WOVOdat documents

2 Summary of 1st WOVOdat workshop, IAVCEI General Assembly, Bali, July 23-24 2000

Summary of 1st WOVOdat workshop, IAVCEI General Assembly, Bali, July 23-24 2000

The state-of-the-art in eruption forecasting improves each year. But, at the same time, public expectations rise that volcanologists can give precise warnings, neither too early nor too late, and neither overstated nor understated. Much of the societal relevance and funding of volcanology depend on accurate and precise forecasts, so we must ask ourselves, "In addition to current work, what else might we do to improve the accuracy and precision of forecasts?"

Historically, eruption forecasting has relied on repetition of patterns in seismicity, ground deformation, and gas emission before eruptions at the particular volcano of concern. Pattern recognition works well at volcanoes that erupt frequently and with roughly the same precursors to each eruption; it fails where eruptions are infrequent, where the character of unrest changes, or where the same unrest can lead to eruption or not. During the early stages of unrest, patterns are not yet well developed, the outcome of the unrest is not determined, and small differences in rate of magma supply, rate of degassing, magma viscosity, and other parameters can spell the difference between eruption or not, or between different timing, style, or magnitude of eruption.

Better forecasts require that we understand these processes, conditions, and controls, i.e., how the volcano works. All volcanologists know what their own specialized data imply about subvolcanic conditions and processes. Observatory teams consider all lines of current data (e.g., seismic, geodetic, and gas data together), and the constraints that each line of data imposes on interpretations of the next. Process-oriented working hypotheses guide collection of new monitoring data, helping to focus limited equipment and time on the most promising parameters.

In-depth studies of the past and present of a restless volcano become the main basis for forecasts of its future. This is as it should be. Yet, the focus inward on the specific volcano risks loss of some valuable context and lessons that can be gleaned only from study of a much larger sample of volcanoes and unrest. The analogy in medicine is that doctors must rely in the first instance on a patient's own history and symptoms to diagnose an illness, but their job is made easier by the work of epidemiologists who have studied the occurrence of these same symptoms in large populations. The larger data set may show some variability in the symptoms of a single illness, or several illnesses that have one or more symptoms in common. Also, correlations between the illness, its symptoms, and background information such as genetic or environmental factors may help greatly in efforts to understand the disease process and controls on that process.

Epidemiologists pioneered the use of large digital data sets to find sometimes subtle relationships between parameters in complex systems. Now, large databases are central driving forces in a number of sciences, as well as in business. Within the geosciences, seismology and meteorology led the way but nearly every geoscience is following suit.

This note reports on a recent workshop that begins development, by the World Organization of Volcano Observatories, of a modern database of worldwide volcanic unrest.

Current state of the data.

Over the past century and especially over recent decades, volcanologists collected a tantalizing volume of data on potential and confirmed eruption precursors. Most current data are recorded and stored digitally, but in various formats. Older data are in hard copy tables, graphics, and pen traces. All volcano observatories prepare time-series and, as appropriate, spatial distributions of seismic, geodetic, gas, and other data. But few can plot all parameters together vs. time, or together in common x-y-z geographic coordinates. Even fewer observatories have linked their data sets for broader access and comparison. As a result, it is currently impossible for researchers and for those responding to volcanic crises to search for all close matches to some features of unrest in question, or to use modern data mining techniques to find patterns that we might previously have missed.

Previous compilations of data on volcanic unrest.

The Smithsonian Institution's Global Volcanism Program has for years built and maintained an up-to-date, reliable, searchable database of historical eruptions (Simkin and Siebert, 1994; http://www.volcano.si.edu/gvp). It contains almost no data about eruption precursors, or false alarms, but Smithsonian scientists are keen to see those data added, or linked, to their data on eruptions.

We are aware of only two global compilations of non-eruptive unrest. One, in book rather than digital form, reviews historical unrest at large calderas of the world up to the mid 1980's (Newhall and Dzurisin, 1988). All available sources were used, resulting in some inhomogeneity of the data. A small searchable "index" database shows the occurrence or nonoccurrence of various types of unrest during each episode of unrest. More importantly, this compilation shows graphically the spatial and temporal patterns of unrest.

The other global compilation, in hard copy and digital form, reviews seismic swarms during the period 1979-89 (Benoit and McNutt (1996). This Global Volcanic Earthquake Swarm Database (GVESD), culled from the Bulletin of Volcanic Eruptions (Volcanological Society of Japan), contains summary information about seismic swarms such as starting date, duration, maximum magnitude, and type(s) of earthquakes, and remarks. It does not include a catalog of individual seismic events. GVESD data are available on line at http://www.avo.alaska.edu/dbases/swarmcat/v_table.html. A search utility is pending.

The WOVOdat concept.

To enable searches, comparisons, and other analyses of unrest at many volcanoes, we need detailed catalogs of seismic events, time-stamped vectors of ground deformation, time-stamped fluxes of various gases, and similarly detailed data on other changes. All data must be translated into standardized units and formats and brought together in a unified, digital database. In recognition of the enormous intellectual and practical potential value of such a database, the World Organization of Volcano Observatories (WOVO) has decided to create WOVOdat, a fully searchable database of all aspects of volcanic unrest, accessible via the Web. So that the outcomes of unrest are also detailed, WOVOdat will be linked with the Smithsonian's database of historical eruptions.

Only a few years ago, WOVOdat might have been hopelessly ambitious and might have stretched the resources and the will of contributing observatories. Today, neither hardware nor software is a limiting factor, and the value of data sharing has been amply demonstrated by projects in the seismological, genomics, and many other science communities, and embraced by funding agencies such as NSF, NASA, and NIH. A recent review of the US Geological Survey's Volcano Hazards Program (Committee on the Review …, 2000) strongly recommended that the USGS bring its volcano data into a more accessible database. (That review also recommended that data be made available in real-time, which WOVOdat does not seek). Some wills may still be stretched and some culture change may be implied, but most WOVO observatories have responded enthusiastically. Now, the principal requirement is that each observatory and a WOVO working group devote time, still a small fraction of that spent gathering new data, to bring data from their observatories into WOVOdat.

First WOVOdat planning workshop.

40 participants from 18 countries and 17 WOVO observatories, plus the Smithsonian Institution's Global Volcanism Program, met after the 2000 IAVCEI General Assembly in Bali, July 23-24. A list of participants is included at the end of this note. Goals of the workshop were

- to review the rationale and uses for a larger, unified WOVOdat
- to review existing, related, or smaller databases of volcanic unrest
- to decide parameters and general structure of WOVOdat
- to choose formats for data entry so that contributors can begin the large job of data compilation
- to map out further steps, timetable, and responsibilities for the project.

The agenda followed these goals. Ade Djumarma, Director of the Volcanological Survey of Indonesia, welcomed participants and offered his perspectives on how a database of volcanic unrest could help

VSI in its tasks. The author outlined his vision of a Web-based WOVOdat and how it can be used both for volcanic crises and research on volcanic processes. Lee Siebert and colleagues of the Smithsonian Institution described their database of historical eruptions; Steve McNutt described John Benoit's and his development of a database of 10 years of volcanic earthquake swarms; Warner Marzocchi of Observatorio Vesuviano gave a forward-looking view of how advances in pattern recognition might be well suited to the diverse content of WOVOdat. Then, most countries offered brief summaries of their monitoring and particularly of their internal databases, most of which are not linked across disciplines or between observatories. We closed Day 1 with a review of the proposed content of WOVOdat (parameters, levels of detail).

On the morning of Day 2, participants broke into four working groups: Geological background, Volcanic Seismicity, Ground Deformation, and Volcanic Gases and Thermal Waters. Each working group considered parameters for their own specialization, and, especially, how best to represent data quality for each parameter.

During the final afternoon, ignoring (well, postponing) the siren call of Bali beaches, we discussed strategies for data entry and retrieval, lessons learned by IRIS (Incorporated Research Institutions for Seismology), and potentially sensitive issues of proprietary vs. public data. We closed by discussing next steps, including a pilot project described below.

Principal issues and decisions

Participants were cautiously enthusiastic, realizing that the job ahead is large and that some matters had to be defined more clearly before they could begin work on WOVOdat. Issues that commanded special attention were:

What is the right balance between enough detail that the database will be useful yet not so much detail that the task will discourage contributors?

A draft list of parameters (available to readers upon request) had been circulated prior to the meeting. For each broad type of unrest (seismicity, ground deformation, other geophysical changes, gas emission, thermal change, and changes in concurrent eruptions), the draft recommended catalog-style data, i.e., time, location, spectral character or composition if applicable, and sense and magnitude of change. Data on individual events, e.g., individual earthquakes, periods of deformation, or periods of gas emission, could be combined in time series, spatial distributions, and other graphical presentations. Raw data such as earthquake arrival times or amplitudes at individual seismic stations, station corrections, or unreduced geodetic and gas data, would not be included. When raw data are needed, they can be obtained from the sources for the reduced, catalog data.

Specialist groups considered parameters that were recommended in the draft and, in general, did not suggest changes. This was undoubtedly influenced by the short time available for discussion, so some changes may yet occur.

Should data be entered only for episodes of unrest, or also for baselines during quiet ("boring") periods?

Participants recognized the value of baselines, but also the large extra effort to enter data from much longer periods of quiet than of unrest. As a compromise, a pilot project described below will include one episode of unrest and one period of representative "quiet." Then, future data entry might focus on episodes of unrest.

How best to show uncertainties and data quality?

The various working groups approached this problem in several different ways. The volcanic seismicity working group noted existing methods for portraying uncertainty in earthquake locations, e.g., error ellipses. Discussion then moved on to errors that are not so easily quantified, such as the effects of wind noise on precision of picks of arrival times, and quantification of error in composite measures of seismicity like Realtime Seismic Amplitude Measurement (RSAM). The group also strongly recommended more standardization of volcano seismometers, including placement of at least one broadband instrument per volcano at a distance of about 3 km from the vent. Such standardization of instruments and instrument placement will make comparisons of data from different volcanoes simpler.

The volcanic geodesy working group offered definitions of A, B, and C quality data for tilt, EDM, theodolite, and GPS data. These suggestions had the virtue of simplicity, but may in some instances be too simple. Also, suggestions are still needed on estimation and representation of error for INSAR interferograms. We will seek additional input, especially with the UNAVCO Volcanic Geodesy group and other volcano geodesists who could not attend the Bali meeting.

The working group for fumarolic and thermal water data emphasized documentation of the collection and analytical methods. Users could judge the quality of samples on the basis of how carefully those samples were chosen and what safeguards were taken to prevent contamination or other artifacts. Details of filtration and dilution would be included in the latter. Analytical error can be estimated from repeat analysis of standards and samples of the current batch, or from previous estimation of the precision of that instrument and method. Estimation of error in plume measurements must also include error of windspeed and other meteorological factors.

What incentives can be offered for observatories to contribute actively to WOVOdat?

The single biggest incentive to contribute will be if observatories see that their own capabilities will advance faster if data are translated into standard formats and merged with similar data from other volcanoes. Certainly, seismologists, geneticists, particle physicists, and scientists from many other fields are making major advances as their cultures shift to wide data sharing. A second incentive will be that when some observatories lead by example, others follow. A third incentive is that a WOVOdat consortium can develop tools for data entry and archiving to make the job of each individual observatory easier.

Funding and access can be used as incentives, too. Some scientific consortia and major funding sources (e.g., NSF, NASA) require data sharing as a condition for funding, or for access to the larger pool of data or to a pool of instruments. USGS volcano observatories are under increasing pressure to make their data publicly available. However, because volcano observatories obtain their funding from many funding sources, and are mostly self-sufficient in terms of equipment and data, restriction of funding and access may be less effective for WOVOdat than the incentive of a better scientific product. If money and access are to be used as incentives, we judge that they will be more effective as "carrots" than as "sticks."

Fortunately, most countries in the workshop were keen to contribute and seek little or no added incentive. The greatest reluctance came, perhaps understandably, from the countries with the greatest volumes of data and thus the largest task ahead, but even these countries are willing in principle to participate.

Should there be a grace period during which observatories work on their own data before contributing them to WOVOdat, and/or a period of embargo during which data are contributed but not publicly available. If yes, how long should those periods be?

Workshop participants supported a roughly 2 year "grace period" from the time of unrest (or from the end of an extended period of unrest) to the time when data should be contributed to WOVOdat. This 2 year period would in most instances allow time for the observatory scientists to analyze their data and publish. It would also remove one worry of some observatories, that early or realtime release of data would complicate their job in managing volcanic emergencies. WOVOdat is fundamentally an historical database that will retain most of its utility even if a 2 year grace period is accepted. When important new patterns of unrest arise, we hope that the observatory scientists will make those data available early, at least for collaboration with other WOVOdat contributors.

What reasonably small-scale pilot project or trial could be used to test data entry and data retrieval mechanisms, and to "prove" the concept?

Workshop leaders agreed to select parameter lists, conventions, and formats that require the least possible reformatting from existing data, yet which satisfy the needs of the new database. A data entry shell will then be prepared, using some user-friendly public-domain software such as MYSQL that IRIS uses in its Portable Data Collection Toolkit. **We invite interest from those who might help us develop a similar data collection toolkit for WOVOdat.**

The suggestion was made and accepted that, as soon as the data collection toolkit is ready, each contributing country would format and contribute data from at least one episode of unrest and at least one (month-long?) period of quiet at at least one of their volcanoes. The timetable for this will depend on how quickly the data collection toolkit can be developed; our target is to have the toolkit within 6 months to one year, and to have contributions to the trial be ready by the time of the May 2002 centennial of the Mt. Pelée eruption.

Once constructed, how can the database be maintained?

WOVOdat should be a dynamic database that grows with the rapidly increasing volume and quality of volcano monitoring data. A small permanent staff will be required, supplemented perhaps by visitors from the various WOVO observatories. Given the early stage of WOVOdat discussions and the fact that internet connections now make physical location of a staff and server less important than they once were, we did not attempt to decide exactly who will maintain the database or where it will reside. One attractive possibility is with the Smithsonian's Global Volcanism Program, given its existing database of historical eruptions and existing network of correspondents about current volcanic activity. However, that program would need added resources before it would consider taking on this new responsibility. Close links with IRIS are another possibility. The funding sources will depend in part on the host country and institution.

Next steps

First, we invite comment and participation from the broader volcanological and geoscientific communities. Many of you have experience that was not represented in our workshop in Bali. Please send comments and expressions of interest to the WOVOdat steering group (wovodat@geophys.washington.edu).

A small group will develop a data collection toolkit as described above, to standardize parameters, conventions and terminology, formats, and units. As much as possible, we will use existing formats or neutral formats into which common formats can be easily translated. Most of this effort will start in Seattle (with USGS, IRIS, Univ. of Washington, and other interested scientists), with remote participation by Smithsonian and other interested scientists.

Then, we will ask for volunteers from each volcanic country (from one or more WOVO observatories per country) to contribute data from at least one episode each of unrest and quiet. Those who wish to volunteer should please contact the scientist(s) who represented your country in our Bali workshop, or contact the WOVOdat steering committee.

Eventually, we will seek a much broader pool of volunteers, too, to contribute data from all potential sources and episodes of unrest. Recent, digital data will need some reformatting and documentation;

older data that exist only in hard copy will need digitizing (where numeric) and translations (where anecdotal and not presently in English).

How soon can WOVOdat grow large enough that it will be useful during volcanic crises and for research? Much will depend on commitments of funding and time from contributing countries and from other interested parties. We aim to open it to contributors by 2008 (or earlier, for reference during crises), and to open it to all scientists by 2010.

References cited:

Benoit J and McNutt S, 1996, The Global Volcanic Earthquake Swarm Database 1979-1989. USGS Open-File Report 96-69 and http://www.avo.alaska.edu/dbases/swarmcat/GVESD.HTML

Committee on the Review of the USGS Volcano Hazards Program (J. Fink, chair), 2000, Review of the U.S. Geological Survey's Volcano Hazards Program. Natl Academy Press, Wash. DC, 138 p.

Newhall CG, Dzurisin D (1988) Historical unrest at large calderas of the world. US Geol Surv Bull 1855, 1108 p

Simkin T, Siebert L (1994) Volcanoes of the World (2nd ed) Geoscience Press, Tucson 349 p.

Participants: Name, Institution, Country

Ade Djumarma, VSI, Indonesia
Rudi Hadisantono, VSI, Indonesia
Mas Atje Purbawinata, VSI, Indonesia
Yoshiaki Ida, University of Tokyo, Japan
Jean Louis Cheminee, IPG, France
Francois Beauducel, IPG, France
Bradley Scott, IGNS, New Zealand
Wilfried Strauch, INETER, Nicaragua
Ima Itikarai, Rabaul Volcano Obs., Papua New Guinea
Mark Stasiuk, Geological Survey of Canada, Canada
Ernesto Corpuz, PHIVOLCS, Philippines
T. Frank Yang, National Taiwan University, Taiwan
Bernardo Pulgarin, INGEOMINAS, Colombia
Tatiana Pinegina, IVGG, Russia
Juan Jose Ramirez, Univ. of Colima, México
Vyacheslav Zobin, Univ. of Colima, México
Gabriel Reyes, Univ of Colima, México
Carlos Navarro, Univ. de Colima, México
Mauricio Bretón, Univ. de Colima, Mexico
Juan Carlos Gavilanes, Univ. de Colima, México
Eduardo Malavassi, OVSICORI, Costa Rica
María Martínez, OVSICORI, Costa Rica
Luis Lara, SERNAGEOMIN, Chile
Hugo Moreno, SERNAGEOMIN, Chile
Marino Martini (day 1), Univ. of Florence, Italy (Gas)
Warner Marzocchi, Osservatorio Vesuviano, Italy
John Shepherd, Seismic Research Unit, Trinidad
David Rothery (day 2), Open University, UK
Tom Simkin, Smithsonian Institution, USA
Lee Siebert, Smithsonian Institution, USA
Ed Venzke, Smithsonian Institution, USA
Paul Kimberly, Smithsonian Institution, USA
Jim Luhr, Smithsonian Institution, USA
Joop Varekamp, Wesleyan University, USA (Lakes)
Tim Ahern, IRIS, USA
Steve McNutt, Univ of Alaska, USA (+IAVCEI)
Elisa Koenig, Arizona State Univ., USA
Heidi Guetschow, Univ. of Washington, USA
Roz Helz, USGS, USA
Chris Newhall, USGS, USA