OGC Document 11-144

Science Scenarios

An OGC white paper for the NSF EarthCube

Editors: Edwin Welles, Deltares USA and S.J.S. Khalsa, NSIDC
Dawn Wright, Esri/Oregon State University
Robert Chen and Robert Downs, CIESIN
Yong Liu, NCSA
David Arctur, OGC

Introduction

Using the UML template for describing scenarios, we documented 7 different Science Scenarios. The collected Science Scenarios range from global warming studies in the Antarctic to understanding the earth's interior, and several include the social sciences in addition to the earth sciences. In all cases, the science is multidisciplinary and several common themes emerge regarding cyberinfrastructure requirements.

Many of these scenarios represent situations requiring urgent decision support affecting human health and safety with little advance warning, or under real-time data acquisition regimes. In these conditions it is often not feasible to ground-truth the data sources or validate the results of data processing; the data sources and models must be interoperable and work together correctly under stress. This puts a liability burden on potential solutions, which should be considered in their design.

Including the entirety of all the scenarios would make this white paper too voluminous. Therefore, what we present here are summaries of each scenario and an analysis of common technical and non-technical cyberinfrastructure requirements. More detailed requirements have been entered into the Geo-requirements questionnaire.

Scenario 1: Landslide Forecasts

Landslides constitute a major geologic hazard in the U.S. [Wieczorek and Leahy, 2008; Keefer and Larsen, 2007]. Producing actionable information on landslide probability for slopes adjacent to urban areas will enhance public safety in many regions within the U.S. Research to improve our ability to predict landslides requires data acquired from satellites, aircraft, in situ sensor networks and weather forecast models and thus requires improved cyberinfrastructure.

To predict a landslide, one first develops a risk map of unstable slopes, based on soil type maps (typically available from the regional USGS office) and a highly detailed (~10cm) digital elevation model (available from the municipality or developed via LIDAR). Real-time prediction requires monitoring two triggers, soil saturation and seismic activity. Soil saturation data can be gathered from the NOAA National Weather Service weather and watershed models available from the National Centers for Environmental Prediction and the regional River Forecast Centers. Real-time seismic information can be collected from the Incorporated Research Institutions for Seismology Data Management Center. Slow slope motions can be detected from two additional sources, Interferometric Synthetic Aperture data and GPS data from UNAVCO. Accessing and using these data requires using multiple technologies to transmit and read the data: LIDAR data in LAS format, NWS predictions in GRIB format, seismographic data in SEED format, InSAR data in CEOS format and GPS data in RINEX format. Data may be transmitted via direct satellite feed, ftp and web services. This scenario emphasizes the need to provide actionable information based on accessing and processing multiple real-time data sources with varied formats and transmission paths.

Scenario 2: Understanding the Earth's Interior

With this scenario we consider the fundamental science question posed by the NRC "How does the Earth's interior work, and how does it affect plate boundaries, hotspots, and other surface manifestations?" [NRC, 2011] Our understanding of the Earth's interior is critical to a range of geophysical issues having dramatic effects on society, including earthquake detection, the development of volcano and tsunami warning systems, the role and effect of fluids in the Earth's surface, the verification of groundwater aquifers, hydrocarbon and resource exploration, and even nuclear test monitoring and treaty verification [Forsyth et al., 2009]. Investigation below the oceans of the Earth's uppermost interior has contributed to our knowledge of Earth history through the emerging sciences of paleoceanography and paleoclimatology, and has resulted in discovery of a deep biosphere extending as far as 1.6 km below the seafloor and in sediments older than 100 Ma.

Seismology, geomagnetism, electromagnetism, acoustics and geodesy are the main approaches geophysicists use to understand how the Earth's interior works, and these require long-term, broadband seismic and (electro)magnetic stations equipped with a variety of sensors in the oceans to delineate the structure of the lower mantle and core-mantle boundary, and advance knowledge of the structure of Earth's core and the origin and behavior of Earth's magnetic field (Detrick et al., 2006). This is a fundamental research scenario that highlights the need for substantial computing resources, advanced analysis and visualization tools for large data sets and for sensor integration.

Scenario 3: Adaptive Sensing and Model-based Feedback Control for an Agricultural Observatory

Understanding soil moisture and nitrate dynamics is critical to establishing successful agricultural practices. Deploying mote-based smart sensors into the field can help provide better measurements in space and time of these important variables. However, proper management for power consumption, and optimal sensor placement are needed in order to observe relevant variables when interesting rainfall events happen. Addressing these issues requires model-based adaptive sensing and feedback controls to determine where, when, and how often data should be acquired and delivered. Such a model-based system can also help optimize when, where and how much drainage and/or fertilizer should be used. This data-driven model would extract features from the observation data stream and use them to modify sampling frequency. Feeding the collected data to a physics-based agricultural model will permit the sensitivity analysis so that parameters of interest can be studied. This scenario highlights the need for expanded real-time sensor networks and enhanced data exchange.

Scenario 4: Coastal and Marine Spatial Planning

The United Nations estimates that by 2020, 75% of the world's population will be living within 60 km of the coastal zone [United Nations][Shi and Singh]. We need to answer questions like "What is the appropriate spatial scale of management and under what environmental conditions to: keep commercial fisheries sustainable while setting aside marine protected areas; to find the best places for clean renewable energy generation (e.g., wave energy, tidal energy, wind-on-water energy) given existing shipping lanes, commercial and recreational fishing areas, or oil-gas drilling leases; to protect heavily-populated coastal areas in the wake of storms, tsunamis, and other hazards related to climate change?" Coastal resource managers need to analyze and allocate the spatial and temporal distribution of human activities in coastal and open ocean areas to achieve sustainable ecological, economic and social objectives. These objectives have already been specified through political and social process at state, and national scales [Foley, Halpern et al]

To conduct such studies will require adequate spatial data, interactive mapping capabilities, modeling capacity and decision support systems in relation to human use and climate-change scenarios in vulnerable coastal regions. In the US, regional ocean partnerships such as the West Coast Governors Agreement on Ocean Health (http://westcoastoceans.gov/) provide opportunities to build upon existing individual data portals (aka coastal web atlases) at state levels, along with Integrated Ocean Observing System associations at regional levels (http://www.ioos.gov/regional.html) to provide resources to the states and federal government to support informed ecosystem-based management and coastal and marine spatial planning (CMSP). This scenario highlights the need for easy access to and transparency of data and information, the need for advanced visualization and analysis tools and models, as well as the ability to distribute images and data to non-scientist stakeholders.

Scenario 5: Polar Oceanography and Climate Change

Significant warming is occurring not only in the Arctic but the Antarctic where over the past 50 years, the west coast of the Antarctic Peninsula (Weddell Sea) in particular has been one of the most rapidly-warming parts of the planet. Helly et al. (2011) sought to characterize the melt water field around free-drifting icebergs near the Peninsula and to combine it with synoptic, regional-scale data to investigate the role of icebergs in controlling biological productivity in this region.

In addressing this question, they developed a set of methods and software tools to integrate multi-scale, -source, and -disciplinary oceanographic data over several recent research cruises to the Antarctic. As the information was gathered and processed during their cruise, it provided an increasingly rich basis for planning integrated sampling, optimized across disciplinary teams, which brought welcome agility in near-real-time planning of the expeditions and new scientific insights. The authors needed this new cyberinfrastructure to make the first direct observation and characterization of melt water plumes from individual icebergs and to integrate these individual results with regional- and global-scale data [Wright and Wang 2011]. This scenario illustrates the importance of sophisticated spatial capabilities within a cyberinfrastructure as well as high speed data networks between geographically distant groups and extensive computational methods to assist with fusing data with different spatial and temporal scales and sources.

Scenario 6: Contaminant Transport

The U.S. Coast Guard National Response Center (www.nrc.uscg.mil/nrchp.html) reports that spill incidents of all types in the United States numbered more than 35,000 in 2005. Environment Canada (www.etc-cte.ec.gc.ca) reports that there were 742 large oil-tanker spills worldwide for the period 1974-1997; a large spill is one that involves over 1,000 barrels (136 metric tons) of oil released per event in a non-wartime incident. In U.S. waters an average of approximately 3 million gallons of oil or refined petroleum products are spilled every year (NRC 2003). As seen with the 2010 Deepwater Horizon oil spill in the Gulf of Mexico, the primary response methods consist of the deployment of mechanical on-water containment and recovery systems, such as booms and skimmers, as well as the use of oil dispersants (chemical agents such as surfactants, solvents, and other compounds; NRC 2005) to reduce the effect of the oil by changing its chemical and physical properties. The factors controlling rates of the biological and physical processes that determine the ultimate fate of dispersed oil are poorly understood. Of particular concern is the fate of dispersed oil in areas with high-suspended solids and areas of low flushing rates. There is insufficient information to determine how chemically dispersed oil interacts with suspended sediments, both short- and long-term, compared to naturally dispersed oil (NRC 2005). Attention must also be paid to oil spills in the Arctic, as there is a two-pronged danger there: melting sea ice in the face of global warming will attract more shipping and energy exploration. At the same time, there is very little understanding of the effect of oil spills on ice, where spills are harder to track (because it is not possible to follow the sheen on ice), open water techniques will not work (especially as oil gets trapped under the ice), and harsh, remote conditions make it much harder to get cleanup equipment in place. This scenario highlights the need for more sensor networks and integrating spatially and temporally diverse data.

Scenario 7: Integrated Disaster Risk Assessment

Traditional hazard research is organized primarily by discipline, e.g., hydrologists assess flood risks, climatologists study drought risk, seismologists assess earthquake hazards and sometimes tsunami risk, civil and architectural engineers analyze building fragility, psychologists study risk perception, political scientists examine institutional capacity, and other social scientists study population exposure and vulnerability. There is increasing recognition of the need for integrated disaster research, e.g., as evidenced by the new International Council for Science (ICSU) program on Integrated Research on Disaster Risk (IRDR, see http://www.irdrinternational.org/) as well as by large megadisasters such as the earthquake and tsunami in Japan which led to a major technological system failure. Key issues in promoting integrated disaster research include: the ability to compare the spatial and temporal distribution of diverse hazards within a consistent analytic framework; to address critical differences in units of analysis, scale, resolution, and methods between different types of natural, social, health, and engineering data sources; to understand potential interactions between different hazards, complex environmental and human systems, human behavior, perception, and response, and possible risk reduction and management approaches; and to understand the multidimensional impacts of major alterations to human-environment systems such as climate change. This scenario identifies the need for interoperable access to diverse data systems from natural, social, health, and engineering fields, including meta data and ontologies, as well as the computational and visualization methods identified earlier.

Common Requirements for a Cyberinfrastructure

These scenarios include a wide variety of science and consequently cover a wide variety of required cyberinfrastructure. However, there are similarities in the infrastructure needs across this wide range. There are both technical and non-technical requirements in common listed here. Some of these issues are also raised in the accompanying OGC white papers on cyber-architecture and governance.

Technical Requirements

- All of the scenarios required open access to a broad range of databases, including databases storing not just traditional Earth Science data, but also data to support the social sciences like demographic and economic data. In all cases, the researcher required access via catalogs, that is the capacity for data discovery either with automated or manual search techniques.

- Extensive metadata is also required as users are likely to encounter difficulties in accessing and understanding data availability from different disciplines, given different terminology, unfamiliar units of analysis and measurement, and widely varying observational approaches. This suggests the need for additional sources of meta information including taxonomies, ontologies, and detailed descriptions of methods and instruments. This metadata should include a clear description of the provenance of the data and processes applied.

- Displaying data is always an important step in an analysis and so displays are an important element of the scenarios. Scientists need to be able to construct displays from diverse sources and in diverse display formats such as maps (2-D and 3-D), grids, images, time series, pathways (e.g. glider data), and profiles. In so doing, the issues associated with different time steps, projections, grid sizes and the like must be resolved so they do not hinder the analysis. Though not specifically called out, the capabilities to render in 3-D and to "fly through" the data are requirements for any modern display environment.

- Another important form of communication is the capacity to create and distribute displays to non-technical users. This means annotated displays of specific pre-selected data with capabilities to map spatial data and add geographic features, to pan and zoom and loop, to make x-y and other plot types, to construct formatted pages and link to other resources.

- Data manipulation capabilities are also needed. At a minimum there need to be capabilities to make model-to-model, sensor-to-model and sensor-to-sensor comparisons. Users need to be able to add their own analysis methods building on a library of supported tools.

- In several of the scenarios the need was identified for additional sensors and for easy access to output from existing sensors in real time. Because sensors are necessarily very specific, it is not likely to be possible to develop a general solution for additional sensors. However, better access to the current real time networks may begin to address this need.

- All of these cyberinfrastructure requirements assume access to computing resources and almost unlimited network bandwidth. Exchanging the data sets described in the scenarios will only be possible with good compression algorithms and with fast network capabilities. The same is true for the computations and data blending envisioned – they require fast hardware on which to render displays and make computations.

Non-Technical Requirements

- An interesting and important requirement derived from the scenarios is the need for a well-defined governance structure so users can extend the data standards which currently exist to support interoperability. Once a robust infrastructure begins to create opportunities for massive data exchange, the holes in current standards will become apparent and they will need to be updated.

- Users need sufficient intellectual property rights to use the data products and services that are available. Information about such rights must be easily accessible to the users.

- Users need a clear understanding of all restrictions, constraints, and limitations on using data, e.g., the need for protecting privacy and confidentiality, endangered species, indigenous rights, and national security. Such information must be easily accessible.

- Users need documentation describing available cyberinfrastructure capabilities and how to use those capabilities. Training modules and support staff should also be available.

References

Detrick, R., J. Orcutt, D. Chadwell, A. Chave, J. Collins, G. Laske, B. Romanowicz, A. Schultz, R. Stephen and M. Zumberge, 2006. Open-Ocean Observatories: Earth structure and geodynamics –An ORION Conceptual Proposal. Unpublished manuscript, Woods Hole Oceanographic Institution, Woods Hole, MA.

Foley, M.M., B.S. Halpern, F. Micheli, M.H. Armsby, M.R. Caldwell, C.M. Crain, E. Prahler, N. Rohr, D. Sivas, M.W. Beck, M.H. Carr, L.B. Crowder, J.E. Duffy, S.D. Hacker, K. McLeod, C.H. Peterson, H.M. Regan, M.H. Ruckelshaus, P.A. Sandifer, R.S. Steneck. 2010. Guiding ecological principles for marine spatial planning. *Marine Policy* 34: 955-966.

Forsyth, D.W., Lay, T., Aster, R.C., and Romanowicz, 2009. Grand challenges for seismology, *Eos, Trans. AGU*, 90(41): 361-362.

Halpern, B., 2011. Top priorites for the science, policy, and practice of Coastal and Marine Spatial Planning. Workshop results, Dec. 14-15, 2010. National Center for Ecological Analysis and Synthesis, Santa Barbara, CA, 3 pp.

Helly J, Kaufman RS, Vernet M, Stephenson GR (2011) Spatial characterization of the melt-water field from icebergs in the Weddell Sea. *Proceedings of the National Academy of Sciences*, 108(14): 5492-5497.

Keefer, D. and Larsen, M., 2007. Assessing Landslide Hazards, Science 25 May 2007: 316 (5828), 1136-1138.

NRC, 2003. *Oil in the Sea III: Inputs, Fates, and Effects*, Committee on Oil in the Sea: Inputs, Fates, and Effects, http://www.nap.edu/catalog/10388.html.

NRC, 2005. *Oil Spill Dispersants: Efficacy and Effects*, Committee on Understanding Oil Spill Dispersants: Efficacy and Effects, National Research Council, http://www.nap.edu/catalog/11283.html.

NRC, 2008. *Risk of Vessel Accidents and Spills in the Aleutian Islands: Designing a Comprehensive Risk Assessment - Special Report 293*, Committee for Risk of Vessel Accidents and Spills in the Aleutian Islands: A Study to Design a Comprehensive Risk Assessment, http://www.nap.edu/catalog/12443.html.

NRC 2011. *Critical Infrastructure for Ocean Research and Societal Needs in 2030* , Washington, DC: National Academies Press, 98 pp. (ISBN 0-309-18603-X).

Schofield, O., Tivey, M.K. (eds.), 2005. ORION: Ocean Research Interactive Observatory Networks. A Report of a Workshop, January 4-8, 2004, San Juan, Puerto Rico, Geo Prose, Bethesda, MD, 140 pp.

Shi, H. and Singh, A., 2003. Status and interconnections of selected environmental issues in the global coastal zones, *Ambio*, 32 (2): 145-152.

United Nations, 1992. *Agenda 21: The United Nations Programme of Action from Rio*. United Nations, New York, USA, 147 pp.

Wieczorek G., and Leahy. P., 2008. Landslide hazard mitigation in North America. Environmental & Engineering Geoscience, 14(2):133-144

Wright, D.J. and Wang, S., The emergence of spatial cyberinfrastructure, *Proceedings of the National Academy of Sciences*, 108(14): 5488-5491.