Acquisition, Storage, Retrieval and Dissemination of Disaster Related Data

Shaik A. Ahammed, Karthik Badam, Guddanti V. Bhaskar, Sumit Chhuttani, Anuj Gangwar, Vaibhav Garg, Lokesh Gupta, Anubhav K. Jain, Rohit Jindal, Neeraj Kumar, Kiran K. Manku, Vijay K. Midde, Malempati Nagarjuna, Rahul Patil, Saurabh Singh, Nallapu Sneha, Praveen B. Srinivas, Mudit Tanwani, Sameer Wasnik, Ravindra Guravannavar, Subrahmanyam Kalyanasundaram, Kotaro Kataoka, M V Panduranga Rao, Naveen Sivadasan
Indian Institute of Technology Hyderabad, India

***Abstract*—We report the design and development of a software system that provides a unified platform for acquisition, storage and retrieval of scientific, situational and infrastructural data in the context of disasters. The system supports (i) data acquisition from heterogeneous sources (including sensor data from various research groups, field data from disaster management agencies and citizens, and relevant information from the internet) and ensures interoperability, (ii) high availability and scalability and (iii) a wide range of data visualization and retrieval modes. We believe that such a system would go a long way in assisting researchers and disaster management agencies in enhancing disaster preparedness, mitigation and recovery operations.**

I. Introduction

One of the biggest challenges in natural disaster preparedness, mitigation and relief is the lack of readily available information. Scientific data like those procured from weather and seismic sensors can help in prediction and preparedness for disasters. On the other hand, availability of situational and infrastructural information in real time is essential for mitigation and relief work. This information is also crucial for post-analysis of disasters, and for designing better strategies and protocols for potential disasters in the future.

Given the critical nature of such information, it is important to have a robust, scalable and fault tolerant system of data acquisition, storage and dissemination. The domain of storage and retrieval of disaster data has seen a lot of research and development in the recent past. A fairly large body of work deals with the problem of interoperability of heterogeneous data, see for example [10], [11]. Hristidis et al [9] provide a comprehensive survey of problems and solutions in disaster data management (acquisition, interoperability and extraction of diverse data) and analysis. Other efforts include disaster management systems that provide emergency operational and logistical support during disasters through registration facilities for victims, inventory, shelters etc. The Sahana system deployed in Sri Lanka [13], [5] and the RisePak system in Pakistan [7] are cases in the point. Ushahidi [6] employs the idea of *crowd-sourcing* to gather live disaster related information from the public, and subsequent visualization and mapping of the data.

Taking into account the peculiarities of developing countries where it may not be possible to assume a deep penetration of technologies like high speed internet, GIS etc, we summarize the following list of features expected of a disaster information system that subsumes most of the previous work.

- *Acquisition* of scientific data (for instance, sensor data from weather sensors) and field data. Often, a combination of different scientific parameters are more useful than isolated ones in prediction of disasters. It is therefore important for scientific data to be available across research groups for analysis.
  The system should allow multiple ways of gathering information generated during the disaster itself. For instance, it should be possible to import data collected by various disaster management agencies. In addition, citizen volunteers should be able to contribute disaster related information whenever possible through a user-friendly web interface. Even in the event of a catastrophic failure of all internet and telephone connectivity, it should be possible for first responders to gather data on smartphones, with the intention of exporting it to the centralized server when connectivity is restored.
- *Robustness, scalability and high availability of disaster data* is vital in the immediate aftermath. Therefore, it is important to take into account factors like system failure and down time, the servers themselves getting affected by the disaster and heavy load on the servers.
- *Dissemination* to various research groups (particularly scientific data), and general public (particularly situation data, infrastructural data, data from various social media and news websites, and instructional materials). User friendly web interfaces should allow searching for victims who have contacted emergency services or who have been registered by rescue teams.
  The system should have export facilities and visualization tools for the stored scientific, situation reporting and infrastructural information.
- *Interoperability.* Since all the above information is provided by and shared between heterogeneous en-

---

This work was done while the authors were at IIT Hyderabad.

Corresponding Author: Naveen Sivadasan (email: nsivadasan@iith.ac.in).

This work was supported by the Indo-Japanese project on "Information Network for Natural Disaster Mitigation and Recovery" (SATREPS- DISANET), funded by the Japanese International Cooperation Agency (JICA) and Japan Science and Technology Agency (JST).

tities, it is vital that the data formats are standardized for interoperability [9], [11]. Thus, any data storage and retrieval system of this diversity and scale has to conform to existing standards. When necessary, new ones have to be developed or existing ones enhanced.

In this paper, we report the design and implementation of a unified solution to meet the above objectives. This work is a result of an Indo-Japanese project on "Information Network for Natural Disaster Mitigation and Recovery" (SATREPS-DISANET), funded by JICA/JST. The project consists of collaborating research groups across India and Japan specializing in various disciplines related to disaster research—seismology, meteorology, sensing systems, communication systems and ICT. Indeed, the data stored in the system being reported in the paper includes data from these groups.

The rest of the paper is arranged as follows. The next section presents an overview of the system described in this paper. Sections III and IV discuss acquisition of pre-disaster scientific data and field data during disaster respectively. Section V discusses search and visualization applications. Section VI describes the data standardization schemes used in the system. Section VII discusses the architecture of the system. Section VIII discusses future work, and deployment strategies for effective use of the systems for disaster research, mitigation and recovery.

II. System Overview

Our system has the following four major parts.

Data Storage Platform: This is a central data repository where all disaster related data, including pre-disaster scientific data, logistical data and live data are stored. The storage platform is designed to be highly available and scalable in order to support fail safe operations and a large volume of transactions, which are critical requirements during disaster relief and rescue situations.

Disaster Related Data: A variety of disaster related data can be stored in the system. This includes (i) pre-disaster scientific data such as weather sensor and seismic data (ii) live data about victims such as health status, locations, phone messages, captured videos and audio messages (iii) data related to emergency services such as medical, police, firefighting services and road traffic status and (iv) disaster related news/RSS feeds, educational materials, operation manuals related to disaster preparedness and relief operations.

Data Access APIs: This layer consists of a host of web service based APIs provided to store and access the data based on different criteria. These APIs allow both in-house and third party applications to develop relevant disaster management applications. The data transfer formats conform to major open standards for disaster related data such as PFIF, EDXL etc. The data access layer is designed to provide highly available and scalable services.

Applications: A suite of disaster related web based and smartphone based applications are provided for acquisition, retrieval, and visualizations of pre and post-disaster data, rescue and relief logistics/operations and information dissemination. Example applications include (i) smartphone applications to gather victim data, resource request data etc. from disaster affected areas, (ii) peer-to-peer applications for smartphones to exchange, spread and eventually aggregate disaster data at the central data store, (iii) RFID based victim tracking applications, (iv) applications for sensor data query and visualization, (v) victim search applications, (vi) map based visualization of facilities and victims in the affected regions, and (viii) information portals.

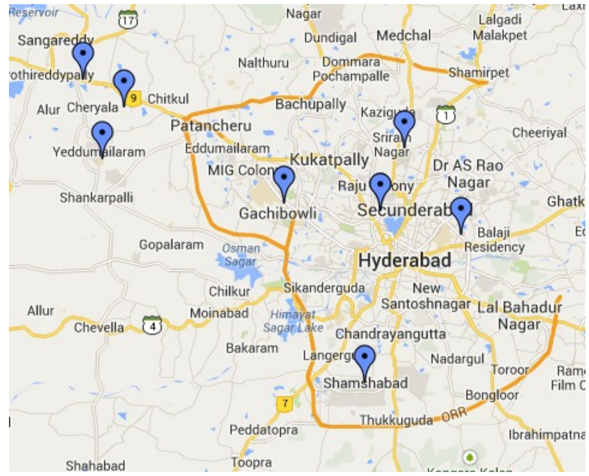

Fig. 1. Map display showing the location of the sensors deployed

In the following sections, we discuss the specific functionalities provided by the system under these subcategories and highlight their salient features.

III. Pre-disaster data

A. Pre-disaster scientific data

Real time monitoring of various atmospheric parameters such as temperature, humidity, rainfall etc. are valuable in predicting natural disasters. Our data platform allows storage of different parameters received from sensors deployed at various locations. The sensors can store the data directly on our system using either http requests or through web services. Presently, our system is receiving data from 11 Automatic Weather Stations (AWS) which were installed as part of the DISANET project in and around Hyderabad city. The installation includes WMO standard Automatic Weather Stations (AWS) in sparse mode (high cost, less number and wide coverage), and non-WMO standard AWS in dense mode (low cost, more number and narrow coverage). These sensors measure various parameters such as rainfall, humidity, temperature, pressure and wind speed. These sensor data can be queried using various query APIs that are provided. This allows for in-house and third party monitoring and alert applications based on various combinations of these sensor parameters. New sensors and sensor types can be added to our system easily. Our system also provides a set of visualizations for these sensor data based on user specified time windows and sensor types (See Figures 1, 2). All visualizations supported by our system are web based and a user can run these visualization services from any remote location using standard web browsers.

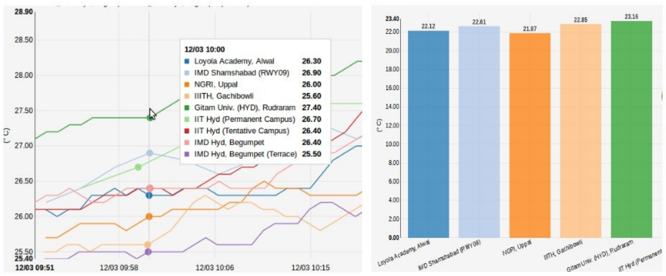

Fig. 2. Some of the sensor data visualizations available in the system. The first figure shows combined plots of temperature from all sensors. The second figure is a histogram showing average temperate measurements for all sensors for a user specified time window.

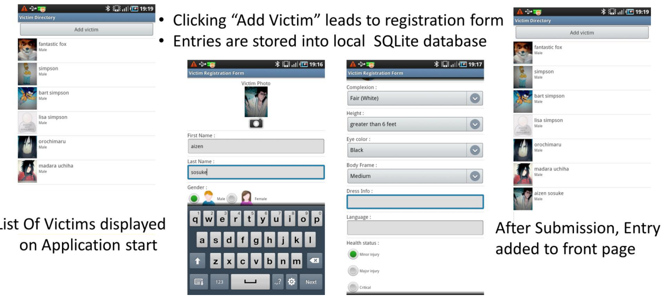

Fig. 3. Screen shots of the victim data app.

B. Static Data

Information regarding existing medical, police, fire force and other emergency service facilities in a locality are essential for effective disaster management. Such data can be easily incorporated in our system. The data conforms to open data standards wherever possible. For instance, our hospital data follows EDXL-HAVE open data standard [2] and stores for each hospital an exhaustive list of attributes such as the medical facilities, bed capacity, emergency department capabilities, medical specializations etc. This data is combined with live disaster data in our search and visualization applications. APIs are also provided for accessing these data from external applications.

IV. Acquisition of Live Disaster Data

In any well coordinated disaster management operation, it is important to gather as much live data from the disaster affected area as possible and make the data readily available to various stakeholders. Devices like smartphones can be quite effective in such scenarios for gathering this data and transmitting these data to the central data store. This will ensure that the various disaster management applications have access to the latest possible data. At a disaster-stricken area, it is quite likely that the local communication infrastructure including land-line telephones, GSM and internet might be either non functioning or partially functioning. Hence the data gathering and transmitting applications should be resilient to such failures. It is also important that such data gathering should be performed in a way that causes minimal hindrance to the actual disaster relief activities. The software tools presented below were developed keeping these requirements in mind.

A. Smartphone based data gathering

A variety of smartphones are available in the market that are reasonably priced and have a host of capabilities such as camera, GPS, 2G/3G, Wi-Fi, RFID, Bluetooth etc. The smartphone applications we have developed make use of all these capabilities to provide effective data gathering and transmitting services. Our smartphone applications are based on the Android platform, though in principle they could be ported to other platforms like Windows Phone or iOS.

The data gathering application can gather a variety of data from the disaster affected area, including detailed data about the victims, request for resources and situation reporting data. Based on the type of data that the user wishes to capture, the application displays a corresponding form through which the data can be entered. For instance, the victim data comprises various fields like name, age, gender, complexion, address, phone number, contact details etc. Other details like severity of the injury can also be recorded. Except for a few mandatory fields, most fields are optional. Phone camera can be used to capture victim photographs. The GPS feature in the phone is used to record the geographic coordinates of the victim. Captured data is stored locally in the phone (in the local DB instance) and would be associated with a unique ID generated for the victim. The victim data gathering application will be used primarily by rescue teams at the disaster spots. Forms for gathering new types of disaster data can be easily incorporated into the application. Locally stored data is eventually uploaded to the central data store.

B. Data acquisition and upload over unreliable communication infrastructure

The data stored locally in smartphones should be uploaded to the central data store to be made available to the various disaster related software tools. The communication infrastructure may be initially unavailable for data upload. Since the phone user usually moves from one location to another, data connection may be obtained at a later stage from a different location. Since timely availability of information is crucial for disaster management, such arbitrary delays could be unacceptable. For resilience to such unreliable communication infrastructure, our share and upload application uses a peer-to-peer gossiping protocol with other smartphones and spread the information among other devices. This increases the possibility of timely arrival of the data as there is more likelihood of one of the phones holding the data to obtain data connectivity soon. In our gossiping protocol, data sharing between pair of phones is done through Bluetooth. There are limiting factors for such a mechanism. Smartphones have limited battery power and limited storage space. Bluetooth communication can drain the battery faster. Flooding the devices with many copies of the data could lead to phones running out of storage space too soon. Slower Bluetooth data rate poses additional challenge. To address these issues, in our protocol, each phone carefully chooses its data sharing partner among the devices in its vicinity based on a variety of parameters including remaining battery power of the paired devices, the pairing history to avoid recently paired devices, partner priorities for faster information spread etc. The paired devices share succinct sketches of their stored data and battery status using bloom filters [8] for faster synchronization and reduced Bluetooth usage. Each phone carries out data sharing in a periodic, autonomous and automated fashion with minimal manual intervention.

C. Live data from other sources

Our system is integrated with other subsystems developed under the DISANET project. One such subsystem is for rapid deployment of communication infrastructure in disaster affected areas. It provides limited communication infrastructure including FM, GSM and satellite connectivity in the deployed region. This provides a base level of communication and broadcast setup in the event of failure of existing communication infrastructure. This communication system allows people in the locality to leave 'I am alive' voice messages, text messages, pictures, videos etc from their phones through limited GSM and Wi-Fi facilities. This data is periodically uploaded to the central data store through satellite connection. This communication subsystem is presently being developed by our fellow group at IIT Madras.

Our system provides web based forms corresponding to most of the smartphone based data gathering forms. Through this, public and disaster response team can provide data directly to the system over the internet. This can significantly improve timeliness and coverage of the received data.

D. RFID based tracking

Up-to-date information about the movement of various entities in the disaster locations is another crucial factor in effective disaster management. Tracking movement of victims from one facility to another, say for specialized medical care or for lack of space in temporary shelters is necessary to provide accurate information about the affected people. Tracking movement of equipments across facilities or shelters helps in better planning of resource allocation. RFID technology provides efficient and cost effective mechanism for such tracking. Our system can receive the RFID tagged tracking data which can later be queried for latest as well as past information on location and status. We have also developed victim tracking applications that can run on NFC (near field communication) enabled smart phones.

An RF tag can be placed on a victim (possibly as a wrist band) and tag it with a unique ID generated by the system. The RF tag can subsequently be scanned by the tracking application running in any of the NFC enabled smart phones. After scanning the RF tag, the application can send the latest GPS coordinates, the health status, victim's location such as the specific relief facility, hospital etc. to the central server. Web based query and visualization applications are provided for this data to facilitate better planning of rescue and relief operations and to provide up-to-date victim information.

E. Live data from the web

Live data in the web from various news agencies and the public in the event of a disaster have proved useful in the past. For example, live situation reporting by citizens and rescue workers through microblogging sites such as Twitter have proved useful in the 2011 earthquake in Japan [12]. Our system collects such live data from news feeds and tweets. RSS feeds are gathered from websites of prominent news agencies, newspapers and disaster management agencies around the world using *Yahoo!* pipes. Disaster related feeds are filtered using a large and configurable collection of filter keywords. Similarly relevant tweets are filtered through a configurable collection of hashtags. The extracted feeds and tweets are made available to the public through a single web portal.

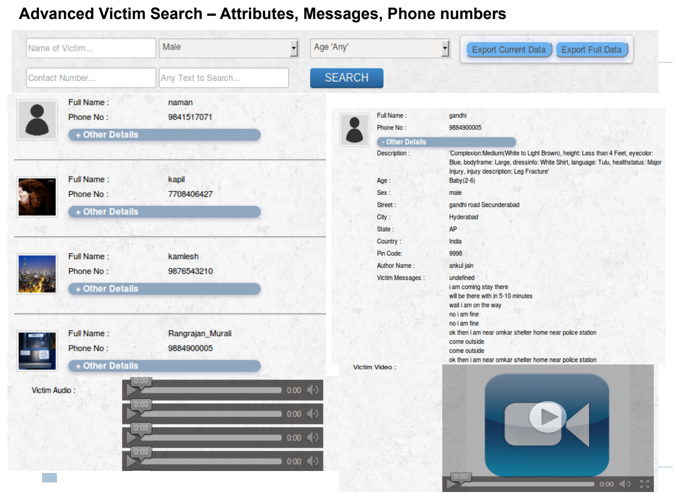

Fig. 4. Victim search results. For each victim, all associated victim data including audio, video and phone messages are displayed.

V. Search and Dissemination Applications

A host of web based applications are provided for search and visualization of the data gathered by the system.

A. Victim Search

Web based tools are provided to search the victim data. The search can be performed on the database using the specified attributes like gender, age and other physical attributes. The system also maintains a full-text index of the entire victim data. This allows the standard keyword based search on the victim data. The search keywords are matched against all victim attributes, SMS messages, phone numbers etc., thereby providing a powerful search mechanism. The matched victim data is displayed in a user friendly manner with all associated data that exists in the system, including audio, video and SMS messages (Figure 4). The search facility is valuable for the disaster management teams and the public alike. Export facility is provided to export the victim data associated with search results in XML format conforming to the PFIF open standard [3].

B. Map based visualizations

Visualization applications are provided to explore the victim data over a map based on the latest victim locations. In addition to victim locations, the map shows locations of various relief facilities, hospitals, shelters and traffic routes to assist the relief operations (Figure 5). Colour codes are used to indicate the capacity, available facilities etc., in the nearby relief centers and hospitals.

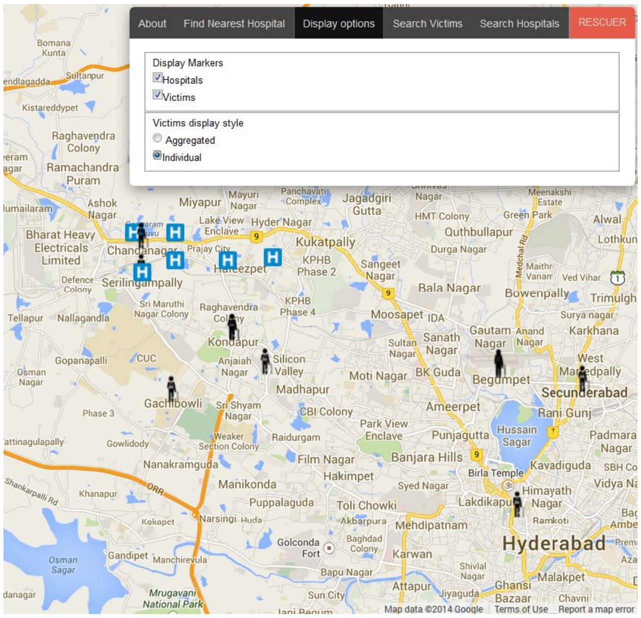

Fig. 5. Map based visualization of victim data. This map shows victim locations and locations of nearby medical facilities.

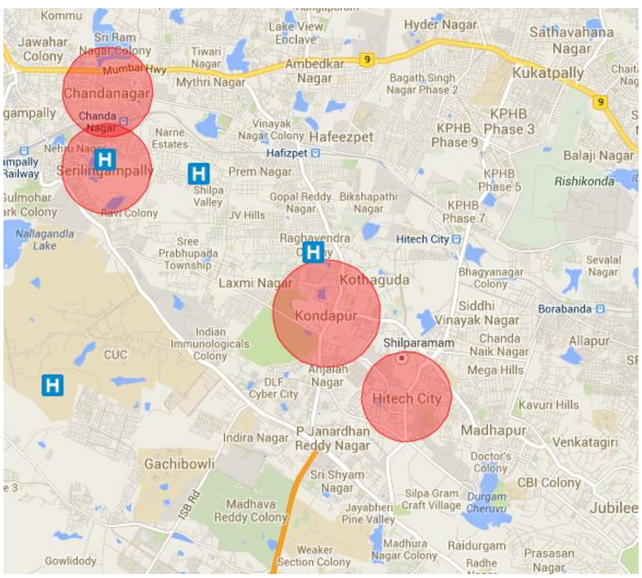

Fig. 6. Aggregate views on maps showing the severity of casualties in various locations.

Map based aggregate visualizations are also provided for viewing the magnitude and severity of casualties in various locations (Figure 6). Colour codes are used to indicate overall severity in each location. The aggregate data can also be viewed through several chart and histogram based displays provided. This is valuable for both disaster relief planning and post disaster analysis. All of the above visualizations provide timelines and the user can view past data based on the user specified time window of interest. The RFID based tracking data collected by the system is used to provide timeline based views. Such inputs can also assist in deployment of relief teams, scheduling and other resource allocation logistics.

C. *Public information portal*

Public information portals are valuable in providing live disaster information, historical information about specific disasters and instructional materials for disaster preparedness. Public can access most of the data gathered through search and visualization applications discussed above through our portal. The portal also displays disaster related news feeds gathered in real time. The portal provides links to various disaster management authorities. Work is underway in providing links to survival guides and instructional materials from by various agencies around the world. Providing the material in one place ensures reduced search time and a wider reach to the public.

VI. Data Standardization

Standardization of disaster related data is crucial for the interoperability and data exchange between various disaster related systems and applications. Various disaster data standards have evolved recently with the goal of standardizing different disaster data types. People finder interchange format (PFIF) [3], Emergency data exchange language (EDXL) [2], Hospital availability exchange (EDXL-HAVE), Situation reporting (EDXL-SitRep), Tracking emergency patients (EDXL-TEP), Resource messaging (EDXL-RM), IEEE 1512 [1] are some of the existing XML based open standards. PFIF, EDXL-HAVE and EDXL-TEP standards are already incorporated in our system. Incorporating EDXL-SitRep and EDXL-RM is in progress. The storage schema of our system conform to these standards. Data import and export functionalities are provided to allow interoperability with other systems. The user can export either the full data or parts of it based on user queries as per these formats.

A. *Metadata*

Pre and post disaster data such as sensor data, victim data, traffic data, resource data, announcement data, crowd sourced data are generated in large volumes. Hence it is important to make such data available in a form suitable for automated processing. This ensures timely delivery of processed results to various stakeholders. Some important aspects to be considered while making such data open are reliability, ownership, distribution, interoperability and access control. For example, different types of AWS sensors give different characteristics in terms of reliability of equipment, accuracy of data, condition of sensor site etc. One approach is to provide metadata comprising such information. We are working on defining XML-based metadata formats for various disaster related data. For instance, metadata associated with sensor data would include fields to specify installation site condition, accessibility, data format standard, data resolution, units for values etc.

VII. System Architecture

In this section, we give a brief overview of the high level architecture of our system. The architecture has been designed keeping in consideration high availability, scalability and performance parameters. Figure 7 depicts the high level architecture of the system.

Data Layer: The data is stored in PostgreSQL. All data access is through web service requests. The data is replicated

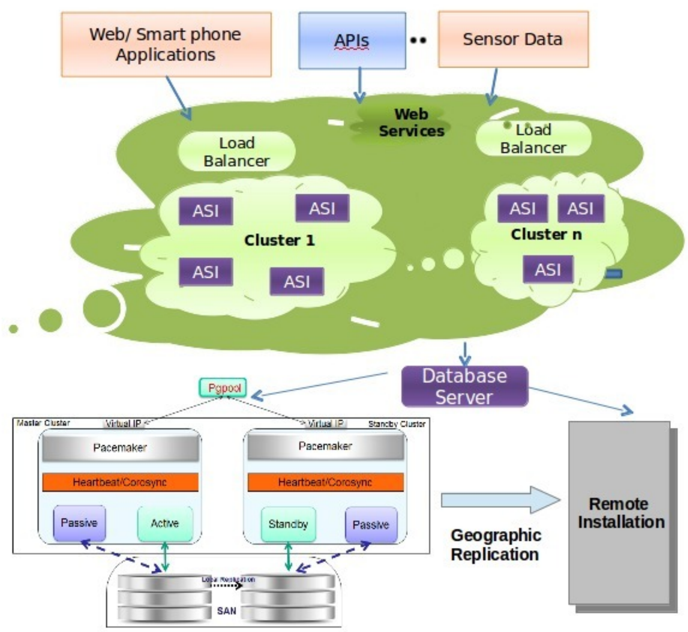

Fig. 7. High level architecture of the system. Data layer and application services layer are designed with high availability, scalability and performance goals.

locally and to a similar installation at a geographically remote location to ensure failsafe operation [4]. This is in addition to the RAID redundancy of the storage system. In our current setup, the main installation is at IIT Hyderabad and the remote installation is located at IIT Madras. For scalability and high availability of DB services, the DB server pool is partitioned into master and standby clusters, each with active and passive servers, to support automatic intra-cluster/inter-cluster failovers. The load balancers distribute DB requests among the clusters.

Application Services: All application and web services are deployed on two or more servers in the server pool. All servers are capable of hosting all services. Load balancers distribute the requests among the servers. Multiple load balancers with shared floating IP are deployed to avoid any single point of failures. All hosted services access the underlying data through web services and this decoupling allows migration of parts of the system (data services or application services or both), for instance to a cloud. Work is underway in migrating parts of the application services to a private cloud that we are setting up as part of this project.

Hardware: Our system is currently deployed on a compute cluster consisting of four high end servers with total of 72 cores and 128 GB RAM. The server pool is connected to a 40 TB SAN storage system via fibre channels.

VIII. Future Work

The system described in this paper has to be deployed extensively for use in disaster management and research. In the following we list some of the aspects that needs to be addressed for the effectiveness of our system.

- Elicitation of scientific data from various research agencies, apart from those involved in the project.
- Design of training programs for various disaster management personnel who can use the applications developed in this project.
- A comprehensive population of infrastructural databases like hospitals for various disaster prone areas.
- Standardization of enhancements to data formats proposed by this work.
- Multiple levels of authentication and access control for different stakeholders.
- Providing privacy and encryption for sensitive data.

IX. Acknowledgements

We would like to thank Prof. Devendra Jalihal, Ms. S. Sampoornam and Mr. Sreehari Nagarajan for their valuable inputs and hospitality during our visits to their group at IIT Madras. Thanks are also due to our collaborating groups from IIT Kanpur, IIIT Hyderabad, NGRI Hyderabad, IMD Hyderabad, Keio University and Tokyo University. Finally, we thank past students who have contributed to the development of this system.

References

[1] IEEE 1512 Standard. http://standards.ieee.org/findstds/standard/1512-2006.html. Accessed: 2014-04-10.

[2] EDXL. https://www.oasis-open.org/committees/tc, 2014. Accessed: 2014-04-10.

[3] PFIF. http://zesty.ca/pfif/1.4/ , 2014. Accessed: 2014-04-10.

[4] PostgreSQL. http://www.postgresql.org/docs/9.2/static/high-availability.html, 2014. Accessed: 2014-06-29.

[5] Sahana. http://sahanafoundation.org/, 2014. Accessed: 2014-04-10.

[6] Ushahidi. http://ushahidi.com/ , 2014. Accessed: 2014-04-10.

[7] S. Amin. Data management systems after the earthquake in pakistan: The lessons of risepak. In S. Amin and M. Goldstein, editors, *Data Against Natural Disasters–Establishing Effective Systems for Relief, Recovery, and Reconstruction*. The World Bank, 2008.

[8] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, 1970.

[9] V. Hristidis, S.-C. Chen, T. Li, S. Luis, and Y. Deng. Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 83(10):1701–1714, 2010.

[10] F. Naumann and L. Raschid. Information integration and disaster data management. In *Workshop on Information Integration*, Position Paper, 2006.

[11] M. Osuchowski. Bringing information management practices to natural disaster risk reduction. *The Australian Journal of Emergency Management*, 24(1):53–59, 2009.

[12] B. D. M. Peary, R. Shaw, and Y. Takeuchi. Utilization of social media in the east japan earthquake and tsunami and its effectiveness. *Journal of Natural Disaster Science,*, 34(1), 2012.

[13] M. A. L. R. Perera. Ex ante preparedness for disaster management: Sahana in Sri Lanka. In S. Amin and M. Goldstein, editors, *Data Against Natural Disasters–Establishing Effective Systems for Relief, Recovery, and Reconstruction*. The World Bank, 2008.