research Network (BIRN) in order to understand diseases such as depression: the integration of distributed geosciences data for use in simulations to understand the formation of mountain ranges, as conducted by the GEOsciences Network (GEON). A third example, described later in this article, is the understanding of lake dynamics across a wide variety of lakes, which is being conducted by the Global Lake Ecological Observatory network. Many other examples exist in areas from ocean observing, earthquake engineering and sensing, and meteorology.

3.2 Enabling Technologies

These technology developments are focused by the applications, but also inform applications of what can be done to change the conduct of the science. Technology examples include use of a common data storage system to track the images and retrieve them (BIRN): the creation of registration of different data sets for the discover of new data (GEON): the streaming of data[7] from sensor to the network(GLEON). Other technologies include the remote control of instruments(e.g. microscopes, telescopes, sensors)and the creation of tools to share images and large data interactively between sites(OptIPuter[8]), and the virtualization of compute and data resources.

3.3 Persistent Infrastructure

Persistent infrastructure refers to both the physical infrastructure (networking, data that are stored, sensors, compute nodes) and the software. Persistence is required to entice scientists change their normal mode of thinking and conducting science and use new tools and distributed infrastructure. This takes time,to learn how to take advantage of the new opportunity, and persistence insure that the infrastructure it will continue to be available. This is a challenge for funding agencies that often do not like to create a mortgage on their funds. Yet this is what is required.

---

6) Motivated by presentation by Tony Hey, GGF 13, Seoul, Korea, March 2005

7) The specific technology is the Open Source DataTurbine, Initiative, http://www.dataturbine.org/

8) http://www.optiputer.net/

The above three components are necessary in all discussions of e-science. But they are insufficient if the long term goal is to build a knowledge community of participants in a potentially transdisciplinary science.

3.4 Education and Training

Part of this component focuses engaging the community on how to use the new tools that are being developed. Another is to expose students to the multicultural, multisciplinary environment, the former since better ideas come from groups of different backgrounds[9], the latter because the nature of many exciting problems requires knowledge from several different communities. These activities are really building the base of the community.

3.5 Community Building

One of the truly revolutionary concepts in e-science is that distance to resources and in particular colleagues is less of a barrier than in any time in the pass. One consequence is that "virtual organizations" can form. In the grid sense these organizations often share resources. However, another focus is on teams of researchers that coalesce around science questions or set of questions and take advantage of the set of distributed resources to do research that none would do individually. For example, groups might begin to ask different research questions, say of a globally distributed set of sensors in lakes, than by just studying a single lake; students will engage mentors from this groups that are geographically distant from their own location. Essential in the component is the notion of "building trust" among members.

4. Two Examples of Communities: PRAGMA and GLEON

Two examples of e-science projects that combine most of the aspects of these five components are the Pacific Rim Application and Grid Middleware Assembly[10][11] (PRAGMA) and the Global Lakes Ecological Observatory Network[12][13][14] (GLEON).

PRAGMA, established in 2002, consists presently of 36 institutional members, who work together to advance the use and development of grid technologies [Enabling Technologies, such as grid middleware] through application drivers [Science Drivers] and to build persistent collaborations. Its own grid is built on the persistent networking and some key software components, its uses applications to drive development and improvement of middleware codes[15][16].

---

9) Ollila J, "Diversity". http://www.nokia.com/link?cid=EDITORIAL_4055 (retrieved 25 July 2006) ["Diverse teams are more creative and find better solutions than homogeneous teams." Nokia CEO Jorma Ollila.]

10) www.pragma-grid.net

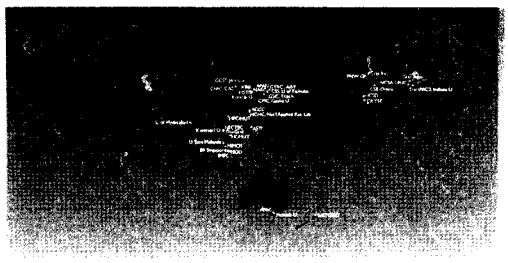
11) Arzberger, P., Papapoulos, P. PRAGMA: Example of Grass-Roots Grid Promoting Collaborative e-science Teams. CTWatch. Vol 2, No. 1 Feb 2006. http://www.ctwatch.org/quarterly/articles/2006/02/

12) gleon.org, Kratz et al, Hanson (Grass roots)

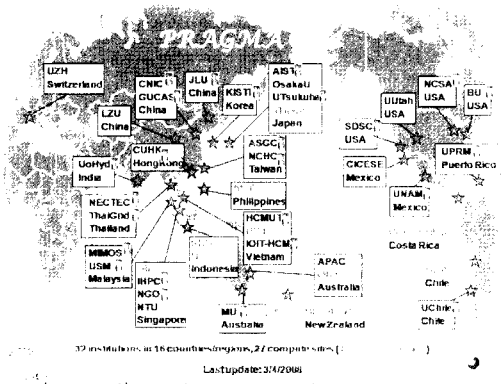
13) Kratz, Timothy K., Peter Arzberger, Barbara J. Benson, Chih-Yu Chiu, Kenneth Chiu, Longjiang Ding, Tony Fountain, David Hamilton, Paul C. Hanson, Yu Hen Hu, Fang-Pang Lin, Donald F. McMullen, Sameer Tilak, Chin Wu. Toward a Global Lake Ecological Observatory Network. Publications of the Karelian Institute 145:51-63 (2006).

14) PC Hanson, Frontiers in Ecology and the Environment, Sept 2007, p343. Discusses advantages of grass roots networks

15) Abramson D, Lynch A, Takemiya H, Tanimura Y, Date S, Nakamura H, Jeong K, Hwang S, Zhu J, Lu



(Image 2) Map of PRAGMA sites (needs to be updated - not included)



(Image 3) PRAGMA Testbed Sites, developed by PRAGMA Working Group. This testbed reflects contribution of compute and data storage resources contributed by PRAGMA members and collaborators.

PRAGMA is organized by working groups in Resources(how to make a heterogeneous grid useful on a routine basis to scientists, examples include areas of climate impact[17], airflow[18], and quantum chemistry[19]), Biosciences(using different middleware on scheduling and data sharing to improv ehigh-throughput screening of potential drugs against avian flu[20]), Telesciences(how to control remote instruments such as microscopes or sensors, and how to use new largetile display ways and OptIPortal technology to collaborate and to view our environment), and GEO(technically how to combine data from different national remote sensing data with different technologies). These working groups interact to share knowledge gained across PRAGMA. PRAG MA also meets twice a year, to update and demonstrate progress, meet and engage new participants, and set plans for the next six to 12 months. These meetings have been essential to build trust among participants and develop and reinforce common principles and goals of this grassroots(bottom-up, voluntary)community.

Over the six years of existence PRAGMA has focused on emphasizing people as part of the grid. PRAGMA has helped catalyze interactions between groups, leading to new programs for training undergraduate students (PRIME[21]), graduate students via the Pacific

Z. Amoreira C, Baldridge K, Lee H, Wang C, Shih HL, Molina T, Li, W, Arzberger P. Deploying Scientific Application on the PRAGMA Grid Testbed: Ways, Means and Lessons. 241 - 248, CCgrid 2006

16) Zheng C, Abramson D, Arzberger P, Ayuub S, Enticott C, Garic S, Katz M, Kwak J, Lee BS, Papadopoulos P, Phatanapherom S, Sriprayoonsakul S, † Tanaka Y, Tanimura Y, Tatebe O, Uthayopas P. The PRAGMA Testbed: Building a Multi-Application International Grid, 57, CCGrid 2006.

17) Lynch AH., Abramson D, Görgen K, Beringer J, and Uotila P. "Influence of savanna fire on Australian monsoon season precipitation and circulation as simulated using a distributed computing environment:, Geophys. Res. Lett., 34, L20801, doi:10.1029/2007 GL030879. 2007. (Savannah PRAGMA Grid application)

18) Ko SH, Kim C, Kim KH and Cho KW. "Investigation of Turbulence Models for Multi-stage Launch Vehicle analysis Including Base flow." Int. Conf. for ParCFD2006. Elsevier 2007.

19) Ikegami T, Maki J, Takami T, Tanaka Y, Yokokawa M, Sekiguchi S, and Aoyagi M. "GridFMO -Quantum Chemistry of Proteins on the Grid." Proc. of Grid2007 (Austin, Texas, Sept 2007).

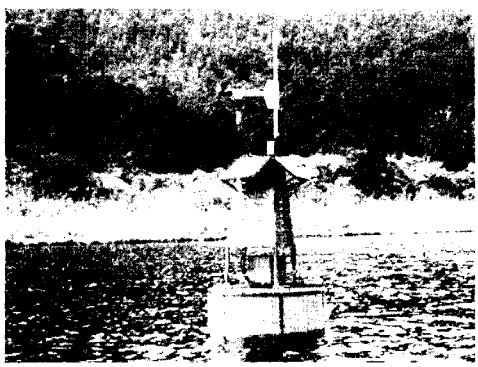
20) A technology supporting the Avian Flu Work: Choi Y, Jung S, Kim D, Lee J, Jeong K, Lim SB, Heo D, Hwang S, and Byeon OH. "Glyco-MGrid: A Collaborative Molecular Simulation Grid for e-Glycomics," in 3rd IEEE Int. Conf. on e-Science and Grid Computing, Banglore, India, 2007. Accepted.

21) prime.ucsd.edu, for video about PRIME students see

Rim International UniverSities(PRIUS[22]) program initiated at Osaka University, and a new program for undergraduate students, launched at Monash University in Australia, the MONASH Undergraduate Research Projects Abroad(MURPA)[Education and Training]. It recently has launched a, PRAGMA Institute aimed primarily at building the base of users in the Pacific Rim community who can use grid technologies]. One such institute focused in particular on a specific region of the Pacific Rim, Southeast Asia, to help build capacity in the region[23]. Both the PRAGMA Institutes and these programs aimed at students participating in research are aimed at growing a community base. Both of these activities had unexpected benefits for PRAGMA by building stronger research ties between partners involved in the educational exchange, improving the understanding of staff at the location of the exchanges, and growing the PRAGMA community via new members who participated in PRAGMA Institutes.

PRAGMA is also a conduit of ideas (new trends and needs) and technologies (how middleware can be used, and how it is improve through usage), and a platform for new collaborations. One new exciting area for PRAGMA is the use of the OptIPortal technology, to use it both as a means of communication between collaborators, but to extend it to other efforts in the environment, such as for observing real time coral and fish ecology (via a nascent effort on CREON[24]), potentially changing how researchers study coral. This involves integration of several technologies and the collaboration across several sites. Also, as noted above, PRAGMA helped launch three education exchange programs. Furthermore, it catalyzed new collaborations between researchers. Once such examples is GLEON, which grew out of an collaboration between researchers U Wisconsin Long Term Ecological Research and those at Taiwan NCHC EcoGrid project.

(Image 4) Data streaming from Kenting Coral Reef Site in Taiwan, via DataTurbine, to Tile Display Wall in California, technology part of OptIPortal. PRAGMA Collaboration between NCHC Taiwan and Calit2]



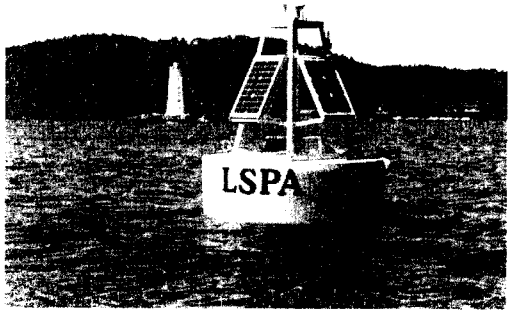
(Image 5) Buoy in Yuan Yang Lake, Taiwan, via collaboration catalyzed by PRAGMA with Taiwan EcoGrid by NCHC and Wisconsin Long Term Ecological Research site.

22) prius.istosaka-u.ac.jp/en/index.html

23) 1st PRAGM Ainstitute(www.ncsa.uiuc.edu/Conferences/PRAGMA13/i-agenda.html), 2ndPRAGMAinstitute (www.nchc.org.tw/event/2007/pragma_institute/)

24) CREON: Coral Reef Environmental Observatory Network. http://www.coralreefeon.org/

(Image 6) Yuan Yang Lake, Taiwan (Dong-Kuan Liao)



(Image 7) Buoy on Lake Sunapee, New Hampshire. Part of GLEON deployment of persistent software

GLEON, the second example of a distributed community, is focused heavily on a specific science domain.. GLEON is a grassroots network of limnologists, ecologists, information technology experts, and engineers who have a common goal of building a scalable, persistent network of lake ecology observatories to understand lake dynamics in the context of global processes. GLEON was established in 2005, following a process similar to that of PRAGMA and it builds on other environmental observing network activities[25]. Two key technologies being deployed currently are a data streaming architecture and a common data framework, based on a controlled vocabulary[26]. Inaddition, this package is being improved to ease the deployment at sites that do not have sufficient technical expertise. With sufficient deployments, a small network of sites sharing data is emerging. Such a functioning network will be used to address questions tha trange over the network, initially in acomparative basis[27].

While GLEON's organizational development and grass-roots nature are similar to PRAGMA, there are several additional differences. One is the focus on data sharing, to do larger scale, network science, which is critical to GLEON's success. Data sharing is one that involves technical, social, and legal challenges[28]. Second, the issue of attribution is veryi mportant, to give credit to those who produced data. A third

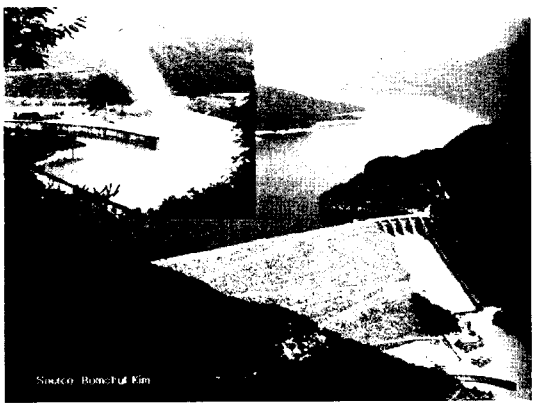
25) Examples are the Long Term Ecological Research Network (LTER), National Ecological Observatory Network (NEON): WATer and Environmental Research Systems Network (WATERS): Ocean Observing Initiative (OOI).

26) The data streaming architecture is DataTurbine (www.dataturbine.org) and the common data model is the Vega Model()…See also Tilak, S., P. Arzberger,D.Balsiger,B. Benson,R. Bhalerao,K.Chiu,T. Fountain,D. Hamilton,P.Hanson,T.Kratz, F.P.Lin,T. Meinke,L. Winslow, Conceptual Challenges and Practical Issues in Building The Global Lake Ecological Observatory Network, Proceedings of the Third International Conference on Intelligent Sensors, Sensor Networks and Information Processing(ISSNIP)2007, Melbourne, Australia

27) For discussion of GLEON by participants see http://www.calit2.net/newsroom/article.php?id=1257and the links to its video.

28) Arzberger, Peter, Peter Schroeder, Anne Beaulieu, Geof Bowker, Kathleen Casey, Leif Laaksonen, David Moorman, Paul Uhlir, Paul Wouters, An International Framework to Promote Access to Data, Science, Vol 303, pp 1777-1778 (19 March 2004).

difference is the engagement of students directly into the working groups of GLEON. Working groups are the key bodies that focus GLEON activities. For the students, they are embedded in distributed, multisciplinary, multicultural teams, all focused on understanding specific science questions. This immersion is the "hands-ontraining" for their future careers, and a way to build their own professional networks. Finally, many of the GLEON participants are very involved in the role of lakes as ecosystems service. By being able to share experiences, we hope to improve those services more broadly.



(Image 8) Soyang Lake, and issues of turbidity for drinking water in Seoul (photo BC Kim)

5. Questions about Virtual Teams and Distributed Knowledge Communities

If one starts from the premise that these distributed teams and virtual communities will become more prevalent in the future, what then are the characteristics that will enable certain teams or communities to succeed and others to fail? Do the rules that apply to organizations that are located at one site apply to groups and communities that are distributed? How do the rules change when the organization consists of individuals who bring their own resources (grass-roots) versus those where the resources come from a single source? Can new technologies and persistent infrastructure replace the needs for groups meeting face-to-face (as is done now in PRAGMA and GLEON) or the immersive experience of an exchange program for students such as in PRIME, PRIUS and MURPA?

These are interesting and important questions for organizational scientists. The important lessons from these early experiments (PRAGMA, GLEON) need to be shared among researcher and in particular among students and junior researchers who will live the longest in this new regime. We note that recently NSF is requesting proposals on Virtual Organizations as Sociotechnical Systems to conduct scientific research directed at advancing the understanding of what constitutes effective virtual organizations and under what conditions virtual organizations can enable and enhance scientific, engineering, and education production and innovation[29].

Of course, the premise is valid only if new science, understanding, and solutions to large problems are found via these groups. In the case of GLEON, its success will be the creation of both a new theory for lake ecology, but new understanding of looking at lakes as a network rather than in isolation. And the ultimate success of PRAGMA is in the continued growth of the collaborating community, the

29) http://www.nsf.gov/pubs/2008/nsf08550/nsf08550.htm

creation of new science with the sharing of the technologies, and ultimately addressing the larger social challenges.

6. Challenge to the e-science Community

Let's return to the SARS example at the beginning of this paper. With all of our tools in cyberinfrastructure, how can we use them to understand and help address the suffering of our neighbors. This was the question that was posed by Dr. Fang-Pang Lin of NCHC in May 2003.

Grid computing researchers around the Pacific Rim quickly mobilized to fight the SARS epidemic, helping establish a cutting-edge communication grid among quarantined hospitals across Taiwan within two weeks. In addition to linking the hospitals to each other, the grid connects doctors to global sources of health information.

What allowed for this success was not only posing the question, but by having a community(PRAGMA) to respond!

But that example is just one of many more that we currently face: the growing impact of global climate change on our planet, and its impact on water supply[30], crop growth, and infectious diseases[31]; the fact that we are living beyond our means with respect to fish, water ,…[32]; and humanitarian imperative to help all humans achieve a level of education and living outside of ignorance and hunger.

In my experience, even if one does not have all (any) of the answers, there is great value in posing the question. Furthermore, as demonstrated above, we will need to know how to work as a community or grassroots community, to use the tools we have, to engage experts and all career levels, to make a difference.

7. Acknowledgements

This paper is based on talks delivered at the HPC Asia conference 2007 in Seoul Korea (September 2007) as well as at the 2nd PRAGMA Institute in Hsinchu Taiwan (December2007). The author acknowledges support from these organizations and their funding organizations.

This work has been supported by the US National Science Foundation awards and the Gordon and Betty Moore Foundation. The author has benefited tremendously from all of the participants of PRAGMA, PRIME, GLEON, NBCR, Calit2, and wishes to acknowledge their contribution. Finally, there are many other people who have been allowing me the opportunity to be involved in these efforts. But only the author is responsible for any misstatements or inaccuracies.

30) Pearce, Fred. When Rivers Run Dry. The Defining Crisis of the Twenty-First Century. Beacon Press - 2007

31) Intergovernmental Panel on Climate Change. Report 2007: http://www.ipcc.ch/.November 2007

32) Living Beyond Our Means. Millennium Ecosystem Assessment. Natural Assets and Human Well-being. March 2005.

Author Profile

Peter Arzberger

Peter Arzberger is Chair of the Pacific Rim Application and Grid Middleware Assembly (PRAGMA; www.pragma-grid.net), an open, institution-based organization of 30 institutions. PRAGMA, founded in 2002, has a mission to build sustained collaborations among researchers around the Pacific Rim by building applications on top of emerging Grid hardware and software. Connected with PRAGMA is PRIME, the Pacific Rim Undergraduate Experiences (prime.ucsd.edu) program, which provides international research and cultural internship experiences to undergraduate students. PRIME, founded in 2004, has admitted 36 students and sent students to four PRAGMA sites. Arzberger is a founding member of the Steering Committee another international activity, GLEON(http://www.gleon.org), the Global Lake Ecological Observatory Network. GLEON is agrassroots network of people, institutions, programs, and data linked by cyberinfrastructure and united by the mission to understand and predict the response of lake ecosystems to natural processes and human activities at regional, continental, and global scales.

In addition, Arzberger is Director of the National Biomedical Computation Resources (http://nbcr.net), an NIH National Center for Research Resource award. NBCR's mission is to develop computing and information technologies(e.g., end-to-end tools in cyberinfrastructure) to catalyze and facilitate biomedical research across a broad range of biological scales. He is also Chair of the National Advisory Board to the U.S. Long Term Ecological Research(LTER)network.

Arzberger is the former Executive Director of the National Partnership for Advanced Computational Infrastructure (NPACI) and a former Program Officer at the National Science Foundation in Computational Biology.