AmericanLifelinesAlliance

A public-private partnership to reduce risk to utility and transportation systems from natural hazards and manmade threats

American Lifelines Alliance
Post-Earthquake Information Systems (PIMS)
Scoping Study

September 2008

FEMA

nehrp

National Institute of BUILDING SCIENCES

American Lifelines Alliance

The American Lifelines Alliance (ALA) is a public-private partnership project initially funded by the Federal Emergency Management Agency (FEMA) of the Department of Homeland Security (DHS) and managed by the Multihazard Mitigation Council (MMC) of the National Institute of Building Sciences (NIBS). The ALA's goal is to reduce risks to lifelines – the essential utility and transportation systems that serve communities across all jurisdictions and locales – from all hazards. To do so, it facilitates the development, dissemination, and implementation of planning, design, construction, rehabilitation, and risk-management guidance and encourages use of this information to improve the performance and reliability of new and existing critical infrastructure. The ALA's key stakeholders are lifeline operators and the communities they serve, standards development organizations, and engineering and risk-management professionals. The ALA provides a forum to address current industry and community needs and crafts unique partnerships to work across lifelines systems. ALA products either are incorporated in national consensus standards documents or are disseminated to key industry stakeholders through relevant associations and industry publications.

ALA Principal Investigator

Douglas G. Honegger, D. G. Honegger Consulting, Arroyo Grande, California

Members Participating in PIMS Project

Stuart Nishenko, Pacific Gas and Electric Company, San Francisco, California
William Savage, U.S. Geological Survey, Menlo Park, California
Stuart Werner, Seismic Systems & Engineering Consultants, Oakland, California

Federal Agency Liaison Members

Edward M. Laatsch, PE, Multihazard Engineering Services Section, Federal Emergency Management Agency
John "Jack" R. Hayes, Jr., PhD, PE, National Earthquake Hazards Reduction Program, National Institute of Standards and Technology
John R. Filson, Scientist Emeritus, U.S. Geological Survey

National Institute of Building Sciences Staff

Claret Heider, NIBS Vice President for MMC Programs
Carita Tanner, Communications Director

American Lifelines Alliance
Post-Earthquake Information Systems (PIMS) Scoping Study

September 2008

National Institute of Building Sciences
1090 Vermont Ave., NW
Washington, DC 20005

This report was prepared under contract EMW-2003-CO-0417 between the Federal Emergency Management Agency (FEMA) of the Department of Homeland Security (DHS) and the National Institute of Building Sciences (NIBS) through its Multihazard Mitigation Council (MMC).

For further information on ALA activities and products, write the ALA at the Multihazard Mitigation Council of the National Institute of Building Sciences, 1090 Vermont Avenue NW, Suite 700, Washington, DC 20005 or visit the ALA website at: www.AmericanLifelinesAlliance.org

Acknowledgements

This report was prepared for the NIBS/MMC American Lifelines Alliance (ALA) by the PIMS Project Team at the University of Illinois at Urbana-Champaign (UIUC). The ALA is grateful for the efforts and expertise of Bill Spencer, who served as the UIUC team's principal investigator, and for those of team members Jim Myers, Blake Andrews, and Bernie Acs. This project would not have been possible without the work of these talented individuals.

The NIBS Multihazard Mitigation Council (NIBS/MMC) is also especially grateful to ALA Project Team Principal Investigator Douglas G. Honegger for his contribution of time and effort to this project and to ALA Project Team members Stuart Nishenko of Pacific Gas and Electric Company, William "Woody" Savage of the U.S. Geological Survey, and Stuart Werner of Seismic Systems and Engineering Consultants. The NIBS Multihazard Mitigation Council wishes to thank Edward Laatsch of the Federal Emergency Management Agency for his efforts in conjunction with this PIMS project as well as those on behalf of ALA since its inception. Finally, NIBS/MMC is grateful to Jack Hayes of the National Institute of Standards and Technology and John Filson of the U.S. Geological Survey for their contributions in support of this PIMS project on behalf of the National Earthquake Hazards Reduction Program.

Table of Contents

Executive Summary

In December 2007, the American Lifelines Alliance, with funding from the Federal Emergency Management Agency (FEMA), tasked a team from the University of Illinois at Urbana-Champaign (UIUC) with performing a scoping study to assess both the infrastructure requirements (e.g., data system architecture, technological needs, and issues) and the implementation requirements (e.g., facilities, expertise, and funding) for establishing a national post-earthquake information management system (PIMS). The Strategic Plan for the National Earthquake Hazards Reduction Program has emphasized the need for a national PIMS for use in achieving national risk reduction and mitigation goals. This report presents the results of a 10-month study that involved both a broad effort to obtain stakeholder input and an in-depth examination of user needs and system requirements, challenges and system-level issues involved in establishing a PIMS, and the design strategy needed to overcome the challenges and satisfy user needs.

User Needs for PIMS

User needs and system requirements were identified through an information collection process that involved a review of existing documentation (see Section 2.2 for documentation reviewed) and interviews with stakeholders (see Appendix H for the list of stakeholders). The user needs and requirements identified can be categorized as follows:

- **User Interfaces —** The interfaces that users want to use to discover and retrieve data using a PIMS.
- **Information Needs —** The types of information that users want to obtain from PIMS including: general information; hazard data; building, bridge, and other lifeline performance data; critical structures performance data; historical data; loss/socio-economic data; pre-event inventory data; and PIMS system data.
- **Data Access, Privacy, and Security Issues —** Policies concerning the removal of personal information from data, the creation of aggregated data sets, restriction of access to certain types of data, and safeguarding the overall PIMS holdings.
- **Direct Ingestion of Data —** Procedures for directly uploading data into a PIMS using various means.
- **Harvesting and Exchanging Data —** Requirements that will enable a PIMS to harvest and exchange data with a variety of existing (and future) electronic databases.

System Requirements and System-Level Issues

In addition to serving the direct needs of users and other stakeholders, a PIMS must address their implicit assumptions about how the system should align with their goals. A PIMS also must address issues related to the cultural, political, technological, and organizational contexts in which it will operate. The system requirements and system-level issues identified relate to:

- Data collection, organization, and storage;

- Data curation and quality assurance;
- Information presentation, discovery, and retrieval;
- Privacy and security;
- Long-term data preservation;
- Data standardization;
- System evolution and change management;
- Coordination and data sharing with public, private, and governmental sources;
- Lack of information for certain types of structures, lifelines, or other entities; and
- PIMS community adoption.

System requirements are summarized in Tables 2.1 and 2.2.

Design Strategy

A recommended PIMS design strategy has been formulated to satisfy user needs and system requirements and to resolve system-level issues. It is based on a service-oriented system architecture (presented conceptually in Figure 3.1) with the following components:

- **A semantic content repository —** The central storage place for all PIMS data organized using a semantic content system that allows for automated processing and retrieval of data.
- **Workflow and provenance management systems —** Systems that organize computing tasks into simple workflows and manage data history tracking.
- **Data ingestion, curation, preservation, and discovery/export mechanisms —** Methods and technologies for ingesting data into the PIMS; verifying or improving data quality; maintaining data accessibility and quality over the life of the PIMS; and finding data efficiently and exporting it to user software programs.
- **User interfaces with analysis and visualization tools —** Methods and technology for user interfaces and data visualization.
- **Security subsystem —** Recommendations for role-based access control and credential management systems.
- **System administration structures —** Standards-based infrastructure to monitor system operations and interfaces to monitor and manage the system.

In addition to the system architecture, the design strategy is based on the following concepts:

- **Content management —** A semi-structured data model that allows efficient searching while retaining the flexibility to dynamically add new metadata and new types of data to the system.
- **Virtual organization-based components —** Components that use standard interfaces and dynamically discover resources and policies and other components in the system that allow independent evolution of system components (e.g., security mechanisms).

- **Virtual machine implementation —** Use of a software layer that makes real computers appear to applications and services as multiple independent virtual machines, which can improve utilization of resources, reduce management costs, and improve overall system reliability.

Other important components of the design strategy are the policies and procedures that work in parallel with the system architecture and design concepts. These relate to data standards and provenance, data privacy and security, persistence of data, and community adoption of PIMS.

To ensure that all user needs and system requirements have been met using the proposed design strategy, a Requirements Traceability Matrix has been constructed that compares required system abilities to provided system functionality. It is presented in Appendix G.

This report provides a detailed discussion of the development, operation, and oversight of PIMS. Potential phases of PIMS development are identified, and the goals, schedule/milestones, and required resources for each phase are given. Because successful PIMS implementation will depend on strong oversight and management, this report also outlines suggested oversight and management approaches, describes potential sources of funding for PIMS, lists key issues to be addressed as oversight and management functions, and discusses issues relating to the data collection component of PIMS.

Next Steps

Although implementing PIMS as a comprehensive persistent national resource and a means to coordinate community standardization efforts related to field reconnaissance will be a complex long-term undertaking that will require ongoing discussion, concrete next steps can be taken to develop and operate a core system and to incrementally define and implement more advanced functionality. This document is intended to provide both guidance on how such an effort could be structured and estimates for scope, schedule, and budget that potential funding agencies can use to realize a PIMS effort.

An initial build of PIMS would not possess full functionality, but it would provide the basic abilities to archive and retrieve data on a subset of infrastructure drawn from existing repositories and would serve as a basic national clearinghouse for data from future events. It is estimated that an investment of approximately $1.5 million over the course of 24 months would be needed to create the PIMS initial build. A viable core set of functionalities for such a system are identified and the key set of discussions and decisions that will be needed to drive further advancement of a PIMS are outlined. These discussions relate to a set of seven relatively independent pilot efforts that would include community discussion, development of concrete requirements and software designs, and implementation of new functionalities that could then be hardened and incorporated into the evolving base PIMS.

1 INTRODUCTION

1.1 Project Background

This report describes the results of a scoping study to establish a national post-earthquake information management system (PIMS). As envisioned in discussions with earthquake community stakeholders and as reflected in the NEHRP strategic plan, PIMS is the end-to-end system for accomplishing the national tasks of facilitating post-hazard data collection, archiving of data, and distribution and use of the data to improve protection against hazards. In this section, the background for the PIMS project is described, including past events leading up to this project, efforts similar to PIMS in scope and scale, existing sources of data for PIMS, a review of existing practices and procedures for PIMS, and a summary of organizations that may have a stake in PIMS.

It has long been recognized that any national effort to reduce economic losses and social disruption resulting from severe natural hazard events requires a mechanism to capture and preserve engineering, scientific, and social performance data in a comprehensive and coherent system that will contribute to our learning from each disaster. Such a resource can play a vital role in efforts to enhance infrastructure and building design and to optimize mitigation, disaster planning, and response and recovery activities. Despite this recognition, no mechanism is currently in place in the United States to ensure that relevant and often perishable data are systematically collected and archived for future use. Further, those data that are gathered often are lost relatively soon after they have been collected rather than being organized and maintained to enable study, analysis, and comparison with subsequent severe natural disaster events that may not occur for many years or even decades.

To stimulate a national discourse to improve collection, archiving, and distribution of data related to the performance of the built environment in natural disasters within the United States, the American Lifelines Alliance (ALA), which operates under the Multihazard Mitigation Council (MMC) of the National Institute of Building Sciences, held a Workshop on Unified Data Collection in Washington, D.C., in October 2006 [1]. The workshop, conducted with funding from the Federal Emergency Management Agency (FEMA), was motivated by the idea of a "Post-Earthquake Information Management System" (PIMS) that had been evolving at numerous places including FEMA and the National Earthquake Hazards Reduction Program (NEHRP), and it served as a forum for open and candid discussion of common needs of the utility and transportation systems (lifelines) community and possible opportunities for cooperation and collaboration in addressing those needs. The findings from the workshop contributed to the establishment of improved post-earthquake information acquisition and management as an objective and development of a national post-earthquake information management system as a strategic priority in the Strategic Plan for the National Earthquake Hazards Reduction Program: Fiscal Years 2008-2012 [2]. The workshop participants recognized that an integrated PIMS needed to include all aspects of the built environment and could potentially be expanded in scope to address all types of natural hazards.

The basic premise for the 2006 ALA workshop was outlined in a white paper distributed to representatives of the lifelines community invited to participate in focused working groups that addressed the following topics:

- Mechanisms and procedures for post-disaster data collection,
- Cooperative data collection between and within the public and private sectors,

- IT management (data archiving and exchange), and
- Long-term administration and maintenance of a data archive.

In preparation for the workshop, the ALA provided workshop leaders of the four working groups a list of issues to consider in the workshop discussions (see [3] for the list). During the workshop, members of the working groups discussed, validated, and added detail to these issues. These issues provide the basis for the challenges that will be involved in implementing PIMS.

As a follow-on to the workshop, the ALA, in consultation with FEMA and NEHRP, decided to proceed with the next step of developing the requirements for an integrated system that could successfully address the long-term informational needs of society to systematically and strategically improve the performance of buildings and lifelines in significant natural disaster events.

In December 2007, the ALA tasked a team from the University of Illinois at Urbana-Champaign (UIUC) with conducting the scoping study reported here to further guide PIMS planning and development efforts. Although the UIUC study focused on earthquake-related information, the findings are expected to be directly relevant to planning information management support addressing other types of natural hazards. The objectives of the study are to assess both the infrastructure requirements (e.g., data system architecture, technological needs, and issues) and the implementation requirements (e.g., facilities, expertise, and funding) for establishing PIMS. In other words, the project is to provide a "road map" or requirements document that:

- Delineates the user and functional requirements of PIMS,
- Outlines the steps that need to be taken to create the needed information management center,
- Estimates likely costs and levels of effort required for each step, and
- Provides possible development schedules with milestones.

A preliminary draft of this document was completed in July 2008 and presented to stakeholders at a PIMS Workshop held in July 2008 in Chicago, Illinois (see Appendix H for the list of stakeholders who attended the workshop). The purposes of the workshop were to present findings to date on various aspects of developing PIMS; to review the preliminary draft of the road map document; and to discuss how PIMS might be established, how it would be organized, how it would function, how it might be supported, and how obstacles to PIMS development might be overcome. Input from individuals received at the workshop was incorporated into this final report.

The PIMS scoping study provides a roadmap for establishing an end-to-end system for post-hazard-event data collection, archiving of data, and distribution of the data for use to improve hazard mitigation.

1.1.1 Efforts Related to PIMS

Once developed, PIMS will function as one of a broad spectrum of community efforts and existing data management and analysis systems. These systems may differ both in purpose and scale as compared to PIMS but, collectively, they manage or use the breadth of data that would be captured in PIMS. Existing systems currently

- Serve as clearinghouses for data collected after hazard events;

- Archive, synthesize, and format data; and
- Support risk assessment and inform post-event response.

System scales range from small (i.e., focused on specific regions of the country or on specific types of information or hazards) to very large (i.e., country-wide data of various types for various hazards).

Systems serving as post-event clearinghouses include:

- **Digital Libraries**, which are repositories that store information collected following hazard events for the long-term purposes of education and application. Information may be either contained in a report format or a database format depending on the type of data; the data collection and ingestion processes for the libraries may be automated or the data may be entered manually into the library. Some examples of earthquake-related digital libraries include:
  - **The EERI Learning From Earthquakes Library —** General information and reconnaissance reports for recent earthquakes. (http://www.eeri.org/site/content/section/6/35/)
  - **COSMOS Geotechnical Virtual Data Center —** A clearinghouse for geotechnical observations following earthquakes. It includes all types of geotechnical observations and is event specific. (http://www.cosmos-eq.org)
  - **COSMOS Virtual Data Center —** See Center for Engineering Strong-Motion Data.
  - **Center for Engineering Strong-Motion Data —** A data center for receiving, processing, and archiving strong-motion (ground shaking) observations from the United States; the Center is jointly operated by the California Geological Survey and the U.S. Geological Survey. (http://www.strongmotioncenter.org) In 2009 it will also encompass the COSMOS Virtual Data Center that provides virtual access to national and international strong-motion data. . **(**http://db.cosmos-eq.org/scripts/default.plx)
  - **MCEER QUAKELINE Database —** A bibliographic database that covers earthquakes, earthquake engineering, natural hazard and disaster mitigation, and related topics. (http://mceer.buffalo.edu/utilities/quakeline.asp)
  - **Earthquake Survey (Social Science) Data at UCLA —** An archive of quantitative survey data collected as part of research at the University of California, Los Angeles, about the knowledge, attitudes and behaviors of individuals in responding to earthquakes. (http://www.sscnet.ucla.edu/issr/da/earthquake/erthqkstudies2.index.htm)
- **State Clearinghouses**, which are repositories that provide a location, real or virtual, after a damaging earthquake, where engineers, geologists, seismologists, sociologists, economists, and other professionals in the affected area can utilize a relatively large, temporary infrastructure (the clearinghouse) to facilitate the gathering of information, maximize its availability, and better use the talents of those participating in data collection. Clearinghouses can take many forms. Recent examples include the Nisqually Earthquake Clearinghouse (www.ce.washington.edu/~nisqually/index.html) and the Wells, Nevada, clearinghouse (http://www.nbmg.unr.edu/WellsEQ/ch/). In California, revised formal plans for clearinghouses (www.eqclearinghouse.org) were influenced by *The Plan to Coordinate NEHRP Post-Earthquake Investigations* (USGS Circular 1242) [11]. Similar formal plans exist in the Central United States. In the West, the Western States Seismic Policy Council has endorsed the concept

and recommends that member states organize clearinghouses. In all cases, web-based archival is part of the clearinghouse package [4].

- **Homeland Security Infrastructure Program (HSIP)**, which is a data collection effort performed by the National Geospatial-Intelligence Agency that focuses on creating a unified homeland infrastructure geospatial data inventory. It is one of the largest and most comprehensive databases that currently exists. HSIP data, however, are available only to the federal community involved with the homeland security mission, and state and local users may view the data only on a DHS web-based geographic information system [5]. The most recent version is the HSIP Gold dataset, which has updated data and is available on DVDs; it is expected to be web-accessible in the future.

- **Hurricanes Katrina and Rita Clearinghouse**, which is a 20 terabyte clearinghouse for storage of field reconnaissance data. It is GIS-based and is limited to information about Hurricanes Katrina and Rita. This is the most recent and largest application of the clearinghouse concept, and many lessons were learned from its formation and operation [6]. (http://www.katrina.lsu.edu)

Existing systems for archiving, synthesis, and visualization include:

- **Ocean and Weather Data Navigator**, which is a service of the National Oceanic and Atmospheric Administration (NOAA). The Dapper Data Viewer allows users to visualize and download in-situ oceanographic or atmospheric data. (http://dapper.pmel.noaa.gov/dchart/)

- **Spatial Hazard Events and Losses Database for the United States (SHELDUS)**, which is a county-level hazard data set for the United States for 18 different types of natural hazard event (e.g., thunderstorms, hurricanes, floods, wildfires, and tornados). For each event the database includes the beginning date, location (county and state), property losses, crop losses, injuries, and fatalities that affected each county. The data set does not include Puerto Rico, Guam, or other U.S. territories. (http://webra.cas.sc.edu/hvri/products/sheldus.aspx)

- **The WATERS Network**, which is an integrated, real-time, distributed observation system that seeks to address deficiencies in current scientific understanding of the dynamics and spatial variability of water processes by developing a collaborative scientific exploration and engineering analysis network. The WATERS Network is currently in a planning phase, but pilot efforts and cyberinfrastructure development projects have produced a range of databases and data integration technologies. (http://www.watersnet.org/index.html)

Finally, systems for the assessment of effects of hazards and loss estimation for risk-analysis and decision support include:

- **Alliance for Global Open Risk Analysis (AGORA)**, an open source software for multihazard risk analysis. (http://www.risk-agora.org)

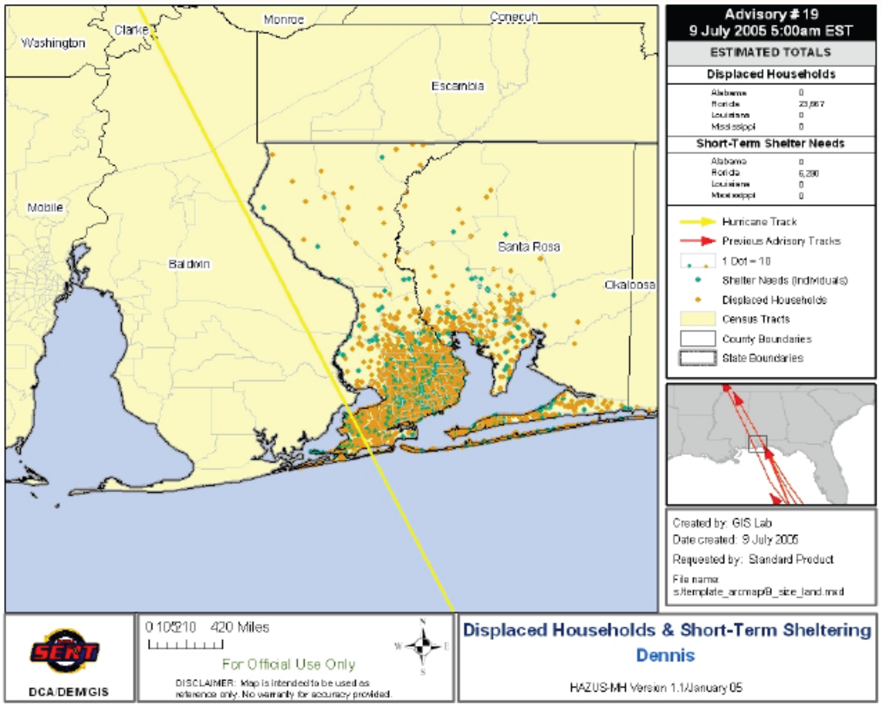


Figure 1.1 Results from HAZUS showing displaced households and short-term sheltering estimates for Hurricane Dennis (from FEMA).

- **HAZUS**, FEMA's software for analyzing risks and potential losses due to earthquakes, floods, and hurricane winds (see Figure 1.1 for an example result from HAZUS). The most recent HAZUS-related release is the Comprehensive Data Management System (CDMS) Version 2.0, which provides users with the capability to update and manage statewide and HAZUS datasets used to support analysis in the HAZUS-MH. (http://www.fema.gov/plan/prevent/hazus/index.shtm)
- MAEViz, open-source seismic risk assessment software to help perform consequence-based risk management. Region-specific data are necessary. (http://mae.ce.uiuc.edu/software_and_tools/maeviz.html)

Although the above list of selected systems or efforts related to PIMS is not exhaustive, it does provide an overview of what already exists (for a complete and detailed list, see Appendix B).

It is expected that PIMS will use existing systems and resources to the greatest extent possible. Moreover, review of these efforts has identified challenges for PIMS implementation in the following areas:

- Data collection, organization, and storage;
- Data curation and quality assurance;
- Privacy and security;
- Information presentation and retrieval;
- Long-term data preservation;

- Data standardization;
- System evolution and change management;
- Coordination/data sharing with public, private, and government sources;
- Lack of information for certain types of structure;[1] and
- Community adoption of PIMS.

When developed, PIMS will join a broad ecosystem of community efforts and existing data management and analysis systems. Ongoing review of these efforts will allow PIMS to utilize existing systems and resources wherever possible and help identify the primary challenges that PIMS will face in implementation.

1.1.2 Existing Types and Sources of Data

Review of existing data reveals what types of data are collected and to what level of detail. Moreover, it provides an overview of the information that PIMS will have to accommodate.

Existing information for PIMS varies significantly both in type and format and is managed in disparate locations by a variety of organizations. Moreover, data volume and availability vary as a function of type and storage location. A good example of this is the differences in data quality and availability between the earth sciences and engineering. Earth science data (e.g., USGS ShakeMaps) are generally readily available in digital database format for public access whereas engineering performance data are relatively sparse and, if collected, are usually stored in a report-based format, not an electronic database format. Within engineering performance data, more information is available regarding the performance of building and bridge structures during earthquakes than for water or sewer networks. In addition, even if data exist, they may not always be accessible (e.g., performance observations for building and bridge structures and utility systems owned by public agencies are often readily available to the public whereas observations related to private industrial buildings and utilities are usually considered proprietary). Finally, much more information regarding the seismic performance of structures has been collected in California and a few other western states than in central U.S. states

The currently available types of data that are relevant to PIMS can be divided into the following basic categories:

- **Maps/photos** (topographic maps, photos before and after events, etc.),
- **Hazards** (ground shaking intensity maps, fault rupture zones, liquefaction and landslides, ground shaking records, wind speed maps, run-up levels, etc.),
- **Buildings/bridges/lifelines** (location, use, plans, date of design and construction, etc.),
- **Loss/socio-economic information** (quantifications of injuries, deaths, displaced people, downtime, economic loss, etc.), and

---

[1] In this report, the term "structure" is used in its most general sense to include building structures and structural components as well as lifeline structures, systems, and components.

- **Critical structures** (designation of structures as hospitals, police, fire, critical bridges, levies, etc.).
- **Field observations** (performance observations from field investigators for the entities listed above and often summarized in a textual field observation report that may not be available in a structured format)

In addition to currently available data, reports from historical earthquakes provide significant information on past hazard events and the performance of the built environment during those events. Comprehensive reports exist for the 1906 San Francisco, 1964 Alaska, 1971 San Fernanco, 1989 Loma Prieta (http://earthquake.usgs.gov/regional/nca/1989/papers.php), and 1994 Northridge earthquakes.

To integrate the various types of data listed above, PIMS must establish collaborative relationships with organizational sources. These are numerous and varied but can be categorized by type as follows:

- Government (federal, state, local),
- Industry,
- Academia, and
- Nongovernmental organizations.

A list of existing sources for data is presented below and although not exhaustive, it illustrates the types and variety of resources available (for a more complete and detailed list, see Appendix B). Some sources, such as the EERI's Learning from Earthquake Program, provide information in report form while others, such as the COSMOS Geotechnical Virtual Data Center, provide information using a map-based interface.

Government Sources

- California Geologic Survey
- California Office of Emergency Services
- California Department of Transportation (Caltrans)
- State departments of insurance[2]
- DHS Homeland Security Infrastructure Program
- FEMA's Multihazard Mapping Initiative
- HAZUS Base Datasets
- LandScan (provides population density estimates based on census data)
- Library of Congress
- Local governments
- NOAA National Geophysical Data Center
- State geologic surveys

[2] These are state agencies that are involved with insurance policy and regulation (see entry in Appendix B for more information).

- USGS (Advanced National Seismic System, Earth Resources Observations and Science, ShakeCast, ShakeMap, National Earthquake Information Center)
- Other federal infrastructure entities (e.g., Bonneville Power, Tennessee Valley Authority, Bureau of Reclamation, Department of Energy)

Industry

- Insurance companies and related organizations (such as ISO and the Institute for Business and Home Safety)
- Utility companies
- Risk analysis and planning organizations

Academia

- Disaster Research Center at the University of Delaware
- Louisiana State University Hurricane Katrina and Rita Clearinghouse
- MCEER and its QUAKELINE Database
- Natural Hazards Center at the University of Colorado at Boulder
- University of California-Berkeley Earthquake Engineering Online Archive
- University of Washington Nisqually Earthquake Clearinghouse

Nongovernmental Organizations

- COSMOS Geotechnical Virtual Data Center
- Earthquake Engineering Research Institute (EERI)
- American Society of Civil Engineer (ASCE)
- Electric Power Research Institute (EPRI)

The organizations with which PIMS will need to collaborate or interface to gather the needed information vary significantly in scale, type of data managed, systems for data management, etc., and include government, nongovernment, industry, and academic entities.

1.1.3 Recommended Practices and Procedures

There exists a broad range of recommended practices and procedures that cover many aspects of data representation, collection and management procedures, and repository policies. For example, there are standardized forms for post-earthquake data collection and standards for performance-based design criteria for buildings. Support for these recommended practices and procedures should be incorporated into the PIMS design as they represent the currently accepted methods used in practice.

Existing standards and input forms can be divided into the following categories:

- Field collection procedures,
- Data standardization and quality,

- Electronic storage and exchange formats, and
- Use of collected information.

Representative sets of standards and forms for each category are presented below (for a more complete and detailed list, see Appendix B).

Field Collection Procedures

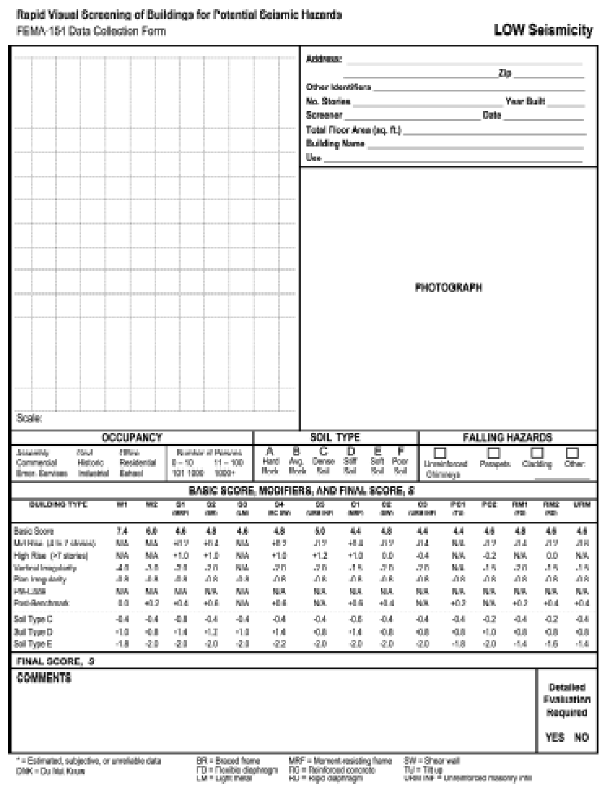
- Caltrans Field Inspection Manuals
- EERI Post-Earthquake Investigation Manual
- FEMA 154 , *Rapid Visual Screening of Buildings for Potential Seismic Hazards* (see Figure 1.2 for a sample form from this handbook)
- USGS Circular 1242, *The Plan to Coordinate NEHRP Post-Earthquake Investigations* (http://geopubs.wr.usgs.gov/circular/c1232/)

Data Standardization and Quality

- ATC-38 *Postearthquake Building Performance Assessment Forms* package (http://www.atcouncil.org/atc38assfms.shtml)
- EERI field reconnaissance forms (http://www.eeri.org/site/content/view/17/31/)
- Standardized Injury Categorization Schemes for Earthquake Related Injuries (http://www.cphd.ucla.edu/pdfs/scheme.pdf)

Electronic Storage and Exchange

- Automated Critical Asset Management System (ACAMS), a web-based tool enabling collection of infrastructure and risk information from owners/operators, law enforcement personnel, and first responders at the state and local levels
- Data Interchange for Geotechnical and GeoEnvironmental Specialists (DIGGS), a coalition of government agencies, universities, and industry partners whose focus is on the creation and maintenance of an international data transfer standard for transportation



Rapid Visual Screening of Buildings for Potential Seismic Hazards
FEMA-154 Data Collection Form
LOW Seismicity

Address:
Zip
Other Identifiers
No. Stories
Year Built
Screener
Date
Total Floor Area (sq. ft.)
Building Name
Use

PHOTOGRAPH

Scale:

OCCUPANCY: Assembly, Commercial, Emer. Services, Govt, Historic, Industrial, Office, Residential, School; Number of Persons: 0 – 10, 11 – 100, 101 1000, 1000+

SOIL TYPE: A Hard Rock, B Avg. Rock, C Dense Soil, D Stiff Soil, E Soft Soil, F Poor Soil

FALLING HAZARDS: Unreinforced Chimneys, Parapets, Cladding, Other:

BASIC SCORE, MODIFIERS, AND FINAL SCORE, S

BUILDING TYPE	W1	W2	S1 (MRF)	S2 (BR)	S3 (LM)	S4 (RC SW)	S5 (URM INF)	C1 (MRF)	C2 (SW)	C3 (URM INF)	PC1 (TU)	PC2	RM1 (FD)	RM2 (RD)	URM
Basic Score	7.4	6.6	4.6	4.8	4.6	4.8	5.0	4.4	4.8	4.4	4.4	4.6	4.8	4.6	4.6
Mid Rise (4 to 7 stories)	N/A	N/A	[illegible]	[illegible]	N/A	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	N/A	[illegible]	[illegible]	[illegible]	[illegible]
High Rise (>7 stories)	N/A	N/A	+1.0	+1.0	N/A	+1.0	+1.2	+1.0	0.0	-0.4	N/A	-0.2	N/A	0.0	N/A
Vertical Irregularity	[illegible]	-3.5	-2.0	-2.0	N/A	-2.0	-2.0	-1.5	-2.0	-2.0	N/A	-1.5	-2.0	-1.5	-1.5
Plan Irregularity	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8
Pre-Code	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Post-Benchmark	0.0	+0.2	+0.4	+0.6	N/A	+0.6	N/A	+0.6	+0.4	N/A	+0.2	N/A	+0.2	+0.4	+0.4
Soil Type C	-0.4	-0.4	-0.8	-0.4	-0.4	-0.4	-0.4	-0.6	-0.4	-0.4	-0.4	-0.2	-0.4	-0.2	-0.4
Soil Type D	-1.0	-0.8	-1.4	-1.2	-1.0	-1.4	-0.8	-1.4	-0.8	-0.8	-0.8	-1.0	-0.8	-0.8	-0.8
Soil Type E	-1.8	-2.0	-2.8	-2.0	-2.8	-2.2	-2.0	-2.0	-2.0	-2.0	-1.8	-2.0	-1.4	-1.6	-1.4

FINAL SCORE, S

COMMENTS

Detailed Evaluation Required
YES NO

* = Estimated, subjective, or unreliable data
DNK = Do Not Know
BR = Braced frame
FD = Flexible diaphragm
LM = Light metal
MRF = Moment-resisting frame
RC = Reinforced concrete
RD = Rigid diaphragm
SW = Shear wall
TU = Tilt up
URM INF = Unreinforced masonry infill

Figure 1.2 FEMA 154 form.

- National Information Exchange Model (NIEM), a federal, state, local, and tribal interagency initiative providing a foundation for seamless information exchange
- Geographic information system-related standards (e.g., shapefiles and Open Geospatial Consortium web services)

Use of Collected Information

- ATC-58 project on performance-based seismic design guidelines for new and existing buildings (http://www.atcouncil.org/atc-58.shtml)
- ATC-63 project on development of a methodology for reliably quantifying building system performance (http://www.atcouncil.org/atc63.shtml)
- FEMA 310, *The NEHRP Handbook for the Seismic Evaluation of Buildings* (www.degenkolb.com/0_0_Misc/0_1_FEMADocuments/fema310/prestnd.html); served as the basis for the ASCE 31 prestandard.
- *NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures* (http://www.fema.gov/library/viewRecord.do?id=2020)
- ASCE 7, *Minimum Design Loads for Buildings and Other Structures*

Inasmuch as these practices will continue to evolve with technology and user needs, PIMS will need to stay abreast of their evolution. For example, consider a current project called The Los Angeles Basin Project, which is being conducted by Robert Wible at FIATECH. The project's purpose is to develop protocols for the linking of disparate hardware and software systems (PDAs, laptops, cell phones, etc.) used by local building officials to develop an interoperable network to gather and disseminate damage assessment and other field inspection data in the wake of a major natural or human-caused disaster. Use of these protocols would rapidly speed damage assessment surveys and provide for efficient methods to transfer information to clearinghouses for storage and possibly directly to PIMS. The recommendations from this project are intended to be completed by October 2008.

A broad range of existing recommended practices and procedures covers many aspects of data representation, collection and management procedures, repository policies, and data application that may be incorporated into PIMS.

1.1.4 Stakeholder Organizations

Collaboration with stakeholder organizations will be crucial in achieving community acceptance of PIMS. These organizations range from large federal government agencies such as FEMA to focused organizations such as the Electric Power Research Institute. They can be categorized by their purpose(s) as related to PIMS as follows:

- Provider of post-event field observations,
- Data source (including data providers who assemble data from original sources),
- Developer of recommended data collection practices,
- User of PIMS, and
- Operation and management of PIMS.

In defining the scope of PIMS and the policies and procedures it will implements, it is critical that stakeholder input be sought and stakeholder expectations be met. The many organizations listed in previous sections should all be considered stakeholder organizations. Furthermore, while most organizations previously mentioned are domestic, international organizations (e.g., International

Association of Earthquake Engineers, the Institution of Structural Engineers, and the Japan Association of Earthquake Engineers) should also be considered, as PIMS could eventually operate in the international arena. A list of organizations that would have a broad interest in PIMS is provided in Appendix B. These organizations, along with those involved in the creation and management of the resources discussed in previous sections, should be considered as the stakeholder community for PIMS.

Collaboration with a wide variety of stakeholders and stakeholder organizations will be critical to successful implementation of PIMS.

1.2 Overall PIMS Framework

This section describes the overall framework for PIMS. First, PIMS is defined, including a description of the three components of the PIMS framework: data collection, information management, and adoption and usage. The following sections describe each component in detail and conclude with a discussion of how PIMS will benefit both the national interest and its users.

1.2.1 PIMS Defined

The scope of PIMS is addressed in the NEHRP Strategic Plan for 2006-2010 [2], where it is stated:

> NEHRP will work with the earthquake professional community to improve post earthquake reconnaissance and detailed and structured data collection; develop a national post earthquake information management center; and stimulate the use of this information management system by researchers, practicing engineers, and government and business leaders.

Thus, PIMS is the end-to-end system for accomplishing the national tasks of facilitating post-hazard event data collection, archiving those data, and distributing the data for use to improve protection against hazards. While PIMS may be defined by NEHRP as relating to earthquake hazards only, implicit in the notion of PIMS is that the system can eventually be expanded to include other hazards.

Given the definition of PIMS, the three components of an overall PIMS framework are:

- Data collection activities and procedures,
- Development and operation of a national information management center, and
- User acceptance and application of PIMS.

PIMS, divided into its three components, is shown schematically in Figure 1.3. Details of each component are given in the following sections.

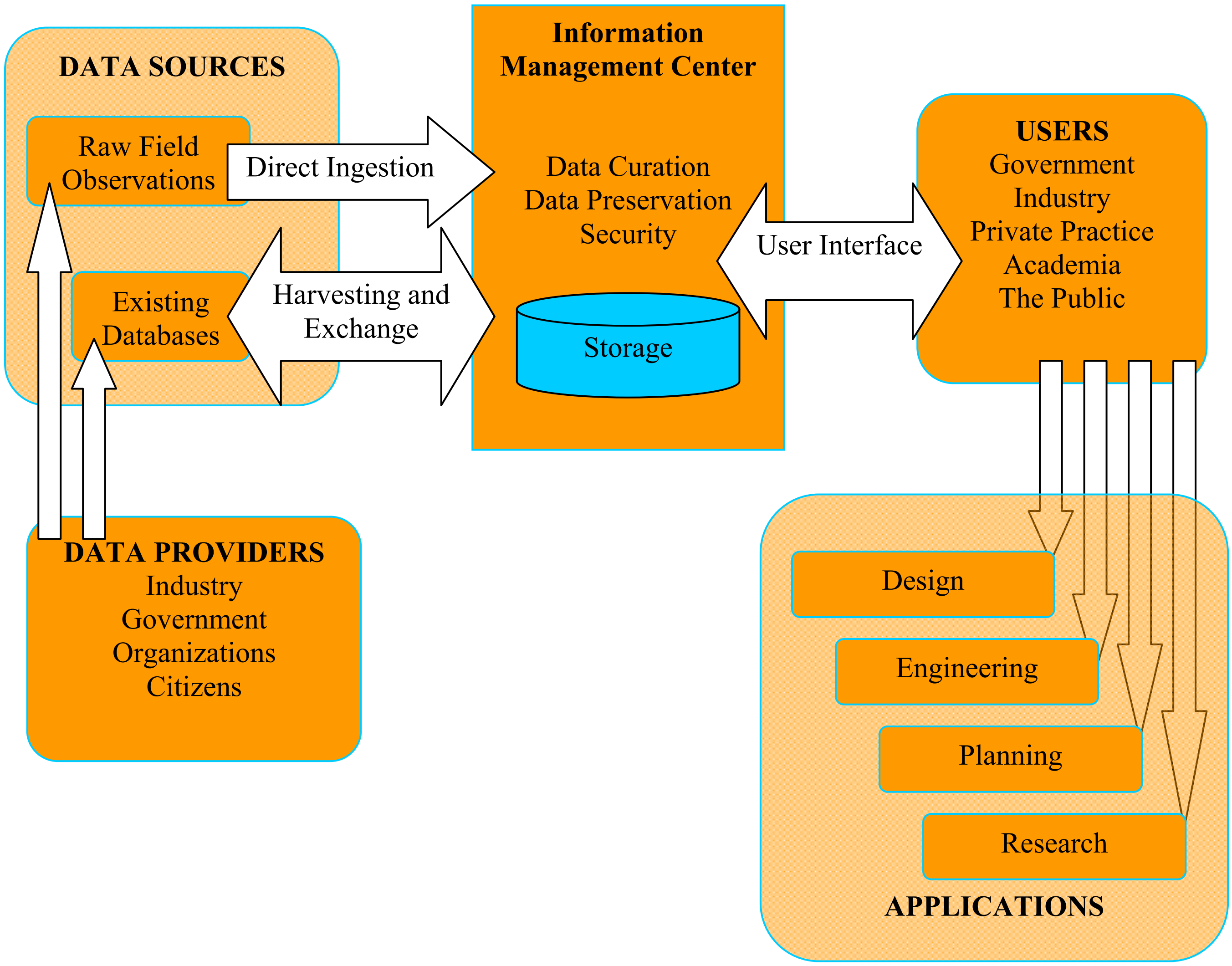


Figure 1.3 PIMS framework.

In general, PIMS must collect and synthesize information from multiple sources, provide curation and long-term preservation services, and facilitate access to its holdings for a variety of purposes. Sources of data for and potential users of PIMS represent a broad spectrum of interests and include all levels of government, industry, academia, private sector organizations and design professionals, and the public. Practical applications of the data obtained from PIMS include design, engineering, planning, and research.

While PIMS is charged with being the end-to-end system for collection, archiving, and use of data, it is important to note the distinction between the information management center component of PIMS and PIMS as an entity that both operates the information system and plays a role in the community to help improve data collection and stimulate use of the collected data. This distinction is important in terms of this report in that, while the information system and the larger sense of PIMS as a community participant are intertwined, and while the report does discuss the larger PIMS concept, outlining a wide range of challenges and issues in realizing the overall PIMS vision, it argues for a specific design of the information system given the larger concept and the structure of the community as a whole. The scope, schedule, and budget estimates in section 4 focus primarily on the development and operation of the information management system component. None-the-less, the community discussion captured in this report and the analysis presented do represent progress in refining the larger vision of PIMS as a contributor to the overall NEHRP goal of improving protection against hazards.

1.2.1.1 *Data Collection*

PIMS will facilitate the collection of data including the creation of data collection plans, development of standards, provision for funding, etc., and it will support the collection of both coarse-grained and fine-grained data. Data collection plans and funding are discussed in Chapter 4, and the development of standards for data collection is outlined in Chapter 2.

Coarse-grained data are obtained rapidly with little curation and are typically used to inform post-event emergency response and recovery (e.g., including emergency operations, research investigations, and recovery operations). Coarse-grained data often are later refined into fine-grained data. For example, coarse-grained data include gross estimates of structure damage and loss over wide areas, global tagging of structures as to their occupancy safety, gross estimates of casualties, and initial and unorganized damage and performance observations; these initial estimates may later be edited, amended, and refined as time allows. Note that PIMS is viewed not as a primary system for use by emergency management offices to respond to hazard events (systems for this purpose already exist) but rather as a secondary system for use by emergency responders, investigators, and researchers to generally inform their response efforts. For example, PIMS may not be able to provide real-time estimates of structure damage or rapidly updated information to response teams but it may provide non-real-time data (e.g., structure inventories, pre- and post-event aerial and satellite imagery, estimates of hazard intensities from organizations such as USGS, and raw data from responders and investigators entered as they are able) and should support FEMA's life-safety mission. In this capacity, PIMS should work closely with state clearinghouses to provide the infrastructure to support raw data collection, storage, and dissemination. However, PIMS should be used simply as a tool to aid these processes and should not be an operations center for the clearinghouses.

Fine-grained data are used to support the overall goals of learning and mitigation and may be collected prior to a hazard event or in the days and months after hazard events. The key to fine-grained data is systematic data collection, which involves considering all components of a structure or lifeline that can affect its performance during hazard events and insuring that all of these components are fully characterized. For example, for a building structure, it is important to identify elements of the lateral-force-resisting systems (i.e., beams, columns, walls, floors, connections, etc.) and quantify the performance of each. Systematic data collection also requires that information be collected for both structures that were damaged or failed and those that were not, such that the differences in the structures can be analyzed to identify the cause of the damage or failure. Finally, fine-grained data collection requires knowledge of structure inventories so that the number of structures damaged can be compared to the overall number of structures affected by the hazard event to assess such things as the effectiveness of mitigation measures and vulnerability to loss. Additional discussion of the types of data and level of detail required for PIMS is presented in Chapter 2.

The first component of the PIMS framework is data collection, in which PIMS acts to facilitate the collection of data through the creation of data collection plans, development of standards, and support for both coarse-grained data (rapidly collected data with little curation) and fine-grained data (data collected over the long term in a more systematic way).

1.2.1.2 *Information Management Center*

The end-use vision for the information management center component of PIMS is to provide users of the system with a means to assure the secure, long term, technologically compatible archiving of data and the ability to retrieve and utilize such data in an intuitive and interactive manner.

For the information in PIMS to be most useful, it must be made available and presented in a coherent fashion. Thus, PIMS is envisioned not only as a single source that provides uniform mechanisms for accessing data from multiple underlying sources but also as a mechanism for organizing and cataloging this information to enable rapid discovery of all information relevant to a user's interests.

Today, most post-earthquake data are contained in field observation reports that usually present anecdotal-type information about hazard event effects and structure performance. These reports are usually compiled by the team that performed the observations and are not necessarily organized by location or by the type of structures inspected, etc. While these reports contain a wealth of useful information, finding specific information about the performance of a specific structure or class of structure may be very difficult. It also is difficult to determine whether one has found all of the relevant reports due to differences in report organization and vocabulary. To improve the situation and ensure that relevant data can be found easily, data must be collected, archived, and presented in a systematic fashion with high quality metadata including information about where and when data were collected and the entity or structure being documented.

The expertise, knowledge, and technical aptitude of PIMS users will vary widely. Users will include designers, engineers, planners, and researchers from government, industry, private practice, and academia. In general, though, their interactions with PIMS will involve a discovery phase during which they define their interests and find relevant information and a retrieval phase during which they view and/or download information to perform their analyses.

Although user interests will vary, it is expected that the user interface will be the same or similar for all users. PIMS is envisioned to have an intuitive interface combining a map with a simple form to enable users to quickly specify the combination of event, location, structures of interest, and other relevant parameters. This interface will allow users to locate structure or lifeline performance data based on many criteria including, as a minimum, the following:

- Location of structure or hazard event,
- Hazard source characteristics,
- Experienced hazard level at a site or within an area,
- Structure or structure component type or classification exposed to hazards, and
- System or component performance level.

Once the relevant data are identified, the interface should be capable of supporting all manner of information overlays (e.g., aerial photos, satellite imagery, hazards maps, and ground shaking intensity maps) as well as providing graphs and tabular data displays (this might look something like the Google Earth example shown in Figure 1.4). Utilizing the envisioned user interface, a system user will be able to pose and hopefully answer the following type of generalized question (initial developments of PIMS may not have the functionality to answer a question with this level of detail): During hazard event A, how many structures of type B with design feature (or component) C were damaged due to hazard level D? What levels of damage and/or failure occurred for these structures?

To answer this question, the user would proceed as follows:

1. Identify the local area of hazard event A.
2. Perform a system inquiry to display all structure type B in the local area.
3. Screen the list of structures to display only those with design feature C.
4. Overlay a hazard map for the area.
5. Select all structures located within hazard level D.
6. Use the system to create a listing for the selected structures that might include for each structure or component: general structure information (type of construction, construction date, building code used for design, structure height and configuration, etc.), links to design drawings, performance notes for the various components of the structure, and other notes from the field investigation.

Using the above information, the user could identify the number of structures that were damaged or failed and then be able to extract the information values for further analysis when desired.

Users also will be able to inquire about the performance of structures independent of specific hazard events. For example, a user will be able to select the subset of structures that were damaged due to a certain ground shaking intensity regardless of the hazard event that caused the ground shaking.

While this example is a generalized use case for PIMS, it covers the most typical uses of PIMS. For a complete description of user needs for PIMS and a list of specific use cases, see Section 2.3.

The second component of the PIMS framework is the information management center comprising the people, processes, and technologies that ensure the secure, long-term, technologically compatible archiving of data and the ability to retrieve and utilize such data in an intuitive and interactive manner.

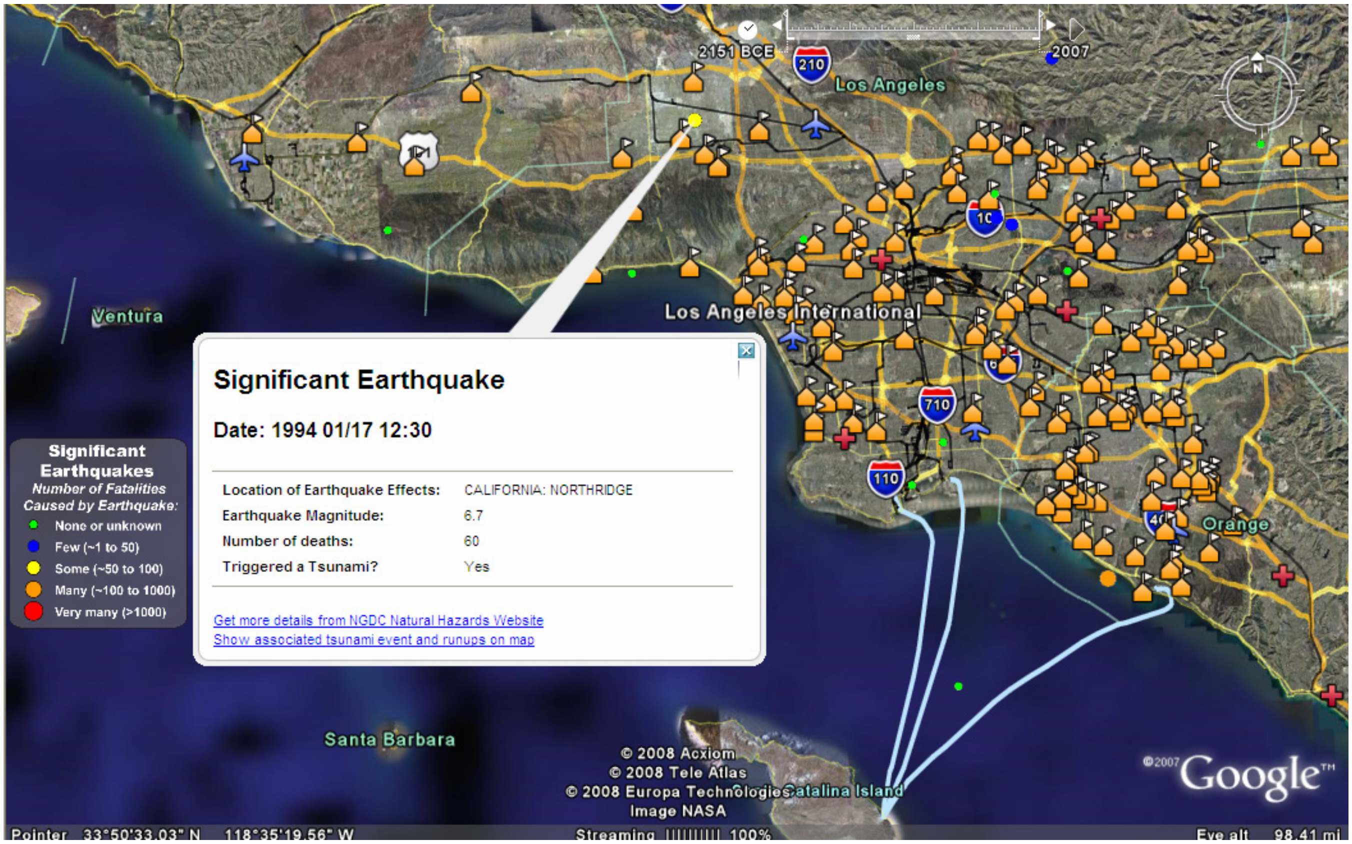


Figure 1.4 Example of a map-centric user interface showing political, transportation, and critical structure overlays (Google Earth) and significant earthquakes [7].

1.2.1.3 *Adoption and Use of PIMS*

Adoption and use of PIMS by the earthquake hazard and other hazard communities is obviously necessary for PIMS to fulfill its larger mission as a community resource helping to improve protection against hazards. To ensure community adoption and use, the PIMS development effort should include outreach, education, and other efforts as described in Chapter 2. It should also be designed to complement existing community infrastructure and to work within the community's cultural, social, and political context as discussed in Chapter 3.

1.2.2 Benefits of PIMS

NEHRP has recognized the benefits of establishing PIMS and made it a priority in their strategic plan [2]. Moreover, the NEHRP Advisory Committee also has cited the significance of the opportunity to develop PIMS [8]. Although the resources required to develop PIMS are extensive, the are believed to be far less than the benefits to the nation of PIMS that can contribute measurably by facilitating improved design, response, and mitigation — the cost-benefit ratio of investments in NEHRP programs such as PIMS could be as high as 10 to 1 [8]. PIMS users could employ the system to ingest, archive, and use coarse-grained data to improve response operations. For example, cross-referencing structure and lifelines inventories and vulnerabilities with imposed hazard demands can be used to predict likely areas of damage and guide responders and researchers. Moreover, raw data obtained from first responders can be used to guide further operations and help researchers locate areas to investigate.

Beyond coarse-grained data applications, the large community of end-users of PIMS urgently needs to obtain and apply fine-grained, long-term data to improve mitigation. For example, quantitative comparison of detailed structure design with actual performance would serve as the basis for improved engineering and design requirements for building codes and standards. Detailed analysis of lifeline response to hazards would inform future design and planning for hazards, and accurate and specific performance data would allow researchers and owners to calibrate their models to better predict performance in future hazard events. Finally, analysis of economic and social impacts of hazards (displaced people, business interruption, etc.) would provide for planning to improve mitigation and response. A complete list of potential uses of PIMS is given in Chapter 2.

PIMS will benefit not only system users but also all those who interact with it, including data providers and system operators. To illustrate this, consider the roles and benefits of the various individuals or organizations that interact with PIMS as outlined in Table 1.1 below.

Table 1.1 Benefits of PIMS

Individual/Organization	Role	Benefits
Data Provider	Provides data for PIMS either through collection and ingestion or through sharing of databases	• Means to create and enforce standards for collection and ingestion • Place to store data • Backup of database • National attribution and use of their data • Data retrieval and application
PIMS Operator	Operates the various functions of the PIMS information management center	• Means to develop best-practices for operation and management of long-term data repositories
Public User	Uses PIMS for personal reasons	• Information on hazardous areas to avoid • Education and learning • Benefits from improved mitigation and response
Government User (e.g., FEMA/DHS, USGS, etc.)	Uses PIMS for government purposes	• Improve planning for response • More accurate loss estimates for relief funds • Calibrate hazard and risk models
Industry User	Uses PIMS as an agent of industry (power, water, manufacturing, etc.)	• Improve product, network, or structure design • Improve response and recovery • Analyze interaction between lifelines and structures
Private Practice User	Uses PIMS for private business purposes	• Improve engineering designs or reduce costs • More efficient planning for business delays

Academic User	Utilizes PIMS for research purposes	• Provides comprehensive data for research purposes • Attribution if data is provided
Nongovernment Organization	Uses PIMS to accomplish missions of organization	• Improve hazard response planning • Investigate social and economic impacts
Oversight Agencies (e.g., NEHRP agencies)	Provides the management, leadership, and funding for PIMS	• Ability to facilitate national standards for data collection, storage, and use • Accomplishes program goals

Once PIMS is developed, it could be used for disasters of all scale, from small, non-federally-declared disasters to large regional or catastrophic disasters. Moreover, PIMS will provide significant guidance in developing infrastructure to manage data related to other hazards and, given the design proposed in Chapter 3, could provide the core infrastructure necessary for additional repositories or simply be expanded to support other hazards. As government agencies and other organizations realize the benefits of PIMS in improving protection against hazards, it is expected that community resources will increase, leading to greater quantities of data, a need for enhanced infrastructure, and increased support for activities that make use of PIMS.

In addition to providing a clear national benefit, PIMS provides significant benefits to all those individuals or organizations that use, operate, or otherwise interact with the system.

1.3 Objectives of the Scoping Project

The purposes, goals, and deliverables of this scoping study are described below. In addition, the tasks performed to accomplish the goals are listed.

The purpose of this project is to perform a scoping study for the envisioned national post-earthquake information management system. The objectives of the project are to assess both the infrastructure requirements (e.g., system architecture, technologies, and logistic issues) and the implementation requirements (e.g., facilities, expertise, and funding) for establishing PIMS as envisioned in the NEHRP Strategic Plan. Due to the overall breadth of the complete PIMS framework, this scoping study focuses primarily on the information management center function of PIMS; however, the other two PIMS components, data collection and use, are discussed, particularly as they relate to the information management center.

The primary deliverable from the project is this "road map" or requirements document that:

- Delineates the user and functional requirements of PIMS,
- Outlines the steps that need to be taken to create the needed information management center,
- Estimates likely costs and levels of effort required for each step, and
- Provides development schedules with milestones.

This road map is a scoping document, not an implementation plan. It will not be used to build the system but rather will serve a first step in developing PIMS. The document is expected to be used to secure funding for PIMS, to provide a basis for development of a PIMS implementation plan, and to aid further planning efforts.

To accomplish the objectives outlined above, the scoping project involved the following tasks:

- Researching existing resources for PIMS including efforts related to PIMS, sources of data, standards and forms, and resource organizations;
- Identifying potential stakeholders in the earthquake and other hazard communities with an interest in one or more aspects of PIMS;
- Performing stakeholder interviews to obtain input regarding user needs, system requirements, and challenges for implementing PIMS;
- Using stakeholder input to develop use cases, which are specific envisioned uses of PIMS to answer a question(s) or solve a problem;
- Defining system requirements and functionality from use cases and stakeholder input;
- Identifying challenges to implementing and operating PIMS;
- Specifying an appropriate high-level architecture for PIMS;
- Creating a phased schedule for the development of PIMS; and
- Composing an initial cost and level of effort estimate for PIMS.

The purpose of the project is to scope the overall PIMS and provide as a deliverable a road map or requirements document that outlines what the community wants to do with the system, identifies challenges to PIMS development, and specifies how a system might be built (design, development schedule, milestones, and required resources).

1.4 Report Outline

This report describes the scope of PIMS in relation to user needs (what users want to be able to do with PIMS), challenges for implementing PIMS, and a design strategy for creating PIMS to satisfy user needs and overcome challenges. Following this introductory chapter, it is organized into four additional chapters as follows:

Chapter 2, PIMS User Requirements, identifies the specific user needs for PIMS and outlines the problems and challenges for implementing PIMS. The chapter contains the following sections:

- **Information Collection Process** – the process by which stakeholders input was obtained, catalogued, and incorporated into this report.
- **User Needs** – the required abilities of PIMS as indicated through interviews with potential users of the system.
- **Functional Requirements and System-level Issues** – the capabilities the system should possess as inferred from an analysis of user needs and the challenges that must be overcome to implement PIMS including coordination/data sharing with all sources, data privacy and security, data collection/standardization/curation, system evolution and change management, long-term data preservation, and community adoption of PIMS.
- **PIMS Requirements Summary** – a detailed summary of user and functional requirements for PIMS presented in a tabular form.

Chapter 3, PIMS Design, describes the design of PIMS to provide for user needs, to satisfy functional requirements, and to overcome the identified challenges. It includes the following sections:

- **Design Strategy** – design concepts for PIMS used to satisfy user needs and functional requirements including potential solutions to the system-level issues and challenges mentioned above.
- **System Architecture** – layout and description of the various components of PIMS.
- **Policies and Procedures** – policies that work in parallel with the system design to make PIMS function as desired and that relate to data standards and provenance, data privacy and security, and persistence of data.
- **Community Adoption of PIMS** – discussions on methods and procedures to encourage community adoption of PIMS.
- **Assessment of Requirements Traceability Matrix** – presentation of the PIMS requirements traceability matrix (RTM), a tool used to ensure that PIMS is designed to satisfy all user needs and functional requirements and to avoid unnecessary design features.

Chapter 4, Development, Operation, and Oversight, covers issues and requirements related to the oversight, management, and leadership of the overall PIMS. It includes the following sections:

- **Phased Development –** a discussion of the phases of PIMS development including required functionality, schedule and milestones, and required resources for each.
- **Oversight and Management –** discussion of issues related to oversight and management including potential management structures, funding sources, and data collection plans.

Chapter 5, Conclusions and Next Steps, summarizes this aspect of the scoping study.

2 PIMS USER REQUIREMENTS

2.1 Overview

By definition, the PIMS must serve a broad range of users, meet stakeholder expectations as a persistent national resource, and complement related government and industry efforts. In this chapter, an analysis of the scope of PIMS is presented from these three overlapping perspectives.

From a user perspective, PIMS must provide the data and interfaces necessary to support the overall processes and high-quality analyses. From a stakeholder perspective, PIMS must be scalable, reliable, maintainable, evolvable, available, and secure. While requirements in these areas could be derived directly from analysis of potential users' interests, looking at them from a system-level stakeholder perspective makes them more explicit. Similarly, an "ecosystem" perspective that looks at PIMS in the context of other efforts helps to identify nontechnical requirements arising from social and political concerns.

2.2 Information Collection Process

The process of collecting information to develop the scope for PIMS involved review of existing documentation and interviews with potential PIMS stakeholders (experts and professionals in the hazards communities). A review of existing documentation helped to capture the consensus conclusions of earlier discussions in the hazards communities related to PIMS. The interviews with stakeholders provided the current, broad, concrete views of individuals highly interested in PIMS. Through this information collection process, the scope of PIMS has been analyzed in terms of user needs, PIMS requirements, and system-level issues or challenges for implementing PIMS.

2.2.1 Existing Documentation

The following documents were analyzed to derive user needs:

- *ALA Workshop on Unified Data Collection* [1] — This report summarized deliberations at the 2006 ALA workshop that served as a precursor to the PIMS project (as discussed in Section1.1). This document outlines the basic scope of PIMS and broadly articulates user needs, system requirements, and challenges for PIMS implementation.
- *Collection and Management of Earthquake Data: Defining Issues for An Action Plan* (EERI) [9] — This document was based on a workshop hosted by EERI in 2002 with 70 experts in the fields of earthquake engineering, earth sciences, and the social and policy sciences to identify the major issues in developing an action plan for an earthquake damage and loss data collection and management framework. Most user needs and system requirements derived from the workshop relate to the types and formats of data that PIMS might ingest from post-earthquake field investigations as well as how it might ingest them.
- *Structural Engineers Association of California's (SEAOC) New Post-Disaster Performance Observation Committee* [10] — This document provides a succinct overview of the Post-Disaster Performance Observation Committee established in 2007 by the Structural Engineers Association of California. The committee is charged to obtain more in-depth information about the performance of structures during natural and man-made hazard events. The committees intent is not to duplicate the functions of others such EERI, the Applied Technology Council (ATC), the Pacific Earthquake Engineering Research Center (PEER), and other organizations

already doing this but rather to supplement those efforts with prioritization from practicing structural engineers. The report argues for systematic data collection and interpretation. Along with this analysis of the concept of systematic data collection and interpretation, it provides a comprehensive overview of current projects, organizations, and other existing resources that are working on these or related efforts.

- *The Plan to Coordinate NEHRP Post-Earthquake Investigations* (USGS Circular 1242) [11] — This circular outlines a plan to coordinate domestic and foreign post-earthquake investigations supported by NEHRP. It deals with identification, collection, processing, documentation, archiving, and dissemination of the results of post-earthquake work in a timely manner and easily accessible format. Therefore, it provides information relevant in developing user needs and PIMS requirements relating to field-observed data ingestion, processing, archiving, interpretation, and presentation.

Beyond this core set of references directly addressing post-event information management, the UIUC project team integrated findings from design documents related to community data systems and cyberinfrastructure developed elsewhere in earthquake engineering research and other scientific disciplines. These documents address topics ranging from user interactions with a system and challenges for system design and operations to lessons learned in operating large multidisciplinary projects and system design and technologies appropriate for long-lived community-scale cyberinfrastructure. These references, also cited in their relevant sections of the report, include:

- *Science, Education, and Design Strategy for the WATer and Environmental Research Systems Network* [12] — This document describes the vision, research opportunities, science requirements, and design requirements for the WATERS network (described in Section 1.1.1). Sections of the document related to network design and networking and informatics cyberinfrastructure requirements were referenced for PIMS.
- *Adaptive Middleware for Scientific Cyberinfrastructure* [13] — A general discussion of the basics of semantic content management, workflows, virtual organizations, etc., to create scalable scientific cyberinfrastructure and cyberenvironments.
- *Information Technology within the George E. Brown, Jr. Network for Earthquake Engineering and Simulation: A Vision for an Integrated Community* [14] — Presents the information technology vision for the Network for Earthquake Engineering Simulation (NEES) to enable on-site and remote research, data sharing, distributed testing, and model-based simulation, etc. Most relevant to PIMS are the descriptions of information technology applications and services in NEES, information technology processes in NEES, and information technologies in NEES.

A review of existing documents was performed as a first step in determining PIMS user requirements. Documents reviewed included contemporary planning documents for post-earthquake data collection, archiving, and use as well as design documents related to community data systems and cyberinfrastructure developed elsewhere.

2.2.2 Interviews with Stakeholders

While existing documentation provided a strong foundation of identifying user needs, discussions with potential PIMS stakeholders proved invaluable in understanding the entirety of user needs. Stakeholders were identified as people who: were potential users of PIMS, were potential operators of a PIMS, might own and maintain data for PIMS, have previous experience with virtual clearinghouses and other systems similar to PIMS, have experience with post-hazard field data collection, and might provide funding or oversight for a system such as PIMS. The list of stakeholders is provided in Appendix H.

In obtaining stakeholder input, a systematic process involving the following steps was consistently applied:

1. **Initial communications with stakeholder —** Initial communications usually took place via email. During the initial contact, the stakeholder was told why they were being contacted and who referred them; the PIMS project was described; input was requested about the scope of PIMS in reference to the areas of expertise of the stakeholder; and a teleconference or face-to-face meeting with the stakeholder was organized. One example of an initial communication email is presented in Appendix C.

2. **Teleconference or face-to-face meetings to obtain stakeholder input —** The second step in the stakeholder input process was a teleconference or meeting during which the stakeholder was encouraged to share his or her thoughts and opinions regarding the appropriate scope for PIMS. Typical meetings included introductions to the PIMS Project team, stakeholder introductions, a summary of the PIMS Project background, and discussions about the professional background and experience of the stakeholder. Input from the stakeholder then was sought in regards to user needs and system requirements for PIMS, challenges for PIMS implementation, and existing resources with which PIMS might collaborate or interface. Each stakeholder also was asked to identify other individuals who might be interested in providing input for PIMS. A typical agenda for these stakeholder meetings is presented in Appendix C.

3. **Creation of summary documents by the UIUC project team from meetings with stakeholders —** As the interviews were conducted, it became clear that stakeholder input could be best summarized using three types of documents: use cases, system requirements and issues entries, and existing resources entries. These summary documents generally were created after meetings with the stakeholders; however, in a few cases, the stakeholders themselves created and submitted the documents.

   A use case is a detailed description of an envisioned use of PIMS to solve a specific problem or to answer a particular question regarding the past performance of the built environment during an earthquake or other hazard event. It includes the following components:

   - Explanation of the background situation and indication as to why the question cannot be answered through currently available means;
   - A statement of the specific question to be answered;
   - Description of the steps that might be taken in using an envisioned PIMS to answer the question;
   - A list of the tools that would be helpful in extracting information from PIMS, or a description of how the information in PIMS should best be presented; and
   - Identification of the important items of information that would be required to answer the question using PIMS in the manner described.

In addition to use cases, system requirements and issues entries were created from stakeholder input. These were simply descriptions of the required abilities of PIMS and the potential challenges for implementation of PIMS that could not be appropriately described in any envisioned use cases.

The third type of summary document was existing resources entries, which were used to catalogue the existing resources for PIMS mentioned by the stakeholders.

Typical use cases are presented in Section 2.3, and all catalogued use cases, system requirements and issues entries, and existing resource entries are presented in Appendices D, E, and F, respectively.

4. **Review and editing of summary documents by the stakeholder —** After creation of the summary documents by the project team, the documents were emailed to the stakeholders for review, editing, and approval.
5. **Posting of the summary documents to the PIMS Wiki site —** Following review and approval of the summary documents, they were posted to the PIMS Wiki[3] for dissemination to, review by, and comment by PIMS stakeholders and others involved in the project.
6. **Synthesis of information in summary documents to derive user needs, system requirements, and challenges for implementation —** The information in the use cases, system requirements and issues entries, and existing resource entries then was synthesized as input to this report. This process, outlined in Figure 2.1, involved review of the summary documents, review of stakeholder meeting notes, organization of use cases into use case families, compilation of user needs and identification of user needs categories, synthesis of system requirements from system requirements and issues entries by category, and synthesis of existing resources from existing resource entries by category.

[3] Wiki is an online technology that allows for the creation of a collective information website on the internet – in this case, the Wiki was created by the UIUC project team. See http://pims.ncsa.uiuc.edu/wiki/index.php/PIMS:Community_Portal.

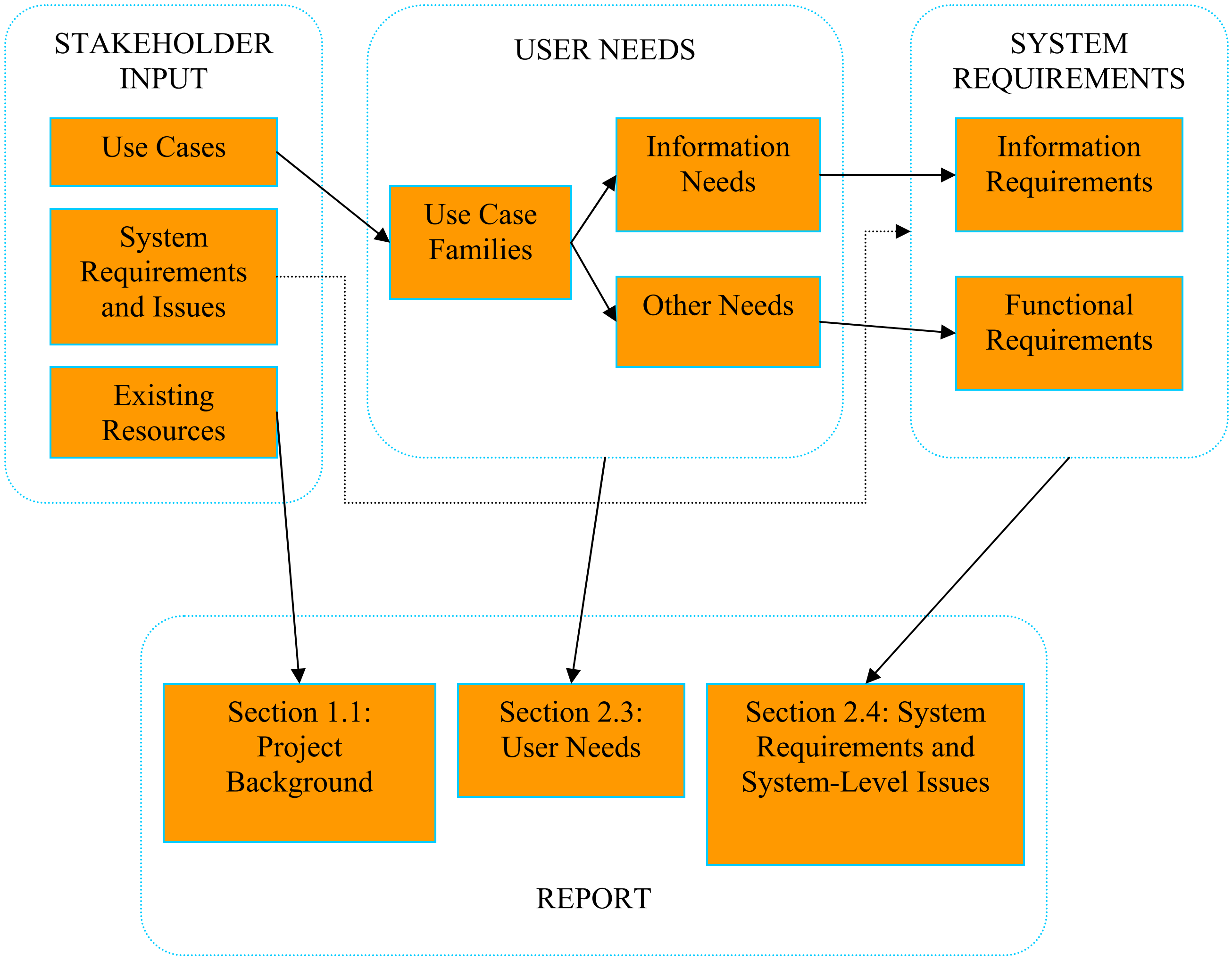


Figure 2.1 Information synthesis process.

While the same general procedure was followed with all stakeholders, the specific process was tailored for each stakeholder. Meeting agendas were adjusted to include discussion items relevant to each stakeholder's background, and stakeholders were encouraged to elaborate on their particular areas of expertise or interest.

While existing documentation provided a strong foundation of user needs, discussions with PIMS stakeholders proved invaluable in understanding the entirety of user needs. Moreover, the process of obtaining stakeholder input was systematic, rigorous, and consistently applied.

2.3 User Needs

This section focuses primarily on requirements for PIMS that derive from productive use of the system by practitioners and researchers. (Requirements related to operations, outreach, interaction of PIMS with other systems, etc., are addressed in Section 2.4 as part of system requirements.) First, a discussion is provided of the high-level categories of use (defined as use case families below) that have been envisioned in prior documents and elicited through our interview process. These uses are then analyzed to derive more concrete requirements for interfaces, functionality, and data products that are direct consequences of the envisioned use of the system (these concrete requirements are described in terms of categories of use). In this way, the derived requirements for PIMS are user-driven, and technical capabilities have a direct link to use cases(s) that require them.

2.3.1 Use Case Families

Use cases can be grouped into the following families:

1. **Evaluation of structure performance —** PIMS data are used to evaluate the performance of structures (buildings, bridges, etc.) or structural components.
2. **Evaluation of geotechnical and geological failures —** PIMS is used to locate and mark the extent and severity of geotechnical or geologic failures (foundation failure, liquefaction, ground rupture, settlement, earthen dam and embankment failures, slope failure, etc.), to evaluate soil and foundation performance, and to record other earth-science-related phenomena.
3. **Lifeline network response to hazards —** Data from PIMS regarding the system performance of lifelines networks (transportation, power, water, etc.) is used to analyze the vulnerability of these systems to hazards.
4. **Loss assessment, risk-assessment, and planning —** PIMS data are utilized to quantify socio-economic effects and losses (injuries, deaths, displaced people, evacuations, business interruption, etc.) due to hazard events, assess the risks due to certain hazards, and plan hazard mitigation strategies.
5. **Inform post-event response and emergency management —** Real-time rapid information collected by PIMS directly after a hazard event and during the post-event emergency response is used by emergency managers. This example is the coarse-grained or short-term data usage described previously.
6. **Information exchange among databases —** PIMS is used to facilitate the sharing of data among databases that would not otherwise be able to exchange data.
7. **General hazard event information source —** PIMS acts as a general organizing force and information source in the hazards communities. It provides publicly accessible information (e.g., via a website) including areas affected by the hazard, what organizations are conducting post-event reconnaissance, links to other databases that store hazards data, etc.

For use case families 1 through 5 above, generally, information derived from PIMS may be used to:

- **Expedite emergency response assistance —** Data in PIMS are used to infer areas of greatest damage to help guide response efforts.
- **Inform post-event reconnaissance —** Rapidly ingested data from field observation teams are used to inform post-event reconnaissance efforts (i.e., to help identify where people have investigated, where the most heavily damaged areas are, etc.).
- **Aid the immediate rebuilding phase —** Data are used to identify those areas that would benefit from immediate rebuilding, the optimum process of rebuilding, etc.

- **Improve design practice —** Entity (structure, lifeline, foundation, soil) performance is compared to the applicable design code to evaluate the effectiveness of the design code or practice and suggest improvements.
- **Validate and improve models —** Results from PIMS are used to validate structural analysis, loss-assessment, lifeline network performance, and other models and to calibrate parameters within these models to improve their accuracy. PIMS also can be used to develop new fragility relationships.
- **Investigate the dependencies and relationship between structural, geotechnical, and lifeline performance —** Information on the performance of the various entities is compared and correlated to better understand how one affects the other.
- **Enable new research directions —** Information not previously available is now available in PIMS and may be used to perform research that previously could not be performed.

Many separate use cases exist within each family. For example, within the evaluation of structure performance family, separate use cases exist for different types of structure or structural component and for different hazards. The full database of use cases created from stakeholder input is presented in Appendix D.

As a first step in classifying user needs, use case families were developed to provide an understanding of the various purposes for which stakeholders wish to use PIMS.

2.3.2 User Needs Categories

The purpose of grouping use cases into families and analyzing use cases was to define lower-level user needs. These lower-level user needs relate to specific abilities of the system, span the use case families, and are intended to be general and capable of serving more uses cases than the specific ones from which they were derived. User needs may be divided into the following categories:

- **User Interface —** Based on the geographic-heavy content of the data in PIMS, a map-based interface may be the preferred solution to provide for efficient data presentation, discovery, and retrieval. Regardless of the interface type, user interface should provide users the ability to:
    - Interactively filter and select entities based on location, overlay values, and entity attributes, etc.;
    - List entity attributes (data and metadata) both graphically (with the ability to highlight an entry and bring up all additional information about it) and in a tabular format;
    - Display maps, photos, and other overlays on top of entities;
    - Perform searches by event, by location, by structure type, by structure component, etc.; and
    - Extract/export information to other programs such as Excel, HAZUS, and MAEViz.
- **Information Needs —** This category relates to the types of information that users want to obtain from PIMS. Information needs may be categorized as follows:
    - General information relating to the area affected by the hazard (e.g., topographic maps, road maps, political subdivisions, aerial photographs) that can be overlaid on top of other

data to provide a background on which to perform further data searching and abstraction (e.g., pre-event and post-event photos can provide much information on the extent of damage and inform further inquiries).

- Geological/geotechnical information relating to the location and distribution of geological or geotechnical failures including ground ruptures, liquefaction, and landslides.
- Quantitative and visual representations of the hazard effects that occurred or the potential for hazard effects to occur (e.g., ground shaking intensity maps, locations of wind speed recording stations, actual ground acceleration time-history records).
- Information about building-type structures, such as type, use, date of design and construction, type of lateral resistance system, performance observations, etc.
- Information about bridge structures (e.g., type, annual daily traffic, lateral-resistance mechanisms, date of design and construction, performance observations).
- Information about other lifelines facilities and components (transportation or utility system networks) including their location, alignment, design features, performance observations, etc.
- Information about critical structures such as hospitals, police stations, fire stations, airports, critical bridges, and communication networks. This information includes the location and designation of critical structures as well as how the entities function and perform as critical structures (i.e., the purposes they serve and how well they served them during hazard events). Moreover, information in this category also relates to systems and components that allow the structures to function both during and following a hazard event (e.g., emergency diesel generators, system controllers, emergency communications).
- Data from historical earthquakes (such as those mentioned in Section 1.1.2) and other hazard events that often are stored in libraries or published in journals such as *Spectra* but also may reside with individuals, state agencies, and private organizations.
- Data quantifying losses during hazard events including injuries, deaths, displaced people, evacuations, structure downtime, repair costs, and business disruption. It is important to identify the socio-economic effects of earthquakes and other hazards as these effects are often the primary drivers to improved policies and practices.
- Certain basic information (location, use, alignment, height, etc.) should, if possible, be collected before the hazard event on inventories of buildings, bridges, other lifeline components, and critical facilities (fire, police, hospital, etc.). For example, system users may want to query based on whether structures have been labeled critical or not or find performance information based on pre-event structure descriptions.
- PIMS system data that relate to PIMS itself and the use of the system. This includes system performance information such as system utilization and audit trails of system operation, outreach and training materials, and policies and procedures.

- **Data Access, Privacy, and Security —** While user requirements related to data access, privacy and security could be wide ranging, two of the primary needs include:
  - Providing data from which personal identification information has been removed and

   - Restricting access to proprietary information or other potentially sensitive information based on concerns related to liability, terrorism, business advantages, etc.

- **Direct Ingestion of Data —** These needs relate to the ability of PIMS to directly receive field observations. In general, data providers would like PIMS to be capable of:
   - Directly accepting data using an internet-based service capable of handling scanned paper notes and reports; electronic notes, forms, audio, and video from portable devices; files from laptops; and files in web folders.
   - Processing paper forms.
   - Capturing structured data based on standardized procedures.
   - Accepting nonstructured data such as images, notes, and reports that have related metadata.
   - Accepting data from the public that might later need to be reviewed and curated.

- **Harvesting and Exchanging Data —** Needs here relate to the requirement that PIMS be able to interface with external organizations or sources to acquire the information needed to satisfy the user information needs defined above. External sources include government, industry, academic, and private sector organizations. Data to be obtained include both pre-event data (e.g., inventories, design drawings, schematics, maps, population counts) and post-event data collected for a specific event but not directly ingested by PIMS (e.g., hazard quantifications, structure performance observations, lifeline/network performance observations, loss quantifications).

To further classify and organize user needs, categories of use were identified relating to the PIMS user interface, needs for specific types of information, harvesting and exchanging of data as well as data access, privacy, and security.

2.3.3 Example Use Cases

To illustrate how PIMS would be used in a real-world situation, example use cases collected during the course of this study are presented below that describe possible applications of PIMS to analyze structural performance, lifelines performance, and earthquake lossess and socio-economic effects.

STRUCTURAL PERFORMANCE EXAMPLE USE CASE

Use Case Title: Evaluation of Overall Building Seismic Performance and Fragilities Development

Explain the background situation and indicate why your question cannot be answered through currently available means:

Little high-quality information exists on the overall performance of building structures during earthquakes (see Figure 2.2 for an example failure of a building structure) that may be used to evaluate the adequacy of code design provisions.

Figure 2.2 Column failure at Olive View Hospital in 1971 San Fernando earthquake (from USGS Library).

State, succinctly, the specific questions to be answered:

What is the correlation between structure performance and experienced hazard levels?

Here, the experience hazard levels should be quantified by shaking time histories or elastic spectra, not just PGA or PGV regional estimates.

Answers to this question may be used to evaluate the adequacy of structure design provisions and codes.

Describe, in a step-wise fashion, how you might envision using the PIMS to answer the question:

(1) Select an event of interest and zoom to the affected area

(2) Identify strong motion recording stations

(3) Identify structures of interest near strong motion recording stations

(4) Generate a data listing regarding the selected structures that provides

a. The codes or specifications used for design

b. Building performance (using appropriate metrics)

c. Structural information (construction material, lateral resistance system, etc.) and possibly drawings

(5) Correlate performance to experienced hazard levels.

LIFELINES EXAMPLE USE CASE

Use Case Title: Buried Pipeline Damage Rates

Explain the background situation and indicate why your question cannot be answered through currently available means:

Damage rates for buried pipelines (see Figure 2.3 for an example pipeline failure) are currently based upon empirical correlations between post-earthquake pipe repairs and estimates of ground motion or permanent ground displacement. Most of the available correlations are based on damage to water pipelines from ground shaking and the extension to oil and gas pipelines and other buried conduits is questionable. Existing relationships for ground shaking damage cannot account for differences in pipe diameter, wall thickness, pipe material, depth of burial, or soil conditions as this information has not been collected in a systematic fashion. As a result, the uncertainty in damage estimates is very high (85% confidence bounds captured by more than plus or minus a factor of 2 on the mean estimate). Relationships for damage from ground displacement are even cruder as they do not account for the direction of ground displacement relative to the pipeline. Existing ground shaking and ground displacement relationships do not provide a means to gauge the severity of pipe damage (e.g., minor leak vs. full break).

Figure 2.3 Compression failure in 49.5 inch diameter steel pipe during 1971 San Fernando earthquake (from USGS Library).

State, succinctly, the specific question to be answered:

How do common pipeline design parameters such as diameter, wall thickness, burial depth, and backfill conditions affect the level of damage experienced by buried pipelines?

Describe, in a step-wise fashion, how you might envision using the PIMS to answer the question:

Improved relationships could be developed by overlaying a pipeline system with critical pipeline attributes that include ground motion maps (either actual motions or estimated motions that account for local soil conditions such as SHAKEMAP) and documented locations of pipeline damage and damage severity. With these overlays, a database can be extracted that is suitable for regression analyses.

LOSS/SOCIO-ECONOMIC EXAMPLE USE CASE

Use Case Title: Post-Event Loss Assessment Validation

Explain the background situation and indicate why your question cannot be answered through currently available means:

PIMS should be able to be used to obtain information and data to validate and refine post-event loss estimates as well as to improve and validate existing modeling tools such as HAZUS-MH. HAZUS-MH utilizes default values based on the San Francisco Bay Area or national averages. Utilizing data captured immediately following an event provides a mechanism to tailor modeling parameters for specific areas within the United States.

State, succinctly, the specific question to be answered:

How do actual losses and social impacts due to a hazard event compare to those predicted by HAZUS-MH (see Figure 2.4 for an example loss estimate from HAZUS)? These actual losses, damage assessments, and social impacts should be used to calibrate and improve existing models.

Describe, in a step-wise fashion, how you might envision using the PIMS to answer the question.

(1) Collect near real-time field data from sensor networks (e.g., ShakeCast).

(2) Import this ground motion information into loss estimation model (e.g., HAZUS-MH).

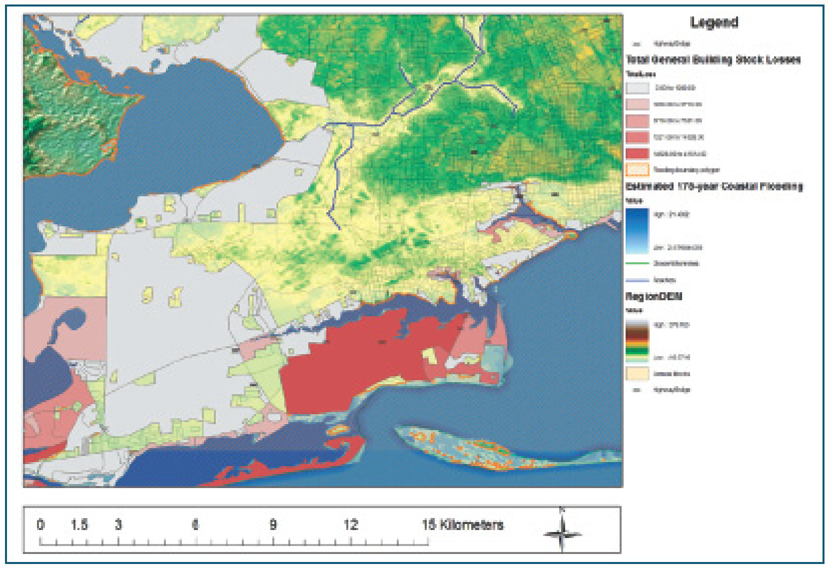


Figure 2.4 Coastal flooding and overall building stock losses in Escambia County, Florida, from Hurricane Ivan (from FEMA).

(3) Collect post-event data from remote sensing (satellite imagery; aerial and field photos or video) and other methods such as casualty reports, displaced and sheltering populations, and demographics.

(4) Use post-event data to either refine or corroborate loss estimates from simulation model.

(5) Display information via internet (e.g., Google Earth).

2.4 System Requirements and System-level Issues

In addition to serving the direct needs of users, PIMS must address their implicit assumptions about system scope and the goals of stakeholders. PIMS also must address issues related to the cultural, political, technological, and organizational context in which it will operate. This section elucidates requirements and issues resulting from these perspectives on PIMS role as a comprehensive long-term national resource. As with user needs, the following sections are organized in terms of general categories which then break down into specific requirements and concerns for a PIMS project. This analysis represents an integration of direct system requirements and issues submissions, information derived through creation of use cases, and investigation of the requirements identified for similar systems reported as existing resources for PIMS and noted in the documents discussed in Section 2.2.

2.4.1 Data Collection, Organization, and Storage

PIMS is envisioned as a comprehensive repository of pre- and post-event performance data across the areas outlined in Section 2.3. Implicit in this statement of scope are requirements that PIMS directly or indirectly collect all relevant data, that it provide sufficient storage to house the data, that it implement procedures to avoid data loss, and that it organize and document its holdings to support users. The core functionality needed is described below.

2.4.1.1 **Data Collection**

PIMS will need interfaces to collect the broad set of information listed in Section 2.3. As noted in the introduction, existing systems contain a significant amount of data and are likely to continue to collect data on future events. PIMS also will need to serve a role as the primary collection point for information from researchers and practitioners for types of information not covered by other systems. As an extension of this, PIMS also may serve as the collection point for supplemental information (e.g., images or movies of damage) provided by ”citizen scientists.” In addition, PIMS must support collection of certain data prior to hazard events as discussed in Section 2.3.

With these extensive data collection requirements, PIMS must ensure that it collects data only at the level of detail (or granularity) required to support its purpose. Coarse-grained data to be used in post-event clearinghouses by emergency managers and others should be collected at a lesser level of detail than fine-grained, systematically-collected data to be used for long-term learning purposes. Otherwise, significant efforts and resources may be wasted (unless, of course, the coarse-grained data can be refined into fine-grained data later as discussed below). Moreover, collection efforts should focus on perishable data (that become lost quickly after a hazard event). Finally, in regards to historic data, PIMS should prioritize the ingestion and storage of the data based upon the likelihood that the data will be lost and the need for those data to support current initiatives.

To support coarse-grained, short-term data collection, PIMS should capture raw data and make it available immediately. Quality assurance and quality control may be performed later and are asynchronous with data collection. Moreover, there is no extra cost for immediate data access; however, the raw data do need to be tagged as to quality.

To satisfy the need for fine-grained, systematic data collection, the data should be collected at such a level of detail that one would be able to generate repair estimates and basic repair designs or plans for the specific structures, lifelines, or entities for which data were collected. In addition, it is important that the fine-grained data collection effort be consistent among the total inventory of entities (e.g., buildings, bridges, pipelines, etc.), not just those that experience damage in order to assess differences in performance. While these requirements seem extensive, some level of systematic data collection to generate accurate loss assessments is currently required by FEMA to obtain federal funds for relief and rebuilding (per the Stafford Act). Moreover, the FEMA-funded ATC-58 project defines standards for conducting the systematic data collection needed to develop the next generation of performance-based earthquake engineering guidelines for buildings. It is possible that the existing FEMA data collection requirements related to receipt of relief and rebuilding funds could be altered to align with ATC-58 requirements and PIMS requirements for fine-grained, systematic data collection.

To support these roles and goals, PIMS will need to have multiple types of collection capabilities ranging from user interfaces (e.g., web-based forms) to programmatic interfaces (e.g., web-services) to direct repository-to-repository harvesting, mirroring, and metadata exchange mechanisms. For each particular source of data, PIMS will have the flexibility to employ the most appropriate collection capability depending on the characteristics of the source (e.g., direct ingestion versus harvesting from existing databases). Finally, PIMS collection mechanisms also will need to ensure that sufficient context (metadata) is captured to support the organization/curation needs outlined in later sections of this report.

2.4.1.2 **Data Storage**

PIMS must provide storage means of sufficient size and distribution to handle data from many large future hazard events. Additional storage will be needed for metadata, indices, audit logs, redundant copies of data, and derived data products. PIMS must consider issues related to how the data will be stored. Will all data be stored in one central location (with back-ups) or will the data be stored in locations distributed across the nation? Moreover, as is discussed further in later sections, PIMS also should anticipate growth in data volume per event as sensors improve and more sensors are deployed. While it is not possible to identify the specific amount of storage that will be needed without concrete plans for PIMS's scope, current clearinghouse sizes can be used to estimate storage demands. For example, the clearinghouse used to store all post-event data from Hurricanes Katrina and Rita is approximately 20 terabytes in size. Although this is significantly larger than what has been collected per U.S. earthquake to date, this amount of data should probably be considered representative of the storage requirements of a near-future large hazard event.

2.4.1.3 **Data Persistence and Availability**

As the ultimate guarantor that the data necessary for post-event analysis is captured and preserved, PIMS must incorporate redundant storage to avoid single points of failure throughout its architecture. It also must include policies and procedures that ensure data integrity and proactively monitor data availability. Depending on the specifics of PIMS's scope, particularly the extent to which it must support event response activities in real-time, PIMS also may need to address operations in the face of power outages, regional network failure, and other external failures such as outages of affiliated repositories. (Real-time requirements also may affect aspects of data collection such as the frequency of data ingest from affiliated repositories.)

2.4.1.4 **Data Organization**

PIMS will be a multipurpose community resource and, as such, it must support organization of the data in ways that ensure users will be able to find and use all relevant data despite the broad range of goals users are pursuing and the evolving set of data formats. While PIMS probably will have a role in community standards processes, PIMS software will need to be capable of gathering metadata and data in multiple, externally determined vocabularies and formats and support new vocabularies and formats over time. Further, PIMS must be capable of synthesizing information from across these vocabularies and formats and providing for the export of information into popular formats. Internally, PIMS must ensure that information is organized to enable reasonable response times to user queries (i.e., through the development of relevant indices).

> *Data collection, organization, and storage requirements for PIMS are extensive. PIMS must collect the data necessary to satisfy the information needs outlined in Section 2.3, provide long-term storage for data obtained from future hazard events, and organize the data such that users can access it in ways most convenient to the users.*

PIMS will require extensive metadata, and standards for data collection (such as those listed in Section1.1.3) should be referenced to develop a core set of required metadata. For example, EERI publishes reconnaissance forms (available at http://www.eeri.org/lfe/recon_forms.html), which are currently used to obtain data for California's Earthquake Clearinghouse (see Section 1.1.1), pertaining to:

- Architectural and nonstructural elements
- Emergency management
- Engineered buildings
- Geoscience
- Ground deformation
- Industrial facilities
- Instrumentation
- Societal impacts (see Figure 2.5)
- Transportation structures
- Tsunami

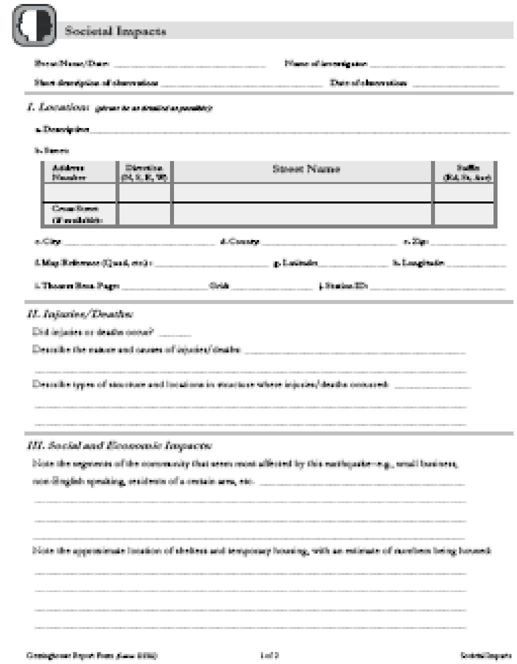


Figure 2.5 EERI form for societal impacts.

In addition, ATC provides the following forms, standards, or guidelines that may guide development of metadata requirements and standards for PIMS [15]:

- ATC-20, *Building Safety Evaluation Forms and Placards*
- ATC-38, *Post earthquake Building Assessment Forms*
- ATC-45, *Field Manual: Safety Evaluation of Buildings after Wind Storms and Floods*
- ATC-58, The next generation of performance-based seismic design guidelines for new and existing buildings
- ATC-63: The purpose of this project is to establish and document a recommended methodology for reliably quantifying building system performance and response parameters for use in seismic design.

2.4.2 Data Curation and Quality Assurance

For PIMS to serve as the source of information regarding engineering, scientific, and social performance data for natural hazards in the United States, it must maintain or improve the quality of the data that it receives and should strive for continuing improvement. Thus, a process is needed to decide what levels of data quality/completeness and retention schedules for data are needed. Moreover, while PIMS should try to obtain high quality data, it cannot ignore lower quality data because such data often are the only data available. For example, in reference to general damage and loss assessments, field observations provide high quality information about the extent and distribution of damage whereas pre- and post-hazard aerial photographs provide lower quality information. However, field observations require extensive resources to obtain and are limited in their scope, whereas aerial photos are more readily obtainable.

To document, maintain and improve data quality, PIMS should capture and preserve metadata that document data provenance (the origin and processing that has happened to the data) and quality assertions about it. By reading the metadata, PIMS users should be able to easily recognize differences in data quality. For example, PIMS users should easily be able to distinguish between remote-sensed and field-verified information and between field observations made by citizens and those made by professionals. PIMS also should incorporate mechanisms to support a best practice manual and automated quality assurance testing. Examples of such mechanisms include peer review of submissions, community annotation and error reporting mechanisms, automated metadata capture, and automated validation of file formats and contents (e.g., to verify that the image data in the file is truly 300x400 pixels as the metadata states).

In addition to making metadata available and accessible, PIMS also must accommodate both short-term (coarse-grained) and long-term (fine-grained) information needs. Owners and operators of facilities and infrastructure have different needs from the scientific and engineering communities. In the hours, days and even weeks following an earthquake, the owners and operators are concerned primarily with life-safety, damage assessment, and service restoration. By contrast, the scientific and engineering communities are usually focused on "lessons learned" studies that result in design or other improvements. Both needs are important in different contexts. Thus, rapid data curation and archiving is necessary for data needed in the short-term whereas more exhaustive, slower data curation methods may be used for long-term data. Therefore, lower quality data standards may be required in the short-term whereas higher quality standards may be enforced in the long-term. Implicit in this requirement is that PIMS provide for data quality evolution where users are permitted to provide low-quality data (i.e., sparse metadata) in the short-term but, when time allows, go back and fill-in the missing metadata.

PIMS must maintain or improve data integrity through quality reviews, metadata documentation of data provenance, and allowance for data refinement with time (where coarse-grained data becomes fine-grained).

2.4.3 Information Presentation, Discovery, and Retrieval

The basic functional requirement in this category is that PIMS provide the means for a user to discover and retrieve data in a manner most suited to his/her particular needs. PIMS should be designed to support multiple interfaces that focus on the needs of particular users grouped by level of sophistication or domain. Users can be broadly grouped as public users and as professional users. The former includes members of the general public who may wish to use PIMS to obtain general information on hazard events and effects. The latter includes users trained and learned in specific fields such as the sciences, engineering, planning, and emergency management. Users in this group, due to their specific knowledge sets, have detailed and specific needs for obtaining information from PIMS. For example, engineers may wish to obtain data from PIMS to analyze the effectiveness of their structure designs in a detailed and component-based form whereas planners and emergency management personnel may wish to assess the global vulnerabilities of the same structures.

The interface for professional users should provide full access to all data (with special provisions for access-restricted data) and tools (this type of interface is described in Section 2.3). In addition, while the initial development of PIMS may focus on creating one common interface for the entire professional user group, future development of PIMS may involve creation of separate interfaces or tools for the subgroupings within the professional user group (i.e., for scientists, engineers, planners, etc.).

The interface for public users, based on the information requirements of these users, may not need to provide full data access and functionality. In fact, a full-access interface would probably hinder the public's use of the system. Instead, an interface to which the public is accustomed (such as a website) should be used. Regardless of the interface type, it should provide access to community-type information, such as:

- Notification of hazardous areas;
- Links to websites of organizations that are managing and performing post-event investigations;
- Links to a website where people can log-on and communicate their status or situation during a hazard event (e.g., to indicate if they need assistance from emergency personnel, have evacuated the affected area);
- Links to PIMS tools used to directly ingest data; and
- Derived or pre-sorted datasets to display common information, such as hazard quantification maps, damage indication maps, etc.

In addition, the public interface should provide a simple means to upload data from the public into PIMS (e.g., pictures of structures, personal accounts of damage, etc.). While this public data may be coarse-grained, unrefined and unverified, PIMS should provide the ability for such data to be reviewed and refined (e.g., by adding high quality metadata, as discussed in Section 2.4.2).

This public interface may be utilized to support the community adoption efforts described in Section 2.4.10.

For both the public and professional interfaces, the initial focus should be on supporting data retrieval and basic map-based and tabular displays. Usability studies in which interface features are tested and feedback is received from users can be used to incrementally improve the interface and to identify where additional data filtering, analysis, summarization, format conversion, and other functionality would be most helpful. This phased approach is outlined further in Chapter 4. Pilot projects that can prioritize requests for PIMS functionality based on real-world needs will provide a venue for assessing the overall functionality of PIMS and a realistic context in which to perform usability studies.

Ancillary issues and requirements related to information presentation, discovery, and retrieval overlap with other system issues. For example, data collection for PIMS should consider the specific needs of the identified user groups, and PIMS must ensure that it only collects data at the level of detail required to support the specific user groups' purposes. In addition, privacy, liability, and security issues all involve trade-offs with usability. The PIMS interfaces must consider the policies, procedures, and technical mechanisms used to identify sensitive information and to specify how that information may be used and by whom, such that PIMS is capable of restricting user access correctly. Moreover, PIMS should incorporate mechanisms to provide aggregate and/or anonymized data to users when full details are not required or cannot be provided. Given the complex and intertwined nature of these issues, this is an area where PIMS should obtain community advice on an ongoing basis.

PIMS must provide the means for users to discover and retrieve data in ways most suitable for those particular users.

2.4.4 Privacy and Security

Providing access to the types of data identified as necessary for PIMS raises significant privacy, liability, business sensitivity, and security concerns. Many privacy issues are the result of the *Privacy Act of 1974*, which limits the unauthorized disclosure of personal information. Names, addresses, and descriptions of specific structures generally cannot be shared because of the restrictions set forth in the *Privacy Act*. Aggregate, anonymized information, however, can be shared. For example, State Farm Insurance cannot share information about individual policies or homes even if the owner's name and property address are removed. It can, however, state that a certain number of houses in a specific zip code were damaged. The medical field deals directly with these same issues as HIPAA national standards and regulations for patient privacy [16] strictly limit the types of medical information that can be shared (e.g., for research purposes).

Data access is further limited by legal liability. Although the *Privacy Act* does allow for sharing of personal information if consent from the information provider is given, providers often are hesitant to share information because of liability resulting from use of that information. For example, a building owner may be hesitant to provide information to the public regarding the damage state of the structure and the possibility of its collapse because, should the structure collapse, the owner then might be liable for injuries, casualties, and other damage. Liability may arise further if the data provided misrepresents the truth or may relate to negligence. For example, issues may arise if a utility company makes public statements regarding the state of its infrastructure but data in PIMS contradicts this claim. Finally, legal liability may arise from the inadvertent transfer of data from PIMS to unauthorized users.

Business sensitivity issues may cause organizations to withhold data from PIMS. The loss of competitive advantage caused by providing proprietary information to PIMS is a primary disincentive for business to provide data. For example, risk assessment and management companies may not be willing to provide data for PIMS because data creation, management, and analysis are their primary profit-generating services. Similarly, insurance companies may not wish to provide the information they use to set their premiums for fear that competing companies could use that information to their advantage. In addition to these concerns, PIMS must deal with issues of copyright infringement as data provided to PIMS may be subject to copyright agreements. PIMS may be required to notify users of and restrict data access, use, and dissemination based on the copyright agreements.

Issues related to security also hinder the sharing of information. Namely, the sharing of information on critical structures (locations, detailed descriptions, and drawings of police stations, fire stations, hospitals, bridges, etc.) may not be provided by government organizations such as DHS because of concerns that this information could be utilized in an attack on the critical infrastructure. In addition, access to loss data for hazard events may need to be restricted because such information also could be used to plan malicious acts. Finally, private data providers may not wish to provide data about their buildings and infrastructure to the public for fear that this information could be used by criminals to gain access.

Although specific procedures or policies to resolve the issues mentioned above are not enumerated in this report, best practices and potential solutions exist and should be considered when PIMS is implemented. These include:

- **Aggregation/anonymity methods —** These involve removal of identifying information (e.g., the address and owner information for a structure) and implementing policies for aggregating raw data so as to obscure detailed, single-entity information while still yielding valuable, meaningful data (as in the insurance company example mentioned previously).

- **Nondisclosure or privacy agreements —** These might focus on such things as PIMS promising to share information only with certain parties or system users. To enforce these agreements, PIMS must be capable of restricting data access such that only specified users are capable of accessing specified information.

- **Release of liability agreements —** These may render data providers free of liability resulting from use of the information that they provide to PIMS and possibly transfer the liability to PIMS.
- **Data provider subsidization —** Agreements may be formulated where PIMS users or PIMS itself pays data providers for their information on a subscription, data usage, or lump-sum basis.

To implement such practices, PIMS will need to provide policies, procedures, and technical mechanisms to identify sensitive information, to specify how that information may be used and by whom, and to restrict access to that information. Using these methods, PIMS users may only access that information for which they have access privileges, which may be limited by nondisclosure agreements, liability agreements, or security concerns. In the basic use case, a full-access user may be able to obtain detailed, individual-entity data whereas a limited-access user would be able to obtain only less detailed, aggregate-type data. For example, a user with a security clearance would be provided access to detailed information regarding critical structures (e.g., their individual locations, responsibilities, staffing) whereas a user without security clearance would only be provided information as to the number and type of critical structures in a given area. In the private arena, a PIMS user who pays subscription fees to data providers would be provided access to detailed information proprietary to that provider whereas those users who do not wish to pay the fees would be provided less detailed and/or aggregated information.

Providing access to the types of data identified as necessary for PIMS raises significant privacy, liability, business sensitivity, and security concerns, and some best practices exist that PIMS could employ to resolve these issues have been identified.

2.4.5 Long-term Data Preservation

Long-term preservation of data in PIMS will be a significant challenge. During the indefinite lifetime of PIMS, there will be rapid technologic evolution, and PIMS will need to preserve data as hardware, software, data formats, and metadata vocabularies become obsolete. For data to be preserved adequately, both the data itself (the digital bits being stored), the descriptive information needed to interpret and assess its quality and the means to analyze, visualize, and export it with then-current software must be preserved. To do this, PIMS should use best-practice policies, procedures, and technology for data preservation and evolution.

Practices to maintain data quality include continuous data quality reviews where data are analyzed to determine if they meet current quality standards. If they do not, policies must specify if the data are to be repaired or removed. These practices must be continually reviewed and updated as technology and PIMS functionality changes.

Quality data are useless unless the technology is available to access them. To maintain data accessibility, either the data format must be continually updated or the original technology used to access the data must be maintained. Currently, there are major information technology projects working on both options. Therefore, as PIMS hardware and software ages, for each component in the system, the pros and cons of migrating the data versus emulating the system component must be weighed, understanding that each migration introduces potential errors/misinterpretations while emulation can introduce performance problems. Different decisions may need to be made for different components.

For data to be preserved adequately, both the data quality and the means to interpret, visualize, and retrieve the data must be maintained.

2.4.6 Data Standardization

Past efforts have shown that challenges associated with data standardization are significant, and this situation is still true today. In general, observations after past earthquakes appear to be disjointed, serendipitous, and nonsystematic such that form-filling and data entry by PDA or laptop for recording the data are not compatible with intelligence gathering in the field. Handwritten notes and digital voice recorders still seem to be the preferred methods for data collection as compared to other more sophisticated systems. Moreover, transcribing and generating data in the evenings after a day in the field may still be the preferable approach for most observers. For example, after the Northridge earthquake, it was believed that a great amount of useful data would have been collected. Instead, data were collected in different formats that were not compatible or else needed data were left uncollected [17].

Development of standards, forms, and tools for diverse groups and needs is very challenging, not only because groups have diverse needs, but also because groups store and arrange data in formats most relevant to their specific applications. Creation of data collection forms and standards is a balance between designing an all-inclusive form and keeping the form sufficiently short so that they will be used. Regardless of the length of the form, users will find it incomplete in certain aspects. Past efforts in this area have met with little success. For example, the California Office of Emergency Services Post-Earthquake Information Clearinghouse has tried several computer systems in mock trial field exercises and found that no single system could meet all the needs of a variety of disciplines [18]. However, efforts to create uniform standards for single disciplines have been more successful. For example, the Geo-Engineering Earthquake Reconnaissance organization (GEER) has been successful in developing standards and methods for post-earthquake collection of geotechnical and geological information and has produced multiple post-earthquake reports (see Figure 2.6) that document geotechnical and geologic failures (available at http://research.eerc.berkeley.edu/projects/GEER/index.html).

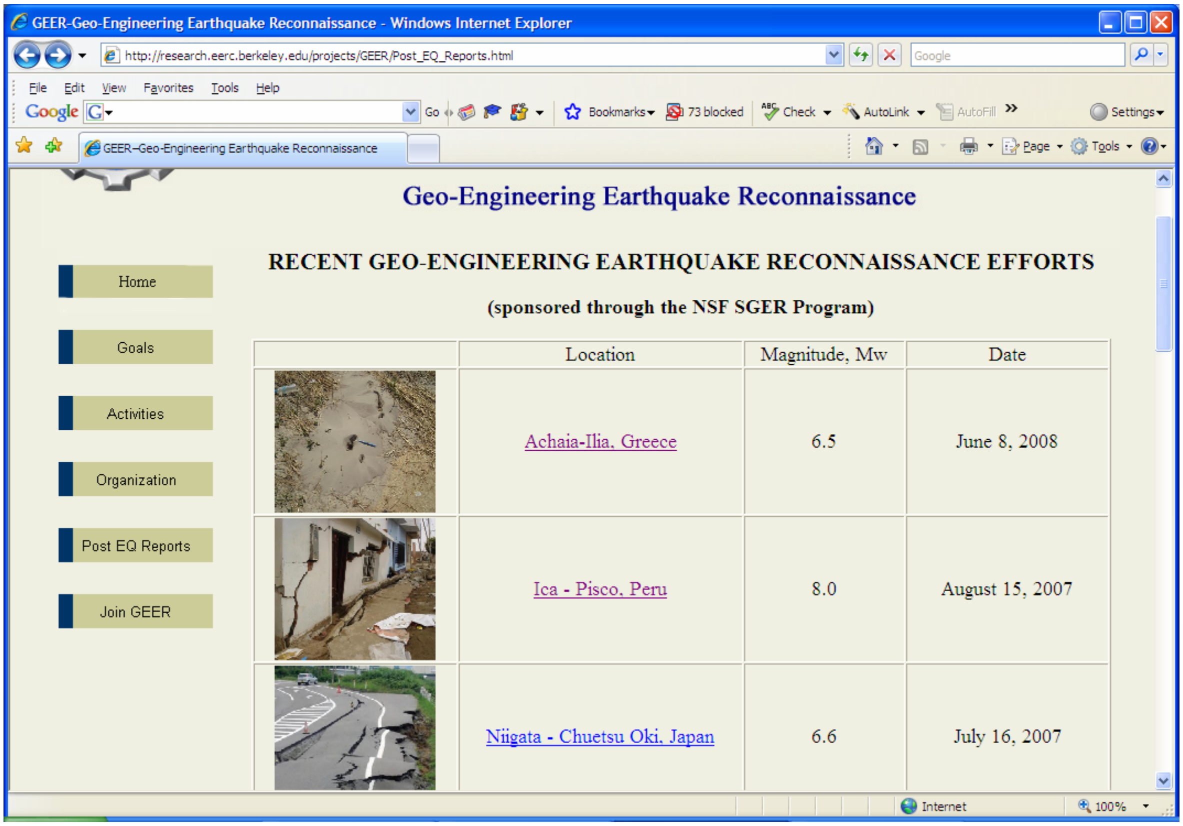


Figure 2.6 Recent GEER efforts and reports.
(http://research.eerc.berkeley.edu/projects/GEER/Post_EQ_Reports.html)

The solution for these issues may involve two components:

- A leading role for PIMS in development and support for national standards related to data collection, ingestion, and storage and
- A requirement that data entered into PIMS be adequately described using metadata.

PIMS, as the national repository for hazards-related data, should play a role in helping the community to develop standards. This role is discussed in Section 4.2.3. It should also directly support those standards through tailoring its ingestion, harvesting/exchange, presentation, and export features to follow those standards. Initial development of standards and PIMS features should be based on contemporary standards followed by large users groups. Some such standards, such as those of EERI, ATC, FEMA and USGS, are listed in Section 1.1.3. In addition to contemporary standards, PIMS should consider future standards for data collection, ingestions, storage, etc. For example, future standards may include those resulting from projects listed in Section 1.1.1.

To address the limitations and challenges involved in developing national standards and to provide a solution where no standards exist, data entered into PIMS should be adequately described by metadata, which means that metadata of significant quantity and quality exist to allow for complete data use and evolution. PIMS should develop policies regarding required levels of description of the collection process, of the observations itself, of the file formats, etc., that are required to pass quality assurance testing (which are described in previous sections about data quality). As described previously, existing standards can be referenced to develop initial policies. Specifically, EERI standards [9] indicate that metadata shall describe the data in relation to the following protocols:

- Collection protocols that address how the data were collected (e.g., standardized forms, interview protocols, sketches, images, etc.)
- Data access protocols that address who may be granted full access to the data
- Data protocols that describe the data structure (i.e., how the data are organized)
- Data storage protocols that stipulate how the data should best be stored, when the data will become outdated, and what to do with the data when they become outdated
- Dissemination protocols that address how to best disseminate the data
- Document protocols that explain how to aggregate or interpret the data
- Identification protocols that identify the data by type (e.g., statistical, graphical, audio, ethnographic)
- Timelines that describe when the data will be needed or most useful (e.g., some data, such as a delineations of the areas affected by the hazard, are more useful directly following a hazard event to inform response and recovery operation whereas other data, such as detailed performance information, are more useful for the long-term purposes of education and learning)

PIMS also might develop policies for which standards are accepted by default. For example, data submitted that conform to EERI or ATC forms/standards might automatically be quality-certified. In addition, while PIMS may wish to enforce high data quality standards, sometimes poor quality data not well-described using metadata are better than no data at all. For these cases, PIMS might accept the data but note that they do not meet quality standards and, as mentioned in Section 2.4.2, provide a means for adding more metadata with time.

To overcome the significant challenges related to data standardization, PIMS should act as a leader in development and support for national standards related to data collection, ingestion, and storage.

2.4.7 System Evolution and Change Management

A fundamental requirement is that PIMS be operational for at least the 50 to 100 years that might pass between severe natural hazard events. This requirement will prove very challenging to satisfy due to the pace of technological evolution and changes in user needs. The pace of technologic evolution is frantic, and the technology that will be available in the future cannot be predicted. In general, technological developments relating to computing, the internet, and information technology can be foreseen with some level of accuracy up to five years in the future. However, it is very difficult to predict with any certainty the technology that PIMS will utilize in 10 years or longer. Consider, for example, that the internet was developed less than 20 years ago. At the time of its development, few could foresee the ways in which the internet would be used and how it would revolutionize the way information is shared. Moreover, this same period of time has also seen incredible development in hardware and software.

While changes in technology over the life of PIMS will be great, changes in user needs and associated system requirements also will be significant. As more advanced buildings, bridges, roads, etc., are constructed, new user needs relating to their performance will develop. For example, currently very few structures are continuously monitored with sensors. In the future, however, it is possible that the majority of structures will be heavily instrumented. Moreover, the type and functionality of these sensors will continue to develop. PIMS users will have a need to obtain information from these more numerous and advanced sensors. As data volumes grow, PIMS users will need more advanced capabilities to search, summarize, visualize, and analyze the data. In addition, as scientific knowledge and analytical tools develop, users will require a larger volume of and more detailed information about the built environment. Finally, should use of PIMS grow to include other countries, PIMS may need to consider, in addition to growth in the number of users, the need for translational capabilities that provide users the ability to enter and retrieve data in their native language.

Challenges associated with technologic evolution and changes in user needs cannot be addressed with a present-day solution. Only as technology evolves and user needs change can solutions be tailored to fit the needs of PIMS. Therefore, PIMS must include means and methods for system evolution and change management. These may include physical resources such as scalable hardware and software as well as policies and procedures and support from those who administer and lead PIMS. For example, PIMS should incorporate best practices for software modularity and be built using an incremental model that assumes periodic technology refreshes over time. In addition, as mentioned previously, PIMS must include a means for adding to and updating metadata as they becomes available.

To satisfy the requirement that PIMS be operational for the 50 to 100 years or more that may pass between severe natural hazard events, PIMS must include means and methods for system evolution and change management.

2.4.8 Coordination/Data Sharing with Public, Private, and Governmental Sources

PIMS as a singular location for storage and maintenance of all hazards-related information is impractical for a variety of reasons. Hazards-related information spans many scales, from structure specific, to event specific, to regional, to global, and it relates to many disciplines, such as engineering, seismology, meteorology, planning, emergency management, etc. For these reasons, it is collected, stored, and managed by a large variety of organizations with disparate goals, methods, and technologies. Even as technology increases our ability to share information between scales and between disciplines, it is unlikely that a single organization such as PIMS will emerge to collect and manage all hazards-related information.

Due to the impracticality of PIMS serving as a single source for all hazards-related information, PIMS will be required to coordinate and share data with a variety of organizations and sources. These include, but are not limited to:

- Public sources of data — libraries and the internet
- Government sources of data — Department of Homeland Security (specifically Homeland Security Infrastructure Program data), Library of Congress, National Earthquake Hazards Reduction Program, U.S. Geologic Survey
- Nongovernment organizations — Central United States Earthquake Consortium (CUSEC), Earthquake Engineering Research Institute (EERI), Geo-engineering Earthquake Reconnaissance (GEER), Mid-American Earthquake Center (MAEC), MCEER, Pacific Earthquake Engineering Research Center (PEER)

- Private organizations and industry — utility companies, insurance companies, risk management firms, national and regional digital libraries and clearinghouses, HAZUS datasets, MAEViz datasets, maps and photos, critical structures, loss information, information archives of the stakeholder organizations such as FEMA

To effect coordination and data sharing with this variety of sources, PIMS may need to utilize technical, economic, and sociopolitical approaches:

- **Technical approaches** relate to means and methods of efficient data sharing, such as the technology and standards used to facilitate exchange. In this regard, it may be prudent for PIMS to coordinate with the standards of existing large data providers such as DHS/FEMA or USGS.
- **Economic approaches** encourage organizations to share their data by providing needed resources, such as primary storage location for data, data backup services, mechanisms for exchange among multiple databases (such as state and multi-state technical clearinghouses), or other services that the organization requires.
- **Sociopolitical** approaches involve using social or political organizations and regulations to encourage data providers to share data. For example, agreements to share data with PIMS may be required for data providers to renew government or other kinds of licenses. Moreover, the vast majority of data gathering and warehousing activity supported by federal funds are publically available as a result of those federal funds. Beyond this, other incentives may be studied to assure access to critical data.

Given that the various organizations with which PIMS must interact will make different choices in regards to data collection, ingestion, storage standards, etc., PIMS should incorporate the technical means for interoperation and, as a project, include the means to negotiate and monitor agreements with other organizations regarding all of the functional areas of PIMS mentioned previously (data collection, curation, quality, privacy, security, etc.). For example, agreements for data exchange between PIMS and sources of data must be formulated that specify what data are to be exchanged, the exchange process, and frequency of exchange. Organizations that provide data likely will not provide all of their data, instead volunteering subsets of the data for exchange, and the specific data or data subsets to be exchanged must be clearly delineated in sharing agreements.

It is impractical for PIMS to serve as the single source for all hazards-related information; therefore, PIMS should provide incentives (technical, economic, and sociopolitical) that will result in other organizations entering into agreements to provide for data sharing and general coordination with PIMS.

Unfortunately, many disincentives for sharing information exist. Most organizations must be motivated to share proprietary information, as described previously, because the sharing of their information may cause them to lose a competitive advantage. For example, USGS ShakeCast datasets are a significant resource for PIMS but, while the majority of the data are publicly-distributable, the datasets still contain some proprietary information related to specific users, such as network designs and network fragility relationships. One possible solution for this is to require that all data collected with federal funds, no matter how remotely traced, should be treated as public and available regardless of any constraints the providing group has placed on it.

Similar to private organizations, members of academia must be motivated to share the information that they have collected. Researchers need credit for their work that they can use for their professional development. In particular, most researchers need to publish data and papers in journals relevant to their fields. As a result, PIMS may need to be capable of tracking the use and distribution of data so that PIMS can negotiate various mechanisms with entities who provide funds for PIMS, system users, journals, etc., to make sure that credit accrues to data providers and that proper attribution is given.

Even if organizations are not limited in sharing their data because of lack of personal benefit, they may not be motivated to share because they do not believe that the information will be used for important purposes. Therefore, filling in gaps in current knowledge should be of mutual interest to both PIMS and data providers. For example, PIMS might incorporate mechanisms such as reports or alerts (that inform data providers that their data are being used) to assure those data providers of the value of their contributions. Data attribution may be so important both for PIMS and for the data providers that it may be wise to require it of all PIMS data.

While the majority of data sharing and coordination with organizations providing data for PIMS will occur in the United States, PIMS should consider coordination with other countries and international organizations. This coordination is necessary because the majority of significant earthquakes and other hazard events do not occur in the United States, and, eventually, PIMS should collect data from these events. In addition to issues discussed thus far relating to domestic coordination, there are significant impediments for international coordination and data sharing. Beyond the obvious issues of language, cultural and standards barriers, agreements must be reached between PIMS and foreign countries to allow PIMS and field reconnaissance teams to operate in those countries. Exploration of all the issues related to international reconnaissance is beyond the scope of this study (and it is unlikely that these issues will be realized in initial PIMS development), but organizations such as EERI, USGS, FEMA, and NSF that have experience with performing investigations in and coordinating with foreign countries should be consulted.

2.4.9 Lack of Information

Information availability and information quality limit the ability of PIMS to function as desired. Both information quality and quantity vary as a function of the type of information and when it is collected. In some information categories, very few data are available. For example, uniform, reliable, and complete pre-event inventories of building structures (residential, commercial, industrial, etc.) do not exist. Tax assessor and other local-government records provide the most comprehensive source for obtaining building stock inventories, and a majority of the data are in digital form [19]. Of course, the quality of these sources varies with the local government provider. While insurance companies may be able to provide pre-event information, even their datasets are limited as they obtain some of their data from local government sources. Another example is the lack of information about recovery from disasters. This includes quantification of the time taken for recovery and repair and the associated costs such as time for repair/rebuild, repair/rebuild costs, opportunity costs incurred, etc. Finally, often the most useful data are perishable, and access to affected areas may be highly controlled immediately after an event, which means perishable data frequently are not collected.

The variation in quality and quantity of different types of data may be severe. For example, hazard quantification data for recent earthquakes may easily be obtained from the USGS in a usable electronic form (e.g., recordings of ground motion); however, data regarding structure performance, lifeline performance, and human and economic losses for the same recent earthquakes is difficult to obtain. Again, while insurance companies may be a good source of post-event data, even their datasets may be limited. For example, only a small portion of people (perhaps less than 15% in California) have earthquake insurance (in part because it is not required to obtain a mortgage). Finally, funding often limits the amount of information collected following an event. For example, aerial photos are very expensive, particularly for large areas, and often cannot be purchased even though they are incredibly useful for identifying structure damage and ground failures.

In addition, not all types of information can be directly observed. Directly observable data such as crack widths and locations, water-run-up levels (see Figure 2.7 for example), and liquefaction provide more certain and possibly higher quality information than nondirectly observable data such as damage indices, socio-economic losses, and other performance metrics. For example, consider how damage to a structural component in the field is assessed. Who makes sure that the initial observer is trained the same way as the next? If photos are used to assess damage, what algorithms or other methods are used to quantify damage and determine the performance metric? Does PIMS store the raw observations notes and photos, the results from damage assessments methods, or both?

Figure 2.7 Stripped forested hills showing clear record of tsunami run-up during 2004 Indian Ocean tsunami (from the International Tsunami Information Centre).

Information availability and information quality are critical for PIMS to function as desired. For some types of information, few data are available or are not easily accessible (structure inventories) while for others many data are available in an easily accessible format (earth sciences data).

2.4.10 Community Adoption of PIMS

While the majority of system requirements and challenges enumerated thus far deal with challenges related to PIMS itself, challenges relating to the PIMS community also exist. The PIMS community may include users, data providers, operators, administrators, and other actors. PIMS will languish without the support of all these community members. It is necessary for PIMS to have the support of data providers in order to collect the required data. Without their involvement and ongoing support, whatever PIMS system emerges will be of little use.

For PIMS to be successful, the PIMS community must support and use the system – researchers and practitioners must buy into the concept of PIMS and acknowledge that PIMS is useful and important to their professions. Moreover, proponents of PIMS, such as the NEHRP agencies, must be convinced of the system's worth. To accomplish this, the development of PIMS shall include outreach (presentations, workshops, etc.), consensus-building efforts, and pilot projects intended to demonstrate to users both what PIMS specifically can do for them and how PIMS contributes to national hazard loss reduction. For example, California's Earthquake Clearinghouse potentially could be used as a test bed to incrementally improve practices in cooperation with a variety of organizations that participate in the clearinghouse. It could then be expanded to include the Western States Seismic Policy Council stakeholders and then other regions with more moderate seismic risk [18]. In addition, CUSEC has developed a plan for implementing a multi-state technical clearinghouse (MSC) during large earthquake events [20]. The purpose of the MSC is to address the interaction between and among the various state clearinghouses set up during a large earthquake and the proposed MSC to ensure coherent and methodical investigations of the physical impacts of an earthquake, timely gathering of perishable data, and central tracking of field investigations. As part of this effort, CUSEC is planning a pilot implementation of the MSC, and it is possible that a preliminary version of PIMS could be involved in this pilot project [21].

Some specific challenges exist for community adoption of PIMS. PIMS could be presented as a tool to help inform post-event response (both emergency and research responses) and, as such, could be seen as valuable to the disaster assistance and operations side of FEMA [22]. Another obstacle is assuring data providers that the data they collect will prove useful. This can be challenging, as patterns in damage observations often do not emerge until the data gathered by many observers are aggregated. As discussed previously, PIMS may include reports or other means to notify data providers of how their data are being used. This, itself, is also a challenge as new procedures and technologies will need to be developed and/or applied to accomplish this. Finally, hazard mitigation, engineering, and other related fields are experiencing changes in personal/professional responsibilities, attitudes, and resources (e.g., increased workload and smaller budgets) and, consequently, PIMS may not be able to rely on the current level of volunteer assistance for data collection, ingestion, curation, etc. and may have to commit its own resources to accomplish these tasks.

Community adoption of PIMS is critical to its success.

2.5 PIMS Requirements Summary

Whereas previous sections have described in general user needs and system requirements, this section provides a complete, detailed compilation of user needs and system requirements for PIMS. In general, system requirements for PIMS were generated directly from user needs considering the three perspectives discussed in Section 2.1 and discussed in terms of user needs, system requirements, and system-level issues in Section 2.3 and Section 2.4. System requirements can be divided into information requirements and functional requirements, as shown in Figure 2.8. Information requirements are the required types of information or data that PIMS must have to function as desired; these translate directly from the user information needs described previously. Functional requirements are the means and methods by which PIMS must operate to achieve the desired performance. They translate more indirectly from user needs than information requirements and are more technical in nature than user needs. Table 2.1 lists the information needs organized by the user information needs categories enumerated in Section 2.3. Table 2.2 lists all system functional requirements obtained through the information collection and synthesis processes. These are organized by the system requirements and system-level issues categories described in Section 2.4. As the user needs and system requirements listed in the tables are detailed and all-inclusive, it may prove difficult to implement all system requirements in the initial phases of PIMS development.

A detailed compilation of the requirements for PIMS is presented in Table 2.1 and Table 2.2.

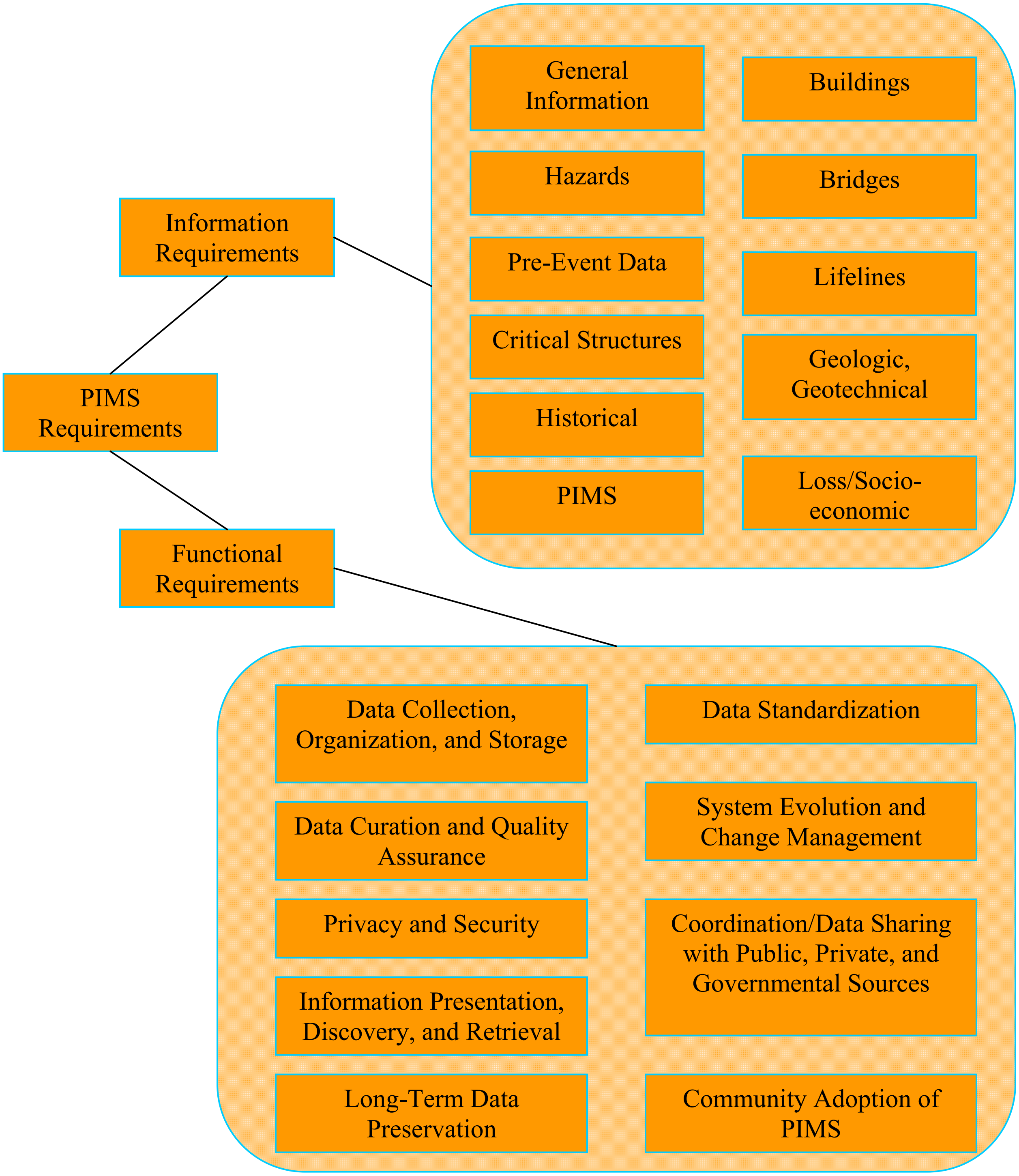


Figure 2.8 PIMS requirements summary.

Table 2.1 Information Requirements Summary

Bridges	
Location	
Traffic loads	
Height	
Construction type	
Completion date	
Design date and applicable design code	
Structure type	
Drawings	
Structure damage	Nonstructural
Structure damage	Beams
Structure damage	Columns
Structure damage	Connections
Structure damage	Deck
Structure damage	Residual drift
Structure damage	Collapse
Structure damage	Foundation
Geotechnical damage	Settlement
Geotechnical damage	Liquefaction
Geotechnical damage	Failure
Casualties	
Injuries	
Downtime	
Repairs	List
Repairs	Cost
Sensor readings	
Buildings	
Location	
Use	
Height	
Construction type	
Construction date	
Site class	
Design date and applicable building code	
Lateral-resistance system	
Drawings	
Single-family residential building information	Floor plan
Single-family residential building information	Cladding

Single-family residential building information	Windows (percent of wall area)
Single-family residential building information	Doors / garage doors (especially garage door wall plane details)
Single-family residential building information	Roof / roof shape /complexity of roof design (dormers) / stick framed vs. truss
Single-family residential building information	Building appurtenances (porches, decks, lanais, etc.)
Non-structural elements	Type
Non-structural elements	Connections to lateral-resistance system
Structure damage	Cladding
Structure damage	Non-structural
Structure damage	Beam
Structure damage	Columns
Structure damage	Connections
Structure damage	Slabs
Structure damage	Residual drift
Structure damage	Collapse
Structure damage	Foundation
Fire Damage	
Estimate of maximum drift	
Geotechnical damage	Settlement
Geotechnical damage	Liquefaction
Geotechnical damage	Failure
Casualties	
Injuries	
Displaced people	
Downtime	
Repairs	List
Repairs	Cost
Sensor readings	
Critical Structures	
Designation as critical structure	
Functions	
Buildings	
Bridges	
Lifelines	
Reservoirs	
Dams	
Performance as critical structure (information not related to structure classification)	

Performance of systems and components that allow the structures to function both during and following a hazard event (e.g., emergency diesel generators)	
General	
Aerial photos	Pre-event
Aerial photos	Post-event
Geologic maps	
Hydrographic maps	
Political maps	
Satellite photos	Pre-event
Satellite photos	Post-event
Soil shear wave velocity maps	
Soil maps	
Soil moisture maps	
Topographic maps	Current
Topographic maps	Historic
Transportation maps	
Population distribution	
Geologic/Geotechnical	
Ground ruptures	Location
Ground ruptures	Displacements
Ground ruptures	Velocity
Liquefaction distribution	
Landslides distribution	
Hazards	
Advanced hazard quantification tools from USGS	
Did You Feel It Data (USGS)	
Earthquake energy	
Earthquake magnitude	
Earthquake shaking maps	Intensity
Earthquake shaking maps	Peak ground acceleration
Earthquake shaking maps	Peak ground velocity
Earthquake shaking maps	Peak ground displacement
Earthquake shaking maps	Other intensity measurements
Epicenter location	
Event designation	
Event general area	
Fault slip distribution	
Faulting characteristics	
Focal depth	
Hazard information for historical earthquakes	

ShakeCAST data (USGS)	
Strong motion recording stations	Location
Strong motion recording stations	Designation
Strong motion recording stations	Records
Water run-up levels	
Wind measurement stations	Location
Wind measurement stations	Records
Wind measurement stations	Peak wind speeds
Hazard potential	Earthquake
Hazard potential	Landslide
Hazard potential	Liquefaction
Hazard potential	Tsunami
Hazard potential	Wildfire exposure
Historical Data	
Data from past earthquakes published in libraries and journals such as Spectra	
Data from past hazard events stored digitally	
Lifelines	
Type (transportation, electric, gas, water, sewer, communication, etc.)	
Location and alignment (system map)	
Construction type	
Construction date	
Design date	
Drawings	
Key design parameters	Depth
Key design parameters	Size
Key design parameters	Material
Contents	Type
Contents	Height/Volume
Damage reports	Location
Damage reports	Description
Photos of damage	
Site-specific soils data	
Permanent ground displacement	
Disruption time	
Repairs	List
Repairs	Costs
(See Helco Inventory Worksheet in Appendix for detailed information requirements for power industry.)	
Sensor readings	

Loss/Socio-Economic	
Casualties	
Injuries	
Displaced people	
Downtime	
Repair costs	
Business/Service disruptions	
Economic cost predictions	
Fragility relationships	
PIMS System Data	
System performance	Usage statistics
System performance	Resource utilization
System performance	Audit trails of system operation
Outreach and training materials	Public website
Outreach and training materials	Community information
Outreach and training materials	Tutorials on system use
Policies and procedures	Collection
Policies and procedures	Organization
Policies and procedures	Storage
Policies and procedures	Curation
Policies and procedures	Quality assurance
Policies and procedures	Privacy
Policies and procedures	Security
Policies and procedures	Evolution and change management
Policies and procedures	Community adoption
Policies and procedures	Agreements with other organizations
Policies and procedures	
Recognized standards	Data collection
Recognized standards	Ingestion
Recognized standards	Storage
Recognized standards	Presentation
Pre-Event Data	
Number and types of structures affected by hazard	
Location of critical structures and lifelines	
Foundation types	
Location of instrumented structures	

Table 2.2 Functional Requirements Summary

Community Adoption of PIMS	
Buy-in of users, data providers, and administrators	
Consensus-building efforts	
Outreach efforts	
Pilot projects	
Coordination/Data Sharing	
Abide by data sharing plans developed by controlling organizations	
Develop the technical means, agreements, and standards for interoperation	Government sources
Develop the technical means, agreements, and standards for interoperation	Public sources
Develop the technical means, agreements, and standards for interoperation	Non-government organizations
Develop the technical means, agreements, and standards for interoperation	Private organizations and industry
Facilitate data sharing between state technical clearinghouses and multi-state technical clearinghouses.	
Facility data sharing among information provider databases	
Maintain data ownership and copyright information	
Provide data backup services	
Provide means to track use and distribution of data	To provide credit to data provider
Provide means to track use and distribution of data	To inform provider of usefulness of data
Provide primary location for data storage	
Utilize social-political approaches/solutions for coordination	
Coordinate PIMS standards with the standards of major data providers	
Develop the ability to operate internationally	
Data Collection, Organization, and Storage	
Assure that information is organized to enable reasonable response times to user queries	
Collect data only at the level of detail required by users	
Support coarse-grained data collection through rapid ingestion	
Support fine-grained data collection by requiring detailed data collection standards	
Harvest directly from MAEViz and HAZUS data sources	
Incorporate redundant storage mechanisms	

Obtain data necessary to satisfy information requirements	
Obtain metadata sufficient to provide for full use and evolution of data	
Perform harvesting operation to obtain data from electronic databases	
Provide a service for entering metadata when uploading to PIMS	
Provide a web-service for uploading data	
Provide means for system reliability in the face of power outages, communication outages, etc.	
Provide storage space for data, metadata, and ancillary information many large hazard events	
Retrieve data from laptops	
Retrieve data from mobile devices	
Retrieve data from web folders	
Serve as the collection point for information provided by 'citizen scientists'	
Serve as the primary collection point for information not covered by other systems	
Support electronic notes, forms, reports, audio, and visual	
Support GIS and other similar file formats	
Support scanned notes, forms, and reports	
Utilize accepted standards, methods, and protocols for data collection and storage	
Provide pull-down menus or similar functionality for fast specification of metadata	
Data Curation and Quality Assurance	
Ensure that data meets quality and formatting standards.	
Ensure that metadata meets quality standards	
Metadata shall delineate to users the differences in data quality	
Metadata shall describe data provenance	
Perform rapid data curation and archiving for short-term data needs	
Perform slow, more thorough data curation for long-term information	
Develop a process to decide what levels of quality/completeness and retention schedules are needed	
Provide mechanisms for quality assurance testing	
Translate street addresses to geo-coordinates	
Data Standardization	

Be a leader in standards development	
Create policies which specify which existing standards meet PIMS quality requirements	
Data shall be adequately described by metadata to allow for complete use and evolution of that data	
Metadata	*Collection protocols*: Field investigation team or individual identifier, time of collection, area investigated, notes regarding collection process
Metadata	*Data Access:* To whom full access to the data may be provided
Metadata	*Data Protocols:* Descriptions of the data structure – how the data is organized
Metadata	*Data Storage:* How the data should best be stored; when the data will become outdated and what to do with the data when it becomes outdated
Metadata	*Dissemination protocols*: How to best to disseminate the data
Metadata	*Document protocols*: How to aggregate or interpret the data
Metadata	*Identification protocols:* The type of data, such as statistical, graphical, audio, ethnographic, etc.?
Metadata	*Timeline*: A description of when data should be used
Support developed standards tailoring PIMS tools to those standards	
Information Presentation, Discovery, and Retrieval	
Identify specific user groups and create/calibrate interface for each	
Perform pilot projects	
Perform usability studies	
Professional Interface	Full data access
Professional Interface	Access to all tools
Professional Interface	Map-based
Professional Interface	Interactively screen and select entities such that entities may be selected based on location, overlay values, and entity attributes, etc.
Professional Interface	List entity attributes both graphically (with the ability to highlight an entry and bring up all additional information about it) and in a tabular format
Professional Interface	Display maps, photos, and other overlays on top of entities
Professional Interface	Perform searches by event, by location, by structure type, by structure component, etc.
Public Interface	Restricted data access
Public Interface	Limited tools
Public Interface	Common interface such as website

Public Interface	Notification of hazardous areas
Public Interface	Links to websites to organizations that are managing and performing post-event investigations
Public Interface	Links to website where people can indicate if they do not need assistance from emergency personnel
Public Interface	Links to PIMS tools used to directly ingest data
Public Interface	Pre-sorted datasets to display common information, such as hazard quantification maps, damage indication maps, etc.
Long-Term Data Preservation	
Maintain access to data as technology evolves	
Review data quality and repair or remove data as necessary	
Privacy and Security	
Manage data access rights and restrict access to data	
Perform data aggregation	
Remove identification information (or personal information) from data	
System Evolution and Change Management	
Develop and maintain plans or methods for change management and system evolution	
Incorporate other hazard information with time	
PIMS shall be operational for 50 to 100 years	
Provide for ability to enter additional metadata with time	

3 PIMS DESIGN

Operating and maintaining cyberinfrastructure[4] with the scope, scale, and duration envisioned for PIMS can be done cost-effectively only if significant attention is paid to design and architecture to create a robust yet evolvable system. PIMS, as a persistent infrastructure for the community, faces many challenges. As noted previously, it will live longer than the specific technologies used to build it and must interface with other systems that will make independent technology choices and independent decisions about data formats and metadata. Further, PIMS must co-evolve with the community. Implicit in the long-term vision for PIMS are assumptions that more comprehensive data will be collected and that new efforts will emerge that can use that data to, for example, improve building practice. A base PIMS with limited functionality and an ability to grow would increase the value of existing data and help catalyze additional data collection, community standardization efforts, and the development of projects that use PIMS data. Such an incremental model for PIMS development reduces technical risk, fosters community adoption, and ultimately maximizes the community return on the investment in PIMS. Addressing these challenges and supporting the scalability, evolvability, and extensibility that will be required to meet them leads to specific design and architectural decisions that are different than those one would make if PIMS were built with a large, fixed scope up front. Sections 3.1, 3.2, and 3.3 present a best-practice design and architectural model for PIMS and review current technology options for a PIMS implementation. Additional material in sections 3.4, 3.5, and 3.6 then discuss the match between the presented model for PIMS and the requirements discussed in Chapter 2.

To maximize return on investment, PIMS must be scalable, starting with focused core capabilities, catalyzing data acquisition and post-event performance analysis, and growing in response. To do so, PIMS must adopt appropriate design and architecture.

3.1 Design Strategies

3.1.1 Overview

In this section, current and emerging best practice design strategies that have been developed in industry and scientific projects to support systems such as PIMS are reviewed. It draws upon cyberinfrastructure analyses for a number of cutting edge projects currently developing national infrastructure including the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) [14], the National Ecological Observatory Network (NEON) [23], the Ocean Observatory Initiative (OOI) (www.orionprogram.org/OOI/default.html), and the Water and Environmental Research Systems (WATERS) Network [12, 24]. This review also incorporates strategies currently being applied in bioinformatics [25] and general discussions within the e-science and grid communities [26, 27].

3.1.2 Service-Oriented Architecture

One of the most obvious design strategies that should be incorporated into PIMS is the use of a service-oriented architecture (SOA). An SOA works by separating its functions into distinct units, or services, that are made accessible over a network. Multiple services are then used together to create the application-level functionality seen by users. SOA is the latest in a series of techniques for

4 A glossary of cyberinfrastructure terms is presented in Appendix A.

modularizing software and is well suited to the development of large flexible systems. SOA focuses on defining the interfaces between individual capabilities within a system and providing a standard protocol to invoke services, which allows the interoperation of components built in different languages. Exposing system capabilities as programmatically-available web services that can be called from within third-party software has become a standard supplement to providing web user interfaces, enabling integration of systems across the internet. SOAs can also be valuable for internal system interfaces, facilitating independent evolution of system components as requirements and/or technologies evolve. Each of the components could be built using different technologies (e.g., using Python, Java, and .Net) and could be upgraded or replaced without changes to the others. In a PIMS context, a SOA design would include, for example, a web service for data discovery (search) and download that could be used internally to support the PIMS web interface for data discovery and download, as part of pilot data curation services, and as a way for third parties to integrate data into applications such as Microsoft Excel.

Service-oriented architecture (SOA) is a current best practice for long-lived systems that serve diverse, distributed customers and that involve components developed and/or operated by multiple organizations.

3.1.3 Content Management

Content management (CM) is computer software used to create, edit, manage, and publish content (e.g. documents, images, videos, and data files) in a consistently organized fashion. It provides a uniform set of services for data storage, versioning, description, access control, translation, provenance tracking, annotation, etc., that is independent of data type and storage mechanism. A CM system is similar to a file system in that it can store arbitrary types of content (e.g., in different file formats), but it adds the capability to associate structured metadata (which is additional information such as key/value pairs recording the authors or creation date that describes the data with which it is associated) and provides mechanisms to browse and search content based on this metadata. CM systems bridge the world of databases and file systems at the metadata level as well; they can be configured to require a minimal set of metadata that must be provided with a data file but that set can be different for each type of data. In addition, CM systems can be configured to accept arbitrary metadata. Further, processing for curation, translation, publication, and similar tasks can be customized for each data type. In the PIMS context, CM would allow the system to accept new types of data (e.g., new image formats, video, sensor streams, and 3-D structures) as they are adopted in the community, to store new types of metadata gathered by reconnaissance missions before standards exist, and to apply more stringent quality control and curation processes to core data and metadata than might be applied, for example, to citizen scientist inputs.

CM systems are critical in contexts where the repository owner cannot or does not wish to constrain the types of data that will be managed, where continual change in data types is expected over the lifetime of the system, and where the particulars of the technology used to implement the system are expected to change. CM has found widespread adoption in enterprise document management and web site management and is gaining traction in community-scale scientific endeavors. As will be discussed in Section 3.1.6, combining semantic grid technologies with CM abstraction can provide additional value in systems where data may be managed by several organizations or may be moved between repositories during its life cycle.

Content management will enable PIMS to keep pace with the community as new data formats arise and metadata standards evolve.

3.1.4 Workflow and Provenance

Workflow and provenance (data history) management provide capabilities for creating explicit descriptions of processes and repeatedly executing such descriptions with minimal human intervention. At its most basic level, a workflow is the ability to execute defined sequences of operations but, in computing, workflow software also often provides end users with an easier way to orchestrate or describe complex processing of data in a visual form, similar to flow charts, that limits the need to understand computers and traditional programming languages. As with CM, a workflow-based design separates common required core functionality from the functionality that depends on the specifics of the data being managed. With workflow, the workflow engine itself manages the overall sequence of processing steps and ensures that data move between them and that each subprocesses is launched, monitored, and has its actions recorded as provenance, independent of the particular types of data and algorithms being managed. Neither the purpose of the processes, the algorithms used, or, in many systems, the language used to implement the algorithms are restricted by a workflow system and, consequently, the set of processes being run within a system such as PIMS could expand and evolve over time.

Capturing provenance information about which processes created particular data products lowers the barrier to managing systems in which processes change over time or in which different subcommunities require different derived data products. In these cases, workflow and provenance serve both as management tools and as communication mechanisms. More advanced workflow and provenance systems manage process descriptions at multiple levels of abstraction (e.g., scientific, mathematical, algorithm, and software instance) and thereby provide a very structured framework for understanding related processes and evolving processing to retain scientific and mathematical correctness while incorporating new software components. Many workflow and provenance systems support distributed processing of information, requiring the invocation of remote web services within the workflow or transfer of the whole workflow description to remote, often higher performance, resources for execution.

In the context of PIMS, workflow and provenance can simplify tracking of data ingestion from remote repositories and can be used to structure and automate curation, quality assurance, data translation, harvesting (the process where PIMS exchanges or obtains data from other electronic databases), backup, and other processes. It also could potentially be used to provide a means to generate custom-derived data products such as statistical summaries specified by users as workflow templates. While such processes could be hard-coded into the system and evolved through standard software development mechanisms (i.e., creating new code and versioning the system), workflow and provenance provide a more facile means to perform such evolution and allow old and new workflows to coexist. Moreover, workflows can be customized as needed for different data sources or types of data. Provenance then provides the means for documenting which processes were being applied to PIMS data over time, for communicating that information to users, and for assuring reproducibility of results based on PIMS data.

Workflow will enable PIMS capabilities in areas such as data curation and the creation of derived data products to evolve over time. Provenance capabilities will ensure that users can understand the processing that has been applied to a data set and discover its source.

3.1.5 Virtual Organization-based Security and Configuration

Stand-alone software applications encode a wide range of decisions about design, including what security mechanisms are needed, where to store data, and how to optimize performance; in some cases, they also expose such choices as local configuration options. In large systems such as that envisioned for PIMS, such choices should be coordinated across all components and alternate choices should be implementable without requiring each component to be modified. In the grid community, the design strategies supporting such separation of concerns are associated with the concept of virtual organizations (VO). VOs are organizations built dynamically to address a specific set of goals by specifying the subset of resources (i.e., people, software, hardware) from other organizations that will comprise it. Implicit in this notion is the assumption that all resources brought into the VO will work together and will modify their operations to conform to the policy, procedure, and resource context of the VO. In the enterprise and grid contexts, VO-centric design manifests itself as an assemblage of components that assume that they will interact with an external service to authenticate users and discover which groups a user belongs to. In web portals (see, for example, http://www.liferay.com), VO-centric design involves the interaction of portlets with the portal framework to discover security, display, and other information, which allows portlets built by independent parties to coexist with each other in the same web page. In all cases, instead of standardizing the set of resources and components that will make up the system, one standardizes the mechanisms for registering new resources with the system and for components to discover each other and the global policy and procedure choices with which they must work. This extra level of indirection dramatically lowers the barriers to evolution of the system over time and reduces the coupling between components and, thus, the amount of coordination required between developers of different components.

In a PIMS context, VO-based design, along with the use of content and workflow management, will enable an incremental approach to development that can begin providing services early while still providing a means to expand the system over time. Security, including authentication, authorization, auditing, accounting and encryption, is a key area for VO-based design given the likelihood of new security exploits in the coming decades which would necessitate rapid, systemwide upgrades of security components. PIMS could start, for example, with basic local username/password authentication and later migrate to stronger and more scalable systems that involve public-key certificates and/or authorization of PIMS users based on their login to their local university or company infrastructure (see, for example, http://gridshib.globus.org). VO-based design can also provide the group context framework necessary to allow groups of PIMS users to specify the set of resources and processes that they trust or accept in terms of quality or completeness. This approach would enable PIMS to support groups that have stringent requirements for validation while also serving those who are willing to accept data that come from less reputable sources, have less documentation, or have been derived using unproven algorithms.

Virtual organization centric design allows customization of PIMS services for particular groups of users and the evolution of systemwide policies such as those related to authentication and access control.

3.1.6 Semantics

Another design element PIMS should consider is the emerging framework of the semantic web and its extension to the semantic grid. In the semantic web, all data are completely defined in a manner that enables computers to interpret and use the data without human intervention, thus allowing a computer to perform tasks that previously required human action. Semantic web technologies make explicit the meaning of data values and metadata terms and separate this discussion from that of how that information will be organized and stored as digital bits. In this sense, semantic technologies are the right level at which to define data and metadata standards related to PIMS. Semantic standards (ontologies) would support agreement and transfer of well-defined information between the PIMS repository and other data providers without requiring all of them to adopt the same technologies for their data storage.

At the most basic level, semantic designs address two issues. First, from their roots in the web community, semantic designs require globally unique identifiers for resources (e.g., content as discussed previously as well as non-web entities such as people and organizations) and metadata (e.g., representing the concept of damage state and the damage state value of "low"). Global identifiers can avoid situations where different data sources are recording the same types of information but are incompatible due to different local choices for identifiers. At a minimum, global identifiers provide a common hub for mapping between database identifiers. The second core advantage of a semantic approach is the underlying model of triples, which identify the subject, predicate (verb), and object in relationships. Most people are familiar today with the advantages of languages such as XML, which is a general purpose specification for creating custom markup languages (an artificial language that uses a set of annotations to text that describe how text is to be laid out, formatted, structured, etc.) and provides a common syntax for describing hierarchical information (e.g., tree structures).

The resource description framework (RDF) plays a similar role in the semantic web, providing a common syntax for describing networks (graphs) of relationships as triples. In RDF triples, the subject denotes the resource, and the predicate (verb) denotes traits or aspects of the resource thus expressing a relationship between the subject and the object ("buildingA" "hasDamageState" "low" or "datasetB" "hasCreator" "John Smith" where each quoted entry would be a global identifier rather than a text string). Such a model is more powerful than the XML model in that the predicates of relationships are explicit and because networks are a superset of the hierarchical trees that XML provides. These benefits are already being recognized in the disaster management community. The capabilities of the RDF are critically important in multidisciplinary situations where there is more than one logical way to organize information and where different data providers may know different facts about the same objects [13]. The semantic web and semantic grid hold additional promise for systems such as PIMS in that they can provide mechanisms analogous to XML schema and XSLT (which can be used to transform XML documents into other XML documents or those that can be read easily by humans) that allow verification of information against standard vocabularies and translation between vocabularies. It is likely during the lifetime of PIMS that a robust infrastructure for performing logical inferences and implementing rules that can be used to assess data quality and trustworthiness will develop that could replace less automated, lower-level mechanisms that are current best practice.

Semantic web concepts can be applied across the other design concepts presented in this section. CM systems can incorporate global identifiers and store RDF metadata and thus gain the abilities to more easily federate information across repositories and to combine metadata and data from multiple sources without custom programming. Such a semantic CM can also serve as the system wide registry needed for VO-based designs. Semantic technologies provide a clean mechanism for workflow and provenance systems to manage process descriptions at multiple levels of abstraction and to automate conversions between them (i.e., taking a mathematical description of the desired process from a user and discovering and invoking specific services to execute the desired plan). Of all the design concepts that are relevant to PIMS discussed in this section, semantic technologies are probably the least mature. However, current best practice in multidisciplinary efforts such as biomedicine [28] already incorporate basic exports of semantic information, and many systems make effective use of semantic technologies to enhance their functionality and lower development costs.

Semantic web technologies allow evolution of how and where data and metadata are stored without affecting the meaning of that information.

3.1.7 Virtual Machine Implementation

Deployment of production services for PIMS also would benefit from deployment within a virtual machine environment. A virtual machine (VM) is a software implementation of a physical computer that executes programs as if it were a real computer. VMs are a key element of *cloud computing*, the concept that an organization's data and software can be pushed to service providers “in the cloud” who would own and operate the actual computing hardware and network infrastructure and are a standard element of internet-based business infrastructure. They can provide very cost-effective hosting because they allow deployment of multiple independent services on a single physical machine without the issue of software incompatibilities. They also can provide very robust and scalable service deployments because services can be moved between physical machines dynamically if hardware fails. Further, as services become popular, they can be replicated across many VMs to handle high loads. Deployments on VM infrastructure can be managed with off-the-shelf software and can incorporate best practices for SOA, including redundant disk arrays and mirrored load balancers. They also can support services that require shared back-end databases and file systems. Such an infrastructure can be rapidly and cost-effectively grown if demand for services goes beyond the level budgeted or, in the case of PIMS, in response to hazard events that will grow the data storage requirements and cause a spike in system usage. A PIMS infrastructure based on VMs would be in a position to accept donated hardware and/or cloud computing services very quickly after an event. In addition to providing a robust central capability, a VM infrastructure also provides a convenient mechanism for replicating services at multiple sites and for rapid deployment of experimental services. Given that VM images of working services can be retrieved over the internet and installed on local machines to quickly create running services without the need to compile, install, or assemble multiple components as would be required with traditional software deployment strategies, such an infrastructure would reduce the costs of operating a back-up site and reduce the effort required for partners to develop PIMS compatible services and test data transfer capabilities. Similarly, a VM infrastructure would allow rapid creation of multiservice test infrastructures that duplicate required portions of systems such as PIMS and allows testing of new components under realistic circumstances. Such capabilities would reduce the costs of operating a back-up site and reduce the effort required of partners to develop PIMS compatible services and test data transfer capabilities.

Virtual machine based design would help optimize PIMS’s use of computation and storage hardware.

3.2 System Architecture

The design principles discussed in the previous section result in a high-level architectural view of PIMS as shown in Figure 3.1. Within the constraints of the architecture, there is still significant room for variation, and it should be expected that the exact technologies proposed for PIMS will vary according to the proposer, the decisions made regarding its final project scope, the potential synergies that may be possible with existing systems, and the amount of time that elapses before a functional PIMS is implemented. Such choices may affect the development and operations costs for PIMS and their trade-offs and could affect the ease with which future modifications can be made.

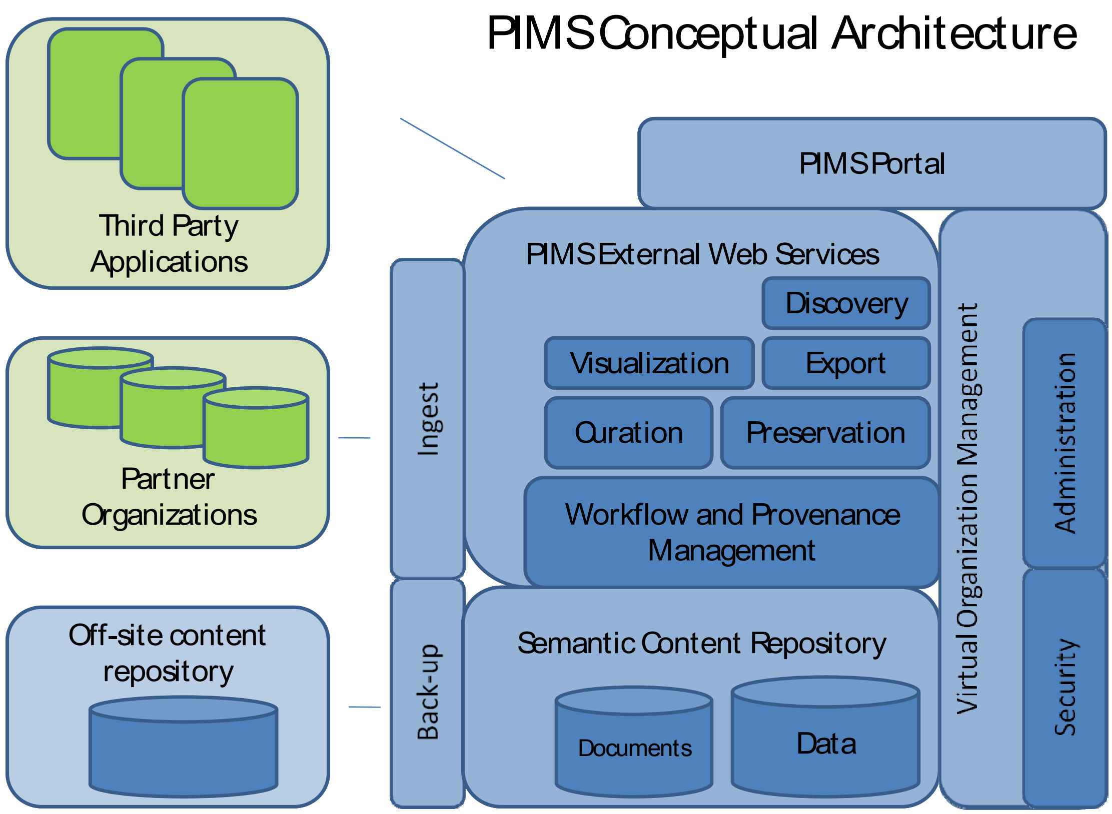


Figure 3.1 PIMS conceptual architecture.

The requirements outlined in this report and the design responses to them discussed above do allow some level of analysis of options for specific software and approaches to PIMS if it were to be developed in the near future. Various agents are producing advances that should be considered as PIMS plans are updated, including trends such as Moore's Law[5] and its analogs for data storage and network technologies; private sector interest in geospatial data and in streaming observations related to localizing information for cell phone users and for tracking RFID-tagged goods; and the wide range of developments being pursued under the auspices of NSF's environmental observatories [23, 24], under their associated pilot and regional observatory efforts, and by the broader community of researchers and practitioners using geospatial information.

For purposes of this discussion, core technologies for PIMS are presented under the categories of content, workflow and provenance, and virtual organization management but these ideas are often

[5] The number of transistors that can be inexpensively placed on an integrated circuit is increasing exponentially, doubling approximately every two years [Wikipedia].

intertwined in software products (i.e., content flows through processes, processes create content, and user and group management is embedded somehow in almost every multiuser product). Thus, some options for PIMS may involve tighter coupling between components than discussed here, which could reduce initial costs but limit future functionality.

3.2.1 Content Repository

Content management (CM) systems are widely used in business and employ a wide range of commercial and open source products. A number of projects have implemented CM in a scientific context. Standardized programming interfaces such as the Java Content Repository (JCR) specification (JSR 170 and the JSR 283 update) and protocols such as WebDAV (http://www.webdav.org/specs/) are widely supported. (For example, the experts defining the JSR 283 specification includes over 20 companies and open source repository groups such as IBM, Oracle, EMC, SAP, Sun, Filenet, Day, Apache, Xythos and Alfresco, all of which have products supporting the JCR specification.) The advantages of these principles are evident in a number of cyberinfrastructure (CI) projects today. For example, the Collaboratory for Multiscale Chemical Science (CMCS) [29] is built upon a CM abstraction and supports automated metadata extraction, content translation, and provenance browsing. The MAEviz earthquake risk management cyberenvironment (http://mae.ce.uiuc.edu/software_and_tools/maeviz.html) is based on a similar abstraction and incorporates CM, workflow elements, and an Eclipse rich-client-based dynamic plug-in mechanism. Semantic content management systems are less mature, but a variety of tools that provide core capabilities are available from companies like Oracle and open source projects such as the National Center for Supercomputing Application's (NCSA) Tupelo[6] semantic content middleware [30]. Comb-e-Chem (http://www.combechem.org) is another leading example of a system in which all information from laboratory notes on tablet computers through analyzed results that are stored in reference libraries, and all of the provenance and metadata that links them, is managed using semantic repositories.

3.2.2 Workflow and Provenance Management

Basic workflow systems, like CM systems, are a mature technology with a wide range of commodity solutions from industry and academia. However, workflow systems that support semantic module descriptions are relatively rare [31]. Many workflow systems have some form of internal provenance recording capability, but standard models for provenance, such as the proposed Open Provenance Model (OPM) [32], and workflow systems that support them are only now emerging [33]. Similarly, while some workflow toolkits and CM systems with workflow capabilities provide mechanisms for running defined workflows on a periodic basic or in response to events (such as the arrival of new data), such capabilities are not currently available for systems with advanced semantic and provenance capabilities. Thus, workflow may be an area where the trade-off between development cost and long-term maintainability is very apparent and where the suitability of a given product will be dependent on choices elsewhere in the infrastructure.

3.2.3 Data Curation

As discussed in this report, data curation, the process of taking raw data and turning it into data to be retrieved by the end-user, includes a range of underlying procedures including input validation and cleaning, quality assurance/quality control (QA/QC), metadata extraction and data translation, indexing,

[6] Tupelo was initiated as part of the NEESgrid effort and is now being used in the development of digital watersheds for NSF's environmental observatories.

and the creation of derived data products. Given the broad range of data that will be collected in PIMS and the variation in the amount of validation and quality control that will be done before PIMS acquires these data, PIMS will need a very flexible curation subsystem that is based upon an underlying workflow and provenance capabilities. PIMS curation workflows involve simple checks to validate inputs against minimum metadata constraints and ensure that values are within anticipated ranges, workflows that implement manual expert evaluation or peer review, or complex processing that statistically assess the data in relation to other observations and modeled results. As data are ingested and assessed, they also may undergo cleaning and have embedded metadata extracted and indexed. Data ingestion also may trigger the creation of translated versions of the data as well as derived data products including summary reports and statistical analyses across data sets. Across the set of processes, PIMS should ensure that data products are linked with the original data, calibration information, and other inputs through provenance. The curation subsystem in PIMS must be capable of executing such an ensemble of processes, documenting them with provenance, and robustly handling any errors and/or alerting operators when manual intervention is needed.

3.2.4 Data Preservation

To ensure that collected data are useful over the next 50 to 100 years, PIMS must manage data preservation at several levels. At the most basic level, PIMS must ensure that the digital bits entered are not lost or corrupted over time as storage media fail and are replaced, as disasters strike, or as the PIMS faces cyber attacks. A wide range of mechanisms are available at this level ranging from relatively simple redundant arrays of inexpensive disk (RAID) systems to ensure that single or double failures of disks can be automatically corrected using redundant encoding of information across the arrays to provisions for mirroring the contents of entire repositories to remote sites that are available as part of CM systems or their underlying database and file system components. Common techniques such as recording a cryptographically signed and time-stamped "hash" of the data combined with periodic recomputation of the hash can detect changes in the data that then can be restored from back-up copies.

Long-term data preservation also requires preservation of the structure and accessibility of data as hardware, operating systems, and software services and applications evolve. Solutions at this level are usually custom-built following one of two design options, but they also may be provided as part of a CM system. In the first option, system operators strive to maintain working copies of old hardware and software in perpetuity or to keep old virtual machines alive. In the latter, operators manage a process of migration to move each type of file to newer versions over time and implement testing procedures per file type to ensure that conversions do not affect content. Neither is optimal. An emerging approach that attempts to combine the benefits of both involves the use of a format description language to characterize each file type in terms of its logical structure and the maintenance of generic (format-neutral) software that uses the descriptions to parse files and display them. Such an approach, which is currently being developed by the EU SHAMAN project [34] and its U.S. partners at the San Diego Supercomputing Center (SDSC) and NCSA, requires a minimal amount of defined software to be maintained and only a single format description document per file type. As development of the Data Format Description Language (DFDL) and the Defuddle parser for it [35] has shown, such description files can be used to support metadata extraction, translation, and the creation of human readable format documentation as well as for preservation. Parsing also can be made efficient enough for large-scale use.

At the most abstract level, PIMS will need to preserve the meaning of information over time. At the scale of 50 to 100 years, one must consider changes in the meaning of natural language terms and in what is considered common knowledge. For example, definitions of damage states may evolve as training or evaluation protocols change. Similarly, the limitations of old-fashioned sensors may be forgotten by younger generations familiar only with newer approaches. Addressing these concerns involves incorporating term definitions and ancillary data into the system and potentially transforming the data as is done to preserve structure to incorporate new terms or to make implicit knowledge abut terms explicit.

3.2.5 User Interfaces

Implicit in the discussions with many stakeholders is the assumption that PIMS will be a web accessible system and that the majority of PIMS functionality will be accessed via a web browser. Such an assumption makes sense because a web-based solution satisfies many underlying requirements for lightweight, universal access and because there are a wide range of web sites that demonstrate the types of functionality PIMS will require. As noted above, an SOA strategy would compliment web accessibility with a service layer that could be used to incorporate PIMS functionality into desktop tools or other services provided by PIMS or third parties. In other communities, such as NEES, web services are proving to be a critical complement to web form interfaces, with particular value in automating bulk transfers of data between desktop data acquisition and analysis software and the central repository.

A PIMS implementation created today should incorporate key technologies (e.g., Asynchronous JavaScript and XML, AJAX [36]) that allow different parts of a web page to be independently updated, providing functionality much closer to what is expected from desktop applications than traditional web applications where each user interaction with the page requires a refresh of the entire page. There are a number of toolkits available in this area including popular ones such as the Google Widget Toolkit (described at http://code.google.com/webtoolkit/). A PIMS implementation also might incorporate a portal framework to support aggregation of multiple independent interface components onto the same page, to support management of group security and preference settings, and to provide basic collaboration and communications mechanisms such as wikis and blogs to the community. Alternatively, PIMS might employ newer technologies such as Google Gadgets (see http://www.google.com/webmasters/gadgets/). These would allow users to place PIMS gadgets for data download on their personal pages alongside gadgets from other organizations (see, for example, http://www.google.com/ig). While such techniques are less mature, they allow tremendous flexibility for creating "mash-ups" that integrate functionality from multiple providers, which in turn can drive community adoption and harness users' creativity in developing new capabilities.

Since much of the information that PIMS will gather is either spatial in nature or about spatially distributed infrastructure, a PIMS implementation will need to harness recent innovations for the display of spatial information in the browser. Map interfaces available from companies such as Google and Microsoft are probably the most well-known of these new capabilities. Through the use of a standard input format (e.g., the Keyhole Markup Language, KML), new layers of information can be added to maps showing satellite images, road networks, buildings, or other infrastructure provided by the underlying map service. Using such maps, one could overlay ground motion estimates and provide pop-up links to show structure-by-structure observations of damage. Maps also can be used as input controls enabling users to specify a region of interest as part of a search for data. Along with map interfaces, a number of standards for managing geospatial data on the web are emerging from such organizations as the Open Geospatial Consortium (http://www.opengeospatial.org/). Incorporating such standards into PIMS would allow use of existing open source and commercial components for spatial data management.

Although the web provides a very simple way to access information, displays within browsers are still limited compared to rich desktop applications. One obvious example is the difference between Google Maps displayed in a browser (Figure 3.2) and the 3-D interactive capabilities available in the downloadable Google Earth software (Figure 3.3). Further discussion may be needed to decide whether PIMS should support richer desktop capabilities in addition to web interfaces. Third parties may be a more natural source of such tools but, in either case, PIMS should consider providing web service interfaces that can be used within other programs that offer equivalent capabilities to those available in its web user interfaces.

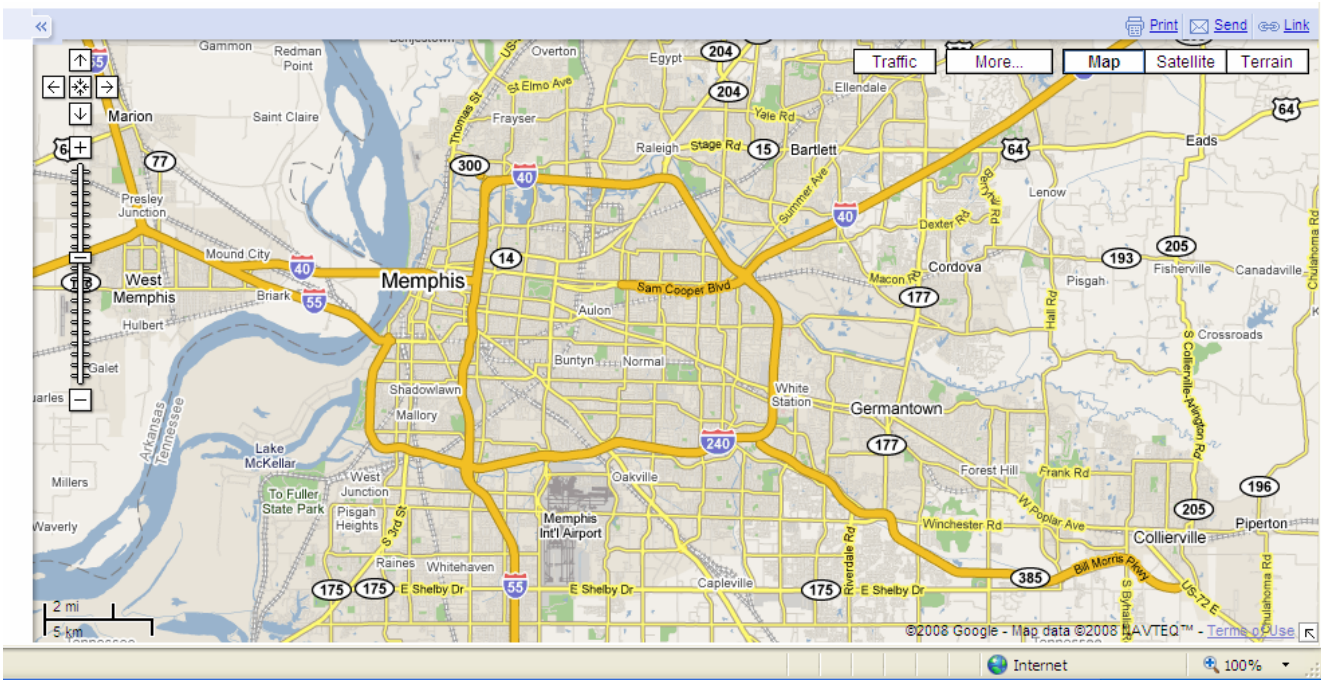


Figure 3.2 Google Map interface showing view of Memphis, Tennessee.

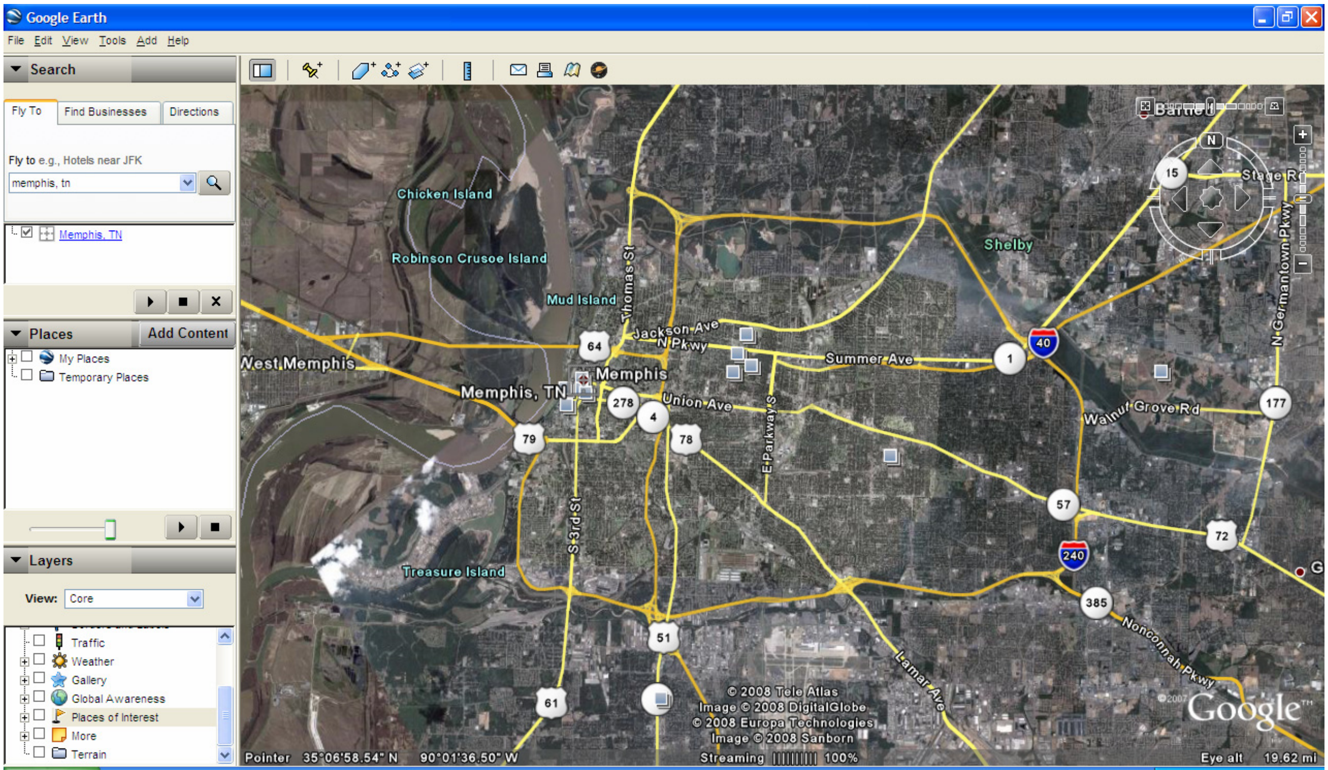


Figure 3.3 Google Earth interface showing view of Memphis, Tennessee.

3.2.5.1 Data Ingestion

PIMS is expected to acquire data from both direct community input and through agreements with other data repositories. Further, PIMS may acquire data from both trained observers and "citizen scientists" who may be unfamiliar with observation protocols. PIMS also will handle a wide range of data and will deal with metadata entered manually as well as metadata embedded in file formats. Thus, PIMS will likely require multiple types of ingest (data entry) interfaces focusing on different types of infrastructure and targeting different classes of observers. (One also might consider these as a single interface that dynamically adapts based on the data types and user.) While the most obvious way to implement such interfaces may be to create long web forms and to require users to complete all fields, experience within NEES and more generally throughout the digital library community confirms the common sense conclusion that requesting large amounts of metadata via forms can be a significant barrier to use. Thus, more sophisticated approaches that would allow for incremental additions to metadata over time (i.e., keeping track of a user's entries in progress in assembling required metadata and allowing them to complete entries at a later date), provide default values drawn from previous entries or from observational protocols, allow entry of metadata that applies to multiple data sets at the level of projects, and provide scaffolding (just-in-time training) on best practices for entering information will be important for the long-term success of PIMS. Similarly, mechanisms that track data entry and report statistics on the amount of information uploaded per person and per organization, as well as feedback on the use of uploaded data, also can encourage data submission and lead to broad adoption.

Many PIMS data sets will relate to others. For example, one user may upload a schematic for a bridge while another uploads a damage report for the same structure. Thus, PIMS will need interfaces that support the creation of relationships between data sets and that allow users to discover and use identifiers for known infrastructure and other information in PIMS. Such interfaces may combine elements of the data discovery interfaces discussed below.

PIMS also may need interfaces to support community involvement in processes such as data curation (e.g., that present newly entered information for peer review or enable community members to associate additional data or metadata with existing entries). Such interfaces will likely be required as part of the administration tools available to operators to support editing to correct errors.

3.2.5.2 Data Discovery/Export

The second major focus for PIMS interfaces is to support discovery (the process wherein a user finds data) and retrieval of data held by the system. As noted in Chapter 2, different users may wish to retrieve data based on spatial region, event type, class of infrastructure, or any number of other metadata attributes. Thus, PIMS will require general search and browse capabilities that allow for sophisticated inquiries and precise retrievals. However, the success of modern web search engines shows the utility of simple discovery mechanisms that may have less precision. In PIMS, such an approach could be manifested as a single text entry search box or, as many stakeholders have suggested, a map-based interface that allows spatial filtering of data. As with other map-based services, a filtering metaphor, where users can select/deselect various categories of information to be displayed (e.g., based on structure type, information source, event, tags, etc.), could potentially support the bulk of users without requiring use of a more structured (e.g., form-based) definition of a query. Given the general nature of the underlying PIMS infrastructure, an implementation could potentially support multiple user interfaces targeting different sub-communities or providing "experimental" capabilities that could be dropped or moved to production based on their adoption. One potential area for exploration along these lines would be the development of interfaces that exploit provenance information to allow users to find all inputs used to draw specific conclusions or to find all information derived from a given source.

Once specific sets of data are identified, users are primarily expected to retrieve the data for use in other applications. PIMS will likely need to support both ”raw” downloads of information where users retrieve information at the same granularity in which it was entered (e.g., retrieving files and associated metadata) as well as providing support for online viewing and retrieval of summary and statistical information as discussed in the next section. Raw download capabilities should include options to translate data and metadata into an evolving set of default formats whose selection is driven by community interests. The export interface of PIMS should present the available options to users and, depending on the trade-offs between computational and storage costs, the infrastructure should either retrieve pre-computed translated products or perform the required translations dynamically.

3.2.5.3 Analysis and Visualization

While further discussion may be needed to define the appropriate scope for PIMS in terms of the level of analysis and visualization services to be provided, some capability is clearly needed for creating summary information and statistical (and anonymized) derived data products as well as preview-style visualization capabilities. In addition to being able to overlay information on maps as discussed previously, PIMS will need to be able to show information in x-y graphs and tables and to generate previews of images, movies, 3-D structures (Figure 3.4), and other information. It is likely that commodity components can be found and incorporated to support these needs, but development work may be required if significant interactivity is required within individual visualizations (e.g., to rescale graphs or fit points in a graph to a curve) or between them (i.e., to highlight entries in a table that correspond to selected entries on a map or in a graph). Such interactivity is possible today as can be seen in web-based digital watersheds being developed by NCSA and others or the environmental observatories (see http://his.cuahsi.org/ for an example).

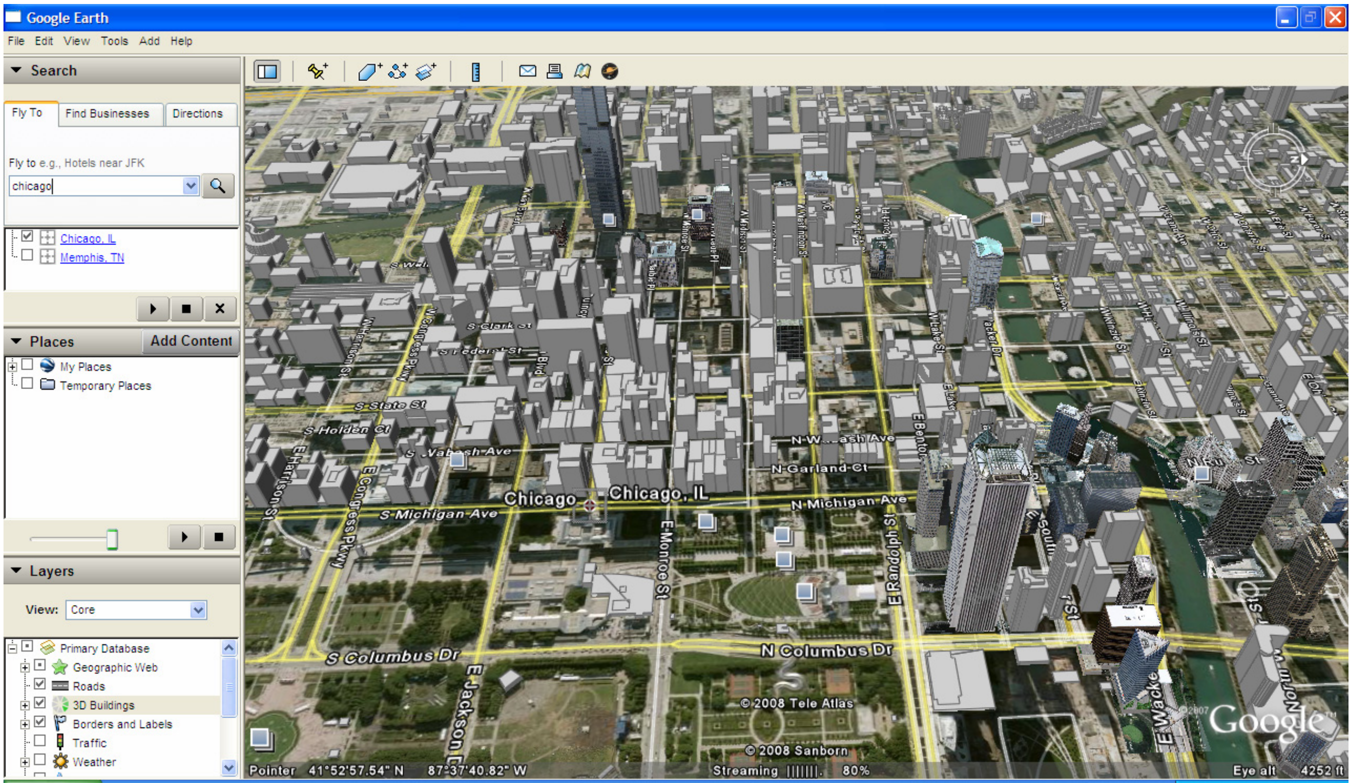


Figure 3.4 3-D building view of Chicago (Google Earth).

3.2.6 Security (Authentication, Authorization, Auditing)

As a national resource, PIMS must adopt evolving industry best practices to protect its data and computational resources against physical and cyber threats through appropriate access controls and intrusion detection and response measures. In doing so, PIMS must balance security with ease of access. Security measures should not impede access to public data but must effectively limit access to sensitive data. As noted above, PIMS will need to restrict access to different classes of information to particular groups of users; therefore, security mechanisms are likely to involve some form of role-based access control (RBAC) and would benefit from single-sign-on (SSO) and distributed credential management capabilities. With these technologies, a user would enter a password or other proof of their identity, would have one or more roles (e.g., public user, member of Project ‘X’), and would have access to functionality and data within the system based on that role (e.g., data items could be accessible to “all public users” or restricted to “members of Project X” as desired). Applications of these and related technologies in the web and grid communities include the Liberty Federation SSO framework (http://www.projectliberty.org/), the security assertion markup language (SAML) [37], and Shibboleth federated identity management middleware (http://shibboleth.internet2.edu/). If PIMS is required to dynamically forward user request for information to other systems, credential delegation mechanisms developed for grid computing such as MyProxy [38] also may be relevant.

In addition to technologies focused on authenticating and authorizing users, PIMS security infrastructure must address capture of auditing information (to capture the actions of malicious actors), encryption of data in transport and on disk, physical and cyber monitoring, and restricted-access networks. A wide range of technologies are available to implement this functionality, and there are a number of de facto standards such as the HTTPS protocol for encrypted communications that are widely used in industry that can be adopted.

3.2.7 System Administration

PIMS, as envisioned in this report, is a highly functional, multicomponent system that will require significant automation of operations, administration, and maintenance activities to be cost-effective. Managers of the system will need interfaces to support system-wide configuration, monitoring of its health status, analysis of its use (i.e., reporting capabilities), periodic testing of new and existing components, and recovery from component failures. More detailed functionality will be required for each component (i.e., to define metadata requirements within the CM subsystem and to implement and execute new workflows to create new derived data products or implement new back-up, curation, and preservation procedures). While the full specification of system administration functionality is dependent upon the particular technologies selected for an implementation, the same general categories specified at the system level will apply at the component level. Publication of component level information through common protocols such as the Simple Network Management Protocol (SNMP) (http://www.ibr.cs.tu-bs.de/projects/snmpv3/) and general publish/subscribe messaging frameworks such as the Java Messaging Service (JMS) (http://www.jcp.org/en/jsr/detail?id=914) will be required to enable overall system-level capabilities.

State-of-the-art design principles are employed to develop a high-level architectural framework for PIMS that addresses the content repository, workflow and provenance management, data curation, data preservation, user interfaces, security, and system administration. This framework should provides a firm foundation for subsequent development.

3.3 Policies and Procedures

A significant effort will be required to develop community consensus policies and procedures that can then be implemented in the software system. With design strategies specified, these efforts can proceed in parallel with system development; further, the system can be made operational before or without complete agreement between all stakeholders. While agreement is highly desirable and can help drive community adoption, the ability to support divergent views and to incrementally improve standardization over time can help avoid the technical and community issues associated with premature standardization (i.e., a system that cannot meet the needs of all users or fracturing of the community into opposing camps).

As a long-term national resource, PIMS will play an important role in the community. While PIMS will likely not be large enough to impose standards on the community, it should be able to initiate discussions of standardization when it would benefit the community and to influence community adoption of standards through its endorsement of specific standards as supported defaults. The following subsections highlight specific areas where policies and procedures will need to be developed with the community.

3.3.1 Data and Metadata Standards

As described in this report, PIMS will be capable of supporting data in multiple formats and arbitrary metadata, but its practical utility will depend on significant standardization by the community. Most critical for PIMS is standardization of the core metadata necessary to support data discovery. These include geospatial metadata, structure classification metadata, collection protocol metadata, bibliographic metadata (including attribution), and provenance. In many of these areas, PIMS should collaborate to develop domain-independent standards (coordinated across many science and engineering fields) whereas, in others, PIMS may take a leadership role or partner with other hazard community organizations to develop standards.

Data format standards for file-oriented transfer of data to and from PIMS are also important. With the design discussed, PIMS can easily support industry standards such as ESRI Shape files, GML, and KML and can provide capabilities to translate between them and to map into them formats that may be specific to particular subdisciplines. PIMS will significantly reduce tensions related to standardization and prove more productive than a mandated, top-down approach if it supports an organic approach to standardization, allowing each subdiscipline to develop its own techniques and to adapt them quickly to unique circumstances and priorities emerging from each disaster, yet also providing the means to map that information to emerging community standards over time. [18]

3.3.2 Data Privacy and Security

As with metadata and data standardization, decisions about the policies and procedures related to data privacy and security issues represent trade-offs between costs and benefits. In the case of privacy and security, these trade-offs are strongly influenced by community norms and, hence, privacy and security policies and procedures are another key area for community discussion. Physical storage and protection mechanisms and operating procedures can draw heavily on existing security policies, procedures, and mechanisms in use in other disciplines and can be informed by ongoing national and international community discussions on best practices within projects such as the TeraGrid and within organizations such as the Open Grid Forum (OGF). Data access and privacy policies can similarly draw upon efforts in other areas of science and engineering (e.g., the Science Commons, http://sciencecommons.org/) but will need to be tailored to the specific circumstances and concerns in the PIMS community. PIMS will manage access to a wide range of data and associated software, documentation, workflows, and related services. Some PIMS data will be appropriate for public release, while other information may be proprietary and/or sensitive due to national security or privacy concerns. PIMS also may provide derived data products that summarize or anonymize information such that derived products may be releasable more widely than the original data. Further, metadata generated by the system (e.g., usage information) may require protection. Given this range and the differing amounts of control the community and individual groups will have over these various types of information, a single static, blanket policy is unlikely to serve the community. PIMS should instead seek to standardize major categories of data and other resources and implement appropriate default policies and procedures developed in collaboration with stakeholders. These policies and procedures should also be vetted with legal experts representing stakeholder interests.

3.3.3 Persistence of Data

Required policies and procedures related to data persistence should incorporate best practices from other communities to deal with issues related to such things as disk failures, power and network outages, and intentional data corruption and should address specific scenarios of concern to PIMS stakeholders such as assuring data collection in the event of earthquake related network outages or failures of externally managed data collection systems. While basic infrastructure such as RAID systems contributes to the solution, policies and procedures must be in place to monitor proper system function, to require periodic testing of equipment and failure recovery procedures, and to require periodic reassessment of overall operations. Analogous policy and procedure development effort will be required for longer-term preservation functionality related to format translations and preservation of semantics.

PIMS is a long-term national resource that will play an important role in the community, serving as a catalyst for development of effective policies and procedures.

3.4 Community Adoption of PIMS

PIMS management must be strongly focused on providing immediate and lasting value to its stakeholders. The amount of effort that will be required to maintain community engagement throughout the planning and development process and to provide value to individual users and organizations that contribute to PIMS operations or use its services should not be underestimated. As acknowledged in training workshops for large-scale infrastructure managers such as the Project Science series (http://www.projectscience.org), failure to recognize the magnitude of "social" issues in the development of large infrastructure is one of the major factors cited by managers of large projects as contributing to project delays, cost overruns, and low adoption rates. To maximize its success, the PIMS project should be conducted in a very public manner and should provide value to individual contributors and users as well as to the community as a whole. For example, providing reports that summarize the contributions made by individuals or reporting usage resulting from a specific individual's contributions, while not strictly necessary to fulfill PIMS's mission, may be critical in fostering adoption. Similarly, leveraging the PIMS infrastructure to provide a guide to other community capabilities and to support community collaboration and outreach efforts, while not strictly necessary as part of providing long-term access to data, ultimately may focus attention on PIMS and lead to more use of PIMS data. While PIMS should not endeavor to be all things to all people, continuing dialog with stakeholders and attention to providing coherent services that support use of PIMS data should be core elements of a PIMS effort.

PIMS also will need to maintain a focus on relationships with data providers. While the sponsors of PIMS might use their powers to require private organizations to provide their data (e.g., via a government requirement to collect and provide data as a condition for the renewal of government-controlled licenses and awarding of research grants), other potential solutions include paying the information provider for the data to offset the loss of competitive advantage and the use of nondisclosure agreements wherein the information provided is distributed selectively to users. Further, PIMS may be able to offer incentives to data providers, including guaranteed embargo periods, preferred access to data from others, value added services for data entry, data translation and/or analysis, the minting of Digital Object Identifiers (DOIs) or other forms of persistent identifiers useful for data citation in papers, and a neutral, precompetitive forum for industry. PIMS might even function as a value-added reseller of industry data, increasing data value through aggregation and enhanced services. In considering these options, the role of PIMS as national infrastructure and its ability to harness the network effect should provide significant leverage.

Community engagement throughout the planning and development process is essential; an important key is to provide value to the individuals and organizations contributing to PIMS operations and using its services.

3.5 Assessment of Requirements Traceability Matrix

A high-level requirements traceability matrix (RTM) for the PIMS infrastructure effort is provided in Appendix G. This matrix, despite its length, is relatively coarse-grained compared to what will be needed to guide development during an implementation project. It is nonetheless useful to provide an overview and introduce the concept. As noted, an RTM is one mechanism used to guide development efforts. In particular, an RTM is designed to assure that all requirements are met by identified components of the system and that all components in the system are necessary to support one or more requirements. Any requirement that does not map specified infrastructure indicates that additional components are required. Conversely, any components that do not map to identified requirements indicate either that there are implicit requirements that are not yet well defined or that the system is overdesigned. An RTM can be evaluated to interpret overall system interactions as well (e.g., clusters of features often map to particular subsystems in the architecture, and cases where requirements map across subsystems indicate areas where coordination between developers of those subsystems will be required).

Use of a requirements traceability matrix (RTM) to guide development efforts will ensure that all requirements are met by identified components of the system and that all components in the system are necessary to support one or more requirements.

3.6 Summary

PIMS development will pose both engineering and research challenges. While meeting the hard requirements for network bandwidth and storing and serving data are primarily an engineering challenge (cost/benefit analysis of existing technologies, deployment and professional operation of core infrastructure), the less quantitative requirements for PIMS to effectively serve a broad range of stakeholders with varying degrees of technical sophistication will require both a research mentality (e.g., analysis of current research processes, experimentation with user interfaces, and iterative development based on community feedback) and significant systems engineering to harden emerging technologies before widespread community use. Perceived ease of use, acceptance of PIMS by the broad community, the community's sense of ownership of PIMS and its mission, and the ultimate success of PIMS depend on success in meeting both types of requirements. Understanding when and where an engineering versus research approach will be needed and structuring the project to focus research and development efforts to feed into engineering decisions will be an important in managing costs and ensuring that PIMS is both robust and responsive to evolving user needs.

4 DEVELOPMENT, OPERATION, AND OVERSIGHT

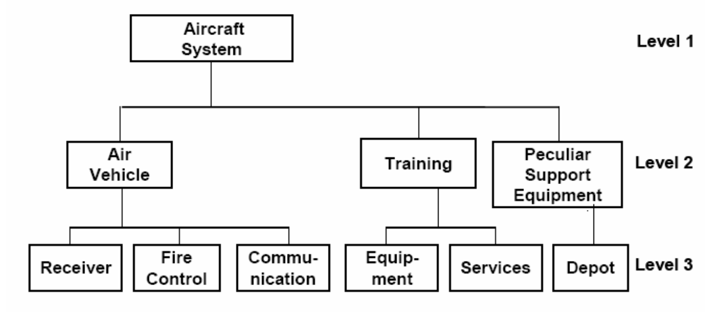


Figure 4.1 Simple example work breakdown structure (Wikipedia).

Moving from the conceptual PIMS design outlined in this report to a detailed implementation strategy for the system will involve a number of activities to refine the anticipated scope, evaluate and select technologies, and develop a formal project plan. PIMS can leverage and use relatively standard tools and techniques for project management but will need to train the community tasked with developing PIMS to use the tools. These tools include use cases, requirements traceability matrixes, work breakdown structures (Figure 4.1), and architectural and component interaction diagrams. Decision matrices, which are helpful in analyzing the cost-benefit ratios of alternative designs and of individual subsystems, also will be very relevant. Use of these tools will help ensure that all science requirements can be met by the infrastructure, that all elements of the infrastructure are justified by their value to the science, that risks are identified early and actively managed, that all development tasks and decision points are identified, that components are specified well enough to integrate into the overall system, and that designs are appropriate to handle the anticipated scaling and evolution of the driving scientific needs over time.

Adoption of an overall framework for project management, such as the Department of Defense Architecture Framework (DoDAF)[7] selected by the Ocean Observatory Initiative (OOI), would be one mechanism for identifying and producing a coherent set of documentation. A more organic approach like that recently used in National Ecological Observatory Network (NEON) also has merits in terms of limiting the new jargon that must be introduced to the community and increasing community ownership of the infrastructure being created.

PIMS can leverage and use relatively standard tools and techniques to aid its development, including use cases, work breakdown structures, requirements traceability matrixes, and system interaction diagrams.

As planning and development for the PIMS specification proceed, several key concepts should guide the community and the project team:

- PIMS should integrate relevant resources managed by other projects and organizations, co-exist with the range of other post-event data collection and archiving efforts, and live in the context of the broad development of national cyberinfrastructure for science and engineering (i.e., learn from other CI projects, incorporate methods and technology used in the projects, present PIMS findings to the national CI community, etc.). Interoperability and re-use of designs, components,

[7] See http://en.wikipedia.org/wiki/Department_of_Defense_Architecture_Framework for an overview.

and standards across this ecosystem is highly desirable both in terms of maximizing the value to the community and minimizing development and operations costs.

- Requirements and the system scope for PIMS should be specified not only in terms of functionality but also in terms of when that functionality will be needed. For example, the storage capacity and computational power to deal with 50 years of data are not required at the beginning of operations, and it would be wasteful to invest in hardware resources too far in advance of when they will be needed. Similar issues exist at the software level. Building software with well-defined interfaces that allow for more capable components to be developed and integrated over time can be much more cost-effective than developing all the desired functionality up front. Recognizing this allows for a design that is scalable but not initially scaled. Further, it will enable PIMS to begin operation much sooner and for plans for future enhancements to be reassessed based on operational experience. It should be expected that the amount and variety of data, the number of users, and the sophistication of analysis, modeling and visualization required will all increase during the project lifetime. Modular and flexible design strategies as outlined in this report can help ensure that PIMS will be able to evolve to meet changing needs and technologies.
- Planning should include decision points at which key requirements will be frozen and technology decisions made. Such a strategy is critical in incrementally resolving dependencies in the design. For example, specification of the maximum data volume that must be managed and the maximal system load (number of contributors and users) as a function of time allows evaluation and selection of computing and storage hardware and network infrastructure (or required levels of virtual machine services) and the development of testing procedures to ensure the system meets the specified level of performance. Similarly, decisions about programming languages and selection of major system components feed into hiring decisions and the selection of additional system components.
- Pilot efforts and investigation of technologies from other projects and the larger cyberinfrastructure community should focus on areas where there are significant risks and opportunities and/or long lead times. Planning should identify areas where PIMS requirements exceed the limits of current technology or go beyond the scope of what has been deployed in other projects and where costs escalate dramatically for higher levels of functionality. It also should identify areas such as data format standards and query protocols where long lead times result from the need to negotiate the standards and evangelize their adoption and for third parties to implement them in community software. In such cases, PIMS should launch or participate in activities to gain sufficient experience in these areas to enable timely, informed decisions.
- Given its long operational lifetime, PIMS must be considered to be a living system that will be iteratively improved. Initial PIMS functionality should target a coherent subset of the overall scientific goals of the full system (i.e., it could start by simply federating existing data available from regional and sub-discipline specific repositories and provide for archiving and basic data discovery and retrieval). As community support grows and experience is gained, PIMS can employ more advanced tools and features and might expand to incorporate data relating to other types of hazards. This type of incremental development should be pursued both within an initial development project and through periodic technology refresh efforts throughout the life of the PIMS.

Several key concepts have been identified to help guide the community and the project team in further developing and implementing PIMS.

4.1 Phased Development of PIMS

A proposed phased development approach to PIMS is introduced below in terms of overall considerations based on past experiences and efforts, proposed phases of PIMS development, an estimate of required resources, and a discussion of how PIMS might support the community during development should a major earthquake happen in the next few years.

Community-scale projects often have significant planning and development phases that can result in overall timescales from project initiation to mature operational capability of 5 to 10 years. Such a timescale is typical for major NSF initiatives (e.g., the proposed national environmental observatories). To some extent, this timescale reflects the time required for change in community practice and, thus, would be applicable to the development of a fully functional PIMS (described as Phase 2 below). However, given that an initial PIMS capability (described as Phase 1 below) could be better described as supporting current practice at a national scale rather than as requiring the development of new practices, development of core capabilities should be possible in a shorter period of time. Given the maturity of technologies necessary for PIMS and the current state of planning within the community, it is reasonable to assume that an initial PIMS capability (PIMS Phase 1) could be implemented within two years, while a fully-functional PIMS (PIMS Phase 2) would take 5 to 10 years. Although the technical capabilities could be implemented more quickly through either a series of pilot projects or a skunk works style effort, a more compressed schedule would increase risk and potentially lead to a perception of technology push (i.e., "build it and they will come") within the community.

Although community scale projects similar to PIMS typically have timescales from initiation to operational capability of 5 to 10 years, an initial PIMS capability (PIMS Phase 1) could be available in approximately 2 years.

The development of PIMS represents a significant engineering effort and, if serious consideration is given to implementing a system capable of cost-effectively preserving data for 50 to 100 years, one with a significant research element. In terms of data volume and breadth of input sources and data types, PIMS is similar to NSF's national ecological observatory network (NEON) development efforts. Initial plans for NEON, for example, estimated raw data volumes of up to 5 terabytes per year with up to 50 terabytes per year required in total (raw data plus, for example, metadata and derived data products). At this level, hardware costs are minimal relative to labor costs. For example, commodity disk drives cost less than $200 per terabyte today and servers capable of supporting multiple virtual machines are available for well under $10K. Based on these costs, the overall hardware infrastructure for an initial PIMS could cost less than 1 FTE (full-time equivalent) of labor. Thus, PIMS's resource requirements are largely for labor associated with management, requirements analysis and community engagement, design, development, operations, and user support. While detailed analysis of resource requirements requires further definition of the PIMS scope, a comparison with efforts such as NEES, NEON, and OOI suggests PIMS would require a full-time project leader, one to two FTEs of effort in requirements analysis and community engagement, a project manager, one FTE of administrative support, three to five FTEs for design and development activities, and two to three FTEs for testing, operations (including data curation), and user support. Such an estimate assumes that the mix of activities will shift during the course of the effort (i.e., that requirements analysis will ramp down over time as other forms of community engagement such as training increase). Thus, based on a comparison with existing systems, a PIMS effort would involve an annual cost of roughly 9 to 13 FTEs with additional funding for infrastructure, travel, workshops, advisory board activities, and overhead. Additional FTEs would be required for PIMS to play a significant role in community standardization efforts or to be involved in research and development of advanced functionality, although research efforts potentially could be funded as part of cross-disciplinary cyberinfrastructure research and development efforts such as those funded through NSF's Office of Cyberinfrastructure. While the comparison with other projects is subjective, it is roughly consistent with, though slightly larger than, the more direct estimate derived below (assuming all parts of the phased plan would be pursued) and thus serves to indicate that the direct estimate is reasonable but may not include all costs that a more detailed analysis would identify.

Consistent with the concept of PIMS as a catalyst that evolves with community use of PIMS data, the consensus among stakeholders is that the development of PIMS must be phased and that value must be added in each phase in proportion to the resources required to develop that phase. To accomplish this, the required functionality of PIMS and goals for PIMS development for each phase must be clearly articulated and metrics for tracking this development must be identified. The cost-benefit ratio of PIMS functionality developed in each phase should be quantified (e.g., by using a decision matrix methodology that compares the value obtain from meeting a requirement with the incremental cost of meeting it), and those capabilities of the system that provide the greatest cost-benefit ratio should be developed first. Moreover, it is important that the required abilities developed in each phase be attainable so that the development of each phase is successful. Thus it is wiser to develop a Phase 1 PIMS with lesser capabilities than to set high goals and not be able to accomplish them.

As noted, the development of PIMS is envisioned as occurring in two phases. Phase 1 would be completed within two years and would include development of an initial PIMS capable of harvesting data from a few key sources, basic ingestion and archiving capabilities for hazard events in the near future, and a simple interface to provide for data discovery and retrieval. Phase 2 would take from 5 to 10 years. It would be informed by Phase 1 but would involve development of a more advanced, "full-function" PIMS capable of harvesting data from a wide variety of sources, providing advanced tools for ingestion and archiving, and offering sophisticated user interfaces for data discovery and retrieval.

The consensus among stakeholders is that the development of PIMS should be phased with Phase 1 resulting in an initial core system capable of some basic system requirements, and Phase 2 resulting in a fully functional PIMS. This incremental PIMS development process will reduce technical risk, foster community adoption, and ultimately maximize the community return on the investment in PIMS.

4.1.1 Phase 1

As noted above, development of the Phase 1 PIMS would require approximately 24 months. The key goals for the initial system are presented below.

1. Integrate selected major existing collections of data in specified areas (e.g., for selected classes of buildings or bridges):

   a. Begin by collecting public sector data that are readily available and then move toward including private-sector data over time, which may be more complicated, time consuming, and controlled than public data.
   b. Negotiate agreements with current data holders concerning data management (whether PIMS will replace current store, serve as a backup, work as an independent harvester of public information, etc.) and the data in question (public, private, number of types, metadata existence, existing search mechanisms, etc.).
   c. Collaborate with existing systems such as ShakeCAST and IRIS Data Management Center and discuss leveraging their ideas and technology and obtaining their data.
2. Provide operational capabilities for ingestion and preservation of data for the next significant hazard event:
   a. Define core required metadata – develop web forms and/or extractor mechanisms to acquire metadata.
   b. Be ready to capture data as they present themselves (e.g., NEES data may need to be captured and saved for ultimate use in PIMS).
3. Provide for basic discovery and retrieval of data:
   a. Geo-coded data and a map-based interface are required for this phase.
   b. Advanced analysis tools (e.g., for the creation of derived data sets) are not required in this phase. Phase 2 and future phases will consider analysis tools.
4. Test PIMS using historical data:
   a. Test PIMS ingest, harvest, archiving, and discovery/retrieval capabilities.
   b. Possible historic earthquakes include Northridge, San Fernando, and Loma Prieta(advantage of recent earthquakes is the availability of good digital earth science data).
   c. Select one recent earthquake and use it as a model for the rest.
5. Test PIMS with an exercise:
   a. Purposes
      i. Practice mobilization of data collection teams, enactment of procedures, and interactions between organizations
      ii. Demonstrate the national value of PIMS
      iii. Develop list of data that is desired in the full PIMS
   b. The 2011 central U. S. magnitude 7.7 event is one possibility, but it must be noted that this exercise will concentrate only on the short-term clearinghouse components of PIMS (i.e. coarse-grained data).
   c. The exercise has limitations because it is an "exercise" – the actual data produced will be unlike a real event.
   d. PIMS may or may not begin development soon enough to meet this exercise's planning cycle.

PIMS Phase 1 would result in a system capable of harvesting data from a few key sources, basic ingestion and archiving capabilities for hazard events in the near future, and a simple interface to provide for data discovery and retrieval.

Table 4.1 Phase 1 Milestones

1	Final requirements and overall design (data sources, major technology choices, list of interfaces, etc.)	3rd month
2	Initial user interface designs (mock-ups reviewed)	6th mo.
3	Ingest mechanisms defined/draft Memorandums of Understanding (MOUs) with partners	6th mo.
4	Initial repository (for internal use in development)	9th mo.
5	Interface demonstration on mock data	12th mo.
6	Scaling tests to 20 TB	15th mo.
7	Ingest mechanisms developed (ingest demonstrations)	15th mo.
8	Initial system test (alpha/prototype period) – test with historic data	18th mo.
9	Public beta – test with exercise	21 th mo.
10	Operational system	24 th mo.

An estimate of the development costs for PIMS Phase 1 is presented below. Two alternative designs are presented in the estimate: XML/webDAV/standard Content Management System versus a RDF/Semantic Content Management System. This provides flexibility to choose between a less expensive system with limited future flexibility (i.e., the XML/webDAV/standard Content Management System) as compared to a more expensive (up-front) system with greater future flexibility (i.e., the RDF/Semantic Content Management System). Also note that the cost estimate does not include funds for data collection for PIMS. These costs are significant, and this issue is discussed in detail in Section 4.2.3. Although a more precise and detailed cost estimate will need to be prepared before PIMS implementation begins, the cost estimate below provides an initial analysis of the costs associated with the proposed PIMS Phase 1.

- Personnel
  - Project lead - 0.5 FTE (~$120K[8])
  - Administrative assistant/outreach/community logistics - 0.5 FTE (~$70K)
  - Programmers
    - Option A — XML/webDAV/standard Content Management System (relational database back-end) – 2 FTE (~$280K)
    - Option B[9] — RDF/Semantic Content Management System - 3 FTE (~$420K)
  - System management - 0.50 FTE ($70K)

---

[8] Personnel costs are calculated as base salary * FTE * 2.0 (fringe and overhead estimate).
[9] Option B is preferred, as it provides a stronger foundation for development of the fully functional system.

    - Testing and Documentation - 1.00 FTE ($100K)
- Equipment
    - $30K – includes 20TB+ RAID drive array, redundant dual core servers, space charges
- Travel - $25K
- Commercial software licensing fees and/or open source support contracts - $100K
- Total Development Costs (salary costs per year * 2 years + equipment costs + travel+licensing):
    - Option A -- $1,435K
    - Option B -- $1,715K

PIMS Phase 1 would require two years to complete and would cost between $1.4 and $1.7 million. This basic PIMS will be able to grow and is expected to help catalyze additional data collection, community standardization efforts, and the development of projects that use PIMS data.

4.1.2 Phase 2

Phase 2 would occur over a 5- to 10-year period. Its goal would be a PIMS capable of the required functionality initially envisioned and outlined in this scoping report. Although the schedule for this phase is not outlined explicitly herein, it can be described in terms of a number of key tasks or pilot projects. Each pilot project would be divided into a design phase during which the specifications for the particular component of PIMS addressed by the pilot project would be developed and an implementation phase during which the system component would be built. The design phase of each pilot would last approximately six months and would require one FTE of resources plus workshop funds, leading to a total cost of approximately $90K for the design phase. The implementation phase would last from 12 to 18 months and would require 1 to 2 FTEs of resources, leading to a total cost of $140-$420K for the implementation phase.

PIMS Phase 2 would result in a PIMS capable of the required functionality initially envisioned and outlined in this scoping report and would require between 5 and 10 years to complete. Although the schedule for this phase is not outlined explicitly, it can be described in terms of a number of key tasks or pilot projects.

Phase 2 development will involve the following pilot projects/tasks:

- **Preservation —** Development of mechanisms for managing format translations over time, procedures for validating translations, and policies about keeping original data.
- **GUI (graphical user interface)/user input from the field —** Development of mechanisms, procedures, policies to allow PIMS to support field reconnaissance missions via semi-automated means (e.g., via PDAs, using store/forward mechanisms to avoid network dependence, allowing post-trip additions/editing of info).

- **GUI – advanced user interface —** Incorporation of more sophisticated search queries, additional analysis/statistical summarization capabilities within PIMS, and more advanced visualization and export capabilities. Individual disciplines may be engaged in this pilot to help develop tools related to their needs.
- **Open protocol for harvesting from anywhere —** Standardization of the mechanisms, policies, procedures for incorporation of data from additional sources; taking the set of agreements developed in the initial phase and creating a small set of standard agreements; defining a common protocol for harvesting/ingesting data; defining web-service interfaces for programmatic querying of PIMS from other systems; and conferring with organizations such as USGS and EERI on the development of national post-hazard data collection plans and standards.
- **Management of proprietary data —** Conduct of workshops and negotiations discussing development of mechanisms, policies, and procedures to support integration of proprietary information within PIMS, including definition of access controls (what classes of data exist and who can access each class) and data security (determining whether PIMS will redirect to holders of proprietary data or will develop security mechanisms such as encryption on disk, robust intrusion detection and activity monitoring, to hold the data).
- **Community curation —** Development of the mechanisms, policies, and procedures needed to support curation/validation/quality evaluation of data from sources of varying quality within PIMS (e.g., development of standard curation procedures and a validation test, user feedback mechanisms to report quality issues, branding of different levels of data (PIMS-certified through caveat emptor)).
- **Oversight and managemen**t **—** Identification of the long-term oversight and management structure for PIMS, including leading organizations. The associated FTEs for this pilot may involve people with skills different from those involved in the other pilots because this pilot does not involve the design and implementation of software but rather policies and procedures.

A few more pilot projects/tasks are probably will evolve during Phase 2 relating to enhancing support for new data types, making international connections, managing privacy issues, etc.

Operations of the new capabilities developed in the pilots as part of the core PIMS system would require integration and hardening that would require additional development per pilot that is collected as a 1 FTE ($140K/yr) addition to operations costs to the overall project team throughout Phase 2. This raises operations costs to an estimated $600K/yr during Phase 2. Following Phase 2, operations costs are estimated at approximately $460K/yr. As with PIMS Phase 1, this cost estimate for PIMS Phase 2 is only an initial analysis and does not include funds for data collection; a more detailed and precise analysis of costs will need to be performed before implementing PIMS Phase 2.

Phase 2 Costs Summary

- ***Phase 2 Development Costs (per pilot project)***
  - Development Phase – 1 FTE ($90K/yr)
  - Implementation Phase – 1-2 FTEs ($140-$420K)
  - Total Cost – $230-$510K per pilot project
- ***Operations Costs (Phase 2 and after)***

- Project lead – (can also manage pilot efforts/next phases) – 0.5 FTE ($120K)
- Admin assistant/outreach/community logistics – 0.5 FTE ($70K)
- System management – 0 .25 FTE ($35K)
- User support – 1.0 FTE ($70K)
- Maintenance Programmers – 2.00 FTE during Phase 2 ($280K), 1.00 FTE after Phase 2 ($140K)
- Data expansion – estimate additional 4 TB /yr : <$5K
- Travel – $20K /yr
- **Total Operations Costs:** $600K/yr during Phase 2, $460K/yr after Phase 2

PIMS Phase 2 would cost between $230K and $510K per pilot project (for 7 to 9 pilot projects total) to develop. PIMS would cost $600K/yr to operate during PIMS Phase 2 and $460K/yr to operate after completion of PIMS Phase 2.

4.1.3 Beyond Phase 2

Completion of Phase 2 represents completion of the current view of a fully functional PIMS, and it is difficult to define in detail how PIMS might progress beyond Phase 2. Further developments of PIMS might include:

- Incorporation of data for other hazards,
- Development of additional user interfaces and tools to provide more customized support for particular types of users or groups addressing specific problems, and
- More advanced analysis and summary tools with the ability to create and share derived data sets (i.e., the ability to manipulate the raw data to produce value-added data sets) and analyses themselves through PIMS.

Although it is impossible to predict PIMS will become after Phase 2, it is important that the implementation plan for building PIMS (one of the next steps in developing PIMS) include the all-encompassing, long-term vision of PIMS as a system serving all stakeholders needs regardless of how complex they are. Even if this vision is never fully realized, or it evolves as the community engages with an operational PIMS, it is important that the end-goal be stated and used to guide design and technology choices. Keeping the end-goal for PIMS in community conversations during the early phases will also be important in maintaining engagement with those stakeholders whose required functionality is not fully implemented in initial implementations of PIMS.

Although completion of PIMS Phase 2 represents completion of the current view of a fully functional PIMS, it is important that the all-encompassing, long-term vision of PIMS continue to be discussed and evolved to guide technical choices during the project and to maintain community momentum towards the goal of improving protection from hazards beyond the end of Phase 2.

Given the rate of change of technology, the broad range of technologies involved in PIMS and the need to integrate with external projects that are evolving independently, an argument can be made that the level of effort necessary to ensure that PIMS can respond to changing user needs and maintain and evolve core infrastructure is roughly the same size as the minimal team required for initial development. As user interest and the data holdings grow, the effort in user support and outreach may actually need to increase slightly. In an environment where other communities are also actively evolving cyberinfrastructure, significant leverage across efforts may be possible. As in the development phase, additional leverage may be available through participation of PIMS in ongoing research and development efforts.

4.1.4 Preparedness for Next Earthquake

What happens if a significant earthquake occurs before PIMS Phase 1 is complete? Will any components of PIMS be mobilized to collect data as able? The consensus of the PIMS workshop participants is that PIMS will be a significant source of expertise, capabilities, and resources even before Phase 1 is complete and thus it will be better to respond in some measure to the next significant earthquake than to not do so at all. While it may be appropriate for PIMS staff to engage with the community in response to any event after the start of the project, to assist with operational capabilities will require a core set of functionality to be in place including:

- The core repository operating reliably and at sufficient scale to manage one event's worth of data, and
- A basic means for manual web-ingestion of data into the repository.

Planning specifically related to event response would also be very important and would potentially include planning related to

- Using third-party/donated virtual machine hosting services for the repository and web services to quickly scale services after an event,
- Providing pre- and post-event rapid training on how to manually upload data into the repository, and
- Support for use of third-party/donated mobile data collection tools (i.e., camera, GPS, PDA, etc.).

The following are more advanced levels of functionality that may also be deployable before Phase 1 is complete:

- Simple standard forms for data collection like those developed for the ROVER program (Figure 4.2) or ATC-38 that can be employed using mobile data collection tools,

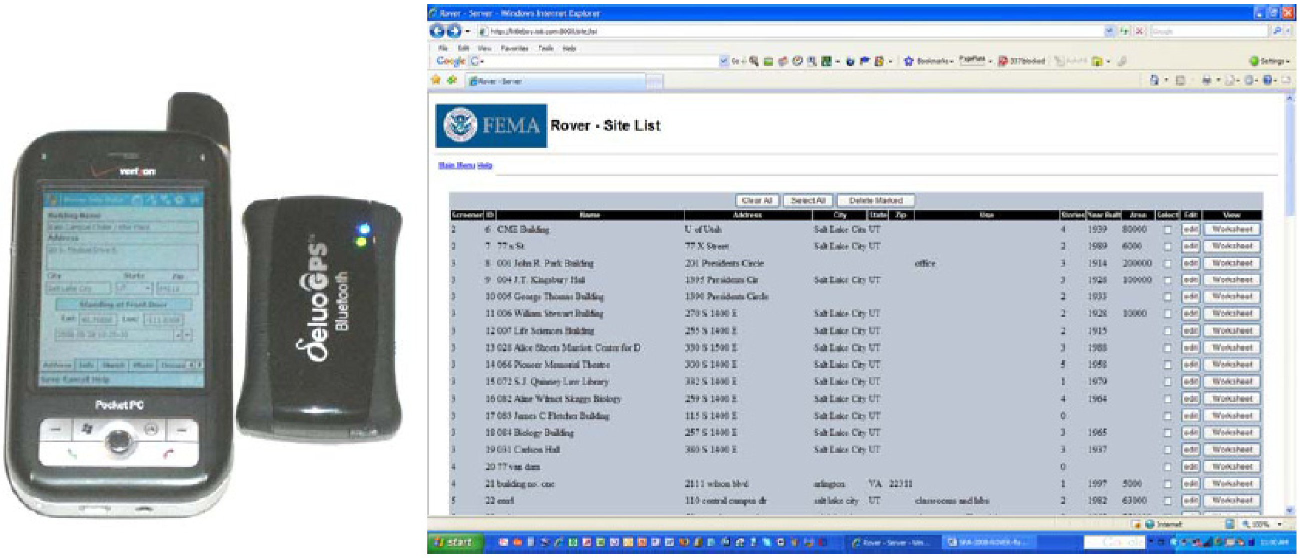


Figure 4.2 FEMA's ROVER program mobile data collection tools (left) and database server (right) (from http://www.sparisk.com/pubs/ATC67-2008-ROVER-flyer.pdf)

- Field centers/stores where data from mobile data collection tools can be automatically ingested into PIMS (i.e., bulk data ingestion via web services as opposed to post-collection manual entry using web forms),
- Mechanisms to encourage or ensure the use of the standard forms for data collection, and
- Provision for dissemination of federal data sets (such as the Homeland Security Infrastructure Program dataset and USGS hazard maps).

Analysis of the best options for supporting response during development will be required after an implementation plan is developed though, given that the technology choices made may make it more or less expensive to deliver specific levels of early capability related to ingest and storage, and therefore may affect the trade-off involved in being prepared for early events versus remaining focused on delivering PIMS as designed on time and on budget. Decisions in this area should be considered by the PIMS project team and PIMS community as implementation planning and development progress PIMS.

PIMS should operate in some form if the next earthquake occurs before PIMS Phase 1 is developed, which will require prioritization of basic ingest and storage capabilities along with event-response-specific planning and community discussion of the potential value in providing additional early capabilities.

4.2 Oversight and Management

In projects such as PIMs that involve significant scope and ongoing coordination with a broad community, effective oversight and management are critical in keeping technical developments on track and in assuring that success is achieved in terms of value to the community (i.e., going beyond success in terms of meeting scope, schedule, and budget). This section identifies a relevant structure and strategies for oversight and management and presents a list of key issues and questions to be addressed or resolved during the development of PIMS. Considerations related to community data collection and PIMS interactions with other community organizations are also discussed.

4.2.1 Oversight and Management Structures

The PIMS project should be structured as an integrated effort by a core team. Experiences in other scientific cyberinfrastructure development programs have shown that very strong management is required to ensure that IT development remains focused on delivering operational capabilities. It is beneficial that this core team remain the same from project inception through at least the initial phases of PIMS development to avoid conflicts that could arise due to changes in leadership. Moreover, the core team must have a clearly identified leader. Requiring the head of the project to be a stakeholder is one mechanism for ensuring that the focus remains on delivering an operational system [39]. It also may be beneficial to separate responsibility for requirements analysis, design and development, and system integration. Separating specification from development forces accountability within the project, giving the project lead the ability to request alternate designs from third parties and issue subcontracts as needed to pursue the best strategy. The lead also should have the authority to monitor subcontract performance and make changes as needed. Further, empowering the project team to leverage expertise and capabilities from multiple organizations and to reallocate funding to address performance issues helps align responsibility and authority within the project. Supplementing the core team with separate advisory boards (in contrast to a governing board) can help ensure that the lead receives needed advice while still centralizing authority to make changes. These advisory boards should make periodic assessments of the progress of PIMS and the effectiveness of management and suggest improvements.

In addition to the ongoing management team and advisory boards, PIMS must incorporate incident response teams to be mobilized when significant hazard events occur. Their purpose would be to supplement the core management and operations teams to operate PIMS in periods of high demand by ensuring that data collection standards are followed, aiding ingestion of data into PIMS, facilitating curation of data, etc. For example, additional PIMS teams will need to mobilize to aid in data ingestion at hazard sites, and additional personnel will be required to provide rapid curation of the large amounts of data that come in from field investigations.

The PIMS management team should be structured as an integrated effort by a single core team with a clearly identified lead. The team should be supplemented by permanent advisory (not governing) boards and incident response teams that mobilize when significant hazard events occur.

4.2.2 Key Issues

Many issues related to PIMS' capabilities and its role in the community have arisen in the course of this scoping study that cannot be fully addressed at this stage. Some of these issues can be addressed through further community discussion leading to consensus (e.g., through workshops), whereas others can be addressed only by the central PIMS management team with specific implementation decisions or inter-organization agreements. In both cases, the community discussion and scoping analysis that have contributed to this report highlight relevant directions and general solutions that must be made concrete and specific as PIMS becomes reality. Table 4.2 summarizes key questions to be answered or issues to be addressed in the implementation plan for PIMS, through community consensus, or by central PIMS leadership. The table includes those issues discussed in Chapter 2 along with additional issues relating to oversight and management. Discussion of these issues and potential answers to the questions are given in the relevant sections of this report.

Many issues related to PIMS' capabilities and its role in the community will need addressed through community discussion and project decisions as PIMS moves into implementation and operations. A list of key questions or issues is given in Table 4.2.

Table 4.2 PIMS Key Questions and Issues

Oversight and Management
How and when will the specific scope and requirements for each phase of PIMS be decided? Is it necessary to wait until a project team is in place? What is minimally necessary to be prepared to respond to the next significant earthquake? How much will need to be sacrificed to be ready for this event? Where will PIMS be housed? What organization will administer PIMS? If PIMS will be housed in the private sector, a series of issues need to be addressed; this is true if PIMS is housed in the public sector but to a lesser extent. Who will provide funding for PIMS? What are the basic requirements for a testable PIMS (e.g., for the testing in Phase 1)? This needs to be addressed as a first step in the detailed implementation plan for PIMS. Is PIMS the ultimate organization for creation of data collection plans and operations of data collection?
Data Collection
Is PIMS a secondary or primary system for input? Or both? Should PIMS serve as the primary collection point for information from researchers and practitioners for types of information not covered by other systems? Should PIMS serve as the collection point for supplemental information (e.g., images or movies of damage) provided by "citizen scientists"? Should historical data be collected immediately? What mechanism should be used to determine which data to collect and when?
Data Storage
What are the appropriate data storage requirements for PIMS now and how will these storage requirements increase? How much additional storage is needed for metadata? Will all data be stored in one central location (with back-ups) or will the data be stored in locations distributed across the nation?
Data Persistence and Availability

What policies and procedures are needed to ensure data integrity and pro-actively monitor data availability? How will PIMS address operations in the face of power outages, regional network failure, and other external failures such as outages of affiliated repositories?
Data Organization
What is an appropriate design for PIMS to support organization of the data in ways to ensure that users will be able to find and use all relevant data despite the broad range of user goals and the evolving set of data formats? How should PIMS be designed to be capable of gathering metadata and data in multiple, externally determined vocabularies and formats and support new vocabularies and formats over time? How should information be organized to enable reasonable response times to user queries?
Data Curation and Quality Assurance
How should PIMS incorporate best practices for both manual and automated quality assurance testing? How will PIMS provide for data quality evolution if users are permitted to provide low-quality data (i.e., sparse metadata) for use in the short term but then return to fill in missing metadata when time permits? How will decisions be made about whether metadata and/or data are harvested/cached/gathered from other sources for initial phases of PIMS, and who will make them?
Privacy and Security
How will PIMS overcome the constraints that names, addresses, and descriptions of specific structures generally cannot be shared because of the restrictions set forth in the Privacy Act? How will PIMS avoid data access issues related to legal liability? What mechanism will be used to ensure that data are not inadvertently released to unauthorized users? What will be done to avoid the loss of competitive advantage that occurs when organizations provide proprietary information to PIMS? How will PIMS deal with issues of copyright? How will PIMS obtain information on critical structures (locations, detailed descriptions, and drawings of police stations, fire stations, hospitals, bridges, etc.) given government security concerns?
Long-term Data Preservation
How will PIMS maintain both data quality and the means to interpret, visualize, and retrieve the data over the long term?
Data Standardization
Is PIMS a data service or a driver of best practices and standards for post-event data collection efforts? How do you standardize data from groups that have diverse needs and that store and arrange data in formats most relevant to their specific applications? Which standards will PIMS support through tailoring of its ingestion, harvesting/exchange, presentation, and export features?
System Evolution and Change Management

How will PIMS accommodate, technically and organizationally, the fundamental requirement that PIMS be operational for at least the 50 to 100 years that may pass between severe natural hazard events? How will PIMS evolve (e.g., through periodic refreshes) to keep pace with changes in technology and user needs over this time span?
Coordination/Data Sharing with Public, Private, and Governmental Sources
From which organizations should PIMS collect data? What approaches should be used to encourage coordination and data sharing? How will PIMS motivate members of academia to share their information?
Lack of Information
How will PIMS deal with lack of information in certain areas (e.g., building structure inventories)? How will PIMS address the variations in quality and quantity over different types of data?
Community Adoption of PIMS
Who are the sponsors/stakeholders for PIMS? How will the PIMS community be caused to buy into the system?

Other oversight and management considerations:

- PIMS itself should be considered to be separate from new directions in system research (e.g., PIMS may participate in new research directions but probably should not fund them.)
- The more PIMS considers all hazards, the more it is likely to be attractive to federal agencies.
- When PIMS becomes all-hazard, it may need to be administered by an organization that represents the interest of all hazards, not just earthquakes.
- The physical location of data for PIMS can be independent of the organization that manages PIMS. Software can be independent of the organization as well.
- There exist a number of programs and organizations that might cooperatively operate PIMS (e.g., NEHRP agencies), and the proposed PIMS design allows for federated data repositories, not just a central repository.
- PIMS funding is a major issue that needs to be addressed as a next step in PIMS development. At least initially, it appears that federal funding will be needed, and it may be possible for PIMS to utilize private sector funding as the system evolves.

4.2.3 Data Collection

With respect to the data collection component of the overall PIMS framework (some similar issues are discussed in Section 2.6), the primary question is whether PIMS should be the ultimate organization for creation of data collection plans and operation of data collection? No overall consensus has been reached on this topic. Some believe that another organization should serve in this role and that it would consider the PIMS information management center in the process. Others believe that PIMS should take a more active role in the data collection planning and operation processes. Specifically, the majority appears to believe that PIMS will be required to collect inventory data before hazards events because no other organization currently does this. In general, the PIMS framework includes data collection planning and operation, but it is unclear to what extent PIMS will control the process. This issue needs to be addressed as PIMS is developed.

Although it is not clear whether or not PIMS should lead the effort to create a data collection plan, it is important that the organization controlling the planning process consider PIMS in relation to:

- Forms/standards to be used for data collection;
- Requirements for systematic data collection for PIMS because the transition from anecdotal data collection practices to more systematic collection practices requires a far greater commitment in terms of time and funding from the earthquake engineering community than currently exists;
- The location where data are to be stored or uploaded; and
- Implementation of clearly defined agreements between PIMS and data providers such as memoranda of understanding (MOUs) and service level agreements (SLA) prior to a natural disaster.

Agreements with data providers should identify the types of data to be shared, how the data are to be used, what restrictions will be placed on access and use of the data, and what benefits will accrue to the private entity for supplying the data [9]. In many instances, establishing MOUs with cities may be more beneficial than with states because it is easier to collect fine-grained data from cities and easier to provide incentives and value given the more limited resources of cities.

While no consensus was reached on whether PIMS should be the central entity to plan for and operate data collection efforts, a majority of the community appears to believe that PIMS should act to facilitate the development of standards for data collection and ingestion but that it should not control (or micromanage) the entire process. While PIMS as a project may not be able to drive standardization on its own, its support in terms of making standards a part of the community discussion around PIMS and implementing operational support for new standards in PIMS could add significant momentum to such efforts. Individual professional societies (ASCE, EPRI, TCLEE, etc.) may be the best drivers of standards, a role they have today, but PIMS should play an active and organizing role in the community by providing a framework for creation of standards that would include efforts to map standards to driving PIMS use cases and to operationalize standards through the creation of templates and forms, development of incentives to provide standardized data and metadata, and inclusion of standards in PIMS training materials and activities. Standards should be pursued for both coarse-grained data collection (hence be coordinated with state clearinghouses) and fine-grained data collection.

Funding for data collection for PIMS is a significant consideration. Currently, no single source provides funding for comprehensive collection of data following hazard events; rather, funding is provided on an ad hoc basis by individual government agencies, professional societies, private sector industries or organizations. As such, the interests and aims of the individual entities control what data are collected and to what level of detail they are collected, often leading to gaps in information. Moreover, the funding provided by these organizations is generally not sufficient for the collection of data at the level of detail required for PIMS to be able to provide the value envisioned in this report (i.e., fine-grained, systematic data collection). For example, government agencies and professional societies may pay for their employees or members to travel to hazard event sites, but they often do not pay for the time necessary to collect comprehensive detailed data. An exception to the typical practice exists when there is an especially unique aspect to the observed damage (e.g., damage to moment-frame connections in the

Northridge earthquake) which may prompt one or more government agencies to invest in detailed data collection and evaluation.

A first step in improving data collection is to identify realistic resource requirements for such things as personnel, equipment, and funding. The second step involves identifying or creating a single government source of funding with the authority to prescribe what data are tp be collected and to what level of detail. Although expanding how to take these two steps is beyond the scope of this report, it should be considered a priority in the implementation of PIMS. Whichever government or other entity leads the data collection planning efforts described above, it must consider these important issues relating to funding for data collection.

While it is unclear if PIMS should act as the entity that plans for and collects data, PIMS should be considered as plans are made and should be active in the development and support of standards for data collection and ingestion by individual professional societies. For PIMS to perform as envisioned in this report, funding for data collection must be increased and dispensed in a coordinated way.

5 CONCLUSIONS AND NEXT STEPS

The scope for PIMS has been presented in terms of user needs and system requirements, challenges or system-level issues involved in PIMS development and operation, and a design strategy to overcome the challenges and satisfy user needs and requirements. The need for PIMS is clear, and the hazards community has long recognized that any national effort to reduce economic losses and social disruption resulting from severe natural disasters requires a mechanism that will capture lessons learned from disasters; preserve engineering, scientific, and social performance data; and provide a coherent, comprehensive national resource for analyzing performance. An overall PIMS framework has been described that includes the collection of data for PIMS, an IT-based information management center, and the distribution of data in PIMS for use to improve protection against hazards. Implicit in this description is the need for data collection practices to evolve beyond anecdotal-type observations and to become a systematic method of collecting the data needed to support in-depth statistical and other analyses of performance data.

User needs and requirements for PIMS have been analyzed in terms of three overlapping perspectives: serving a broad range of users, meeting stakeholder expectations as a persistent national resource, and complementing related state, local, and industry efforts. Specific needs and requirements have been captured through review of pertinent national documentation on the subject and discussions with individuals highly interested and experienced in the subject. User needs range from higher-level needs to be able to search for, visualize, and extract information in PIMS to lower-level, concrete needs for specific types and quantities of data. Specific types of information required for PIMS have been identified as relating to hazard quantification, buildings, bridges and other lifeline system components, critical structures, physical and economic losses, historic data, and PIMS itself. Moreover, higher-level needs have been inferred from the lower-level needs to use PIMS. These include the need for PIMS to directly ingest data from field observations, exchange information with other organizations and databases, provide security measures, and have the means to deal with privacy concerns.

In addition to satisfying user and stakeholder needs, PIMS development must address issues related to the cultural, political, technological, and organizational context in which it will operate. These system-level issues elucidated through review of existing documentation and discussions with stakeholders are significant. Issues identified relate to data collection, organization, storage, curation, quality assurance, and standardization; privacy and security; information presentation and retrieval; long-term data preservation; system evolution and change management; coordination/data sharing with public, private, and governmental sources; lack of information for PIMS; and community adoption of PIMS.

The design strategy for PIMS has been formulated to satisfy user needs and address the system-level issues. It is based on a service-oriented system architecture whose components include a content repository; workflow and provenance management systems; data ingestion, curation, preservation, and discovery/export mechanisms; user interfaces with analysis and visualization tools; security system; and system administration components. In addition to the system architecture, the design strategy is based on the concepts of content management abstraction, workflow and provenance, virtual organization-based security and configurations, semantics, and virtual machine implementation. Other important components of the design strategy are the policies and procedures that work in parallel with the system architecture and design concepts. These relate to data standards and provenance, data privacy and security, persistence of data, and community adoption of PIMS. To ensure that all user need and system requirements have been met using the proposed design strategy, a Requirements Traceability Matrix has been constructed that compares required system abilities to provided system functionality.

In addition to a design strategy, PIMS development, operation, and oversight have been explored. The rationale supporting a phased development approach has been developed, and the goals, schedule/milestones, and required resources for each identified phase are given. Further, the need for strong PIMS oversight and management has been emphasized. This report outlines suggested approaches for oversight and management, describes potential sources of funding for PIMS, lists key issues to be addressed in PIMS development, and discusses issues relating to the data collection component of PIMS.

The next steps in PIMS development are to:

1. Discuss/evangelize PIMS with stakeholder constituencies which will involve identifying links/relationships with initial data sources for PIMS (portals to existing data)
2. Develop a source of financial support for development and obtain backing for the long-term operation of PIMS as a national resource
3. Create a detailed PIMS implementation plan
    a. Determine user requirements priorities (leverage this report to do this)
    b. Identify data that are easy to obtain for the initial development of PIMS (e.g., ShakeMap, strong motion data, HAZUS data, IRIS data)
    c. Define the functional requirements of PIMS
    d. Eliminate the gaps in user defined requirements
4. Involve the PIMS community in the implementation plan
5. Develop PIMS Phase 1
    a. Test with historic earthquake data
    b. Test with an exercise
6. Develop PIMS Phase 2
7. Continue development to create the long-term vision of supporting improved protection from hazards

A firm foundation has been laid and a clear set of next steps in PIMS development have been identified.

Finally, some general considerations should be taken into account in the next stages of PIMS development. Knowledge gained during conduct of this scoping study (e.g., what users want to do with PIMS, the types of technology and methods that exist which PIMS may utilize) should be leveraged to create the implementation plan for PIMS and develop PIMS Phases 1 and 2. Throughout each development cycle, outreach efforts should be conducted to the extent possible. These include presentations at technical conferences, organizational planning meetings, and national conferences that communicate the concept of PIMS, implementation plans, and ways in which people can get involved with PIMS. In addition, community input regarding PIMS should continually be encouraged and incorporated into the planning for PIMS.

References

[1] National Institute of Building Sciences. (2007). *ALA Workshop on Unified Data Collection.* Available at <http://www.americanlifelinesalliance.org/pdf/ALAdatawkshprpt.pdf>.

[2] National Earthquake Hazards Reduction Program. (2007). Program Overview. <http://www.nehrp.gov/pdf/nehrp_acehr_ppt.pdf>.

[3] National Institute of Building Sciences. (2007). *ALA Workshop on Unified Data Collection Appendices.* Available at <http://www.americanlifelinesalliance.org/pdf/ALAwkshpAppendices.pdf>.

[4] Interview with Thomas Holzer. July 23, 2008.

[5] The Domestic Preparedness Branch. (2007). New 'HSIP Gold' Offers More Data, Dramatic Improvements. *Pathfinder: The Geospatial Intelligence Magazine*, Vol. 5, No. 4, pg. 8-9.

[6] Mills et al. (2008). The clearinghouse concept: a model for geospatial data centralization and dissemination in a disaster. *Disasters*, In Publication.

[7] National Geophysical Data Center. Significant Earthquake Database. Accessed July 2008. Available at <http://www.ngdc.noaa.gov/hazard/earthqk.shtml>.

[8] Advisory Committee on Earthquake Hazards Reduction. (2008). *Effectiveness of the National Earthquake Hazards Reduction Program*. Available at <http://www.nehrp.gov/pdf/2008ACEHRReport.pdf>.

[9] Earthquake Engineering Research Institute. (2003). *Collection and Management of Earthquake Data: Defining Issues for An Action Plan.* <http://www.eeri.org/lfe/eq_collection_data.html>

[10] Tuner, Fred and McCormick, David L. (2007). SEAOC's New Post-Disaster Performance Observation Committee. *Proceedings of the 2007 SEAOC Convention.*

[11] Holzer et. al. (2003). The Plan to Coordinate NEHRP Post-Earthquake Investigations. *U. S. Geological Survey Circular 1242*. U. S. Geological Survey, Reston, Virginia. <http://geopubs.wr.usgs.gov/circular/c1242/ >

[12] *WATERS* Network. (2007). *Draft of Science, Education, and Design Strategy for the WATer and Environmental Research Systems Network.* February 27, 2008 draft. Available at < http://www.watersnet.org/index.html >.

[13] Myers, J., McGrath,R.E. (2007). Cyberenvironments: Adaptive Middleware for Scientific Cyberinfrastructure. *6th Workshop on Adaptive and Reflective Middleware* (ARM'07), Newport Beach, November 26, 2007.

[14] Task Group on Information Technology Vision of the Board of Directors of NEES, Inc. (2007). *Information Technology within the George E. Brown, Jr. Network for Earthquake Engineering Simulation: A Vision for an Integrated Community*. The George E. Brown, Jr. Network for Earthquake Engineering Simulation. Available at < http://www.nees.org/Governance/Committees/TGoITVision_v1_DVD.pdf >.

[15] ATC Active Projects and Forms. Available at <http://www.atcouncil.org/#>.

[16] HIPAA. *National Standards to Protect the Privacy of Personal Health Information*. Available at < http://www.hhs.gov/ocr/hipaa/>.

[17] SRI Phil Schneider. Available in Appendix E.

[18] SRI Fred Turner. Available in Appendix E.

[19] ER Steve French. Available in Appendix F.

[20] Central United States Earthquake Consortium. (2008). *Multi-State Clearinghouse Coordination Plan: A Plan to Bridge State Post-Earthquake Technical Information Clearinghouses in the Central United States (Version 3 Draft).* Available upon request through CUSEC.

[21] ER Jim Wilkinson. Available in Appendix F.

[22] SRI Jim Wilkinson. Available in Appendix E.

[23] National Ecological Observatory Network. *Networking and Informatics Baseline Design for the National Ecological Observatory Network.* (2006). Available at < http://www.neoninc.org/milestones/2006/nibd.html>.

[24] CUAHSI. (2007). Recommendations from Environmental Observatories Operations Meeting, September, 2007. Available at <http://www.cuahsi.org/meetings/wtb-presentations.html>.

[25] Wang X., Gorlitsky R., and Almeida, J. S. "From XML to RDF: how semantic web technologies will change the design of "omic" standards*." Nat. Biotech,* 2005, Vol. 23 No. 9, pg. 099–1103.

[26] Foster, I. "Service-oriented science." *Science*, Vol. 308, pg. 814-817.

[27] De Roure D., Hendler J. A. "E-Science: The Grid and the Semantic Web." *IEEE Intelligent Systems* ,Vol. 19, No. 1, January/February 2004, pg. 65-71.

[28] Neumann, E. A. "Life science semantic web: are we there yet?" *Science Signal Transduction Knowledge Environment (STKE)*, Vol. 283, 2005, pg. 22.

[29] Myers, J.D., Allison, T.C., Bittner, S., Didier, B., Frenklach, M., Green, W.H. Jr., Ho, Y.-L., Hewson, J., Koegler, W., Lansing, C., Leahy, D., Lee, M., McCoy, R., Minkoff, M., Nijsure, S., von Laszewski, G., Montoya, D., Pancerella, C., Pinzon, R., Pitz, W., Rahn, L.A., Ruscic, B., Schuchardt, K., Stephan, E., Wagner, A., Windus, T., Yang, C. "A Collaborative Informatics Infrastructure for Multi-scale Science." *Proceedings of the Challenges of Large Applications in Distributed Environments (CLADE) Workshop*, 2004.

[30] Futrelle J., Myers J., Minsker B., and Bajcsy P. "Community-based Metadata Integration for Environmental Research." *Proceedings of the 7th International Conference on Hydro-science an Engineering (ICHE),* Philadelphia, PA, Sept. 10-13, 2006.

[31] Gil Y., Deelman E., Ellisman M., Fahringer T., Fox G., Gannon D., Goble C., Livny M., Moreau L., and Myers J. "Examining the Challenges of Scientific Workflows." *Computer*, Vol. 40, No. 12, December, 2007, pg. 24-32.

[32] Moreau L., Friere J., McGrath R. E., Myers J., Futrelle J., and Paulson,P. (2008). The Open Provenance Model. Provenance Challenge Review, January 1 - February 14, 2008.

[33] “Provenance and Annotation of Data.” *International Provenance and Annotation Workshop IPAW*, 2006. Revised Selected Papers (Lecture Notes in Computer Science Vol. 4145). Chicago, IL, USA: Springer. See http://twiki.ipaw.info/bin/view/Challenge/WebHome.

[34] Watry P. "Digital Preservation Theory and Application: Transcontinental Persistent Archives Testbed Activity." *International Journal of Digital Curation*, Vol. 2, No. 2, 2007.

[35] Tara D., Talbott K. L., Schuchardt E. G., Myers S., and Myers J. D. “Mapping Physical Formats to Logical Models to Extract Data and Metadata: The Defuddle Parsing Engine.” *IPAW'06 International Provenance and Annotation Workshop,* Vol. LNCS 4145, pg. 73-81. Springer-Verlag Berlin, Heidelberg, Germany.

[36] Jesse James Garrett (February 18, 2005). "Ajax: A New Approach to Web Applications". <www.AdaptivePath.com>. Retrieved on June 19, 2008.

[37] Ragouzis N. et al. Security Assertion Markup Language (SAML) V2.0 Technical Overview. OASIS Draft, February 2007. Document ID sstc-saml-tech-overview-2.0-draft-13. Available at <http://www.oasis-open.org/committees/download.php/22553/sstc-saml-tech-overview-2%200-draft-13.pdf>.

[38] Basney J. MyProxy Protocol. Global Grid Forum Experimental Document GFD-E.54, November 26, 2005.

[39] Spencer, B., Butler, R., Ricker, K., Marcusiu, D., Finholt, T., Foster, I., Kesselman, C. "Cyberenvironment Project Management: Lessons Learned." (September 2006). Available at < http://www.nsf.gov/od/oci/CPMLL.pdf >.