WOVOdat documents

3 Summary of Discussions at 2nd WOVOdat Workshop, Menlo Park, California, Dec 11-13 02

Summary of Discussions at 2nd WOVOdat Workshop, Menlo Park, California, Dec 11-13 2002

The second WOVOdat workshop had three principal aims:

- To review existing volcano database development projects, to determine whether any would suffice for the WOVOdat design or, if not, to explore possible sharing of database design elements.
- To engage additional observatory scientists who were unable to attend the first workshop in Indonesia.
- To revisit preliminary recommendations of the first workshop, for confirmation or revision.

Each of the formal presentations described an existing database or data management scheme. Steve McNutt described pioneering compilations of volcanic earthquake swarms and tremor. From Europe, Radu Gogu and Florian Schwandner described GEOWARN; Pierre Briole described data management of the Observatoires Volcanologiques, IPG Paris and also of the European Mobile Volcano Surveillance for Early Warning (EMEWS) project; Marcello Martini described data management at INGV's Osservatorio Vesuviano (see also, data management at INGV Catania); and Juergen Neuberg described the goals and data management of MULTIMO (Multidisciplinary Monitoring, Modelling, and Forecasting of Volcanic Hazards). Moving west across the Atlantic, Stephan Husen described data management at the Yellowstone Volcano Observatory; Dan and Peter Cervelli showed the VALVE (Volcano Analysis and Visualization Environment) system from the Hawaiian Volcano Observatory; Maurizio Battaglia showed previews of his (now published) CD with Long Valley data; Tom Murray described data management at the Alaskan Volcano Observatory, and Steve Malone and Steve Schilling described data management at the Cascades Volcano Observatory. Chuck Meertens described the UNAVCO archive for GPS data, including data from 20 volcanoes worldwide. Yoshiaki Ida briefly described volcano data management in Japan, which is handled at individual volcano observatories and at the Japan Meteorological Agency (JMA). JMA will be the primary link between Japanese volcano observatories and WOVOdat. Finally, Paul Kimberly described the Smithsonian's Global Volcanism Program and, specifically, their pioneering Volcano Reference File to which WOVOdat will be closely linked. Thanks very much to these presenters and to the rest of the 38 participants for their helpful comments during plenary and breakout discussions. A special thanks to those who traveled from overseas to share their work.

Following these presentations, we split into three Working Groups:

1. Parameters, Data quality, and Metadata
2. Schema -- relationships between data types
3. Data access (who, when, how)

The main recommendations from each Working Group are as follow:

I. Parameters, Data Quality, and Metadata:
    1. All data will be time-stamped and geo-referenced.
    2. In general, the appropriate level of data for WOVOdat is the most basic (i.e., least processed) data that are needed for comparative studies and pattern recognition. This level and specific parameter recommendations may be found in the draft list of WOVOdat parameters. Rates and changes of rates can be calculated rather than entered separately. Raw data can be sought directly from observatories.
    3. Similarly, the appropriate level of metadata for WOVOdat is the minimum needed by users for comparative studies. Additional details such as calibrated instrument responses can be obtained from observatories if needed.

4. Some "knowledge products" can also be included, especially those that would be awkward for users to recalculate, e.g., focal plane solutions. One section of WOVOdat will be explicitly interpretive, noting published interpretations of unrest in keywords for easy retrieval, without any implied warranty that the interpretations are correct. This will simply record “who thought what, and when.”
5. Data quality should be expressed whenever possible in terms of precision (g). In this way, value judgments about the usefulness of the data are minimized and comparability is maximized.
6. Data with low precision will be retained but flagged with the appropriate caution and a link to pertinent metadata. The reason for including data of less-than-current standards is that we want to capture lessons from early monitoring efforts, to get a sense of variability in patterns of unrest from one episode to the next or from one volcano to the next. There are many more episodes of unrest and eruption in a century than in just a few years, and for macro-scale changes sample size is more important than data precision.
7. The primary responsibility for data quality assessment will remain with the contributing observatory. Users can include data quality in their search parameters, as appropriate to the problem they are studying.

II.Database Schema

1. There was a strong consensus that, for recent and future data, we should "undefine unrest,” i.e., to include continuous data without specifying or segregating that judged to be from “unrest.” This avoids arbitrary or inconsistent judgments about what constitutes unrest and will also be more efficient for many observatories.
2. At the same time, we recognized that in many instances of older unrest, only short periods of data have been preserved. Also, even if continuous data from the pre-digital age were available, the time investment to enter those data manually for long periods of quiet would not be worth the effort.
3. WOVOdat will offer a copy of its database structure to every observatory in open source software. This may solve some of the data management needs of smaller observatories and will certainly facilitate local uploading of data for WOVOdat.
4. WOVOdat will use numeric primary keys in each table to facilitate linking between tables and linking with other databases.

III.Access Working Group:

1. Free public access to observatory data is a sensitive matter that we discussed at length. All participants favor rapid public release of interpreted data (warnings and updates), most favor public release of minimally processed (publishable) data after a suitable grace period, and few favor open public release of raw data. On the matter of raw data, most participants recommended that such data be shared on a person-to-person collegial basis, rather than through WOVOdat, and this is entirely consistent with the recommendation of the parameters working group that WOVOdat not store raw data except a few types that can be used directly for comparisons between volcanoes or unrest.
2. One primary reason for a grace period before release of minimally processed data is to minimize public and official confusion during volcanic crises. Volcanologists have an unenviable record of public disagreements that ought to have been worked out behind the closed doors of scientific discussions. Such disagreements confuse officials and the public and result in credibility loss for all scientists. The other primary reason for a grace period is to give those scientists who collected the data a fair first chance to interpret and publish the data. Participants were in consensus that the appropriate length of a grace period is 2 years.
3. If it is more efficient for an observatory to send data to WOVOdat soon after the data are gathered, that observatory may request WOVOdat to filter (not deliver) those data for up to 2 years. Each observatory, not WOVOdat, will choose an appropriate grace period or serving filter for its own data not to exceed 2 years. Some observatories already make

data available in near-real-time because the risk of data rip-off has proved less than the value of transparency and enhanced collaboration. We expect more observatories to move in this direction but accept that local factors must dictate their pace.

4. WOVOdat will not have different levels of access for contributors and outside users, nor will there be passwords for read-only, search, or download access. The only use of passwords will be for data maintenance.
5. Each observatory will tag data with guidance on how to cite the source, and the quality and idiosyncrasies of its data.
6. Researchers who will publish papers will be encouraged to contact original sources to avoid duplication and to build collaboration. We encourage inclusion of local co-authors in publications.

We discussed the pros and cons of centralized vs. distributed models for WOVOdat. The centralized model is easier and faster to query; distributed models might be easier and retain more control for observatories but they are harder to organize and query and require fast internet connections. We decided to proceed with a centralized system and then reconsider a distributed system in the future when data storage and formats become more standard and internet connections are faster. We also discussed potential linkages to other major geoscience databases, including those of the Smithsonian, IRIS, UNAVCO, Earthscope, NEPTUNE, and GEON. Linkages could include sharing of selected data and of utilities for data I/O and visualization and, potentially, co-hosting. We do not wish to reinvent any wheels, have had initial discussions with most of these groups, and are keen to talk further with potential "dance partners!" The project has been several years in design and gestation and is moving towards the pilot build phase and testing. Then, the large tasks of populating and maintaining the database will begin. The job ahead is large and could make difficult demands on already-busy observatory staff. To minimize these demands, we plan that the WOVOdat project will have staff that, upon request, can travel to observatories and play an active role in data reformatting and transfer to WOVOdat. We also plan to offer timesaving data management solutions that will serve individual observatories' own data management needs as well as serve the large expected WOVOdat user community.

Workshop Organizers:
Dina Venezky, Menlo Park
Chris Newhall, Seattle

Participants in 2nd WOVOdat Workshop, Menlo Park, December 11-13, 2002

Maurizio Battaglia (LVO), USA/Italy
Pierre Briole, IPG Paris, France
Peter Cervelli, HVO, USA
Dan Cervelli, HVO/UW, USA
Bob Christiansen, YVO, USA
Radu Gogu, Geowarn, Zurich, Switzerland
John Ewert, USGS/VDAP, USA
Stefano Gresta, Univ. of Catania, Italy
Marianne Guffanti, USGS/Reston USA
Roz Helz, USGS Reston – 1st day only USA
Dave Hill, LVO, USA
Stephan Husen, YVO, USA
Yoshiaki Ida, Himeji Tech, Japan
Malcolm Johnston, USGS Menlo USA
Paul Kimberly, Smithsonian, USA
Jake Lowenstern, YVO, USA
Steve Malone, UW, USA
Marcello Martini, INGV, Italy
Warner Marzocchi, INGV, Italy
Steve McNutt, AVO, USA
Chuck Meertens, UNAVCO, USA
Tom Murray, AVO, USA
Juergen Neuberg, MULTIMO – 1st day only UK
Chris Newhall, CVO @ UW, USA
John Pallister, USGS/VHP, USA
Jim Quick, USA
Laura Sandri, Univ Bologna, Italy
Steve Schilling, CVO, USA
Florian Schwandner, Geowarn, Switzerland
Willie Scott, CVO, USA
Paul Segall, Stanford, USA
Tom Simkin, Smithsonian (1st day only) USA
Don Swanson, USGS/HVO, USA
Glenn Thompson, MVO (1st day only) UK
Dina Venezky, WOVOdat/USGS, USA
Randy White, USGS/Menlo Park USA
Meredith Williams, Stanford, USA
Jeff Wynn, USGS/VHP, USA