WOVOdat – An online, growing library of worldwide volcanic unrest

C.G. Newhall, F. Costa, A. Ratdomopurbo, D.Y. Venezky, C. Widiwijayanti, Nang Thin Zar Win, K. Tan, E. Fajiculay

1. Introduction

Volcanoes exhibit a complex suite of geophysical, geochemical, geologic, and hydrologic changes (unrest) as magma ascends and prepares to erupt. For example, many small earthquakes will mark where magma pressures are fracturing the host rock through which magma will rise (McNutt, 2005; Benoit and McNutt, 1996a, 1996b). Earthquakes can also be induced by volumetric compression of aquifers along pre-existing tectonic faults (Terakawa et al., 2013). Although volcanic unrest can at times beewilder and lead to no eruption, it also offers a window into the volcanic subsurface and the potential to forecast eruptions (e.g., Newhall and Dzurisin, 1988; Diefenbach et al., 2009; Segall, 2010; Moran et al., 2011; Bell and Kilburn, 2013; Papale, 2015; Loughlin et al., 2015 and many others).

Seismic, geodetic, gas, and other volcano monitoring data are collected by about 80 observatories around the world, most of which are members of the World Organization of Volcano Observatories (WVO). Unfortunately, ground-based data have a myriad of different formats, and are stored mainly within each observatory (sometimes, only in the files of individual researchers!). As such they cannot currently be accessed and visualized together. A growing number of data sets are open, but many others are still inaccessible except through the published literature and direct queries to those who collected the data.

Other geodetic, gas, and thermal data are remotely collected by satellites (e.g., Francis and Rothery, 2000, and papers in Mouginis-Mark et al., 2000). Satellite data are more standardized, and both spatial and temporal resolutions are improving rapidly, but some are still costly to obtain and we must be careful to not oversimplify when extracting simple parameters suitable for (time-series) comparison with ground-based data. Moreover, care must be taken to distinguish gas plumes from nearby but independent volcanoes, and to associate distal plumes with their correct source volcanoes.

The lack of standardization in data formats and database architectures makes it difficult to compare episodes of unrest, to find data for analogues to any current unrest, to search for subtle but instructive...
patterns, to find correlation within multi-parameter precursory datasets, or to test hypotheses about unrest at a large number of volcanoes. This fragmented state of affairs utterly fails to take advantage of the intellectual power of worldwide observatory experience and of galloping information technologies.

Potentially, more than a century's worth of volcano monitoring data 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.

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 bulging outward at >1 m/day and would soon be unstable and that a large landslide was possible (Lipman et al., 1981; Voight et al., 1981). What was not so clear was that the hydrothermal system was probably pressurized. As a result, a complex interplay between hydrothermal pressures, steam explosions, over-steepening, and gravity was taking place. This 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. Potentially analogous behavior at Bandai-san in 1888 and Bezymianny in 1956, virtually unmonitored, was discussed before collapse at Mount St. Helens but not taken as a definite analogue (Siebert et al., 1987; Belousov et al., 2007).

Now volcanologists know better 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 Soufrière Hills Volcano on Montserrat, including Barry Voight with Mount St. Helens experience, 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 (Belousov et al., 2007). 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.

To make volcano monitoring data more readily accessible, and to enable searching and comparisons of those data, WOVO is developing WOVOdat, a web-accessible database of seismic, geodetic, gas, hydrologic, and other unrest from volcanoes around the world. WOVOdat stores historical data that are already processed by the contributor; the original raw data remain with that contributor. Although the principal contributors and users of WOVOdat are member observatories and associated, WOVOdat is open to all and we anticipate wide usage for academic research and teaching. Here we present the data format and structure of WOVOdat, a brief history of its development, the current status and data population level, and some examples of applications and future plan.

2. Brief history of how WOVOdat came to be and recent evolution

The concept of WOVOdat dates back to the eruption of Mount St. Helens in 1980 and the unrest at Long Valley, Campi Flegrei, and Rabaul calderas in the early 1980’s. The key developments are listed here in chronological order:

- In 1980, when Mount St. Helens was threatening to erupt, 3 potential analogues were considered—Bezymianny, Bandai, and Usu (e.g. see comment above about the value of historical experiences).
- In 1981, WOVO was formed to promote communication between observatories. About 70 WOVO observatories now have email connections and most have their own websites (see www.wovo.org/directions and www.iavcei.org). About the same time, researchers in various groups around the world started exploring how to represent accounts of past unrest from papers published in many languages, into a simple database. C. Newhall (USGS) and H. Okada (Hokkaido University) designed what was the equivalent to a spreadsheet of volcanic unrest, but they put the project on hold because it was too simplistic to give an adequate description of each episode of unrest.
- A better picture of unrest was provided by “cut and paste” compilations of unrest at Japanese volcanoes (Hiromu Okada and colleagues, unpublished), of unrest at large calderas of the world including Long Valley (Newhall and Dzurisin, 1988), and several reviews of specific types of volcanic phenomena (e.g., phreatic eruptions, Barberi et al., 1992).
- The Long Valley case and simultaneous unrest at Campi Flegrei and Rabaul calderas showed the urgent need for better ways to access historical records. It took Newhall and Dzurisin (1988) five years (!) of library research to find a then-comprehensive set of examples of caldera unrest with known outcomes. Given the time required to search through literature, especially during times of volcanic crisis when rapid answers are needed, and advances in computer and database technology, a group of scientists from WOVO observatories resolved to create a central, relational database called WOVOdat. The new database would contain processed or “catalog” data on seismicity, ground deformation, gas emissions and other monitoring data types.
- In a renewed effort to encode the essence of volcanic unrest into a database, Benoit and McNutt (1996a, 1996b) took summaries of volcanic earthquake swarms from Japan’s Bulletin of Volcanic Eruptions and put them into another equivalent of a spreadsheet, from which they drew generalizations about earthquake swarms. Descriptions were of each swarm as a whole, without records of individual earthquakes within the swarm.
- Two workshops were convened to plan WOVOdat and to invite observatory and other community “buy-in.” The first was held in conjunction with the IAVCEI General Assembly in Bali in 2000, and the second in Menlo Park following the Fall 2002 AGU meeting. Details of discussion in those workshops may be found at http://www.wovodat.org/about/WOVOdat-Bali-2000.pdf and http://www.wovodat.org/about/WOVOdat-Menlo-2002.pdf
- We then sought and received seed funding from the USGS-USAID Volcano Disaster Assistance Program, from which one of us (Venezky) was hired to organize all early ideas into a detailed schema, tables, and table relationships. Early development focused on design of the database and defining common data formats and data entry protocols to facilitate interchangeability of monitoring instruments, data, and data processing tools. The result was a document, WOVOdat 1.0 (Venezky and Newhall, 2007), describing but not yet actually coding the required tables.
- At the same time that this early design work was proceeding, we formed two WOVOdat Steering Committees, one for policy and scientific oversight led by the first author, and the other for technical matters led by Florian Schwandner. Members were drawn from observatories, government agencies, and university research groups. Subsequently, the two steering groups were merged into one.

In 2008, WOVOdat received a much-needed boost with major funding from the Government of Singapore, through the Earth Observatory of Singapore (EOS), Nanyang Technological University. That enabled hiring of a staff to encode the design into a MySQL database and to get on with the huge job of populating WOVOdat. Since then there have been several stages of development (see details in Fig. 1). WOVOdat was first opened to the public during the 2013 IAVCEI meeting in Kagoshima (Japan).

3. Content and structure of WOVOdat

WOVOdat contains only processed data with physical meaning, i.e., interpretable and publishable. Examples include earthquake locations, daily earthquake counts, GPS positions or relative position changes,
and daily gas fluxes. Moreover, as much as possible, WOVOdat data are continuous in time, and include periods of both quiet and unrest, since both are critical for anticipating eruptions and building probabilities for event trees (e.g. Newhall and Pallister, 2015). Data during times of quiet define baselines (e.g., background levels of activity). Data of pre-eruption unrest show possible eruption precursors; monitoring data during eruptions capture any further changes before later phases of eruptions; and post-eruption data capture magma recharge and/or tectonic relaxation. By using continuous data rather than data from immediately before an eruption, it is possible to obtain a more complete view of magma and gas input to volcanic systems years or even decades before eruptions, and it also avoids our prejudging what is relevant.

WOVOdat is a relational database, keyed in the first instance to Volcano, then to network and instruments, and finally to the data themselves. WOVOdat also logs metadata about station location, dates of station operation, instruments, and data acquisition settings. WOVOdat also offers basic tools for searches, visualization, and data analysis as described in later sections.

The general schema of the database consists of the following (Fig. 2):

- Background information of the volcano including maps, rock compositions, tectonic setting, etc.
- Inferred processes of unrest, from the literature
- Eruption information including date, type, and size. Most eruption data come directly from the Smithsonian Institution’s Global Volcanism Program (GVP) but we are also adding some details, e.g., onset of magmatic phases, if those are not already reflected in the GVP catalog.
- Alert level changes (from Winson et al., 2014 and other sources) are also included.
- Data on seismicity, ground deformation, gas emission, and other unrest, time-stamped, geo-referenced, and suitable for plotting in time series and 2-D or 3-D visualizations. While WOVOdat does not presently include raw data, we do include a small number of digital seismic waveforms to illustrate earthquake classifications used at specific volcanoes.
- Supporting metadata on instruments used, data quality, links to images, bibliographic references, etc.

Table 1 is an example of the metadata and fields for tilt data. The current list of tables and parameters (WOVOdat 1.1) may be found at http://www.wovodat.org/doc/database/1.1/wovodat11_doc.pdf, organized mainly by types of unrest (e.g., seismic, ground deformation, etc.).

Finally, we have developed a WOVOdat standalone software version that can be run within an observatory. This has been already adopted, with local modifications, by PHIVOLCS (Philippines), CVGHM (Indonesia), RVO (Papua New Guinea), NIED (Japan) and soon, observatories in several more countries. Mirror servers such as the first one at NIED (Japan) will also run the WOVOdat software. The standalone package uses the WOVOdat database schema and data formats for daily, in-house volcano monitoring and in-house archiving, which can then feed seamlessly into WOVOdat when the data are made publicly available. This will help WOVOdat to have sustainable data population, via automated data uploads to the WOVOdat main server when the data are ready to share publicly.

4. Current status of WOVOdat and data sources

The current main menu on the WOVOdat webpage (www.wovodat.org) points users to three general activities:

- Exploration of the WOVOdat system: database schema & structure, and online user interface. Users can browse the documents online or download them. Users can also download the standalone software package.
- Interactive browsing of what kinds of data are available, visualizations of the data in graphs and tables, and the option for users to download selected data. To download data, users simply register with their contact details (email, name, institution), and agree to the WOVOdat Data Policy. The flow of data uploads and downloads is shown in Fig. 3.
• Interactive data input through online forms and file conversion tools (e.g., to convert a file into WOVOdat format before uploading).

Estimating what percentage of available data is already entered into WOVOdat is complicated. We know, for example, that while some large data sets (e.g., from Japan, New Zealand, and the US) are already included, other large data sets are still in the earliest stages of addition. One way to estimate is to compare the number of eruptions (since 1950) for which there was one degree or another of monitoring, with the number of eruptions since 1950 for which at least one type of monitoring data is already included in WOVOdat. Using this approach, we count 2711 eruptions, of which 955 or about 25% have at least one monitoring data set already in WOVOdat. Of those 955 data sets, about 55% are from observatories, about 10% from open catalogues, about 15% are from research projects, and the rest (20%) are from published literature. The largest potential data contributions in terms of eruptive episodes (counting all 2711 cases) are 17% from Indonesia, 15% from Japan, 11% from USA, and 9% from Russia, the rest from many other countries (Fig. 4).

An alternative approach to estimating completeness is to compare the volume of data entered vs. that known to exist. Some countries have large volumes of data on a relatively small number of volcanoes and episodes of unrest – the result of long histories of monitoring and dense instrument networks. We know how much data has already been entered into WOVOdat, but we don’t know the volumes of data sets not yet seen, so we cannot yet estimate volume % completion by this approach. The volume of data (and volume % completion) in WOVOdat will increase significantly as data from these countries are added.

With regard to data types, WOVOdat at present contains mainly seismic and deformation data (Table 2 for details), but we continue to add new data of all types. We encourage participation by specialists in particular types of data, as these individuals will know both sources and nuances of these data.

WOVOdat offers two categories of data access as chosen by the contributors: Category A, available for search, visualization and download immediately after being uploaded, and Category B, available for searches and visualization immediately, and download 2 years after the time of data collection.
5. WOVOdat visualization, search, and download tools

WOVOdat can display data for single volcanoes and side by side comparisons of two volcanoes (Fig. 5). Displays include shaded relief and monitoring stations, 2D or 3D hypocenter display, and time-series plots of default or user-selected multiple parameters. Plots of temporal evolution of the various parameters will include the onset of eruptions and, where available, individual phases of eruptions. In addition, it is

Table 1

Table of data in WOVOdat. “dd_tlt” indicates digital (electronic) tilt data. The third modifier after “dd_tlt,” e.g., “srate,” indicates a detail of tilt data, in this case, sampling rate. Field names starting with “cc” refer to contacts for further information.

| #  | Column              | Type          | Unit         | Comments                              |
|----|---------------------|---------------|--------------|---------------------------------------|
| 1  | dd_tlt_id           | Mediumint(8)  |              | Tilt data identifier                  |
| 2  | dd_tlt_code         | Varchar(30)   |              | Tilt data code                        |
| 3  | ds_id               | MEDIUMINT(8)  |              | Deformation station identifier        |
| 4  | di_tlt_id           | Smallint(5)   |              | Tilt/strain instrument identifier     |
| 5  | dd_tlt_time         | Datetime      |              | Measurement time                      |
| 6  | dd_tlt_time_seconds | Decimal(2,2)  |              | Centisecond precision for measurement time |
| 7  | dd_tlt_time_unc     | Datetime      |              | Measurement time uncertainty          |
| 8  | dd_tlt_time_unc_seconds | Decimal(2,2) |              | Centisecond precision for measurement time uncertainty |
| 9  | dd_tlt_srate        | Double        | second       | Sampling rate                         |
| 10 | dd_tlt1             | Double        | μrad         | Radial tilt measurement for component-1 or component-X (positive is down to the north) |
| 11 | dd_tlt2             | Double        | μrad         | Tangential tilt measurement component-2 or component-Y (positive is down to the east) |
| 12 | dd_tlt_err1         | Double        |             | Tilt 1 error                          |
| 13 | dd_tlt_err2         | Double        |             | Tilt 2 error                          |
| 14 | dd_tlt_proc_flg     | Enum(‘P’, ‘R’)|              | Flag: P = processed, R = raw          |
| 15 | dd_tlt_temp         | Double        | °C           | Soil temperature                      |
| 16 | dd_tlt_mot          | Double        |              | State of the battery                  |
| 17 | dd_tlt_ori          | Enum(‘D’, ‘O’) |              | A flag for source of data, D = digitized, O = original from observatory |
| 18 | dd_tlt_com          | Varchar(255)  |              | Comments                              |
| 19 | cc_id               | Smallint(5)   |              | First owner ID                        |
| 20 | cc_id2              | Smallint(5)   |              | Second owner ID                       |
| 21 | cc_id3              | Smallint(5)   |              | Third owner ID                        |
| 22 | dd_tlt_loaddate     | Datetime      |              | The date the data was entered (in UTC) |
| 23 | dd_tlt_pubdate      | Datetime      |              | The date the data become public       |
| 24 | cc_id_load          | Smallint(5)   |              | Contact ID for the person who entered the data |
| 25 | cb_ids              | Varchar(255)  |              | List of bibliography ID, link to bibliography table (cb). Can be multiple, separated by a comma. |

Indexes & links

| Keyname      | Unique | Column   | Null | Links            |
|--------------|--------|----------|------|------------------|
| PRIMARY      | Yes    | dd_tlt_id| No   |                  |
| CODE         | Yes    | dd_tlt_code| No   |                  |
| STATION      | No     | ds_id    | No   | ds ds_id         |
| INSTRUMENT   | No     | di_tlt_id| No   | di_tlt di_tlt_id |
| OWNER 1      | No     | cc_id    | No   | cc cc_id         |
| OWNER 2      | No     | cc_id2   | Yes  | cc cc_id         |
| OWNER 3      | No     | cc_id3   | Yes  | cc cc_id         |
| CONTACT ID   | No     | cc_id_load| No   | cc cc_id         |
| BIBLIOGRAPHY | No     | cb_ids   | Yes  | cb cb_id         |

Fig. 3. WOVOdat data flow.
also possible to search for volcanoes with specified characteristics, or features of unrest. Finally, users can download data of Category A in csv format.

We are also currently planning to develop tools that would allow testing, in hindsight, of various methods for eruption forecasting.

More details about the different menus and tips for searching can be found in Supplementary Appendix A.

6. Links of WOVOdat with other global volcano databases

WOVOdat is being built in parallel with other volcanological databases. The Smithsonian Institution’s Volcanoes of the World (VOTW, Simkin et al., 1981; Simkin and Siebert, 1994; Siebert et al., 2010; Global Volcanism Program, 2013) is the premier source of data about historical and Holocene volcanism. The Global Volcano Model’s LaMEVE (Crosdwell et al., 2012; Brown et al., 2014) database contains additional details of selected large volcanic eruptions, most before the era of volcano monitoring. WOVOdat partners with the Smithsonian’s GVP so that dates of eruptions can be plotted within WOVOdat’s records of unrest, making it easy to see which unrest led to eruptions. WOVOdat captures the data of inter- and pre-eruption unrest, while VOTW captures the outcomes of that unrest. Additional details of those eruptions, also mainly from VOTW, allow for Boolean searches of WOVOdat keyed to eruption type or Volcanic Explosivity Index (VEI; Newhall and Self, 1982).

There are a number of other databases for specific types of data such as InSAR (e.g., COMET project; https://volcanodeformationblogs.irlt.org/), lava dome growth (DomeHaz; Ogburn et al., 2015), and a Deep Carbon Observatory/GVP database on volcanic emissions (DECade; Clor et al., 2013). WOVOdat is in discussion with these efforts to bring time-stamped data into WOVOdat where it can be viewed alongside other data of unrest. Conversely, users of those databases can turn to WOVOdat for the temporal and spatial context of observations in those databases.

Individual observatories or networks of national volcano observatories have developed a variety of data management solutions for their own operations. In principle, these can interface with WOVOdat after preparation of scripts to map between data fields and structures. We certainly do not wish to duplicate the databases of individual observatories, nor of specialized disciplines, but, instead, to bring key parameters of these varied datasets together where they can be searched and studied in the same holistic way that volcano observatories do as they try to understand and forecast the outcomes of unrest. We recognize that it is often easier to get funding to develop one’s own specialized database than it is to get funding to support a larger, community database. But we also see great synergies if those who are developing specialized databases make them compatible with WOVOdat, so that data can be freely shared.

WOVOdat links each volcano to the GVP page through the new Volcano number (VNUM), and the GVP and others can link to WOVOdat in the same way. Any database developer wishing to link to a WOVOdat “single volcano page” can use the following URL that contains the VNUM-based variable; i.e., http://www.wovodat.org/precursor/index_unrest_devel_v6.php?vnum=300260 for Redoubt volcano. Although there is potential for real time connection between different volcano databases through web link services, this is still something for the future.

7. Possible applications of WOVOdat, from simplest to more complex

We foresee a myriad of applications of WOVOdat, which we have broadly grouped into four.

7.1. Data standardization, organization, and visualization

The first and simplest level of application is organization, visualization, and archiving of multiparameter data that are currently available.
only in separate files and formats. This seems an obvious and simple objective but it remains a major task. A first level of visualization of multiparameter information is a stack of data plots on the same time scale (Fig. 6) which allows one to track the time evolution of various parameters. An even more useful way to visualize the multi-parameter data is to overlap all of them on top of each other in a single diagram (Fig. 7).

We also recognize that some preselected sets of data will help novice users to find the most useful parameters to plot. We invite observatories to contribute multiparameter data sets that they judge best characterize specific episodes of unrest, including but not limited to those leading to notable eruptions (see the example of Pinatubo 1991 in Fig. 8). Those who have collected the data and been responsible for warnings will have already devoted much thought to the significance of the data and are in the best position to define data sets needed to capture the essence of an episode of unrest.

7.2. Data search and comparisons

The next level of application is a Boolean search to find episodes of unrest with analogues at the Volcano level (e.g., at volcanoes of specified type, magma composition, repose period, or any other characteristics of interest) and/or in characteristics of their unrest (e.g., unrest with VT earthquakes ≥ M4, or with strain rates or gas flux rates within specified ranges). Searches that key on volcano characteristics can be used to anticipate what unrest might be expected before eruption of a nearby reactivated volcano with similar characteristics. Searches that key on one or more characteristics of unrest can be used to answer the question, “Given observations of X, Y, Z at a newly-reactivated volcano, where else have these been seen and how did the unrest culminate?”

In the current version of WOVOdat, Boolean searches produce listings of volcanoes and episodes of unrest that met the search criteria. To illustrate, we might search for all examples in WOVOdat of SO2 flux > 10,000 t/d. This might seem like an ominously (e.g. threatening) high flux rate, but is it ominous? A Boolean search of current data in WOVOdat yields 12 examples (Supplementary Appendix B), and a quick perusal of that list suggests that, in general, high SO2 flux by itself is not ominous. A better indicator might be some combination of SO2 flux and a measure of stress or pressurization within the volcano.

Fig. 9 shows time series plots of seismicity (RSAM) and SO2 flux from eight different episodes of volcanic unrest, selected from results of a Boolean search for data sets of both RSAM AND SO2 flux. Of the seven volcanoes represented, four are volcanoes in which conduit-filling magma probably solidifies into a plug between eruptions, allowing accumulation of gas in magma if there is fresh magma supply from depth. In this group, we include Redoubt, Augustine, Mount St. Helens, and Pinatubo. In contrast, three are open-vent volcanoes in which only a minimal cap solidifies between eruptions, allowing fresh gas to escape. Volcanoes in this group are Colima, Asama, and (post-2000) Miyake-jima. What might we learn from comparison of these cases? A full treatment is beyond the scope of this paper, but several interesting possibilities arise. First, seismicity (swarms) and SO2 flux correlate better at the plugged or semi-plugged volcanoes than at the open-vent volcanoes. Indeed, there is a distinct lack of correlation at the open-vent volcanoes. Second, at the plugged or semi-plugged volcanoes, there’s a hint of temporary decrease in SO2 shortly before several eruptions (initial magmatic eruptions at Redoubt and Pinatubo; a slightly later eruption at Augustine). Whatever hypotheses one might formulate on the basis of these plots can be tested using additional datasets that will be added soon to WOVOdat.

7.3. Data analysis tools, including unrest indices and probability estimates

The next step up in complexity of application involves “analysis tools” which are currently still in development phase. Here, we might identify thresholds above which unrest must rise before an eruption is possible. Or we might identify different stages in the development of unrest at specific volcanoes, such as those proposed for a Volcanic Unrest Index (VUI) by Potter et al. (2015). Often the analysis tools will require modest further processing of data, e.g., to calculate displacements from sequential position data, or to get inverse RSAM, the 1st or 2nd derivatives of GPS position data (velocity and acceleration), or time-averaged or cumulative values of selected parameters (e.g., long-term average SO2 flux, cumulative seismic energy release, etc.).

Still others might be estimates of eruption probability gain indicated by unrest X, Y, Z as gleaned from the statistics of data in WOVOdat and the Smithsonian’s VOTW4. From inverse RSAM, one can project to an eruption time window directly (Voight and Cornilus, 1991; Cornelius and Voight, 1994), or use it in probabilistic assessments (Fig. 10). WOVOdat’s analysis tools might rely on data from successive episodes of unrest from a single volcano. More often, they will utilize data from the volcano in question and unrest at a number of analogous volcanoes.

An extra word might be added about use of WOVOdat in probabilistic eruption forecasting. Until recently in the field of earthquakes, probabilistic forecasting was seen by many as diametrically opposed to process-related, deterministic prediction. That false dichotomy is now being discarded in favor of probabilistic models that consider both processes and stochasticity in nature (e.g., Jordan et al., 2014). A growing number of volcanologists see the process of estimating probabilities for volcanic events as a way to marry the statistics of empirical experience with monitoring- and process-oriented discussions. In general, empirical experience (including that offered by WOVOdat) will be the basis for initial (a
priori) Bayesian probability estimation, while new monitoring results and interpretations will be the basis for updated (a posteriori) estimates. The process of populating progressively more specific nodes of a probability tree is a great framework for discussion, debate, and further research into the meaning of past and current unrest (Newhall and Hoblitt, 2002; Aspinall et al., 2003; Marzocchi et al., 2008; Sobradelo et al., 2014; Komorowski et al., 2015; Newhall and Pallister, 2015).

7.4. Pattern recognition applications

The fourth and most sophisticated level of applications that we currently envision includes automated, machine searches and other data mining for patterns in unrest that no one has yet suspected or postulated. Pattern recognition algorithms are increasingly sophisticated and well suited to looking for correlations and co-variance of parameters.
that we might not even suspect to be related. Given how rapidly information technology is shooting ahead, we expect that there will be even more complex applications in the not too distant future (e.g., Curilem et al., 2016).

7.5. Questions that could be addressed using WOVOdat

Examples of questions that could be addressed using WOVOdat include the following:

Fig. 6. Stacked graph of 4 different precursory parameters of November 1998 Colima dome forming eruption. (a) Seismicity, (b) SO₂, (c) temperature, (d) EDM. Users can view multiple parameters of unrest in stacked time-series graphs. Onsets of phreatic eruptions (in gray arrow) and magmatic eruptions (in red arrow) are shown. Data compiled from Galindo and Domínguez, 2002; Ramírez-Ruiz et al., 2002; Taran et al., 2002; Zobin et al., 2002a and Zobin et al., 2002b; Engberg, 2009).

Fig. 7. Same data as Fig. 6, but displaying the 4 different parameters in one composite graph. Simultaneous display of several parameters on a common timescale can be more useful than separate plots of individual parameters.
What are the most common precursors of volcanic eruptions? Which of these are the most reliable? How reliable are these, one day, one week, or one month before an eruption? Is there a combination of precursory phenomena that is an especially reliable predictor of eruptions?

What are the most diagnostic precursors to eruptions of a particular volcano, type of volcano, or type of eruption?

How does current activity of my hometown volcano compare with its baseline activity (quiescent state)? Is the current activity common or unusual?

Given current unrest, how soon might an eruption occur? At the earliest? Most likely? At the latest? Given developing patterns of seismicity, ground deformation, gas emission and other parameters, where else has such volcano unrest been observed? What was the outcome? Using statistical correlations of unrest and outcomes, what are the probabilities of various scenarios (including false alarms)?

How do the various data types (parameters) correlate with one another? For example, how do long-period earthquakes correlate with SO2 emission rates? (A simpler example was already mentioned in relation to Fig. 9, above)

What is the significance of a particular change in one or more parameters, e.g. sudden seismic quiescence? From a number of possible causes, which are best supported by data?
Fig. 9. SO₂ emission and seismicity (swarms) during several episodes of unrest. Please note that this is a simplified chronology of the eruption phases (e.g., compare Fig. 8 with Fig. 9e for Pinatubo). The vertical gray dashed arrow is the onset of (phreatic) eruption as given in Global Volcanism Program (2013) and the red arrow is the beginning of the magmatic eruption from various references cited, applicable for each individual figure panels. We include onsets of magmatic eruptions because, from the perspective of those concerned with volcanic hazards, phreatic eruptions are potential precursors of more hazardous magmatic eruptions rather than the “main event.” To be sure, some phreatic eruptions can kill, but we think it important to show the timing of phreatic eruptions, instrumentally recorded unrest, and warnings relative to the onset of more hazardous magmatic phases.

a. Redoubt, July 2008–July 2010: unrest spanning small ash-and steam explosion (March 15, 2009); lava effusion (March 22–23) accompanied by intermittent explosions (March 23rd–April 4th). Extrusion ceased on April 6th. Data compiled from Dixon and Stihler, 2009; Dixon et al., 2010. Dixon et al., 2011. Werner et al., 2012; Werner et al., 2013; Schafer, 2012. b. Augustine, July 2005–July 2007: unrest spanning phreatic to explosive phase (January 11–28, 2006) followed by effusive phase (March 8–16, 2006). Data compiled from Dixon et al., 2006; Doukas and McGee, 2007; Dixon et al., 2008a and Dixon et al., 2008b; McGee et al., 2010; Power et al., 2010. c. St. Helens, 1980–1984: unrest spanning phreatic explosion on March 27, 1980, giant sector collapse, lateral blast and Plinian eruption on May 18, 1980, and subsequent smaller explosive and dome-building eruptions. Data compiled from PNSN catalog; Endo et al., 1981; Casadevall et al., 1981; Lipman and Mullineaux, 1981; McGee and Casadevall, 1994; Gerlach et al., 2008. d. St. Helens 2004–2005: unrest spanning minor phreatic or phreatomagmatic explosion; mostly, sluggish dome extrusion. Data compiled from PNSN catalog; Endo et al., 1981; Casadevall et al., 1981; McGee and Casadevall, 1994; Gerlach et al., 2008; Sherrod et al., 2008. e. Pinatubo, 1991: unrest spanning phreatic explosions on April 2, 1991; vanguard lava extrusion starting June 7, magmatic explosions June 12–14, and climactic Plinian eruption on June 15. Data compiled from Daag et al., 1996; Murray et al., 1996; Harlow et al., 1996; I. Colina, 1998; unrest spanning lava dome extrusion that started on November 21, 1998, transitions to explosive eruption on February 10, 1999. Data compiled from Zobin et al., 2002a and Zobin et al., 2002b; Taran et al., 2002; Englebrecht, 2009. g. Asama, 2007–2011: unrest spanning series of small explosions in 2008–2009. Data compiled from Japan Meteorological Agency, multyear earthquake catalog; Kazahaya et al., 2011; Ohwada et al., 2013; and Japan Meteorological Agency-Volcanological Society of Japan, 2013. h. Miyake-Jima, 2004–2005: unrest spanning series of small explosions in 2004–2005, following a major off-island dike intrusion and caldera collapse in 2000. Data compiled from Japan Meteorological Agency, multyear earthquake catalog; Kazahaya et al., 2004; and Japan Meteorological Agency-Volcanological Society of Japan, 2013.
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 volcano monitoring data, especially, patterns that have escaped prior notice?

Is magmatic intrusion causing current unrest? If yes, how likely is it to reach the surface, and how soon?

Pre-designed SQL queries will be developed for common questions; users will develop their own queries for other questions.

For many years, seismologists have been contributing software applications to community clearinghouses, starting with the IASPEI’s Toolbox for Seismic Data Acquisition, Processing, and Analysis (Lee, 1989) and continuing with other packages today. In the volcanological community, a similar community spirit is evident in software applications contributed to VHUB (Palma et al., 2014). We invite WOVOdat users to contribute additional applications that others might also use.

8. Who is and will potentially be using WOVOdat?

The variety of applications of WOVOdat underscores a variety of potential users. These would include:

- Researchers seeking to understand magmatic ascent, vesiculation, degassing and other pre-eruption processes and, conversely, identifying what unrest might be expected and taken as evidence for specified subvolcanic processes. Just as epidemiological databases help medical researchers to identify factors in the spatial and temporal patterns of disease, WOVOdat can help volcanologists to interpret unrest patterns in subvolcanic and volcano-tectonic processes.
temporal distribution of diseases, WOVOdat is helping volcanologists to discover new relationships between different variables, e.g., between gas fluxes and seismicity, size of the unrest and size of the eruption, and apparent influence of regional tectonic earthquakes on local volcanic unrest.

- Volcanologists who must forecast the outcome of unrest and need to know of similar unrest elsewhere, its causes and outcomes. Boolean searches can be made for other examples of puzzling or worrisome unrest. Probabilistic event trees (c.f. Newhall and Hoblitt, 2002) and short-term, time-dependent risk analysis often require data from analogues to the restless volcano, and WOVOdat will soon be an easily accessible, authoritative source for these data. The database will be especially important when unrest begins at a volcano that has been dormant for a significant amount of time, or that erupts every half a century or longer. For these cases without previous background information, comparison to unrest and outcomes at analogous volcanoes can serve as a proxy of the past at the volcano in question. Even at volcanoes that erupt every few years and for which there is an existing monitoring database, it is useful to have access to a global unrest database since volcanoes can quickly change their behavior from one eruption to the next, as happened in Merapi 2010 and Kelud 2007. WOVOdat will let users trace, in retrospect, changes in monitored parameters from the start of unrest to any eruption(s) that might occur, and return back to normal.

- Researchers outside the geosciences, e.g. those improving techniques for data mining or studying the general phenomenon of self-organized criticality.

- 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. Some of those inspired may become the next generation of volcanologists. WOVOdat will provide packaged data sets and key questions for some classic, well-monitored eruptions, and

- The general public, especially those at volcanic risk and those wanting to delve personally into data of volcanoes.

9. Are comparisons between volcanoes meaningful? A caution about reference to analogues

Like people, each volcano is a little different from the next, but we also recognize enough similarities that generalizations about people and their behavior can be made. Reference to similar or analogous volcanoes during crises or in research requires the assumption that similarities in properties and processes are sufficiently strong that information from the analogues can inform study of the new subject volcano, and help in forecasting outcomes of unrest (e.g. Sheldrake, 2014). Volcanoes are physical systems that are ultimately governed by physical parameters and laws and thus there is an underlying commonality of parameters and processes between volcanoes that WOVOdat captures well.

One needs only to look at classifications of volcanoes, or of eruptive activity, to see that there are far fewer classes of volcanoes and eruptions than there are volcanoes. We generalize all the time. We expect Kilauea, Mauna Loa, and other basaltic ocean island volcanoes to show similarities in products and processes. We expect Mount St. Helens, Bezymianny, Soufrière Hills, Unzen, and other volcanoes that grow viscous domes to show similarities in products and processes. The question is not whether there are similarities, but whether those similarities are sufficiently strong to use experience from one to guide interpretation of another, and whether one’s classification or grouping of analogues captures the essential similarity.

Everyone who uses sets of analogue volcanoes in probabilistic eruption forecasting or hazard assessment chooses a set of analogues based on observed or inferred similarity to the volcano in question. The USGS’ Volcano Disaster Assistance Program routinely chooses sets of analogue volcanoes when estimating a priori probabilities at restless volcanoes, especially at volcanoes that are otherwise poorly known (Pallister, 2015; Newhall and Pallister, 2015). But there is, as noted by Cashman and Biggs (2014), uncertainty and the risk of circular reasoning in doing so.

Rodado et al. (2011) also suggested for statistical reasons that qualitative geologic information be used in addition to mean onset rates, and a subsequent paper by Bebbington (2014) about long-range
forecastability of explosivity gave further support to use of analogues. Prior to grouping of volcanoes into those with open- and closed-conduit degassing behavior, Bebbington (2014) found no volume predictability, but later found robust volume predictability among volcanoes with conduits more or less closed to degassing.

Whelley et al. (2015) classified volcanoes of SE Asia for the purposes of approximating their potential to produce ash. Their classification is a hybrid of descriptive and interpretative, considering geomorphology, known eruptive frequency, and known degassing (all descriptive), and, as in Bebbington (2014), inferences about how efficiently incoming gas is either trapped or bled off. Stratovolcanoes of the region were classed as plugged, semi-plugged, or open-vent. Shield volcanoes and calderas were classed separately, but Acocella et al. (2015) extended the concept of plugged, semi-plugged, and open to calderas as well. Though we made no mathematical test yet of how this improves ultimate results, we can say that in both of the last-mentioned cases, grouping of analogous volcanoes helped build logically consistent, coherent stories.

The critical rule to bear in mind when referring to analogues is that no analogue will be exactly the same as the volcano of current interest. It would be unwise to ever base a forecast solely on a supposed analogy. What one can do is to glean ideas from the analogues about the range of processes and activity that might follow at your volcano of interest, and to make working hypotheses to be tested with further data from your own volcano. This was the thought process behind the successful forecast of the Boxing Day collapse at Soufrière Hills Volcano (B. Voight, cited in Belousov et al., 2007).

10. Fair use of WOVOdat

Contributions of data to WOVOdat may be strongly influenced by how we use and credit the data. Those who collect data at no small personal risk have a right to examine and publish on those data first, before opening them to the world. This is the rationale for the 2-year grace period. Because WOVOdat is intended for comparisons and surveys of unrest across many volcanoes, we encourage those who are interested in a single volcano to seek data directly from the relevant volcano observatory or data owner. At the same time, there are growing requirements for complete openness of data sets funded by EU, NSF, and other national/international funding agencies, and a growing number of volcanologists see more value in early release of data than in holding back. Over the years that we have been developing WOVOdat, we have been delighted witnesses to cultural change in volcanology, toward more open data sharing.

11. Contributing to WOVOdat

The WOVOdat project actively seeks and welcomes contributions of data from all interested parties. Although most of the contributions come from WOVO observatories, we recognize that there are many valuable data sets that have been collected by university groups and research consortia like IRIS, UNAVCO, and EPOS. WOVOdat can only be an authoritative source if we can capture data from all sources. This is a community effort, of and for the volcanological community. Those who contribute don’t get exclusive access, but do get the satisfaction of building this community resource and of showing that their data are included and accessible to the whole community. The WOVOdat project accepts data in several formats. Our preference is in WOVOuml format (http://wovodat.org/doc/index.php), or in de facto community standard formats for which we have already created conversion scripts. Other options are to submit CSV files that will be converted by our online interactive tools into WOVOuml, or to enter small volumes of data manually into webforms. The last option is to submit files in their original observatory data format and help the WOVOdat team to understand the data and convert it to WOVOuml. Please refer to our documentation (WOVOdat 1.1) for the exact names of the fields and other details of required formats.

Software applications that will help WOVOdat users are equally welcome.

12. Challenges of building and sustaining WOVOdat

We would like to make WOVOdat a living, enduring tool for the community. After the initial population phase, it will be maintained and grow as the two-year window of latency advances.

For the immediate future, the Earth Observatory of Singapore will continue partial support if matching contributions can be made (in cash or in kind) by others. To make in-kind contributions easier, we’ll take advantage of the internet to have a distributed work model. We thank all contributors to this distributed model, and invite new observers and volunteers to join. Automated, periodic data feeds from observatories into WOVOdat will greatly reduce the time required to keep it up to date and growing.

It is not easy to fund a volcano database and even harder to keep it funded. Although the Earth Observatory of Singapore is continuing support for WOVOdat, it needs to see tangible evidence of uptake within the volcano community. The best way you can show your support for WOVOdat is to contribute your time, data and/or software modules. Those who have more time are invited to spend time with us in Singapore or one of the other WOVOdat hubs to join in the effort. Contact us at wovodat@wovodat.org if you like the concept of WOVOdat and can help!

Acknowledgments

We offer hearty thanks to the USGS/USAID Volcano Disaster Assistance Program (C. Dan Miller, J. Pallister) for seed funding to get WOVOdat going, and to the Earth Observatory of Singapore (K. Sieh, Director) for major funding to turn the concept into reality. We also thank WOVOdat participants who have been “launch customers” by contributing early data sets and by helping to beta test WOVOdat utilities, and the Smithsonian’s GVP for being a partner in this project.

We also thank early WOVOdat staff members aside from those listed as co-authors: Alexander Baguet, Chin Mei Liu, and Li Dong Chen. Many other colleagues have helped along the way. Special thanks to Jacopo Selva, Warner Marzocchi, Paolo Papale, Giuseppe Puglisi and Danilo Reitano (INGV); Florian Schwandner (ETH, NTU and currently at JPL); Etsuake Fujita and Hideki Ueda (NIED), PHIVOLCS-volcano monitoring division; CVGHM-volcano monitoring division; Noritake Nishide, Satoshi Harada, Hiroshi Yamamoto, Takeshi Koizumi, Yohko Igarashi, Koji Kato, Yoshiaki Fujiwara (JMA); Nico Fournier, Greit Jolly (GNS); Shinji Takarada and Joel Bandihas (GS); Freystein Sigmundsson (Univ. of Iceland) and Kristin Vogfjord (IMO), Simon Carn and Verity Flower (MTU), Matt Pritchard (Cornell), Juliet Biggs, Steve Sparks and Jo Gottsmann (Univ of Bristol), Steve Malone (Univ of Washington), Steve McNutt (Univ of Alaska, now at Univ. South Florida); Tom Simkin, Jim Luhr, Lee Siebert, Paul Kimberly, Ed Venzke, Liz Cottrell, and Ben Andrews (Smithsonian); and Chuck Mertens (UNAVCO).

Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jvolgeores.2017.08.003.

References

Acocella, V., Di Lorenzo, R., Newhall, C., Scandone, R., 2015. An overview of recent (1988 to 2014) caldera unrest: knowledge and perspectives. Rev. Geophys. 53, 896–955.
Aspinall, W.P., Woo, G., Voight, B., Baxter, P.J., 2002. Evidence-based volcanology: application to volcanic crises. J. Volcanol. Geotherm. Res. 128, 273–285.
Barberi, F., Bertagnini, A., Landi, P., Principe, C., 1992. A review on phreatic eruptions and their precursors. J. Volcanol. Geotherm. Res. 52, 231–246.
