Space Weather Community Operations Workshop (SpWxCOW) Lessons Learned

W. Kent Tobiska

Space Weather Community Operations Workshop

- **Purpose**:

  bring together industry, agency and university Space Weather operational personnel in a 2-day meeting to promote a robust data exchange capacity within the Space Weather enterprise. Participants share common operational problems and solutions as well as work toward developing common best practices, and eventually standards, for Space Weather operations.

- **Sponsors**:
  - ACSWA: American Commercial Space Weather Association
  - AIAA: American Institute of Aeronautics and Astronautics
  - AMS: American Meteorological Society
  - CCMC: NASA Community Coordinated Modeling Center
  - SIF: Space Infrastructure Foundation
  - SWPC: NOAA Space Weather Prediction Center
  - USGS: U.S. Geological Survey
  - SWC: Utah State University Space Weather Center

Overview

Workshop topics

1. 1) Operations Procedures and Problems
2. 2) Real-time Data Sharing
3. 3) Data Storage and Archiving
4. 4) Breakout sessions (Progress in Data Use and Metrics, Commercial Data Buys, Data Archiving Strategies)
5. 5) Guidelines
6. 6) make SpWx data more accessible

Workshop deliverables

- ✓ **SpWxCOW report** (distributed at SWW)
- ✓ ***Space Weather Journal* article** (spring 2013)
- ○ **AIAA Guidelines document** via AIAA ASETC CoS: lessons for reliability, maintainability, accessibility, dependability, quality, and safety in SpWx ops (1/2014)

Operations Procedures and Problems

- **"99% Rule"**
  - there is a 1% failure rate per product delivery
  - Out of a hundred data streams being ingested per day, there seems to be at least one that has a failure
  - *How can the impacts of those failure rates be minimized* and *how do we quickly communicate when those failures occur between institutions?*
- **Reducing failure impacts**
  - asynchronous, redundant input data streams
  - exception handler routines that produce deadman files and run-time logs
  - SQL databases to manage data
  - limited reliance on 3rd party applications and open-sourced (unsupported) libraries
  - continuous systems administrator support
  - In operational systems labor for development and maintenance are the greatest long-term costs

Operations Procedures and Problems

- **How to determine what the customer wants?**
  - imagine what life is like for other people
  - Through iterative discussions, understand what the customers are looking for
  - These are often specialized, quality products
- **Development strategies**
  - joint operations with other providers
  - agile development
  - early and continuous development and deployment of software
  - automated testing and result tracking
  - software delivery that is immediately useful
  - web-based data for user feedback

Real-time Data Sharing

- **Goal**: timely and accurate data transfer among organizations
- **Methods for keeping real-time data sharing secure**
  - Web feeds create a secure, pull-based data stream, e.g., Rich Site Summary (RSS) feeds or alternatives like Atom
  - no username or password and use standard XML format, supported in most web browsers
  - Users merely subscribe to the web feeds via web pages or applications and instantly have access to updated data as soon as it is available
- **Data product quality**
  - external validation is needed
  - Feedback from users can help developers eliminate model biases based on missing physics
  - attention to output data quality affects how data is initially created and how it will ultimately be distributed
  - As part of this distribution, errors in metadata must be considered as well as the importance of tracking data version numbers

Real-time Data Sharing

- **Test-bed method for data production**
  - iterations with the customer including testing, code management, identifying missing and bad data, as well as developing user guides and documentation
- **Verify the input data stream is critical to the quality of the output data produced**
- **Trade-offs between accuracy and latency**
  - Near real-time, low latency data may be less accurate than processed data, i.e., a consideration for users
  - When input data formats change abruptly, this can cause failure modes by user systems
  - Format changes by data producers may often be an oversight, but it should be a high priority of producers to inform their data consumers of changes
  - A data stream failure mode – evolving computer operating systems – 95% of data loss came from Windows OS updates in one example

Data Storage and Archiving

- **Input data to systems**
  - the availability, reliability, and quality must be assessed
  - Operational users must identify short-term vs. long-term storage requirements, including database design and management
  - Input data gaps must be managed along with anomalies in parsing data that is inserted into a database
  - Reliability can be improved using hardware redundancy and by ensuring control of data format changes for backwards compatibility with announcement of changes ahead of time
- **Data availability is a primary concern**
  - multiple data types are often needed
  - accessibility – if there is no access to the data, it might as well not exist
  - Latency and data transfer speed – most operational users prefer to get data sooner rather than later

Data Storage and Archiving

- **Quality of data affects storage and archiving**
  - secondary quality checks are vital
  - Will archiving methods currently in existence be accessible in 50-years
  - metadata information may be preserved through Quick Recognition (QR) codes
- **Data storage and archiving**
  - Often accomplished via DVD backup, HDD storage, and cloud storage
  - DVD backup is most commonly used; though it isn't the most easily accessible, it is likely to be the easiest method of storing data
  - Hard disk drives (HDD) are most easily accessible
  - Cloud storage is being used more but not for secure data
  - Cloud storage uses networked online storage, which is becoming a more affordable option, and it makes accessing data as easy as accessing data on HDDs

Breakout sessions

- **Progress in Data Use and Metrics**
  - identified real-time data sharing and comparison opportunities
  - recommended using existing databases to compare real-time data related to examples of ISS in-situ electron densities, ΔB, CME speed/magnitude/arrival and Dst
  - Having common data comparisons through a repository enables metrics to be developed for understanding the strengths and weaknesses of our current capabilities

Breakout sessions

- **Commercial Data Buys**
  - who is using and purchasing space weather data
  - specific and custom-tailored data products are purchased and used by a variety of organizations (DoD, NOAA, commercial)
  - university–industry–agency communication should continue and expand in an effort to maximize cost-savings for data production in a budget-constrained environment
  - key products in the hands of the right operational personnel can generate a pull for a data product
  - The Space Weather Workshop and Space Weather Enterprise Forum could be used to expand discussions on this topic
  - Participants agreed to explore commercial data purchases from qualified data providers in order to minimize duplication of scarce community resources, create a better inter-agency and commercial division of labor, and promote a robust operational space weather community

Breakout sessions

- **Data Archiving Strategies**
  - all data doesn't have to be immediately accessible as long as metadata is available
  - useful to insert metadata that describe the location of archival data into the raw data streams
  - If data are archived up-stream, there is no reason to store it down-stream in several places
  - archival and operational databases should be separate
  - important to identify input data constraints that are going to be inserted into a database
  - Using self-defining formats like NetCDF and XML can be beneficial, since many users broadly accept those community standards

Best Practices

- **Best practices are guidelines viewed as beneficial by the space weather community that can enhance its robustness and growth**
- **They make our data products reliable, maintainable, accessible, and dependable with quality and safety**
- **Operations Procedures and Problems**
  - Check any 3rd party software with known inputs and outputs to avoid weeks of troubleshooting someone else's code
  - Maintain and check logs for database errors related to constraint violations
  - Thoroughly test custom code
  - Use an Integrated Coding Environment (IDE) with built-in code syntax corrections
  - Use industry standards for software, coding and testing
  - Use consistent filename conventions
  - Test software upfront
  - Use version control of code to make changes recognizable
  - Develop a community forum to address the problems encountered
  - Minimize data format changes since they can break end-user's software

Best Practices

- **Real-Time Data Sharing**
  - Use quality flags for operational data
  - Develop standard naming conventions for space weather parameters
  - Use consistent community metadata tags and formats for all distributed data
  - Use standardized self-describing metadata (XML, NetCDF)
  - Develop a community space weather data dictionary
  - Identify within metadata raw, partially processed, or final data
  - Develop standards to insure product consistency, allow universal software processing, and provide minimal downtime for customers due to product failure
- **Data Storage and Archiving**
  - Avoid storage duplication if possible
  - Use SQL databases for rapid access
  - Use metadata to describe where actual data is located
  - Archive all raw data even if it is error prone
  - Use constraint testing to determine if data can go into an operational database
  - Separate raw from operational data
  - Identify repository locations in metadata

Best Practices

- **Facilitating SpWx Distribution**
  - Action: Identify relevant parameters for testing and common validations in (e.g.)
    - Total Electron Count (TEC)
    - Equatorial Eastward E-Field
    - Higher frequency spectral content of ground-level magnetic fields
  - Action: Improve data sharing by creating
    - Awareness of data availability
    - Easy accessibility
  - Action: Explore commercial data buys based upon
    - Fully coupled models that may generate data pulls
    - Key products in the hands of operational managers
  - Action: Promote improved government, academic and commercial communications using lessons learned from the terrestrial weather community
  - Action: Develop better interagency–commercial data production divisions of labor
  - Action: Develop real-time data metrics starting with a focused topic area (CMEs)
  - Action: Develop an operational space weather Wikipedia page
  - Action: Set up a Google Group forum to facilitate common problem resolution for distributing quality space weather data

Next Steps

Space Weather Accessibility

Move to new technologies: Google Earth, Twitter while maintaining legacy systems

- Clouds
- spaceweather.com
- RSS automated feed
- Facebook
- Twitter automated feeds
- Wikipedia entry

URL links, KMZ files, subscription lists, products/ services/kiosks

Next Steps

Space Weather Cloud Mall

- Making SpWx easily accessible to each other and outside users; Benefits for each group? (marketing, visibility)
  - ✓ Google Cloud – use for archiving data, marketing data, neutral 3rd party; archive and access possibility; dissemination
  - ✓ Contact spaceweather.com as entry point host (Tony Phillips); Commonly hosted webpage, with links; Linkages, KMZs, subscription pointers, product and service lists

Next Steps

Space Weather Cloud Mall

- ✓ Twitter automated feeds:
  - SEC (@spacenv) and SET (@spacenvironment) implemented Twitter feeds since last year
  - Examine hashonomy.com/hashtag/spaceweather
  - Hashtags in use #spaceweather, #spwx, @spacewx, #sunspot, #sun, #aurora, #cme, #solarflare
  - Recommended list: issuing organization, time, phenomenon, links, URL
  - Put alert threshold high otherwise people drop-off; consider how long since last event
  - Utility of feeds? Too much information? Links to URLs?

Some Space Weather Product Qualities

- Simple to understand – visually
- Accurate representation
- Sufficient information for decisions
- Easy access to information
- Low latency and appropriately rapid updates
- Unambiguous guide to action now and in the future