



## **TECHNICAL MEMORANDUM**

**TO:** Chad LaFontaine  
Executive Director  
Southeastern States Air Resources Managers, Inc. (SESARM)

**FROM:** Regi Oommen, Eastern Research Group, Inc. (ERG)

**DATE:** December 21, 2020

**SUBJECT:** Task 10 – SESARM Data Handling and Sharing Summary

---

### **1.0 INTRODUCTION**

For Task 10, the ERG/Alpine team consulted with SESARM to implement and utilize multiple solutions for the distribution and short-term and long-term archival of project assets (e.g., emissions and air quality modeling outputs, summaries, and other project documentation). Over the course of this project, tens of thousands of data products, images, files, and programming codes were developed to support Regional Haze SIP development for the SESARM member states, as well as to benefit other stakeholders.

For this project, ERG has acted as the data librarian, which was a role that Alpine served in the first VISTAS study. For the first VISTAS study, the most practical solution for data sharing of the data to the stakeholders was by mailing 300 gigabyte (GB) external hard drives to groups requesting large datasets. Since that time, computer communications have become much faster than they were in the mid-2000s, and the size of the datasets has expanded.

### **2.0 DATA HANDLING AND SHARING**

One of the key challenges was distributing these pieces to the multiple stakeholders for review, comment, revision, and usage. ERG and Alpine worked with the SESARM to develop an effective data handling and sharing scheme. Essentially, five solutions were implemented throughout the life of the project and are described in this memorandum.

#### **2.1 File Transfer Protocol (FTP)**

Early in the project, files were posted for storage, access, upload, and download on ERG's password-protected FTP site. The link and password were only distributed to the key

stakeholders. This solution, while simple to execute and control, provided limitations of long-term storage, space, and accessibility for certain users. The organization of the FTP site consisted of placing files into folders specific to each task and subtask. This was beneficial in sharing and review of state-specific emissions files, which would often exceed size limitations of stakeholder e-mail servers. The FTP was also the solution for facilitating the transfer of large emissions and select modeling files between ERG and Alpine.

## **2.2 Contractor Network Storage**

All interim and final data files developed by ERG and Alpine are retained on their respective networks, and will remain active for at least one year after project conclusion. As such, backups and maintenance are routinely applied to maintain the integrity and security of the data files. Access to these files is limited to only project staff at the respective companies, and the files are not publicly available. Alpine will retain project files for five years from the end of the project while ERG will indefinitely retain project files on tape drives and hard drives. Files can be retrieved with little effort in case the SESARM-retained project files are lost or corrupted.

## **2.3 External Hard Drives**

Two sets of external drives that are used for this project. The first set contains all project-related files, excluding modeling input and output files, while the second set contains the modeling input and output files.

### *Project files (excluding modeling inputs and outputs)*

The first set of external hard drives contain all the project files, excluding the modeling input and output files. One external drive is owned by SESARM (one terabyte [TB]) and the other was purchased by ERG specifically for this project (four TB). The files preserved on these drives will be exact copies and will contain all the files needed for transparency and any necessary reproducibility. At the end of the project, the SESARM drive will be sent back to SESARM archival according to federal grant records retention requirements and for provision to member states and other interested parties when no other source of the data is conveniently available.

### Modeling files

The second set of external hard drives are a series of nine Serial Advanced Technology Attachment (SATA) hard drives of varying sizes containing the modeling input and output files for each of the final modeling runs. They include:

- Drive 1 – 2011/2028 CAMx and spatially and temporally merged emissions (10 TB drive);
- Drive 2 – 2028 elv5 CAMx 6.40 emissions and SMOKE processing files in the VISTAS 12-km domain (4 TB);
- Drive 3 – 2011 el outputs, VISTAS 12-km domain (8 TB drive);
- Drive 4 – 2011 el outputs, continental U.S. (CONUS) 12-km domain (10 TB drive);
- Drive 5 – 2011 CAMx 6.32 outputs for EPA and VISTAS modeling (8 TB);
- Drive 6 – 2028 elv3 CAMx 6.40 outputs for CONUS (12 TB);
- Drive 7 – 2028 elv3 CAMx 6.40 outputs in the VISTAS 12-km domain for PSAT tagging rounds 1 and 2 (12 TB);
- Drive 8 – 2028 elv3 CAMx 6.40 outputs in the VISTAS 12-km domain for PSAT tagging rounds 3 and 4 (12 TB); and
- Drive 9 – 2028 elv5 CAMx 6.40 outputs in the VISTAS 12-km domain (8 TB).

The files on these drives can be accessed using a Wavlink docking station, which can be plugged into the USB port, or the data can be more rapidly read by mounting the drive directly to a computer using the SATA interface. These files were prepared by Alpine and shipped to ERG for the duration of the project. At project end, these data files will be physically stored at the SESARM office location or offices of a VISTAS state if determined to be more appropriate, and available for duplication or shipping if a VISTAS state has a need for this level of detailed data. The drives and data are designed to be compatible with Linux EXT-3 filesystems that are easily accessible on the Linux computer systems which are required to run the CAMx model.

Only the first drive (2011/2028 CAMx and spatially and temporally merged emissions) is necessary for the states to run the model on either the VISTAS 12-km domain, or the smaller state-specific domains (as requested in Task 11.1). If the states wish to modify the emission files, then the second disk (2028 elv5 CAMx 6.40 emissions and SMOKE processing files in the VISTAS 12-km domain) will also be required. README files on each of these disks explain the organization of the modeling files.

## 2.4 SharePoint

North Carolina's Department of Environmental Quality (NCDEQ) offered supplemental VISTAS project data handling and storage services on a SharePoint platform. The SharePoint platform allows for collaborative editing, parallel review, and file exchange for multiple users. Its primary functions include:

- A central web site to provide quick access to project assets and news for stakeholders.
- File exchange and storage. Folders within the SharePoint platform have been used to organize project assets.
- Access control. The SharePoint site is not publicly available. Access to this site is granted by NCDEQ. Using SharePoint user groups, NCDEQ customizes permissions so only the properly credentialed individuals have access to specific files and folders.
- Project asset distribution. The SharePoint platform is configured to send out notifications related to file publishing.
- Project asset archives. The SharePoint platform uses version control and a recycle bin for file retention. Version control allows keeping previous versions of files as needed, and the recycle bin provides a mechanism for restoring files when an entire site collection is mistakenly deleted.

The information in the SharePoint platform is organized by task folder. In addition to the final reports, interim and draft reports are archived. The SharePoint platform also houses thousands of images and data files within Class I area folders for stakeholder use.

## 2.5 SESARM Website

The Metro 4/SESARM website hosts information about the project under the "Technical Center" location (<https://www.metro4-sesarm.org/content/vistas-regional-haze-program>). The "Technical Center" will ultimately contain task-level folders housing all task and final reports in .pdf format and for some tasks, data products. For example, the Area of Influence spreadsheets developed for each Class I area within the SESARM domain are posted in the Task 5 folder. Additionally, project materials such as presentations, handouts, and consultation letters are posted for public review.

SESARM staff are responsible for providing access to all data and reports critical to regional haze SIP development and approval. Maintenance of this website and retention of other

supporting information in the various archival resources noted in this report will ensure the availability of information to the VISTAS states, Federal Land Managers, and other interested stakeholders for an adequate amount of time to serve the VISTAS states' regional haze SIP needs and the requirements of the federal grant that funded this VISTAS project.