

# 7. Data Management

(final version: 15 April 2009)

The development and maintenance of a comprehensive and accurate data archive is critical in meeting the scientific objectives of VORTEX2. An archive that allows VORTEX2 investigators to share data with each other and with the rest of the scientific community helps these investigators promptly publish significant findings and transfer new knowledge into operational severe storm prediction. The overall guiding philosophy for VORTEX2 data management is to make datasets available to the scientific community as soon as possible after the field project while at the same time accommodating legitimate interests of investigators who collect the data.

## 7.1 VORTEX2 Data Policy

In order to accomplish the goals described in the VORTEX2 Scientific Program Overview (SPO) and in individual proposals, VORTEX2 investigators will use a combination of field observations and operational data. Policies for using the VORTEX2 field observations are described in the current and next sections while procedures for archiving and accessing operational data are described later.

Individual investigators are devoting significant time and resources in preparing to collect, collecting, and analyzing VORTEX2 field observations. After the 2009 and 2010 field phases, investigators will need ample time to identify cases of interest, preprocess observations (data conversion, quality control, etc.), analyze observations, and to begin publishing results. An **initial data-analysis period** is planned to accommodate the needs of individual investigators. The initial data-analysis period lasts 16 months after the data are collected (June 2009-October 2010 for data collected in 2009 and June 2010-October 2011 for data collected in 2010).

**VORTEX2 Principal Investigators** (PIs) are those who are listed as PIs and/or co-PIs on VORTEX2 grants and/or facility requests that have been funded by NSF and NOAA. The VORTEX2 PIs will participate in **VORTEX2 Research Workshops** in October/November 2009 and 2010 to prioritize cases, discuss research avenues, and develop plans for analyzing VORTEX2 data efficiently in a coordinated and collaborative manner. Having an initial data-analysis period and coordinating research through workshops satisfies the following project needs: (a) significant findings should be published promptly; (b) PIs need time to peruse the data they collected and to prepare them for analysis, sharing, and publication; (c) research niches should be identified for students, who must produce original research and typically require 1-3 years to do so; and (d) research naturally progresses from analyses of data from a single or small number of instruments initially to integrated analyses of multi-platform observations later.

In addition, VORTEX2 PIs will participate in an **End-of-Season Meeting**, which will occur in the field immediately after the last day of operations (i.e., most likely one or two days after the last day of operations). A preliminary discussion of analysis plans will occur, so that PIs can begin analysis and have results in time for summer and fall conferences. The more comprehensive plan for coordinated data analysis will not be developed until the October/November VORTEX2 Research Workshop.

After the VORTEX2 Research Workshops, VORTEX2 PIs shall promptly provide data to other PIs upon request and in accordance with the coordinated analysis plans developed at the workshops. Data sharing can be accomplished either through direct communication among PIs or through the VORTEX2 Data Archive (Section 7.4) with the permission of the PI who collected the data.

After the 16-month initial data-analysis period, all VORTEX2 data will be considered public domain. If the team of investigators who collected a dataset wishes to make the dataset available in the public domain before the end of the 16-month period, they may do so. Until a dataset is in the public domain, the dataset may not be provided to a third party (someone who is not a VORTEX2 PI) for journal articles, presentations, research proposals, etc. unless the team of investigators who collected the data provides consent.

Users (including both VORTEX2 PIs and third parties) of VORTEX2 field data should acknowledge and/or offer co-authorship to the investigators who collected the data.

## 7.2 Data Sharing in the Field

The VORTEX2 mission depends on coordinated yet somewhat autonomous deployment of multiple vehicles in a rapidly-evolving, complex environment. Because the value of the dataset depends on the degree to which coordination is achieved, it is in the interests of the VORTEX2 community to, as quickly as practical, examine the collected data such that a full understanding of the quality, coverage, scope, duration, and other parameters of data sets collected may be assessed. In this manner, the successes and failures of each mission will be used to strengthen the possibility for success in future missions by enabling us to dynamically adapt observational strategies, to prioritize project goals, and to detect instrument issues or incompatibilities or discrepancies. Furthermore, development of a database containing the known attributes of all cases will allow the earliest possible discussions among PIs concerning analysis collaborations. In order to achieve this goal, the teams that are fielding instruments are strongly encouraged to share their data according to the following guidelines.

It is encouraged strongly that unedited raw and translated data from all platforms will be shared within 24 hours of collection when there are down days between field operations. It is understood that exceptions may occur in unusual circumstances. It is encouraged that in situ measurements (e.g., mobile mesonet, StickNet, tornado in-situ probe, etc.) will be provided in Excel readable ascii spreadsheet format. It is encouraged that radar data will be provided in both raw and commonly readable (dorade sweep, uf, and/or netcdf) formats. It is encouraged that this data sharing be as complete as possible, encompassing the entirety of the collected data.

To facilitate this process, a standardized metadata log sheet for field logs/notes, navigational information (time, lat/lon, pointing angles, etc.) will be provided to each platform. It is encouraged that crews of platforms use these forms. VORTEX2 will likely provide a hotel room/office that can serve as a data sharing/archiving center during down days. Any hand-written or non-standardized notes can be scanned and uploaded. Community data-archiving Windows and Linux computers will be available so instrument leaders can physically, or through ftp, upload and download datasets. Copies of the data can be uploaded to individual PI's disks/computers and uploaded to their home institutions. The intention is good faith sharing of data as complete as possible to enable preliminary discussions/evaluation of collaborative research, to identify quality control concerns, and to evaluate coordination in the shortest practical time.

Data from instruments that are not included in the 24-hour sharing list above are still expected to be shared freely, in a timely fashion, with all VORTEX2 PIs. Instrument leaders from these latter groups are encouraged to follow the 24-hour policy.

It is understood that quality control will not have been completed.

It is agreed by all PIs who participate in this field sharing of data that analysis teams and collaborations will not be finalized and cases or aspects of cases will not be allocated until the post-field phase meetings (End-of-Season Meeting in June and Research Workshop in October/November) described above. Put in other terms, PIs agree not to begin publication-quality analyses with other instrument data until the distribution of cases is finalized at these meetings.

## 7.3 On-Line Field Catalog

An on-line **Field Catalog** (implemented using a WWW browser interface; <http://catalog.eol.ucar.edu/> ) will be functional during the VORTEX2 field project to support real-time planning. The real-time Field Catalog will contain data of three basic types: operational (NOAA/NWS, etc.) data, images from experimental real-time numerical weather prediction models, and field reports. A summary of real-time data that will be available in the Field Catalog is being compiled in Appendix D.

Data-collection information about both operational and field datasets (including metadata and overview documentation) will be entered into the system in near real time beginning on the first day of the field phase in May 2009. The catalog will permit data entry (data-collection details, field-summary notes, certain operational data, etc.), data browsing (listings, plots), and limited catalog information distribution. Designated VORTEX2 participants will prepare daily summaries containing information about operations (sampling times for major instrument systems, weather forecasts and synopses, etc.). These summaries will be entered into the on-line catalog electronically (via WWW interface and/or e-mail). It is important and desirable for the PIs to contribute

graphics (e.g., plots in gif, jpg, png, or postScript format) and/or data for retention on the catalog whenever possible. Updates of the status of data collection and instrumentation (on a daily basis or more often depending on the platforms and other operational requirements) will be available. Public access to status information, mission summaries, and selected datasets will be available.

## 7.4 Distributed Data Archive

The primary **VORTEX2 Data Archive** will be located at NCAR EOL in Boulder, CO, and data will be accessed through a WWW browser interface. The Data Archive offers scientists access to operational data, field documentation (e.g., situational awareness graphics, daily operations summaries, mission summaries, status reports, and mission scientist reports), and field data. It provides the means to identify datasets of interest and automatically obtain data and metadata via Internet file transfer. The user may browse data to preview selected datasets prior to retrieval. Data displays include time series plots for surface parameters, skew-T/log-P diagrams for soundings, and gif images for model analysis and satellite imagery. Users can download data via the Internet directly to their workstation or personal computer or request delivery of data on magnetic/optical media. Data may be selected by time or location and can be converted to one of several formats before delivery. The Data Archive automatically includes associated documentation concerning the data itself, processing steps, and quality-control procedures.

Some datasets will be archived at the home institutions of the instrument managers rather than in the EOL Data Archive center. These distributed data centers will typically be necessary to manage raw datasets that are too large to be stored in EOL Data Archive center (e.g., datasets from mobile radars and high-resolution model output). Instrument managers who archive data at these distributed centers will provide contact information (name, e-mail address, phone number, and an ftp link to the data archive at the instrument manager's home institution) on the Data Archive and must adhere to all policies and guidelines described in this Data-Management Plan.

NCAR EOL Field Project Services will perform any necessary processing for the operational datasets only (e.g., satellite, upper air soundings, surface observations, and model output). Otherwise, individual PIs will be responsible for the final processing, quality control, and submission of their own datasets to the Data Archive since they are best qualified to do so. The time required to prepare "final" observations suitable for analysis depends greatly on the observation type. For example, the quality control of EOL Mobile GPS Advanced Upper-Air System (MGAUS) observations is expected to take 3-6 months after the field phase. Radar datasets particularly require significant preprocessing before a quantitative analysis can be produced, and it is not feasible for instrument managers to prepare quality-controlled radar observations for other PIs. Thus, the standard for exchanging "final" radar data will be raw dorade sweep files, plus sufficient metadata (radar-deployment information, etc.) so that a quantitative analysis such as dual-Doppler analysis could eventually be produced after data-quality control.

All PIs participating in VORTEX2 must promptly (within 16 months after data are collected) submit their "final" data to the VORTEX2 Data Archive. (For most datasets, "final" refers to quality-controlled data; for radar datasets, "final" refers to raw dorade sweep files for the VORTEX2 intensive observing periods.) The requirement for PIs to submit their final data to the Data Archive will facilitate intercomparison of results and an integrated interpretation of the combined VORTEX2 dataset. The PIs will greatly benefit from further collaborative analysis of their datasets within the VORTEX2 community. The scientific community will be interested in VORTEX2 datasets for at least 10 years, probably more, after the field phase ends, so having the data archived in a central location will make it possible to obtain and analyze data for the lifetime of the dataset. Complete metadata (including dataset descriptions, documentation, calibrations, quality assurance results, etc.) must accompany the submitted data. Upon submission, unless otherwise specified by the PI, these data will be available to the general scientific community. The PI does reserve the right to request that the Data Archive provide password protection for these data and/or send notifications when a request for these data is received during the initial 16-month data-analysis period.

## 7.5 Responsibilities

The VORTEX2 Steering Committee and NCAR EOL Field Project Services are working together to prepare and implement the VORTEX2 Data-Management Plan. Contact information is provided in Appendix A.

Joint responsibilities of both the VORTEX2 Steering Committee and EOL Field Project Services are:

- develop the Data Management Plan
- handle disputes regarding VORTEX2 data.

The following are responsibilities of the VORTEX2 Steering Committee specifically:

- maintain a list of and contact information for VORTEX2 PIs
- coordinate the October/November 2009 and 2010 VORTEX2 Research Workshops.

The following are responsibilities of NCAR EOL Field Project Services specifically:

- survey the VORTEX2 PIs to determine what information will be provided in the Field Catalog and Data Archive
  - operational data and experimental NWP products
  - real-time status reports and other field documentation
  - field data
- develop and maintain the Field Catalog and Data Archive, including processing operational data
- provide in-field support
  - make products needed for real-time planning available through the Field Catalog
  - ensure that VORTEX2 PIs contribute the required operations summaries within a few days after each operations day
- make arrangements for data distribution (e.g., cost, if any, method of distribution, etc.) and coordinate data orders with the requestor
- provide current inventories at least every three months for two years following the field phase so that VORTEX2 data users are aware of new dataset availability.

Individual PIs are responsible for providing their "final" field data to the Data Archive, as described earlier in the Data Management Plan.

---

Created by: [admin](#) last modification: Wednesday 15 April 2009 [15:04:11] by [dowell](#)

The original document is available at <http://rasmsys.com/tiki-index.php?page=Chapters%204-7>