Data Management Plan

Technical expertise and oversight for data management – We will organize a Data Management Board to help plan policies and processes for the management of SEGA data. This group will be chaired by Aaron Ellison (Sr. Personnel), of Harvard University who has published in ecology, ecological modeling, and the management of environmental data. Other members of the group include Paul Flikkema (co-PI), Paul Heinrich (Sr. Personnel), Kelly Redmond, Michael Crimmins, Nancy Selover and David Simeral. Dr. Redmond is a research professor at the Desert Research Institute and serves as the Regional Climatologist for the western United States. Dr. Redmond has published extensively in the areas of climatology, climate change, instrumentation and data management. Dr. Crimmins is an associate professor and Climate Science Extension Specialist at the University of Arizona. Dr. Crimmins conducts research and outreach in climatology, drought impacts, fire science and land management. Dr. Selover is the Arizona State Climatologist. Dr. Selover is responsible for the management and dissemination of climate data and forecasts for the state. Dr. Simeral is an associate research meteorologist with the Desert Research Institute, who has extensive experience with the management of a large network of weather stations situated in the southwestern deserts.

Cyberinfrastructure – Data from the weather stations, the watering system, and WiSARDs will be sent via cellular or satellite modem to a database server at NAU (Yamamoto et al. 2007), allowing data access by researchers. We have demonstrated round-trip data transmission times under 30 seconds using this system in the field in N. Arizona. This capability affords near real-time monitoring of the activity and performance of each site's infrastructure, dramatically reducing the cost of site visits for routine inspections. Detection of infrastructure activity can be used to help maintain the array and allow rapid response to problems. The system will be able to send messages to individuals notifying them of the activity. For example, the system might detect a loss of power or water pressure and notify garden managers to the problem. Additionally, the system can be used to notify researchers of interesting environmental events, such as heavy rainfall or unusual temperatures. More importantly, it transforms the SEGA instrument into a sensor/actuator internetwork that will support experiments where natural temporal precipitation or soil moisture patterns at higher elevations can be accurately replicated and verified in near real-time at lower elevations.

Data management -SEGA staff will develop a formal data management policy (Mitchner & Brunt 2000; Osterweil et al. 2010) for all environmental and engineering data collected by the SEGA instrument that will address data access, quality control, storage and security. To the extent possible given variable funding source and partner institution policies, we will apply the SEGA data policy to user-collected data as well. Our data access framework will include public and private access levels that reflect the needs of both investigators (e.g., prior to publication) and the public. More levels of data access will be added as needed to address special circumstances, such as archeological issues. All data will be classified as "public" within 3 years of collection. Most climate, weather and some engineering data collected by the SEGA instrument will be publicly available in near real time via the web. Raw data from all sites will be automatically assimilated, checked for anomalous data (with automatic reporting and logging of potential problems and failures) and archived in replicated RAID-based database

servers in NAU's Wireless Networks Research and GRAIL Laboratories (with periodic tape backups) (Cook et al. 2001, Mitchner & Brunt 2000, Osterweil et al. 2010). Raw data will be made available through web portals and possibly Open Database Connectivity connectors in a read-only mode. Access to SEGA web portals will be controlled through a standard user/roles system where users are assigned a role which defines their rights within the system. For example, a public role might only allow a user to view public data, a researcher role would allow a researcher to control irrigation on his or her experiment and a manager role would allow a manager to assign roles to other users. All user commands and queries will be logged within the database. SEGA web portals will generate both tabular and graphical products for researchers and the general public. The SEGA portals and control interface will be available anywhere with internet connectivity including the SEGA sites themselves. All spatial data associated with the SEGA project will be stored in a geodatabase in the NAU GRAIL facility, and will be presented online through mapping applications.

We will enforce use of standards-based metadata (Cook et al. 2001, Ellison 2010) for all variables to 1) publicize SEGA datasets through the Knowledge Network for Biology, National Biological Information Infrastructure and Merriam-Powell Center for Environmental Research (MPCER) metadata clearinghouses, 2) classify data according to our data access policy, 3) maintain knowledge of the composition, organization, and quality of the data, and 4) document detailed data schema to ensure easy and complete understanding of the dataset. We will use the NCEAS-developed Ecological Metadata Language (EML) standard for most metadata, the Morpho editor, and Metacat for storage and management (Higgins et al. 2000). For spatially explicit datasets such as site maps and project geodatabases we will use the Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (CSDGM v. 2) and transition to the new ISO 19115 international standard as it is adopted. These metadata will be uploaded to the MPCER metadata clearinghouse and digested by both the National Spatial Data Infrastructure and by the NBII metadata portals automatically. MPCER has been a NBII and NSDI node and has conducted federally-funded training in spatial data documentation for several years.

Project Documentation – We will formally document all array processes using standards-based methods whenever possible. This will include a site management document to define site-level infrastructure responsibilities and methods, an instrument management document to specify methods and responsibilities for sensor maintenance, a sustainability document to describe the methods used to ensure long-term sustainability of the array, a data management document detailing how data will be collected, archived and distributed, a data ownership policy describing the rights and responsibilities of the array managers and researchers (following NSF data ownership policies), a data liability statement defining the limits of data liability for the array, researchers and the cooperators, and an interface control document describing all details of how data is passed among SEGA components and subsystems and that defines all formats, codes or algorithms used within the array instrumentation and software. Software development will use version control software licensed by the university. All code, documentation, format definitions and other documentation will be made available on a project website as soon as those documents are finalized. Working versions of some documents may be made available before formal adoption.