JOINT URBAN 2003 DATABASE/WEB DESIGN

Scott F, Halvorson*, Donald P. Storwold, Erik M. Vernon
U.S. Army West Desert Test Center, Dugway UT

1.   INTRODUCTION

The U.S. Department of Homeland Security (DHS) and Defense Threat Reduction Agency (DTRA) conducted the Joint Urban 2003 (JU2003) field meteorological and dispersion experiment in Oklahoma City (OKC), Oklahoma in July 2003. The objective of the experiment was to advance the knowledge about movement of contaminants in and around cities and into and within building interiors. The resulting data are currently being used to improve, refine, and verify computer models that simulate the atmospheric transport of contaminants in urban areas (Allwine 2004).

Over 150 researchers from federal agencies, academia, and the private sector deployed instrumentation during the JU2003 campaign. With the large number of test participants and the vast array of fielded instrumentation, data management is an important issue. At the request of the JU2003 science team, the West Desert Test Center (WDTC) Meteorology Division at Dugway Proving Ground (DPG) designed and constructed a relational database and website interface for all data collected during JU2003. This paper focuses on the database/website design, security, output, and quality control.

2.   COMPONENTS

The hardware platform for the JU2003 database/website is a dual-processor server connected to a 3 terabyte (TB) Redundant Array of Independent Disks (RAID). Two identical systems were built – one as a primary server and the other as a redundant system in the event of failure of the primary system.

The database/website design is based on the Linux, Apache, MySQL, PHP (LAMP) model, which is described in Figure 1. Red Hat Linux serves as the operating system, the website software is provided by Apache, and MySQL is the database management system (DBMS). PHP (PHP: Hypertext Preprocessor) provides an interactive interface between MySQL and the website. All of these software components have been optimized for security and performance. The advantage of this system is that it is available 24/7 and has built in redundancy, which increases the reliability of the system.

*Corresponding author address: Scott F. Halvorson, West Desert Test Center, U.S. Army Dugway Proving Ground, Dugway UT 84022. email: scott.halvorson@us.army.mil

3.   DATABASE

The purpose of using a relational database was to provide permanent data archival, accelerate data retrieval, and provide organization to all JU2003 datasets. The database also provides data format standardization (e.g. – UTC time coordinates), which reduces confusion and data processing by an end user. Experiment participants provided datasets with a wide variety of formats and time stamps for input into the JU2003 database. The database unifies these various formats and outputs a single format for any given instrument. Table 1 provides a sample list of the types of meteorological data archived in the database, which are available online through the website.

4.   WEBSITE

The JU2003 database website provides a graphical user interface (GUI) for retrieving data from the database. It was designed specifically for easy navigation and streamlined data retrieval. The home or main page contains 14 links that are grouped by specific types of data sets (e.g. – surface, aircraft, tracer, publications, etc.). Users also can send comments to the JU2003 developers and make special data requests through the website. Clicking on one of the links on the homepage takes the user further into subgroups of available datasets. Through the hierarchy of groups and subgroups, the user can find data organized by data type, time, or organization responsible for the data collection. Figure 2 shows an example of a group (Profile data in this case) subdivided down to the LIDAR instrumentation. To retrieve the data, the user selects a station(s) (if multiple stations are available), date, and time range, and then clicks on the adjacent retrieve button. The request is sent to a queue for processing. Once the data have been retrieved, an email message is sent to the user with a download link. To retrieve other data within the data subgroup, links are provided just below the subgroup heading. Users also can go back to the main page by clicking on the "main page" at the end of the subgroup links.

Prior to 31 January 2005, JU2003 data were only available to experiment participants. However, data from JU2003 now are available to the general public. Table 2 summarizes database usage statistics since the website has been available to non-participants.

Table 1. Example listing of available data on the JU2003 website.

	Surface	Vertical Profile	Other
.5km MM5 Model Runs			X
Documentation			X
FM-CW Radar		X	
Lidar			X
Mesonets	X		
NWS Data			X
Publications			X
Portable Weather Stations	X		
Radiometers		X	
Radiosonde		X	
Samplers	X		
Sodars		X	
Sonic Anemometers	X		
Tethersonde		X	
Tall Tower (multilevel)		X	
Wind Profilers		X	

Table 2. JU2003 database usage statistics since January 2005.

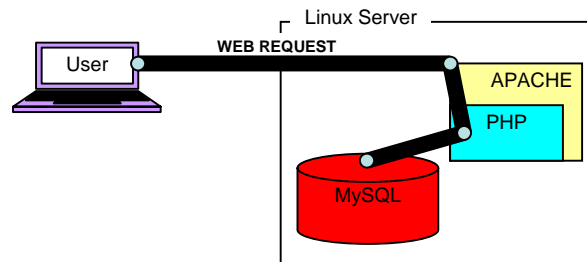| Web Page	
• User Accounts	125
• States	24
• Countries	6
Usage	
• Average Hits per day	56.7
• Average logins per day	3.1

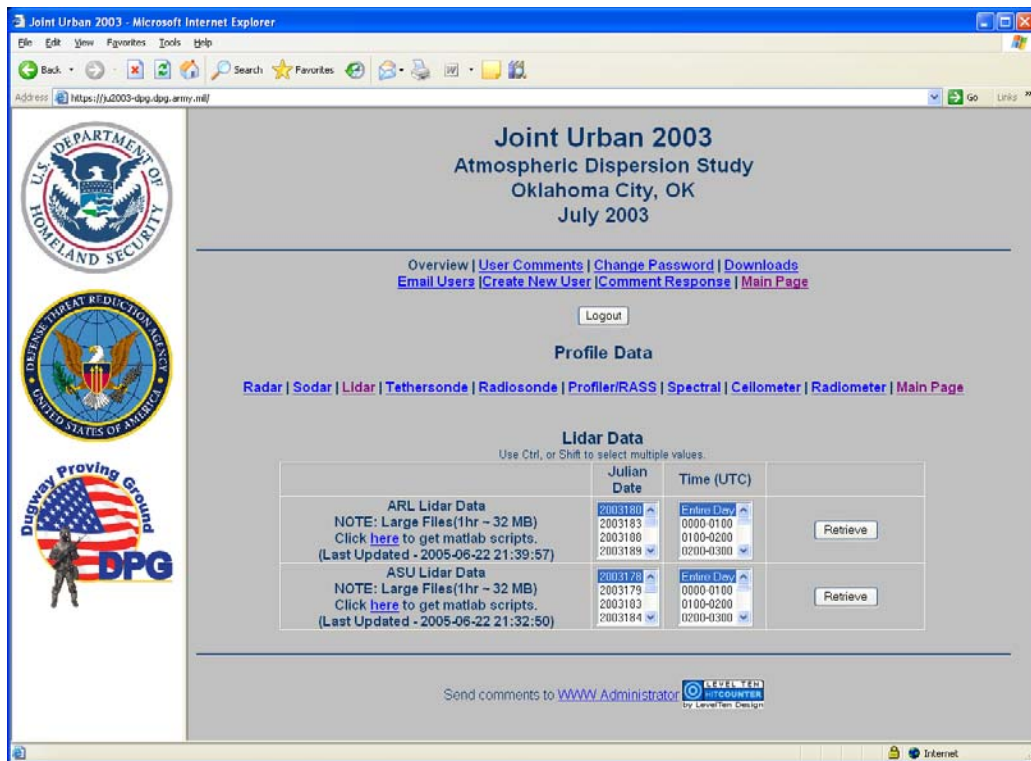
Figure 1. LAMP model diagram

Figure 2. An example of retrieving lidar data.

5. SECURITY

While the data in the JU2003 database are available to the general public, several security measures have been taken to prevent the loss of data and/or interruption of service. These security measures include:

- Servers reside behind a DPG firewall
- Transactions between client and web server are encrypted (https://)
- User accounts are required to access the website; account requests are reviewed by the JU2003 science team before approval
- Measures have been implemented to detect and prevent hacking attempts (e.g. – turning off unused functions in PHP and modifying MySQL user permissions)

6. OUTPUT

The JU2003 database provides two types of data packages, textual (ASCII) and binary (e.g. – Radar, Microsoft Word, Adobe PDF, etc.). Textual output is packaged in a compressed format and contains two or more files. A "Readme" file is included, if provided by the experiment participant for that particular dataset. The "ReadMe" file generally contains the metadata (e.g. – originator of the data, site location, instrumentation, point of contact, etc.). The data files also contain abbreviated metadata at the top of the data set that account for 60 lines. Column headers, which immediately follow the metadata, describe each column of data. Each row of data has a time stamp in Coordinated Universal Time (UTC) in the first column, followed by the data. This organization of textual output eliminates confusion by the recipient and also streamlines software applications that interpret the data.

Binary formatted data are used for any dataset that is not in human readable format, such as LIDAR, model, images, etc. Although these datasets also have an attached "Readme" file provided by the supplier of the data, the output of the data is not organized similarly to textual data. Common programs, such as Microsoft Word or Adobe Acrobat Reader, are used to view most of the binary files. In the case of special binary formats, the JU2003 website has links to algorithms that can display or analyze the unique formats.

7. QUALITY ASSURANCE

Quality assurance (QA) can be broken down into data quality control (QC) and database/website QA. Data QC was performed by the organization that collected the data. This information is always included in all retrieved datasets as a separate data column. Database/website QA is the process of verifying that the data available to the end user exactly match the original data. With over 2 billion records in the database, it is not feasible to examine each individual record for correctness. However, an independent QA contractor performs quarterly audits in which it randomly samples 5 percent of the data for comparison with the original data submitted to the database. The audit reports summarize any discrepancies discovered and suggest corrections to the JU2003 database/website staff.

Another feature of the JU2003 website is a section that allows users to provide comments on any discrepancies that they find in the data. When comments are input, the JU2003 database/website staff is notified via email, and corrective actions are taken to rectify any problems.

8. ACCOUNT REQUEST

To obtain a JU2003 database user account, please visit https://ju2003-dpg.dpg.army.mil. Near the bottom of the page, click on "FORM" and complete the required information. The user will receive email notification after the account is created. This process generally takes about 24 hours.

9. SUMMARY

The JU2003 database and website have proven to be powerful data retrieval tools. Feedback from end users has been overwhelmingly positive. The most frequent comment is on the convenience of having a single source for all JU2003 data. Datasets from the JU2003 database have a common time base (UTC), a standardized output, and are available 24/7. The system is reliable, secure, and relatively easy to use. An interagency Joint Action Group (JAG) established by Office of the Federal Coordinator for Meteorology (OFMC) recently recommended that future data warehouses for field experiments should follow the JU2003 database/website model.

10. POINTS OF CONTACT

Donald Storwold – Project Manager
   donald.storwold@us.army.mil
Scott Halvorson – Database Administrator
   scott.halvorson@us.army.mil
Erik Vernon – Webmaster
   erik.vernon@us.army.mil

11. REFERENCES

Allwine, K.J., Leach, M.J., Stockham, L.W., Shinn, J.S., Hosker R.P., J.F., Bowers, J.F., Pace, J.C. (2004, January). Overview of Joint Urban 2003 An Atmospheric dispersion study. Paper at the Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone, Seattle, WA. J7.1.

Apache-SSL. Available from http://www.apache-ssl.org/

MySQL AB Database Server. Available from
http://www.mysql.com

PHP. Available from http://www.php.net/

Storwold, D.P., (03 November 2004). Explanation and
Quality Control of the Joint Urban 2003
Database/Website.   Meteorology   Division,
Dugway Proving Ground, Technical Note 04-90-
5.   Available via email from Storwold, D.P.,
(donald.storwold@us.army.mil).

Storwold, D.P., Halvorson, S.F., Vernon, E.N., and
White J. M. (2005, July).   Joint Urban 2003
Database/Website. Poster presented at the 9th
Annual Gorge Mason University Conference on
Atmospheric   Transportation   and   Dispersion
Modeling, Fairfax, VA.

The Apache Software Foundation. Available from
http://www.apache.org