

# Prospects for Instrument Integration

## With TeraGrid Resources

Gregory G. Pike, John W. Cobb, and James A. Rome  
Oak Ridge National Laboratory

---

This paper presents a partial survey of efforts to integrate instruments and facilities to distributed computing environments. We examine several current instrument integration projects to determine common principles for classification based on instrument size, complexity, and integration technology. Using these classification principles, we evaluate the types of projects that are well suited to integration in a grid-enabled environment. We further assess which types of projects are well suited for deployment on the TeraGrid in terms integration with and advantage of common software stacks, software environment homogeneity across computational resources, data storage and transport tools, data collection integration, and other tools offered in the TeraGrid environment.

---

### 1. Background

The integration of computing resources and instruments is not only a natural relationship, it has become a necessity. When Henry Ford began producing Model T cars in 1914 using an improved version of Ransome Eli Olds' assembly line, the entire process was monitored and controlled by people. Modern day factory floor automation systems can contain thousands of sensors and individual controls, making it cost-prohibitive to manage a facility without the integration of computing resources. Cyber infrastructure plays a critical role in monitoring and controlling large facilities by reducing costs and errors associated with human interaction. By consolidating monitoring services, automating responses, and providing access to controls in harsh environments that are inhospitable to humans, instrument integration provides facilities with the competitive advantage required in today's research and business environments.

The popularity and stability of Linux has brought low-cost commodity control systems within the reach of even the smallest facilities. The computing power provided by readily available computing systems allows the consolidation of multiple integration efforts within a single system. Shared cyber infrastructure resources such as the TeraGrid[1] bring the promise of even greater cost savings and efficiency by leveraging state of the art computing, software,

and work-flow systems under a common framework and infrastructure.

### 2. Defining Instruments

In surveying instrument integration efforts, we take a very broad view of what constitutes an instrument. We consider the simplest instruments consisting of a single tool with a very small data stream to large facilities consisting of many control and monitoring systems with numerous individual tools and sensors. By examining the broadest range of instruments, we can distill common properties that differentiate tools, classify them according to those properties, and determine their suitability for integration into a shared cyber infrastructure resource such as the TeraGrid. For the purposes of this survey, we explicitly ignore systems that are completely autonomous (such as pipeline pigs[26]) and have no external data stream interface since they provide no opportunity for integration.

### 3. Classifying Instruments

#### a. Interactivity

Some instruments can be considered purely as sensors because they provide no means for changing the state of the tool (read-only instruments). The CMU Coke machine[2] provided feedback as to whether the vending machine was empty, and if not, how cold the products were. A user was able to read the state

of the machine, but there was no means to remotely change the state. In contrast, the Bradford Robotic Telescope[3] allows registered users to submit requests that reposition a telescope and obtain images of the sky. In this case, the user is allowed to change the state of the instrument, and then receive the resulting output (read-write instruments). There is a third special case of instruments that provide only an input interface (write-only instruments). Consider the case of a missile self-destruct system. This type of system is completely independent of the other missile systems, and provides no readable data stream. Its sole function is to respond to an input data stream, and do so only once.

#### b. Complexity

The complexity of an instrument can be classified by a number of different criteria. If we consider only the data stream required for an instrument, then we can consider instruments with small data streams to be simple. Personal weather stations that report to The Weather Underground[4] can be considered simple instruments because they produce a small amount of data such as temperature, wind speed and direction, and barometric pressure. Instruments which produce large volumes of data, or manage data from a large number of data sources such as the Spallation Neutron Source[5] can be considered complex instruments simply because of the volume of data that needs to be managed from multiple independent sources.

It's important to note that many simple instruments may be managed by a single control point which then becomes a complex instrument simply by aggregating a multitude of simple data streams. The SensorNet project[6] gathers data from multiple remote sensors to build a comprehensive incident management system and disseminate critical data to appropriate incident response agencies. Likewise, the LEAD[7] project processes meteorological information from multiple remote sensors to predict mesoscale weather events and distributes weather warnings to appropriate agencies. Smart dust sensors[24] and motes[23] are examples of simple devices that self-assemble into complex instruments using wireless networks[21]. All of these projects gather data from relatively simple sensors to build a complex system.

We also need to consider the data streams of instruments that provide remote access to enable interactive control. The simplest are instruments that provide simple binary controls such as assembly line conveyor belts (start/stop), or remote entry systems (open/close). More complex systems include large manufacturing facilities where operators can control operations with single commands that, in turn, produce multiple commands affecting several individual instruments on the factory floor. Allowing write-access to an instrument adds an extra level of complexity by requiring access and safety controls.

Complex instruments can also be combined into even more complex distributed systems. Federal Express[8] can locate your package from a central location by querying complex systems at multiple shipping hubs, which in turn use simple barcode scanners to determine the location of a package. WalMart[11] has built one of the most sophisticated inventory control systems in existence by gathering data on over 20 million transactions per day from systems such as cash registers, warehouses, and delivery vehicles, and processing that data to initiate automatic inventory delivery. Each of these examples can be considered an instrument system built from the aggregation of complex instruments.

#### c. Cyber infrastructure requirements

All of these systems require some degree of cyber infrastructure to support their function. Data storage facilities are required to manage working information and to archive historical data. Computational resources are required to analyze instrument data and distill it into a form useful for decision making. High speed and high reliability data transport mechanisms are required to move data from the instruments to computational resources, and to send the results to a facility that can act on that data. Most instruments require some computational resource embedded within the instrument to perform simple tasks like converting sensor data to a data stream. Higher level analysis of these data streams provides the opportunity for using shared cyber infrastructure, perhaps external to the facility. The success of projects like SETI@home[9] and Folding@Home[10] have made the use of external cyber infrastructure an integral part of their operations.

#### d. Confidentiality concerns

Of special concern in many situations is the issue of proprietary data. Commercial facilities need to keep much of their data private to maintain a competitive advantage in the marketplace. Research facilities commonly retain some degree of privacy on data until the research is peer reviewed and published. Instrument integration efforts must be aware of proprietary data issues and protect information in an appropriate manner, both in terms of confidentiality and integrity.

e. Protection concerns

Another special concern for instrument integration efforts is safety. Instrument integration efforts must carefully consider which capabilities to expose when the manipulation of an instrument could result in loss of life, destruction of property, loss of data, or loss of instrument availability. For example, an online electron microscope such as the Telepresence Microscopy Collaboration[12] might allow horizontal stage movement to all remote users, but restrict vertical movement to skilled operators. Projects that provide interactive control of very expensive instruments such as the Mars Rover[14] may elect to have all control functions limited to a private network. These concerns will direct the choice of computational resources and network transport mechanisms for the integration effort.

#### 4. Appropriateness for TeraGrid Deployment

The TeraGrid provides a semi-public research-oriented network that may be appropriate for efforts like the LEAD Project, but probably is not appropriate for commercial integration efforts like the Madison Paper Twist Wrapper[13]. Projects that are appropriate for the TeraGrid must be approved through a peer reviewed process. Additionally, projects should carefully consider issues such as data confidentiality and safety when considering the TeraGrid as a resource.

#### 5. Instruments already on the TeraGrid

LEAD – Interactive (Doppler radar stations can be retasked to focus on weather events), complex (many sensors), non-proprietary data, life-safety issues

SNS – Semi-interactive (processed data directs subsequent experiments), complex (many sensors), semi-proprietary data, safety issues are explicitly denied

TeleScience – Interactive (telemicroscopy), complex (many data streams), instrument safety issues

NEESGrid – Interactive (teleobservation and telecontrol), complex (many data streams, many instruments), safety issues addressed with NEES Telecontrol Protocol (NTCP).

SPRUCE – A framework enabling on-demand access to computational resources; may also be adapted to schedule instruments.

Others

### 6. Conclusions

Instrument and device integration efforts have traditionally been *ad hoc* custom interfaces, but these efforts are starting to move out of the “back room” and onto mainstream computing resources. The TeraGrid provides a rich variety of resources that all share a common software environment. Technologies currently deployed on the TeraGrid such as Web Services show promise for standardizing instrument interfaces. For integration efforts that have flexible requirements for data confidentiality and safety issues, the TeraGrid provides a large investment in leadership class cyber infrastructure, in both computational hardware and common software environments, that can be leveraged by projects and facilities in lieu of investing in their own dedicated resources.

### 7. Acknowledgements

This material is based upon work supported by the National Science Foundation under the following NSF programs: Partnerships for Advanced Computational Infrastructure, Distributed Terascale Facility (DTF) and Terascale Extensions: Enhancements to the Extensible Terascale Facility.

Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC for the U.S.

Department of Energy under contract number DE-AC05-00OR22725.

## 8. References

- [1] The TeraGrid Project, <http://www.teragrid.org/>
- [2] CMU Coke machine, <http://www.cs.cmu.edu/~coke/>
- [3] M. J. Cox and J.E.F. Baruch, Robotic Telescopes: An Interactive Exhibit on the World-Wide Web, Proc. IASTED Conf. on Robotics and Applications, pages 158-162, Santa Barbara, 28-30 Oct. 1999  
<http://www.telescope.org/index.php>
- [4] The Weather Underground, <http://www.weatherunderground.com/>
- [5] The Spallation Neutron Source, <http://www.sns.gov/>
- [6] SensorNet, <http://www.sensornet.gov/>
- [7] LEAD, <http://lead.ou.edu/>
- [8] Federal Express, <http://www.fedex.com/>
- [9] E. Korpela, D. Werthimer, D. Anderson, J. Cobb, and M. Lebofsky. "SETI@home - Massively Distributed Computing for SETI", IEEE: Computer Science and Engineering, 2001, 3(1), 77-83 <http://setiathome.ssl.berkeley.edu/>
- [10] S.M. Larson, C.D. Snow, M.R. Shirts, and V.S. Pande, "Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology", Computational Genomics, 2003. <http://folding.stanford.edu/>
- [11] C. Babcock, "Parallel Processing Mines Retail Data", Computer World, 6, 1994. <http://www.walmart.com/>
- [12] TelePresence Microscopy Collaboratory, <http://tpm.amc.anl.gov/>
- [13] Madison Paper Industries, <http://www.pulpandpaper.org/html/profiles/madison.html>
- [14] L. Boissier, B. Hotz, C. Proy, O. Faugeras, and P. Fua, "Autonomous Planetary Rover: On-board perception system concept and stereovision by correlation approach" in Proc. IEEE Int. Conf. Robotics Automat., Nice, France, May 1992, pp. 181-186. <http://marsrovers.nasa.gov/home/index.html>
- [15] K. Goldberg, J. Santarromana, G. Bekey, S. Gentner, R. Morris, C. Sutter, J. Wiegley, and E. Berger. The telegarden. In SIG-GRAPH(1995), <http://www.usc.edu/dept/garden/>
- [16] C. Kesselman, T. Prudhomme, and I. Foster, "Distributed Telepresence: The NEESgrid Earthquake Engineering Collaboratory," in The Grid: Blueprint for a New Computing Infrastructure (2nd Edition), I. Foster, Ed.: Morgan Kaufmann, 2004. <http://www.neesgrid.org/>
- [17] e-Science, <http://www.e-science.clrc.ac.uk/>
- [18] ISIS Pulsed Neutron & Muon Source, <http://www.isis.rl.ac.uk/>
- [19] P. Beckman, S. Nadella, I. Beschastnikh, N. Trebon, "SPRUCE: Special PRiority and Urgent Computing Environment", 26 April, 2006. <http://spruce.uchicago.edu/>
- [20] National Animal Identification System, <http://animalid.aphis.usda.gov/nais/>
- [21] Songhwai Oh, Phoebus Chen, Michael Manzo, and Shankar Sastry, "Instrumenting Wireless Sensor Networks for Real-time Surveillance", Proc. of the International Conference on Robotics and Automation, Orlando, FL, May 2006.
- [22] N. Eagle and A. Pentland, "Social Serendipity: Proximity Sensing and Cueing", MIT Media Laboratory Technical Note 580, May 2004.
- [23] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TAG: A Tiny AGgregation service for ad-hoc sensor networks. In Fifth Symposium on Operating Systems Design and Implementation (OSDI '02), Boston, Dec. 2002.
- [24] J. M. Kahn, R. H. Katz, and K. S. J. Pister. Next Century Challenges: Mobile Networking for Smart Dust. In Proc. of Intl. Conference on Mobile Computing and Networking (MOBICOM), August 1999.
- [25] P. Saucy and F. Mondada, "KhepOnTheWeb: One year of access to a mobile robot on the internet", Workshop WS2, Robots on The Web, IROS'98, Victoria, Canada, October 1998. <http://khepontheweb.epfl.ch/>
- [26] Pipeline Pigs, <http://www.pigtek.com/>