

UNAVCO Annual Report

EarthScope Comprehensive SAR Archive

Report Date: July 2011

For period of performance September 2010 – August 2011

NSF Grant EAR-0952375

Submitted to:

Russell Kelz

National Science Foundation, Instruments and Facilities Program,
EarthScope Science Utilization

Craig Dobson

NASA Headquarters

Elizabeth Lemersal

USGS

Submitted by:

Fran Boler

UNAVCO

1.0 Executive Summary

In February, 2011, NSF and NASA approved a realignment of WInSAR and EarthScope Comprehensive SAR archive funding to allow for significant investments in the SAR archive infrastructure at UNAVCO. The impetus for the infrastructure upgrade request was the reduced need for data funding due to the European Space Agency's (ESA) revised open data policy, with the corresponding freeing of funding previously set aside for data tasking and purchase as set out in the grant proposals. The WInSAR Executive Committee requested part of the upgrade to include an application programming interface (API) and associated client software to allow rapid data download from UNAVCO. Upgrades to hardware, the SAR database, and the UNAVCO SAR dynamic search and informational web pages are also part of the upgrade plan. An InSAR specialist has been hired in a temporary part-time capacity to provide domain knowledge for database improvements and the API development. The work is progressing well, with hardware purchases made, database schema changes being implemented, and API design and implementation underway. In addition, a website survey was conducted to obtain users' input prior to overhauling the overall SAR information and data access web sites.

NASA's contribution towards this award has been directed towards support of the International Group on Earth Observations (GEO) Geohazard Supersites Initiative. The Group on Earth Observations is coordinating efforts to build a Global Earth Observation System of Systems, or GEOSS. A principal aim of GEO is international cooperation in data sharing to support disaster preparedness planning and disaster mitigation. The Geohazard Supersites provide access to spaceborne and in-situ geophysical data of selected sites prone to earthquake, volcano or other hazards. UNAVCO's role in the SAR area has been to provide SAR data ordering from ESA; data management for the ordered data; and web site hosting and event Supersites content management.

2.0 Expanding the content of the EarthScope and WInSAR SAR Archives

2.1 The Proposed Work

The EarthScope Comprehensive SAR Archive proposal set out a vision for building on the extensive archive content developed through the GeoEarthScope InSAR archive acquisitions and ongoing WInSAR data orders. Funds were requested for imagery acquired by ESA largely after October 2008, the end of the GeoEarthScope component of the EarthScope facility construction project. Funds were also requested to task the Envisat satellite until October 2010 when Envisat's InSAR capability was expected to end. The proposed new imagery combined with the available GeoEarthScope and WInSAR imagery was expected to form a complete, unprecedented 1992-2011 SAR data set for the WInSAR core region, and lead to new scientific discoveries within the EarthScope region as well as to the development of new InSAR methods aimed at improving the accuracy of the measurement. The new data would facilitate the ability to estimate the InSAR vertical for the EarthScope region.

2.2 Change in Data Policy at ESA Drives Change in Priorities for Support

At the end of 2010, ESA confirmed that costs were waived for all data ordered in 2010, and that henceforth data could be ordered and would be delivered without charge. Tasking fees were also waived. UNAVCO had incurred approximately \$130,000 in charges both for tasking and data throughout 2010 that were waived. The Comprehensive SAR Archive proposal had been in budget negotiations until mid 2010 with precise funding amounts for data purchase still in question. UNAVCO held off on additional data orders in 2010 until the budget negotiations were completed. With the change in ESA data policy to the free data model, the remainder of the ordering planned for 2011 under this project to complete the data purchase that was the main purpose and intent of the proposal could be accomplished without expending the funds allocated in the grant.

Meanwhile, the WInSAR Executive Committee (EC), as the body that drove the original proposal and that acts as a standing committee advisory to the UNAVCO Board on SAR matters, evolved a plan in conjunction with UNAVCO to address some longstanding needs within the information technology hardware and software supporting the InSAR operational activities at UNAVCO. The need for a user controllable Application Programming Interface (API) to allow automated, GUI independent search and data download was identified as a high priority need by the WInSAR EC..

To address both emerging and longstanding needs, in early 2011 the WInSAR EC vetted and tentatively approved an approach for infrastructure upgrade at UNAVCO. The new work plan and budget realignment were presented in a teleconference with Russ Kelz and Craig Dobson, Program Managers, and the plan was approved in early February, 2011. A target timeline of one year to complete the major elements of the upgrade was envisioned. Section 3 of this report covers the infrastructure upgrade that is currently in progress.

3.0 UNAVCO SAR Archive Infrastructure Upgrade

Since the infrastructure upgrade plan was put in place by the WInSAR EC and the budget realignment was approved by the funding agencies in February 2011, significant strides have been made towards the improved infrastructure envisioned. The upgrade process can be divided into the following major elements: project management, hardware, software (including database and API), and web site.

In the subsections below, this report will detail the pre-upgrade status of the SAR infrastructure at UNAVCO, and the status of activities in each of upgrade elements.

3.1 Pre-Upgrade SAR Archive IT Infrastructure

To appreciate the scope of the upgrade project, it is helpful to review where the infrastructure technology stood at the start of 2011. The WInSAR Archive software system was obtained from Stanford and installed at UNAVCO in 2005. The system included the functions of providing an archive for preservation and distribution of the raw data, a database for metadata storage and retrieval, a second database for membership and data orders management, and a website with both informational content and dynamic tools to show the WInSAR data catalog and allow for data access. UNAVCO contributed one server for ingestion processing, running the database, and hosting the website, and about 4 Tb of storage RAID for holding the raw files. This was older hardware that had been retired from other UNAVCO activities.

At the time, the support for WInSAR through the NSF, NASA and the USGS was dedicated solely for data acquisition through purchase from the European Space Agency. The WInSAR grant included minimal budget for personnel, and no budget for hardware. The GeoEarthScope SAR Archive was instituted in 2007, and \$93,000 was spent for about 48Tb of RAID storage hardware. The storage came with processing servers for archiving and maintaining the GeoEarthScope collection including the database. Because it was simpler to host the GeoEarthScope SAR web site alongside the WInSAR website, the WInSAR hardware was used for both web sites. As the GeoEarthScope SAR archive was being built, the need to upgrade the WInSAR database was recognized and effort was put in to make that a reality and to institute the ESA - requested minicat method of approving data access. This was sufficiently different from the WInSAR model that significant schema differences were needed. Meanwhile, the WInSAR production system and all of its elements needed to continue as they were to continue to serve the WInSAR community. Thus the software, hardware, and database supporting the infrastructure had bifurcated as of the initiation of GeoEarthScope.

Since the end of the construction phase of PBO, the GeoEarthScope infrastructure has remained relatively maintenance free. However, with the changed data policy at ESA, UNAVCO has more freedom to serve the EarthScope SAR collection in new ways that are more useful for data access.

3.2 Project Management

To achieve a wide-ranging hardware and software upgrade such as this in a short period of time while drawing upon numerous individuals for short periods of focused effort requires careful attention to project management throughout the upgrade process. Technical management for this upgrade project has been undertaken by Dr. Lou Estey, long-time UNAVCO employee with extensive knowledge with the parallel GNSS archive IT infrastructure at UNAVCO. Dr. Estey is coordinating the specification and configuration of new hardware, the allocation of tasks to individuals with appropriate expertise, and establishing timeframes for completion of tasks and the overall upgrade.

One of the shortcomings of hosting the SAR infrastructure at UNAVCO has been the absence of hands-on scientific expertise within UNAVCO staff with SAR data use and

InSAR processing. Without this expertise, UNAVCO technical staff is left to do some research and some checking in with community experts from time to time, but the ability to evaluate critical capabilities such as the SAR metadata held in the schema that allows effective search by users can be hit and miss using this process. As we approached this infrastructure upgrade we determined that it was of critical importance to have SAR expertise on staff, and the WInSAR EC agreed. UNAVCO moved ahead with hiring Scott Baker, University of Miami PhD student working with Falk Amelung on InSAR processing and analysis with application to volcano deformation, as a temporary part-time staff person participating on the activities related to the SAR infrastructure upgrade. Scott Baker is funded at about .35 FTE. His temporary appointment with UNAVCO cannot exceed 1000 hours.

Other technical personnel contributing to this project, in addition to Dr. Estey and Scott Baker are: Susanna Gross (database and operations), Stuart Wier (database and operations), Jon Davis (database), Andy Gorman (API development), Jim Riley (web), Matt Beldyk (advisory on API development, database install), Jeff McWhirter (advisory on API development).

UNAVCO staff share the approx 1 FTE available for technical developments plus 0.5 FTE for operations. Some of the individuals are expected to provide just a few days of effort at the right time, with the key API developer devoting the most total hours to the project.

3.3 Hardware Upgrades

A critical need in the SAR archive project has been to have additional robust, maintainable storage. The WInSAR and GeoEarthScope systems had together utilized the 48 Tb of RAID storage purchased from GeoEarthScope funding along with 4 Tb of WInSAR RAID for long term data storage, preservation, and for data distribution functions. There was little scratch space for data management and no room for expansion.

Data ordering has continued under the WInSAR order request system and data have been added at a rate of about 0.5 Tb per year. This ordering does not yet include the major infill of about 5 Tb of post-2008 SAR data for the western US that was proposed in this grant. Ordering of all of the data sets desired are now limited not by cost, but by the responsiveness of the ESA data production system. Over the summer of 2011, new interactions by the WInSAR EC with the German Space Agency (DLR) is leading to new prospects for orders of TerraSAR-X data. In summary, the expectation is that the SAR archives at UNAVCO will continue to have a high need for robust, managed RAID storage.

In addition to requirements for storage of new data orders, the community has identified the mission and geographically sorted structure (relative orbit and frame) on an ftp server provided in the WInSAR archive as a goal for the GeoEarthScope archive, at least for the ESA data. With the change in ESA data policy, there is no longer a strict requirement of segregation of these two collections. Providing part of the GeoEarthScope collection with

access analogous to the WInSAR system, allowing for download via links (after authentication), simplifies some data delivery aspects of the API development. To provide this geographic sort copy of the ESA data while retaining the long-term storage and preservation archiving requirements, translates to the need for a minimum of 5 Tb of additional storage to be used both for long term preservation and for data management operations.

With all of the above needs, plus anticipated future needs for storage, UNAVCO has elected to devote the bulk of the hardware funds available through the budget realignment to upgrading the existing storage systems by replacing 92 0.5 Tb drives with 1 Tb drives, effectively adding 60 Tb of usable storage space at a cost of \$30,000. These drives have been ordered and received, and are currently being configured within the existing (GeoEarthScope-purchased) RAID systems.

In addition to the storage capacity deficit addressed above, a second critical hardware need has been the replacement of the aging server that has been the heart of both the GeoEarthscope and WInSAR databases and web sites. (The upgrade activities for the revamped web site is covered below.) The server running the database engine will either be the processors associated with the RAID storage or another suitable powerful server to be determined.

3.4 Software, Database, and API

As part of the SAR upgrade project, UNAVCO has identified a number of steps that must be completed so that we can provide a flexible, responsive API that serves the user community needs. Our InSAR domain expert, Scott Baker, has been contributing heavily to the upgrade activities within this area. The WInSAR EC provided descriptions of the desired functionality for an API that would meet their needs. Scott turned these into a formal Software Requirements Specifications document.

Scott has also been helping to design and vet the database schema revisions that are the underpinnings of the fully-functional API. The SAR system database at UNAVCO is a hybrid of an inherited WInSAR database and the UNAVCO-developed GeoEarthScope database. The database engine has been mysql. A major database schema merge (for the WInSAR and GeoEarthScope schemas) and evolution has been an ongoing activity for about 6 weeks. Part of the schema evolution allows accommodation of TerraSAR-X data within the archive. In order to be able to accommodate spatial queries for polygon overlap, we have elected to migrate from the MySQL database used up to this point, to the open source Postgres database engine with the PostGIS spatial extension. Implementation of the new schema in test mode is underway within the Postgres database engine. Many components of the production ingestion and data access web pages depend on the database, so migration requires careful testing of all components at each stage. When the database schema merge with enhancements, plus the Postgres/Postgis migration have each been implemented, the result will be faster return from queries, and the user will be able to pinpoint more precisely the data that are desired.

UNAVCO has been engaged in development of a flexible web-services based architecture for the NASA ROSES ACCESS project to modernize the Geodetic Seamless Archive Centers (GSAC) distributed geodetic data access system. The main component of the software is a GSAC Service Layer and GSAC Repository that provide for interaction with a database and with resources such as data files being served. In the GSAC system this architecture includes an API. UNAVCO has elected to investigate whether utilizing the SAR web services and API from a generalization of this GSL, further leveraging the GSAC project software development, would meet the requirements of this project. A test implementation of this concept with a test SAR database indicates that this reuse of the GSAC software will meet the requirements of the SAR community for an API. Final testing awaits the completion of the database migration that is in progress.

During the requirements gathering and API development so far, UNAVCO has been in touch with the Alaska Satellite Facility, which is developing their own API for SAR data access. Ideally, a coordination of development would lead to interoperability among APIs; however, to achieve full interoperability, close coordination would be necessary and could require additional resources.

3.5 Web Sites

There are currently two sites where users can gain information about UNAVCO's SAR holdings corresponding to the WInSAR and GeoEarthScope collections. The main WInSAR web site structure was inherited along with the database and software from Stanford in 2005. There have been some incremental improvements to the web site content, and several major improvements to the dynamic web pages that support data search and access including a Google Maps based spatial display search page. The look of the WInSAR web site has remained the same as what was inherited. The GeoEarthScope SAR project was a major effort that required visibility through the web site and that has provided a rich data set utilized by many. It was also subject to some constraints based on the agreement with ESA on how the data would be distributed (through the "minicat" process which allows access after an application is accepted). With a change in ESA data policy, some of the need for distinction has been removed.

UNAVCO conducted a survey in May to get feedback on directions for a SAR web site and tools. This was helpful in prioritizing how to approach the work and how each of the tools and pages are utilized. Among other conclusions, the survey indicated that a unified web site would benefit the users by consolidating background information and access points. The planning for evolution of the unified site following the insights from the survey is in progress.

The web site will be migrated to UNAVCO's main web server, which will provide robust service and allow leveraging of UNAVCO look and feel.

4.0 Additional Activities in Support of SAR Data Management, Access, and Distribution at UNAVCO

4.1 Interferometric SAR Computing Environment

UNAVCO has for several years hosted short courses on nearly an annual basis to provide professional training in SAR processing to community researchers. The processing software ROI_PAC has been the open source processing package taught during the short courses. The developers of ROI_PAC (from JPL and Stanford), who also participate as short course instructors, have recently released a newly revamped version in their Interferometric SAR Computing Environment package. They requested and the WInSAR Executive Committee approved that the license management and code distribution for this software be managed by UNAVCO. The final preparations for this activity are nearly completed.

4.2 Geohazard Supersites Support

NASA had identified as a key priority for funding to UNAVCO under the Comprehensive SAR Archive grant, the continued operational support of GEO's Geohazard Supersites initiative, including website hosting, Supersites events page setup and content management, data ordering, management of upload of data into the Supersites archive, and other miscellaneous tasks. With the budget realignment, these tasks continue to be supported by UNAVCO staff and hardware resources.

Data ordering (from the European Space Agency), along with data management (download and repackaging) of the orders received, and upload of the orders to the ESA supported cloud storage (Level 3 and Hetzner storage services) is the major activity. UNAVCO has placed orders for over 6000 scenes from ESA during the report period in locations including Japan, Greece, Turkey, Etna, Iceland, and Hawaii, and received 5200 files for repackaging and upload. In addition, data previously ordered by US scientists for their research in global Supersites locations are being contributed to the Supersites initiative; data management for more than 3000 scenes has been provided.

In addition to this data management operational activity, UNAVCO provides web site content management and hosting for Supersite Event pages, the most recent case being the Japan earthquake site. Content management for the event pages requires constant attention over the days following an event as scientists have utilized this mechanism to get out early results. Susanna Gross, UNAVCO's SAR data engineer, has acted as the point of contact and content curator for these pages, which have received significant attention as a place to find comprehensive early results based on SAR.

