OPENING RISK ANALYSIS

C. Scawthorn[1]

[1] Professor (ret.), Dept. Urban Management, Kyoto University, Kyoto Japan
President, SPA Risk LLC, San Francisco CA 94119 USA
cscawthorn@sparisk.com

ABSTRACT :
Open Risk Analysis (ORA) refers to risk analysis based on open methods, data and tools. A prime focus are the tools, especially open source software. In 2007 the Alliance for Global Open Risk Analysis (AGORA, www.risk-agora.org) was founded to support the development of ORA, and is sponsoring the development of OpenRisk, an open source generalized risk analysis software code, and a Natural Emergency Experience Database (NEED). Lastly, in order to better understand and promote ORA, this paper proposes an Open Risk Rating System (ORRS).

KEYWORDS:
Risk, open source, software, data, analysis, rating

1. INTRODUCTION

While natural hazards risk analysis is nearly as old as humankind, the development of formal risk analysis only emerged in the latter half of the 20th century although the need and methods were foreseen and discussed earlier (Scawthorn 2007) The fire insurance industry had lead the way in accumulating detailed accurate exposure data (eg, Sanborn maps) as well as loss data since the Civil War. Its methods for risk analysis were relatively primitive however, with the primary measure being the ill-defined PML (Probable Maximum Loss). Regarding earthquakes, during the 1960s Cornell, Davenport and Esteva developed formal methods for analyzing seismic hazard, as did Algermissen, who then collaborated with Steinbrugge in addressing seismic risk. In the 1970s at MIT the SDDA (Seismic Design Decision Analysis) project, under the leadership of Whitman was developing a fully probabilistic methodology for integrating vulnerability over hazard, accounting for uncertainty. This methodology, in varying forms, is the standard model used today. Because of the data and computational demands for hazard and risk analyses, computer software emerged as an integral and necessary part of seismic risk assessment. Friedman was probably the first to use computers to calculate natural hazards losses (for wind and flood, Friedman, 1966), In the 70s Friedman, the Wiggins Company (1976), Algermissen (1978), Scawthorn (1979) and others had developed software for natural hazards loss estimation purposes. By the early 1980s Steinbrugge had a proprietary computer program for use by insurance companies (Steinbrugge, personal communication). In the 1980s, three loss-modeling companies emerged (Applied Insurance Research, EQE International, and Risk Management Solutions, referred to as *the modelers*) to serve the global insurance industry. The three companies have developed into sizable corporations with global capabilities, providing proprietary modeling software that primarily serves the finance, insurance and real estate (F.I.R.E.) industries. The modelers tend to have sophisticated treatment of probability and uncertainty (we term this *actuarial-quality* modeling), and generally have models of varying detail for earthquake or tropical cyclone modeling in the US, Canada, Japan, China, Mexico, parts of Europe and Latin American, New Zealand and Australia. On a selected basis, they have also developed flood, hailstorm, wildland fire, terrorism and other models. Smaller modeling firms include Risk Engineering, Applied Research Associates (ARA), Risk Frontiers and SPA Risk LLC.

While the insurance industry is a major user of risk analysis, the public sector also has a major need to estimate the potential for future losses, for planning and development needs. To serve these needs, in the early 1990s two seismic risk analysis programs emerged *EPEDAT* (Early Post-earthquake Damage Assessment Tool) and *HAZUS* (HAZards United States). The EPEDAT development was supported by California OES (Eguchi et al,

1997) but appears to have fallen into desuetude with the further development of HAZUS. HAZUS currently is the only free software with the ability to perform end-to-end multihazard catastrophe loss modeling, and permits relatively quick and easy estimates of potential seismic, flood and wind losses, including social and indirect economic impacts, for any community in the U.S. The estimates can be modified for various mitigation strategies, and the results can be used for mitigation and emergency response planning. The results are clearly communicated in map and tabular formats. HAZUS is gradually winning adoption within state and local governments for planning and emergency response purposes, and is a substantial contribution to national earthquake hazards reduction - see (Scawthorn et al. 2006) for more detail. Several other programs similar in some ways to HAZUS have been developed or are being developed:

- *CATS*: Developed for the Defense Threat Reduction Agency (DTRA) and the Federal Emergency Management Agency (FEMA) by SAIC, CATS is available to Federal, State, and local government emergency response organizations nationwide, and includes models for data fusion analysis that supports disaster and other emergency management. It also treats some technological hazards. (http://cats.saic.com/cats/models/cats_earthquake.html)
- *EXTREMUM* was developed for the Russian Ministry of Extreme Situations. It is GIS-based, has a global default inventory on various scales, seismic hazard/seismic zoning and microzoning maps, elements at risk/population, buildings and structures, lifeline systems, hazardous facilities and algorithms for combining the parameters of mathematical models for estimation of population distribution, buildings' damage and damage distribution, human casualty, etc. A version of *EXTREMUM* is *QUAKELOSS* (www.wapmerr.org/publications.html).
- *NERIES* is a recent EU project that will seek to promote improved access to distributed databases, common protocols, standardized procedures and strategies for long-term archiving and distribution of seismological data; and develop a new generation of hazard and risk assessment tools designed to improve monitoring and understanding of the earthquake process. One of the products is intended to be a HAZUS-like software tool. (http://www.orfeus-eu.org/neries/NERIES-proposal.htm).
- *PAGER* (Prompt Assessment of Global Earthquakes for Response) is an automatic system to estimate human impact following significant earthquakes, in development at the USGS (Golden CO). In a sense, it is a global extension of the successful ShakeMaps from hazard to risk. (http://earthquake.usgs.gov/eqcenter/pager/).
- *ZEUS-NL* has been developed by the Mid Americas Earthquake Center (MAE) and is a free 3D static and dynamic structural analysis platform for earthquake engineering applications, as well as *MAEviz*, a Consequence-Based Risk Management methodology using a visual, menu-driven system to generate damage estimates from scientific and engineering principles and data, test multiple mitigation strategies, and support modeling efforts to estimate higher level impacts of earthquake hazards, such as impacts on transportation networks, social, or economic systems. It enables policy-makers and decision-makers to ultimately develop risk reduction strategies and implement mitigation actions. (mae.ce.uiuc.edu/software_and_tools/zeus_nl.html).

In other developments, the UN IDNDR several years ago funded the RADIUS (Risk Assessment Tools for Diagnosis of Urban Areas Against Seismic Disasters, see http://www.geohaz.org/radius/RADIUSIntro.htm) project, a spreadsheet-based earthquake loss estimation tool. It suffers from a distinctly inferior technology to HAZUS, and distribution limitations, despite its claim to be freely available.

Concurrent with all these development is risk analysis and risk software has been the emergence of the open source movement. Open source is a development methodology which offers practical accessibility to a product's source (goods and knowledge). The open source model allows concurrent input of different agendas, approaches and priorities, and differs from the more closed, centralized models of development – these two modes of development are respectively referred to as the *Bazaar* and the *Cathedral* (Raymond 1999). The rise of open-source culture in the late 20th century resulted from a growing tension between creative practices that involve appropriation, and therefore require access to content that is often copyrighted, and increasingly restrictive intellectual property laws and policies governing access to copyrighted content. The principles and practices are commonly applied to the development of source code for software that is made available for public collaboration, usually released as open-source software.

In summary, a number of valuable risk-modeling tools currently exist, but appear to be ill suited for research to examine the societal impacts of new knowledge or alternative approaches in earth science, geotechnical and structural engineering, and disaster social sciences. Many of the models follow an empirical approach, although Performance Based technology is increasingly being adopted. The most advanced models are proprietary and confidential, are focused on the F.I.R.E. industries, and only available at significant cost. While HAZUS is distributed free of charge, its adoption has been modest to date, for various reasons and despite excellent documentation of its methodologies. HAZUS has required a significant investment on the part of FEMA, and major changes to the platform, such as to permit incorporation of physics-based modeling of ground motion, are likely to be painfully slow. Currently, there is no risk analysis software freely available and readily adaptable to advances in technology.

This paper introduces the Conference session on, and the concepts of Open Risk Analysis, which is the combination of open methods of risk analysis, open-source software using those methods, and open data for use in that software, towards to goal of transparent widely available estimation of multihazard risk. We briefly discuss the Alliance for Global Open Risk Analysis (AGORA), the Open Risk software code, and a rating system for quantifying the degree of 'open-ness' of any open risk analysis.

2. ALLIANCE FOR GLOBAL OPEN RISK ANALYSIS (AGORA)

In response to the needs discussed above, the Alliance for Global Open Risk Analysis (AGORA) was founded in Pasadena in 2007. AGORA (www.risk-agora.org) supports ORA through promotion of the basic tenets (see Box 1), facilitating communication among ORA practitioners, providing a forum for development and exchange, and sponsoring the development of OpenRisk (discussed below). The reader is referred to the AGORA website for further information.

Box 1. The AGORA framework

Whereas, risk analysis is the foundation of risk reduction, and
Whereas, we all share natural and technological hazards risk, and
Whereas, risk analysis is most broadly accepted and used when

- Risk methodologies are open
- Risk software is open source, and
- Fundamental risk data (such as hazard and vulnerability) are open,

Therefore, we support

- Open risk analysis
- Open risk methodologies
- Open risk tools, including open-source software
- Open risk data (such as hazard and vulnerability)
- Shared and collaborative development of methods, tools and data, and
- Open communication of risk results

While respecting individual initiative and the individual's right to the fruits of that initiative.

3. OPEN RISK

OpenRisk is an effort to develop object-oriented, web- and GUI-enabled, open-source, and freely available software code for conducting multihazard risk analysis. The resulting body of code and applications is also referred to as OpenRisk. OpenRisk's goal is to provide a framework where any arbitrarily complex model of hazard, vulnerability, and loss can "plug in" for analysis without having to change what's being plugged into. Hazard refers to the effects of natural and other phenomena (e.g., earthquakes, tropical cyclones) that can lead to damaging impacts on the built and social environments. By vulnerability is meant the relationship between these effects and loss through damage to the built and social environment. An example of a vulnerability function would be repair cost for a particular category of buildings, as a function of damped elastic spectral acceleration response. Loss refers to the undesirable financial and social impacts resulting from damage. Examples of loss are repair costs, casualties, and loss of use (i.e., dollars, deaths, and downtime). OpenRisk applications may be stand-alone or may be distributed over the Internet in what the OpenSHA developers refer to as a community modeling environment, all tied together with a user-friendly web interface.The purpose of developing OpenRisk is to provide researchers with tools to perform and incrementally improve quantitative study in any of the domains of risk modeling, and to enable them to explore the risk implications of new knowledge, without requiring universal mastery of all the relevant scientific, engineering, and social-science disciplines. Toward that end, OpenRisk is intended to comprise open-source software using open-source software licenses such as those distributed by the Open Source Initiative (http://www.opensource.org/), and participation from anyone in this development is welcome. For more details, see (Porter 2008).

4. NATURAL EMERGENCY EXPERIENCE DATABASE (NEED)

An additional problem for development of a risk analysisis the loss of valuable data on natural and technological disasters, collected by disparate investigators and organizations. Experience shows these data are typically lost after a very short period, because of the lack of a central, standardized archive. The author has proposed development of a *Natural Emergency Experience Database (NEED)* to address this need.

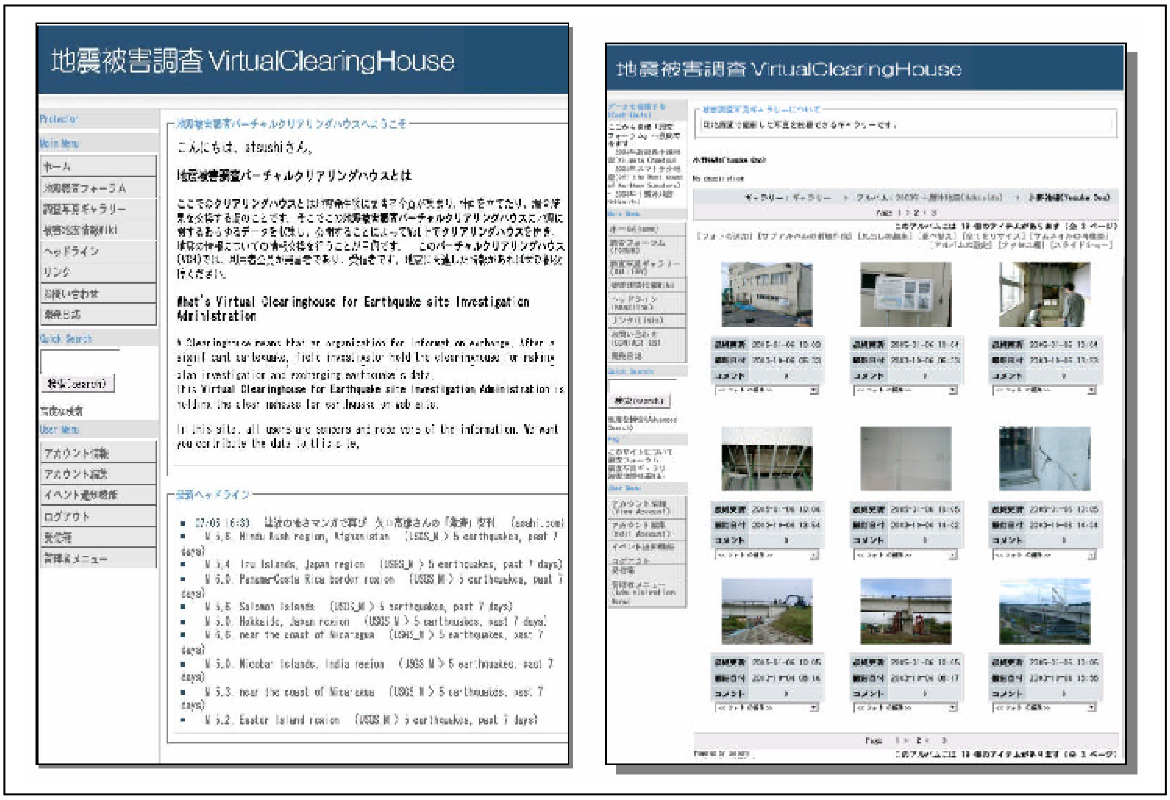

Figure 1 Screen shot of prototype Virtual Clearinghouse

The overall objective is the design, development, prototyping, and pilot testing of a mechanism to preserve and disseminate emergency experience data, that is, data about natural and man-made disasters. There is a wide variety of data related to disasters, ranging across the natural, technological, built and social environments. Perhaps the most important class of these data—and the focus of this proposal—is the detailed information from data-gathering in the field: the raw data that are typically used as a basis for drawing general conclusions, but that are too voluminous to publish conventionally. In the example of earthquakes, these perishable raw data have historically included earth-science information; observations of the structural and nonstructural performance of building, bridges, and other facilities; and near- and long-term data on the economic and human consequences and response. Analogs exist for all other natural and man-made disasters.

Experience has shown that raw and first-reduction data usually become lost or otherwise unavailable in the years following the emergency. The conclusions derived from the data are often published in durable form: journal articles, reports, etc. However, the conclusions are often presented in a form that precludes analysis by others, or comparison with other studies. There are two principal reasons for this: researchers develop new, nonstandard terms to explain the data; and the published summary information does not address aspects of the underlying data that are only later recognized as important. In either case, because the underlying data are lost or unavailable, lessons learned by different researchers cannot be compared, and studies that ask new questions must always gather new data or make questionable assumptions about the old, now missing data.

Historically, the loss of the raw data has been due to its volume and, typically, hardcopy format. With the digital revolution, however, the opportunity exists to archive inexpensively much of the raw data, which is often field-recorded digitally, via PDAs, digital voice-recorders, digital video-taping, GPS, GIS, etc. The loss of these valuable, and now readily storable and retrievable, data from recent and future emergencies can be prevented by creating a national, public, durable, electronic archive for permanent storage and dissemination. More than an online database of past emergency information, the archive (i.e., the *Natural Emergency Experience Database, NEED*) can be the locus for dissemination of data-collection protocols and software, as well as the IT backbone of a system for propagating common definitions and goals for future data gathering, both in the field and in the laboratory.

NEED is not a site to store journal articles online; such resources already exist. Rather, it is a place to store the raw data that underlie emergency research, for example databases containing every response to every question on a survey of human responses to a disaster, or the characteristics and digital images of every building examined in a safety survey. These are the data that are analyzed to produce publishable findings, but that are not themselves typically published. The concept for such an archive emerged in the earthquake community, and has been called for by NEHRP in its *Plan to Coordinate NEHRP Post-Earthquake Investigation*s (USGS 2003) and by participants at the Earthquake Engineering Research Institute's (EERI) Invitational Workshop: An Action Plan to Develop Earthquake Damage and Loss Data Protocols (2002).

The problem is fundamentally that of the inefficient utilization of information in the mitigation of disasters. The current process of information flow is shown in

Figure 1. Basically, prior to, during, and after disasters, individual researchers and organizations gather data on the disaster. The data may be depth of flooding, number of earthquake-collapsed buildings, hurricane-caused sheltered persons, or anthrax-related hospital admissions. In some cases, the actual data may have constraints on its use, such as personal privacy issues. When the researcher has collected the data, it is used in some manner: in response organizations, to improve their response; for academics, to create new knowledge, etc. In some manner, the data are usually reduced, and a summary of the data and the researcher's findings are published in a report, refereed journal etc. (the article in

Figure 1). Other researchers, or users themselves, synthesize more application-oriented findings from one or more articles, which are then translated into information for users. The process is similar, whether it is social scientists researching crowd behavior and ultimately improving emergency messages, or structural engineers researching beam-column connections and ultimately improving building codes. The inefficiency arises in the lessons being founded on only one or a few datasets. Scatter in the data, lack of broad sampling, and assumptions about the (inaccessible) underlying data, result in very fuzzy knowledge and large inefficiencies.

The proposed solution is shown in

Figure 2. It consists of a voluntary pooling of data in NEED. Data contributions are open to all disciplines and all natural and technological disasters, so that cross-cutting behavior and performance, whether of crowds or of steel beams under inelastic deformations, can be extracted, compared, synthesized, and more efficiently used. Larger datasets enhance the signal-to-noise ratio and reduce bias, thus enhancing efficiency. Data gaps are more easily identified: if the data do not reside in NEED, they need to be gathered. Errors are more easily identified: if a particular dataset has a significantly different trend than others, the question arises as to *why*. In some cases, the difference in trends may be spurious because of differing terminology or initial conditions, but in others, important errors may be found.

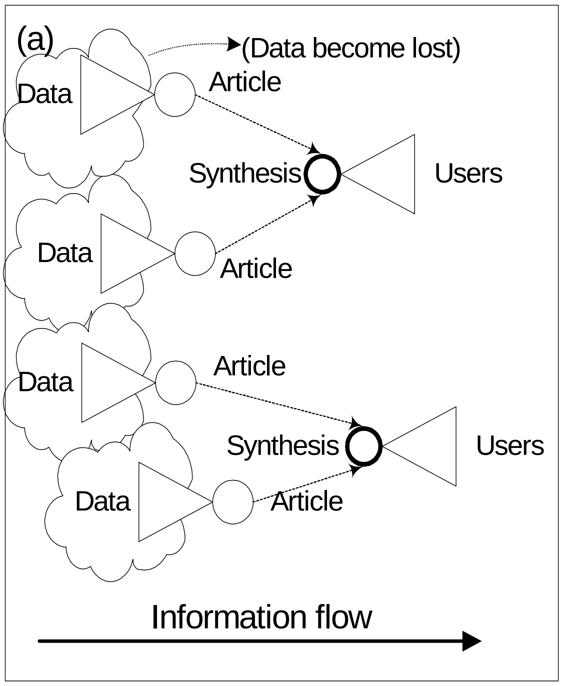

Figure 1 Information flow without NEED

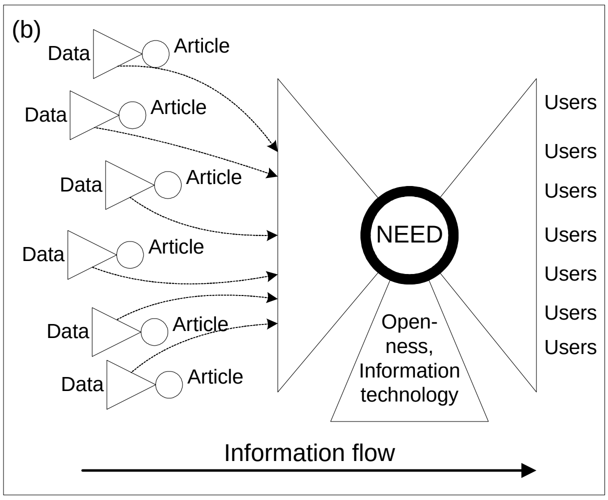

Figure 2 Information flow with NEED

5. RATING OPEN-NESS

"*What gets measured, gets managed*" is a management maxim – in order to better define what 'open-ness' is, we propose an Open Risk Rating System (ORRS), which quantifies an Open Risk Score (ORS) as follows:

$$ORS = (OMS + ODS + OTS) / 3 \tag{5.1}$$

where

$$OMS = \text{Open Methods Score} = (MDOC + MPROT + MREPRO) / 3 \tag{5.2}$$

where

MDOC = open-ness of documentation, scored from 0 (no documentation of methods available) to 10 (full documentation of all methods and references widely available, no gray literature)

MPROT = Methods Protection, scored from 0 (methods fully proprietary and not available for use by others) to 10 (methods in public domain)

M = Reproducible by others, scored from 0 (methods and results cannot be reproduced by others) to 10 (methods and results fully reproducible by others, including availability of test cases and data)

ODS = Open Data Score = (ODATA + DPROT + DREPRO) / 3 (5.3)

where

ODATA = open-ness of underlying data, scored from 0 (data not available) to 10 (data fully available). In this data refers to hazard and vulnerability data, and asset data in public studies – private asset data is acknowledged to be private and protectable.

DPROT = Data Protection, scored from 0 (data fully proprietary and not available for use by others) to 10 (data in public domain)

DREPRO = Reproducible by others, scored from 0 (data cannot be reproduced by others) to 10 (data fully reproducible by others)

OTS = Open Tools Score = (OT + TPROT + TREPRO) / 3 (5.4)

where

OT = open-ness of tools, especially computer code, scored from 0 (tools and/or computer code not openly available) to 10 (full availability of computer source code, for example via licenses such as GPL or LGPL)

TPROT = Tools Protection, scored from 0 (tools and/or computer code fully proprietary and not available for use by others) to 10 (tools in public domain)

TREPRO = Tools products (e.g., results) reproducible by others, scored from 0 (tool products cannot be reproduced by others) to 10 (tool products fully reproducible by others, including availability of test cases and data)

For example, using this system, HAZUS and one of the modeler's proprietary risk analysis computer codes might score as shown in Table 1.

Table 1 Example Open Risk Rating System

		HAZUS-MH (EQ)	**modeler**	
METHODS	MDOC	8	2	HAZUS methods well documented, but some key references unavailabile. Modelers documentation available only to licensees, and even then limited.
	MPROT	8	0	All HAZUS methods available but Copyrighted;
	MREPRO	8	0	
	OMS	**8**	**0.67**	
DATA	ODATA	10	0	HAZUS very good in furnishing much data; other data from USGS etc.
	DPROT	9	0	All HAZUS data copyrighted;
	DREPRO	8	0	
	ODS	**9**	**0**	
TOOLS	OT	5	0	HAZUS software available at no cost, but source code not available

	TPROT	8	3	HAZUS-MH code available but Copyrighted.; modeler code not available except to licensees.
	TREPRO	6	0	
	OTS	**6.33**	**1**	
Open Rating Score	**ORS**	**7.78**	**0.56**	

6. CONCLUDING REMARKS

Risk Analysis is a vital tool for reducing natural hazards risk, and should be available to all. Currently it is not, not so much due to 'greedy' commercial interests as to lack of transparency and lack of access to data and tools. . To improve this situation, a number of risk analysis practitioners from around the world have formed the Alliance for Global Risk Analysis, and are in the process of developing OpenRisk, NEED, ORRS and other methods, data and tools to provide the benefits of open risk analysis to all. The Open Risk Rating System will be featured on the AGORA website and hopefully will become a global standard for promoting ORA.

REFERENCES

Porter, K. "OpenRisk: new applications of open-source risk estimation software." *14th World Conference on Earthquake Engineering*, Beijing.

Raymond, E. S. (1999). *The Cathedral & the Bazaar*, O'Reilly.

Scawthorn, C. (2007). "A Brief History of Seismic Risk Assessment." Risk Assessment, Modeling and Decision Support. Strategic Directions, A. Bostrom, S. P. French, and S. J. Gottlieb, eds., Springer-Verlag, Heidelberg New York,Tokyo

Scawthorn, C., Schneider, P. J., and Schauer, B. A. (2006). " Multihazards Loss Estimation and HAZUS." *Natural Hazards Review, Special Issue*,, 7(2), 39-103pp.

USGS. (2003). "The Plan to Coordinate NEHRP Post-Earthquake Investigations,." USGS Circular 1242, Thomas L. Holzer (Chair), Roger D. Borcherdt, Craig D. Comartin, Robert D. Hanson, Charles R. Scawthorn, Kathleen Tierney, T. Leslie Youd,, ed., prepared in coordination with the Federal Emergency Management Agency, National Science Foundation, and National Institute of Standards and Technology, USGS Western Region, Menlo Park CA.