WOVOdat documents

1 Development of a database of worldwide volcanic unrest (WOVOdat)

Development of a database of worldwide volcanic unrest (WOVOdat)

PROJECT SUMMARY

The World Organization of Volcano Observatories (WOVO) seeks to build WOVOdat, a modern database of worldwide volcanic unrest.

The term, *volcanic unrest* mean significant changes (usually but not always increases) in seismicity, ground deformation, gas composition, fumarolic activity, or other parameters, within or adjacent to a volcanic system between and in the early stages of eruptions. Data from observatories around the world will be brought for the first time into common formats and Web accessibility. Data will be geospatially and time-referenced, to provide 4-D pictures of how volcanic systems respond to magma intrusion, regional strain and other disturbances, and as they prepare to erupt.

The WOVOdat database will be the central resource of a system to link to other databases such as that of IRIS (global waveforms), UNAVCO (global GPS coordinates and strain vectors) and the Smithsonian's Global Volcanism Program (historical eruptions). WOVOdat will serve volcanologists and other earth scientists as epidemiological databases serve the medical community. It will support searches for subtle, previously unrecognized patterns of unrest, research into the processes of volcanic unrest and improvements in eruption forecasts.

A summary of initial planning (Phase-I) may be found at http://www.wovo.org/WOVOdat/plansum.htm. Work in Phase-II, design, and Phase-III, pilot testing, will follow promising new directions in defining and categorizing volcano data (tentatively using XML and a modern relational database engine), for integrating data from multiple sources, each of which has its own instrument characteristics, data formats, and local data reduction methods. Sample data input methods and sample output views using data from several cooperating observatories will be used in the design and testing. Phase-IV, populating the database, and Phase-V, maintenance and enhancements, will take place in separate follow-on project handled primarily by WOVO observatories.

PROJECT DESCRIPTION

<u>Introduction</u>

Volcanoes exhibit a complex suite of geophysical, geochemical, geologic, and hydrologic changes (*unrest*) as magma ascends and prepares to erupt. For example, small earthquakes will mark where magma pressures are fracturing the rock through which magma will rise. Slight swelling of the ground surface will also reflect increased magma pressure. Gases that were soluble under higher pressures while magma was at depth exsolve and escape through fumaroles as magma rises and confining pressures decrease. Additional unrest may occur even without magma ascent, reflecting either in-situ changes of the magma or interaction between magma, hydrothermal systems and regional tectonic stress. Although volcanic unrest can at times bewilder and lead to false alarms, it also offers a window into the volcanic subsurface and the potential to forecast eruptions.

Increasingly, as populations on volcanic slopes expand and observatory experience and technologies improve, the public expects that both the accuracy and precision of eruption forecasts will rise. Volcanologists must now monitor the wide range of parameters and changes noted above, and project and interpret trends in the data. Eruption forecasts -- both empirical and process-oriented -- are based on comparing current unrest with that of prior unrest and its outcomes.

A large and valuable body of data has been acquired over the past century. Potentially this could be studied in the same way that epidemiologists study the occurrence, symptoms, and origins of disease. A whole new field of *volcano epidemiology* awaits, and we anticipate that it will significantly improve eruption forecasts as well as address a variety of research questions.

The data of volcanic unrest are currently so widely scattered that they are largely inaccessible except through the published literature -- usually only summaries - and direct queries to those who collected the data. Data files exist only in individual observatories and with individual researchers. There is no centralized, standardized database.

This fragmented state of affairs utterly fails to take advantage of the intellectual power of worldwide observatory experience and of galloping information technologies. There is no way at present to search through that collective experience for close matches or for subtle but instructive patterns or to test hypotheses about unrest at a large number of volcanoes.

The value of historical experience on volcanoes:

Many languages and cultures of the world emphasize the value of remembering history. In geodynamics the past and present are the key to the future. At volcanoes that have not erupted for many years we must extend our view of the past to other, similar volcanoes around the world. For example, at Mount St. Helens in 1980 volcanologists readily saw that the volcano was restless, that the north flank of the volcano was unstable and that a large landslide was possible. What was not so clear was that the hydrothermal system was probably pressurized and that, as a result, a complex interplay between hydrothermal pressures, steam explosions, over-steepening, and gravity could lead to a giant landslide and blast at any moment without the usual acceleration of creep and microseismicity that occurs before many normal landslides. Even in retrospect there was not enough monitoring before similar events at Bandai-san in 1888 and Bezymianny in 1956 to have helped much at Mount St. Helens.

But now volcanologists know exactly what to watch for. We now identify collapse-prone volcanoes from deposits made distinctive by Mount St. Helens. We no longer expect immediate precursors to collapse, and we know the roles that over-steepening and pressurized aquifers can play. Scientists working on the crisis of Montserrat volcano (including one with Mount St. Helens experience, Barry Voight) saw in 1996-97 the warning signs of potential collapse of the south wall of English Crater. Collapse on December 30, 1997, triggered a phreatomagmatic blast just as at Mount St. Helens. The smaller but still devastating combination of avalanche and blast swept through homes on the south slopes of the island. Fortunately, all residents had been evacuated and absolutely forbidden to return.

We propose to remedy this anachronism by building a modern database of worldwide volcanic unrest. Because most data on such changes would come from members of the World Organization of Volcano Observatories (WOVO), and many of the users would be from the same observatories, we propose that this new database of volcanic unrest be called WOVOdat. However, WOVOdat will have a Web interface and be open to all. We anticipate wide usage for academic research and teaching as well.

Uses of WOVOdat

The principal uses of WOVOdat would be:

- Research on magmatic ascent, vesiculation, degassing and other processes that control volcanism as manifested in changes that are monitored by volcano observatories.
- Reference by volcanologists who must forecast the outcome of unrest and need to know of similar unrest elsewhere, its causes and outcomes.
- Research on regional strain events as seen at sensitive, metastable volcanic systems. A number of recent observations, e.g., following the 1992 Landers earthquake (Hill et. al., 1993), point to greater sensitivity and participation of volcanic systems than previously believed.
- Reference by risk analysis professionals and the insurance industry.
- Data source for K-12 and university students. Volcanoes capture student interest and are wonderful vehicles for teaching basic science concepts and lessons about inquiry in complex systems. Learning about the volcanoes is a bonus.
- Reference for the general public. Well beyond the student years there is strong public interest in volcanism as a vivid reminder of the forces of nature, and the Web surely encourages this. Some also want to know more about volcanoes in their backyard.
- A resource for researchers from outside the geosciences, e.g. those improving techniques for data mining or studying the general phenomenon of self-organized criticality.

An added benefit, though not a use *per se*, is that adoption of common data formats and data entry protocols will facilitate exchange/interchangeability of monitoring instruments, data, and data processing tools.

Scope of WOVOdat

At least five general types of unrest will be included in WOVOdat:

1. Unrest that leads directly to an eruption and is thus a clear precursor to that eruption.
2. Unrest that does not lead immediately to an eruption but reflects one of a series of events over an extended time period (e.g., repeated intrusions) that collectively lead to an eruption.
3. Unrest that occurs between phases of an extended eruption.
4. Early phases of an eruption as indicators of what will follow.
5. Unrest that cannot be related to any eruption, e.g. regional earthquake activity occurring near a volcano or thermal changes that result only from development and fracturing of a siliceous self-seal in a hydrothermal system.

A user who is concerned with why a volcano reawakened this year or decade rather than the last one might consider all unrest since the previous eruption. A user who is concerned with predicting the immediate onset of an eruption might consider only the unrest that leads over a matter of days or weeks to eruptions. A user who is interested in subtle and not-so-subtle precursors of second or later phases of an eruption might consider only unrest since the previous phase.

In relating unrest to an individual eruption the user will be able to look at as wide or as narrow a window as data permit. Certainly complications will arise in generalizations because some eruptions have multiple phases. Phases that are closely spaced are often combined as a single phase in *Volcanoes of the World* (Simkin and Siebert 1994) unless details about characteristics and dates of each phase were available. Phases that were separated by more than 3 months are generally shown as separate eruptions in *Volcanoes of the World*, even though they might be from the same batch of magma. WOVOdat will provide time stamps on everything and the flexibility to look for relations between any set of events.

Information about the frequency or type or vigor of early phases of an eruption provides important information about what is to follow. Much of the information about eruptions themselves is already contained in a Smithsonian database (Simkin and Siebert, 1994). Additions that WOVOdat will include are quantification of eruption parameters that are important for forecasting. For example, maximum daily extrusion rate would be included because fast extrusion rates leave little time for gas escape and thus gas pressures and explosive potential may increase.

A few sample questions to illustrate uses of WOVOdat.
Each question will be translated into database queries and grouped into reports. Custom queries will also be possible.

Examples of questions from volcanologists:

- Given developing patterns of seismicity, ground deformation, gas emission and other parameters, where else has such unrest been observed?
  - What happened?
  - On this basis, what are the probabilities of various scenarios (including false alarms)?
- What are the most diagnostic precursors to eruptions of a particular volcano, type of volcano, or type of eruption? This is a deceptively complicated query as it requires examination of temporal, spatial and spectral patterns of multiple parameters and examination of unrest rates and changes in those rates.
- How does one particular aspect of unrest (e.g. long-period earthquakes) correlate with another (e.g. SO_2 fluxes)?
- What is the significance of a particular change, e.g. sudden seismic quiescence? From a number of possible causes which are best supported by the data of WOVOdat?
- What does the character of volcanic unrest imply about the coupling and interaction between magmas, hydrothermal systems and regional tectonics?
- What interesting patterns exist in the data of volcano monitoring, especially, patterns that have escaped prior notice? (This is data mining.)
-

Examples of questions for (and from) students or the general public:

- What does volcanic unrest tell us about why volcanoes erupt? What are the necessary and sufficient conditions for eruption?
- What are the most common precursors of volcanic eruptions? Which of these are the most reliable? What combination is the most reliable? How reliable are these one- (day, week, month, etc). before an eruption?
- What systematic differences exist between unrest at volcanoes of one type vs. another, and how might these relate to magmatic properties and processes?
- How often is the volcano in our backyard restless, and how does its current activity compare with baselines and periods of greater unrest?

Parameters and types of unrest

A preliminary list of parameters to be included has been developed following suggestions made at the WOVO Planning Workshop and, are detailed on the WOVO WEB page (http://www.wovo.org/WOVOdat/parameters.htm). A final list, together with suggested units, data formats and reporting protocols will be drawn up after Phase-III of the work plan, described below. The following table summarizes general classes of data into which the large variety of individual parameters can be fit.

Data Type	Examples
***Non-gridded time series*:**	
Continuous high or low rate	Continuous seismic trace; tilt data every 10 minutes
Triggered high or medium rate	Triggered seismic traces (event waveforms); tilt data every 15 sec if a certain rate is exceeded in the 10-minute data.
Occasional data that apply to the volcano as a whole	SO_2 flux; magma discharge rate e.g., rate of dome growth
Calculated, cumulative data	Cumulative seismic energy release, running b-values
Gridded:	
Uniform sampling	DEM's, geologic maps, seismic tomography, InSAR interferograms, GOES thermal image data
Time-space point data	Earthquake hypocenters, fault-plane solutions, x-y-z coordinates of GPS benchmarks, elevations of precise leveling benchmarks, fumarole T's, spring discharges,
Images:	
Photos, remote sensing images	Photos of volcano and its eruptions; images rectifiable to a DEM, snapshots of seismograms that aren't otherwise digitized.
Video	Time-stamped video clips of explosions and other events that can be correlated with and explain monitored unrest.
Text:	
Primary index fields	Volcano name and catalogue number; eruption years, types and magnitudes; bibliographic data, author, title, etc; name, telephone, address, email of contacts, etc.
Free-form text	Notes, longer narrative descriptions.
Definitions	Definitions of earthquake types VT, LP, etc.
Station information	Location, dates of operation, site characteristics
Network information	Seismic velocity model, hypocenter determination technique, geodetic reduction technique, gas analysis technique
Instrument information	Model, gain, calibrations; response; for satellite remote sensing data, spectral and on-the-ground spatial resolution.

<u>Plan for Phases II and III</u>

System Design

The WOVOdat system can be thought of as having three parts, input modules, output modules and the database itself. The design of the fundamental schema for information to be contained in WOVOdat will be at the heart of this project. A computer specialists in modern database engineering will be hired to do the technical design and will interface with the computer specialists at IRIS and UNAVCO on technical details of data exchange. The schema will be developed to retain as much high-level generality as possible to allow the easy addition or modification of lower level elements as the system develops and matures. Once a basic framework is defined, simple example input and output models will be written. These can then be modified and adapted to other types of data or for the same type of data from different sources. The fundamental database engine will be MySQL for this development work. This DBMS is OPEN SOFTWARE, relatively powerful and runs on many different computer platforms. It has many of the advanced features needed for this project such as fixed and variable-length records, large number of data types, *FULLTEXT search*, multi-level security, high scale-ability, easy IP connectivity and a large suite programming interfaces (APIs).

The overall WOVOdat project has five phases:

I. Concept development and approval from member observatories (completed)
II. Database design and development of preliminary data entry tools
III. Pilot testing and revisions to design as needed
IV. Populating the database
V. Maintenance and Enhancement (stable use)

Phase I, Concept development and approval from member observatories: (completed)

At the first WOVOdat Planning Workshop in Denpasar, Indonesia (July 23-24, 2000), forty participants from 18 countries, 17 WOVO observatories and the Smithsonian Institution's Global Volcanism Program discussed general and specific issues pertaining to WOVOdat. Based on this workshop (*ed- and a second workshop,held Dec 11-13, 2000*) (http://www.wovo.org/WOVOdat/plansum.htm) a detailed list of parameters relevant to volcanic unrest was drawn up (http://www.wovo.org/WOVOdat/parameters.htm).

In addition, a list of potential user groups and approximately 50 common searches of WOVOdat has been developed based on the workshop and subsequent communications and is available on the WOVO Web pages.

Phase II, Database Design

Part IIA -- Requirements analysis: Interviews with volcanologists will check and supplement our preliminary understanding of how they use, enter and store their data; what their current needs are; what questions they currently ask of their data; what they anticipate their future needs to be (future technologies and techniques); what types of information gained from WOVOdat would be beneficial to them; how quickly they need their queries returned; and what their current database skills are. Throughout this process we will collect sample data sets. We will then continue the interview process with a group of educators and students to determine their needs, ideal data formats and database skill levels. In addition, a small group of scientists from other fields will be interviewed.

Part IIB -- Data Analysis/Logical Model/Schema Design: We will review the preliminary data types and scientific *business* questions in part IIA to expand our initial logical model. At this stage we will prioritize the list of anticipated questions and expand the set of basic reports (views and view summaries) that will define the database and presentation interface. We will ensure that the basic reports answer the highest priority questions and that the multiple lower priority questions can be answered by custom (handwritten) queries. As with any design process, we aim to design a flexible model such that future questions can be incorporated into standard reports.

One of the primary challenges of designing a database of this complexity is to determine how to identify and analyze patterns and then how to display the returned information. We will begin by translating our set of common questions (examples in sidebar 2) into formal queries to determine what data is needed and how it must be organized. To answer a simple question such as where have specified gas emissions been observed and what was the outcome, we could write a query to select all episodes of unrest in which emission of gas *g* is greater than mass *m* during period *t* and link to the Smithsonian's file of historical eruptions to learn the outcome. From these basic queries, we will organize the database for multi-component questions that examine changes in seismicity, ground deformation (x, y and z coordinates), gas emissions, etc. for a period of time. Rates and acceleration or deceleration of rates will always be queried, as they are often more informative than simple magnitudes. We anticipate that the translation of scientific questions into queries and reports will take considerable time.

In creating the schema we will be working on balancing varied user requirements, database maintenance and query performance. Preliminary conversations have indicated that interactive query returns on the order of a few minutes are most useful and common. Our schema will therefore be developed to handle these types of queries most efficiently. Preliminary conversations about the data types at each of the volcanic centers have shown the data collection to be uneven in space, time and quality. We will explore using XML as a data definition agent to allow for this type of variability and for future adaptability. One of the goals of this project is to help set voluntary standards for future data collection.

Part IIC -- Sample Data Population - We anticipate work with several volunteer WOVO observatories and a small sample of recent data to create basic scripts for data extraction, loading, mapping, cleansing and validating.

Part IID – Presentation Interface Development – Initially all queries will be hand-written and grouped into reports for test purposes. Once we are satisfied with the standard set of reports we will work to simplify query development and refine the presentation interface. Access to the database will be provided through a standard web interface.

We will also make sure that basic documentation is available for use by the QA team in Phase III. Additional training materials will be developed as parts of Phase IV and Phase V with the goal of making the database easily accessible to all user groups.

Phase III: Sample testing

Preliminary sample testing will have occurred as part of the presentation interface development process. Once we are satisfied with the results for our standard reports we will begin a larger QA cycle using the pilot cases of volunteer observatories.

Specifically we will test to be sure that the schema and data management utilities are optimized for data entry, archiving and both standard and custom queries. Utilities will be improved as needed, as will scripts for standard queries. The overall test of the design and utilities will be that a user can quickly gain a clear overview of pertinent data, including the character of an episode of unrest and easily execute all standard and most custom queries.

Phase IV: Populating the database:

At the end of Phase II and III we plan to have WOVOdat in a state to be able to ingest large amounts of data from many different sources. Through presentations at regular meetings and through communications to WOVO members the system will be demonstrated and be available for general use. We anticipate a large and lengthy task of data recovery, compilation, reformatting and database loading. It may be as much as 5 years before the database is large enough to be significantly useful and longer before most pre-digital data are included and the project is relatively stable and ready for Phase V.

Most effort in Phase IV will be by scientists at WOVO observatories. The WOVOdat team will provide modest help as needed, e.g., with scripts to translate a unique observatory data format into the WOVOdat standard. The WOVOdat team will also check contributions as they are readied, to catch any problems early in each observatory's effort.

Phase V: Maintenance and Enhancement, - three parallel tasks.

- Ensure that data flow smoothly from volcano observatories and other contributors into the database, with all of the necessary metadata. We anticipate that WOVOdat will be so useful that observatories and others will gladly continue their contributions, but we also know from IRIS and other experience that designated data managers will be needed to remind contributors to maintain the metadata as well.
- Serve data as requested. Most of the data serving will be automatic, controlled interactively by the users. Some commonly used data products might be prepared regularly by the WOVOdat staff.
- Develop and provide additional tools to assist users. We anticipate that these will include teachers' modules for K-12, state-of-the-art data modeling and interpretation modules, and updates of the data entry toolkit to reflect changes in monitoring and communications technology. Development of these tools could be started during Phase III or IV. Basic data query and teacher tools would be developed jointly by the WOVOdat team and teachers; most modeling and interpretation tools would be developed and contributed by user scientists.

End of summary, 05/2002